Contact in the Classroom: A Field Experiment on Virtual Intergroup Contact in Elementary Schools

Pre-Analysis Plan

Salma Mousa*

February 14, 2019

Contents

*PhD Candidate, Political Science, Stanford University. smousa@stanford.edu.

1 Overview

1.1 Motivation

Can virtual contact build tolerance toward out-groups? The classic "contact hypothesis" suggests that positive contact across social lines has the potential to build tolerance, if subjects share an equal power status within the intervention, work cooperatively toward a common goal, and if the relationship is supported by authorities, laws or customs (Allport, Clark and Pettigrew, 1954; Pettigrew and Tropp, 2006). While positive intergroup contact typically reduces prejudice (Pettigrew and Tropp, 2006; Paluck, Green and Green, 2017), less is known about the potential for virtual contact (i.e., conversing over the Internet) to do the same. Studies of virtual contact tend to involve simulated or fictional out-group members, rather than contact with real-life peers (Alvídrez et al., 2015; Cameron et al., 2006; Mallett and Wilson, 2010; Husnu, Mertan and Cicek, 2018; Stathi et al., 2014; Vezzali et al., 2012; Crisp and Turner, 2009). A related research agenda focus on empathy-building interventions. These interventions often involve putting oneself in the shoes of a fictional or composite out-group member, to great effect (Simonivitz, Kezdi and Kardos, 2017; Adida, Lo and Platas, 2018). Causal evidence on virtual contact, however, remains elusive.

This study build on the insights from the inter-group contact and empathy-building agendas. To do this, I evaluate a virtual exchange program aimed at elementary classrooms. The virtual exchange program (hereafter "VEP") connects classrooms around the world via real-time video sessions (similar to Skype). The VEP is aimed at building the social and emotional learning among children, as well as facilitating curiosity about other cultures. Teachers of students aged seven to eleven can sign up to the platform for free. Based on their scheduling availability and topics of interest (e.g. the local environment, or celebrations around the world), they are then matched with a partner classroom. Teachers can also apply to become fellows through a dedicated VEP program. Teaching fellows share learnings and insights with each other and receive structured support from the VEP team (e.g. research training, conference participation, and administrative support). Fellowships last one academic semester. Fellows are required to complete at least five virtual exchanges during this time.

2 Research Design

What is the *causal* effect virtual exchanges on the attitudes of students? To answer this question, I compare students that do and do not partake in the VEP at two time periods four months apart (t1 and t2). I create a control group in two ways. First, many VEP fellowship target teachers teach multiple classes of students at once. These teachers offer us an ethical opportunity to randomly choose which class will partake in the exchanges.[1] Multi-class teachers who enroll in the study allow us to create a *treatment group* (the class that uses the VEP) and *control group* (the class that does not use the VEP) – leading to causal estimates of the VEP's effects.[2] By comparing the outcomes of students taught by the same teacher, we hold constant teacher traits like personality, experience, and teaching style. Any

[1] Because the VEP currently only allows one class per teacher to use the platform, we exploit this pre-existing constraint to the number of classes that can be enrolled in the VEP to ethically conduct an RCT.

[2] About 30 – 40% of study participants teach multiple classes.

differences observed between students can then be attributed to the VEP, rather than the teacher.

I have designed a second approach to creating a control group for teachers that teach just one classroom. Single-class teachers nominate a "companion class" to participate in the study. A companion class is another class in the same grade, in the same school. Students in the companion class serve as the control group. A difference-in-differences design accounting for baseline data causally identifies treatment effects among students.

In the U.S., students are often randomly assigned to a particular classroom at the start of the school year. This means that assignment to treatment – a classroom with a teacher that chooses to enroll in the VEP – is as-good-as-random. I rely on another assumption. In order for treatment assignment to be random, teachers should not select classes to partake in exchanges on the basis of traits like empathy or knowledge of other cultures. I validate this assumption based on qualitative evidence from early submissions of teacher applicant surveys ($n = 27$). 24 applicants for the VEP Winter fellowship explained their decision to choose a particular class on one of three bases: scheduling availability, a set of students fitting the target age-range, and the classroom that they spend most time with (e.g. homeroom). The remaining three teachers reported that learning about other cultures aligns with this class's social studies curricula or learning progression. It seems safe to assume that most teachers do not select particular classes to partake in the VEP because of traits correlated with the outcomes of interest. Moreover, assignment to particular classrooms at the start of the year is random, there is no ex-ante reason to believe that some classrooms are more empathetic or knowledgable about other cultures to begin with.

Among students that *do* conduct exchanges, what is the added effect of the partner classroom's identity? I answer this question by leveraging a natural experiment built-in to the VEP platform. The VEP matching algorithm automatically finds a partner class based on scheduling availability and topics of interest (e.g. the local environment, or holidays around the world). This setting provides a natural experiment: social distance to the partner classroom is randomly assigned. I define socially different as a partner classroom where the majority of students belong to a different ethnic or national group. This particular dimension of difference is the easiest to perceive among children of the target age range. This measure is divided into three levels: 75% or greater co-ethnic (3), 25 – 75% co-ethnic (2), and fewer than 25% co-ethnic (1). Pairs that bring together classrooms in two different countries are categorized as non co-ethnics (1).

I thus collect t1 and t2 outcomes about control and treated students. In addition to this group, I also compare t1 and t2 outcomes among teachers (a descriptive analysis). Teachers can also opt-in to a research track embedded into the VEP fellowship. As part of this research track, teachers may choose to survey parents. If teachers survey parents from a minimum of 5 control classrooms and 5 treated classrooms, I will conduct an exploratory analysis on spillover effects among parents. Teachers may also choose to conduct lab games with their students as part of the research track. I take the results from this lab game as an additional exploratory analysis.

2.1 Research Questions

Can virtual contact increase out-group tolerance, empathy, and knowledge among students? Condition on experiencing virtual contact, how does social distance shape these effects? How long do these effects last? How does dosage shape outcomes?

2.2 Hypotheses

I predict that out-group empathy, tolerance, and knowledge will increase among students and teachers, while in-group bias will decrease. For teachers and parents, we further expect a reduction in discrimination on the basis of religion and nationality. The teacher and parent analyses are secondary, however, and descriptive (over-time changes) rather than causal.

H1: Relative to students that do not partake in virtual exchanges (control condition), **students that partake in virtual exchanges** will be:

Tolerance

(a) More likely to report "somewhat wanting to" or "very much wanting to" meet a kid who "just moved to your area from another country";

(b) More likely to report "somewhat wanting to" or "very much wanting to" exchange messages with a foreign penpal; and

(c) More likely to use more positive language when describing those in other countries, as measured via automated sentiment analysis.

Empathy

(d) More likely to report that they "sometimes" or "often" put themselves in others' shoes.

Pro-Social Behavior

(e) More likely to anonymously donate stickers to another student in the class (a dictator game adapted for children, as in (Benenson, Pascoe and Radmore, 2007)). Only for research track teaching fellows;

(f) Less likely to engage in bullying (measured by teacher surveys – 1 to 10 scale of how serious of a problem bullying is in his/her classroom);

(g) More likely to have parents that agree or strongly agree that the student "stops others who are unfair or disruptive" (measured by the parent survey, only for research track teaching fellows); and

(h) More likely to have teachers that agree or strongly agree that students in his/her classroom "stops others who are unfair or disruptive" (measured by the teacher survey).

(i) More likely to have teachers that agree or strongly agree that students in his/her classroom "Students from different social classes and races get along well" (measured by the teacher survey).

Knowledge

(j) More likely to report knowing "a little" or "a lot" about life in other parts of the world.

Norms

(k) More likely to report that their friends would be "a little happy" or "very happy" if they started playing with a "kid who just moved to your area from another part of the world."

(l) More likely to report that their parents would be "a little happy" or "very happy" if they started playing with a "kid who just moved to your area from another part of the world."

In-group Bias

(m) More likely to report that they share "a little" or "a lot" in common with ethnic out-groups (depicted using a photo of age-matched peers of Black, Hispanic, and White descent).[3]

H2: Relative to t1, **teachers at t2** will be:

Tolerance

(a) More likely to agree or strongly agree that their country of residence would be better society if "people treated each other as equals first, rather than divide people by citizenship, religion, or ethnicity" ; and

(b) Less likely to discriminate toward religious and national out-group (as measured by the conjoint experiment)

Empathy

(c) More likely to agree or strongly agree that "there are two sides to every situation, and I always try to look at them both."

In-group Bias

(d) More likely to agree or strongly agree that "I share a lot in common with people in other countries."

[3] Out-groups are defined in relation to the student's own ethnicity, roughly proxied using a survey item asking student to choose which Emoji skin tone best fit their own.

H3: Relative to parents of students assigned to the control condition, **parents of students assigned to the treatment condition** will be:

Tolerance

(a) More likely to agree or strongly agree that their country of residence would be better society if "people treated each other as equals first, rather than divide people by citizenship, religion, or ethnicity" ; and

(b) Less likely to discriminate toward religious and national out-group (as measured by the conjoint experiment)

3 Outcomes

3.1 Attitudinal Outcomes

Students and teachers take t0 and t1 surveys. The t1 surveys take place four months after the intervention begins, and two weeks after the final exchange Where feasible for teachers to track students during the new academic year, students are also surveyed at t3, seven months after the final exchange. The student instrument uses validated items measuring cognitive empathy (perspective-taking), in-group bias (preference for one's own group), self-other overlap (perceived similarity with a specific out-group), sentiment toward out-groups in other countries, norms around cross-group friendships within their friendship group and household, and knowledge of life in other countries (Alvídrez et al., 2015; Cameron et al., 2001; Rowley et al., 2008; Scacco and Warren, 2018; Tropp, O'Brien and Migacheva, 2014; Beelmann and Heinemann, 2014). We analyze these items separately, as well as running an unsupervised hierarchical clustering model to collapse them indices and maximize statistical power.

The teacher survey instrument measures tolerance using a conjoint experiment described below, in addition to items capturing cognitive empathy. I also use the teacher survey to gather classroom level data. Classroom-level variables are: school country, school state (if in the U.S.), school area (rural, suburban, or urban), classroom ethnic composition, classroom ethnic composition, and classroom socio-economic status – Title 1 status (U.S. only), % students receiving free or reduced lunch (U.S. only), and the income composition of students' families (mostly high income, middle income, low income, or mixed income).

Lastly, I run a sentiment analysis on open-ended questions asking students and teachers what they think of people from other countries.

3.2 Behavioral Outcomes

I test whether students that participate in the VEP (the treatment condition) are more likely to indicate interest in having a penpal relative to similar students that do not use the VEP (the control condition).

Among students that use the VEP, I also measure how many messages they actually send to penpals in their partner classroom.[4] I do this through teacher-given prompts at the end of the semester, one week after the final exchange: "if you would like to send a message to a student in our partner classroom, please write it out on a piece of paper, and I will take a photo of it to send to them." Using a convenient file naming system, I can track which students choose to write out a message and subsequently merge this data with the student's survey responses.

The teacher survey also includes three items capturing student behavior: the degree to which bullying is a problem, the degree to which students stop others from being unfair or disruptive, and the degree to which students get along with students of other social classes and races. These three outcomes are measured at the classroom rather than individual level.

A subset of teachers may opt-in to the fellowship's research track. As part of their research project, teachers may choose to conduct lab games among the treated and control classrooms. I train teachers to run a dictator game adapted for children, as in (Benenson, Pascoe and Radmore, 2007). Teachers give 30 stickers to each student, alone. The student chooses their favorite 10 stickers. The teacher then carefully explains that the stickers belong to the student, but that they may choose to give some stickers to another boy or girl in the classroom. The teacher stresses that this decision is totally anonymous, and ensures that the student understands this point. Donated stickers are placed in a white envelope in front of the student, with the name of the recipient labeled on the front.

3.3 Implicit Discrimination Conjoint Experiment

We measure implicit discrimination toward religious and national out-groups using a conjoint experiment, included in the teacher and parent surveys. Respondents are asked to compare profiles of classrooms that differ on five attributes: teacher's age, teacher's gender, country, majority religion, and school area (rural, suburban, or urban). The values each attribute can take are listed below:

Attributes and Values

- **Teacher Age:** 21 years old, 38 years old, 55 years old
- **Teacher Gender:** Female, Male
- **Country:** Mexico, U.S., China, Iraq, Nigeria
- **Majority Religion:** Muslim, Christian, No religion
- **School Area:** Rural, Suburban, Urban

Respondents then rank their preference for connecting with Teacher A and Teacher B on a scale from 1 – 7. Respondents are then asked, "if you had to choose just one of these teachers to connect with, which would you choose?" The survey presents four conjoint questions in a row. The data from the

[4]A meta-analysis of 211 contact interventions among children concludes that choosing to spend time with out-group members (e.g. via offline communication) is a valid behavioral measure of prejudice (Davies et al., 2011), as is indicating sustained interest in communicating (West et al., 2014).

conjoint experiment allows us to estimate the per cent "penalty" or "advantage" respondents give, on average, for different religious and national out-groups.[5]

3.4 Mechanisms

If we do find treatment effects, what explains them? The teacher and student surveys include items that test for three mechanisms in particular:

- **Reduced in-group bias.** Especially when it comes to children, inter-group contact may not shift attitudes toward the out-group, but instead weaken the strong preference for one's in-group. This is measured by the in-group bias survey item outlined above.
- **Increased information.** Accurate information replaces missing or incorrect information, prompting subjects to overturn stereotypes and notice commonalities (Scacco and Warren, 2018). This is measured by the knowledge survey item outlined above.
- **Strengthened empathetic norms.** Those observing inter-group contact – parents, friends, teachers – may begin to normalize inter-group friendships. Alternatively, these peripheral groups may strengthen norms against hostile behavior or discrimination. This is measured by two norms-related survey items outlined above (relating to parents and friends).

4 Estimation

4.1 Hierarchical Clustering of Outcome Variables

I take pre-survey responses as covariates to increase statistical precision and collapse individual survey outcomes into more stable indices to extract the most power out of the t0 survey. I use an unsupervised hierarchical machine learning algorithm to identify latent clusters in the pre-survey data. Below is sample code to show how the clusters will be created. Once 25% of the baseline data are collected, I will run this code to create the real indices and supplement this PAP.

The following are the attitudinal outcomes of interest from the student survey instrument:

```
outcome_variables <- c("knowledge_foreign_t1", "empathy_general_t1",
                "penpal_interest_t1", "meet_foreign_t1" ,
                "common_wht_t1" ,  "common_hisp_t1" ,
                "common_blk_t1" ,  "common_foreign_t1", "norm_parents_t1",
                "norm_friends_t1")
```

Below is the code used to generate factor loadings associated with each item using a hierarchical clustering method.

[5]The teacher survey gathers information on teacher religion and school location, enabling me to infer which profiles presented in the conjoint are religious and national out-groups). Where no responses are given, I use the district's majority religion (measured according to the most recent census).

```
# Ensure package recognizes the variables as "qualitative" (ie, categorical)
df <- df %>% mutate_at(vars(one_of(outcome_variables)), funs(as.factor(.)))

# Apply clustering algorithm
tree <- hclustvar(X.quali = df[outcome_variables])
plot.hclustvar(tree, type = "tree")
plot.hclustvar(tree, type = "index")

# Cut the tree at n clusters (based on dendrogram)

cut <- cutreevar(tree, n)
clusters <- cut$cluster
cut$var
```

4.2 Estimating ATE

I use an OLS model with standard errors clustered at the class level and with the covariates mentioned above, as shown below.

```
# Function for estimating treatment effects. Requires package "lfe" for
# clustering the standard errors at the class level.
ate <- function(dta, dv, treat, covars, cluster = "class_id", extravars_extract =
     NULL) {

  # Create OLS formula
  rhs <- paste(c(treat, covars), collapse = " + ")
  f <- paste(dv, "~", rhs, "| 0 | 0 |", cluster)
  cat("Formula used:", f, "\n")

  # Estimate regression
  m <- felm(formula(f), data = dta)

  # Extract results of interest
  tidy(m, conf.int = T) %>%
    filter(term %in% c("(Intercept)", treat, extravars_extract))

}
```

The main analyses will be estimated as shown below. I analyze the treatment effect on indices by subtracting the t0 index from the t1 index, and adding demographic controls outlined in section 4.3. The indices below are to demonstrate the estimation code; I will replace these with the clusters computed based on a sample of the baseline data.

```
# Index #1
```

```
ate(df, dv = "dv_norms", treat = "treat_PAP", covars = covariates)

# Index #2
ate(df, dv = "dv_empathy", treat = "treat_PAP", covars = covariates)

# Index #3
ate(df, dv = "dv_ingroup", treat = "treat_PAP", covars = covariates)

# Index #4
ate(df, dv = "dv_knowledge", treat = "treat_PAP", covars = covariates)
```

The teacher survey includes three items capturing classroom-level pro-social behavior. I compute a simple difference in means comparing these items across treatment and control classrooms.

4.3 Covariates to use in Regression Adjustment

I will include the following demographic variables from the baseline pre-survey as covariates to predict the outcome and increase power. These covariates include the outcome variable measured at t1, and demographic traits at the individual and classroom levels.[6] I add to the vector of covariates below the t0 measure of the outcome in question.

```
covariates <- c("ethnicity", "lang_home", "grade", "country", "siblings", "
    done_exchanges_before", "class_ses_status", "class_ethnicity", "
    class_eth_composition", "majority_eth", "distant_match")
```

4.4 Treatment Effect Heterogeneity by Subject Attributes

Detecting interaction effects typically requires a ten-fold increase (or more) in the sample size needed to estimate a main effect. Interaction effects are therefore only suggestive. I expect that positive ATE estimates would be amplified among three groups:

1. Classrooms matched with socially different partner classrooms;

2. Only children, given that recent research suggests that children with siblings have higher baseline empathy (Jambon et al., 2018);

3. Individuals that belong to the classroom's ethnic majority, relative to those that belong to ethnic minority groups (Beelmann and Heinemann, 2014; Feddes, Noack and Rutland, 2009; Aboud et al., 2012). We measure ethnic identity by asking student respondents to choose which skin tone fits them the closest, and asking parent or teacher respondents to self-report their ethnicity.

[6] Class SES is computed using the teacher-reported items on the income level of students. For U.S. classrooms, I substitute this measure with Title 1 status and % students receiving free or reduced lunch, as a robustness check. Social distance is measured by whether the classroom is matched with a majority co-ethnic, non co-ethnic, or mixed co-ethnicity partner classroom. Class ethnicity is a set of continuous variables: % White, % Black, % Hispanic, % East Asian, % Pacific Islander, % South Asian, % American Indian or Alaska Native, % Middle Eastern or North African, and % Other. Classroom ethnic composition is a continuous measure of how many ethnic groups are represented in the classroom.

We ask this question last to avoid priming effects associated with reminding respondents of their immutable ethnic identities (Bigler and Liben, 2007; Alvídrez et al., 2015).

As an example, the code below demonstrates how I will estimate heterogenous effects:

```
# Heterogeneous effects by having a homogenous friendship group
covariates_new <- c(covariates, "siblings")

# Interacting sibling dummy with the treatment effect, main dv analysis
ate(df, dv = "dv_norms", treat = "treat_PAP", covars = covariates_new,
    extravars_extract = c("siblings"))

 # Heterogeneous effects by social distance of match
covariates_new <- c(covariates, "distant_match")

# Interacting social distance with the treatment effect, main dv analysis
ate(df, dv = "dv_norms", treat = "treat_PAP", covars = covariates_new,
    extravars_extract = c("distant_match"))

# Heterogeneous effects by whether or not the student belongs to the classroom's
    majority ethnic group
covariates_new <- c(covariates, "majority_eth")

# Interacting majority ethnicity with the treatment effect
ate(df, dv = "dv_norms", treat = "treat_PAP", covars = covariates_new,
    extravars_extract = c("majority_eth"))
```

5 Other

5.1 Missing Values

I impute missing baseline values using the default predictive mean matching method in the `mice` package. I will present the results with and without this imputation as a robustness check. All variables are taken as predictors for the imputation algorithm, except student and classroom ID.

5.2 Power Calculation

Below is a power calculation taking into account the clustered design, and based on the ten outcome questions asked in a pilot of n = 982 students. As demonstrated below, the outcome that requires that largest sample size (perceiving oneself to have something in common with Black children) yields an effective sample size of n = 2,478 students. The minimum predicted sample size for this study is n = 2,760 students divided into 60 classrooms, half of which are treated.

```
## calculating design effect while adjusting for clusters

## outcome 1: interest in meeting a new kid who moved from another country
```

```
## deff(df1$meet_foreign_t1, df1$class_id1)

         n    clusters        rho       deff
968.00000000 70.00000000 0.07604358 2.49761869

## formula: (1 + rho *(14 -1))

> 1 + (0.07604358*22)
[1] 2.672959

## effective sample size

> 2760/ 2.672959
[1] 1032.564

## outcome 2: putting yourself in someone else's shoes in a disagreement

> deff(df1$empathy_general_t1, df1$class_id1)

         n    clusters        rho       deff
955.00000000 70.00000000 0.04161579 1.81636679

## formula: (1 + rho *(14 -1))

> 1 + (0.04161579*22)
[1] 1.915547

## effective sample size

> 2760/ 1.915547
[1] 1440.842


## outcome 3: interest in sending messages to a penpal

> deff(df1$penpal_interest_t1, df1$class_id1)

        n   clusters       rho      deff
960.0000000 70.0000000 0.1015721 2.9958917

## formula: (1 + rho *(14 -1))

> 1 + (0.1015721*22)
[1] 3.234586

## effective sample size
```

```
> 2760/ 3.234586
[1] 853.2777

## outcome 4: knowledge of what life is like in other countries

> deff(df1$knowledge_foreign_t1, df1$class_id1)

         n    clusters       rho       deff
960.00000000 70.00000000 0.03876198 1.76167294

## formula: (1 + rho *(14 -1))

> 1 + (0.03876198*22)
[1] 1.852764

## effective sample size

> 2760/ 1.852764
[1] 1489.666

## outcome 5: something in common with white kids

> deff(df1$common_wht_t1, df1$class_id1)

         n    clusters       rho       deff
956.00000000 70.00000000 0.01142291 1.22489747

## formula: (1 + rho *(14 -1))

> 1 + (0.01142291*22)
[1] 1.251304

## effective sample size

> 2760/ 1.251304
[1] 2205.699


## outcome 6: something in common with hispanic kids

> deff(df1$common_hisp_t1, df1$class_id1)
         n    clusters       rho       deff
957.00000000 69.00000000 0.02264975 1.44570542

## formula: (1 + rho *(14 -1))
```

```
> 1 + (0.02264975*22)
[1] 1.498295

## effective sample size

> 2760/ 1.498295
[1] 1842.094

## outcome 7: something in common with black kids

> deff(df1$common_blk_t1, df1$class_id1)
          n     clusters         rho        deff
959.000000000 70.000000000 0.005192875 1.102016438

## formula: (1 + rho *(14 -1))

> 1 + (0.005192875*22)
[1] 1.114243

## effective sample size

> 2760/ 1.114243
[1] 2477.018

## outcome 8: something in common with foreign kids

> deff(df1$common_foreign_t1, df1$class_id1)
         n    clusters       rho       deff
956.00000000 70.00000000 0.09102027 2.78632040

## formula: (1 + rho *(14 -1))

> 1 + (0.09102027*22)
[1] 3.002446

## effective sample size

> 2760/ 3.002446
[1] 919.2505

## outcome 9: parents would be happy if i played with a foreign kid

> deff(df1$norm_parents_t1, df1$class_id1)
         n    clusters       rho       deff
962.00000000 70.00000000 0.05741171 2.12830122

## formula: (1 + rho *(14 -1))
```

```
> 1 + (0.05741171*22)
[1] 2.263058

## effective sample size

> 2760/ 2.263058
[1] 1219.589

## outcome 10: friends would be happy if i played with a foreign kid

> deff(df1$norm_friends_t1, df1$class_id1)
          n    clusters        rho        deff
967.00000000 70.00000000 0.09261444 2.82240694


## formula: (1 + rho *(14 -1))

> 1 + (0.09261444*22)
[1] 3.037518

## effective sample size

> 2760/ 3.037518
[1] 908.6366
```

5.3 IRB

This study has been approved by Stanford University's Institutional Review Board (Protocol #47727).

References

Aboud, Frances E, Colin Tredoux, Linda R Tropp, Christia Spears Brown, Ulrike Niens, Noraini M Noor et al. 2012. "Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review." *Developmental review* 32(4):307–336.

Adida, Claire L, Adeline Lo and Melina R Platas. 2018. "Perspective taking can promote short-term inclusionary behavior toward Syrian refugees." *Proceedings of the National Academy of Sciences* 115(38):9521–9526.

Allport, Gordon Willard, Kenneth Clark and Thomas Pettigrew. 1954. "The nature of prejudice.".

Alvídrez, Salvador, Valeriano Piñeiro-Naval, María Marcos-Ramos and José Luis Rojas-Solís. 2015. "Intergroup contact in computer-mediated communication: The interplay of a stereotype-disconfirming behavior and a lasting group identity on reducing prejudiced perceptions." *Computers in Human Behavior* 52:533–540.

Beelmann, Andreas and Kim Sarah Heinemann. 2014. "Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs." *Journal of Applied Developmental Psychology* 35(1):10–24.

Benenson, Joyce F, Joanna Pascoe and Nicola Radmore. 2007. "Children's altruistic behavior in the dictator game." *Evolution and Human Behavior* 28(3):168–175.

Bigler, Rebecca S and Lynn S Liben. 2007. "Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice." *Current Directions in Psychological Science* 16(3):162–166.

Cameron, Jessica A, Jeannette M Alvarez, Diane N Ruble and Andrew J Fuligni. 2001. "Children's lay theories about ingroups and outgroups: Reconceptualizing research on prejudice." *Personality and Social Psychology Review* 5(2):118–128.

Cameron, Lindsey, Adam Rutland, Rupert Brown and Rebecca Douch. 2006. "Changing children's intergroup attitudes toward refugees: Testing different models of extended contact." *Child development* 77(5):1208–1219.

Crisp, Richard J and Rhiannon N Turner. 2009. "Can imagined interactions produce positive perceptions?: Reducing prejudice through simulated social contact." *American Psychologist* 64(4):231.

Davies, Kristin, Linda R Tropp, Arthur Aron, Thomas F Pettigrew and Stephen C Wright. 2011. "Cross-group friendships and intergroup attitudes: A meta-analytic review." *Personality and Social Psychology Review* 15(4):332–351.

Feddes, Allard R, Peter Noack and Adam Rutland. 2009. "Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study." *Child development* 80(2):377–390.

Husnu, Shenel, Biran Mertan and Onay Cicek. 2018. "Reducing Turkish Cypriot children's prejudice toward Greek Cypriots: Vicarious and extended intergroup contact through storytelling." *Group Processes & Intergroup Relations* 21(1):178–192.

Jambon, Marc, Sheri Madigan, André Plamondon, Ella Daniel and Jennifer M Jenkins. 2018. "The Development of Empathic Concern in Siblings: A Reciprocal Influence Model." *Child development* .

Mallett, Robyn K and Timothy D Wilson. 2010. "Increasing positive intergroup contact." *Journal of Experimental Social Psychology* 46(2):382–387.

Paluck, Elizabeth Levy, Seth Green and Donald P Green. 2017. "The Contact Hypothesis Revisited." *SSRN* .

Pettigrew, Thomas F and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90(5):751.

Rowley, Stephanie J, Margaret R Burchinal, Joanne E Roberts and Susan A Zeisel. 2008. "Racial identity, social context, and race-related social cognition in African Americans during middle childhood." *Developmental Psychology* 44(6):1537.

Scacco, Alexandra and Shana S Warren. 2018. "Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria." *American Political Science Review* pp. 1–24.

Simonivitz, Gabor, Gabor Kezdi and Peter Kardos. 2017. "Seeing the World Through the Other's Eye: An Online Intervention Reducing Ethnic Prejudice." *American Political Science Review* pp. 1–8.

Stathi, Sofia, Lindsey Cameron, Bonny Hartley and Shona Bradford. 2014. "Imagined contact as a prejudice-reduction intervention in schools: The underlying role of similarity and attitudes." *Journal of Applied Social Psychology* 44(8):536–546.

Tropp, Linda R, Thomas C O'Brien and Katya Migacheva. 2014. "How peer norms of inclusion and exclusion predict children's interest in cross-ethnic friendships." *Journal of Social Issues* 70(1):151–166.

Vezzali, Loris, Dora Capozza, Sofia Stathi and Dino Giovannini. 2012. "Increasing outgroup trust, reducing infrahumanization, and enhancing future contact intentions via imagined intergroup contact." *Journal of Experimental Social Psychology* 48(1):437–440.

West, Tessa V, Joe C Magee, Sarah H Gordon and Lindy Gullett. 2014. "A little similarity goes a long way: The effects of peripheral but self-revealing similarities on improving and sustaining interracial relationships." *Journal of personality and social psychology* 107(1):81.