FONDAZIONE BRUNO KESSLER
IRVAPP
INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

FBK-IRVAPP Working Paper Series

Working Paper No. 2024-03

Beyond Sight: Exploring the Impact of a Multifaceted Intervention on Knowledge, Attitudes and Behaviors towards Persons with Visual Impairment

Sergiu Burlacu
Federico Podestà
Davide Azzolini

April 2024

FBK-IRVAPP Working Paper No. 2024-03

Beyond Sight: Exploring the Impact of a Multifaceted Intervention on Knowledge, Attitudes and Behaviors towards Persons with Visual Impairment

Sergiu Burlacu
FBK-IRVAPP
sburlacu@irvapp.it

Federico Podestà
FBK-IRVAPP
podesta@irvapp.it

Davide Azzolini
FBK-IRVAPP
azzolini@irvapp.it

FONDAZIONE BRUNO KESSLER
IRVAPP
INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

Institute for the Evaluation of Public Policies
Fondazione Bruno Kessler
Via Santa Croce, 77
38122 Trento, Italy
https://irvapp.fbk.eu

Beyond Sight:

Exploring the Impact of a Multifaceted Intervention on Knowledge, Attitudes and Behaviors towards Persons with Visual Impairment

Sergiu Burlacu[*‡] Davide Azzolini[†] Federico Podestà[†]

This version: April, 2024

Abstract

We evaluate a multi-faceted intervention aimed at improving social inclusion and reducing prejudice against individuals with visual impairment. The intervention, randomly assigned to upper-secondary school students, consists of an awareness-raising activity and a simulation-based inter-group contact activity. While we find positive effects on knowledge of visual impairment, perspective-taking and empathic concerns, and general societal attitudes toward persons with visual impairment, no improvements are observed in terms of implicit attitudes or multidimensional attitudes. Moreover, the intervention does not improve outcomes measured through incentivized choices, such as the willingness to pay for social interaction with persons with visual impairment, beliefs regarding their performance and outcomes in various domains, and altruism towards them. The evidence suggests that assessing impacts only on knowledge and general attitudes, as is commonly done in the literature, may not suffice to determine the extent to which such interventions are successful at improving social inclusion for persons with visual or other forms of impairment.

JEL Codes: I24, I31, J14, J18, P36

Keywords: Disability, Social Inclusion, Discrimination, Inter-group contact

[*] *Corresponding author;* e-mail: sburlacu@irvapp.it

[†] Research Institute for the Evaluation of Public Policies, Bruno Kessler Foundation (FBK-IRVAPP), Via Santa Croce 77, 38122 Trento, Italy

We thank *AbilNova Cooperativa Sociale* and the participating schools for the fruitful collaboration. We would also like to thank Martina Bazzoli and Gianluca Bortoluzzi for their invaluable support during the data collection, and to Simone Valle for the excellent research assistance during the initial phases of the project. The research was made possible thanks to a grant from *Fondazione Caritro, Trento.* The study was pre-registered on the American Economic Association's registry for randomized controlled trials (ID number AEARCTR-0009131).

1 Introduction

Individuals with disabilities face a high risk of poverty and social exclusion, a reality that holds true also in areas with robust social protection policies, such as in Europe. While increasing employment has traditionally been the focus of policies targeting persons with disabilities, in recent years increased attention has been paid to other areas in need of improvement, which may even facilitate labor market integration, such as education. Youths with disabilities attain lower than average educational outcomes and face higher risks of early school leaving and not reaching a university degree (European Commission, 2021).

In the school year 2021/22 in Italy, the context of the current study, 3.8% of students in primary or secondary education experienced some form of disability (ISTAT, 2022). Nearly all were placed in regular classrooms (i.e. integrated classrooms). However, evidence indicates that being placed in integrated classrooms does not suffice to ensure the social inclusion of students with disabilities (Lindsay and Edwards, 2013). Negative attitudes of their peers, limited acceptance, and social interaction create a school climate that does not favor the academic development of students with disabilities (Nabors and Keyes, 1997; Martinez and Carspecken, 2007; Tavares, 2011). Moreover, teachers often lack the training to foster an inclusive classroom environment and may themselves have limited awareness and biased attitudes (Garrote et al., 2020; Ginevra et al., 2022). Finally, in most settings the focus is placed on the social inclusion of students with disabilities at their class level, ignoring the broader school context where they may be left excluded. In light of this, while the literature has been growing in recent decades, there is a need for evaluating interventions that aim to improve awareness, attitudes, and behaviors towards persons with disabilities (Lindsay and Edwards, 2013; Freer, 2023).[2]

This paper aims to contribute to this literature by evaluating the impact of a multi-faceted school intervention, implemented in upper-secondary schools, aimed at fostering inclusive behaviors of students by improving awareness of visual disabilities, perspective-taking capabilities and attitudes towards persons with visual disabilities. The intervention consisted of two activities: a curriculum class-based activity and a simulation with inter-group social contact activity - a meal in complete darkness, where students were served by and interacted with waiters with visual impairment. Each activity lasted for about one school hour.

The main evaluation follows a within-class randomized controlled trial design. 225 students from 12 classes in 3 schools were randomly assigned to receive the intervention at different moments in time, either two to three weeks before the follow-up survey (i.e. treatment group) or

[2]Out of the 42 papers included in the meta-analysis by Lindsay and Edwards (2013), only half have an experimental design and many have small sample sizes (less than 100 subjects). Furthermore, most studies measure only self-reported attitudes, thus potentially suffering from experimental demand effects. Finally, only a handful focus on visual impairment, the focus of the current study. In an updated review considering studies published between 2012 and 2019, Freer (2023) also highlights the fact that most studies measure effects only immediately post-intervention raising concerns regarding the sustainability of treatment effects.

in the weeks following it (i.e. control group). Given the multi-faceted nature of the intervention, a variety of methods were employed to measure the outcomes of interest, both at baseline and follow-up. As outcomes more immediately related to the intervention, we administered a quiz on knowledge of visual disabilities, developed jointly with the implementing organization, and a scale on perspective-taking and empathic concerns (Davis, 1980; Diotaiuti et al., 2021). We measured both implicit attitudes, through the Implicit Association Test (IAT Greenwald et al., 1998), and self-reported/explicit attitudes through the Attitude Toward Disabled Persons Scale (ATDP Yuker et al., 1970; Yuker and Block, 1986) and the Multidimensional Attitudes toward Disabled Persons Scale (MAS Findler et al., 2007). Finally, we used incentivized choices to measure several other dimensions of interest: willingness to interact socially with persons with visual disabilities, beliefs regarding the performance and outcomes of persons with visual disabilities in various domains, altruism towards persons with disabilities and social value orientation (Murphy et al., 2011; Murphy and Ackermann, 2014).

At baseline, only a subset of the outcomes were measured. Descriptive evidence indicates that students had limited basic knowledge of visual disabilities, responding correctly on average to half of the questions on the quiz. In terms of implicit attitudes, the median score on the IAT is roughly -0.7, indicating a far stronger mental association between being blind and negative words, and being sighted and positive words, compared to when the pairings are reversed. In terms of explicit attitudes, the mean scores on the scales indicate moderately positive attitudes towards persons with visual disabilities.

The results of the evaluation indicate large positive changes mainly in the dimensions more closely related to the activities of the intervention, knowledge of visual impairment, and empathic concerns/perspective-taking abilities. In terms of attitudes, results are mixed. While the ATDP is improved by the intervention, no positive changes are observed on any of the MAS sub-scales or on implicit attitudes. We interpret this latter result as students being more positive towards the integration of persons with visual disabilities: they are more in agreement with the fact that they are capable to experience lives qualitatively similar to those of sighted individuals and that they should be treated and interacted with in the same manner. However, they do not experience any changes in attitudes operating at the subconscious level and do not report imagining having different thoughts, emotions or behaviors in the context of a social interaction with a person with visual disabilities.

In terms of incentivized choices, no differences are observed across treatment arms. Participating in the interventions does not increase the willingness to pay (i.e. giving up part of the endowment) for the chance of having a social interaction with a person with visual disabilities. This result appears in line with the results on explicit attitudes measured through the MAS scale where students imagined a social interaction with a person with visual impairment. Next, while there is substantial variation in the reported beliefs, on neither of the dimensions studied we see any meaningful differences by treatment arms. In the dictator game, irrespective of treatment

status, students are willing to share on average about 40% of the endowment with an anonymous student with visual disabilities. Finally, not surprisingly given the previously stated results, we do not see any indirect effects of the intervention on more general dimensions of social preferences as measured through the Social Value Orientation and giving in the Dictator Game.

There are two potential concerns with the validity of the experimental design. First, treated students may be more prone to give the perceived socially desirable answers, especially in self-reported scales. Second, considering that the randomization was performed within-class, there is the risk of spillover effects on control group students, potentially masking non-null treatment effects. We provide suggestive evidence against both concerns. At follow-up, we administer a scale measuring the propensity for social desirability (Manganelli et al., 2000), and do not see significant difference by treatment arm. For what concerns the possibility of spillover effects, this was considered from the initial design of the intervention. In each school, two additional classes were assigned to serve as pure control classes. Though power is lower given the clustering of the class-level randomization, the analyses suggests that there are no relevant concerns regarding within-class spillovers.

The study has several limitations. First of all, even though the intervention is multi-faceted, it is only a one-time activity, which likely limits its effectiveness as longer duration or repeated interventions generally appear to be more effective (Lindsay and Edwards, 2013). Secondly, the sample size is somewhat limited, even though relatively large compared to many of the existing evaluation studies of related interventions. Thirdly, the incentives used in the games are not strong; only one student in each class was randomly drawn to receive the prize based on one of randomly drawn decisions in one the first three "games" (all except the beliefs elicitation). The beliefs elicitation game was a competition that awarded prizes to the top 3 students in the entire sample based on their guesses in one of the studied domains.

The study contributes to several strands of literature. First of all, we contribute to the body of evidence on the effectiveness of interventions aimed at improving knowledge and attitudes towards disabilities (Lindsay and Edwards, 2013). This is one of the few studies focusing on visual disabilities. To our knowledge, this is the first study measuring such a rich set of dimensions of attitudes and behaviors toward the social inclusion of persons with disabilities. While most studies focus on knowledge or explicit attitudes, we considered also implicit attitudes and incentivized choices to better map the potential causal chain. This is highly relevant, especially in the context of multifaceted interventions. For instance, while our intervention improves some explicit attitudes, it does not reduce the strong implicit bias, which may drive behaviors, especially under conditions of time or mental pressure, leading potentially to microaggressions during social interactions (Bertrand and Duflo, 2017). Secondly, the study adds to the literature on fields experiments studying discrimination (see Bertrand and Duflo, 2017, for a review). While our study does not specifically aim to distinguish between statistical (Phelps, 1972; Arrow, 1973) and taste-based discrimination (Becker, 1957), some of the measurement tools used better capture the

former or the latter. Thirdly, we contribute to the literature on intergroup contact (Allport, 1954; Pettigrew and Tropp, 2005; Paluck et al., 2019; Bursztyn et al., 2021). Ours is a particularly interesting setting given that most of the students in the sample reported not having any contact with persons with visual disabilities. Lastly, we contribute to the growing literature in economics using lab-in-the-field experiments to study social cohesion in the classroom (e.g. Rao et al., 2019; Alan et al., 2021). Outcomes measured through incentivized choices and games have been shown across various domains to be predictive of real-life outcomes (Franzen and Pointner, 2013; Potters and Stoop, 2016; Rao et al., 2019; Alan et al., 2021). The fact that the evaluated intervention affects attitudes towards the integration of persons with disabilities without affecting the willingness to interact socially, suggests that measuring only self-reported attitudes may not suffice to capture to what extent such interventions can improve the social inclusion and participation of persons with disabilities.

This paper is structured as follows. Section 2 presents the context and the program. Section 3 describes the evaluation design and outcomes, while Section 3 discusses the empirical strategy. Section 5 presents summary statistics, balance, and attrition checks, followed by the presentation of results in Section 6. Section 7 discusses potential spillover effects. Lastly, Section 8 concludes and discusses implications for research and policy.

2 Context and Program

2.1 Context

Beginning in the 70s, Italy initiated a series of major reforms towards the integration process of students with disabilities into mainstream schools and the gradual abolishment of special schools. In the following decades, Italy introduced various measures to support individuals with disabilities, including their right to inclusive education, and promoting their integration into mainstream schools. Each student with disabilities typically has an *Individualized Education Plan* that outlines their specific needs, goals, and the support services they receive. Schools are required to make physical and curricular accommodations to ensure that students with disabilities can access and participate in the educational environment. Teachers are trained on inclusive education while specialized educators can work with students who require additional support.

Despite the progress made in the past decades, the implementation of inclusive education faces several challenges and issues. Besides the ongoing presence of accessibility barriers, the training received by teachers may not suffice to support the needs of students with disabilities and to create a cohesive classroom environment. The negative attitudes of peers and teachers still represent a limiting factor to the integration of students with disabilities into mainstream education (Szumski et al., 2020). Furthermore, the focus is on social cohesion in the integrating classrooms with less attention being paid to social inclusion in the wider school context. For these reasons,

schools generally welcome interventions from NGOs operating in the field of inclusion of persons with disabilities. However, there is little evidence on the effectiveness of such interventions.

2.2 Program

The intervention was designed and delivered by a non-governmental organization operating in the field of visual impairment in the province of Trento, Northeast Italy. The NGO conducts training programs for individuals with visual impairment and implements educational initiatives focused on raising awareness and tackling stereotypes in schools and the broader community.

The intervention was composed of two activities which we refer to as the "informational treatment" and the "simulation treatment". The activities took place at school during regular school hours and lasted for about one school hour each, 50 minutes. The first activity was the informational treatment, delivered in class by a sighted facilitator. The aim of it was to increase knowledge and understanding of visual impairments and to address the questions and curiosities that students may have. Some of the topics covered were the following: eye disorders and different degrees of impairment, existing strategies, tools and opportunities that can be adopted to increase the autonomy of persons with visual impairment, how sighted people can (physically and verbally) aid persons with visual impairment to move around, and what jobs and sports persons with visual impairment can do. In general terms, the activity described what living with visual disabilities means, acknowledging the existing challenges and limitations (discrimination being one of them) while highlighting the strengths and opportunities available. The goal was to reduce negative stereotypes through information under the assumption that stereotypes often stem from incomplete knowledge which, in turn, depends on the social environment in which one lives (Roberts and Smith, 1999; Hunt and Hunt, 2004; Nowicki, 2006; Rillotta and Nettelbeck, 2007; Ison et al., 2010).

The simulation intervention was delivered in a mobile restaurant, a truck designed to operate as a restaurant in complete darkness. After a brief introduction, small groups of students were guided by blind waiters inside the restaurant, seated at the table, and offered a snack. During the school hour spent in complete darkness, students had the opportunity to interact with the blind waiters, deepening some of the insights gained from the previous activity or just chatting and asking questions. It is worth highlighting this aspect because it signals that this activity is more than a simulation experience, being in fact combined with inter-group contact which appears to be highly effective at improving attitudes towards people with disabilities (Armstrong et al., 2017). Simulation-based interventions are not always effective at reducing stereotypes, as indicated by a meta-analysis of 41 studies (Flower et al., 2007). There is even evidence that in some cases they may actually have unintended consequences (Silverman et al., 2015; Nario-Redmond et al., 2017). For instance, in Silverman et al. (2015) sighted students were asked to walk around a room blindfolded. This induced them to judge individuals with visual impairment as less capable to work or live independently, likely because the experience highlighted the initial challenges associated with an impairment, rather than the adaptations, competencies and strengths gained through living with

an impairment. This is precisely the reason why we opted for a multifaceted intervention. We hypothesized that by first providing information and allowing students to experience blindness first-hand while being guided by blind individuals (inter-group contact),[3] the complex capabilities that individuals with visual impairment have would become more salient than their limitations which are in many instances misperceived.

At the same time, we also expected that the implemented simulation treatment would enable participants to experience positive inter-group contact by concretely satisfying the key conditions identified in Allport (1954): equal status, common goals, no inter-group competition, and authority sanction.[4] Inter-group contact is widely accepted as an effective tool in reducing inter-group prejudice (Silverman et al., 2015; Kenworthy et al., 2005; Pettigrew and Tropp, 2005; Pettigrew et al., 2011; Armstrong et al., 2017) even though the magnitude of effects varies depending on the target group, larger effects being observed when the outgroup is represented by individuals with disabilities (Paluck et al., 2019). Several studies on diversity in education claim that positive inter-group contact and experiences improve perspective-taking and reduce prejudice (Aberson and Haag, 2007).

3 Evaluation design and outcomes

The program was implemented in three upper-secondary schools in the province of Trento, in North-Eastern Italy. The three schools represent the three main educational tracks in the Italian education system: academic, technical, and vocational. In the initial project phase, members of the implementing organization and research team introduced the program to representatives of each school (principals, administrative staff and teachers). Following this, each school recruited six classes satisfying two eligibility criteria: no prior exposure to a similar intervention and no students with visual impairment in the class.[5]

The evaluation design used is a hybrid - between and within class RCT design (Figure 1). The program uses a partial phased-in approach, all participating classes were guaranteed access to the intervention[6] but informed that the timing will be randomized, at the class level and then also at the individual level for a randomly selected sample of classes. Specifically, in each school, two

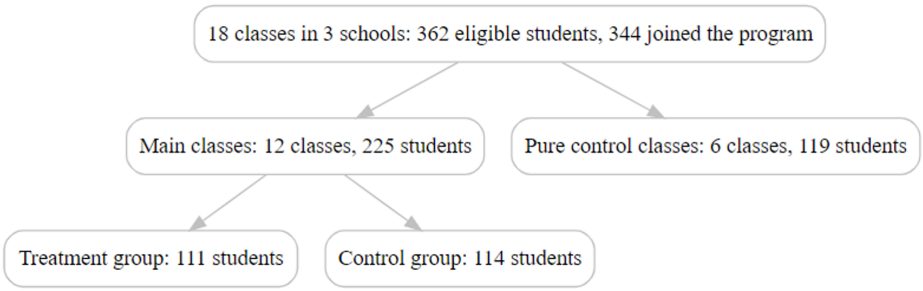
[3]Previous evaluations of multi-faceted interventions have shown that including a social contact component has the potential to improve their effectiveness (Favazza and Odom, 1997; Almond and Currie, 2011).

[4]Pettigrew and Tropp (2005) argue that the conditions are not essential for prejudice reduction while Paluck et al. (2019) highlight the lack of systematic research on the conditions suggested by Allport (1954)

[5]We expected classes with students with visual impairment to have on average higher knowledge of visual impairment, potentially also lower levels of prejudice due to higher inter-group contact and higher likelihood of having participated in related interventions. Given the relatively low number of students with visual impairment in the region, such classes would also not representative of the overall population. In addition, we would have faced the risk of having only one such class in the sample, making it difficult to have balance between the main classes and pure control classes.

[6]Initially, pure control classes were not considered to receive the intervention. However, the implementing organization succeeded in offering the intervention to them as well.

Figure 1: Evaluation Design



Notes: The diagram displays the evaluation design with the randomization of classes and students within the main classes.

classes were selected to serve as 'pure controls' - all students received the intervention only after the follow-up survey. Students in the other four classes were randomly assigned at the individual level to receive the intervention either in between the two surveys (treatment group) or after (henceforth main controls or control group). We will refer to them as the main classes. This hybrid evaluation design aims to maximize statistical power given the available budget[7] while also trying to provide some evidence on within-class spillover effects.

The timeline of the trial is as follows: In January 2022 classes were randomly assigned to either main classes or pure control classes. The baseline data was collected in February, 2022, followed by the individual-level randomization in the main classes, stratified at class level. Two to three weeks after the baseline, students assigned to the treatment group received the intervention, while the others remained in class. For logistical constraints, treated students absent on the day of the intervention were substituted with randomly chosen control students from the same class following a pre-determined waitlist design. After another two to three weeks, in March and April, the follow-up survey was implemented. Finally, students not in the treatment group were offered the intervention in the following days or weeks.[8]

The randomization of classes to main classes and pure control classes could not be effectively implemented in two of the schools. In one of the schools, the two pure controls classes were chosen by the principal based on the fact that could not participate in the intervention during the scheduled week due to being engaged in a field trip. In another school, post randomization of classes and baseline survey, we found out that two of the classes randomly selected as main classes did not satisfy one eligibility criteria, having a large share of students (75% and 85%) which participated in a similar simulation intervention in previous years. To avoid imbalances, one of the classes was randomly selected to be substituted with another randomly chosen class initially assigned to serve as pure control. In spite of this deviations from the initial protocol, there is good

[7]Randomizing treatment only at the class level was not feasible due to low statistical power even with small intra-cluster correlation.

[8]Students could participate in the survey and project activities only if their legal guardians signed the consent forms. About 95% percent of students received the consent to participate. During project activities, all the other students remained in the class with the teachers and were engaged in alternative activities.

level of balance between main and pure control classes, as it will be shown in Section 7.

Both surveys were conducted by the research team in the computer labs of the schools. The questionnaires were implemented using oTree (Chen et al., 2016) and required less than one school hour to be completed. Following a general verbal introduction and instructions, each student proceeded with the survey and received additional instructions and support if needed. Efforts were made to minimize interactions between students by distancing them whenever possible and closely monitoring those hesitant to follow the rules. In all games and tasks, we use the term "blind" in place of "with a visual impairment", even though blindness is only one form of visual impairment. We opted for this because we expected students to be much more familiar with the concept of "blindness", increasing awareness of other forms of visual impairment being actually a goal of the program. The trial was pre-registered on the AEA RCT Registry.[9] Throughout the analysis, we highlight any deviations from the pre-registration.

Most of the existing studies evaluating the effects of similar interventions focus primarily on impacts on knowledge and self-reported attitudes and only a few consider also implicit attitudes. This paper aims to take a step forward in this direction. The more proximal outcomes of the intervention are knowledge of visual impairment and perspective-taking and empathy. As a result of improvements in these dimensions, we expected beliefs and attitudes, explicit and implicit, to improve. Finally, as the most relevant primary outcome of the intervention, we expected an increase in the willingness to interact socially with persons with visual impairment. Despite all the outcomes considered, given the sample size and the nature of our multifaceted intervention establishing the detailed causal chain (the impact of each component of the intervention and underlying mechanisms) is not possible.

Before proceeding with the description of the outcomes, we provide a short methodological note on the scoring procedure for the outcome variables. With the exception of the IAT, all outcomes are normalized - transformed to vary between the theoretical minimum and maximum. In the case of several outcomes the scores represent shares, while for the outcome using Likert scales, the interpretation will depend on the number of response options of the scale.[10] In Appendix B, Tab Table B.2 we provide the Cronbach Alpha scores for all the psychometric scales administered at baseline and follow-up and generally find good reliability, lower only for the Empathic Concerns sub-scale and the Social Desirability scale.

3.1 Knowledge of visual impairments

As emphasized in section Section 3, the intervention aimed to fill a potential gap in knowledge of visual impairment that students in the targeted age group may have due to lack of formal instruction or interactions with individuals with visual impairment. The question is however how

[9] Available at the following link: `https://www.socialscienceregistry.org/trials/9131`

[10] For instance, with a Likert scale with 4 response options, a score of 0.66 would indicate that on average the respondent chose the 3rd response option on the scale.

much of the conveyed information was internalized and remembered after the intervention. With the support of the implementing organization, a multiple-choice quiz was administered both at baseline (shorter version) and at follow-up. A simple scoring procedure was applied: for each question, one point was awarded if the answer was correct and zero otherwise, thus wrong answers or "don't know" answers received no points. The number of correct answers is then summed up and divided by the number of questions, thus the score is the percentage of correct answers and varies between 0 and 1.

3.2 Perspective-taking and Empathic Concerns

Given that the intervention is in part an exercise in perspective-taking (i.e. simulation treatment), we aimed to explore whether changes in behaviors or attitudes may be mediated by an increased ability to perceive the states of mind of other individuals or to experience empathic concerns. Growing evidence points to the role of such abilities in social cooperation (e.g. see Batson et al., 1997; Galinsky and Moskowitz, 2000; Galinsky and Ku, 2004) and social cohesion (Kardos et al., 2018; Alan et al., 2021). We used the Perspective-taking and Empathic Concerns sub-scales of the Interpersonal Reactivity Index (Davis, 1980), validated on a sample of young adults in Italy in Diotaiuti et al. (2021). The scale is a 9-item Likert scale (5 perspective-taking and 4 empathic concerns items, respectively) with 6 response options ranging from "does not describe me at all" to "describes me fully". We explore the dimensions both separately and also as a single index.

3.3 Implicit attitudes

Implicit attitudes are measured through the Implicit Association Test (Greenwald et al., 1998), a computer-based test that assesses attitudes through the speed of association between pairings of concepts. The concepts used in our study were: "blind" and "sighted" as primary concepts; "positive", and "negative" as secondary concepts. Participants were presented with a series of stimuli, words, and images, displayed in the center of the screen, which they had to correctly categorize using the keyboard keys into one of the four concepts, displayed in the top left or top right of the screen. The response time when categorizing each stimulus is recorded. In the initial practice rounds, the primary and secondary concepts were displayed in isolation (e.g. blind on the left and sighted on the right), while in future rounds they are grouped and displayed together (e.g. blind or positive on the left, and sighted or negative on the right). The positions on the screen and the pairings are eventually switched. The implementation and scoring of the task follow the recommendations in Greenwald et al. (2003).

The IAT score indicates the strength and direction of a person's implicit bias. For instance, a negative score is recorded if a participant has a faster response time when "blind" is paired with "negative" and "sighted" is paired with "positive" compared to when the pairings are reversed (blind-positive and sighted-negative). In contrast, a positive score would indicate the reverse, while

a score close to zero indicates the absence of a differential implicit association. IAT has been used extensively in the past decades, and increasingly so as an outcome variable in interventions aimed at reducing discrimination. An early meta-analysis (Greenwald et al., 2009) indicated that the IAT appears to predict better actual behaviors relative to self-reported scales, which may suffer from social desirability bias. While a more recent meta-analysis by Oswald et al. (2013) toned down the findings of the previous one, the IAT appears to reflect something fundamental about psychological processes. The question remains to what extent such processes can shape behavior in a social context.[11] In spite of its issues of interpretation and reliability, Bertrand and Duflo (2017) argue that the IAT remains a useful tool to understand if and how interventions can affect attitudes and behaviors. In a recent study, the IAT has been used as a particularly effective de-biasing tool (see Alesina et al., 2018).

3.4 Explicit attitudes

The two measures adopted in this study for assessing explicit attitudes are the Attitude Toward Disabled Persons Scale (ATDP Yuker et al., 1970; Yuker and Block, 1986) and the Multidimensional Attitudes toward Disabled Persons Scale (MAS Findler et al., 2007). The ATDP is a unidimensional scale developed to assess generalized attitudes toward persons with disabilities. The scale was slightly adapted for this study[12] and includes 15 items from the ATDP-O scale in (Yuker et al., 1970), a 6-point Likert-type scale ranging from strongly disagree to strongly agree. The items generally capture the degree of agreement with the fact that persons with visual disabilities can lead lives similar to those without disabilities in terms of life quality and social interactions, and whether they should be treated in the same manner. The score is computed by first summing up the scores on the individual items, and then normalizing it to vary between 0 and 1 using the theoretical minimum and maximum. The items are recoded such that higher ratings represent more positive attitudes.

MAS is a multidimensional scale developed to assess different aspects of attitudes towards persons with disabilities: affective (emotional) states, cognitive beliefs, and behavioral tendencies. The MAS is based on a vignette, an indirect measure through which participants are expected to project their own emotions, thoughts, and behaviors in a given real-life situation. Following Findler et al. (2007) we randomly assigned whether participants had to imagine the situation as the interaction between a generic female "Anna" or male "Marco" and a blind person. We did not randomize also the gender of the blind person. Participants were asked to rate the likelihood that Anna/Marco would experience specific emotions (affect - 16 items), thoughts (cognition - 10 items), or whether they would display certain behaviors (8 items) when interacting with the blind person. The responses were given on a five-point Likert scale. The score of each sub-scale (Affect,

[11] Implicit attitudes appear to impact more micro-behaviors or behaviors under conditions of pressure (Rudman and Lee, 2002; McConnell and Leibold, 2001; Bertrand and Duflo, 2017).

[12] The scale was rephrased to apply specifically to persons with visual disabilities. In addition, few items were removed or slightly rephrased to better represent the local context.

Cognition, and Behavior) is computed following the same procedure as for the ATDP scale.[13]

3.5 Incentivized choices

In the follow-up survey, students were informed that they would be playing four games in which depending on their choices, they may win higher or smaller prizes. The general instructions were first explained out loud by a member of the research team and were available also in written form within the survey. Then students had to respond correctly to several comprehension questions on the general rules of the games. They could not proceed unless all the answers were correct and did not receive feedback on which answers were wrong or correct. This constrained students to re-read the rules carefully and to ask additional clarification questions. Comprehension questions were also included after the rules of the dictator games and the willingness to pay decisions. The order of the four games was the same for all students.[14]

One student per class was randomly selected at the end of the survey, through a public lottery, to be paid for one of their choices in one of the four games plus a fixed participation fee of 5 euros. The payoff was given in the form of a gift-card from a well-known sports equipment and clothing store. This detail was made very salient to the students, also through the use of comprehension questions, as may impact the interpretation of their choices in the dictator game.

3.5.1 Willingness to interact

One of the main goals of the program was to change preferences towards socially interacting with other students or persons with visual impairment. Since it was unrealistic to expect changes in real-life interactions over such a short period of time, we opted instead to measure experimentally the incentivized willingness to pay (WTP) for the opportunity to interact socially with a person with visual impairment. Students received an endowment of 14 euros and asked how much they would be willing to pay to participate in a short individual meeting at school with a person with visual impairment. For simplicity, we opted for a Multiple-price list (Andersen et al., 2006) instead of a BDM procedure.[15] Each student made 8 choices, prices varying from 0 to 14 euros in steps of 2. One of the choices was randomly selected and implemented for the winners in each class if this game was the one chosen for payment. The outcome is computed as the maximum share of the endowment the respondent is willing to pay for social interaction.[16]

[13] As in Findler et al. (2007), we find that the ATDP moderately correlates with the MAS sub-scales, more strongly with the affect and behavior sub-scales.

[14] The Dictator Game was played first followed by the Social Value Orientation task, the Willingness to Pay and beliefs elicitation.

[15] Rao (2019) opted for the willingness-to-accept procedure to measure the willingness of richer students to socially interact with poorer students. We opted for a WTP instead of WTA as we did not want students to feel that they were getting paid to spend time with a person with disability.

[16] Only two observations were dropped due to highly inconsistent choices. Students with left censoring, refusing also the 0 price, are coded as having the WTP -2.

3.5.2 Incentivized beliefs

The program aimed to potentially correct inaccurate beliefs, highlighting the strengths and competencies developed by individuals with visual impairment, while also acknowledging the possible limitations. In other words, despite the challenges, individuals with visual impairment can lead a normal and happy life and even exceed in various domains. We measured beliefs through incentivized guesses. Students were asked to guess the average outcomes of individuals with visual impairment in: (i) a sports competition (400 meters female sprint in the previous Paralympics), (ii) a memory test (values from Withagen et al., 2013), (iii, iv) a math test (summation and multiplication, values from Dormal et al., 2016) and (v) in a life-satisfaction survey (Brunes et al., 2021). After being explained the context of each guess, students were given as reference the average outcomes of sighted individuals and then made their guesses. One of the five guesses was randomly selected, and the three students in the sample closer to the true value received prizes of 30, 20 and 15 euros respectively. The order of the guesses was randomized at the individual level.[17]

3.5.3 Dictator games and social value orientation

Dictator games are traditionally used as a tool to measure altruism,[18] one dimension of social preferences. Participants are generally grouped in pairs, with one party playing the role of the dictator while the other of the recipient. The dictator is offered a given endowment and asked if and how much of it they would be willing to share part of it with the recipient. Usually, the identities of both parties are kept fully anonymous. A fully self-interested dictator would choose not to share anything and keep the entire endowment. However, empirically many individuals decide to share a positive amount across a wide range of cultures (Cochard et al., 2021).

In our game, all students in the sample played the role of the dictator and had to make three decisions: in each, they started with an initial endowment of 14 euros and asked if they would like to share part of it with another anonymous student, randomly selected from another school in the province. What changes between each decision is the type of recipient, the three categories being: a blind student (primary outcome), a generic student, or a student with motor impairment (secondary outcomes). The order of the first two categories of recipients was randomized at the individual level, while the third one was always played last. Students were informed that their identities would remain fully anonymous and that the recipient would receive the contribution also

[17] Asking students to make several guesses aimed also to reduce the risk of experimental demand effects, which could be present even if the choice was incentivized. With only one guess, students may expect that the choice of the guess by the researchers was intended to emphasize a particular domain where visual impairment may even provide an advantage. While this is true for the memory and math guesses, for the rest (sprint and life-satisfaction) observing that students guessed a more positive outcome would suggest either experimental demand effects or the fact that the nuanced overall message of the program was not well conveyed to the students. Throughout the analyses, we pay close attention to these dimensions.

[18] In practice giving may be influenced by several other factors, such as inequality aversion, social norms, preferences for the group to which the recipient belongs to, experimenter demand effects and several other factors

in the form of a gift card from a sports store. This last detail was made salient also through comprehension questions.

What would a treatment effect on the amount shared with a blind recipient measure? As stated previously, the dictator game with anonymous parties is expected to measure altruism. However, in the context of a vulnerable group, altruism could also be seen as a form of charitable giving motivated by the perceived distress experienced by the other group. Such a motivation runs counter to the message delivered through the intervention which aimed to make students more aware also of the strengths and capabilities of persons with visual impairment. Thus, if charitable giving is the main motivation, one could expect the treatment to even lower giving to a blind student. We aimed at limiting this through the choice of the sports store gift card. We expected students to share their endowment with a blind student if they believed they could be able to use it when practicing sports. Thus, a positive treatment effect may also reflect a change in beliefs concerning the perceived capabilities of blind individuals. At the same time, a positive treatment effect may also indicate a change in tastes. Finally, for the positive effect to be observed these drivers have to be stronger than any reduction in charitable giving motivations.

Through the other two decisions - sharing with a generic or a student with motor impairment - we aimed to measure possible positive or negative spillovers. Specifically, we wanted to understand (i) if the intervention fosters or hinders altruism in general/towards other vulnerable groups, or (ii) if it changes the beliefs regarding the capabilities of other vulnerable groups to practice sports. The outcomes are computed as the share of the endowment sent to each recipient.

To further deepen possible changes in social preferences, following the dictator task students also filled a simplified social value orientation task (Murphy et al., 2011; Murphy and Ackermann, 2014), where they had to choose between various allocations of prizes for themselves and other anonymous students. Even though the slider measured proposed in Murphy et al. (2011) offers a more granular and precise measurement, for simplicity, we opted for binary choices given the risk of students struggling to understand the task and that this is only a secondary outcome. We will explore both the SVO score computed using the procedure in Murphy et al. (2011), and also look at each individual choice. In particular, we aim to explore the intervention made students more prosocial or altruistic as opposed to individualistic or competitive. The four types are more likely to maximize: the payoff of the other (altruistic), the total payoff of both parties (prosocial), own payoff (individualistic) and own payoff relative to the payoff of the other (competitive). In addition to the six main choices, students made an additional choice between two third-party allocations for two randomly chosen anonymous students (adapted from Rao et al., 2013).

3.6 Social desirability

One major threat to the internal validity of the study is that differences in attitudes and behaviors between groups are driven in fact by differences in social desirability. All students were aware of

which was the underlying topic of the research project and thus possibly prone to anticipate what would a socially desirable behavior be and report to comply with such behaviors. The risk is that in addition to that, through participation in the intervention, treated students become even more prone to give socially desirable answers. This is particularly the case in the self-reported scales, less so in the incentivized choices, and likely doesn't play a major role in the IAT or knowledge test. While we cannot directly test for social desirability in each scale, we can explore differences in general propensity to give socially desirable answers. At the end of the follow-up survey, students were administered the brief version (9 items) of the Crowne and Marlowe (1960) validated in the Italian context in Manganelli et al. (2000). Students reported on a scale from 1 "Absolutely false" to 6 "Absolutely true" to what extent the statements apply to themselves (e.g. "It doesn't matter with whom I speak, I am always a good listener"). The answers are summed up (reversed wherever needed), with higher values indicating a higher propensity to give socially desirable answers.

4 Empirical strategy

4.1 Main analytical sample and variable construction

We restrict the main analyses to the main classes (i.e. randomized within-class to receive the treatment). In section Section 7, we return to the "pure control" classes when addressing potential spillovers, discussing the specific empirical strategy in the dedicated section. The analytical sample is composed of students present at follow-up. Among them, for those with missing baseline data, we impute missing covariate values with the sample mean for continuous variables and adding the missing level to categorical variables (i.e. creating the category "Missing").

4.2 Estimator

Several classes in two of the three participating schools had imperfect treatment compliance (i.e. students absent during the day of the intervention). For logistical reasons,[19] we adopted a waitlist design: treated students absent on the day of the intervention were substituted with control group students selected at random. The waitlists were generated before the actual intervention, including also control group students eventually absent during the day of the intervention.[20] With this design, we can opt between two estimators: the initial-offer estimator (IO) and the ever-offer estimator (EO) (de Chaisemartin and Behaghel, 2019) to estimate Intention-to-treat effects (ITT) and Local Average Treatment Effects (LATE).

The IO estimator considers as assigned to treatment students allocated initially to be

[19]Since eventually all control group students could participate in the intervention, given the space constraints in the truck restaurant, the implementing organization requested treating half of the class during the intervention day and the other half after the follow-up.

[20]Students on the waitlists absent during the intervention are considered as being offered the treatment if they were supposed to substitute a treated student.

offered the treatment, prior to making substitutions using the waitlists. Assuming no "defiers", those absent for the intervention ("never takers") make-up the non-compliers group together with those who were randomly chosen and accepted to replace the absent students ("always takers"). Differently, the EO estimator considers as offered the treatment all students ever invited to take part in the intervention (initially allocated and substitutes), including those absent during the intervention if they were among the treated or among the randomly chosen substitutes. In contrast to the IO, de Chaisemartin and Behaghel (2019) show that the EO estimator is not consistent without adjusting for the fact that, by construction, the share of takers will be higher among the treatment group, since substitutes stop when the last available spot is filled. In given settings, the doubly-robust EO (DREO) proposed by de Chaisemartin and Behaghel (2019) can even be more efficient. For simplicity and given that both IO and DREO estimators are unbiased, in the main body we run the analyses (including balance and attrition checks) using the IO estimator. In the Appendix, we replicate the main analyses using the DREO estimator and only comment on them briefly in the main body.

For some of the outcomes, in particular, those with empirical distributions displaying fat tails or shapes distant from the Gaussian, we perform also non-parametric tests.

4.3 Empirical specification

The main empirical specification used throughout the analysis is the following:

$$y_i = \beta_0 + \beta_1 Z_i + X_i'\gamma + \delta + \epsilon_i \tag{1}$$

where y_i is the outcome of interest for student i, Z_i is the indicator of treatment assignment, X_i' is a vector of covariates (socio-demographics, other characteristics at baseline including the outcome measured at baseline whenever available, and outcome specific covariates wherever applicable)[21] δ is a vector of strata fixed effects (school fixed effects). ϵ_i is the error term. The estimate of intention to treat effect (ITT) is the estimated β_1 . Given the lack of full compliance, we also report local-average-treatment-effects (LATE) estimates for the main outcomes.

We do not run or report any analysis of heterogeneity given the lack of power but make the data available for meta-analyses or exploratory studies.

5 Data: Descriptive Statistics, Balance and Attrition check

With the exception of the incentivized choices, PT-EC scale, and social desirability, all outcomes were measured both at baseline and at follow-up. Basic demographics were also measured at baseline (age, gender, highest level of education of parents or legal guardians), prior exposure

[21] An example of outcome specific covariate is the order of the pairing of categories in the IAT.

to similar interventions and level of contact with blind individuals (at home, at school, in the neighborhood or friends in general). At follow-up, we recovered this information for students absent at baseline. Table 1 presents the descriptive statistics and balance checks by treatment status - the initial allocation to the treatment group.

Table 1: Descriptive statistics and balance checks

	(1) N	(2) Control Mean (SD)	(3) Treated Mean (SD)	(4) Difference [p-value]
Socio-demographics				
Male	220	0.49 (0.5)	0.45 (0.5)	-0.01 [0.87]
Female	220	0.5 (0.5)	0.54 (0.5)	0.01 [0.91]
Gender Other	220	0.01 (0.09)	0.01 (0.1)	0 [0.86]
Age	220	18.1 (0.86)	18.4 (1.29)	0.25 [0.03]
Education: tertiary or above	220	0.33 (0.47)	0.33 (0.47)	0.01 [0.93]
Education: upper secondary	220	0.41 (0.49)	0.41 (0.49)	0.01 [0.89]
Education: at most lower secondary	220	0.16 (0.37)	0.2 (0.4)	0.02 [0.7]
Education: refused to answer	220	0.11 (0.31)	0.07 (0.25)	-0.03 [0.35]
Speaks mainly Italian at home	220	0.88 (0.33)	0.85 (0.36)	-0.03 [0.55]
Prior dined in the dark	220	0.2 (0.4)	0.25 (0.44)	0.04 [0.4]
Has any contact with blind persons	220	0.25 (0.43)	0.21 (0.41)	-0.05 [0.41]
Outcome variables				
Knowledge Quiz	202	0.54 (0.15)	0.53 (0.15)	-0.02 [0.42]
Implicit Association Test	201	-0.7 (0.37)	-0.64 (0.4)	0.04 [0.42]
MAS Affect	202	0.69 (0.17)	0.68 (0.16)	0 [0.95]
MAS Cognition	202	0.68 (0.18)	0.69 (0.16)	0.01 [0.61]
MAS Behavior	202	0.76 (0.16)	0.78 (0.15)	0.01 [0.59]
ATDP	202	0.72 (0.15)	0.73 (0.13)	0 [0.8]

Note: Columns 2 and 3 display the means and standard deviations (in parentheses) of baseline variables in the control and treatment groups. Column 4 shows the estimated difference in means which is obtained from regressing the variable of interest on the treatment dummy and strata (school) fixed effects, and using robust standard errors. P-values are reported in brackets. All baseline outcomes except the Implicit Association Test are normalized (0-1). The higher N for the socio-demographic variables is due to information being recovered in the follow-up survey for students absent at baseline. Prior dined in the dark indicates whether the student had previously participated in a similar simulation experience. Education indicates the highest level of education attained by the parents or legal guardians of the student. ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities. The sample used in the analysis is composed of students present for the follow-up survey.

The average age of students in our sample is roughly 18, slightly higher among the treated, the only variable different with a p-value below 10%. While in two of the schools, the participating classes are predominantly female or male, the overall sample has a balanced gender representation, with about 52% females and 47% males. 23% of students participated in the past in a lunch or dinner in complete darkness. The same proportion of students reports having some form of contact with individuals with visual disabilities have any form of contact with individuals with visual disabilities (at school, at home, in the neighborhood, or any other type of friendship). For roughly 33% of students, the highest level of education completed by their parents or legal

guardians is tertiary education. The vast majority of students, 86% report speaking Italian or an Italian dialect at home, with the remainder speaking other languages.

Moving on to the outcome variables measured at baseline, we note first that students correctly responded to about 54% of the questions on the quiz. The average value of the IAT is -0.67, close to the -0.7 lower threshold for strong preference for sighted over blind persons. The explicit attitudes scales, ATDP and MAS (3 sub-scales), have normalized average values ranging from 0.68 to about 0.77, indicating slight to moderately positive attitudes towards persons with visual impairment. All outcomes are very well balanced across arms as shown also Figure A.1 in the Appendix which plots the empirical distributions of the variables by treatment arm.

The difference in the rate of attrition at follow-up, reported in Table A.1 in Section A, is relatively small. Students in the treatment group are 3 percentage points less likely to be absent at follow-up relative to 13% attrition in the control group (p-value = 0.45). In addition to this, in Table B.1 in Section B we show that the groups remain very well balanced also when restricting the sample to students present at follow-up.

6 Results

In the following, we examine treatment effects on students' choices and outcomes, following the order presented in Section 3, that is based on the expected proximity to the treatment in the theory of change.

6.1 Treatment effects on knowledge of visual impairments, perspective-taking and empathic concerns

Table 2 reports the ITT and LATE estimates of the effects of the intervention on knowledge of visual impairment and perspective-taking/empathic concerns.

At baseline, students responded correctly to 54% of the items on the quiz, on average. Across all items in the follow-up test, we estimate an ITT of 0.048 percentage points (ES = 0.3, p-value = 0.013). The results indicate that part of the information conveyed during the intervention was internalized and recalled 2 to 3 weeks after. At the same time, they also suggest that, at least on this dimension, it is unlikely that there were substantial positive spillover effects given the stable observed scores for the control group across time.[22]

Positive effects are observed on the Perspective-taking (PT) and Empathic Concerns (EC) dimensions, larger and estimated precisely only on the latter. The ITT on the combined scale is roughly 0.044 (ES = 0.283, p-value = 0.052). The larger effect on empathic concerns than on

[22]The follow-up quiz was identical with the addition of six items. Across items repeated at baseline and follow-up, the performance of control group students is unchanged, while the treated improved on average by roughly 4 percentage points.

Table 2: Treatment effects on visual impairment knowledge, perspective-taking and empathic concerns (initial offer)

	(1) Knowledge Quiz	(2) PT	(3) EC	(4) PT-EC
ITT	0.048	0.031	0.057	0.044
	(0.019)	(0.031)	(0.025)	(0.023)
	[0.013]	[0.315]	[0.022]	[0.052]
LATE	0.062	0.04	0.075	0.057
	(0.025)	(0.04)	(0.032)	(0.029)
	[0.013]	[0.311]	[0.02]	[0.047]
Control Mean	0.558	0.637	0.658	0.647
Control SD	0.161	0.201	0.17	0.156
Observations	199	179	179	179

Note: ITT indicates intention to treat effects, estimated via OLS regressions. LATE indicates local average treatment effects obtained through 2SLS regressions, instrumenting actual treatment take-up with the treatment offer. Robust standard errors in parentheses. P-values in brackets. The model controls for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics). The outcome variables are the scores on: the visual impairment knowledge test (Knowledge Quiz), the perspective-taking sub-scale (PT), empathetic-concerns sub-scale (EC) and the combined PT-EC scale. The lower number of observation for some of the outcomes is due to a class not completing the entire survey in the allocated time.

perspective-taking was not a target of the intervention. However, an increase in empathic concerns could be a channel contributing to an improvement in attitudes and behaviors.

6.2 Treatment effects on implicit and explicit attitudes

The intervention appears not to be effective at changing implicit attitudes, as measured through the IAT.[23] The first column in Table 3 reports the estimated treatment effects.[24]. Though imprecise (p-value = 0.133), the ITT is actually negative (ITT = -0.08; ES = -0.182). The specific features of the intervention - the fact that the simulation treatment is preceded by the information treatment and includes also social interaction with the waiters with visual impairment - does not offset what has been observed in previous studies (Silverman et al., 2015; Nario-Redmond et al., 2017). Future studies could investigate if repeated exposure to such interventions could mitigate any initial negative shocks and eventually convert to positive effects also at an implicit level.

Column 2 in Table 3 presents the estimated treatment effects on the uni-dimensional ATDP scale, while Columns 3 to 5 show the estimates on the 3 MAS sub-scales: affect, cognition,

[23]Students did not receive their IAT scores in either survey as it could have an impact on behaviors (see Alesina et al., 2018).

[24]In addition to the standard set of covariates used throughout the analyses, the model also controls for the randomized ordering of the four categories across rounds in the test (4 possible orderings) in both the baseline and the follow-up test.

Table 3: Treatment effects on implicit and explicit attitudes (initial offer)

			MAS		
	(1)	(2)	(3)	(4)	(5)
	IAT	ATDP	Affect	Cognition	Behavior
ITT	-0.08	0.037	0	-0.011	0.011
	(0.053)	(0.018)	(0.021)	(0.025)	(0.027)
	[0.133]	[0.039]	[0.994]	[0.666]	[0.669]
LATE	-0.102	0.049	0	-0.014	0.015
	(0.069)	(0.024)	(0.027)	(0.032)	(0.034)
	[0.14]	[0.042]	[0.994]	[0.667]	[0.669]
Control Mean	-0.634	0.739	0.681	0.677	0.738
Control SD	0.438	0.155	0.146	0.2	0.186
Observations	197	179	179	179	179

Note: ITT indicates intention to treat effects, estimated via OLS regressions. LATE indicates local average treatment effects obtained through 2SLS regressions, instrumenting actual treatment take-up with the treatment offer. Robust standard errors in parentheses. P-values in brackets. The model controls for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics). The outcome variables are the scores on: Implicit Attitudes Test (IAT), the Attitudes Towards Disabled People scale (ATDP) scale, and the 3 sub-scales of the Multidimensional Attitudes Scale Toward Persons With Disabilities (MAS) scale. With the exception of the IAT, all outcomes are normalized (0-1). The lower number of observation for some of the outcomes is due to a class not completing the entire survey in the allocated time.

and behavior. All outcomes are normalized (0 and 1 indicating the theoretical minimum and maximum) with higher values indicating more positive attitudes. Results point towards large positive effects on ATDP (ITT = 0.037; ES = 0.241; p-value = 0.039) and no effects on any of the MAS sub-scales. Treated students are more likely to agree with the fact that blind individuals can have lives similar to those without disabilities in terms of life quality and social interactions, and that they should be treated in the same manner. In contrast, they are not more likely to report more inclusive thoughts, feelings or behaviors when imagining a social interaction context.

6.3 Treatment effects on incentivized choices

6.3.1 Willingness to interact, giving in the dictator game, and social value orientation

We start by exploring treatment effects on the willingness to interact with a blind person, the main outcome of interest aiming to proxy changes in actual social behaviors. Column 1 in Table 4 reports the ITT and LATE estimates on the maximum share of the 14 euros endowment that students would be willing to pay to participate in an individual meeting with an anonymous blind person at school. Results indicate that the intervention did not contribute to an increased willingness to socially interact with blind individuals (p-value = 0.453).[25] Given the censored

[25]Confirmed by a Kolmogorov-Smirnov test (p-value = 0.61).

shape of the distribution of the outcome, especially at the upper threshold (25% would give up the entire endowment, the modal value), we also estimate a Tobit model, which confirms the previous findings, and the effects on the probability of giving up the entire endowment.[26] The treatment group is almost 8 percentage points less likely to agree giving up the entire endowment, a somewhat large effect relative to the 32% share in the control group though we are not powered to detect such effects. One possible explanation is that treated students already had contact with individuals with visual disabilities during the simulation treatment and as a result they may have a lower marginal utility compared to control group students. However, the goal of the intervention was to impact the utility function which does not appear to be the case, at least as measured through the willingness to pay for social interaction.

Table 4: Treatment effects on willingness to interact socially, giving in the Dictator Game and Social Value (initial offer)

		Giving in the DG (%)			
	(1)	(2)	(3)	(4)	(5)
	WTP	Visual Impairment	Motor Impairment	Generic	SVO
ITT	-0.042	0.025	0.021	0.036	-0.068
	(0.056)	(0.037)	(0.039)	(0.033)	(3.061)
	[0.453]	[0.509]	[0.582]	[0.279]	[0.982]
LATE	-0.054	0.031	0.027	0.045	-0.087
	(0.072)	(0.048)	(0.05)	(0.042)	(3.911)
	[0.452]	[0.51]	[0.583]	[0.281]	[0.982]
Control Mean	0.538	0.375	0.425	0.289	27.454
Control SD	0.379	0.249	0.267	0.218	20.894
Observations	197	199	199	199	199

Note: ITT indicates intention to treat effects, estimated via OLS regressions. LATE indicates local average treatment effects obtained through 2SLS regressions, instrumenting actual treatment take-up with the treatment offer. Robust standard errors in parentheses. P-values in brackets. The model controls for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics). The outcome variables are the willingness to pay for social interaction with a persons with visual impairment, expressed as a share of the endowment (WTP), the share of endowment allocated to three anonymous recipients (students) - one with visual impairment, one with motor impairment and one generic - in the Dictator Game (DG), and the score on the Social Value Orientation Task (SVO).

As highlighted in the Section 3, giving towards a persons with visual disabilities in the Dictator Game has a nuanced interpretation, and could be seen as a form of altruism which acknowledges the capability of the recipient to perform sports, since the prize could only be used to purchase goods in a sports store. Treatment effects estimates are reported in Table 4, Column 2. Giving in the control group, as a share of the endowment, is roughly 37.5% on average. The share is slightly higher in the treated group but the estimates are very imprecise (p-value = 0.509). Columns

[26] Results available upon request.

3 and 4 report the estimates on giving when the recipient is a student with motor impairment and a generic anonymous recipient, respectively. The two dimensions represent secondary outcomes aimed at measuring potential indirect effects on generosity towards groups not directly targeted by the intervention. Similar to the estimates presented previously, we observe positive yet imprecise estimates with p-values higher than 0.25. Interestingly, even though the payoff was in the form of a gift card in a sports store, students show higher generosity on average towards recipients with disabilities (visual or motor) than towards a generic recipient.

Social value orientation. Column 5 in Table 4 shows the estimated effects on the indicator of social value orientation (Murphy et al., 2011). Higher values of the indicator indicate more prosocial and altruistic attitudes and less individualistic or competitive. The average social value orientation of students in the sample lies within the interval of prosociality. We do not see any difference between the two treatment arms. In Table B.4 in Appendix B, we zoom in and look at each individual choice. Students in the treatment are not consistently more likely to choose the more prosocial or altruistic allocations.

6.3.2 Incentivized beliefs

Since the empirical distribution of beliefs deviates from the Gaussian distribution substantially for some of the choices, we investigate potential treatment effects non-parametrically. Figure A.4 in Section A plots kernel densities by treatment arm for the 5 guesses in the four studied domains (sprint, memory, life satisfaction and math skills). The figure plots also the average values for sighted individuals which were offered as references to students when making the guess, and the average values for individuals with visual impairment which they had to guess. In all dimensions, we observe a strong overlap between the empirical distributions of the two groups. For all outcomes, the median values by group are almost identical. Smirnov-Kolmonrov tests fail to reject the null hypothesis, with associated p-values larger than 0.5 for all outcomes. Interestingly, the median guesses are generally very close to the true values for some outcomes (memory, sprint, life satisfaction), while for the others they appear anchored to the known reference values for sighted individuals. We repeated the above-mentioned analyses also considering as outcomes the absolute deviations of the guesses from the average value for sighted individuals and for blind individuals. Not surprisingly, giving the strong overlap between the distributions shown in Figure A.4, we do not observe any differences by treatment arm.

6.4 Social desirability

We move on to asessing whether there are signs of differences in social desirability by treatment arm. Given the fact that, besides the effect on knowledge, we only observe positive effects on self-reported scales (ATDP and PT-EC), and no or imprecise effects IAT and incentivized choices, the issue of social desirability differential by treatment arm becomes particularly relevant. We

find treated students to show slightly higher social desirability (IT = 0.014, ES = 0.093), effect is imprecisely estimated (p-value = 0.531) and small relative to the main treatment effects detected. Nonetheless, it does point out that the positive effects on the ATDP and PT-EC may be an upper bound of the true effects. Indeed, including the social desirability index in the main models as an endogenous control, reduces the estimated effects of the treatment, but only by about 10% for the ATDP and by about 20% for the PT-EC scale.

7 Spillover Effects

In this section, we aim to provide evidence on the potential existence of spillover effects within the main classes. The premise is that at least in part the effects of the intervention on the treated students may have been transferred to their classmates, making the estimated previously most likely lower bounds, in absolute terms, of the true treatment effect. Negative spillover effects may also exist: the outcomes of the control group may have been pushed in the opposite direction (e.g. worsened attitudes) due to the fact that they did not participate in the intervention. We consider this less likely, given that students were aware that everyone would receive the opportunity of taking part in the intervention but at different times and who would do it earlier or later would be chosen at random.

In practice, we restrict the analysis to control group students, that is, control students in the main classes and the pure control classes, and run a similar specification as in Equation 1 on all outcomes, using as reference group the pure controls.[27] Given the class-level assignment, we cluster standard errors at the class level. However, considering the small number of clusters (Cameron and Miller, 2015), we also run Fischer's permutation tests and provide exact p-values. This second approach also has a caveat. Due to deviations from the class randomization protocol in two of the schools (see Section 3), even though the initial assignment process is known, the final assignment is not, condition required to run permutation tests. To limit this issue, we also perform a baseline check and as before, we include a wide range of covariates, for most outcomes this means including also their baseline level.

Before discussing potential spillover effects, we first test for differential attrition and the differences in observable characteristics across main controls and pure controls. In terms of attrition, pure controls are 6 percentage points less likely to be present at follow (86% relative to 80%), imprecisely estimated difference with a p-value of 0.25. In terms of baseline across observable characteristics (Table A.3 in Section A), there are 3 variables with associated p-values close to or below 0.1 - the share of students having participated in the past in a lunch or dinner in the dark - 21 percentage points higher in the pure control group, ATDP and knoledge quiz score - both higher

[27] Due to the waitlist design used, the share of compliers may not be balanced between the two groups, smaller for the controls in the main classes. We repeat the analyses considering the last student on the waitlist accepting the treatment as control student to recover potential upper bounds of spillover effects. Results are consistent with the ones reported and available upon request.

in the within group. All the other variables appear very well-balanced across groups. Nonetheless, as for the main models, we control for all these dimensions in the estimations.

Table A.4 in Appendix A presents the results for all dimensions besides beliefs which are discussed in the following paragraph. For the dimensions where within-class treatment effects were observed (quiz, ATDP, and PT-EC) we do not see evidence of spillover effects, neither negative which would imply that the measured treatment effects are upper biased, nor positive which would imply the reverse. The only two dimensions with spillover estimates having p-values in the lower range are IAT and giving in the Dictator Game, in particular giving towards other generic students. The estimates on the IAT are positive with p-values close to 0.1. This would suggest that the imprecise negative effects observed in Section 6 may be biased upwards. Speculatively, one could imagine that talking about the treatment with peers who experienced it can impact implicit attitudes. However, more likely this difference is due to chance. The negative spillover effects on giving, in particular the imprecisely estimated effects on giving towards a generic recipient, would indicate that the positive and imprecise estimates presented in Table 4 are biased upwards.

Turning to the incentivized beliefs, we repeat the non-parametric analyses performed for the main treatment effects analyses, reported in Section 6. Again, we do not see significant differences between the groups. The only difference that appears more pronounced is in the distribution of guesses regarding the reported life satisfaction of individuals with visual impairment. While the median guess is similar, the guesses of students in pure control classes tend to be closer to the true average value (Kolmogorov-Smirnov test p-value = 0.18).

Overall we do not see robust evidence of spillover effects within class. However, we cannot exclude the potential existence of spillover across classes, even though this is unlikely given that classes were often from different grades and tracks. Nonetheless, we can investigate, for the outcomes measured before and after the intervention, if there is evidence of change across time. While seeing a change across time may not necessarily reflect across-class spillover effects and instead reflect natural changes in time or some sort of placebo effects (simply being part in a study on visual impairment may trigger changes in attitudes), the tests may still be informative, especially if no change is observed across time. We perform two-sample paired t-tests for the within-class controls and pure controls present at baseline and follow-up, using the outcomes measured in both survey waves: quiz, IAT and MAS. There appear to be no relevant differences across time in the IAT or the Cognition dimension of MAS. However, there is small increases in the quiz score (possibly due to test-retest effects), whereas the ADTP score is significantly larger at follow-up. Conversely, the other two dimension of the MAS scale (cognition and behavior) are actually significantly lower at follow-up. Based on the different directions of the changes, it is unlikely they can be attributed to spillovers.

8 Conclusion

We assess the effectiveness of a multi-faceted intervention, implemented upper secondary schools, aimed at improving the social inclusion of persons with visual impairment. The intervention consisted of activities that have been shown to be generally effective in previous studies. First, students were engaged in a curriculum class-based activity focused on improving awareness and knowledge of visual impairment. The second activity combines a simulation (perspective-taking) activity with inter-group contact: students had a meal in complete darkness, in a dedicated mobile restaurant where they were served by and interacted with waiters with visual impairment. 225 students in 12 classes from 3 schools were randomly assigned, at individual level, to receive the intervention. Additionally, students from other 6 classes were assigned to serve as pure controls to investigate potential within-class spillovers.

The experimental evaluation reveals significant positive effects of the intervention primarily in dimensions that are more closely related to the intervention's core activities, such as knowledge of visual impairment, perspective-taking and empathic concerns. The intervention also improved generalized attitudes, treated students being more prone to report that persons with visual impairment are not dissimilar to persons without disabilities, can similarly lead fulfilling lives and should be treated in the same manner. However, the intervention was not effective at reducing the strong implicit bias documented at baseline, suggesting that while explicit attitudes may evolve, subconscious perceptions are harder to tackle.

The intervention did not improve students' willingness to socially interact with persons with disabilities, neither when measured through the incentivized willingness to pay for social interaction or through a multidimensional scale of attitudes, assessed through a vignette depicting a social interaction with a person with visual impairment. Furthermore, no effects are observed on incentivized beliefs regarding the performance and outcomes of persons with visual disabilities in various domains, or on altruism towards them, measured through the Dictator Game. Lastly, we do not find indirect effects on broader social preferences, including social value orientation and giving in the Dictator Game towards other groups.

The results of the paper have implications both for research and policy. Most of the existing studies evaluating the impacts of interventions in the field of disability inclusion assess their impacts only on knowledge and generalized attitudes. Our results show that positive effects on these dimensions may not be mirrored by improvements in implicit attitudes or in other dimensions that may be closely related to behaviors, such as the willingness to interact socially with persons with visual impairment. This raises questions regarding the potential of such interventions to improve inclusive behaviors, especially considering the fact that explicit attitudes have been shown to only moderately correlate with actual behaviors. While we can only speculate, our results appear to suggest that the intervention may foster support for inclusion policies but may not trigger social contact. From a policy perspective, the study emphasizes the need for policy experimentation for

the social inclusion of persons with visual disabilities.

References

Aberson, C. L. and S. C. Haag (2007). Contact, perspective taking, and anxiety as predictors of stereotype endorsement, explicit attitudes, and implicit attitudes. *Group Processes & Intergroup Relations 10*(2), 179–201.

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021). Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics 136*(4), 2147–2194.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018, December). Revealing Stereotypes: Evidence from Immigrants in Schools.

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

Almond, D. and J. Currie (2011). *Human Capital Development before Age Five*, Volume 4 of *Handbook of Labor Economics*, Chapter 15, pp. 1315–1486. Elsevier.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006, December). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383–405.

Armstrong, M., C. Morris, C. Abraham, and M. Tarrant (2017). Interventions utilising contact with people with disabilities to improve children's attitudes towards disability: A systematic review and meta-analysis. *Disability and health journal 10*(1), 11–22.

Arrow, K. J. (1973). The theory of discrimination. 3–33 in discrimination in labor markets, edited by orley ashenfelter.

Batson, C. D., M. P. Polycarpou, E. Harmon-Jones, H. J. Imhoff, E. C. Mitchener, L. L. Bednar, and T. R. Klein (1997). Empathy-induced altruism: A research perspective. *The Empathy-Altruism Hypothesis 35*(1), 51–62.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In *Handbook of Economic Field Experiments*, Volume 1, pp. 309–393.

Brunes, A., M. B. Hansen, and T. Heir (2021). General self-efficacy in individuals with visual impairment compared with the general population. *PLoS One 16*(7), e0254043.

Bursztyn, L., T. Chaney, T. A. Hassan, and A. Rao (2021). The immigrant next door: Long-term contact, generosity, and prejudice. Technical report, National Bureau of Economic Research.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources 50*(2), 317–372.

Chen, D. L., M. Schonger, and C. Wickens (2016, March). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Cochard, F., J. Le Gallo, N. Georgantzis, and J.-C. Tisserand (2021). Social preferences across different populations: Meta-analyses on the ultimatum game and dictator game. *Journal of Behavioral and Experimental Economics 90*, 101613.

Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology 24*(4), 349.

Davis, M. H. (1980). Interpersonal reactivity index.

de Chaisemartin, C. and L. Behaghel (2019, September). Estimating the Effect of Treatments Allocated by Randomized Waiting Lists.

Diotaiuti, P., G. Valente, S. Mancone, A. Grambone, and A. Chirico (2021). Metric goodness and measurement invariance of the italian brief version of interpersonal reactivity index: A study with young adults. *Frontiers in Psychology 12*, 773363.

Dormal, V., V. Crollen, C. Baumans, F. Lepore, and O. Collignon (2016). Early but not late blindness leads to enhanced arithmetic and working memory abilities. *Cortex 83*, 212–221.

European Commission (2021). Union of equality: Strategy for the rights of persons with disabilities 2021-2030. `https://ec.europa.eu/social/main.jsp?catId=1484&langId=en/`.

Favazza, P. C. and S. L. Odom (1997). Promoting positive attitudes of kindergarten-age children toward people with disabilities. *Exceptional children 63*(3), 405–418.

Findler, L., N. Vilchinsky, and S. Werner (2007). The multidimensional attitudes scale toward persons with disabilities (mas): Construction and validation. *Rehabilitation Counseling Bulletin 50*, 166–176.

Flower, A., M. K. Burns, and N. A. Bottsford-Miller (2007). Meta-analysis of disability simulation research. *Remedial and Special Education 28*(2), 72–79.

Franzen, A. and S. Pointner (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics 16*, 155–169.

Freer, J. R. (2023). Students' attitudes toward disability: A systematic literature review (2012–2019). *International Journal of Inclusive Education 27*(5), 652–670.

Galinsky, A. D. and G. Ku (2004). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology 86*(5), 713–727.

Galinsky, A. D. and G. B. Moskowitz (2000). The professional identity of a "street cop": A defense against physical threats. *The Journal of Research in Personality 34*(4), 427–447.

Garrote, A., F. Felder, H. Krähenmann, S. Schnepel, R. S. Dessemontet, R. S. Dessemontet, and E. M. Opitz (2020). Social acceptance in inclusive classrooms: The role of teacher attitudes toward inclusion and classroom management. *Frontiers in Education*.

Ginevra, M. C., I. Di Maggio, I. Valbusa, S. Santilli, and L. Nota (2022). Teachers' attitudes towards students with disabilities: The role of the type of information provided in the students' profiles of children with disabilities. *European Journal of Special Needs Education 37*(3), 357–370.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74*(6), 1464.

Greenwald, A. G., B. A. Nosek, and M. R. Banaji (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology 85*(2), 197.

Greenwald, A. G., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji (2009). Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology 97*(1), 17.

Hunt, C. S. and B. Hunt (2004). Changing attitudes toward people with disabilities: Experimenting with an educational intervention. *Journal of Managerial Issues*, 266–280.

Ison, N., S. McIntyre, S. Rothery, H. Smithers-Sheedy, S. Goldsmith, S. Parsonage, and L. Foy (2010). 'just like you': A disability awareness programme for children that enhanced knowledge, attitudes and acceptance: Pilot study findings. *Developmental Neurorehabilitation 13*(5), 360–368.

ISTAT (2022). L'inclusione scolastica degli alunni con disabilità (eng. the educational inclusion of pupils with disabilities). `https://www.istat.it/it/files/2022/12/Alunni-con-disabilita-AS-2021-2022.pdf`.

Kardos, P., G. Kezdi, and A. Simonovits (2018). Religion and support for public policies: Evidence from a representative survey in hungary. *European Sociological Review 34*(3), 271–283.

Kenworthy, J. B., R. N. Turner, M. Hewstone, and A. Voci (2005). Intergroup contact: When does it work, and why? In J. F. Dovidio, O. P. John, P. Glick, and L. A. Rudman (Eds.), *On the Nature of Prejudice: Fifty Years After Allport*, pp. 278–292. Malden: Blackwell Publishing.

Lindsay, S. and A. Edwards (2013). A systematic review of disability awareness interventions for children and youth. *Disability and Rehabilitation 35*, 623–646.

Manganelli, A.-M., L. Canova, and R. Marcorin (2000). La desiderabilità sociale. un'analisi di forme brevi della scala di marlowe e crowne [social desirability. analysis of the short forms of marlowe's and crowne's scales]. *Testing, Psicometria, Metodologia 7*, 5–16.

Martinez, R. S. and P. Carspecken (2007). Effectiveness of a brief intervention on latino children's social acceptance of peers with special needs. *Journal of Applied School Psychology 23*(1), 97–115.

McConnell, A. R. and J. M. Leibold (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology 37*(5), 435–442.

Murphy, R. O. and K. A. Ackermann (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review 18*(1), 13–41.

Murphy, R. O., K. A. Ackermann, and M. J. Handgraaf (2011). Measuring social value orientation. *Judgment and Decision making 6*(8), 771–781.

Nabors, L. A. and L. A. Keyes (1997). Brief report: Preschoolers' social preferences for interacting with peers with physical differences. *Journal of Pediatric Psychology 22*(1), 113–122.

Nario-Redmond, M. R., D. Gospodinov, and A. Cobb (2017). Crip for a day: The unintended negative consequences of disability simulations. *Rehabilitation Psychology 62*(3), 324.

Nowicki, E. (2006). A cross-sectional multivariate analysis of children's attitudes towards disabilities. *Journal of Intellectual Disability Research 50*(5), 335–348.

Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock (2013). Predicting ethnic and racial discrimination: a meta-analysis of iat criterion studies. *Journal of personality and social psychology 105*(2), 171.

Paluck, E. L., S. A. Green, and D. P. Green (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy 3*(2), 129–158.

Pettigrew, T. F. and L. R. Tropp (2005). Allport's intergroup contact hypothesis: Its history and influence. In J. F. Dovidio, O. P. John, P. Glick, and L. A. Rudman (Eds.), *On the Nature of Prejudice: Fifty Years After Allport*, pp. 262–277. Malden: Blackwell Publishing.

Pettigrew, T. F., L. R. Tropp, U. Wagner, and O. Christ (2011). Recent advances in intergroup contact theory. *International journal of intercultural relations 35*(3), 271–280.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review 62*(4), 659–661.

Potters, J. and J. Stoop (2016). Do cheaters in the lab also cheat in the field. *European Economic Review*.

Rao, G. (2019, March). Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools. *American Economic Review 109*(3), 774–809.

Rao, G. et al. (2013). Familiarity does not breed contempt: Diversity, discrimination and generosity in delhi schools. *Job Market Paper*.

Rao, G. et al. (2019). Familiarity does not breed contempt: Diversity, discrimination and generosity in delhi schools. *AER Forthcoming*.

Rillotta, F. and T. Nettelbeck (2007). Effects of an awareness program on attitudes of students without an intellectual disability towards persons with an intellectual disability. *Journal of Intellectual and Developmental Disability 32*, 19–27.

Roberts, C. and P. Smith (1999). Attitudes and behaviour of children toward peers with disabilities. *International Journal of Disability, Development and Education 46*, 35–50.

Rudman, L. A. and M. R. Lee (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes & Intergroup Relations 5*(2), 133–150.

Silverman, A. M., J. D. Gwinn, and L. Van Boven (2015). Stumbling in their shoes: Disability simulations reduce judged capabilities of disabled people. *Social Psychological and Personality Science 6*(4), 464–471.

Szumski, G., J. Smogorzewska, and P. Grygiel (2020). Attitudes of students toward people with disabilities, moral identity and inclusive education—a two-level analysis. *Research in Developmental Disabilities 102*, 103685.

Tavares, W. (2011). An evaluation of the kids are kids disability awareness program: Increasing social inclusion among children with physical disabilities. *Journal of Social Work in Disability and Rehabilitation*.

Withagen, A., A. M. Kappers, M. P. Vervloed, H. Knoors, and L. Verhoeven (2013). Short term memory and working memory in blind versus sighted children. *Research in developmental disabilities 34*(7), 2161–2172.

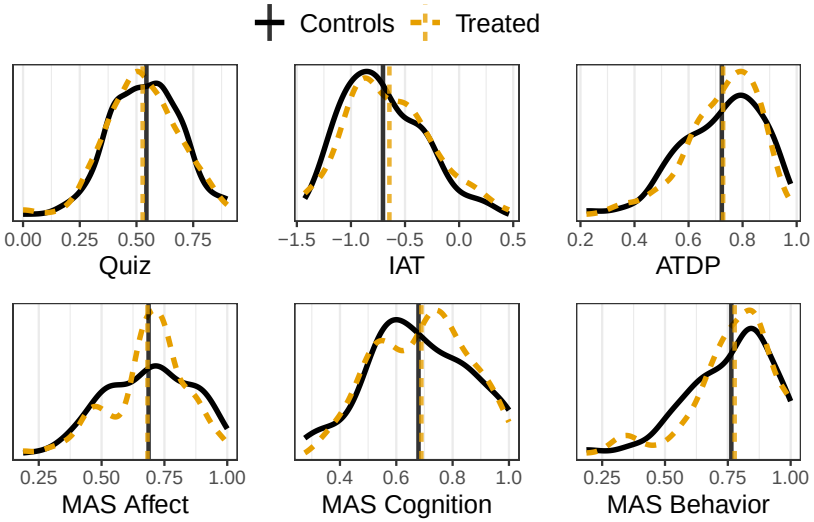
Yuker, H. E. et al. (1970). The measurement of attitudes toward disabled persons.

Yuker, H. E. and J. R. Block (1986). Persons scales (atdp) 1960-1985. *The Measurement of Intergroup Attitudes 32*, 145–154.

Appendices

A Appendix

Figure A.1: Kernel densities by treatment status at baseline (initial offer)



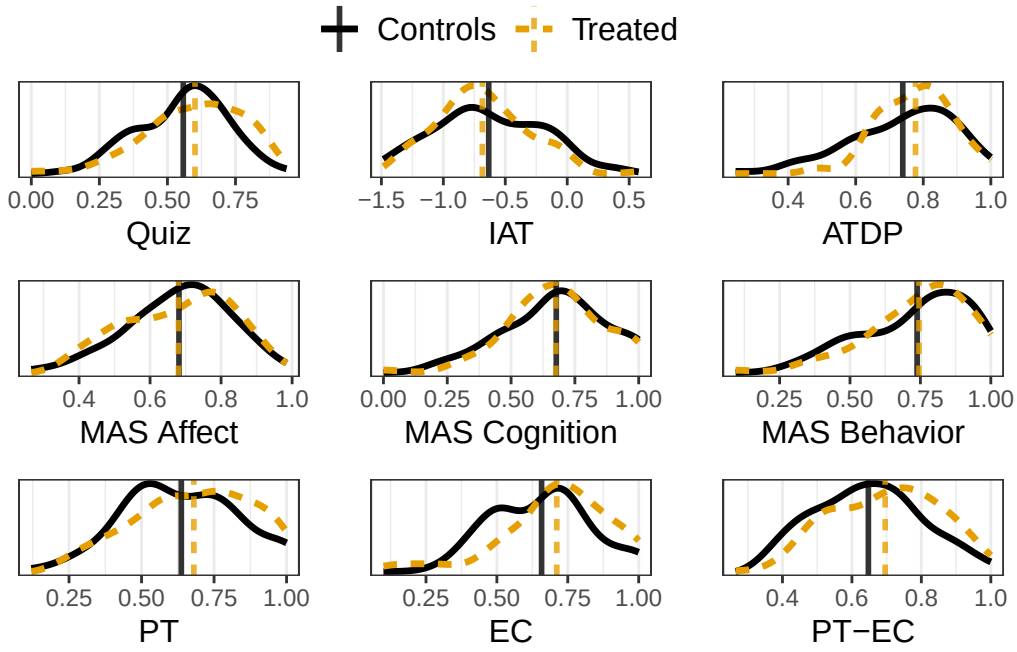
Notes: The vertical bars indicate the means by treatment status. The outcome variables are the scores on: the visual impairment knowledge test (Quiz), the Implicit Association Test (IAT), the Attitudes Towards Disabled People scale (ATDP) scale, and the 3 sub-scales of the Multidimensional Attitudes Scale Toward Persons With Disabilities (MAS) scale.

Table A.1: Attrition by treatment status (initial offer)

	(1) Control Mean	(2) Estimate (SE) [p-value]
Initial Offer	0.13	-0.03 (0.04) [0.45]

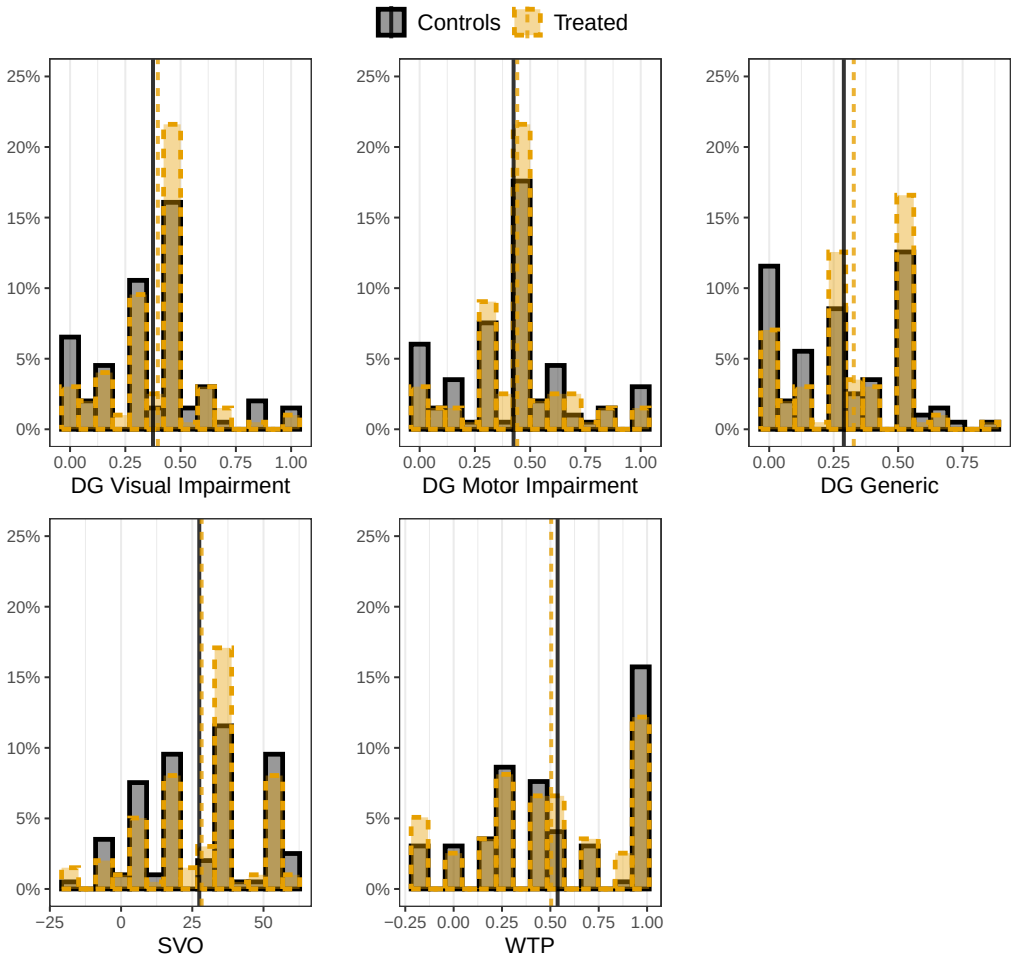
Note: Column 1 report the mean of the control group. The estimate in column 2 is obtained through OLS regression. Robust standard errors in parentheses. P-values in brackets.

Figure A.2: Kernel densities by treatment status at follow-up (initial offer)



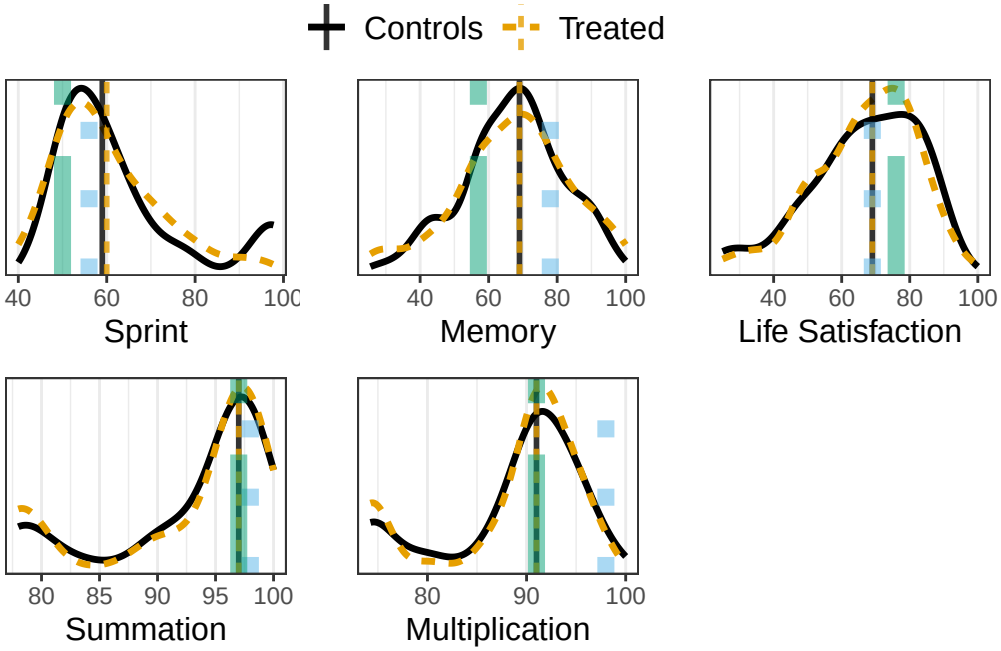
Notes: The vertical bars indicate the means by treatment status. The outcome variables are the scores on: the visual impairment knowledge test (Quiz), the Implicit Association Test (IAT), the Attitudes Towards Disabled People scale (ATDP) scale, and the 3 sub-scales of the Multidimensional Attitudes Scale Toward Persons With Disabilities (MAS) scale, the perspective-taking sub-scale (PT), empathetic-concerns sub-scale (EC) and the combined PT-EC scale.

Figure A.3: Histograms by treatment status at follow-up (initial offer)



Notes: The vertical bars indicate the means by treatment status. The outcome variables are the willingness to pay for social interaction with a persons with visual impairment, expressed as a share of the endowment (WTP), the share of endowment allocated to three anonymous recipients (students) - one with visual impairment, one with motor impairment and one generic - in the Dictator Game (DG), and the score on the Social Value Orientation Task (SVO).

Figure A.4: Kernel densities by treatment status at baseline (initial offer)



Notes: The figure plots kernel densities of the incentivized beliefs, guesses of students regarding the average performance of individuals with visual impairment in various tasks, knowing the average performance of sighted individuals in the same tasks. The thinner vertical lines indicate the median guess by treatment arms. The thicker vertical lines indicate the true average values for: sighted individuals (long-dashed green line) and individuals with visual impairment (dotted blue line). The ploted distributions are censored for values above the median + 1.5 x Inter-quartile-range (IQR) and below the median - 1.5xIQR.

Table A.2: Comparison between initial-offer (IO) and ever-offer (EO) estimates

	Initial Offer (IO)		Ever Offer (EO)	
	(1)	(2)	(3)	(4)
	ITT	LATE	ITT	LATE
Knowledge Quiz	0.048	0.062	0.069	0.079
	(0.019)	(0.024)	(0.021)	(0.023)
Perspective-Taking (PT)	0.031	0.04	0.065	0.075
	(0.031)	(0.04)	(0.03)	(0.034)
Empathic Concern (EC)	0.057	0.075	0.073	0.084
	(0.025)	(0.032)	(0.024)	(0.029)
PT-EC	0.044	0.057	0.069	0.08
	(0.022)	(0.029)	(0.025)	(0.029)
Implicit Association Test	-0.08	-0.102	-0.047	-0.053
	(0.053)	(0.069)	(0.067)	(0.077)
ATDP	0.037	0.049	0.019	0.022
	(0.018)	(0.024)	(0.017)	(0.018)
MAS Affect	0	0	0.001	0.001
	(0.021)	(0.027)	(0.015)	(0.017)
MAS Cognition	-0.011	-0.014	0.015	0.017
	(0.025)	(0.032)	(0.019)	(0.021)
MAS Behavior	0.011	0.015	0.021	0.024
	(0.027)	(0.034)	(0.02)	(0.023)
DG Visual Impairment	0.025	0.031	0.035	0.04
	(0.037)	(0.048)	(0.031)	(0.035)
DG Motor Impairment	0.021	0.027	0.005	0.006
	(0.039)	(0.05)	(0.045)	(0.051)
DG Generic	0.035	0.045	0.054	0.062
	(0.033)	(0.042)	(0.037)	(0.042)
Social Value Orientation	-0.068	-0.087	1.977	2.244
	(3.06)	(3.91)	(3.804)	(4.316)
Willingness to Pay	-0.042	-0.054	-0.064	-0.073
	(0.056)	(0.072)	(0.062)	(0.073)

Note: Columns 1 and 2 report the ITT and LATE estimates obtained through OLS and 2SLS using as treatment assignment the initial offer, reported also in the results section. Columns 3 and 4 report the doubly-reweighted ever-offer (DREO) ITT and LATE estimates considers as assigned to treatment those receiving the initial offer and also those on the waitlist who eventually received, or were supposed to receive the offer. ITT indicates intention to treat effects, while LATE indicates local average treatment effects obtained by instrumenting actual treatment take-up with the treatment offer. All models control for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics). Standard errors in parentheses. The outcome variables are listed in the row column (ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities, DG indicates choices in the Dictator Game.

Table A.3: Descriptive statistics and balance checks across within-class and pure controls

	(1)	(2)	(3)	(4)
	N	Pure Control Mean (SD)	Within Control Mean (SD)	Difference [p-value]
Socio-demographics				
Female	215	0.45 (0.5)	0.48 (0.5)	0.09 [0.26]
Age	215	18.15 (1.39)	18.1 (0.86)	-0.03 [0.94]
Education upper secondary	215	0.37 (0.48)	0.42 (0.5)	0.04 [0.41]
Education at most lower secondary	215	0.21 (0.41)	0.13 (0.34)	-0.08 [0.16]
Education refused to answer	215	0.08 (0.27)	0.1 (0.3)	0.02 [0.61]
Speaks Italian at home	215	0.86 (0.35)	0.9 (0.3)	0.03 [0.5]
Prior dined in the dark	215	0.37 (0.48)	0.16 (0.37)	-0.18 [0.09]
Has any contact with blind persons	215	0.15 (0.36)	0.23 (0.43)	0.09 [0.19]
Outcome variables				
Knowledge Quiz	199	0.5 (0.16)	0.55 (0.15)	0.05 [0.1]
Implicit Association Test	199	-0.72 (0.32)	-0.73 (0.38)	0 [0.97]
MAS Affect	199	0.7 (0.16)	0.69 (0.18)	-0.02 [0.25]
MAS Cognition	199	0.69 (0.17)	0.67 (0.18)	-0.01 [0.51]
MAS Behavior	199	0.77 (0.14)	0.76 (0.16)	-0.01 [0.6]
ATDP	199	0.69 (0.15)	0.72 (0.15)	0.04 [0.06]

Note: Columns 2 and 3 display the means and standard deviations (in parentheses) of baseline variables in the control and treatment groups. Column 4 shows the estimated difference in means which is obtained from regressing the variable of interest on the treatment dummy and strata (school) fixed effects, and using robust standard errors. P-values are reported in brackets. All baseline outcomes except the Implicit Association Test are normalized (0-1). The higher N for the socio-demographic variables is due to information being recovered in the follow-up survey for students absent at baseline. Prior dined in the dark indicates whether the student had previously participated in a similar simulation experience. Education indicates the highest level of education attained by the parents or legal guardians of the student. ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities. The sample used in the analysis is composed of students present for the follow-up survey.

Table A.4: Spillover analyses - difference betweenwithin-class and pure controls

	(1) Estimate (SE)	(2) p-value	(3) Fisher's p-value	(4) ES	(5) N
Knowledge Quiz	-0.019 (0.013)	0.16	0.213	-0.137	180
Perspective-Taking (PT)	-0.016 (0.03)	0.608	0.686	-0.082	166
Empathic Concern (EC)	-0.028 (0.028)	0.322	0.484	-0.145	166
PT-EC	-0.022 (0.025)	0.386	0.457	-0.142	166
ATDP	-0.01 (0.02)	0.62	0.631	-0.072	166
MAS Affect	0.018 (0.012)	0.161	0.332	0.123	166
MAS Cognition	-0.014 (0.033)	0.675	0.72	-0.083	166
MAS Behavior	-0.012 (0.029)	0.675	0.663	-0.073	166
Implicit Association Test	0.098 (0.054)	0.085	0.115	0.273	179
Willingness to Pay	-0.04 (0.06)	0.51	0.515	-0.117	178
DG Visual Impairment	-0.052 (0.034)	0.143	0.243	-0.204	180
DG Motor Impairment	-0.042 (0.039)	0.3	0.298	-0.157	180
DG Generic	-0.077 (0.018)	0.001	0.001	-0.346	180
Social Value Orientation	-3.397 (3.182)	0.301	0.3	-0.177	180

Note: Columns 2 reports the estimates obtained through OLS on an indicator of which takes the value 1 for within-class controls (control students in classes where the intervention was randomized at individual level) and 0 for pure controls (classes in which no student was treated). The same column reports clusterred standard errors in parentheses and associated p-values in square brackets, and Fisher's exact test p-values in curly brackets. ES reports effect sizes computed using the means and standard deviations of the control groups. All models control for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics). Standard errors in parentheses. The outcome variables are listed in the row column (ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities, DG indicates choices in the Dictator Game).

B Online Appendix

Table B.1: Descriptive statistics and balance checks for the sample present at follow-up

	(1) N	(2) Control Mean (SD)	(3) Treated Mean (SD)	(4) Difference [p-value]
Socio-demographics				
Male	220	0.49 (0.5)	0.45 (0.5)	-0.01 [0.87]
Female	220	0.5 (0.5)	0.54 (0.5)	0.01 [0.91]
Gender Other	220	0.01 (0.09)	0.01 (0.1)	0 [0.86]
Age	220	18.1 (0.86)	18.4 (1.29)	0.25 [0.03]
Education: tertiary or above	220	0.33 (0.47)	0.33 (0.47)	0.01 [0.93]
Education: upper secondary	220	0.41 (0.49)	0.41 (0.49)	0.01 [0.89]
Education: at most lower secondary	220	0.16 (0.37)	0.2 (0.4)	0.02 [0.7]
Education: refused to answer	220	0.11 (0.31)	0.07 (0.25)	-0.03 [0.35]
Speaks mainly Italian at home	220	0.88 (0.33)	0.85 (0.36)	-0.03 [0.55]
Prior dined in the dark	220	0.2 (0.4)	0.25 (0.44)	0.04 [0.4]
Has any contact with blind persons	220	0.25 (0.43)	0.21 (0.41)	-0.05 [0.41]
Outcome variables				
Knowledge Quiz	202	0.54 (0.15)	0.53 (0.15)	-0.02 [0.42]
Implicit Association Test	201	-0.7 (0.37)	-0.64 (0.4)	0.04 [0.42]
MAS Affect	202	0.69 (0.17)	0.68 (0.16)	0 [0.95]
MAS Cognition	202	0.68 (0.18)	0.69 (0.16)	0.01 [0.61]
MAS Behavior	202	0.76 (0.16)	0.78 (0.15)	0.01 [0.59]
ATDP	202	0.72 (0.15)	0.73 (0.13)	0 [0.8]

Note: Columns 2 and 3 display the means and standard deviations (in parentheses) of baseline variables in the control and treatment groups. Column 4 shows the estimated difference in means which is obtained from regressing the variable of interest on the treatment dummy and strata (school) fixed effects, and using robust standard errors. P-values are reported in brackets. All baseline outcomes except the Implicit Association Test are normalized (0-1). The higher N for the socio-demographic variables is due to information being recovered in the follow-up survey for students absent at baseline. Prior dined in the dark indicates whether the student had previously participated in a similar simulation experience. Education indicates the highest level of education attained by the parents or legal guardians of the student. ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities. The sample used in the analysis is composed of students present for the follow-up survey.

Table B.2: Cronbach alpha reliability scores at baseline and follow-up

	Cronbach Alpha	
	(1)	(2)
Scale	Baseline	Follow-up
MAS Affect	0.88	0.87
MAS Cognition	0.87	0.91
MAS Behavior	0.82	0.86
ATDP	0.79	0.82
Perspective-taking (PT)		0.82
Empathic concerns (EC)		0.58
PT-EC		0.77
Social desirability scale		0.68

Note: Cronbach alpha scores at baseline and follow-up reported in the columns. The sample used is composed of the pupils present for the data collection in the 18 participating classes. Missing values indicate that a given outcome was not administered also at baseline. ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities

Table B.3: Descriptive statistics and balance checks across within-class and pure controls for the sample present at follow-up

	(1)	(2)	(3)	(4)
	N	Pure Control Mean (SD)	Within Control Mean (SD)	Difference [p-value]
Socio-demographics				
Female	179	0.43 (0.5)	0.48 (0.5)	0.1 [0.28]
Age	179	18.2 (1.46)	18.04 (0.83)	-0.19 [0.69]
Education upper secondary	179	0.34 (0.48)	0.4 (0.49)	0.06 [0.31]
Education at most lower secondary	179	0.21 (0.41)	0.12 (0.33)	-0.11 [0.05]
Education refused to answer	179	0.07 (0.26)	0.11 (0.31)	0.02 [0.7]
Speaks Italian at home	179	0.85 (0.36)	0.88 (0.33)	0.02 [0.64]
Prior dined in the dark	179	0.38 (0.49)	0.18 (0.39)	-0.17 [0.13]
Has any contact with blind persons	179	0.16 (0.37)	0.24 (0.43)	0.1 [0.17]
Outcome variables				
Knowledge Quiz	163	0.51 (0.16)	0.56 (0.14)	0.05 [0.06]
Implicit Association Test	163	-0.73 (0.32)	-0.7 (0.39)	0.04 [0.58]
MAS Affect	163	0.71 (0.15)	0.7 (0.16)	-0.02 [0.5]
MAS Cognition	163	0.69 (0.17)	0.68 (0.18)	-0.01 [0.57]
MAS Behavior	163	0.78 (0.14)	0.75 (0.17)	-0.03 [0.23]
ATDP	163	0.69 (0.14)	0.71 (0.15)	0.03 [0.18]

Note: Columns 2 and 3 display the means and standard deviations (in parentheses) of baseline variables in the control and treatment groups. Column 4 shows the estimated difference in means which is obtained from regressing the variable of interest on the treatment dummy and strata (school) fixed effects, and using robust standard errors. P-values are reported in brackets. All baseline outcomes except the Implicit Association Test are normalized (0-1). The higher N for the socio-demographic variables is due to information being recovered in the follow-up survey for students absent at baseline. Prior dined in the dark indicates whether the student had previously participated in a similar simulation experience. Education indicates the highest level of education attained by the parents or legal guardians of the student. ATDP - Attitudes Towards Disabled People scale, MAS - Multidimensional Attitudes Scale Toward Persons With Disabilities. The sample used in the analysis is composed of students present for the follow-up survey. The sample used in the analysis is composed of students present for the follow-up survey.

Table B.4: Treatment effects on choices between the allocation in the social value orientation task (initial offer)

(Payoff self, Payoff other)						
		(1)	(2)	(3)	(4)	(5)
Allocation A	Allocation B	Estimand	Estimate (SE) [p-value]	Control Mean (SD)	ES	N
(7,7)	(7,1)	ITT	0.048 (0.044) [0.279]	0.879 (0.328)	0.146	199
		LATE	0.061 (0.056) [0.281]	0.879 (0.328)	0.186	199
(7,7)	(7,1)	ITT	0.026 (0.044) [0.548]	0.879 (0.328)	0.080	199
		LATE	0.034 (0.056) [0.549]	0.879 (0.328)	0.103	199
(7,1)	(9,4)	ITT	0.055 (0.037) [0.141]	0.051 (0.22)	0.249	199
		LATE	0.07 (0.047) [0.143]	0.051 (0.22)	0.317	199
(7,1)	(9,4)	ITT	0.044 (0.034) [0.205]	0.051 (0.22)	0.198	199
		LATE	0.056 (0.044) [0.206]	0.051 (0.22)	0.253	199
(4,9)	(7,7)	ITT	-0.049 (0.041) [0.23]	0.121 (0.328)	-0.149	199
		LATE	-0.062 (0.052) [0.229]	0.121 (0.328)	-0.190	199
(4,9)	(7,7)	ITT	-0.049 (0.039) [0.21]	0.121 (0.328)	-0.150	199
		LATE	-0.063 (0.05) [0.209]	0.121 (0.328)	-0.192	199
(4,9)	(7,1)	ITT	0.053 (0.072) [0.464]	0.525 (0.502)	0.106	199
		LATE	0.068 (0.092) [0.464]	0.525 (0.502)	0.135	199
(4,9)	(7,1)	ITT	0.037 (0.077) [0.636]	0.525 (0.502)	0.073	199
		LATE	0.047 (0.099) [0.635]	0.525 (0.502)	0.093	199
(4,9)	(9,4)	ITT	-0.014 (0.062) [0.821]	0.273 (0.448)	-0.031	199
		LATE	-0.018 (0.079) [0.821]	0.273 (0.448)	-0.040	199
(4,9)	(9,4)	ITT	-0.028 (0.066) [0.671]	0.273 (0.448)	-0.063	199
		LATE	-0.036 (0.084) [0.671]	0.273 (0.448)	-0.080	199
(9,4)	(7,7)	ITT	-0.009 (0.066) [0.893]	0.313 (0.466)	-0.019	199
		LATE	-0.011 (0.084) [0.893]	0.313 (0.466)	-0.024	199
(9,4)	(7,7)	ITT	0.007 (0.067) [0.917]	0.313 (0.466)	0.015	199
		LATE	0.009 (0.085) [0.917]	0.313 (0.466)	0.019	199
Passive choices - allocations of payoffs for two anonymous students						
(4,4)	(8,2)	ITT	-0.086 (0.05) [0.085]	0.899 (0.303)	-0.283	199
		LATE	-0.109 (0.064) [0.087]	0.899 (0.303)	-0.360	199
(4,4)	(8,2)	ITT	-0.077 (0.05) [0.13]	0.899 (0.303)	-0.253	199
		LATE	-0.098 (0.064) [0.13]	0.899 (0.303)	-0.323	199

Note: Estimates obtained via OLS regressions using as outcome variables are indicators of the choices in the social value orientation task between the allocations listed in the first two columns - 1 for Allocation A and 0 for Allocation B - the first value in the parentheses indicate the payoff for self while the second the payoff for the other player (with the exception of the last choice where the decision is passive and involves two other players). Robust standard errors in parentheses. P-values in brackets. ITT indicates intention to treat effects, while LATE indicates local average treatment effects obtained through 2SLS regressions, instrumenting actual treatment take-up with the treatment offer. The model controls for strata fixed effects and individual level covariates (baseline values of the outcome variables, outcome specific covariates wherever applicable and socio-demographics).