American Economic Journal: Economic Policy 2025, 17(3): 107–130
https://doi.org/10.1257/pol.20230620

Transmitting Rights: Effective Cooperation, Inter-gender Contact, and Student Achievement[†]

By Sultan Mehmood, Shaheen Naseer, and Daniel L. Chen*

We provide experimental evidence of teacher-to-student transmission of gender attitudes in Pakistan. We randomly show teachers a pro-women's rights visual narrative. Treated teachers increase their and students' support for women's rights, unbiasedness in gender implicit association tests (IATs), and willingness to petition parliament for greater gender equality. Students improve coordination and cooperation with the opposite gender. Effects are larger when teachers teach a gender-rights curriculum. Mathematics achievement increases for classrooms assigned to form mixed-gender study groups treated with an intense program (visual narrative and curriculum), while absent in same-sex study groups. Gender attitudes are transmissible and cooperation improves student outcomes. (*JEL* I21, I28, J16, K38, O15)

A teacher affects eternity.

—Henry Adams in *A Letter to Teachers* (1910)

How are rights revolutions transmitted in societies? The last century witnessed an extraordinary extension of civil rights and freedoms to women and to religious, sexual, and ethnic minorities. Particularly impressive progress was achieved on women's rights, especially in economically developed societies. Yet in much of the world, women still have fewer labor market and educational opportunities; lower physical mobility; less autonomy to run for political office or to make their own decisions about marriage, divorce, and finances; and even less freedom to choose their friends (Doepke and Tertilt 2009; Duflo 2012; Fernández 2014; Fernandez and Wong 2014; Giuliano 2020; Field et al. 2021; Seror 2022). These disparities

*Mehmood: New Economic School (email: smehmood@nes.ru); Naseer: University of Oxford (email: shaheen.naseer@oxfordmartin.ox.ac.uk); Chen: Toulouse School of Economics (email: daniel.chen@iast.fr). Stephen Ross was coeditor for this article. Chen acknowledges IAST funding from the French National Research Agency under the Investments for the Future (Investissements d'Avenir) program (grant ANR-17-EUR-0010). This research received financial support from the research foundation TSE Partnership and ANITI funding. We would like to thank Director of Training Sumera Morris and Director of Academics and Curriculum Development Faria Hafeez at the Progressive Education Network (PEN) for their invaluable support and cooperation throughout the experiment. We also thank Avner Seror, Christina Brown, Seema Jayachandran, Peter Hull, Michela Carlana, Marjorie Sweetko, Alberto Bisin, Matthias Doepke, Gautam Rao, Roberta Ziparo, Hosny Zoabi, Abhijit Banerjee, Esther Duflo, Thierry Verdier, and Eliana La Ferrara for their comments and feedback. Shicheng Xia, Sameen Tariq, Sajwaar Khalid and Bakthawar Ali provided excellent research assistance and on-the-field support. Pre-registered with AEA RCT Registry with ID: AEARCTR-0007465. We received IRB approval from the Lahore School of Economics with IRB Number RERC-062021-03.

[†]Go to https://doi.org/10.1257/pol.20230620 to visit the article page for additional materials and author disclosure statement(s).

are often rooted in cultural attitudes and transmitted from generation to generation (Maurin and McNally 2008; Bisin and Verdier 2011; Alesina et al. 2013; Doepke and Zilibotti 2017; Giuliano and Nunn 2021).

Given these entrenched disparities, our study advances the discourse on gender parity by empirically substantiating the transformative power of education in reforming societal norms. We corroborate findings from influential randomized control trials (Alan et al. 2021; Dhar et al. 2022), which demonstrate that curricular reforms can significantly foster equitable gender attitudes and enhance social cohesion. Building on this literature, our study offers three core insights into the effectiveness of such interventions. First, non-gender-specific interventions, such as those aimed at fostering overall empathy, fail to significantly alter gender attitudes or improve student test scores, suggesting the ineffectiveness of broad-based strategies. Second, gender-focused interventions, even light-touch ones like a gender-themed film screening with structured discussion, can positively influence gender attitudes, suggesting the efficacy of targeted approaches for influencing attitudes. Third, to significantly influence attitudes and improve academic outcomes, intensive measures such as pairing a gender-rights movie with a related curriculum are needed to foster inter-gender cooperation in education and enhance students' test scores.

We implement a randomized control trial in collaboration with PEN, one of the largest networks of charter schools in the world, operating schools across Pakistan. We randomly assign female teachers to one of the following four treatment arms. The first two treatment arms are inspired by economic theory: utilitarianism and identity theory. The principle of utilitarianism is that actions should be evaluated by their utility for oneself and for society as a whole. By emphasizing the importance of empathy, teachers may be encouraged to embrace a more balanced approach to gender equality. Teaching the malleability of one's identity—in particular, regarding growth in empathy—may increase empathy toward outgroups (see, e.g., Weisz and Zaki 2017). In the third treatment, teachers are presented with a visual narrative (Riley 2019; Banerjee et al. 2019) developed at Johns Hopkins University—an award-winning movie—emphasizing the importance of women's rights (Benabou, Falk, and Tirole 2018). Finally, the fourth treatment reinforces the visual narrative with a gender-rights curriculum that the teachers then teach to their students in a semester-long class.

We measure the impacts both on the teacher's own attitudes and on their students' attitudes, as well as on student achievement and gender gaps in student achievement. The gender-rights lessons included structured discussions where students envisioned the rights of men and women: The curriculum encouraged self-reflection by students. To create the opportunity for reflective equilibrium—a deliberative mutual adjustment through inter-gender interactions—we cross-randomized teachers to form either mixed-gender or same-gender student study groups. This cross-randomization enabled us to experimentally examine a causal mediating channel (of inter-gender contact) through which achievement may be boosted. Outcomes are measured six months and one year after the treatment.

The visual narrative presented in the *Bol* movie influences gender rights attitudes among teachers in the short and medium term. Treatment effects are observed after both the stand-alone visual narrative and the joint visual narrative and gender-rights curriculum. Support for equitable gender rights among teachers was roughly 0.15 standard

deviations higher in the group treated with the visual narrative alone and increases to 0.25 standard deviations for teachers when the visual narrative is combined with the gender-rights curriculum. Effects were persistent, with coefficient estimates remaining similar about six months and one year after the treatment. The shift in attitudes is also reflected in the teachers' decisions to petition the government. The visual narrative treatment alone led to at least a 5 percentage point increase in the likelihood of signing pro-women's rights petitions. When this narrative was paired with the gender-rights curriculum, the coefficient estimates suggest that the likelihood of petitioning rose by approximately 20 percentage points. This shows at least a doubling of petitions over the sample mean relative to the control group. Teachers' implicit attitudes, as measured by IAT scores, are also affected: The joint treatment reduced implicit gender bias by about 0.35 standard deviations in both the short and medium term.

Consistent with teachers' internalizing the shift in attitudes through their teaching, the stand-alone visual narrative treatment spilled over to students, and impacted their gender attitudes. The visual narrative treatment on its own led to students becoming 0.1 standard deviations more supportive of gender rights. When it was combined with the semester-long gender-rights curriculum, students became more than twice as likely —about 0.25 standard deviations—more supportive of gender rights. These two results together illustrate the teacher-to-students transmission of gender rights attitudes, especially striking is the impact of the standalone visual narrative treatment on students since this treatment squarely focused on teachers. Students' behavior in cooperation and coordination games when the interaction partner is the opposite gender shifted by 0.2 (in the visual narrative treatment arm) to 0.3 standard deviations (in the visual narrative plus gender-rights curriculum treatment arm). Students in the joint treatment arm (visual narrative plus gender-rights curriculum) also score 0.12 standard deviation higher in mathematics exams. All of these effects, the elevation in cooperation and coordination to the opposite gender due to visual narrative treatment or the joint treatment as well as the elevation in math achievement due to the joint treatment, arise only in mixed-gender study groups.

The rest of the paper is organized as follows. Section I provides the background and experimental design. Section II describes the data and empirical specification, while Sections III and IV present the main results. Section V reports the results of our experimental mediation analysis. A final section provides some concluding remarks. Supplemental Appendixes discuss a series of sensitivity tests and provide the corresponding results and additional experimental details.

I. Background and Study Design

A. *Background*

We collaborate with PEN and embed a large field experiment within their teacher training drives in 2021. PEN is a nonprofit organization that aims to improve the quality of education via a public-private charter. The approach is similar to charter schools in the United States. These schools are privately managed using public funds in a public-private partnership. We implement a randomized evaluation in all charter schools in Punjab, the largest province of Pakistan, where the network

employs 607 teachers and oversees about 15,000 students. According to the PEN mission statement, the focus is not on infrastructure investments but rather on raising the quality of education by improving the quality of teachers:

> We believe that instead of investing resources in 'bricks and mortar', we can leverage the existing infrastructure of public schools to focus on what goes on inside the classroom. The highest percentage of our resources goes towards academic improvement and educational initiatives for the children; which entails teachers' training.[1]

B. *Study Design*

Using a random number generator, we randomly assigned 607 teachers to one of the following treatment arms: (i) utilitarian treatment (121 teachers); (ii) malleability treatment (121 teachers); (iii) visual narrative treatment (122 teachers); (iv) joint visual narrative and gender-rights curriculum treatment (121 teachers); and (v) the control treatment was provided information on procedures to open a bank account in Pakistan (122 teachers). The complete transcripts of the utilitarian and malleability treatments are reported in Supplemental Appendix Tables A1 and A2, respectively. Transcripts to structure the discussion were identical for each treatment arm, including the control arm, and are reported as Table A3 in Supplemental Appendix A. Supplemental Appendix Figure B1 presents a comprehensive summary of the outcomes, treatment specifics, timeline, and principal findings of this study, alongside other studies conducted in partnership with PEN in Pakistan. This integrated perspective aids in the comparative analysis and meta-interpretation of the current research.

C. *Experimental Implementation Details*

Baseline, midline, and endline surveys were conducted in February, September, and March, respectively. We organized classroom-level study groups from four months before the first mathematics examination until the math finals. We cross-randomized teachers to organize students within their classes in mixed-gender or same-gender study groups (from April to July). Figure 1 provides an illustration of the study groups during authors' random spot visits. We conducted the first round of data collection 6 months after the intervention (short-term effects), while the second round of data was collected 12 months later. In both rounds, the data collection was supervised by school administrators and the PEN staff. We worked with a local training department within PEN to manage the surveys with support from our research enumerators.

D. *Utilitarian and Malleability Treatments*

This pair of two-hour interventions targeted broad-based compassion that highlighted global attentiveness to all groups, rather than being limited to gender attitudes. Specifically, the utilitarian intervention was designed to underscore the

[1] Progressive Education Network (2022). PEN mission statement. Retrieved from http://www.pen.org.pk/our-approach-9464.

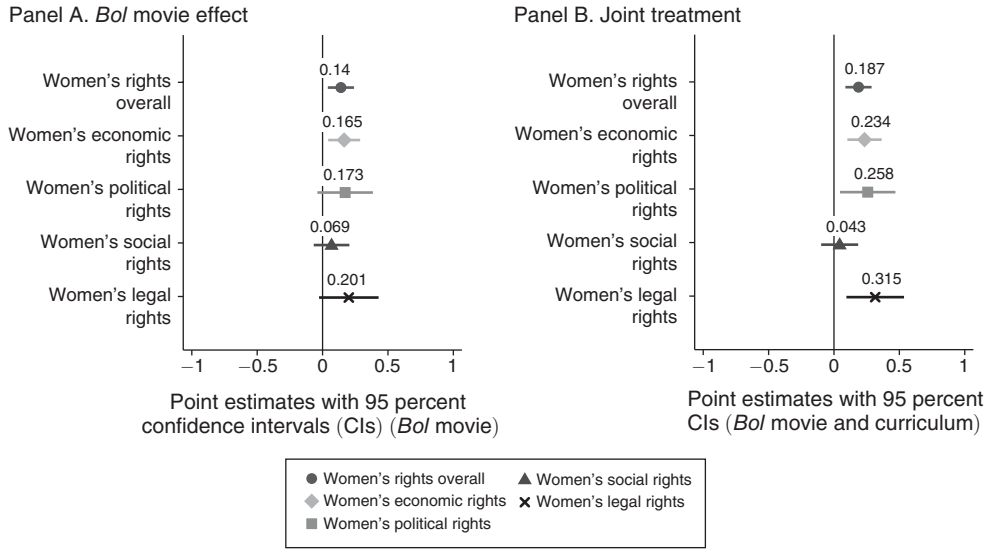

FIGURE 1. IMPACT OF MOVIE AND JOINT MOVIE-CURRICULUM TREATMENT ON TEACHERS' GENDER ATTITUDES

Notes: The dependent variables in the figure are the gender rights indices that were constructed from survey questions listed in Supplemental Appendix D. *Women's rights overall* is an average of all the statements concerning women's economic, social, legal, and political rights. *Women's economic rights* is an index combining women's rights to education and to work outside of the home, based on answer to the following statements: "Women should be allowed to work outside the home"; "Women and men should have equal rights to jobs"; "I have no problem with my sister or female cousin working outside the home"; "Daughters should have the same right to inherit property as sons"; "Women and men should have equal rights to get an education"; "Wives should not be less educated than their husbands"; and "Boys should not have more opportunities and resources for education than girls." *Women's political rights* is based on the following statements: "It would be a good idea to elect a woman as the village Sarpanch (local politician)" and "Women and men have equal rights to be President or Prime Minister." *Women's social rights* is based on the following statements: "Domestic violence by husbands cannot be justified"; "Parents should seek their daughter's consent before fixing her marriage"; "A woman should not necessarily get married before her twenty-fifth Birthday"; "Women who give birth to a son need not be honored in the family"; and "A woman with five daughters should not be under social pressure to bear a son." Finally, the *women's legal rights* index is based on the following statements: "Laws should be passed to ban dowry" and "Under Article 35 of the Constitution of Pakistan and Judgment of Federal Shariat Court, the consent of 'Wali' is not required and a sui juris Muslim female can enter into a valid Nikah/Marriage under her own free will without the consent of Wali. To what extent do you approve of this legal right of women to enter marriage under their own free will." Equation (1) is estimated with all controls, but the coefficient estimate corresponding to the movie treatment is displayed in panel A, and the joint movie-curriculum treatment is displayed in panel B of the figure. Effects are measured 12 months posttreatment. The treatments are compared relative to the placebo treated control group.

practical benefits of fostering empathy in educational settings (Mehmood, Naseer, Chen 2024). By emphasizing the individual benefit of being empathetic, we aimed to align the teachers' self-interest with the promotion of empathy toward others and, by extension, gender-positive attitudes. Similarly, the malleability of empathy training intervention was based on the understanding that empathy is a trait that can be cultivated and developed. Such training could potentially extend to fostering more gender-progressive attitudes. These treatments hoped to leverage a common human capacity for understanding of the "outgroup" and a rational compassion, which can be impactful in shaping attitudes toward gender equality. Full transcripts of these interventions that both hoped to shift empathy by relying on narratives and evidence are presented in Tables A1 and A2 of Supplemental Appendix A, respectively.

E. *Visual Narrative Treatment*

Our third treatment group attended a live screening of a movie—an emotionally charged social drama lasting 2 hours and 45 minutes. The visual narrative was the 2011 film *Bol* (meaning "speak up"), an Urdu-language social drama coproduced by a Pakistani director Shoaib Mansoor and Johns Hopkins University. In this film, a strong female lead, who is on death row, explains why she found it necessary to murder her father, as her "right to exist as a woman" was subverted. The movie explores how she and her sisters deal with her father's obsession with having a son, his perpetuation and strict enforcement of regressive gender attitudes (such as limiting his daughters' rights to employment, inheritance, education, and public spaces), and his staunch rejection of his existing intersex child. This quote from the movie sums up the theme:

> The distance between the rights and freedoms enjoyed by boys in the school next door and us girls in this house is ostensibly one of a wall of a few inches, but the real distance between us and them extends thousands of miles. (Mansoor 2011)

The movie was followed by a 30 minute structured discussion on the movie with application to women's rights and gender attitudes in society. During this discussion, the teachers empathized with the strong female character and discussed the movie's portrayal of the gender gap in rights such as education, work, politics, going outside the home, and accessing public spaces.

F. *Joint Visual Narrative and Gender-Rights Curriculum Treatment*

Our fourth treatment combined the movie *Bol* with a teacher training session on how to conduct a semester-long gender-rights course. This is inspired by theory and empirical evidence on the efficacy of social-emotional learning and teaching as an instrument of self-persuasion (Eskreis-Winkler et al. 2019). In this treatment, the movie was followed by a three-hour workshop on teaching a semester-long course titled "Gender Equality in Child Development for Social Skills." The course commenced in April 2021 and spanned four months. It consisted of three-hour lectures held weekly, amounting to a total of 48 hours of classes over the four-month period. The course was designed by gender activists, the authors of this study, educators, pedagogical consultants, and multimedia developers. The teachers were instructed to discuss gender rights in a context applicable to the child and to his immediate family, and to organize in-class exercises. The classroom exercises involved reading, drawing, and other activities encouraging students to reflect on gender rights, women's place in society, and the rights and freedoms women enjoy. Typical tasks noted by the teachers in their activity logs included the following:

- Task 1: Draw all the work that your father does.
- Task 2: Draw all the work that your mother does.
- Task 3: Which of these are the same? Which are different?
- Task 4: Why is that?

Further examples of lesson plans and teachers' activity logs are reported in Figure B2 of Appendix B. Within this treatment arm, a 30 minute structured discussion similar to the standalone visual narrative treatment on the *Bol* movie was also held.

G. *Implementation Details*

The training was conducted live on Zoom by the same team of eight research assistants located in Lahore, using the same set of questions, resulting in similar discussion duration across groups. The screening of the *Bol* movie was also conducted live on Zoom, and in adherence with PEN training requirements, the teachers had their cameras on. The curriculum cost was similar as it was developed by the research team and a group of pedagogical volunteers, and the teachers did not charge any fees due to our collaboration with PEN.

H. *Control Group*

The control group receives training on generic procedures to open a bank account in Pakistan. This includes readily known facts such as going to the bank reception, requesting to open a bank account and presenting identification documents. Administrative data from PEN reveal that every teacher in our sample already had a bank account at time of this intervention, so this control treatment is unlikely to have a direct effect on gender attitudes.

I. *Mixed- and Same-Gender Study Groups*

Building on recent scholarship in pedagogy that effective cooperation may be a key pathway to improve math achievement (Gutiérrez 2002; Weissglass 2001; Corte 2004) and that intergroup contact reduces prejudice (Rao 2019; Lowe 2021), we cross-randomized teachers to form either mixed-gender or same-gender student study groups in their mathematics classes. The instructions provided to teachers were to make a group of two students (that were randomly assigned either to be mixed or of the same gender). They were to meet once weekly for 30 minutes and discuss any past homework assignments in the assigned group of two. This was done within the mathematics class under teachers' supervision for three months before the mathematics final exam. There was no further instruction for the students or teachers of either study group on the structure of the group discussion except we will do random spot visits in school to ensure enforcement.

J. *Behavioral Games with Students*

We employ administrative data to evaluate student achievement in mathematics exams and gauge their gender attitudes through attitudinal surveys and behavioral games. By analyzing students' behavior in games that involve either same-gender or opposite-gender interactions, we seek to determine the impact of gender on students' treatment of others. To this end, we implement four incentivized economic

games: competition, cooperation, coordination, and redistribution, inspired by the methodology employed in Kosse et al. (2020) in their work with primary school students. Adapting these games to fit cultural norms, we introduce the "milk bank" version where winners receive "milk carton coupons" redeemable for milk cartons at school canteens.[2]

II. Data and Empirical Strategy

A. *Data*

Sample.—The sample consists of all 607 teachers and their 13,911 students across all 52 schools chartered by PEN in the state of Punjab. As is common in most public schools in Pakistan, all teachers are female and teach every class from kindergarten to grade 6. The students, however, are of mixed gender in public schools of Pakistan until they "graduate" from grade 6. Nevertheless, this is not a universal practice. Our sample consists of 7,107 boys and 6,804 girls ranging from age 5 to 12. The PEN network organizes several training workshops for teachers each year, and our experiment took place within the PEN teacher training drives. As a result of our experimental intervention embedded within the PEN's regular training programs, we essentially have zero attrition. All 607 PEN school teachers in the state of Punjab participated in the experiment. The baseline survey was carried out in the second week of January, the midline survey in August, and the endline in March 2022. We utilize detailed administrative data on teachers and students, enabling us to match teachers to their students at the classroom level.

Outcome Variables on Teachers' Gender Attitudes.— Our first set of outcome variables concerns teachers' attitudes toward gender rights as assessed about six months and a year after our intervention. To summarize teachers' gender attitudes, we use the gender rights index that averages across all components of gender rights survey questions listed in Supplemental Appendix D. "Women's rights overall" is an index combining all the statements concerning women's economic, social, legal, and political rights. "Women's economic rights" is an index combining women rights regarding education and work. "Women's political rights" is based on the right to hold political office, while "women's social rights" and "women's legal rights" concern the rights to choose whether or not to conform to social and legal discrimination, respectively. For more details, see our survey instrument in Supplemental Appendix D2, while information on index construction is reported in Supplemental Appendix D3.

Outcome Variables on Teachers' Decisions and IATs.—Our second set of outcomes involves revealed preference measures of gender attitudes in the form of teachers' willingness to sign and send petitions asking parliament to repeal discriminatory laws, and the IAT. The gender IAT measures implicit associations regarding women. The IAT involves categorizing words by placing them on the left or right

[2] Supplemental Appendix Table A4 details research ethics, whereas Supplemental Appendix Tables A5 and A6 discuss curriculum treatment and teacher workshops, respectively.

of a computer screen and measures the strength of association between two concepts based on response times. We use a standard gender-career IAT test to see if respondents associate women with "family" and men with "career". The IAT hence measures female-sounding names and gender stereotypes.[3] The gender that we administered IAT was the standard career-family word association task based on seven questions. Too short or long latencies were automatically dropped according to the algorithm determined in Greenwald et al. (2009).[4]

Outcome Variables on Students.—We compute students' gender attitudes by fielding a five statement survey 6 and 12 months following the treatment. The survey statements are chosen so that it may be understood by primary school students of different ages that populate our sample. The survey statements are reported in Supplemental Appendix D6. We are also able to obtain results on standardized mathematics examinations that the students gave, also 6 and 12 months following the treatment. We use these outcomes to assess both students' gender attitudes and academic achievement. We also play a total of four games with the students that include cooperation, coordination, redistribution and competition. Following Kosse et al. (2020) we adapt the standard games into a "piggy bank" version, which we call the "milk bank" game due to cultural reasons. In particular, we offer winners milk carton coupons that could be redeemed at the school canteens for milk cartons. More details on the games, including the transcripts the students saw can be found in Supplemental Appendix D7.

Main Explanatory Variables.—Our key explanatory variables are dummies for the four treatments. U_i and M_i denote dummies that switch on if the teachers were assigned to the utilitarian or malleability treatments, respectively. $BolMovie_i$ (BM) and $BolMovie\&Curriculum_i$ (BMC) are dummies that switch on if the teachers were assigned to the visual narrative or joint visual narrative and gender-rights curriculum treatments, respectively. The control group receives training on generic procedures to open a bank account in Pakistan. This includes readily known facts such as going to the bank reception, requesting to open a bank account and presenting identification documents.[5]

B. *Attrition and Balance*

Close cooperation and support from the leadership of the PEN organization on administrative data and the fact that the experiment was embedded within PEN's regular trainings meant that attrition was zero for teachers (except the attrition artificially triggered by the IAT algorithm for IAT scores), and student attrition amounted

[3] The use of IATs also reduces concerns about experimental demand. First, it is hard to respond to IATs in a socially desirable way, as this would require strategically speeding up or slowing down in certain blocks of associations (Alesina et al. 2018). Second, consistent with the psychological literature, our IAT algorithm discards observations that are too slow or too fast (Greenwald et al. 2003). Last, it is highly unlikely that teachers within our sample —primary and middle school teachers in Punjab—would know about IATs.

[4] We administer the IAT in Urdu online over oTree and report the text that the teachers saw on their screens in Supplemental Appendix D5, while exact template for petition text is provided in Supplemental Appendix D4.

[5] Administrative data from PEN reveal that every teacher in our sample already had a bank account at time of this intervention, so this control treatment is unlikely to have direct effect on outcomes.

to only 21 students in our midline and endline surveys (held about six months apart).[6] Nevertheless, a lack of balance might still complicate causal interpretation of our results. We therefore examine whether our randomization was successful in creating balance among teachers and students. Table B1.2 in Supplemental Appendix B shows individual characteristics, with panel A reporting the treatment balance over teacher characteristics and panel B regarding student characteristics. Differences across treatment groups are small in magnitude, and almost all estimated *p*-values are larger than 0.10, suggesting that the randomization was effective at creating balance between the groups. For instance, teachers' education, experience, class size, number of hours of teaching, marital status, and pretreatment gender attitudes are balanced across treatment and control groups. Likewise, from panel B of Supplemental Appendix Table B1.2, we observe that the pretreatment gender attitudes and mathematics test scores of students are balanced. Supplemental Appendix Table B1.1 presents the descriptive statistics of the main outcomes used in the study. Supplemental Appendix Tables B1.2 and B1.3 provide evidence of balance for the baseline specifications, while Supplemental Appendix Table B1.4 also demonstrates balance for our cross-randomized school study groups in both same-gender and mixed-gender groups. To ease comparisons with baseline estimates, we also report results in standardized units. In all these instances, the treatments appear to be balanced across teacher and student characteristics.

C. *Estimation Strategy*

The impact of our four treatments can be evaluated by comparing outcomes across groups in a simple regression framework. For each outcome, the estimation equation is

$$Y_i = \alpha + \beta U_i + \gamma M_i + \delta BM + \omega BMC_i + \theta_s + \mathbf{X}_i'\boldsymbol{\mu} + \epsilon_i\,, \tag{1}$$

where Y_i is the outcome for a teacher i; U_i is a dummy variable equal to one if the teacher is assigned to the utilitarian empathy treatment; M_i is a dummy variable equal to one if the teacher is assigned to the malleability empathy treatment; BM_i is a dummy variable equal to one if the teacher is assigned to the visual narrative (movie *Bol* promoting more equitable gender rights) treatment; and BMC_i if the teacher is assigned to the joint visual narrative and gender-rights curriculum treatment. θ_s represents school fixed effects. $\mathbf{X}_i$ is a vector of individual-level teacher and student controls. In equation (1), β measures the effect of the utilitarian treatment, γ the effect of the malleability treatment, δ the effect of the visual narrative treatment, and ω measures the effect of the joint visual narrative and gender-rights curriculum treatment. Following Chetty et al. (2014) in student-level regressions, we also always control for the student's prior test scores. We cluster standard errors at the teacher level for students since that is our level of randomization. Since randomization is at the individual level, Newey-West robust *p*-values are included as baseline.

[6] This remarkable take-up was only possible due to gracious support and cooperation of the Director of Training and Research Sumera Morris and her staff at PEN. They provided invaluable suggestions and support throughout this intervention.

In addressing the inherent challenges of multiple hypothesis testing and the associated risk of inflated type 1 errors, our empirical analysis employs two statistical methods to control the false discovery rate (FDR) and the family-wise error rate (FWER). We adopt Anderson's (2008) two-stage refinement of the Benjamini-Hochberg FDR procedure, which sequentially adjusts significance levels to reduce false discoveries. For FWER control, we apply the Romano-Wolf step-down procedure via the rwolf2 package, leveraging its resampling mechanism to accommodate test correlations. All main tables include Newey-West standard *p*-values as well as FDR *q*-values and FWER-adjusted *p*-values.

III. Impact on Teachers

A. *Effects on Teachers' Attitudes*

We measure teachers' attitudes 12 months following the treatment and observe the quantitatively and qualitatively significant impact of visual narrative treatments, both on its own and when combined with the semester-long gender-rights curriculum. Table 1 reports the results on the impact of all our treatments on gender attitudes, while Figure 1 visualizes all of these impacts, including the stand-alone visual narrative treatment and the impact of the joint visual narrative and curriculum treatment. Figure 1 presents the estimated coefficients for the medium-term impact, measured at 12 months posttreatment. For the short-term effects, observed at a six-month interval after the treatment, refer to Table C1 in Supplemental Appendix C. From column 1 of Table 1, we observe that the visual narrative alone increased support for more equitable gender rights by about 0.15 standard deviations. When the visual narrative is combined with the curriculum, the impact almost doubles, with teachers' support for women's rights increasing by roughly 0.20 standard deviations. The effects are enduring, evident in both the short term (as shown in Supplemental Appendix Table C1) and medium term (as indicated in Table 1), with the robustness of the results confirmed by Newey-West *p*-values, sharpened *q*-values adjusted for multiple hypotheses, and Romano-Wolf correction for FWER *p*-values. Table 1 also contains results on the impact of the utilitarian and malleability treatments: These do not appear to have much impact on gender attitudes either in the short or medium term. These overall gender attitudes index is also disaggregated into economic, political, social, and legal rights indices. The coefficient estimates suggest that the visual narrative and joint treatments likely shifted attitudes concerning women's economic, political, and legal rights. As a point of comparison, the impact of the joint treatment of visual narrative and four-month gender studies curriculum is roughly half the effect size as that found for random roommate assignment (Corno et al. 2022) and a little over half the effect as that found for a 12-week training program (Devine et al. 2012).

B. *Effect on Teachers' IATs*

Next, we assess the impact of the standalone visual narrative and the joint visual narrative and curriculum treatments on IAT scores. From Figure 2, we observe that the visual narrative treatment alone reduces implicit gender bias by about 0.20 standard deviations, and when it is combined with the gender-rights curriculum, implicit

Table 1—Impact on Teachers' Attitude

	Gender rights overall (1)	Economic rights (2)	Political rights (3)	Social rights (4)	Legal rights (5)
Visual narrative (movie)	0.140	0.165	0.173	0.0687	0.201
p-value	(0.0062)	(0.0082)	(0.1102)	(0.3226)	(0.0838)
Sharpened *q*-value	[0.022]	[0.026]	[0.166]	[0.303]	[0.141]
Romano-Wolf corrected *p*-value	{0.030}	{0.026}	{0.5864}	{0.9880}	{0.061}
Joint movie and curriculum	0.187	0.234	0.258	0.0434	0.315
p-value	(0.0003)	(0.0005)	(0.0171)	(0.5472)	(0.0051)
Sharpened *q*-value	[0.004]	[0.004]	[0.033]	[0.460]	[0.022]
Romano-Wolf corrected *p*-value	{0.012}	{0.012}	{0.054}	{0.9770}	{0.018}
Utilitarian	0.0607	0.0805	0.0783	0.0365	0.0345
p-value	(0.1725)	(0.1760)	(0.4280)	(0.5743)	(0.7113)
Sharpened *q*-value	[0.239]	[0.239]	[0.422]	[0.460]	[0.477]
Romano-Wolf corrected *p*-value	{0.4345}	{0.4166}	{0.8472}	{0.9880}	{0.9880}
Malleability	0.0897	0.102	0.155	0.0290	0.132
p-value	(0.0916)	(0.1162)	(0.1836)	(0.6739)	(0.2758)
Sharpened *q*-value	[0.145]	[0.166]	[0.239]	[0.477]	[0.299]
Romano-Wolf corrected *p*-value	{0.1209}	{0.1529}	{0.7502}	{0.9880}	{0.3147}
Controls and School FEs	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	607	607
R^2	0.138	0.118	0.097	0.102	0.125
p-value ($BM = BMC$)	0.388	0.318	0.445	0.725	0.360

Notes: The figure summarizes our main results—effect of the treatments on the different gender attitude indices summarizing attitudes toward women. Standardization to mean 0 and standard deviation are performed to each dependent variable. The treatments are compared relative to the placebo treated control group. The outcomes are recorded 12 months after the treatment. *Women's rights overall* is an average of all the statements concerning women's economic, social, legal, and political rights. *Women's economic rights* is an index combining women's rights to education and work outside home, based on reactions to the following statements: "Women should be allowed to work outside the home"; "Women and men should have equal rights to jobs"; "I have no problem with my sister or female cousin working outside the home"; "Daughters should have the same right to inherit property as sons"; "Women and men should have equal rights to get an education"; "Wives should not be less educated than their husbands"; and "Boys should not have more opportunities and resources for education than girls." *Women's political rights* is based on the following statements: "It would be a good idea to elect a woman as the village Sarpanch (local politician)" and "Women and men have equal rights to be President or Prime Minister." *Women's social rights* is based on the following statements: "Domestic violence by husbands cannot be justified"; "Parents should seek their daughter's consent before fixing her marriage"; "A woman should not necessarily get married before her twenty-fifth Birthday"; "Women who give birth to a son need not be honored in the family"; and "A woman with five daughters should not be under social pressure to bear a son." Finally, the *women's legal rights* index is based on the following statements: "Laws should be passed to ban dowry" and "Under Article 35 of the Constitution of Pakistan and Judgment of Federal Shariat Court, the consent of 'Wali' is not required and a sui juris Muslim female can enter into a valid Nikah/Marriage under her own free will without the consent of Wali. To what extent do you approve of this legal right of women to enter marriage under their own free will." Equation (1) is estimated with all controls. *Utilitarian* variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the *Malleability* treatment indicator. *Visual narrative* (*movie*) is similarly a dummy turning on for subjects assigned the *Bol* movie. The *Joint movie and curriculum* indicator turning on for teachers assigned the joint *Bol* movie and the gender-rights curriculum treatment. *p*-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR *q*-values in square brackets and FWER-adjusted *p*-values in curly braces. Further details on this are provided in Supplemental Appendix D8 (considering 36 hypotheses with 4 treatments × 9 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., kindergarten, nursery, prep, one, two, three, four, five, and six class) and pretreatment math scores.

gender bias reduces by at least 0.32 standard deviations. These effects hold for both the short (panel A) and mid term (panel B) and when we consider the IAT in original units or raw scale (Figure B3 in Supplemental Appendix B). We interpret the greater effect of the joint treatment as a sign that teaching the gender-rights

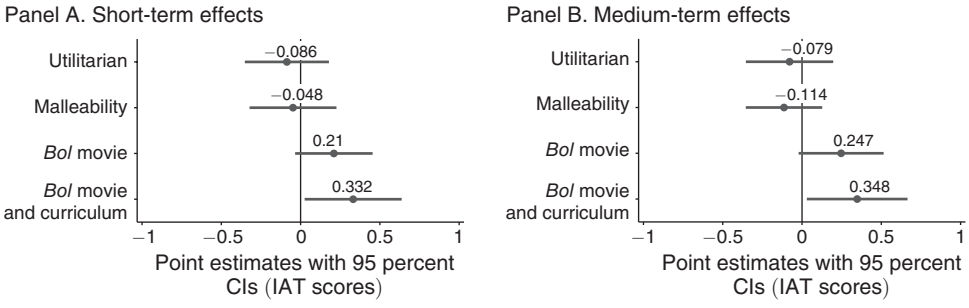

FIGURE 2. IMPACT ON STANDARDIZED IAT SCORES

Notes: The figure reports estimates from equation (1) with all main explanatory variables and controls with gender IAT—standardized to mean 0 and standard deviation 1—as the dependent variable. Controls include all available individual characteristics and school fixed effects. 95 percent confidence bands are also reported. Short-term effects in panel A are recorded 6 months after the treatment and medium-term effects in panel B are recorded 12 months after the treatment. The treatments are compared relative to the placebo treated control group.

curriculum reinforced the treatment through self-persuasion, consistent with recent evidence that even short amounts of time spent discussing principles can result in self-persuasion (Eigen and Listokin 2012; Schwardmann, Tripodi, and van der Weele 2022).

C. *Effects on Teachers' Decisions* (*Petitions*)

Finally, we provide causal evidence that self-reported attitudes are reflective of behavioral change in high-stakes settings. We offer all the teachers the opportunity to sign a petition that is then sent to the Pakistani parliament seeking the abolishment of discriminatory laws allowing polygamy, as well as a petition seeking to make a man's demand for a dowry a criminal offense. The results here are even more striking than what we found for self-reported gender attitudes. From Table 2, we observe that the visual narrative of the movie *Bol* alone increases the likelihood of petitions sent by teachers by about 10 percentage points. This is economically significant and equivalent to roughly a doubling of petitions over the sample mean. The coefficient estimates for the joint treatment are even larger, with petition signing increasing by about 20 percentage points. These results strongly suggest that our results from self-reported surveys are likely to have real-world implications.

D. *Effect of Previously Seeing Bol Movie*

About half of the teachers stated that they had watched the *Bol* movie after it was released a decade earlier in 2011—therefore, we investigate if those teachers who had previously watched *Bol* are more likely to be impacted by the visual narrative or the joint treatment, notwithstanding the endogeneity of the teachers' prior exposure to the movie. Supplemental Appendix Figure B4 shows little evidence for the heterogeneous effect of treatment on those who had previously watched the movie *Bol*. These results hold for gender attitude surveys, IAT, and the two gender rights petitions. We first assess whether exposure to the film *Bol* influences the

Table 2—Impact of Teachers' Training on Petitions to Parliament

	Petition to criminalize dowry		Petition to abolish polygamy	
	(1)	(2)	(3)	(4)
Visual narrative (movie)	0.104	0.115	0.0598	0.0593
p-value	(0.0113)	(0.0055)	(0.0134)	(0.013)
Sharpened *q*-value	[0.029]	[0.022]	[0.029]	[0.029]
Romano-Wolf corrected *p*-value	{0.036}	{0.023}	{0.036}	{0.036}
Joint movie and curriculum	0.173	0.180	0.0866	0.0870
p-value	(0.0002)	(0.0001)	(0.0004)	(0.0005)
Sharpened *q*-value	[0.004]	[0.004]	[0.004]	[0.004]
Romano-Wolf corrected *p*-value	{0.01}	{0.01}	{0.013}	{0.013}
Utilitarian	0.0083	0.0070	−0.0117	−0.0106
p-value	(0.8038)	(0.832)	(0.1972)	(0.2622)
Sharpened *q*-value	[0.477]	[0.479]	[0.246]	[0.296]
Romano-Wolf corrected *p*-value	{0.9880}	{0.988}	{0.586}	{0.724}
Malleability	0.0089	0.0189	−0.0055	−0.0033
p-value	(0.7959)	(0.5844)	(0.567)	(0.7514)
Sharpened *q*-value	[0.477]	[0.46]	[0.46]	[0.477]
Romano-Wolf corrected *p*-value	{0.988}	{0.987}	{0.987}	{0.988}
Individual controls	No	Yes	No	Yes
School fixed effects	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R^2	0.124	0.140	0.188	0.200
Mean of dependent variable	0.114	0.114	0.030	0.030
p-value ($BM = BMC$)	0.185	0.206	0.433	0.409

Notes: The dependent variable in columns 1 and 2 is a dummy variable that switches on if the teacher signed a petition seeking criminalization of dowry while the dependent variable in columns 3 and 4 is a similar dummy variable turning on for a petition seeking to abolish laws allowing polygamy in Pakistan. The outcomes are recorded 12 months after the treatment. *Utilitarian* variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the *Malleability* treatment indicator. *Visual narrative* (*movie*) is similarly a dummy turning on for subjects assigned the *Bol* movie. The *Joint movie and curriculum* indicator turning on for teachers assigned the joint *Bol* movie and the gender-rights curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Supplemental Table A3. The treatments are compared relative to the placebo treated control group. *p*-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR *q*-values in square brackets and FWER-adjusted *p*-values in curly braces. Further details on this are provided in Supplemental Appendix D8 (considering 36 hypotheses with 4 treatments × 9 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., kindergarten, nursery, prep, one, two, three, four, five, and six class) and pretreatment math scores.

attitudes of control group teachers, finding no association with our measured outcomes (Figure B4, Supplemental Appendix B). Subsequently, we analyze treatment effects among teachers who had previously seen the movie, confirming that having watched the *Bol* movie is not correlated with gender attitudes in the control group (Figure B5, Supplemental Appendix B). Lastly, we demonstrate consistent effects of the visual narrative and joint treatment on teachers who had previously seen the *Bol* movie; these teachers also exhibit a similar change in gender attitudes (Figure B6 and Table B10). Taken together, the results suggest that watching the *Bol* movie beforehand is unlikely to overturn our findings. This may be due to many factors. For

Table 3—Impact of Teachers' Training on Students' Attitudes and Math Test Scores

	Student attitudinal survey		Maths	
	(1)	(2)	(3)	(4)
Visual narrative (*movie*)	0.145	0.137	−0.00811	−0.00482
p-value	(0.0057)	(0.0093)	(0.8949)	(0.8433)
Sharpened *q*-value	[0.018]	[0.021]	[0.687]	[0.687]
Romano-Wolf corrected *p*-value	{0.001}	{0.001}	{0.992}	{0.992}
Joint movie-curriculum	0.254	0.245	0.160	0.119
p-value	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
Sharpened *q*-value	[0.001]	[0.001]	[0.018]	[0.001]
Romano-Wolf corrected *p*-value	{0.001}	{0.001}	{0.001}	{0.001}
Utilitarian	0.0713	0.0723	0.105	0.0293
p-value	(0.1767)	(0.1722)	(0.1055)	(0.2544)
Sharpened *q*-value	[0.245]	[0.245]	[0.178]	[0.318]
Romano-Wolf corrected *p*-value	{0.0709}	{0.0599}	{0.027}	{0.1429}
Malleability	0.00171	−0.00252	0.00324	0.00419
p-value	(0.9776)	(0.9666)	(0.9584)	(0.8669)
Sharpened *q*-value	[0.687]	[0.687]	[0.687]	[0.687]
Romano-Wolf corrected *p*-value	{0.992}	{0.992}	{0.992}	{0.992}
Individual controls	No	Yes	No	Yes
School fixed effects	Yes	Yes	Yes	Yes
Observations	13,911	13,911	13,911	13,911
R^2	0.038	0.044	0.090	0.596
p-value ($BM = BMC$)	0.020	0.023	0.007	$p < 0.01$

Notes: The dependent variables are standardized to mean zero and standard deviation for mathematics test scores and student attitudinal survey. The outcomes are recorded 12 months after the treatment. The corresponding survey statements from students are reported in Supplemental Appendix D4. *Utilitarian* variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the *Malleability* treatment indicator. *Visual narrative* (*movie*) is similarly a dummy turning on for subjects assigned the *Bol* movie. The "Joint movie and curriculum" indicator turning on for teachers assigned the joint *Bol* movie and the gender-rights curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Supplemental Appendix Table A3. The treatments are compared relative to the placebo treated control group. *p*-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR *q*-values in square brackets and FWER-adjusted *p*-values in curly braces. Further details on this are provided in Supplemental Appendix D8 (considering 16 hypotheses with 4 treatments × 4 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., kindergarten, nursery, prep, one, two, three, four, five, and six class) and pretreatment math scores.

instance, if the teachers watched the movie a decade earlier when it premiered, and the effects of the movie on teachers may have dissipated after ten years. It could also be the case that the structured discussion on the gender rights themes of the movie among peers reinforced the message of the movie beyond just watching the movie.

IV. Impact on Students

A. *Effects on Students' Attitudes*

To investigate whether gender-rights attitudes were transmitted from teachers to students, we surveyed all students that the 607 teachers taught, roughly seven months and one year later. Columns 1 and 2 of Table 3 report these results. We

observe that the visual narrative alone made students more supportive of gender rights by about 0.10 standard deviations. The results hold for both the short and medium term, suggesting that gender-rights transmission is likely persistent. These results are also particularly striking since the standalone visual narrative treatment squarely focused on the teachers. Nevertheless, consistent with results on teachers, the coefficient estimates from Table 3 also indicate more than a doubling of the effects under the joint visual narrative and curriculum treatment: Students whose teachers both experienced the visual narrative and taught the gender-rights curriculum are at least 0.25 standard deviations more supportive of more equitable gender rights in both the short and the medium term.

B. *Effects on Students' Test Scores*

One of our treatments also impacts student achievement, as shown by the students' mathematics test scores. We observe that math test scores are positively impacted when the teachers were assigned to the joint visual narrative and curriculum treatment. These results also allay potential concerns of experimental demand since the treatment focused on gender attitudes, not math achievement. Columns 3 and 4 of Table 3 report these results: Students' math scores are about 0.1 standard deviations higher under the joint treatment. To put this into perspective, the observed 0.1 standard deviation rise in mathematics test scores implies a 3.5-point enhancement on the 0 to 100 grading scale used for the math assessment. This improvement reflects a 5 percent increase above the average score of the sample of students in our study or student outcomes for grade B+ students increase to A− due to our treatment. While the standalone *Bol* movie treatment of teachers does not affect student test scores, gender-focused interventions, such as film screenings joint with a curriculum, can shift gender attitudes and even student test scores, indicating that more intense interventions are needed to alter student test scores.

It is also worth nothing that the smaller and statistically insignificant point estimates of the standalone movie treatment on academic test scores do not support the hypothesis that exposure to the visual narrative, specifically the *Bol* movie, independently contributes to improvements in academic test scores. This finding implies that interventions of a light-touch nature, which are limited to teacher engagement, are not potent enough to elicit changes in the academic performance of students. In the next section, we investigate a mechanism behind this increase, deploying an experimental causal mediation design, to investigate how students learned more effectively in the joint treatment arm.

C. *Effect on Stress of Students*

In a separate analysis of the same intervention, Mehmood, Naseer, and Chen (2023) found that while the visual narrative treatment arms succeeded in shifting teacher's gender attitudes, they adversely impacted mental health and had adverse consequences in terms of domestic violence, though the negative effects on mental health disappeared if a large enough share of teachers in the school were

also exposed to the visual narrative treatment—a phenomenon we call "moral bandwagoning." It is important to note that the increase in teacher stress has not led to a decline in student academic performance. Access to national mathematics exam results allows us to determine that the implementation of our visual narrative, alone or in conjunction with other treatments, has not adversely impacted academic achievement in the short run (6 months after treatment) or in a more extended period (12 months after the intervention). We further observe that while teacher stress levels rose in response to the visual narrative and the combined visual narrative with curriculum treatment, this pattern was not mirrored among students. Student stress levels are statistically similar across the treated groups when compared to the control group. These results are reported in Table B3 in Supplemental Appendix B.[7]

Our results remain robust across numerous sensitivity tests, as detailed in a robustness section presented in Supplemental Appendix A2. Here, we address potential concerns such as experimental demand and spillover effects, concluding that our results are unlikely to be influenced by these factors. This conclusion is also supported by the use of the IAT and the Marlowe-Crowne social desirability scale, coupled with observed the sustained effects of our treatments over time.

V. Mediation Analysis

The design of our experiment lends itself to causal mediation analysis to study inter-gender contact driving our results. Typically, a sequence of behavioral data is observed to infer how earlier actions mediate final outcomes. Here, we use an experimental set-up of randomization across teachers and re-randomization of teachers to form mixed-gender or same-gender student study groups to causally isolate the effect of working with the opposite gender up to one year after the training intervention. The results of our causal mediation analysis help uncover experimentally why students' test scores increased for the joint visual narrative and curriculum treated group.

Building on the emerging causal evidence that intergroup contact may reduce prejudice (Rao 2019; Lowe 2021), we hypothesized that more equitable gender-rights attitudes would be formed through interactions with the opposite gender, and that these interactions would lead to teacher transmission of equitable gender-rights attitudes to students, and improved theory of mind via increased cooperation and coordination with the opposite gender. Two months before the math exam, we randomly assigned teachers to organize students into mixed-gender or same-gender study groups that lasted until the final examination held about a year after the treatment. Random spot checks by the authors, field assistants, and PEN administrators confirmed that our study group randomization was followed by the teachers.

[7] Further heterogeneity analysis reveals that the impact of visual narrative and joint treatment on gender attitudes remains relatively uniform across various grades of students. Nonetheless, the point estimates reveal that a combined intervention might exert a greater influence. These findings are detailed in Table B4, located in Supplemental Appendix B. A similar heterogeneity analysis by gender is also performed, reported in Table B5, with the results being similar between boys (panel A) and girls (panel B).

Table 4—Impact of Teachers' Training on Standardized Student Math Test Scores

	Math test scores				
	(1)	(2)	(3)	(4)	(5)
Visual narrative (movie)	−0.0384	0.0194	−0.0116	0.0106	0.0202
p-value	(0.3016)	(0.5803)	(0.7111)	(0.7533)	(0.5431)
Sharpened *q*-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected *p*-value	{0.6084}	{0.9271}	{0.9730}	{0.9730}	{0.9041}
Joint movie and curriculum	0.164	0.0463	0.137	0.110	0.0548
p-value	$p < 0.01$	(0.2258)	$p < 0.01$	(0.0015)	(0.1159)
Sharpened *q*-value	[0.001]	[0.999]	[0.001]	[0.01]	[0.971]
Romano-Wolf corrected *p*-value	{0.001}	{0.4635}	{0.001}	{0.001}	{0.1578}
Utilitarian	0.0379	0.0388	0.0256	0.0354	0.0375
p-value	(0.2782)	(0.2834)	(0.4647)	(0.3045)	(0.2972)
Sharpened *q*-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected *p*-value	{0.5894}	{0.5894}	{0.8422}	{0.6084}	{0.6084}
Malleability	0.00168	−0.0204	0.0203	0.00591	0.0095
p-value	(0.9592)	(0.5847)	(0.5333)	(0.8618)	(0.7926)
Sharpened *q*-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected *p*-value	{0.9730}	{0.9271}	{0.9041}	{0.9730}	{0.9730}
Movie-curriculum × mixed study					0.126 (0.0476)
Mixed study group					0.0191 (0.0310)
U × mixed study group					−0.0147 (0.0489)
M × mixed study group					−0.00969 (0.0475)
Movie × mixed study group					−0.0535 (0.0487)
Students are girls			No	Yes	
Mixed-gender study group sample	Yes	No			
Individual controls and school FE	Yes	Yes	Yes	Yes	Yes
Observations	6,959	6,952	7,107	6,804	13,911
p-value ($BM = BMC$)	$p < 0.01$	0.485	$p < 0.01$	0.005	0.346

Note: Dependent variable is standardized to mean 0 and standard deviation for math test scores. The outcomes are recorded 12 months after the treatment. *Utilitarian* variable is a binary indicator that assumes a value of 1 upon the teacher's receipt of the corresponding treatment, analogous to the *Malleability* treatment indicator. *Visual narrative* (*movie*) is similarly a dummy turning on for subjects assigned the *Bol* movie. The *Joint movie and curriculum* indicator turning on for teachers assigned the joint *Bol* movie and the gender-rights curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Supplemental Appendix Table A3. *p*-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR *q*-values in square brackets and FWER-adjusted *p*-values in curly braces. Further details on this are provided in Supplemental Appendix D8 (considering 20 hypotheses with 4 treatments × 5 outcomes). "Mixed study group" is a dummy that switches on when the student group is mixed-gender. U × Mixed study group, M × Mixed study group, Movie × Mixed study group, and Joint movie-curriculum × Mixed study group are interaction terms of mixed study group with utilitarian, malleability, movie, and joint movie-curriculum treatments, respectively. The treatments are compared relative to the placebo treated control group. The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., kindergarten, nursery, prep, one, two, three, four, five, and six class) and pretreatment math scores. Supplemental AppendixTable C2 reports the corresponding short-term results six months posttreatment.

A. *Effects on Students' Math Scores by Study Group*

The increase in test scores is driven by students whose teachers were assigned to the joint visual narrative and curriculum treatment and those rerandomized to form opposite-gender study groups. The math achievement of students assigned to same-gender study groups is unaffected. The estimates from column 1 of Table 4 indicate that students randomly assigned to teachers treated with the joint visual

narrative and gender-rights curriculum and opposite-gender study groups increased their test scores by roughly 0.18 standard deviations.[8] The impact persists over both the short and the medium term. We observe similar results from an interaction specification (Table 4, column 5): There is a qualitatively and quantitatively significant impact on math scores when the teachers treated with the joint visual narrative and gender-rights curriculum treatment formed the mixed-gender study groups. Supplemental Appendix Figure B7 gives raw mathematics scores out of 100, revealing the same impact. Overall, the results presented in Table 4 and Supplemental Appendix Figure B7 indicate that desegregation within classes may enhance learning. Importantly, desegregation has no independent effect in the control group.

B. *Effects on Students' Math Scores by Gender*

Intriguingly, there is a positive impact on math achievement for both boys and girls (Supplemental Appendix Figure B8). The male student scores increase by about 0.15 standard deviations, while the female scores increase by about 0.1 standard deviation, both in the short and medium term. The mathematics achievement gender gap decreases by at least 20 percent. Supplemental Appendix Figure B8 shows that both boys' and girls' overall math scores increase from about 65 to 70 out of 100, following our joint narrative and curriculum treatment—that is, student outcomes improve from grade B+ to A− for the treated group. The greater improvement in test scores for boys, as compared to girls, may suggest that the joint narrative and curriculum treatment is particularly effective in addressing the learning needs of male students; it is likely that since boys had lower baseline achievement levels in mathematics at an early age compared to girls, the treatment might have had more room for impact, leading to a larger gain for boys as they "catch up" to their female peers. This is consistent with a recent meta-analysis performed by Kersey et al. (2018), which synthesized data from six studies to explore gender differences in early mathematical cognition among children. The findings indicated that there were no significant differences in mean performance and variability in mathematical abilities between boys and girls at these early stages. If anything, girls' scores in the early years were slightly higher than those of boys.[9] This mirrors our case (where children aged from 6 to 12 years old are studied), indicating the treatment could facilitate greater gains for boys as they catch up to girls.

C. *Effects on Students' Behavioral Games by Study Group*

Table 5 separates the results of the behavioral games according to whether students played against the opposite gender or against the same gender (Supplemental Appendix Table B2). Students playing against the opposite gender showed greater

[8] The short-run results, observed six months posttreatment, are consistent with medium-term findings and can be reviewed in Table C3 of Supplemental Appendix C.

[9] Similar results are found in the studies by Felson and Trudeau (1991) on gender differences in mathematics performance between girls and boys, in Pomerantz et al. (2002), Duckworth and Seligman (2006), and Hyde et al. (2008).

Table 5—Impact of Treatments on Social Behavior Games Played against Opposite Gender

	Redistribution (1)	Competitiveness (2)	Cooperation (3)	Coordination (4)
Movie × mixed study group	−0.0375	−0.0666	0.171	0.184
p-value	(0.5956)	(0.3455)	$p < 0.01$	$p < 0.01$
Sharpened *q*-value	[0.999]	[0.939]	[0.001]	[0.001]
Romano-Wolf corrected *p*-value	{0.9820}	{0.8472}	{0.001}	{0.0010}
Movie-curriculum × mixed study group	−0.0406	−0.0358	0.299	0.333
p-value	(0.5457)	(0.6251)	$p < 0.01$	$p < 0.01$
Sharpened *q*-value	[0.999]	[0.999]	[0.001]	[0.001]
Romano-Wolf corrected *p*-value	{0.9820}	{0.9820}	{0.0010}	{0.0010}
U × mixed study group	−0.0661	−0.0219	−0.00630	−0.0168
p-value	(0.3066)	(0.7428)	(0.8673)	(0.6270)
Sharpened *q*-value	[0.939]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected *p*-value	{0.8052}	{0.9820}	{0.9820}	{0.9820}
M × mixed study group	−0.0812	−0.0961	−0.0230	0.0122
p-value	(0.2064)	(0.1513)	(0.5444)	(0.7247)
Sharpened *q*-value	[0.703]	[0.571]	[0.999]	[0.999]
Romano-Wolf corrected *p*-value	{0.5894}	{0.4266}	{0.9820}	{0.9820}
Playing with opposite gender	Yes	Yes	Yes	Yes
Individual controls and school FE	Yes	Yes	Yes	Yes
Observations	9,145	9,145	9,145	9,145
R^2	0.008	0.013	0.610	0.331

Notes: Dependent variables are outcomes on redistribution, competitiveness, cooperation, and coordination games, respectively standardized to mean zero and standard deviation one. The outcomes are recorded 12 months after the treatment. The outcomes are recorded 12 months after the treatment. *Utilitarian* variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the *Malleability* treatment indicator. *Visual narrative* (*movie*) is similarly a dummy turning on for subjects assigned the *Bol* movie. The *Joint movie and curriculum* indicator turns on for teachers assigned the joint *Bol* movie and the gender-rights curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Supplemental Appendix Table A3. The treatments are compared relative to the placebo treated control group. *p*-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR *q*-values in square brackets and FWER-adjusted *p*-values in curly braces. Further details on this are provided in Supplemental Appendix D8 (considering 16 hypotheses with 4 treatments × 4 outcomes). Each student plays the game twice, with the same gender and with the opposite gender. The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., kindergarten, nursery, prep, one, two, three, four, five, and six class) and pretreatment math scores.

cooperation and coordination in the behavioral games. Both the visual narrative treatment and the joint treatment positively impact cooperation and coordination in strategic dilemmas commonly used to measure social preferences of children. The coefficient estimates imply that the visual narrative alone increased cooperation and coordination by about 0.15 standard deviations, while the joint treatment increased cooperation and coordination by about 0.3 standard deviations (Table 5). Redistribution and competitiveness are unaffected. The behavior of students playing the strategic dilemma against the same gender (Supplemental Appendix Table B2) also does not seem to be impacted by any of our treatments, while Supplemental Appendix Figure B9 provides corresponding raw averages of the four treatments on games fielded with the opposite gender and Supplemental Appendix Figure B10 provides these effects by mixed-gender and same-gender study groups. Taken together, these findings suggest that the gender-rights curriculum combined with mixed-gender study groups may have enhanced students'

theory of mind toward the opposite gender—the ability to take another person's perspective. More gender-equal attitudes may have arisen through interactions with the opposite gender, promoting both theory of mind and transmission of progressive gender attitudes. The two reduced form effects on cooperation and on student achievement corroborates a body of scholarship that advocates cooperative learning enhances mathematics learning (Gutiérrez 2002; Weissglass 2001; Corte 2004). These findings suggest a potential mechanism explaining why inter-gender contact may improve mathematics achievement. Effective cooperation across genders may yield benefits in student achievement.

D. *Comparison with Studies of Single versus Mixed-Gender Schools*

We observe positive effects of cross-gender interaction in Pakistani primary schools, a finding that stands in contrast to Briole (2021), who found in France that while a higher proportion of female peers boosts girls' academic performance and future educational success, it concurrently steers boys toward vocational paths and diminishes their likelihood of graduating. Likewise, Jackson (2021) also reported benefits from transitioning coeducational schools to single-sex schools in Trinidad and Tobago, noting improved exam results for boys, more rigorous coursework for both sexes, and declines in arrests and teen pregnancies. Our study's distinct context, focusing on primary education for children aged 5 to 12, and explicit focus on cultivating progressive gender norms in a child's formative years may explain these different results. During these critical early years, implementing forward-thinking gender interventions can nurture positive attitudes toward the opposite gender and effectively challenge the harmful behaviors and stereotypes that may be observed in earlier work. Furthermore, the students in the early teenage years analyzed in the studies mentioned previously may have already been affected by established social norms that discourage cross-gender interaction. Consequently, concentrating on a younger age group that may be more receptive to increasing inter-gender cooperation may also account for the differing outcomes observed in our study.

VI. Conclusions

The past century has witnessed tremendous growth in recognition of rights and freedoms across group boundaries. This paper explores how teachers may transmit gender attitudes to their students, through a field experiment in Pakistan. We implement a randomized control trial testing different methods of shifting teachers' and students' views regarding equitable gender rights. We field teacher-training treatments based on the utility of empathy, malleability of empathy, a visual narrative arguing for female rights, and a joint visual narrative and gender-rights curriculum, as well as a control training.

We find that training teachers using a visual narrative shifted the teacher's attitudes toward more equitable gender rights. Teacher attitudes impacted both student attitudes and students' behavior toward the opposite gender. The effect sizes are substantial. Teachers' attitudes measured in gender IATs shifted by 0.2 standard

deviation. The teachers also became 10 percentage points more likely to sign a petition sent to the Pakistani parliament to criminalize mens' demands for a dowry and abolish polygamy laws. Students' attitudes shifted toward gender equality by at least 0.1 standard deviation. Reinforcing the visual narrative shown to teachers with the gender-rights curriculum improved student achievement and magnified the shifts in students' attitudes. Teachers assigned to the joint visual narrative and gender-rights curriculum treatment shifted their own attitudes and behavior by an additional 0.35 standard deviations in IAT scores; they were 18 percentage points more likely to sign gender rights petitions sent to the parliament. We interpret this heightening of treatment effects by teachers exercising self-persuasion via teaching. Students' attitudes on more equitable gender rights, students' behavior toward the opposite gender in games reflecting theory-of-mind, and students' math test scores were all also impacted under the joint visual narrative and curriculum treatment.

Our causal mediation design experimentally identifies a mechanism explaining the increase in student achievement. Cross-randomizing teachers to assign students into mixed-gender or same-gender study groups shifts students' attitudes. Transmission of attitudes occurred only when classrooms were assigned to mixed-gender study groups, as reflected in behavioral games and math scores. Inter-gender cooperation and coordination increased in games involving social dilemmas, suggesting improved theory of mind regarding others' actions.

The gender gap in math test scores essentially disappeared in classrooms assigned to form mixed-gender study groups, likely due to the treatment increasing cooperation and coordination with the opposite gender. Overall, our results provide experimental evidence that gender attitudes are transmissible from teachers to students and underscore the potential benefits of contact with the opposite gender in learning environments.

While gender separation of school children in Pakistan is a potential practice post grade 6 education, it is by no means a standard across all educational institutions in Pakistan, with many private schools favoring a mixed-gender education. Although, we have documented the immediate and intermediate effects of gender separation, it is important to consider the possibility that these effects may not be enduring or could reverse in the context of higher education, particularly when the practice of mixed-gender schooling is not sustained. A more complete understanding long-term impact of gender separation in education merits additional scrutiny.

REFERENCES

Andrabi, Tahir, Natalie Bau, Jishnu Das, Naureen Karachiwalla, and Asim Ijaz Khwaja. 2024. "Crowding in Private Quality: The Equilibrium Effects of Public Spending in Education." *Quarterly Journal of Economics* 139 (4): 2525–77.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking." *Quarterly Journal of Economics* 136 (4): 2147–94.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. 2018. "Revealing Stereotypes: Evidence from Immigrants in Schools." NBER Working Paper 25333.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128 (2): 469–530.

Banerjee, Abhijit, Eliana La Ferrara, and Victor Orozco. 2019. "Entertainment, Education, and Attitudes toward Domestic Violence." *AEA Papers and Proceedings* 109: 133–7.

Bénabou, Roland, Arim Falk, and Jean Tirole. 2018. "Narratives, Imperatives, and Moral Reasoning." NBER Working Paper 24798.

Bisin, Alberto, and Thierry Verdier. 2011. "The Economics of Cultural Transmission and Socialization." In *Handbook of Social Economics*, Vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 339–416. North-Holland.

Mansoor, Shoaib, director. *Bol*. Johns Hopkins University, 2011**.**

Briole, Simon. 2021. "Are Girls Always Good for Boys? Short and Long Term Effects of School Peers' Gender." *Economics of Education Review* 84: 102150.

Chen, Daniel, Sultan Mehmood, and Shaheen Naseer. 2022. "Intergenerational Transmission of Gender Norms: Experimental Evidence from Pakistan." AEA RCT Registry. https://doi.org/10.1257/rct.7465-1.3.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–632.

Corno, Lucia, Eliana La Ferrara, and Justine Burns. 2022. "Interaction, Stereotypes, and Performance: Evidence from South Africa." *American Economic Review* 112 (12): 3848–75.

Corte, Erik De. 2004. "Mainstreams and Perspectives in Research on Learning (Mathematics) from Instruction." *Applied Psychology* 53 (2): 279–310.

Crowne, Douglas P., and David Marlowe. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24 (4): 349–54.

Devine, Patricia G., Patrick S. Forscher, Anthony J. Austin, and William T. L. Cox. 2012. "Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention." *Journal of Experimental Social Psychology* 48 (6): 1267–78.

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2022. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." *American Economic Review* 112 (3): 899–927.

Doepke, Matthias, and Michèle Tertilt. 2009. "Women's Liberation: What's in It for Men?" *Quarterly Journal of Economics* 124 (4): 1541–91.

Doepke, Matthias, and Fabrizio Zilibotti. 2017. "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica* 85 (5): 1331–71.

Duckworth, Angela Lee, and Martin E. P. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology* 98 (1): 198–208.

Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature* 50 (4): 1051–79.

Eigen, Zev J., and Yair Listokin. 2012. "Do Lawyers Really Believe Their Own Hype, and Should They? A Natural Experiment." *Journal of Legal Studies* 41 (2): 239–67.

Eskreis-Winkler, Lauren, Katherine L. Milkman, Dena M. Gromet, and Angela L. Duckworth. 2019. "A Large-Scale Field Experiment Shows Giving Advice Improves Academic Outcomes for the Advisor." *Proceedings of the National Academy of Sciences* 116 (30): 14808–10.

Felson, Richard B., and Lisa Trudeau. 1991. "Gender Differences in Mathematics Performance." *Social Psychology Quarterly* 54 (2): 113–26.

Fernández, Raquel. 2014. "Women's Rights and Development." *Journal of Economic Growth* 19 (1): 37–80.

Fernández, Raquel, and Joyce Wong. 2014. "Unilateral Divorce, the Decreasing Gender Gap, and Married Women's Labor Force Participation." *American Economic Review* 104 (5): 342–7.

Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore. 2021. "On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms." *American Economic Review* 111 (7): 2342–75.

Giuliano, Paola. 2020. "Gender and Culture." *Oxford Review of Economic Policy* 36 (4): 944–61.

Giuliano, Paola, and Nathan Nunn. 2021. "Understanding Cultural Persistence and Change." *Review of Economic Studies* 88 (4): 1541–81.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin 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, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97 (1): 17–41.

Gutiérrez, Kris D. 2002. "Studying Cultural Practices in Urban Learning Communities." *Human Development* 45 (4): 312–21.

Hyde, Janet S., Sara M. Lindberg, Marcia C. Linn, Amy B. Ellis, and Caroline C. Williams. 2008. "Gender Similarities Characterize Math Performance." *Science* 321 (5888): 494–5.

Isensee, Laura. 2016. "US Charter School Model Goes Global in Pakistan." Houston Public Media. https://www.houstonpublicmedia.org/articles/news/educationnews/2016/11/29/178120/us-charter-school-model-goes-global-inpakistan/ (accessed March 12, 2021).

Jackson, C. Kirabo. 2021. "Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood?" *Journal of Human Resources* 56 (1): 1–39.

John Hopkins Center for Communication Programs. 2012. "Film Co-Produced by JHU-CCP Wins at Pakistani Awards Show." https://ccp.jhu.edu/2012/08/03/film-co-produced-by-jhu%E2%88%99ccp-sweeps-prestigious-pakistani-awards-show/ (accessed February 10, 2021).

Kersey, Alyssa J., Emily J. Braham, Kelsey D. Csumitta, Melissa E. Libertus, and Jessica F. Cantlon. 2018. "No Intrinsic Gender Differences in Children's Earliest Numerical Abilities." *npj Science of Learning* 3: 12.

Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz, and Lian Sanbonmatsu. 2004. "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment." Preprint, SSRN. https://dx.doi.org/10.2139/ssrn.588942.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2): 434–67.

List, John A., Azeem M. Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4): 773–93.

Lowe, Matt, 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review* 111 (6): 1807–44.

Maurin, Eric, and Sandra McNally. 2008. "Vive la Révolution! Long-Term Educational Returns of 1968 to the Angry Students." *Journal of Labor Economics* 26 (1): 1–33.

Mehmood, Sultan, Shaheen Naseer, and Daniel L. Chen. 2023. "Why Are Rights Revolutions Rare?" Unpublished.

Mehmood, Sultan, Shaheen Naseer, and Daniel L. Chen. 2024. "Altruism in Governance: Insights from Randomized Training for Pakistan's Junior Ministers." *Journal of Development Economics* 170: 103317.

Mehmood, Sultan, Shaheen Naseer, and Daniel L. Chen. 2025. *Data and Code for: "Transmitting Rights: Effective Cooperation, Inter-gender Contact, and Student Achievement."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E208265V1.

National Center for Education Statistics. 2019. "Digest of Education Statistics." https://nces.ed.gov/programs/digest/d19/tables/dt19_105.50.asp (accessed June 21, 2021).

Pomerantz, Eva M., Ellen Rydell Altermatt, and Jill L. Saxon. 2002. "Making the Grade but Feeling Distressed: Gender Differences in Academic Performance and Internal Distress." *Journal of Educational Psychology* 94 (2): 396–409.

Progressive Education Network. 2022. "Our Mission." https://pen.org.pk/ (accessed December 13, 2020).

Riley, Emma. 2019. "Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment." *Review of Economics and Statistics* 106 (2): 334–51.

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

Seror, Avner. 2022. "Child Development in Parent-Child Interactions." *Journal of Political Economy* 130 (9): 2462–99.

Schwardmann, Peter, Egon Tripodi, and Joël J. Van der Weele. 2022. "Self-Persuasion: Evidence from Field Experiments at International Debating Competitions." *American Economic Review* 112 (4): 1118–46.

Weissglass, J., 2001. "No Compromise on Equity in Mathematics Education: Developing an Infrastructure." In *Changing the Faces of Mathematics: Perspectives on Multiculturalism and Gender Equity*, edited by Walter G. Secada, Judith E. Jacobs, Joanne Rossi Becker, and Gloria F. Gilmer, National Council of Teachers of Mathematics. https://www.researchgate.net/publication/242590206.

Weisz, E., and J. Zaki. 2017. "Empathy Building Interventions: A Review of Existing Work and Suggestions for Future Directions." In *The Oxford Handbook of Compassion Science*, edited by E. M. Seppälä, E. Simon-Thomas, S. L. Brown, M. C. Worline, C. D. Cameron, and J. R. Doty, 205–17. Oxford University Press.