**SPECIAL SECTION**

# Testing the effectiveness of the *Developing Inclusive Youth* program: A multisite randomized control trial

**Melanie Killen**[1] | **Amanda R. Burkholder**[2] | **Alexander P. D'Esterre**[1] |
**Riley N. Sims**[1] | **Jacquelyn Glidden**[1] | **Kathryn M. Yee**[1] | **Katherine V. Luken Raz**[1] |
**Laura Elenbaas**[3] | **Michael T. Rizzo**[4] | **Bonnie Woodward**[5] | **Arvid Samuelson**[1] |
**Tracy M. Sweet**[1] | **Laura M. Stapleton**[1]

[1]Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland, USA

[2]Furman University, Greenville, South Carolina, USA

[3]Department of Psychology, University of Rochester, Rochester, New York, USA

[4]University of Illinois, Urbana-Champaign, Illinois, USA

[5]Department of Psychology, University of Maryland, Baltimore, Maryland, USA

**Correspondence**
Melanie Killen, Department of Human Development and Quantitative Methodology, University of Maryland, 3942 Campus Drive, Suite 3304, University of Maryland, College Park, MD 20742, USA.
Email: mkillen@umd.edu

**Abstract**

The *Developing Inclusive Youth* program is a classroom-based, individually administered video tool that depicts peer-based social and racial exclusion, combined with teacher-led discussions. A multisite randomized control trial was implemented with 983 participants (502 females; 58.5% White, 41.5% Ethnic/racial minority; $M_{age} = 9.64$ years) in 48 third-, fourth-, and fifth-grade classrooms across six schools. Children in the program were more likely to view interracial and same-race peer exclusion as wrong, associate positive traits with peers of different racial, ethnic, and gender backgrounds, and report play with peers from diverse backgrounds than were children in the control group. Many approaches are necessary to achieve antiracism in schools. This intervention is one component of this goal for developmental science.

A central and important developmental science question is how to reduce prejudice and enable children to change group norms that promote unfair and inequitable treatment of others. To achieve these goals, it is necessary to examine norms and practices in children's worlds that exclude groups of individuals from having access to resources and opportunities (Jost & Kay, 2010; Kendi, 2016; Killen & Dahl, 2021; Roberts & Rizzo, 2020; Turiel et al., 2016). Recent research and scholarship in sociology and educational theory on anti-racism have focused on how to dismantle racism and other forms of injustice by changing institutional and societal level infrastructure (Bonilla-Silva, 1997, 2015; Kendi, 2016, 2019; Lewis et al., 2019). This includes understanding how anti-racism approaches can be integrated into educational institutions to promote intergroup friendships and reduce prejudice in childhood (Killen & Rutland, 2022) as well as how schools are structured, how teachers are trained, what students learn, and how parents and guardians are involved in the process (Bonilla-Silva, 2015).

---

The title for this Special Section is **Advancing Scholarship on Anti-racism within Developmental Science**, edited by Shauna Cooper, Noelle Hurd, and Aerika Brittian Lloyd.

Working toward social and racial justice for children within schools requires change at multiple levels, from an institutional perspective to one that also focuses on the child. In addition to focusing on a top-down dissemination approach in which the expectation for change lies with teachers and parents as the socializing agents, we propose that addressing social justice also necessitates a child-centered perspective. Thus, the current study takes a child-centered, developmental perspective, focusing on children as agents of change (Killen & Dahl, 2021).

## Social reasoning developmental (SRD) model

The social reasoning developmental (SRD) model theorizes that children are active participants in their world. They evaluate, interpret, and make decisions about how to treat others based on many sources of input, including information from adults and peers (Killen & Rutland, 2011; Rutland & Killen, 2015). Research from the SRD model studies how children and adolescents conceive of fairness, equality, and rights (Ruck et al., 2011; Smetana et al., 2014; Turiel, 2002) in intergroup contexts involving group identity, group norms, and group dynamics (Elenbaas et al., 2020; Nesdale & Lawson, 2011; Rizzo et al., 2018; Rutland et al., 2010; Verkuyten et al., 2019). Concepts of fairness, equality, and rights emerge early in childhood and guide children's actions, but often conflict with competing considerations about group identity and group norms (McGuire & Rutland, 2020).

Extensive research in developmental science has focused on children's evaluations and interpretations of *peer* interactions to study the origins of racism and other forms of prejudice and bias in childhood (Burkholder et al., 2019; Elenbaas & Killen, 2016). We assert that creating a program to address social and racial injustice requires facilitating peer conversations among children about what prejudice looks like in their social environments, why it occurs, and what should be done to create fair and equitable interactions and relationships.

A child-centered approach to development is not a new theoretical viewpoint. Constructivist theories, such as those proposed by Piaget (1932) and Turiel (1983), rejected the view that children are passive agents in their learning and development. In fact, Piaget (1932) documented the important role of peer exchanges in acquiring concepts about justice. Yet, the developmental approach proposed in the current study is novel to the goal of determining how to enable children to change group norms to promote positive social relationships among peers from different backgrounds. Most developmental perspectives for promoting change focus primarily on top-down strategies such as those that train teachers to understand children's social-emotional learning. The current perspective is also distinct from social psychological perspectives on intergroup attitudes that focus

primarily on implicit bias (see Levy et al., 2016, for a review). Research has shown that children's racial biases and other forms of social prejudices are constructed as they engage with their social environments, and evolve as a function of their opportunities for cross-group friendships, along with other factors (Baron, 2015; Brenick et al., 2019; Gaias et al., 2018; Rutland et al., 2010).

Moreover, a child-centered approach to ending prejudice and promoting inclusive, anti-racist group norms in the classroom has rarely been included in school-based programs (Killen & Rutland, 2022; Losinski et al., 2019). Yet, school environments that are unwelcoming, or exclusive create negative consequences for all children, particularly those from groups who are marginalized (Losinski et al., 2019; Okonofau et al., 2016). Thus, it is a missed opportunity not to implement programs designed to address equality, equity, and justice across multiple contexts in elementary schools (Losinski et al., 2019).

*Social exclusion. Interpersonal* approaches to peer rejection focus on personality "deficits" and implement interventions to teach social skills to children identified as bullies or victims. In contrast, interventions on *intergroup* social exclusion focus on changing the group norms that perpetuate exclusionary behavior to maintain the status quo (Hitti et al., 2014; Killen et al., 2013; Mulvey et al., 2016). This strategy provided the basis for the design of the current intervention program, which created opportunities for children to have extensive discussions in the classroom about peer exclusion exchanges. In order to extend this program to addressing multiple forms of prejudice present in childhood, the tool included scenarios focusing on exclusion based on race, ethnicity, gender, and other group memberships.

Furthermore, children take multiple roles in intergroup social exclusion contexts: victims, perpetrators, and resisters. In many cases, children who are victims are those from marginalized groups, often reflect the numeric minority, and lack social status. Children who are the perpetrators are often, but not always, from higher status groups and exclude others to maintain their social power in the peer group. More recently, research has documented children who are the resisters; these are children who reject unfair treatment, challenge stereotypes, and rectify inequalities (Elenbaas et al., 2020; Killen & Dahl, 2021). Yet, group dynamics in childhood are complex and often curtail the rejection of unfair treatment when the perceived cost involves being excluded from the group (Abrams & Rutland, 2011). A central goal of the intervention program was to provide children with the opportunity to talk with each other about their intergroup interactions in a guided context facilitated by the teacher. It is proposed that this approach will aid in understanding how to protect and support the victims of racism and prejudice, reduce negative group norms espoused by perpetrators, and encourage children to challenge unfair treatment (Bonilla-Silva, 2015).

Children, especially children from racial majority status backgrounds, develop prejudicial attitudes as early as preschool and into late childhood (Levy et al., 2016). Additionally, children from different backgrounds increasingly become aware of group dynamics, social inequalities, discrimination, and want to remediate what they perceive as unjust (Conry-Murray & Turiel, 2020; Elenbaas et al., 2020). Thus, by anti-racism in this context, we refer to enabling children to become equipped to recognize the detrimental consequences of prejudicial attitudes and the need for change. We view this as a first step toward creating an anti-racism curriculum designed specifically for children, prior to early adolescence (Killen & Rutland, 2022).

*Consequences of intergroup exclusion.* Addressing and changing prejudicial attitudes and exclusionary behavior is an urgent issue because children who *experience* prejudice and discrimination (e.g., name-calling, bullying, exclusion, relational aggression) are subject to compromised well-being (Neblett et al., 2008; Yip, 2015), stress and anxiety (Fisher et al., 2000; Neblett et al., 2013), sleep disorders (Yip, 2015), and low academic achievement (Alfaro et al., 2006; Benner & Graham, 2007; Chavous et al., 2008). Moreover, individuals who *hold biases* about social groups that restrict their social interactions also experience health-related stress associated with negative intergroup relationships (Levy et al., 2016; Mendes et al., 2007; Pauker et al., 2016). Interventions designed to reduce prejudice have positive attitudinal, health, emotional, and academic outcomes for all children. Currently, there are very few opportunities for children to discuss intergroup social exclusion exchanges during the school day even though such exchanges occur with regularity (Costello & Dillard, 2019). Thus, the program assessed in this current investigation was one that provided multiple opportunities for reflection, discussion, and social exchange about intergroup exclusion (Killen et al., 2013).

## Mechanisms of change

Two mechanisms for reducing prejudice and promoting social and racial justice in children's lives are indirect and direct intergroup contact. Indirect contact refers to children reading about or witnessing a child who shares their same social identity become friends with someone of a different social identity (Johnson & Aboud, 2017; Turner & Cameron, 2016). The current study provided children with indirect contact opportunities featured in an online program in which characters become friends with those from different backgrounds and who challenged inequalities and unfair treatment based on gender, race, ethnicity, and other social memberships (Figure 1). Children watched and engaged with eight different social exclusion and inclusion scenarios that highlighted both experiences of prejudice as well as characters' rejection of prejudice. Specifically, in each vignette, at least one character voiced a reason to exclude someone based on group identity while a different character rejected exclusion and argued for an inclusive approach. Children became friends with those from diverse backgrounds after the exclusionary encounter had been rectified. Thus, witnessing intergroup friendship as well as observing children discuss the unfairness of exclusionary behavior provided children with unique opportunities to reflect on how to reject exclusion and what it means to be inclusive (Gaias et al., 2018; Graham & Echols, 2018).

In addition to experiencing indirect contact, the present intervention also gave children direct intergroup contact experiences. Direct contact refers to positive experiences with peers from different racial and ethnic backgrounds (Brenick et al., 2019; Crystal et al., 2008; Gaias et al., 2018; Tropp et al., 2014). While much of the literature on direct contact has focused on developing intergroup friendships, the aim of this study was to provide an opportunity for all classmates to discuss strategies for rejecting stereotypes and biases within intergroup peer
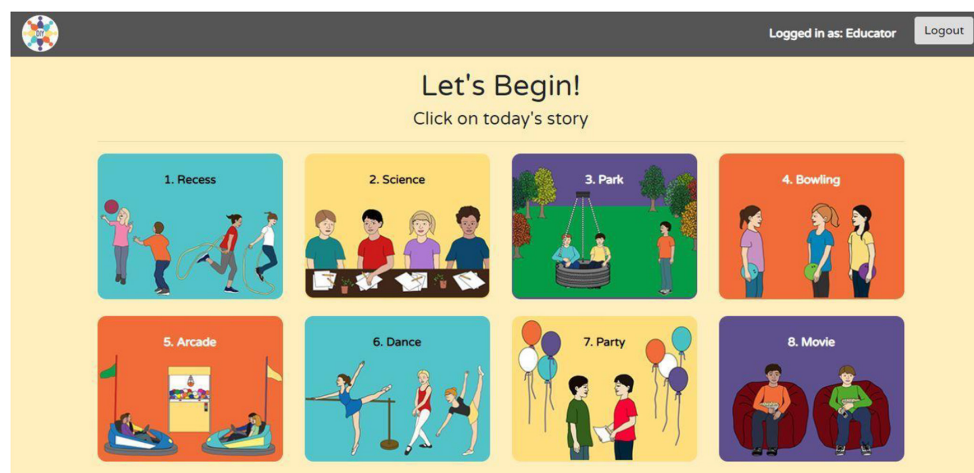


**FIGURE 1** Homepage for the Developing Inclusive Youth tool

interactions in a supportive environment. Specifically, after engaging individually with the curriculum tool, children discussed with their classmates the inclusion/exclusion encounter which included what happened, what they thought about each character's position, along with volunteering whether they had similar experiences to the ones observed. Teachers were trained to be facilitators in this discussion, to create a safe space for children to express their viewpoints, and to encourage children to listen to one another. Teachers prompted children to think about solutions.

## Outcome measures

As social exclusion was a central form of prejudice in the intervention, one of the outcome measures centered on whether participants viewed intergroup social exclusion as wrong and how likely they thought intergroup inclusion occurs (Burkholder et al., 2021; Cooley et al., 2019; Ruck et al., 2011) (see Figure 2). We also measured trait attributions and competency beliefs about diverse peers, as these biases have been theorized to improve as a function of direct and indirect intergroup contact (Tropp et al., 2014). In this study, trait attributions (e.g., friendly/mean, hardworking/lazy, and smart/not smart) were assigned to peers of different racial and gender backgrounds (Liben & Bigler, 2002). In addition, because one of the scenarios centered on inclusion in a science project context, children were assessed on their beliefs about math and science competency for characters from different racial and gender backgrounds (Liben & Bigler, 2002). Finally, as exposure to positive intergroup contact during the intervention program was expected to lead to more positive intergroup contact experiences outside the program, we measured children's self-reported play with peers of different racial and gender backgrounds (modified from Bierman & McCauley, 1987). Thus, the outcome measures reflected the expectations for change in the current study.

## The current study

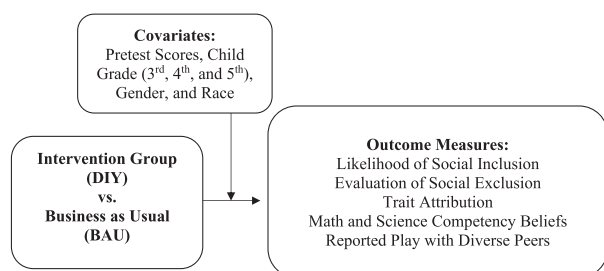The current study was a multisite within-school randomized control trial designed to test the effectiveness of the intervention program *Developing Inclusive Youth* (DIY) relative to a counterfactual (the business-as-usual, BAU, control condition). The DIY program drew on well-established theoretical and empirical lines of research on prejudice and social exclusion in childhood. The program included two components: (1) a web-based curriculum tool; and (2) a teacher-led classroom discussion. Once a week for 8 weeks, children individually logged into an interactive web-based curriculum tool featuring a different target group (Figure 1). The peer scenarios included social encounters between children from different backgrounds in everyday, familiar peer settings (see Table S1). The scenarios depicted in the web-based curriculum tool provided the basis for teacher-led classroom discussions that occurred immediately after the use of the curriculum tool. Importantly, the scenarios were drawn from more than two decades of research on how children evaluate peer social inclusion and exclusion situations that occur in their everyday lives (Killen & Rutland, 2011; Levy et al., 2016; Munirah et al., 2021).

Multiple target groups (e.g., race, ethnicity, gender, wealth status, immigrant status) were included to broaden the recognition of what makes prejudice wrong by exposing children to different experiences and perspectives (Bucchianeri et al., 2016). Children shared group memberships with characters from different backgrounds, increasing the opportunity for all children to relate to forms of social exclusion (Mulvey, 2016). Representing multiple target groups may also alleviate the pressure that individual children may feel when a program focuses only on prejudice against their group, particularly when their group is a numeric minority within the school or classroom.

The goals of the program were to enable children to identify exclusionary, discriminatory, and ostracizing behaviors, what to do when it happens, how to reject these behaviors, and how to work toward changing the norms of the peer culture in ways that directly result in more fair and equal treatment of others (Losinski et al., 2019; Rogers, 2019). Importantly, the program was designed to provide students with the tools and opportunities to talk about solutions for dealing with negative experiences and interactions before they occur, and not in "the heat of the moment." The program did not aim to teach children in-depth content about each identity group, however, since this would require a different type of design which has, to date, focused on adolescent samples (Umaña-Taylor et al., 2018).

*Age and grade sample.* The study focused on elementary school-aged children in third, fourth, and fifth grades, between 8 and 11 years of age. Previous programs have often focused on one age group; including children of multiple grades has the advantage of charting developmental change and targeting the ideal age for intervention. We chose to implement the intervention in elementary schools because these students spend most of the day in their home room, creating an optimal peer group community and a continuity of experience



**FIGURE 2** Study design

for the development of teacher–child relationships. The program addresses children's beliefs before inter-racial friendships decline in middle school (Elenbaas & Killen, 2016; Mulvey, 2016; Turner & Cameron, 2016).

*School composition.* There are a number of factors to consider regarding the racial/ethnic composition of a school when designing studies to change attitudes. For this first test of the program, we identified schools that were racially and ethnically diverse with a White numerical majority (58.5%). Changing prejudicial attitudes and promoting anti-racism orientations is important in schools where there exists a White numerical majority of students. Second, rather than target schools that had a ho-mogeneous White racial composition, we tested the pro-gram in schools with a substantial racial/ethnic minority group of students (41.5%). This composition provided opportunities for children from different backgrounds to voice their interpretations, experiences, and perspectives on social exclusion based on a range of target groups in addition to race (ethnicity, gender, immigrant status, and wealth status), and also created opportunities for direct intergroup contact between classmates. Children learning from their peers and hearing their experiences provides a powerful lever for change. The intervention tool was designed to improve classroom environments for all students by reducing prejudice, increasing positive peer relationships, providing a safe forum to discuss per-sonal experiences of exclusion, and motivating children to identify and address discriminatory attitudes and be-haviors in peer interactions and relationships.

## Hypotheses

There were three central hypotheses. First, we predicted higher positive intergroup attitudes and reported play with diverse peers for children in the DIY (intervention) group than for those in the BAU (control) group, after controlling for initial pretest levels of attitudes and ex-pectations, participant gender, race, and grade (H1).

Second, we predicted that the DIY program would be more effective for children in the fifth grade than for chil-dren in the third and fourth grades (H2). Older children are exposed to more negative outcomes of intergroup exclusion than are younger children; group identity be-comes more salient as children move into higher grades and social exclusion becomes more frequent (Abrams & Rutland, 2011; Mulvey, 2016).

Third, given that racial majority status (White) chil-dren are more likely to display bias and stereotypes than are racial minority status children (Aboud & Brown, 2013; Brown, 2017; Cooley et al., 2019; Dunham et al., 2011; Killen et al., 2007), we predicted that while the DIY program would increase positive attitudes among children of all racial groups, it would produce larger in-creases for racial numeric majority status children than for racial/ethnic numeric minority status children (H3).

# METHOD

## Participants

Participants included $N = 983$ students in third ($n = 323$, females = 172, $M_{age} = 8.64$ years, $SD_{age} = 0.36$), fourth ($n = 337$, females = 176, $M_{age} = 9.65$ years, $SD_{age} = 0.38$), and fifth ($n = 323$, females = 154, $M_{age} = 10.63$ years, $SD_{age} = 0.36$) grades (see Table 1). Participants were from racial and ethnic majority and minority backgrounds (see Table 1). The program was implemented under rou-tine conditions in a large public school district in a major metropolitan area in the Mid-Atlantic region of the United States. The six participating schools had a mean of 8.1% students on Free and Reduced Priced Meals (FARMS) with a range from 5% to 11.4%. This project received approval from the University of Maryland Institutional Review Board Approval #1093717. The data were collected during fall of 2018 and fall of 2019.

In designing the study, we attempted to have a min-imum detectable effect size (MDES) for the overall standardized treatment effect of .31 at a power of .80. Specifically, we determined (using Optimal Design v3.01; Raudenbush et al., 2011), that including six participat-ing schools as blocks, each with six classrooms (i.e., one treatment and one BAU classroom at each grade—third, fourth, and fifth) with 25 students per classroom, result-ing in 900 students total, would yield an MDES of .31. This estimate is based on assumptions that the school site blocking variable explained .40 of the variance of

**TABLE 1** Demographics of students participants

| Student characteristic | Total | BAU | DIY |
|---|---|---|---|
| Grade level | | | |
| 3rd | 32.9% | 33.0% | 32.7% |
| 4th | 34.3% | 34.8% | 33.8% |
| 5th | 32.9% | 32.2% | 33.5% |
| Gender | | | |
| Female | 51.1% | 52.8% | 49.6% |
| Male | 48.6% | 47.0% | 50.0% |
| Not reported | 0.3% | 0.2% | 0.4% |
| Race/ethnicity | | | |
| European American | 58.5% | 62.5% | 55.1% |
| African American | 5.6% | 4.9% | 6.2% |
| Latinx | 4.2% | 3.5% | 4.7% |
| Asian American | 8.3% | 8.2% | 8.5% |
| Multiethnic | 17.5% | 13.7% | 20.7% |
| Other | 0.6% | 0.5% | 0.8% |
| Not Reported | 5.3% | 6.7% | 4.1% |
| Total | 983 | 451 | 532 |

*Note*: Race/ethnicity and gender of the participants was provided by parents in the consent forms. All demographic measures were equivalent at baseline ($ps > .05$).

the outcome measure and classroom-level predictors explained .70 of the variance of the outcomes at the classroom level. Additionally, the ICC value was assumed to be .10 and the standardized treatment effect size variation across sites was assumed to be .01. The minimum detectable effect sizes will be larger for tests involving moderation. The eventual design matched the planned design, except for a lower average number of students per classroom (20.45 students per classroom) and for two schools, four classroom participated at each grade.

After receiving school district approval, invitations were sent to 10 principals regarding participation and six agreed to participate (one declined due to new staff at the school and three cited special programs already being implemented). Written parental consent and verbal child assent were collected prior to the onset of the study. The return rate for participation was high (83.6%), and students without parental consent went to the media center/library to read or do homework during pretest/ posttest assessments and DIY program. The individual child attrition rate was also low (out of the total sample, $n = 54$ were missing: 28 repeated absences, 5 non-English speakers, 6 moved out of the school district, 5 technical issues, 10 other). This resulted in a mean of 20.45 ($SD = 3.89$) participants across the 48 classrooms.

## Design of the study

Within each school and grade level, classrooms were randomly assigned to participate as a DIY (Intervention) or a BAU (Control) classroom. Across six schools, there were 24 DIY classrooms and 24 BAU classrooms, evenly divided by third, fourth, and fifth grades, resulting in eight classes at each grade in each condition. Blocking within school, with randomization at the classroom level, controlled for school-level characteristics. The within-school randomization was preferred over a between-school design to control for demographic differences that exist for schools across the school district and due to the lack of principals interested in serving as a "control" school without the benefit of the program. Children in the BAU control condition were assessed at pretest and posttest with the assessment in the classroom; they did not participate in the weekly DIY program. There were no significant differences for teacher demographics in the DIY treatment and the BAU condition (see Table S2). There were also no significant differences between DIY and BAU groups on outcome variables at pretest (all $p$s > .05) (Table S2).

## Materials and procedure

Students used Chromebooks and headphones provided by the school district for the web-based portion of the intervention. These laptops were used for schoolwork during the day and were familiar to all students. Pretest and posttest data collection efforts took place at Week 1 and Week 10 (Table S3) and were overseen by two trained research assistants who helped the teacher ensure each student successfully logged into the pretest/posttest assessment and answered clarification questions as needed. A research assistant also attended each session for the duration of the DIY program and helped students log in during the web-based tool portion, then sat quietly in the back of the room during completion of the discussion session. Prior to the start of the program implementation, teachers participated in a workshop in which they received materials and training on how to promote discussion in the classroom, create a safe space for discussions, enable students to express their views, and encourage children to engage in conversations (Figures S1 and S2). Teachers were invited to be partners with the university-affiliated research team as part of the program and provided feedback each week regarding comments, reflections, and questions which were discussed by the research team.

## Intervention program: Developing inclusive youth (DIY)

Following pretest data collection, students and teachers in the intervention condition began the DIY program. The eight weekly sessions for the DIY program occurred during a consistent time each week that was identified by the teacher and included two components: a web-based curriculum tool and a teacher-led classroom discussion. Classrooms participated in a lock-step manner, indicating that for all classrooms within each year, data collection began at the end of September and ended the last week of December for a total of 10 weeks; there was one exception whereby one class had to skip a session due to a scheduling conflict.

*Web-based curriculum tool.* The DIY tool included eight scenarios that students viewed in a fixed order, once a week over the course of 8 weeks (Figures 1 and 2; Table S4). The eight scenarios targeted the following social groups: *Recess* (new person at school), *Science* (gender: female), *Park* (race/ethnicity: Latinx), *Bowling* (immigrant status), *Arcade* (wealth status), *Dance* (race/ethnicity: Black), *Party* (race/ethnicity: White), and *Movie* (race/ethnicity: Arab American).

Each portal displayed a short vignette featuring two, three, or four peers. One or two characters discussed excluding a peer from a group activity while another character voiced an inclusive desire. The dialogue included references to stereotypic expectations from characters who wanted to exclude and expressions highlighting commonalities or rejecting exclusive orientations from characters who wanted to include. For example, in the Science scenario, a boy who wanted to exclude a girl from the boys' science project group stated: "Girls aren't

good at science," while his male friend replied: "But my sister is good at science." In a scenario about a ballet group in which a Black girl wants to join, a White girl states to her friend, "Girls like that haven't taken ballet. We want to keep the group as it is," but her White friend tells her "But how do you know she hasn't had lessons if you haven't asked her? Let's see what she can do."

The interactive design of the tool allowed children to watch these indirect intergroup contact experiences and enter prompted responses throughout the scenarios. These prompts included requests to (1) select the feeling states of various characters at key points in the narrative; (2) decide whether the exclusionary statements discussed by the characters were okay or not; (3) make decisions about whether the peers should include or exclude the target child (or activity); and (4) select which reasons reflect their decisions (e.g., stereotypic expectations, moral reasons, practical concerns, group identity).

A unique aspect of the tool is that the story ending depended on participants' decision to "include" or "exclude" the peer. This setup allowed children to witness the direct and immediate consequences of their choice. In most cases, exclusion decisions resulted in a loss of friendship opportunities and sadness displayed by the excluded children and inclusion decisions resulted in friendship and new lessons learned. Importantly, all students watched the opposite outcome after first viewing the one that they chose (after receiving a prompt: "Let's say that the group decided to do X instead…"), such that all participants were able to witness both the benefits of inclusion and the harm of exclusion. A strength of the program from an evaluation perspective was the high fidelity in the administration of the central instrument, the web-based curriculum tool, given that the delivery of the program was the same for all children.

*Teacher-led classroom discussion.* Once all students had individually completed the scenario of the week using the DIY tool, teachers invited the students to sit in a circle on the floor where they participated in the teacher-led discussion. Teachers received training documents and materials that provided reminders and prompts about the content and themes present in the week's vignette (Figures S1 and S2). Teachers were trained to establish a safe space in the classroom, which included agreeing that the discussion must be kept confidential, listening to their classmates without interruptions, and refraining from identifying classmates by names (Figure S1). During the discussion, children were prompted to (1) Make connections between the scenarios and their own experiences; (2) Reflect on how their experiences related to broader themes of inclusivity and anti-prejudice and racism; (3) Reflect on how the story they heard is similar to other weeks' scenarios; (4) Get both sides of the story and discuss why each character made the decisions they did; and (5) Share personal experiences that relate to the week's topic and themes. Teachers thus engaged students in a substantive face-to-face classroom discussion on the topics of inclusion/exclusion and prejudice/bias. One to two research assistants were present to observe each classroom discussion but did not participate or intervene during the session. Afterward, constructive feedback and suggestions for facilitating the discussion were provided to the teacher which reflected the themes in the facilitator guides (Figures S1 and S2). These documents and feedback were derived from critical pedagogy in moral education which encourages teachers to facilitate conversations with children to build mutual respect, equity, and inclusion (Nucci & Ilten-Gee, 2021).

The research assistant also wrote detailed notes regarding children's discussions that pertained to inclusion, exclusion, and personal experiences about exclusion in order to document the types of statements that children exchanged. To assist with interpretations of the findings, the categories that emerged from the observations of the discussions with actual recorded examples (verbatim) are listed in Table S5.

## Measures for the pretest/posttest assessment

*Child demographic variables.* Upon providing their consent, parents of all child participants were given a demographic form. The majority filled out the demographic information, which included students' gender and their race/ethnicity (Table 1). In addition, children's grade level was determined as a function of their classroom. Thus, three variables: *gender*, *grade*, and *race* were included as potential moderators in the models.

*Child outcome variables.* Students completed a 30-min survey-based assessment using Qualtrics, administered at pretest and posttest. Included in this assessment were measurements of (1) children's social reasoning about interracial and same-race peer inclusion and exclusion, (2) trait attributions about race and gender, (3) math and science competency beliefs about race and gender, and (4) reported play with peers of different races and genders. The targets depicted in the measures reflected different racial/ethnic groups as both boys and girls. In terms of reliability, all measures indicated internal consistency with Cronbach's alpha of at least .80 (see Table S6).

*Social reasoning about peer inclusion and exclusion.* Drawn from Cooley et al. (2019), participants were presented with gender-matched illustrations depicting hypothetical contexts of interracial and same-race peer dyads. In each context, participants predicted the likelihood of peer inclusion and evaluated the acceptability or wrongfulness of peer exclusion.

First, participants predicted the likelihood that two characters would decide to include a third character (e.g., "It's Jenny's birthday and she's having a party. She invited all her friends, including her best friend Allison. She can only invite one more person and she's thinking about inviting Rachel, the new kid at school. Allison doesn't think she should invite Rachel. How likely is it

that Jenny will invite Rachel?"). Participants rated their predictions on six-point Likert-type scales ranging from 1 (Really Unlikely) to 6 (Really Likely). Then, participants reported their evaluations of a decision to exclude a peer (e.g., "Let's say Jenny decides not to invite Rachel because she's worried Allison won't like it. How okay or not okay is that?"). Participants responded on six-point Likert-type scales ranging from 1 (*Really Not Okay*) to 6 (*Really Okay*).

Participants responded to these measures for four counterbalanced contexts in which two interracial peer encounters were depicted (White characters excluding a Black peer or Black characters excluding a White peer) and two same-race peer encounters were depicted (White or Black; see Figure S3). Exclusion scores were reversed-coded (so that higher scores indicated exclusion was more wrong). Inclusion and reversed-coded exclusion scores were averaged into a composite based on the racial context of the encounter to create four total outcome variables: Social reasoning about an encounter where White characters excluded a Black peer, an encounter where Black characters excluded a White peer, a same-race Black encounter, and a same-race White encounter.

*Trait attributions for gender and race.* Modified from Liben and Bigler's (2002) gender stereotypes assessment, participants were shown four illustrated drawings: (1) six girls of various races/ethnicities; (2) six boys of various races/ethnicities; (3) six Black children; and (4) six White children (see Figure S4). Participants responded to three prompts per social group (12 prompts total) to determine the extent to which they associated the groups with different traits (smart, friendly, hardworking). For gender, participants were asked: "Do you think these girls/boys are smart or not smart, friendly or mean, hard-working or lazy?" For race, participants were asked: "Do you think kids who look like this are smart or not smart, friendly or mean, hard-working or lazy?" Children were provided with six-point Likert-type scales ranging from 1 (*Really [negative trait]*) to 6 (*Really [positive trait]*). Individual measures were averaged across depicted group membership to create composites, resulting in four trait attribution outcome measures: trait attributions about female characters, trait attributions about male characters, trait attributions about Black characters, and trait attributions about White characters.

*Math and science competency beliefs for gender and race.* Modified from Liben and Bigler (2002), participants were asked to indicate their beliefs regarding math and science skills for five illustrated target groups, which included two gender (female, male) and three racial groups (White, Black, Asian) (Figure S5). Next to the pictures of children, math and science stimuli were depicted as a colorful set of small icons (e.g., calculator, math symbols, test tubes, microscope). For the gender questions, participants were shown silhouettes of four girls or four boys and were asked, "Here are some girls/

boys. How many girls/boys do you think are really good at math and science?" For race questions, participants were shown images of two boys and two girls for each of three racial groups (White, Black, Asian) and were asked, "Here are some kids who look like this. How many kids who look like this are really good at math and science?" Participants responded on five-point scales ranging from 1 (*None*) to 5 (*All*).

*Reported Play with Diverse Peers.* Modified from a task developed by Bierman and McCauley (1987), participants were shown the same illustrated pictures created for the Math and Science Competency Beliefs, without the math and science pictures, for the two gender (female, male) and three racial/ethnic groups (White, Black, Asian). For the gender questions, participants were shown silhouettes of four girls or four boys and were asked, "Here are some girls/boys. How often do you play with girls/boys?" For the race questions, participants were asked: "Here are some kids who look like this. How often do you play with kids who look like this?" (see Figure S6). Responses were recorded on six-point Likert-type scales ranging from 1 (*Never*) to 5 (*All of the Time*).

## Data analytic plan

To determine whether the nested nature of the data required the inclusion of a random intercept and a multilevel framework, Intraclass Correlation Coefficients (ICCs) were calculated, and model comparisons were conducted between models with a random intercept of classroom and without a random intercept for each hypothesized model and for each outcome measure. All models without the random intercept were selected as the better fitting models according to the AIC, and several models could not be fit due to a random intercept variance of 0. Furthermore, the conditional ICCs of the models fit were extremely low, <0.02, and any adjustments to standard errors using the design effect would have been negligible. For clarity, the multiple regression models are reported throughout the manuscript. Moreover, our hypotheses about differences between classrooms pertained to whether they were in the DIY program or BAU control group. We did not have classroom-level moderators of interest (e.g., whether more experienced teachers generated more change than less experienced teachers, or whether more change happened in more diverse classrooms) and given that the ICC was deemed non-problematic, multiple regressions models were the most appropriate.

Analyses for the effectiveness of the treatment and interactions between treatment and student demographic variables utilized a multiple regression framework. To minimize the false discovery rate for multiple comparisons, we performed the Benjamini–Hochberg correction with a false discovery rate of 25% (Benjamini &

Hochberg, 1995). All significant *p* values reported are significant with the Benjamini–Hochberg correction.

While the attrition rate was low, we conducted multiple imputations using linear regression to address missing values (Graham & Hofer, 2000). Specifically, 30 values were imputed for each missing value using linear regression in SPSS; demographic variables (grade, classroom, condition, gender, race, and school) were predictors while pretest and posttest scores were both predictors and imputed values. All analyses used the 30 sets of full data and estimates, and their estimated sampling variances were obtained given the process outlined by Graham and Hofer (2000).

The first hypothesis predicted that treatment would have a significant effect on children's responses to the outcome measures (Table S7). To test this hypothesis, a series of regression models with treatment as a predictor of posttest responses was conducted. In addition, each child's grade, gender, race, and pretest score were included as covariates. Grade was transformed into dummy variables for the model where Grade 4 was coded as 1 if the child was in a fourth-grade classroom and 0 if not, and Grade 5 was coded as 1 if the child was in a fifth-grade classroom and 0 if not. Gender was coded as 1 if the child was female and 0 if the child was male. Due to the proportion of individual racial groups that were the numeric minority in the participating schools, race was coded as 1 if a child was in the racial numerical majority group (White) and 0 if the child belonged to a racial minority group (see Table 1). Finally, treatment was coded as 1 if the child was in the DIY intervention group and 0 if the child was in the BAU control group.

The second hypothesis was intended to determine if students' grade level moderated the effectiveness of the DIY intervention program (Table S8). To that end, a second set of regression models was conducted to determine the significance of an interaction between the condition of the participant and their grade in school, while controlling for all variables that were previously included as covariates.

Finally, the third hypothesis was concerned with moderation of the treatment effect of the DIY program by the race of the student, while controlling for all variables that were previously included as covariates (Table S9).

In addition to these primary hypotheses, we have included results in the supplemental materials for a model testing the moderation of the treatment effect of the DIY program by the gender of the student, while controlling for all variables that were previously included as covariates (Table S10).

## RESULTS

The main effects of treatment and interactions by grade and race are organized by the following outcome variables: social reasoning about peer inclusion and exclusion, trait attributions about race and gender, math and science competency beliefs based on race and gender, and reported play with diverse peers.

## The effect of treatment on social reasoning about peer inclusion and exclusion

Regarding our first hypothesis concerning the overall effectiveness of the DIY program on children's social reasoning about peer inclusion and exclusion (see Table 2), there were significant main effects of treatment for the models testing children's social reasoning about *interracial* peer inclusion and exclusion for scenarios where White characters excluded a Black peer ($t = 6.12$, $p < .001$), and where Black characters excluded a White peer ($t = 5.02$, $p < .001$). Children in the DIY program ($M_{\text{WexB}} = 4.69$, $SE_{\text{BexW}} = 0.05$; $M_{\text{BexW}} = 4.59$, $SE_{\text{BexW}} = 0.05$) had more positive social reasoning (predicted inclusion as more likely and evaluated exclusion as more wrong) than did children in the BAU control group ($M_{\text{WexB}} = 4.31$, $SE_{\text{WexB}} = 0.05$; $M_{\text{BexW}} = 4.26$, $SE_{\text{BexW}} = 0.06$), controlling for all other predictor variables. There were also significant main effects of treatment for the models testing children's social reasoning about *same-race* peer inclusion and exclusion for Black characters ($t = 6.85$, $p < .001$) and for White characters ($t = 7.63$, $p < .001$). Controlling for pretest scores, grade, gender, and race, children in the DIY program ($M_{\text{B}} = 4.81$, $SE_{\text{B}} = 0.05$; $M_{\text{W}} = 4.86$, $SE_{\text{W}} = 0.05$) had more positive social reasoning (predicted inclusion as more likely and evaluated exclusion as more wrong) than did children in the BAU control group ($M_{\text{B}} = 4.32$, $SE_{\text{B}} = 0.05$; $M_{\text{W}} = 4.40$, $SE_{\text{W}} = 0.05$).

For our second hypothesis concerning the moderating effect of grade on effectiveness of the DIY program (Tables 3, S7, and S11), we found significant interactions of treatment by fifth grade for children's *social reasoning about interracial encounters when White characters exclude a Black peer* ($t = -3.16$, $p = .002$) and *Black characters exclude a White peer* ($t = -3.12$, $p = .002$) as well as for *same-race White encounters* ($t = -3.65$, $p < .001$) and *same-race Black encounters* ($t = -4.17$, $p < .001$). Contrary to our third hypothesis, we did not find that race significantly moderated the effect of treatment (Tables 2 and S12).

Thus, relative to children in the BAU condition, children who participated in the DIY program were more likely to expect inclusion to occur and negatively evaluate exclusion in both interracial and same-race peer encounters, and the effects of treatment on these evaluations were moderated by grade. Specifically, within the DIY condition, children in grade 3 significantly increased their social reasoning about inclusion and exclusion more than did children in grade 5.

**TABLE 2** Treatment and interaction effects for social reasoning about interracial and same-race peer inclusion and exclusion

| Composition of encounter | Interracial: White peers exclude black peer | | | Interracial: Black peers exclude white peer | | | Same race: White peers | | | Same race: Black peers | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | β (SE) | B | CI | β (SE) | B | CI | β (SE) | B | CI | β (SE) | B | CI |
| **Main effect** | | | | | | | | | | | | |
| Treatment | .38*** (0.06) | 0.19 | 0.26, 0.50 | .33*** (0.07) | 0.16 | 0.20, 0.46 | .49*** (0.06) | 0.24 | 0.37, 0.61 | 0.46*** (0.06) | 0.24 | .34, .58 |
| **Interactions** | | | | | | | | | | | | |
| Treatment by Racial Majority | 0.17 (0.13) | 0.17 | −0.08, 0.42 | 0.14 (0.14) | 0.14 | −0.12, 0.41 | 0.13 (0.13) | 0.13 | −0.12, 0.38 | 0.19 (0.13) | 0.19 | −0.06, 0.43 |
| Treatment by Grade 4 | 0.04 (0.15) | 0.04 | −0.25, 0.32 | 0.07 (0.16) | 0.07 | −0.24, 0.38 | 0.17 (0.15) | 0.17 | −0.12, 0.50 | 0.10 (0.15) | 0.10 | −0.19, 0.39 |
| Treatment by Grade 5 | −0.47** (0.15) | −0.47 | −0.76, −0.18 | −0.50** (0.16) | −0.48 | −0.81, −0.19 | −0.62*** (0.15) | −0.62 | −0.91, −0.33 | −0.54*** (0.15) | −0.55 | −0.83, −0.25 |

*Notes*: Table reports unstandardized regression coefficients (β) with standard error (SE) estimates, standardized regression coefficients (B) as a measure of effect size, and 95% confidence intervals (CI) of the unstandardized regression coefficients. The "Treatment" row reports the main effect of differences between the DIY program group and the BAU control group on children's social reasoning about interracial and same-race peer inclusion and exclusion from the main models testing the effectiveness of the DIY program. The "Treatment by Racial Majority," "Treatment by Grade 4," and "Treatment by Grade 5" rows report interaction effects from the follow-up interaction models. Full models can be found in the supplemental materials (Tables S7–S9).
Significant values are denoted with *$p$<.05; **$p$<.01, ***$p$<.001.

**TABLE 3** Treatment and interaction effects of trait attributions for target groups based on gender and race

| Target groups | Female | | | Male | | | White | | | Black | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | β (SE) | B | CI | β (SE) | B | CI | β (SE) | B | CI | β (SE) | B | CI |
| **Main effect** | | | | | | | | | | | | |
| Treatment | 0.13* (0.06) | 0.14 | 0.03, 0.24 | 0.25*** (0.06) | 0.24 | 0.13, 0.36 | 0.18** (0.05) | 0.19 | 0.07, 0.28 | 0.12* (0.05) | 0.13 | 0.02, 0.21 |
| **Interactions** | | | | | | | | | | | | |
| Treatment by Racial Majority | 0.14 (0.11) | 0.14 | −0.09, 0.36 | 0.07 (0.12) | 0.07 | −0.17, 0.31 | 0.15 (0.11) | 0.16 | −0.07, 0.37 | 0.10 (0.10) | 0.11 | −0.11, 0.30 |
| Treatment by Grade 4 | 0.01 (0.13) | 0.01 | −0.25, 0.27 | −0.09 (0.14) | −0.08 | −0.37, 0.19 | −0.09 (0.13) | −0.10 | −0.35, 0.17 | 0.01 (0.12) | 0.01 | −0.23, 0.24 |
| Treatment by Grade 5 | −0.16 (0.14) | −0.16 | −0.42, 0.11 | −0.14 (0.14) | −0.13 | −0.42, 0.14 | 0.02 (0.13) | 0.02 | −0.24, 0.28 | −0.14 (0.12) | −0.15 | −0.10, 0.37 |

*Notes*: Table reports unstandardized regression coefficients (β) with standard error (SE) estimates, standardized regression coefficients. (B) as a measure of effect size, and 95% confidence intervals (CI) of the unstandardized regression coefficients. The "Treatment" row reports the main effect of differences between the DIY program group and the BAU control group on children's endorsement of trait attributions from the main models testing the effectiveness of the DIY program. The "Treatment by Racial Majority," "Treatment by Grade 4," and "Treatment by Grade 5" rows report interaction effects from the follow-up interaction models. Full models can be found in the supplemental materials (Tables S7–S9).
Significant values are denoted with *$p$<.05; **$p$<.01, ***$p$<.001.

## The effect of treatment on trait attributions for gender and race

Next, we tested the effect of treatment on children's trait attributions for gender and race (Table 3). Regarding attributions for gender, there were significant main effects of treatment for the models testing children's predictions of the *trait attributions for females* ($t = 2.42$, $p = .016$) and *trait attributions for males* ($t = 4.25$, $p < .001$). Children in the DIY program ($M_F = 4.92$, $SE_F = 0.04$; $M_M = 4.81$, $SE_M = 0.04$) expected both gender groups to be smarter, more hard-working, and friendlier than did their BAU counterparts ($M_F = 4.78$, $SE_F = 0.05$; $M_M = 4.56$, $SE_M = 0.05$), controlling for all other predictor variables. Regarding race, there were also significant main effects of treatment on children's predictions of *trait attributions for White characters* ($t = 3.30$, $p = .001$) and *trait attributions for Black characters* ($t = 2.29$, $p = .022$). Children in the DIY program ($M_W = 4.93$, $SE_W = 0.04$; $M_B = 5.01$, $SE_B = 0.04$) reported higher positive trait attributions of both racial groups than did children in the BAU condition ($M_W = 4.75$, $SE_W = 0.05$; $M_B = 4.89$, $SE_B = 0.04$), controlling for pretest scores, grade, gender, and participant race. There were no significant interactions between treatment and grade or race (Tables 3, S8, and S13). Thus, overall, children in the DIY program reported higher positive trait attributions for female, male, White, and Black characters, compared to participants in the BAU control condition.

## The effect of treatment on math and science competency beliefs

As reported in Table 4 and corresponding to our first hypothesis, we tested the effect of the DIY program on children's math and science competency beliefs. There was a significant effect of treatment on children's predictions of *math and science competency beliefs about Black students* ($t = 2.49$, $p = .013$). Children in the DIY program ($M_B = 3.66$, $SE_B = 0.04$) rated Black students as better at math and science that did those in the BAU condition ($M_B = 3.53$, $SE_B = 0.05$). There was also a significant effect of treatment on children's predictions of *math and science competency beliefs about White students* ($t = 2.21$, $p = .027$). Children in the DIY program ($M_W = 3.67$, $SE_W = 0.04$) rated White students as better at math and science that did those in the BAU condition ($M_W = 3.56$, $SE_W = 0.04$), controlling for all other predictor variables. There was not a significant main effect of treatment for *math and science competency beliefs about male students*, and only marginal main effects of treatment for *math and science competency beliefs about Asian students* and *female students*.

For our second hypothesis concerning the moderating effect of grade on the effectiveness of the DIY program (Tables 4, S7, and S14), there were significant

**TABLE 4** Treatment and interaction effects of math and science competency beliefs based on gender and race

| Target groups | Female | | | Male | | | White | | | Black | | | Asian | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ (SE) | B | CI | $\beta$ (SE) | B | CI | $\beta$ (SE) | B | CI | $\beta$ (SE) | B | CI | $\beta$ (SE) | B | CI |
| **Main effect** | | | | | | | | | | | | | | | |
| Treatment | $0.10^\dagger$ (0.05) | 0.06 | −0.00, 0.20 | 0.02 (0.05) | 0.01 | −0.07, 0.12 | 0.11* (0.05) | 0.07 | 0.01, 0.21 | 0.13* (0.05) | 0.08 | 0.03, 0.24 | $0.09^\dagger$ (0.05) | 0.06 | −0.01, 0.20 |
| **Interactions** | | | | | | | | | | | | | | | |
| Treatment by Racial Majority | 0.01 (0.10) | 0.01 | −0.19, 0.21 | −0.00 (0.10) | −0.00 | −0.19, 0.19 | 0.17 (0.10) | 0.22 | −0.03, 0.37 | 0.01 (0.11) | 0.02 | −0.20, 0.23 | 0.07 (0.11) | 0.08 | −0.15, 0.28 |
| Treatment by Grade 4 | −0.01 (0.12) | −0.01 | −0.25, 0.23 | −0.10 (0.12) | −0.14 | −0.33, 0.12 | 0.05 (0.12) | 0.06 | −0.19, 0.29 | 0.12 (0.13) | 0.15 | −0.13, 0.37 | −0.01 (0.13) | −0.02 | −0.27, 0.24 |
| Treatment by Grade 5 | −0.16 (0.12) | −0.20 | −0.40, 0.08 | −0.16 (0.12) | −0.22 | −0.39, 0.06 | −0.14 (0.12) | −0.17 | −0.37, 0.10 | −0.27* (0.13) | −0.32 | −0.52, −0.01 | −0.30* (0.13) | −0.36 | −0.55, −0.04 |

*Notes:* Table reports unstandardized regression coefficients ($\beta$) with standard error (SE) estimates, standardized regression coefficients, (B) as a measure of effect size, and 95% confidence intervals (CI) of the unstandardized regression coefficients. The "Treatment" row reports the main effect of differences between the DIY program group and the BAU control group on children's math and science competency beliefs from the main models testing the effectiveness of the DIY program. The "Treatment by Racial Majority," "Treatment by Grade 4," and "Treatment by Grade 5" rows report interaction effects from the follow-up interaction models. Full models can be found in the supplemental materials (Tables S7–S9).
Significant values are denoted with $^\dagger p < .10$; $^* p < .05$.

interactions of the treatment by fifth grade for children's *math and science competency beliefs for Black Students* ($t = -2.07$, $p = .039$) and *math and science competency belief for Asian Students* ($t = -2.28$, $p = .023$). Thus, third graders in the DIY program were more positive about predicted Black and Asian math and science competency than were fifth graders in the DIY program. Contrary to our third hypothesis, there were no significant moderating effects of race on the effect of the DIY program for math and science competency beliefs (Tables 4 and S15).

Thus, children in the DIY program reported more positive math and science competency beliefs about Black, and White students, but not about female, male, or Asian students, than did children in the BAU control condition. There were also significant interactions between treatment and fifth grade, indicating that children in third grade significantly changed their beliefs about Black and Asian peers more than did children in fifth grade.

## The effect of treatment on reported play with diverse peers

As reported in Table 5, for the main effect of treatment on reported play, there was a significant main effect of treatment on children's *reported play with male peers* ($t = 2.08$, $p = .038$). Children in the DIY program ($M = 3.52$, $SE = 0.04$) reported a higher frequency of play with male peers than children in the BAU condition ($M = 3.41$, $SE = 0.05$). There were no significant main effects of treatment for *reported play with female peers* or *reported play with Asian peers*, and only marginal main effects of treatment for *reported play with White peers* and *reported play with Black peers*.

For our second hypothesis regarding the moderating effect of grade on treatment (Tables 5, S8, and S16), there were significant interactions of the treatment by fifth grade for children's *reported play with male peers* ($t = -2.58$, $p = .010$), *reported play with Black peers* ($t = -2.54$, $p = .011$), and *reported play with Asian peers* ($t = -2.15$, $p = .031$). There was also a significant interaction of the treatment by *fourth grade* for children's *reported play with male peers* ($t = -1.97$, $p = .048$). Contrary to our third hypothesis, there were no significant interactions between race and treatment (Tables 5 and S17).

These findings reveal that, overall, children in the DIY program reported more play with male peers than did their BAU counterparts. Additionally, within the DIY condition, children in grade 3 significantly increased their reported play with male peers, Black peers, and Asian peers more than did children in grade 5. Similarly, children in grade 3 also increased their reported play with male peers more than did children in grade 4.

**TABLE 5** Treatment and interaction effects of reported play with diverse peers based on gender and race

| Target groups | Female | | Male | | White | | Black | | Asian | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ (SE) | B CI | $\beta$ (SE) | B CI | $\beta$ (SE) | B CI | $\beta$ (SE) | B CI | $\beta$ (SE) | B CI |
| **Main effect** | | | | | | | | | | |
| Treatment | −0.33 (0.05) | −0.01 −0.14, 0.07 | 0.11*** (0.06) | 0.05 0.42, 0.53 | 0.10$^\dagger$ (0.06) | 0.05 −0.18, 0.22 | 0.09$^\dagger$ (0.06) | 0.05 −0.17, 0.21 | 0.01 (0.06) | 0.00 −0.11, 0.13 |
| **Interactions** | | | | | | | | | | |
| Treatment by Racial Majority | 0.16 (0.11) | 0.13 −0.06, 0.38 | −0.05 (0.11) | −0.04 −0.27, 0.18 | −0.13 (0.13) | 0.13 −0.38, 0.12 | 0.03 (0.12) | 0.03 −0.20, 0.26 | 0.00 (0.13) | 0.00 −0.25, 0.26 |
| Treatment by Grade 4 | 0.12 (0.13) | 0.10 −0.14, 0.38 | 0.26* (0.13) | 0.21 0.00, 0.52 | −0.04 (0.15) | −0.04 −0.04, 0.56 | 0.16 (0.14) | 0.17 −0.11, 0.43 | −0.01 (0.15) | −0.00 −0.30, 0.29 |
| Treatment by Grade 5 | 0.00 (0.13) | 0.00 −0.26, 0.27 | −0.34* (0.13) | −0.26 −0.61, −0.08 | −0.26 (0.15) | −0.26 −0.56, 0.04 | −0.35* (0.14) | −0.37 −0.63, −0.08 | −0.33* (0.15) | −0.30 −0.63, −0.03 |

*Notes:* Table reports unstandardized regression coefficients ($\beta$) with standard error (SE) estimates, standardized regression coefficients. (B) as a measure of effect size, and 95% confidence intervals (CI) of the unstandardized regression coefficients. The "Treatment" row reports the main effect of differences between the DIY program group and the BAU control group on children's reported frequency of play with peers of difference races and genders from the main models testing the effectiveness of the DIY program. The "Treatment by Racial Majority," "Treatment by Grade 4," and "Treatment by Grade 5" rows report interaction effects from the follow-up interaction models. Full models can be found in the supplemental materials (Tables S7–S9).

Significant values are denoted with $^\dagger p < .10$; $*p < .05$; $***p < .001$.

# DISCUSSION

This study was designed to address social and racial biases from a developmental science perspective, one that takes a child-centered approach by enabling children to become agents of change. To achieve these goals, children responded to an interactive web-based curriculum tool, *Developing Inclusive Youth*, that portrayed intergroup peer inclusion and exclusion encounters. Using the tool prompted individual reflection and decision-making and was paired with a teacher-led classroom discussion immediately following the online program. The 8-week program and accompanying discussions were focused on observed intergroup peer scenarios as well as personal experiences of intergroup exclusion at recess, in the park, at school, and at home. The intervention served as a catalyst to have conversations with the expectation that these experiences over 2 months could change attitudes and group norms in the classroom regarding the fair and just treatment of others.

The novel findings were that the *Developing Inclusive Youth* (DIY) program was effective for changing attitudes for children in third, fourth, and fifth grades who received the intervention. This program is one of the first of its kind to directly attempt to change children's prejudice and bias as well as prompt children to challenge unfair treatment by seeking solutions to students' experiences of bias at school, a fundamental goal of an anti-racist curriculum. Programs such as DIY may help to reduce prejudice and promote anti-racism and social justice among children and within schools.

## Why schools are an important context for promoting anti-racism

Schools that are unwelcoming, exclusive, and intolerant have negative consequences for children's mental health, social relationships, motivation to attend school, and academic achievement (Rivas-Drake et al., 2014; Suárez-Orozco et al., 2015). Few school programs focus on a group normative approach for improving peer relationships and the classroom environment (Killen & Rutland, 2022). A focus on changing individual children's social skills to be less aggressive, for example, misses an important opportunity to focus on group-level biases that underlie prejudicial attitudes espoused by children and adolescents. Rather than focus on improving individual children's social skills for reading social cues, DIY focuses on changing group norms in the classroom that reflect societal biases based on group membership (such as race, ethnicity, and gender). Exclusivity and biases about others are often promoted to maintain power structures and social status hierarchies (Dovidio & Gaertner, 2008). Group-level expectations stemming from societal norms are picked up by children, sometimes explicitly or implicitly, and used to exclude others

from social groups and opportunities in the peer world. DIY was uniquely designed to enable children to reflect on peer biases and to discuss with their classmates about peer encounters as well as their own experiences of exclusion based on race, ethnicity, gender, and other forms of a group membership.

## Peer exchanges are effective for promoting change

The social reasoning developmental (SRD) model theorizes that providing children with the opportunity to have conversations with one another about group norms, prejudice and biases enables children to reflect on what makes biases wrong, and consider solutions for change. This premise is based on multiple lines of research: (1) moral reasoning about unfair treatment and social inequalities; (2) the role of children as agents of change; and (3) the power of peer discussions for reducing prejudice and other forms of bias. Research on children's moral reasoning has revealed that children care deeply about the fair and just treatment of others (Smetana et al., 2014; Turiel, 1983). Yet, recognizing that prejudicial behavior is a moral transgression similar to an act of physical harm is not often obvious to children. This is due to the salience of group identity and societal norms that support social status hierarchies. Thus, the DIY program was designed to encourage children to recognize situations in which discriminatory and biased behavior occurs, a central component of anti-racism.

The role of peer interaction has been shown to facilitate change in many domains of children's lives in developmental science including peer discussions that promote concepts of justice (Turiel, 1983), and reduce prejudicial attitudes and biases (Tropp et al., 2014). Intergroup contact research has proposed that the conditions that make intergroup exchanges effective for reducing prejudice include establishing and promoting common goals, equal status, authority support, and cross-group friendships. While intergroup contact was a foundation for the current program, the goal was to take it one step further in order to incorporate an anti-racism perspective. This required children to not only form intergroup friendships but also to detect bias in peer exchanges and create solutions for change.

Taking a child-centered approach to anti-racism means creating the conditions where children can discuss issues of prejudice and bias in a safe context. An advantage of this program was that children were not discussing exchanges in the "heat of the moment" (or shortly thereafter) but as a classroom activity prompted by the program and facilitated by the teacher. Extensive research has demonstrated that teachers have biases and stereotypes about their students' abilities and competence (Okonofau et al., 2016). Thus, a program that is created to ask teachers to teach about bias and prejudice

requires training sessions that will "undo" assumptions held by teachers. Instead, the current program did not ask teachers to teach a lesson about prejudice but to serve as a facilitator of children's discussions, and to learn about their own students' experiences. The web-based curriculum tool provided the lesson in terms of information that children reflected on and discussed in the classroom.

Even so, it needs to be acknowledged that teachers are often limited in their own awareness about implicit and explicit biases. This lack of awareness can affect their ability to facilitate the conversations from an anti-racist perspective that was part of this intervention. Thus, the DIY program did not avoid teacher bias completely given that the teachers lead the classroom discussions following the delivery of the DIY web-based curriculum tool. Strategies were in place, however, to provide teachers with weekly feedback to deliver the program in a way that was consistent with the goals of the program.

## Measuring change as a result of participating in the DIY program

Change was measured with a survey for participants that was administered before and after participation in the study. Assessments were selected that reflected the theoretical goals of the study and standard assessments of prejudice and bias in the literature. An extensive body of research has documented how children evaluate intergroup peer inclusion and exclusion (Burkholder et al., 2019; Mulvey, 2016). Children who participated in the program had a greater recognition of the wrongfulness of interracial and same-race exclusion and thought there was a greater likelihood that social inclusion would occur. That children were more likely to view interracial as well as same-race exclusion as wrong after participating in the program provides support for designing programs that explicitly target intergroup social exclusion peer encounters. While previous research has indicated that White children may be most likely to prefer same-race inclusion to interracial inclusion (Cooley et al., 2019), the present intervention did not differentially impact White versus racially minority participants' interracial and same-race inclusion and exclusion judgments. This is contrary to our original expectation that White participants might benefit more from the DIY program in this regard, as previous research has indicated more "room for improvement." As we detail below in the limitations, future research needs to examine this issue more closely.

A second set of findings was that children who participated in the DIY program assigned more positive traits (such as friendly, hard-working, and smart) toward female, male, Black, and White peers. These findings have implications for the effectiveness of the DIY program for reducing prejudice, as previous research suggests negative trait attributions based on group membership are difficult to change (Baron, 2015). Moreover, when children discover that some of their peers view their group as lazy, mean, or not smart this creates anxiety, depression, and a low motivation to attend school (Rivas-Drake et al., 2014). As these types of trait attributions exist by the elementary school years, interventions such as the DIY program are necessary for changing these attitudes to reduce prejudice and impact change in childhood.

Children were also more likely to attribute positive math and science competency beliefs (smart at math and science) to White and Black characters; younger children's attitudes toward Black and Asian characters became more positive than older children's. Extensive research has shown that adolescents from all backgrounds hold traditional stereotypes that White and Asian students are better at math and science than are Black and Latinx students (Skinner et al., 2021). Most research reports that these stereotypes appear during middle school and are much less prevalent during childhood. Thus, the findings that this program increased positive math and science competency beliefs for all ages in this study and that it improved third-grade children's beliefs for Black and Asian characters more than for older children provide further support for the effectiveness of a child-centered intervention to facilitate change prior to adolescence.

Finally, younger children were more likely to report play with Black and Asian peers than were older children as a function of being in the program. This finding reveals that starting these programs early with children as young as 8 and 9 years of age is important. Not only did younger children's desire to play with diverse peers increase but it increased for two groups that have experienced intergroup social exclusion more than for other groups (Black and Asian peers). Classroom discussions and reflections about social exclusion scenarios had a positive effect on children's reported play choices. As has been demonstrated in the literature (Graham et al., 2014; Tropp et al., 2014), children's intergroup interactions help to increase their sense of safety and support as well as reduce bias. Thus, playing with peers from different backgrounds can provide a means for addressing social and racial biases.

Contrary to expectations, change was more pervasive for children in third grade than for those in fifth grade. It was initially proposed that older children would be more likely to change than would younger children. Perhaps, younger children had more to learn than did older children regarding the implications of being exclusive toward others; the DIY experience gave them the opportunity to understand why it is unfair to act exclusively toward their peers. To this point, Nesdale and Lawson (2011) found age-related changes from 7 to 10 years regarding distinctions between exclusive and inclusive peer norms. In their study, younger children failed to differentiate between inclusive or exclusive norms articulated by their peer group. In contrast, older children were more likely

to react negatively to an exclusive in-group norm than were younger children. More research should examine age-related patterns for change with this type of intervention program. Furthermore, the race/ethnicity of the participants did not moderate the effectiveness of the DIY program for changing attitudes. We believe this finding was related to the school composition, which we discuss below in the section on limitations.

Overall, the findings provide a first step for creating a school-based curriculum program that incorporates an anti-racism agenda. This is an important step given that few prejudice programs have been systematically and empirically tested for their effectiveness, and particularly using a randomized control trial. There remain unanswered questions that require further analyses, new versions of the program, and applications to new school compositions to fully address the goals of anti-racism. These will be discussed followed by more general recommendations.

## Limitations and recommendations

*School composition*. For this project, we targeted schools whose student population was 58.5% White numeric majority and 41.5% ethnic and racial minority. The intention was to target the majority group that often perpetuates bias, similar to studies that have focused on White parents and the extent to which their biases can be changed (Abaied & Perry, 2021; Pahlke et al., 2012; Perry et al., 2019). Anti-racism theory discusses the need to move the burden for change to those who have the power, status, and prestige (Kendi, 2019). Rather than obtaining a critical mass of one minority group, however, our sample reflected a diversity of racial/ethnic minority groups and did not provide a large enough sample for analyses of a specific racial/ethnic group due to the low proportions for each group. The proportions for the racial/ethnic minority participants were distributed across four groups rather than one or two groups. Thus, the school compositions did not provide an opportunity to analyze the effects of the program for each racial/ethnic numerical minority group.

Future research needs to examine how children from different racial/ethnic backgrounds respond to the program and whether there are interracial or interethnic differences regarding the effectiveness of the program. This information would be important for learning how to modify the program to best serve children from historically underrepresented backgrounds (Juvonen et al., 2018). Graham and colleagues (Graham et al., 2014) have utilized Simpson's (1949) diversity index to examine how different types of diversity compositions within schools relate to racial/ethnic minority students' well-being, which would be fruitful to apply to intervention studies aimed at improving racial/ethnic school belonging and experiences of inclusion. Implementing this program in schools powered to detect how different racial and ethnic groups benefit from the program is necessary

as a next step. Further, the study was limited in that non-binary representation was not feasible given that the participating school district did not record this information. Future research could include this category as an option for children to indicate when identifying their gender.

*Teacher perspectives*. We conducted focus groups with teachers during a pilot study to gain information for designing this program. As well, we solicited teachers' input after the first implementation of the program during the pilot test. These steps provided essential input from educators regarding the design of the program. One limitation was that we were not able to collect quantitative data on how teachers implemented the program, nor the extent to which teaching styles and relationships with their students contributed to the program's effectiveness. Conducting focus groups with the teachers who participated in the program revealed that teachers learned from their students ("I heard things that my students experienced that I never knew about"). Further, some teachers recognized that they had not discussed the topics in the program with their students in the past. Thus, more detailed surveys, assessments, and observations are necessary to understand the teacher's role and the benefits of participating in the program. In addition, it would be instructive to collect data on teacher's attitudes about biases including their comfort level with talking about race/ethnicity and other forms of bias in the classroom, their strategies for addressing biases, and their views about whether their students' experience prejudice and bias (Juvonen et al., 2019). Future research could also measure what teachers learned from the experience of being a facilitator, and how this experience might change their attitudes.

*Classroom discussions*. We were not able to audio record the classroom discussions due to school district policy. Instead, we hand-transcribed a selection of the conversations (Table S5), but due to the fast pace of the conversations and the lack of audio recording we were not able to systematically capture the bulk of the qualitative data. Thus, a limitation was that we only recorded a small proportion of the conversation data. The next step for future research would be to collect and analyze audio transcribed recordings of the conversations for systematic documentation of the exchanges and to demonstrate which children by age, gender, and racial/ethnic backgrounds made different types of statements, along with analyses documenting the follow-up responses.

*Forms of intergroup contact*. Furthermore, expanding the intergroup contact measures to reflect different types of contact would be fruitful. The current outcome measures focused on attitudes about gender and some race/ethnicity groups (Black, Asian, and White). Yet, in the DIY program, the scenarios that children watched, responded to, and discussed reflected a broader range of racial and ethnic groups (such as Latinx), and also included exclusion based on immigrant and wealth status. Given that immigrant (from a different country) and wealth (high, low) status is often confounded with race, ethnicity, and gender, including such measures would provide a fuller

picture of the contexts in which children have the capacities to change their attitudes (Elenbaas, 2019). Future research should include outcome measures that cover multiple target groups featured in the program.

In fact, most children hold a myriad of identities (some marginalized and some privileged). The current intervention acknowledges intersectionality (interlocking systems of oppression) and the need for all children to learn to be allies regarding the goals of anti-racism. We recognize that the current intervention may be primarily appropriate for reducing the perpetuation of prejudice against marginalized identity groups and increasing intergroup friendships which is a different aim from enhancing group identities for children from marginalized identities. As an example, *The Identity Project* focuses on enhancing adolescents' identities (Umaña-Taylor et al., 2018). The current program could be implemented in conjunction with other programs modified from adolescent studies that build critical consciousness (Diemer et al., 2020). Most of the research has focused on adolescent populations. Given that identities emerge during childhood, building critical consciousness and strengthening racial/ethnic identity could begin prior to adolescence. We view the current program as important for all children with the goal of potentially advancing equity and justice more centrally rather than for targeting specific racial/ethnic minority students in elementary school contexts.

We also recommend that future intervention studies explore whether children are at different starting points on the outcome measures for the different racial and gender groups. While our analyses indicated no significant differences on the pretest levels for the program and control groups, future research could report on the pretest data only to reveal grade-, gender-, and racial/ethnic patterns and differences.

At a broad level, an anti-racism curriculum program has many goals. These include addressing structural inequities and inequalities (contemporary and historic), engaging in discussions of power, privilege, and status as well as understanding how others experience intergroup social exclusion (Rogers et al., 2015; Rouland et al., 2013). It also involves helping students to act as agents of change for promoting the fair and equitable treatment of others (Elenbaas et al., 2020; Killen & Dahl, 2021; Killen & Rutland, 2022). Providing opportunities to explore one's racial and ethnic identity is a central goal as well (Abaied & Perry, 2021; Bonilla-Silva, 2015; Hurd et al., 2021; Rivas-Drake et al., 2014; Umaña-Taylor et al., 2018). This program did not address all of these goals. The DIY intervention was designed to help children to reflect, judge, and discuss issues about exclusion, bias, and prejudice, along with what constitutes fair and equitable treatment of one another in their everyday social interactions and encounters. Future interventions that incorporate multiple components of anti-racism theory and research into classroom curricula have the potential to create inclusive classrooms that foster a sense of belonging and academic achievement for all children.

## CONCLUSION

The DIY program aimed to engage children to take an active approach in reducing outgroup bias and discrimination in an educational context. Children construct notions about group identity and ingroup preferences, acquire biases based on peer interactions as well as authority-based and societal messages, and develop notions of fairness, equality, and rights (Burkholder et al., 2019; Elenbaas et al., 2020; Mulvey, 2016; Rizzo et al., 2021). The findings of this study suggest that prejudice reduction interventions may be effective at reducing bias and discriminatory behavior, particularly with younger children. This intervention, while valuable, should be combined with other approaches that explicitly focus on addressing racism by addressing larger societal issues of power, privilege, and oppression. Ultimately, it will take a multitude of approaches and efforts to succeed in creating anti-racist schools, which will promote healthy child development.

## ORCID

*Melanie Killen* https://orcid.org/0000-0002-6392-9373
*Amanda R. Burkholder* https://orcid.org/0000-0002-4667-1681
*Alexander P. D'Esterre* https://orcid.org/0000-0002-5722-6719
*Riley N. Sims* https://orcid.org/0000-0003-1619-0219
*Jacquelyn Glidden* https://orcid.org/0000-0002-5931-1116
*Kathryn M. Yee* https://orcid.org/0000-0002-6227-1223
*Katherine V. Luken Raz* https://orcid.org/0000-0003-1666-5704
*Laura Elenbaas* https://orcid.org/0000-0002-4921-251X
*Michael T. Rizzo* https://orcid.org/0000-0002-2446-8340

# REFERENCES

Abaied, J. L., & Perry, S. P. (2021). Socialization of racial ideology by White parents. *Cultural Diversity & Ethnic Minority Psychology*, *27*(3), 431–440. https://doi.org/10.1037/cdp0000454

Aboud, F. E., & Brown, C. S. (2013). Positive and negative intergroup contact among children and its effect on attitudes. In G. Hodson & M. R. C. Hewstone (Eds.), *Advances in intergroup contact* (pp. 176–199). Psychology Press. https://doi.org/10.4324/9780203095461

Abrams, D., & Rutland, A. (2011). Children's understanding of deviance and group dynamics: The development of subjective group dynamics. In J. Jetten & M. J. Hornsey (Eds.), *Rebels in groups: Dissent, deviance, difference and defiance* (pp. 135–157). Wiley Blackwell. https://doi.org/10.1002/9781444390841

Alfaro, E. C., Umaña-Taylor, A. J., & Bamaca, M. Y. (2006). The influence of academic support on Latino adolescents' academic motivation. *Family Relations*, *55*, 279–291. https://doi.org/10.1111/j.1741-3729.2006.00402.x

Baron, A. S. (2015). Constraints on the development of implicit intergroup attitudes. *Child Development Perspectives*, *9*, 50–64. https://doi.org/10.1111/cdep.12105

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benner, A. D., & Graham, S. (2007). Navigating the transition to multiethnic urban high schools: Changing ethnic congruence and adolescents' school-related affect. *Journal of Research on Adolescence*, *17*, 207–220. https://doi.org/10.1111/j.1532-7795.2007.00519.x

Bierman, K. L., & McCauley, E. (1987). Children's descriptions of their peer interactions: Useful information for clinical child assessment. *Journal of Clinical Child Psychology*, *16*, 9–18. https://doi.org/10.1207/s15374424jccp1601_2

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, *62*, 465–480. https://doi.org/10.2307/2657316

Bonilla-Silva, E. (2015). The structure of racism in color-blind, "post-racial" America. *American Behavioral Scientist*, *59*(11), 1358–1376. https://doi.org/10.1177/0002764215586826

Brenick, A., Lawrence, S. E., Carvalheiro, D., & Berger, R. (2019). Teaching tolerance or acting tolerant? Evaluating skills- and contact-based prejudice reduction interventions among Palestinian-Israeli and Jewish-Israeli youth. *Journal of School Psychology*, *75*, 8–26. https://doi.org/10.1016/j.jso.2019.07.001

Brown, C. S. (2017). *Discrimination in childhood and adolescence: A developmental intergroup approach*. Taylor & Francis.

Bucchianeri, M. M., Gower, A. L., McMorris, B. J., & Eisenberg, M. E. (2016). Youth experiences with multiple types of prejudice-based harassment. *Journal of Adolescence*, *51*, 68–75. https://doi.org/10.1016/j.adolescence.2016.05.012

Burkholder, A. R., D'Esterre, A. P., & Killen, M. (2019). Intergroup relationships, context, and prejudice in childhood. In H. Fitzgerald, D. Johnson, D. Qin, F. Villarruel, & J. Norton (Eds.), *Handbook of children and prejudice: Integrating research, practice and policy* (pp. 115–130). Springer Press. https://doi.org/10.1007/978-3-030-12228-7_6

Burkholder, A. R., Elenbaas, L., & Killen, M. (2021). Giving priority to race or wealth in peer group contexts involving social inclusion. *Developmental Psychology*, *57*(5), 651–661. https://doi.org/10.1037/dev0001178

Chavous, T. M., Rivas-Drake, D., Smalls, C., Griffin, T., & Cogburn, C. (2008). Gender matters, too: The influences of school racial discrimination and racial identity on academic engagement outcomes among African American adolescents. *Developmental Psychology*, *44*, 637–654. https://doi.org/10.1037/0012-1649.44.3.637

Conry-Murray, C., & Turiel, E. (2020). Culture and children's reasoning about preferences and gender norms. *Journal of Experimental Child Psychology*, *196*, e104861. https://doi.org/10.1016/j.jecp.2020.104861

Cooley, S., Burkholder, A. R., & Killen, M. (2019). Social inclusion and exclusion in same-race and interracial peer encounters. *Developmental Psychology*, *55*, 2440–2450. https://doi.org/10.1037/dev0000810

Costello, M. C., & Dillard, C. (2019). *Hate at school*. Southern Poverty Law Center.

Crystal, D. S., Killen, M., & Ruck, M. (2008). It is who you know that counts: Intergroup contact and judgments about race-based exclusion. *British Journal of Developmental Psychology*, *26*, 51–70. https://doi.org/10.1348/026151007X198910

Diemer, M. A., Pineso, A., Banales, J., Matthews, C. J., Frisby, M. B., Harris, E. M., & McAlister, S. (2020). Recentering action in critical consciousness. *Child Development Perspectives*, *15*(1), 12–17. https://doi.org/10.1111/cdep.12393

Dovidio, J. F., & Gaertner, S. L. (2008). New directions in aversive racism research: Persistence and pervasiveness. In C. Willis-Esqueda (Ed.), *Motivational aspects of prejudice and racism* (Vol. *53*, pp. 43–67). Springer Science. https://doi.org/10.1007/978-0-387-73233-6_3

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, *82*, 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Elenbaas, L. (2019). Interwealth contact and young children's concern for equity. *Child Development*, *90*(1), 108–116. https://doi.org/10.1111/cdev.13157

Elenbaas, L., & Killen, M. (2016). Children rectify inequalities for disadvantaged groups. *Developmental Psychology*, *52*(8), 1318–1329. https://doi.org/10.1037/dev0000154

Elenbaas, L., Rizzo, M. T., & Killen, M. (2020). A developmental-science perspective on social inequality. *Current Directions in Psychological Science*, *29*, 610–616. https://doi.org/10.1177/0963721420964147

Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, *29*, 679–695. https://doi.org/10.1023/A:1026455906512

Gaias, L. M., Gal, D. E., Abry, T., Taylor, M., & Granger, K. L. (2018). Diversity exposure in preschool: Longitudinal implications for cross-race friendships and racial bias. *Journal of Applied Developmental Psychology*, *59*, 5–15. https://doi.org/10.1016/j.appdev.2018.02.005

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201–218; 269–281). Lawrence Erlbaum Associates Publishers.

Graham, S., & Echols, L. (2018). Race and ethnicity in peer relations research. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 590–614). The Guilford Press.

Graham, S., Munniksma, A., & Juvonen, J. (2014). Psychosocial benefits of cross-ethnic friendships in urban middle schools. *Child Development*, *85*(2), 469–483. https://doi.org/10.1111/cdev.12159

Hitti, A., Mulvey, K. L., Rutland, A., Abrams, D., & Killen, M. (2014). When is it okay to exclude a member of the ingroup?: Children's and adolescents' social reasoning. *Social Development*, *23*, 451–469. https://doi.org/10.1111/sode.12047

Hurd, N., Trawalter, S., Jakubow, A., Johnson, H. E., & Billingsley, J. T. (2021). Online racial discrimination and the role of white bystanders. *American Psychologist*, *77*, 39–55. https://doi.org/10.1037/amp0000603

Johnson, P. J., & Aboud, F. E. (2017). Evaluation of an intervention using cross-race friend storybooks to reduce prejudice among majority race young children. *Early Childhood Research Quarterly*, *40*, 110–122. https://doi.org/10.1016/j.ecresq.2017.02.003

Jost, J. T., & Kay, A. C. (2010). Social justice: History, theory, and research. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.),

*Handbook of social psychology* (pp. 1122–1165). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002030

Juvonen, J., Kogachi, K., & Graham, S. (2018). When and how do students benefit from ethnic diversity in middle school? *Child Development*, *89*, 1268–1282. https://doi.org/10.1111/cdev.12834

Juvonen, J., Lessard, L. M., Rastogi, R., Schacter, H. L., & Smith, D. S. (2019). Promoting social inclusion in educational settings: Challenges and opportunities. *Educational Psychologist*, *54*(4), 250–270. https://doi.org/10.1080/00461520.2019.1655645

Kendi, I. (2016). *Stamped from the beginning: The definitive history of racist ideas in America*. Hacette UK.

Kendi, I. (2019). *How to be an antiracist*. One World.

Killen, M., & Dahl, A. (2021). Moral reasoning enables developmental and societal changes. *Perspectives in Psychological Science*, *16*, 1209–1225. https://doi.org/10.1177/1745691620964076

Killen, M., Henning, A., Kelly, M. C., Crystal, D., & Ruck, M. (2007). Evaluations of interracial peer encounters by majority and minority U.S. children and adolescents. *International Journal of Behavioral Development*, *31*(5), 491–500. https://doi.org/10.1177/0165025407081478

Killen, M., Mulvey, K. L., & Hitti, A. (2013). Social exclusion: A developmental intergroup perspective. *Child Development*, *84*, 772–790. https://doi.org/10.1111/cdev.12012

Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. Wiley-Blackwell. https://doi.org/10.1002/9781444396317

Killen, M., & Rutland, A. (2022). Promoting fair and just school environments: Developing inclusive youth. *Policy Insights from the Brain and Behavioral Sciences*, *9*(1), 81–89. https://doi.org/10.1177/23727322211073795

Levy, S. R., Lytle, A., Shin, J. R., & Hughes, J. (2016). Understanding and reducing racial and ethnic prejudice among children and adolescents. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (2nd ed., pp. 455–483). Psychology Press.

Lewis, A. E., Hagerman, M. A., & Forman, T. A. (2019). The sociology of race & racism: Key concepts, contributions & debates. *Equity & Excellence in Education*, *52*(1), 29–46. https://doi.org/10.1080/10665684.2019.1627958

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, *67*, vii–147. https://doi.org/10.1111/1540-5834.t01-1-00187

Losinski, M., Ennis, R., Katslyannis, A., & Rapa, L. J. (2019). Schools as change agents in reducing bias and discrimination: Shaping behaviors and attitudes. *Journal of Child and Family Studies*, *28*, 2718–2726. https://doi.org/10.1007/s10826-019-01452-2

McGuire, L., & Rutland, A. (2020). Children and adolescents coordinate group and moral concerns within different goal contexts when allocating resources. *British Journal of Developmental Psychology*, *38*, 529–542. https://doi.org/10.1111/bjdp.12342

Mendes, W. B., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. (2007). Why egalitarianism might be good for your health: Physiological thriving during stressful intergroup encounters. *Psychological Science*, *18*, 991–998. https://doi.org/10.1111/j.1467-9280.2007.02014x

Mulvey, K. L. (2016). Children's reasoning about social exclusion: Balancing many factors. *Child Development Perspectives*, *10*(1), 22–27. https://doi.org/10.1111/cdep.12157

Mulvey, K. L., Rizzo, M. T., & Killen, M. (2016). Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science*, *19*, 999–1010. https://doi.org/10.1111/desc.12345

Munirah, A., Tenenbaum, H. R., & Rusconi, P. (2021). How do Saudi children and their mothers evaluate religion-based social exclusion? *Journal of Child and Family Studies*, *30*, 1353–1369. https://doi.org/10.1007/s10826-021-01915-5

Neblett, E. W., Banks, K. H., Cooper, S. M., & Smalls-Glover, C. (2013). Racial identity mediates the association between ethnic-racial socialization and depressive symptoms. *Cultural Diversity and Ethnic Minority Psychology*, *19*, 200–207. https://doi.org/10.1037/a0032205

Neblett, E. W., White, R. L., Ford, K. R., Philip, C. L., Nguyên, H. X., & Sellers, R. M. (2008). Patterns of racial socialization and psychological adjustment: Can parental communications about race reduce the impact of racial discrimination? *Journal of Research on Adolescence*, *18*, 477–515. https://doi.org/10.1111/j.1532-7795.2008.00568.x

Nesdale, D., & Lawson, M. J. (2011). Social groups and children's intergroup attitudes: Can school norms moderate the effects of social group norms? *Child Development*, *82*, 1594–1606. https://doi.org/10.1111/j.1467-8624.2011.01637.x

Nucci, L. P., & Ilten-Gee, R. (2021). *Moral education for social justice*. Teachers College Press, Columbia University. ISBN: 978-0-8077-6562-3.

Okonofau, J. A., Walton, G. M., & Eberhardt, J. (2016). A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, *11*(3), 381–398. https://doi.org/10.1177/1745691616635592

Pahlke, E., Bigler, R. S., & Suizzo, M.-A. (2012). Relations between colorblind socialization and children's racial bias: Evidence from European American mothers and their preschool children. *Child Development*, *83*, 1164–1179. https://doi.org/10.1111/j.1467-8624.2012.01770.x

Pauker, K., Williams, A., & Steele, J. R. (2016). Children's racial categorization in context. *Child Development Perspectives*, *10*, 33–38. https://doi.org/10.1111/cdep.12155

Perry, S. P., Skinner, A. L., & Abaied, J. L. (2019). Bias awareness predicts color conscious racial socialization methods among White parents. *Journal of Social Issues*, *75*(4), 1035–1056. https://doi.org/10.1111/josi.12348

Piaget, J. (1932). *The moral judgment of the child*. Free Press.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X. F., Martinez, A., Bloom, H., & Hill, C. (2011). Optimal design software for multi-level and longitudinal research (version 3.01) [software]. www.wtgrantfoundation.org

Rivas-Drake, D., Markstrom, C., Syed, M., Lee, R. M., Umaña-Taylor, A. J., Yip, T., Seaton, E. K., Quintana, S., Schwartz, S. J., French, S., & Ethnic and Racial Identity in the 21st Century Study Group. (2014). Ethnic and racial identity in adolescence: Implications for psychological, academic, and health outcomes. *Child Development*, *85*, 40–57. https://doi.org/10.1111/cdev.12200

Rizzo, M. T., Britton, T., & Rhodes, M. (2021, September 9). Developmental origins of anti-black bias in white U.S. children: Exposure to and beliefs about racial inequality. *Developmental Science*. https://doi.org/10.31219/osf.io/7s39r

Rizzo, M. T., Cooley, S., Elenbaas, L., & Killen, M. (2018). Young children's inclusion decisions in moral and social-conventional group norm contexts. *Journal of Experimental Child Psychology*, *165*, 19–36. https://doi.org/10.1016/j.jecp.2017.05.006

Roberts, S. O., & Rizzo, M. T. (2020). The psychology of American racism. *American Psychologist*, *76*, 475–487. https://doi.org/10.1037/amp0000642

Rogers, L. O. (2019). Commentary on economic inequality: "What" and "who" constitutes research on social inequality in developmental science? *Developmental Psychology*, *55*(3), 586–591.

Rogers, L. O., Scott, M. A., & Way, N. (2015). Racial and gender identity among black adolescent males: An intersectionality perspective. *Child Development*, *86*, 407–424. https://doi.org/10.1111/cdev.12303

Rouland, K. K., Rowley, S. J., & Kurtz-Costes, B. (2013). Self-views of African-American youth are related to the gender stereotypes and ability attributions of their parents. *Self and Identity*, *4*, 382–399. https://doi.org/10.1080/15298868.2012.682360

Ruck, M., Park, H., Killen, M., & Crystal, D. S. (2011). Intergroup contact and evaluations of race-based exclusion in urban minority

children and adolescents. *Journal of Youth and Adolescence*, *40*, 633–643. https://doi.org/10.1111/j.2044-835X.2010.02021.x

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*(3), 279–291. https://doi.org/10.1177/1745691610369468

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*, 688.

Skinner, O. D., Kurtz-Costes, B., Vuletich, H., Copping, K., & Rowley, S. J. (2021). Race differences in black and white adolescents' academic gender stereotypes across middle and late adolescence. *Cultural Diversity and Ethnic Minority Psychology*, *27*(3), 537–545. https://doi.org/10.1037/cdp0000384

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* (2nd ed., pp. 23–45). Psychology Press.

Suárez-Orozco, C., Casanova, S., Martin, M., Katsiaficas, D., Cuellar, V., Smith, N. A., & Dias, S. I. (2015). Toxic rain in class: Classroom interpersonal microaggressions. *Educational Researcher*, *44*, 151–160. https://doi.org/10.3102/0013189X15580314

Tropp, L. R., O'Brien, T. C., & Migacheva, K. (2014). How peer norms of inclusion and exclusion predict children's interest in cross-ethnic friendships. *Journal of Social Issues*, *70*(1), 151–166. https://doi.org/10.1111/josi.12052

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Turiel, E. (2002). *Culture and morality*. Cambridge University Press.

Turiel, E., Chung, E., & Carr, J. A. (2016). Struggles for equal rights and social justice as unrepresented and represented in psychological research. In J. B. Benson (Series Ed.), *Advances in child development and behavior* (Vol. *50*, pp. 1–29). Elsevier Academic Press. https://doi.org/10.1016/bs.acdb.2015.11.004

Turner, R. N., & Cameron, L. (2016). Confidence in contact: A new perspective on promoting cross-race friendships among children and adolescents. *Social Issues and Policy Review*, *10*, 212–246. https://doi.org/10.1111/sipr.12023

Umaña-Taylor, A. J., Douglass, S., Updegraff, K. A., & Marsiglia, F. F. (2018). A small-scale randomized efficacy trial of the identity project: Promoting adolescents' ethnic-racial identity exploration and resolution. *Child Development*, *89*(3), 862–870. https://doi.org/10.1111/cdev.12755

Verkuyten, M., Thijs, J., & Gharaei, N. (2019). Discrimination and academic (dis)engagement of ethnic-racial minority students: A social identity threat perspective. *Social Psychology of Education: An International Journal*, *22*(2), 267–290. https://doi.org/10.1007/s11218-018-09476-0

Yip, T. (2015). The effects of ethnic/racial discrimination and sleep quality on depressive symptoms and self-esteem trajectories among diverse adolescents. *Journal of Youth and Adolescence*, *44*, 419–430. https://doi.org/10.1007/s10964-014-0123-x

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Killen, M., Burkholder, A. R., D'Esterre, A. P., Sims, R. N., Glidden, J., Yee, K. M., Luken Raz, K. V., Elenbaas, L., Rizzo, M. T., Woodward, B., Samuelson, A., Sweet, T. M., Stapleton, L. M. (2022). Testing the effectiveness of the *Developing Inclusive Youth program*: A multisite randomized control trial. *Child Development*, 93, 732–750. https://doi.org/10.1111/cdev.13785