

Self-efficacy and Citizen Engagement in Development: Experimental Evidence from Tanzania

Supplementary Information for Online Publication*

Evan Lieberman[†]

Yang-Yang Zhou^{‡§}

December 6, 2020

*The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: doi.org/10.7910/DVN/SLYEIJ (Zhou and Lieberman, 2020). Our pre-analysis plans can be found at osf.io/jqzxp for Study 1 and osf.io/9xmjg for Study 2. This research received institutional review board (IRB) approval from MIT COUHES (#1603517857R001). We report no conflicts of interest. All errors and omissions are ours.

[†]Department of Political Science, MIT. Email: evanlieb@mit.edu. Address: 77 Massachusetts Avenue E53-409, Cambridge, MA 02142. Website: evanlieberman.org Twitter: @evlieb

[‡]Department of Political Science, University of British Columbia. Email: yangyang.zhou@ubc.ca. Address: 323 C.K. Choi Building 1855 West Mall, Vancouver, BC, V6T 1Z2. Website: yangyangzhou.com Twitter: @yangyang_zhou

[§]Authors contributed equally. Author order randomized using randomizeauthor.shinyapps.io/shiny

Supplementary Information

Contents

I	Supplementary Information	2
1	Descriptive Statistics and Figures	3
1.1	Study 1 Bukoba Baseline Summary Statistics	4
1.2	Study 2 Kilosa Summary Statistics	5
2	Validated Participation Intervention Details	6
2.1	Information on School Quality, Rights and Responsibilities	6
2.2	Components unique to Validated Participation	7
3	Educational Self-Efficacy Scale	12
3.1	Educational Self-Efficacy Scale Index Questions	12
3.2	Correlation between Educational and General Self-Efficacy Scores	13
4	Study 1 Bukoba Implementation Plan	14
5	Study 2 Kilosa Implementation Plan	18
6	Risk and Ethical Considerations	25
7	Deviations from Pre-Analysis Plans	27
8	Association between Active Citizenship and Educational versus General Self-Efficacy	28
9	Bukoba School-Level Interviews and Outcomes	29
10	Binary Coding of Study 1 Follow-up with Teachers on School-Level Outcomes	42
11	Regression Adjustment using Lin’s Regression	43
11.1	Study 1: Effects on Parent and Student Outcomes in Bukoba Pilot	43
11.2	Study 2: Effects on Parent Outcomes in Kilosa	44
12	Multiple Hypothesis Testing using Benjamini-Hochberg Procedure	45
13	Regression Tables	49

1 Descriptive Statistics and Figures

Figure S1 shows the study schools in our two study districts, Bukoba and Kilosa. We selected these districts because they are largely rural and have a substantial number of schools performing under the national average. We avoided already high-performing schools, because we wanted to identify schools for which we could credibly inform citizens that their schools needed improvement. We also avoided the lowest performing schools, under the assumption that there might be unobserved constraints on their performance, making the effects of our treatment unlikely.

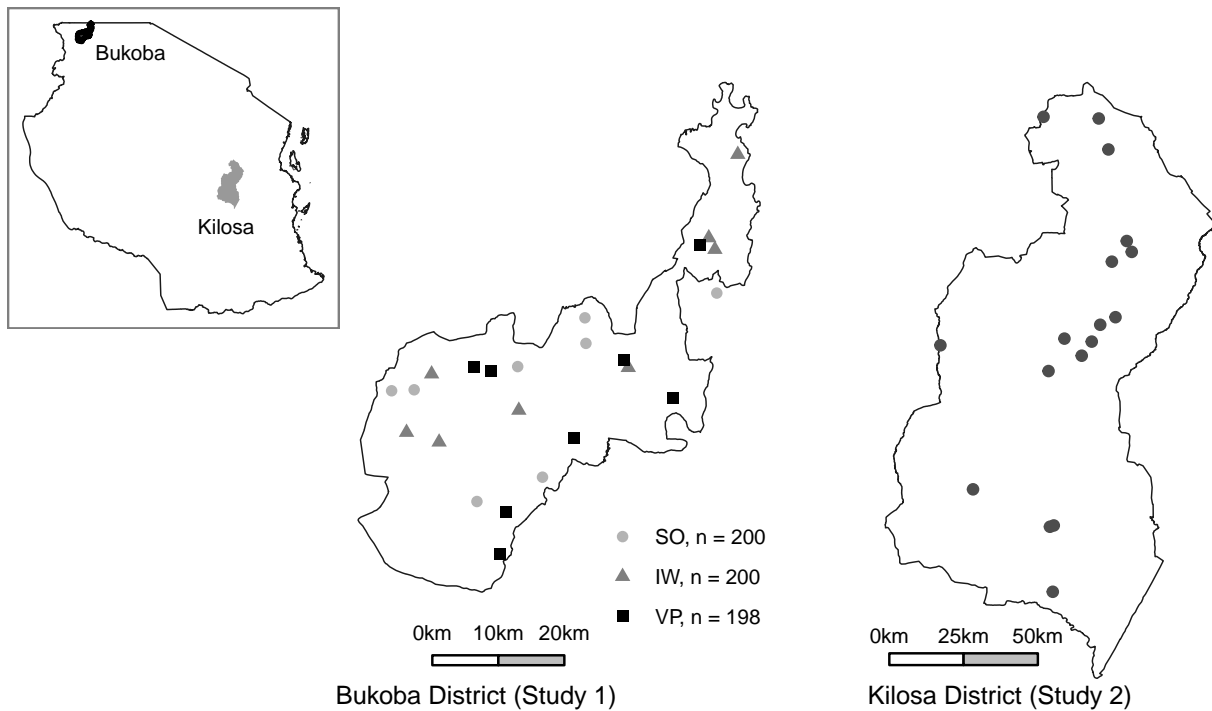


Figure S1: The left map shows the 24 schools in Study 1, located in Bukoba rural district, shown in black in the map of Tanzania (inset). These schools were randomized into the three intervention groups: Survey Only (circle), Information Workshop (triangle), Validated Participation (square), with their respective parent sample sizes. The right map shows the 16 schools in Study 2, located in Kilosa rural district, shown in gray in the map of Tanzania (inset). We randomized the treatments *within* each of these schools.

1.1 Study 1 Bukoba Baseline Summary Statistics

Table S1: Descriptive statistics of Study 1 Bukoba participants at baseline.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Household Size	598	6	2	2	5	6	7	25
Age	598	44	12	19	36	43	51	95
Education Level	596	3	1	1	3	3	3	7
SES Index	598	4	2	0	2.2	4	5	9
No. of Children	598	3	1	0	2	2	3	7
Female	598	0	0	0	0	0	1	1
Household Head	598	1	0	0	1	1	1	1
Literate	598	1	0	0	1	1	1	1
Muslim	598	0	0	0	0	0	0	1
Political	596	1	0	0	1	1	1	1
Parent (vs. Caregiver)	598	1	0	0	1	1	1	1
Child w. Disabilities	598	0	0	0	0	0	0	1
On School Committee	586	0	0	0	0	0	0	1
Edu Efficacy	598	4	1	1	3	4	4	5
Edu Internal Efficacy	598	4	1	1	3.5	4	4.8	5
Edu External Efficacy	598	4	1	1	3	4	4	5
Edu Individual Efficacy	598	4	1	1	3.4	4	4.4	5
Edu Group Efficacy	598	4	1	1	3.5	4	4.5	5
Gen Efficacy	598	31	6	15	27	31	35	40
Checks Homework	596	1	0	0	0	1	1	1
Gives Breakfast	595	1	0	0	0	1	1	1
Checks Teacher Absence	598	0	0	0	0	0	1	1
Child's Test Score (/72)	598	22	11	0	13	21	29	68

1.2 Study 2 Kilosa Summary Statistics

Table S2: Descriptive statistics of Study 2 Kilosa participants.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Household Size	1,633	6	2	2	5	6	7	15
Age	1,633	40	11	18	32	38	46	83
Education Level	1,630	3	1	1	2	3	3	7
SES Index	1,633	4	2	0	2	4	5	10
No. of Children	1,633	2	1	0	2	2	3	8
Female	1,633	1	0	0	0	1	1	1
Household Head	1,633	1	0	0	0	1	1	1
Literate	1,633	1	0	0	1	1	1	1
Muslim	1,633	0	0	0	0	0	1	1
Political	1,633	0	0	0	0	0	1	1
Parent (vs. Caregiver)	1,633	1	0	0	1	1	1	1
On School Committee	1,631	0	0	0	0	0	0	1
Edu Efficacy	1,633	4	1	1	3	4	4	5
Edu Internal Efficacy	1,633	4	1	1	3	3.8	4.2	5
Edu External Efficacy	1,633	3	1	1	3	3	4	5
Edu Individual Efficacy	1,633	4	1	1	3	4	4	5
Edu Group Efficacy	1,633	4	1	1	3	4	4	5
Gen Efficacy	1,633	30	6	11	25	30	34	40
Checks Homework	1,633	1	0	0	0	1	1	1
Gives Breakfast	1,633	0	0	0	0	0	1	1
Checks Teacher Absence	1,633	1	0	0	0	1	1	1
Respected by Parents	1,633	4	1	2	3	4	5	5
Respected by Authorities	1,633	4	1	1	3	4	5	5
Ideas Respected	1,633	3	1	1	2	3	4	5
Raise issue at Mtg	1,633	4	1	1	3	4	5	5
Raise issue w. Official	1,633	4	1	1	3	4	5	5
Support Pro-edu Candidate	1,633	1	0	0	0	1	1	1

2 Validated Participation Intervention Details

Our intervention, Validated Participation, is a series of meetings/or one meeting designed to give participants not only information but also opportunities for successful engagement (mastery experiences), validated by peers and an authority figure. In order to evaluate whether this intervention will affect parents in the manner theorized above, we compare estimated changes in parental attitudes and behaviors as well as changes in student performance among those parents who receive this intervention relative to two other groups: a set of parents who simply receive *information* about the quality of schooling and their rights and responsibilities (Information Workshop); and a set of parents who will receive no intervention (Survey Only/Control). Thus, building on the theoretical discussion above, in this section, we describe the three treatment arms of our study – Validated Participation (VP), Information Workshop (IW), and Survey Only/Control (SO) – which are randomly assigned to school communities. SO schools were not assigned to any meetings, only the baseline and endline measurement, while IW and VP meetings were delivered by trained facilitators with teaching backgrounds over four approximately 90-minute meetings, spaced one week apart.

Next, we describe the components of each treatment arm in detail.

2.1 Information on School Quality, Rights and Responsibilities

For both VP and IW intervention arms, in the context of meetings, trained facilitators provided parents with information on the progress and needs of their school, what they should come to expect from teachers and school administrators, and what they can do to support their children’s education at school and at home. Even if this information is not entirely new to parents, we suspect that its provision is necessary to prompt action. The information provided was gathered from official Ministry of Education reports and from our partner, Twaweza, which specializes in the field of education.

This information emphasized strategies that parents can implement to contribute to their children’s education *without* necessitating either systemic change in schools or resources and skills that parents do not already possess. Specifically, the facilitator was trained to: share information that their schools are performing below the national average (a criteria for inclusion into the study); have the head teacher describe the needs of the school; explain how parents can support the roles and responsibilities of teachers and head teachers; provide materials about how parents can also

improve their children’s learning at home; and explain that education quality is a strong predictor of future economic success.

All *information* delivered in the IW treatment arm was also delivered in the VP arm. The only difference, as we elaborate below, concerned the *style of delivery*. While instructors in both treatment arms were trained to be polite and to answer questions to the best of their ability, in IW meetings, facilitators were instructed to focus on mere delivery of the facts and to present in an authoritative style that is typical of the context: a teacher lecturing the information while students (in this case, parents) listen. In VP meetings, facilitators were instructed to use a more approachable style, and consistent with the intent of the treatment, validate parents’ participation with positive affirmation about the quality of their questions.

Importantly, because both the IW and VP interventions involve meetings that bring parents together at the primary schools of their children, any estimated differences in treatment arms cannot be due to the effects of overcoming obstacles to coordination or collective action.

2.2 Components unique to Validated Participation

We designed Validated Participation to provide efficacy-boosting engagement experiences for targeted parents. Various components are repeated across meetings, but updated with new content, with the hope that repetition would lead to reinforcement of the same objective: *to provide parents meaningful opportunities to envision themselves as impactful agents of change, capable of making decisions using evidence, and being valued for both their ideas and their actions*. Moreover, the proposed intervention was conducted on school premises alongside information about education (discussed above), which aimed to increase parents’ comfort level with active engagement with school officials and with their own children around matters pertaining to education.¹

We view the challenge of boosting parental efficacy in a relatively short period of time to be a substantial hurdle. Perceptions of efficacy are generally built up over a lifetime of experiences. Along these lines one could easily imagine that a more thorough-going intervention conducted over an even longer period of time might have more enduring effects. But as a research study, it

¹While this intervention makes use of discussion and deliberation, our theory is quite distinct from that found in the heart of the scholarly literature on the benefits of deliberation for democratic performance (e.g. Przeworski, 1998; Stokes, 1998), because our outcome of interest is changes in self-efficacy rather than changes in preferences for democracy or other institutions. Empirical research examining how deliberative decision-making and voting affect efficacy finds some impact on external efficacy but no changes to internal efficacy (Finkel, 1985; Morrell, 2005). Thus, Validated Participation not only provides opportunities for discussion and collective decision-making, but also the crucial validation that we theorize will jump-start feelings of both internal and external efficacy.

would also be increasingly more expensive to administer, and would be likely to suffer from greater problems of attrition. The central goal of this experiment is to investigate whether a short-course of intensive validation can substantively affect those perceptions and to evaluate associated behavior change.

1. Validated Participation using Visual Evidence

The first module, *Visual Evidence*, provides parents the opportunity to meaningfully discuss evidence by looking at and engaging with a series of artworks in a facilitated discussion.² The goal of this module is to give all parents, not just the elite, opportunities to offer their ideas and receive validation.

We pre-selected a series of images with the goal of having broadly relevant and interesting narrative potential so individuals can easily offer their analyses while many interpretations are possible.⁴ The subject matters center around family life and education, but they do not provide any new information to parents about schools.

Facilitators were trained to work with the group of parents, in a series of meetings, and focus on getting individuals in that group to respond to three key questions for each work of art presented:

- What is going on in this picture?
- What do you see that makes you say that?
- What more can you find?

Throughout the process, the facilitator repeats and emphasizes what each participant says (e.g. “no right or wrong answers”), *echoes back and validates their ideas* as positive encouragement, and points out relevant details in the image, ultimately helping to broadcast and to reinforce what each person says.

While recognizing that discussing art is an unconventional strategy for promoting efficacy, we believe that this component is an important first step for several reasons. This tool provides opportunities for parents to connect facts and to make inferences about new materials. Because neither outside knowledge, formal schooling, nor literacy is necessary for participants

²This strategy is inspired by and adapted from an existing curriculum called *Visual Thinking Strategies (VTS)*³, which was designed in the U.S. primarily but not exclusively for school-aged children that aims to increase participation and critical thinking skills through guided discussion of art. We modified this tool for use on adults in an East African setting, but with the common aim of providing an opportunity to facilitate engagement of the type we believe is likely to boost efficacy.

⁴We selected these images in consultation with VTS trainers and Twaweza staff.

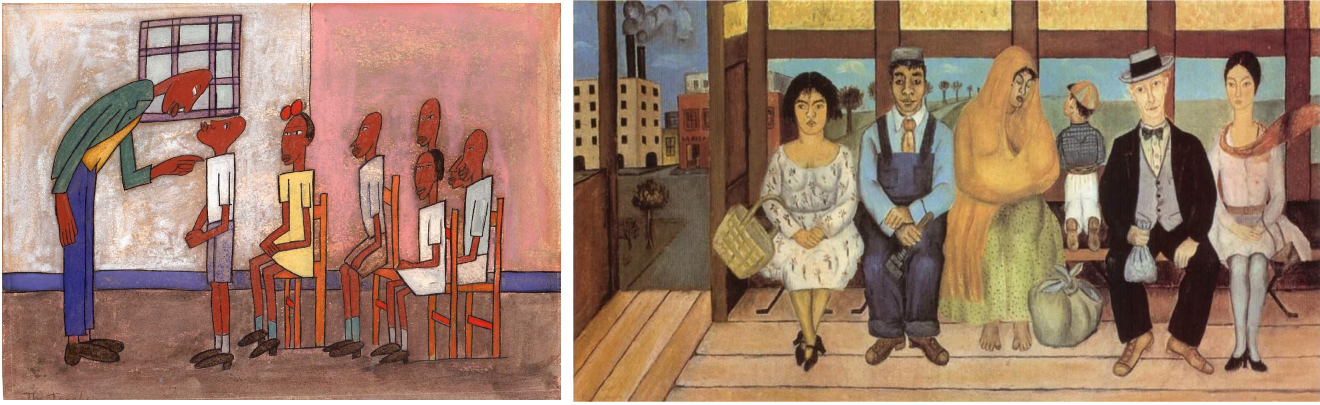


Figure S2: Example images for Visual Evidence

to be successful in discussing art, this tool can serve as a leveler by encouraging participation by all parents not just elite, educated ones. This mitigates concerns of elite capture in which only a few parents, generally older and male with already high efficacy beliefs, will set the agenda for the remaining components of the intervention. To further prevent elite capture, our intervention protocol explicitly instructs facilitators to encourage broad participation across age, gender, religion, etc.

For those who speak up, this exercise provides them an opportunity to see that their ideas are heard and respected by the facilitator (mastery experience). For those who do not actively participate, at least they could observe parents like themselves who do participate and receive positive feedback (vicarious experiences). In short, efficacy in the political arena demands comfort with being able to express one’s opinions about evidence, and Visual Evidence aims to start building that comfort.

2. Validated Participation using Resource Vote:

The second module, *Resource Vote*, sought to make explicit connections between an initial boost of self confidence developed in the Visual Evidence module and parents’ role in the educational sphere. Specifically, it demonstrated to parents who may have little experience with impactful decision-making – particularly in the educational sphere – that their ideas and actions *can* result in the generation of concrete outcomes. Inspired in part by Pradhan et al. (2011) in which parents’ election of school committee members sparked other forms of parent participation in education, Resource Vote was designed to allow parents to directly “elect” how to spend grants for their schools. Ultimately, the idea of Resource Vote is to create opportunities for engagement that over time changes parents’ beliefs about their roles,

responsibilities, and abilities to impact their children’s learning. Specifically, parents were informed that an outside NGO (Twaweza) would like to make a modest contribution to the school, but that the parents must collectively decide how to spend the money. Presented with just two choices, parents were given time to briefly deliberate on the options and then vote in a secret ballot. The option winning the majority of the votes was implemented almost immediately and was physically presented by the facilitator. The resource choices were chalk versus erasers; large rulers versus abacuses; and science posters versus maps of Tanzania. The relatively inexpensive resource options were selected in consultation with local teachers and head teachers as appropriate for responding to local education needs.

Through three iterations of this module, the scale of resources as well as the time from voting to implementation increased slightly. Following the first meeting’s Resource Vote, parents were immediately presented with the chosen resource by the facilitator. The second meeting’s chosen resources were presented during the school tour later in the session. And finally the third meeting’s chosen resources were presented in the fourth meeting a week later. The aim of this scaling-up was to develop beliefs that impact can be realized even if not observed immediately. Although all the resources were modest in scale, the facilitator was instructed to be upfront and emphasize that the aim of the exercise was for parents to practice discussing and deliberating school needs.

Since the goal of this intervention was to prompt parental input and not to measure the impact of the resources themselves, the schools in the other treatment arms received the same resources for which the VP groups voted in similar time intervals.

3. Validated Participation with Peer Validation:

This final module, *Peer Validation*, was implemented only during the last meeting, when the facilitator asked the parents to identify what they can and should do to help their children’s education. Recalling the steps of the previous modules, parents discussed and generated a list of ways they can help, not only echoed and validated by the facilitator but validated by each other as peers. Specifically, each contribution by a parent to this list was further recognized and validated by peers. This list was then publicly displayed in the school, further providing public recognition.

In sum, Validated Participation is a multi-component intervention, but each component was

carefully designed to be conceptually consistent with the theorized sources of efficacy: *parents are given equal opportunities to participate in critical discussions of evidence and decision-making in a successful manner, and their ideas and actions are validated by an authority figure and each other.*

3 Educational Self-Efficacy Scale

3.1 Educational Self-Efficacy Scale Index Questions

The following 9 questions comprise the Educational Self-Efficacy Scale (ESES), with responses ranging on a scale from 1 to 5.

1. If a group of parents are having a discussion about issues at the school, how confident are you that you could make a meaningful contribution to this discussion?
2. How confident are you that you could find out what is being taught in your child's classroom?
3. How confident are you that you could communicate with your child's teacher concerning an issue he faces at school?
4. How confident are you that you could communicate with your child's head teacher about any educational questions or other issues concerning you?
5. Suppose you made a recommendation during a discussion with other parents about issues at the school, how confident are you that the other parents will take action based on your recommendation?
6. Suppose that you had a discussion with your child's head teacher today and that you discussed, for example, the insufficiency of materials available at school for the children. How confident are you that he/she will remember what you discussed one week later?
7. How confident are you that he/she will act on the problem you talked about in your discussion?
8. Suppose that you with a group of other parents have spoken to the head teacher concerning this issue of lack of school equipment. How confident are you that the head teacher would act on the problem you discussed in your talk?
9. How confident are you that people like you can have any influence in making the school better?

Questions 1-4 measure *internal educational efficacy* while 5-9 measure *external educational efficacy*. Partitioned a different way, questions 2, 3, 4, 6, 7 make up the *individual educational efficacy index* as they ask about taking actions by oneself, while the *group educational efficacy index* comprises questions 1, 5, 8, 9, asking about actions taken as a group.

3.2 Correlation between Educational and General Self-Efficacy Scores

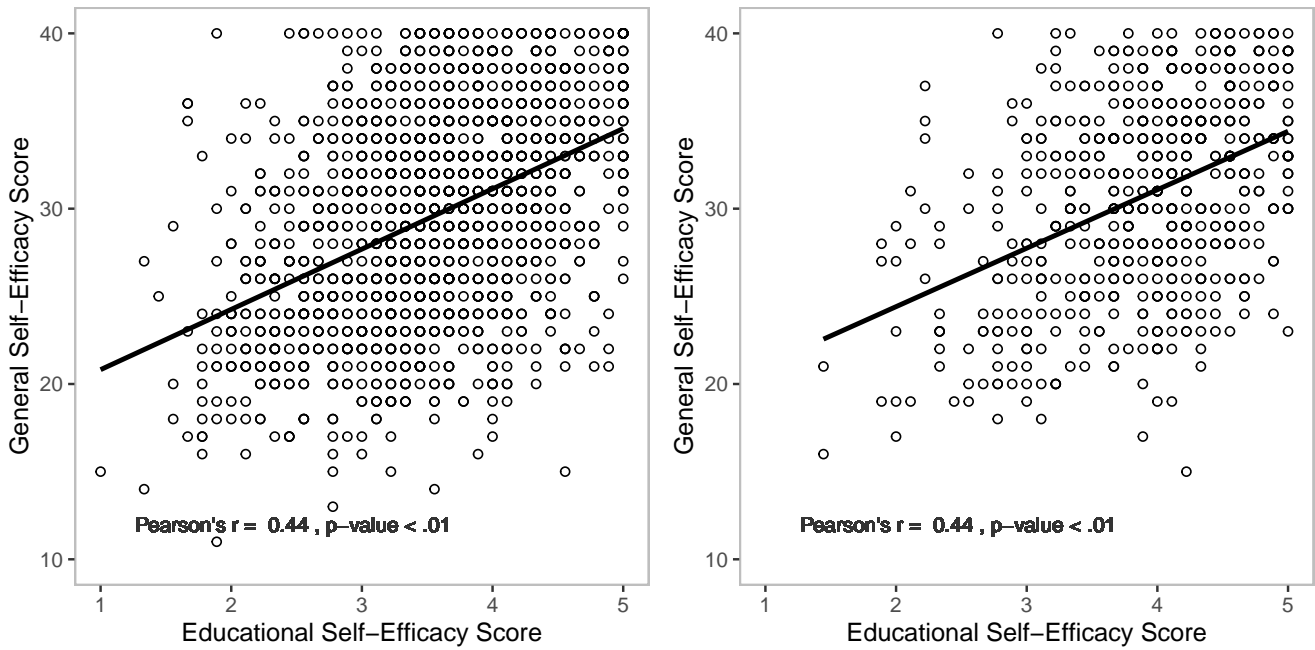


Figure S3: This figure shows the scatterplot and Pearson's correlation between Educational Self-Efficacy Score (x-axis) and the General Self-Efficacy Score (y-axis) for the Study 2 Kilosa (right) and Study 1 Bukoba baseline (left).

4 Study 1 Bukoba Implementation Plan

We conducted Study 1 from August 2016 to January 2017 in Bukoba rural district in Northern Tanzania. This district was selected largely for logistical convenience, because our implementing partner Economic Development Initiatives (EDI) Limited was based in the area. We subsequently learned that Bukoba has a relatively high concentration of people with educational degrees and parents sending their children to school, likely due to the fact it was the site of German missionary education efforts starting in 1910 as part of an effort to evangelize the population. Thus, we observed high levels of parent efficacy and active behavior at baseline, leading to what we believe were ceiling effects, at least for the purposes of our survey. In turn, estimated effects of treatment were likely conservative for this study.

Our target subject population was adult caregivers (following local usage, we refer to as ‘parents’ throughout this paper) of grade 3 public school children. We chose grade 3 in consultation with our partner as a critical age during which children are attaining important and observable learning skills, and it is an age when active parental involvement is likely to be consequential. Additionally, it is not a grade in which students take national standardized exams, so we would not expect extra attention from teachers and head teachers for these students. Based on practice meetings conducted in public schools in Dar es Salaam during summer 2015, we concluded that a meeting size of approximately 25 parents would be optimal.

Given those constraints, and based on cost and logistical concerns, we calculated that we could study 8 schools and the communities they serve (together, we refer to these units as “school communities”) in each of the three treatment arms, for a total of 24 school communities. Using the 2014 primary school data publicly available on the Tanzania Open Data platform⁵, we first subsetted down to 136 Bukoba rural, public schools. Next, given that the average ranking of these schools is 7473.68 (slightly better than the national average of 7933.75), we then subsetted down to 61 schools ranking below this district average. As this intervention is not intended for already high performing schools, and we wanted to identify schools for which we could credibly inform parents that their schools needed improvement. We avoided the lowest performing schools under the assumption that there might be unobserved constraints on performance on such schools, making the effects of our treatment unlikely.

⁵<http://www.opendata.go.tz/>

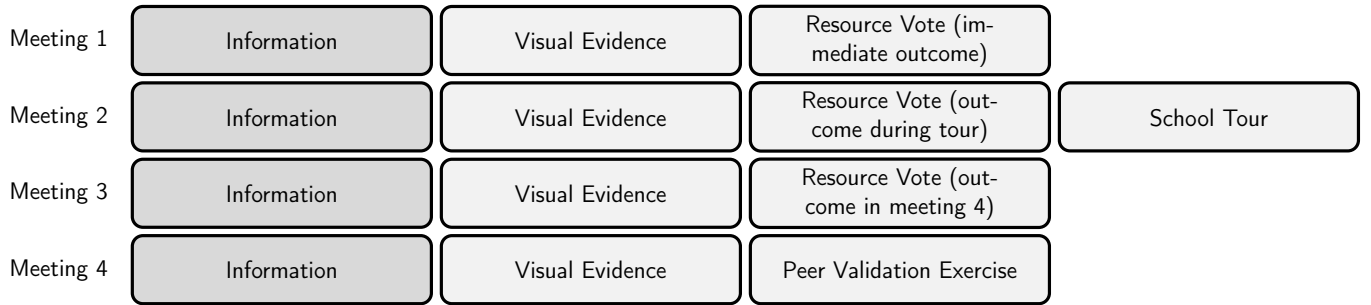


Figure S4: Implementation of Validated Participation meetings. For example, Meeting 1 consists of three modules: the Information intervention, the first Visual Evidence exercise, and the first Resource Vote exercise. Each meeting was approximately 90 minutes.

From this list of schools⁶, we used the R package `blockTools` (Moore, 2016) to match triplets of schools based on pupil/teacher ratio, rank, change in rank from previous year, enrollment, and distance in kilometers from the main urban center, Bukoba town. We took the eight best triplet matches based on smallest mahalanobis distances (for 24 school communities in total). Within each triplet match, we randomly assigned each school community to one of three study arms: VP, IW, and SO.

The head teacher of each school community was asked for permission and informed of the District Education Officer’s approval of the project to carry out the study and in all cases, this was granted. We asked the head teacher to provide us with a list of grade 3 students. From this list, we randomly selected 30 and invited their parents to a meeting regarding education on a given day. We arranged to have our enumeration teams conduct a lottery, stratified by gender, to select only 25 parents in the cases where more showed up. At the end of the study, each parent was given 5,000 TZS (equivalent to about 2USD) as a small token of appreciation and to offset any transportation costs.

For VP schools, Figure S4 shows the implementation of VP components across the four meetings.

Despite block randomization at the school level using school-level variables, the number of school units is relatively small, and across treatment groups we do find individual-level differences that are significant at the 90 percent confidence level. Individual-level variables for parents at each of the study schools were not available to us before the baseline on which we could block randomize. We do find significant baseline differences on our key outcome of interest, Educational

⁶We do not share school names or disaggregated data on schools to maintain anonymity

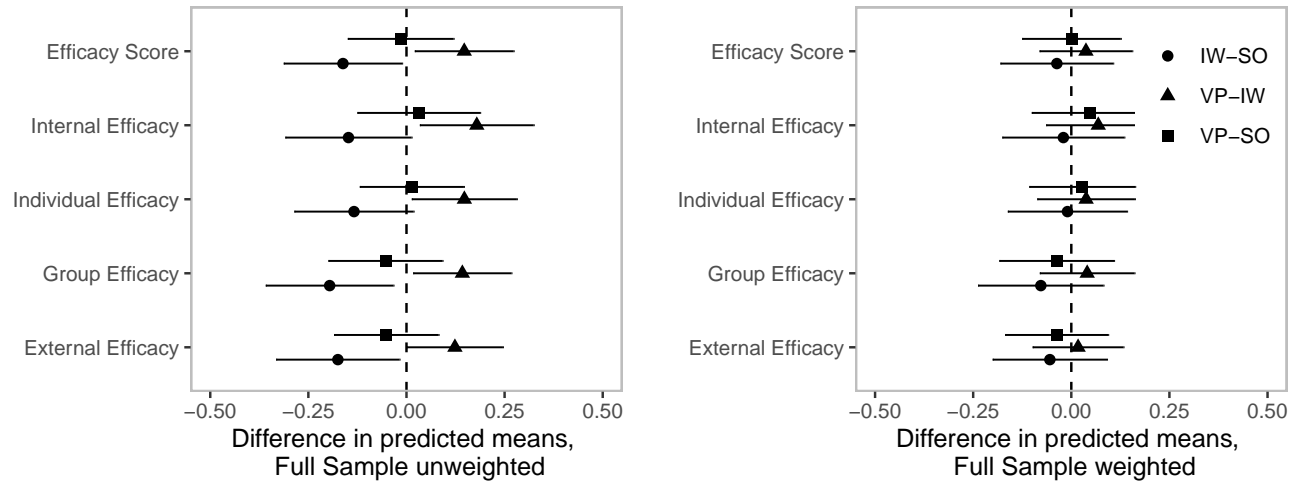


Figure S5: The first plot shows that ESES was unbalanced at baseline with IW parents reported lower efficacy than SO and VP parents reported higher efficacy than IW prior to the interventions. We use covariate balancing propensity score (CBPS) to generate weights for each parent, matched based on baseline ESES. The second plot shows the predicted difference-in-means once we use these weights.

Self-Efficacy Scale, thus we generate weights for each parent, down-weighting those parents whose baseline ESES scores contributed to imbalance. Figure S5 shows that parents randomly assigned to the IW treatment arm reported lower efficacy at baseline than those assigned to SO, and parents assigned to the VP arm reported higher efficacy than those assigned to IW prior to the interventions; thus, IW parents may have had more room to increase their sense of efficacy while VP parents may have encountered ceiling effects. Such important imbalances would weigh heavily against valid interpretation of naive (unadjusted) experimental results (Abadie, 2005). Thus we correct for these statistically by using the R package covariate balancing propensity score (CBPS) to generate weights for each parent, down-weighting those parents whose baseline scores contributed to imbalance at baseline (Fong, Ratkovic and Imai, 2014). The second plot shows the predicted difference-in-means once we use these weights, lending credibility to this approach as a reasonable correction.

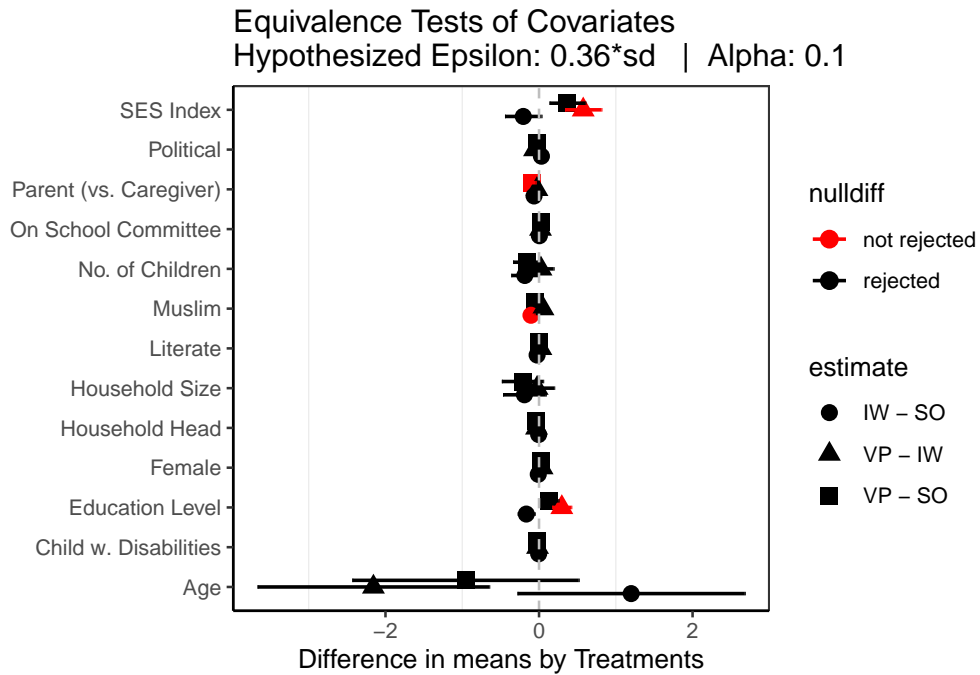


Figure S6: This figure shows the equivalence tests for the treatments across demographic covariates for Study 1. The covariates for which the null hypothesis of statistical difference was not rejected are SES index, Parent vs. Caregiver, Muslim, and Education level.

Using equivalence tests in lieu of less conservative balance tests (Hartman and Hidalgo, 2018), Figure S6 shows that the null hypothesis of statistical difference was not rejected for the following covariates: SES index, Parent vs. Caregiver, Muslim, and Education level. With such a small sample size and randomization at the school level, rather than parent level, we expected that some covariates would not be balanced. We adjust for all covariates in the analyses.

5 Study 2 Kilosa Implementation Plan

For the Kilosa study from May to June 2018, we again evaluated the effects of the Validated Participation (VP) protocol. Using data from OpenData and Uwezo (a Tanzanian NGO focused on education and a partner of Twaweza), in consultation with Twaweza, we selected Kilosa (Morogoro) for our study site. Kilosa District is one of the six districts of the Morogoro Region of Tanzania. Its administrative seat is the town of Kilosa. Kilosa District covers 14,918 square kilometres (5,760 sq mi). According to the 2012 Tanzania National Census, the population of the Kilosa District is 438,175.⁷

While Study 1 provided solid evidence of the plausibility of the theory, as anticipated, the strength of evidence in this study was limited, and we sought additional evidence that would be useful both for testing the overall theory of change, and for gathering information that would be useful from a cost effectiveness standpoint for implementation by organizations in the future. In Bukoba, we tested our novel intervention as a series of four meetings that we hypothesized would lead parents to feel a heightened sense of efficacy, and in turn to actively promote the education of their children at home and at school. While the content of the intervention was largely the same for the Kilosa study, in the latter, the intervention was conducted as a “single dose” in one day. We made this change for research and policy-related purposes – one meeting is less costly and logistically easier than four. However, we recognized that without the reinforcement of additional meetings, the impact might be more limited.

A second key difference from our prior study, in which we assigned treatment at the school level, was that in this study, we randomized at the individual-level within schools, and we planned for a much larger sample of subjects. For this study we chose a district in central Tanzania, Kilosa in Morogoro region to increase external validity and to addresses some of the “ceiling effects” from the prior study (discussed below).

In addition, we designed this study to immediately observe behavioral and behavior-related outcomes that should theoretically follow a boost in efficacy: specifically, a greater willingness to ask questions and to engage with authority figures at a meeting with a classroom teacher, and a greater desire to offer suggestions for how the school might improve. In the wake of a series of studies about information and behavior, our study provides a unique opportunity to observe

⁷http://www.tzdp.org.tz/fileadmin/documents/dpg_internal/dpg_working_groups_clusters/cluster_2/water/WSDP/Background_information/2012_Census_General_Report.pdf

whether behavior change is possible immediately following an intervention. While we recognize that the somewhat artificial setting of our study might have led some to behave in ways that are different from what would be observed in a purely natural setting, this should not have varied across treatment group. Importantly, we set up a “likely” test for a behavioral impact in the sense that subjects were provided a relatively low cost opportunity to take action directly after an intervention that ought to encourage participation.

Figure S7 shows the power analysis for this study. For an effect size of .1, which is similar to what we obtained from the Phase 1 study, with a power of .8, we aimed for 500 parents per treatment group or 1500 parents total. From the Bukoba study, we determined that no more than 25 participants was ideal for the VP activities – a sufficiently small group that would allow parents to engage in a lively discussion, with opportunities for every parent to speak up.

Within Kilosa, we selected 1 pilot school and 16 study schools randomly, conditional that their national rank and pass rates were in the 26-50 percentile. Two weeks prior to the study, an EDI staff member conducted pre-visits to each of these schools to speak with the head teacher to make sure that the schools were accessible. This staff member asked for permission and informed the head teachers of the District Education Officer’s approval of the project to carry out the study and in all cases, this was granted. We asked the head teachers to send out invitation letters to the parents in grades 2, 3, 4, and/or 5 (depending on how large the grades are, we aimed to get a total of 90-120 parents in total). Parents were asked to send back a confirmation slip of their participation with their gender marked so that we could try to achieve a mix of genders in each session.

On the date and time of their session, upon arrival, parents were greeted by the enumeration team, who carried out the screening and consent process to make sure that all potential subjects were over 18 years of age, a parent/primary caregiver to a child attending the school, and that they consented to participation in the study. At that point, the parents were randomly assigned, blocked by gender, to one of the three treatment arms – VP, IW, and SO, each taking place in a different classroom on the school grounds. We did not exclude any eligible parent, so if there were more than 60 parents or less than 45 parents per session, we still conducted the session.⁸

Figure S8 shows the overview of a session and which member(s) of the team conducted which

⁸We learned from the Phase 1 study that conducting a lottery so that we would only have 25 parents exactly per school (the treatments were at the school level) may have led to unforeseen dynamics of social desirability for the parents who ‘won’ the lottery. So for this study, we decided not to conduct another lottery.

Sample Size Estimation for 3 Treatment Groups (ANOVA), $\alpha=0.01$

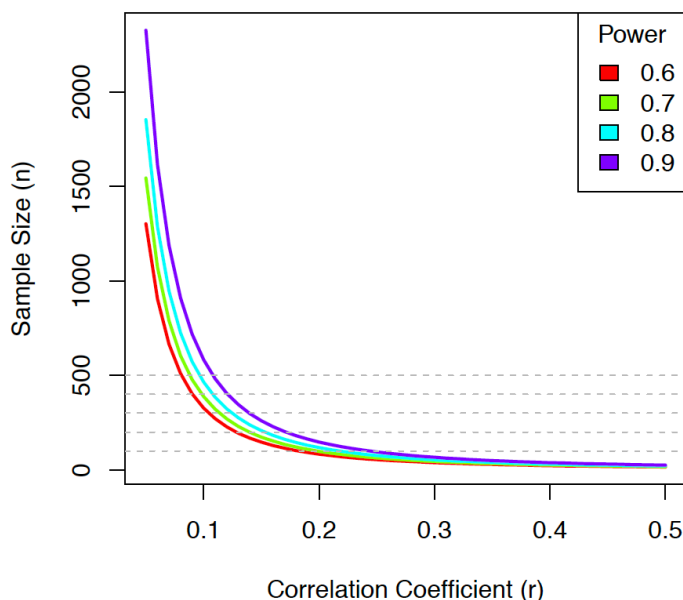


Figure S7: Power analysis for Single Dose study.

activity at a given time. The ordering of activities was deliberate to ensure that measurement occurred only post-treatment in the treatment arms and without treatment in the control arm, while also disguising randomization into different treatment arms to avoid experimenter effects. Subjects were assigned to classrooms, but at the time of measurement, they were not aware that the sequence of activities was different depending on classroom assignment.

Measurement consisted of a 15 minute survey, a group meeting in a classroom, and optional Office Hours with the Head Teacher. The SO (control) group only engaged in the measurement exercise. The IW and VP treatment groups both participated in an information module, including a 24 minute video, the contents of which were clips collected from HakiElimu⁹, which worked with Twaweza to create this content, as well as Minibuzz Tanzania¹⁰, a talkshow that discusses current issues with citizens. Together these clips showed 1) issues with education in Tanzania such as overcrowded classrooms and insufficient resources, 2) citizens discussing the role that parents can play in their children’s education, and 3) interviews with Twaweza staff members on these issues. For the IW treatment, the meeting facilitator simply showed this video prior to measurement. For the VP treatment, the meeting facilitator showed this video *and* conducted the VP activities prior to measurement.

⁹<http://hakielimu.org/>

¹⁰<http://miatv.co/miatv/minibuzz/>

Schedule of Each Session:

Start – End time	Survey Only	Information Workshop	Validated Participation
Randomization:	Eligibility Screening and Consent Procedure Give placard with unique ID # and registration by gender bin		
Direct parent to:	Survey Only classroom (refreshments)	Information Workshop classroom (refreshments)	Validated Participation classroom (refreshments)
0 – 30 min	Enumeration & give token	Info Video	Visual Evidence Info Video + discussion Resource Vote
30 – 60 min	Meeting with Classroom Teacher	Enumeration & give token	
60 – 90 min	Office Hours with HT	Meeting with Classroom Teacher	
90 – 120 min		Office Hours with HT	Enumeration & give token
120 – 150 min			Meeting with Classroom Teacher
150 – 180 min			Office Hours with HT
180 min (3 hours)	Make sure all placards are collected and organized for next session.		

Roles:

VP Facilitator	Logistical Assistant	Enumeration Team
----------------	----------------------	------------------

Figure S8: Kilosa meeting session schedule overview.

The field team consisted of all Tanzanian staff members, two supervisors, a data processing officer, 10 enumerators, and two facilitators – one male and one female. The two facilitators were rotated across sessions as the VP facilitator and the Logistical Assistant to ensure against detecting merely a “facilitator effect.” Figure S8 shows how the staffing was purposely staggered throughout the session, which was designed so parents would not spend too much time waiting.

In practice, one limitation of our design was that the VP group experienced a longer session than the other groups. We contemplated giving the SO group a placebo video to control for the session time, however, given the added logistical burdens and staffing costs that would have implied, as well as the challenge of identifying a placebo that would clearly have no effects but also would be appropriate in the context, we opted against this. Our choice likely biased against finding a treatment effect because if anything, treated parents were more likely to be fatigued by the end of the session.

At the end of the session, each parent was given 3,000 TZS (equivalent to about 1.5USD) as a small token of appreciation and to offset any transportation costs. We intentionally did not include any behavioral measures asking parents to ‘pay’ or ‘contribute’ to anything, which in the past we found to be a noisy measure, and determined that it was unfair to parents to be promised a token only to give them potentially socially desirable opportunities to spend it during the course of the study.

We note a few important differences between our prior Phase 1 study and this “single dose” study. These revisions aimed to bring greater efficiency to the study, to test whether similar effects might be found with less effort (implying less burdensome policy prescriptions should the findings be conclusive and positive), and to maintain a more controlled (but still naturalistic) setting.

1. In our prior study, the “dose” of treatment was delivered over 4 meetings; here we studied our intervention as a “single dose” in one day.
2. Rather than randomization and treatment status occurring at the school level, in this study we randomized at the individual-level within schools to gain efficiency, and all measurements were taken in a single day, immediately following the delivery of the interventions, thus decreasing the possibility of spillovers within schools across treatment arms.
3. We did not have an exclusion lottery.
4. We did not limit to only Standard 3 parents, instead inviting parents from standards 2–5, depending on the number of parents in each standard. We aimed for about 45-60 parents

in each study session, meaning 90-120 parents per school since there will be two sessions in each school.

5. For the information portion of the Validated Participation and Information Workshop groups, instead of facilitators giving informational lectures, we showed an informational video.

Since the intervention and measurement occurred all in one day for this study, we did not conduct a baseline and endline. While such a move requires that we estimate effects between-subjects rather than within-subjects – generally thought to be a less preferable strategy – we believe that this approach likely offered some benefits because questions asked on a baseline survey can affect how respondents answer questions on the endline survey especially if the timeframe between the two is short.

We also did not measure student learning outcomes in this study. While student learning is ultimately a critical outcome, it is not the focus of this study. Our main outcomes of interest in this study were beliefs of self-efficacy and pro-development citizen behaviors, namely parent participation in education. Furthermore, it would not make sense to expect and to measure student outcomes the day of the meeting, since it is downstream from parent efficacy and behavior change. But since this research design relies on randomizing between parents *within* schools, if we measured student outcomes later, we also could not prevent spillover effects between parents of the same school. Instead, we focused on measuring self-efficacy beliefs, attitudes about the importance of parental participation in education, reflections on the extent of respectful or dignified treatment received, and behavioral measures of participation.

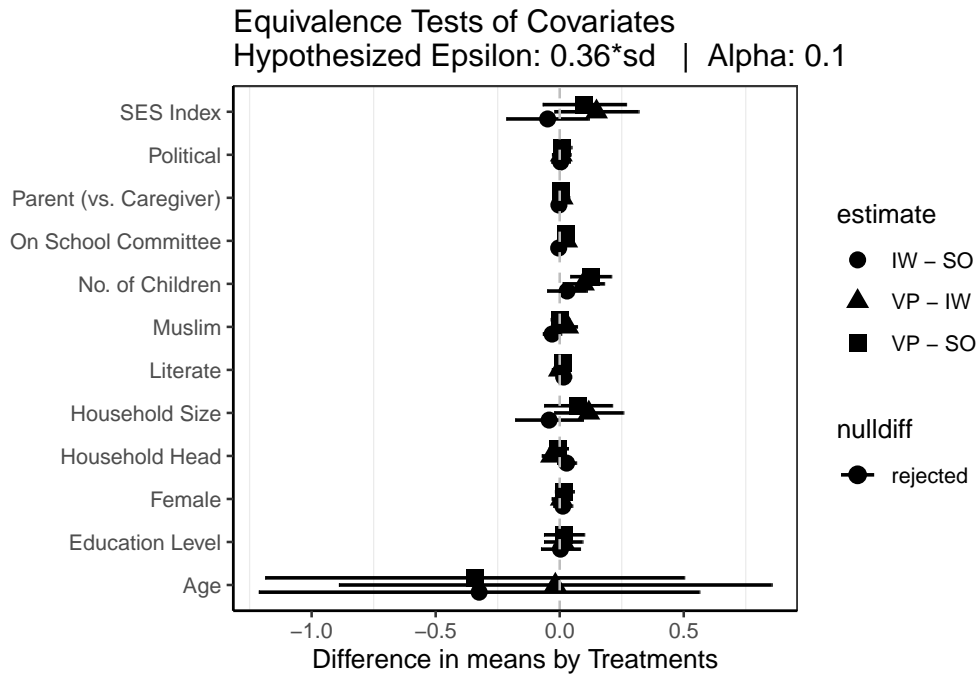


Figure S9: This figure shows the equivalence tests for the treatments across demographic covariates for the Kilosa study. For all covariates, the null hypothesis of statistical difference was rejected.

Figure S9 shows that the null hypothesis of statistical difference was rejected for all covariates. Given the larger sample size for this study and that randomization occurred at the parent level, we expect the covariates to be balanced across treatment groups. Nevertheless, we adjust for all covariates in the analyses.

6 Risk and Ethical Considerations

The central motivation for this study was to better understand how to elicit a set of outcomes we believe to be intrinsically beneficial for our research subjects and for society more generally: more pro-development attitudes (perceptions of efficacy) and behaviors among citizens who are frequently excluded from democratic practice; and to increase engagement, oversight, and quality of education for primary school children. That said, not only do many well-intentioned projects fail to do good, they may do harm. It is important for researchers to consider this possibility.

Although we assessed the likely risk of harm to humans as being very low in this study, we believe that social scientists may be under-emphasizing some of the costs and potential adverse effects caused by social research. In this section, we detail a few such concerns anticipated prior to conducting the research.

Since this intervention did not seek to make structural changes within the education system - such as bringing in new administrators or challenging the school budgeting procedures already in place - our post-treatment parents likely encountered the same systemic barriers to change that they did before the onset of the study. A concern with the proposed experiment was that when facing these barriers, post-treatment parents may feel even lower feelings of efficacy. Giving information and simulating opportunities to engage may only cause increased awareness of bureaucratic and political unresponsiveness.

To be clear, our intervention protocol never delivered a message to parents that they *will* be successful. *Validated Participation* activities simply provided opportunities for them to feel efficacious by showing them that they are *capable* of making inferences and decisions for their children's education. We neither "rigged" the votes in Resource Vote nor engaged in any deception. Thus, while it is possible our participants may feel discouraged if and when the actual system fails them, this is no different from any other project in which the failure to meet raised expectations might generate some frustration. For example, this concern also brings to mind *informed disenchantment* from Gallagher (2006), in which Chinese plaintiffs felt both increased levels of competency and disappointment in the system's fairness upon actively engaging with the legal system.¹¹ Nevertheless, rather than becoming despondent, these citizens channeled their disenchantment into more critical and informed action by educating their friends and family members and finding creative

¹¹Similarly, Soboleva (2019) finds in Eastern Ukraine that cognitive therapy and civic education improved efficacy but also suppressed civic engagement.

ways to mobilize. We hope that the subjects in our study would react similarly. We view this risk as minimal.

Having identified these possible areas of concern, we conducted several pilot studies near Dar es Salaam and in Bukoba to monitor for potential harm. Additionally, after this study, we looked for any evidence of possible to harm to human subjects through analysis of both our quantitative data (i.e., through discovery of adverse effects of treatment) and through qualitative research conducted after the completion of the endline survey of adult participants. We found no such adverse effects.

7 Deviations from Pre-Analysis Plans

Plan (page)	Study	Deviation	Reason
April 26, 2016, Phase 1 PAP, p16	Study 1 Bukoba	We do not analyze the Family Tree project as a measure of parental participation or for spillovers.	While we included instructions for this assignment, teachers offered different instructions across the study schools for how to complete this project and how long students had to turn it in. Thus, we were not able to attribute differences across the schools to our treatments, and so we were not able to use this measure.
April 26, 2016, Phase 1 PAP, p16	Study 1 Bukoba	We do not analyze attendance records as a measure of child learning habits or for spillovers.	We could not attain sufficient records to complete this analysis.
August 1, 2018 Phase 1 PAP update, p21-23	Study 1 Bukoba	We noted support for key hypotheses based on Phase 1 study, but report these in final article as null results.	In our PAP, we specified in Eq. 1 a model in which we would estimate models with change scores $\Delta y_{i,j}$ as our main outcome. But we learned that best practice is to control for pre-treatment outcomes for more precise estimates. ¹² Rerunning our models with this alternate specification has changed our results from when we wrote this PAP.
August 1, 2018 Phase 1 PAP update, p24	Bukoba Qualitative Follow-up	We said we would code all variables on a 0-1 scale, but we coded on a 3-point scale for frequency and change outcomes.	We assumed that the responses would be Yes/No re:parental involvement and None/Positive re:change. However, we found that numerous responses made distinctions between some or multiple instances of parental involvement (frequency), and the situation getting worse relative to baseline (change), necessitating a separate coding category. However, we also analyze the data with binary outcomes and find no substantive difference in the results. See SI Section 10.
June 16, 2018, Phase 2 PAP, p54	Study 2 Kilosa	We do not analyze as outcomes meetings with classroom teachers and office hours with head teacher.	Since parents in the VP treatment arm had the opportunity to talk and engage a teacher-like authority during the intervention, they said they felt no need to speak further with school officials immediately after the long intervention. This was a design flaw on our part, and rendered the test invalid.

Table S3: This table logs deviations from our pre-analysis plans.

8 Association between Active Citizenship and Educational versus General Self-Efficacy

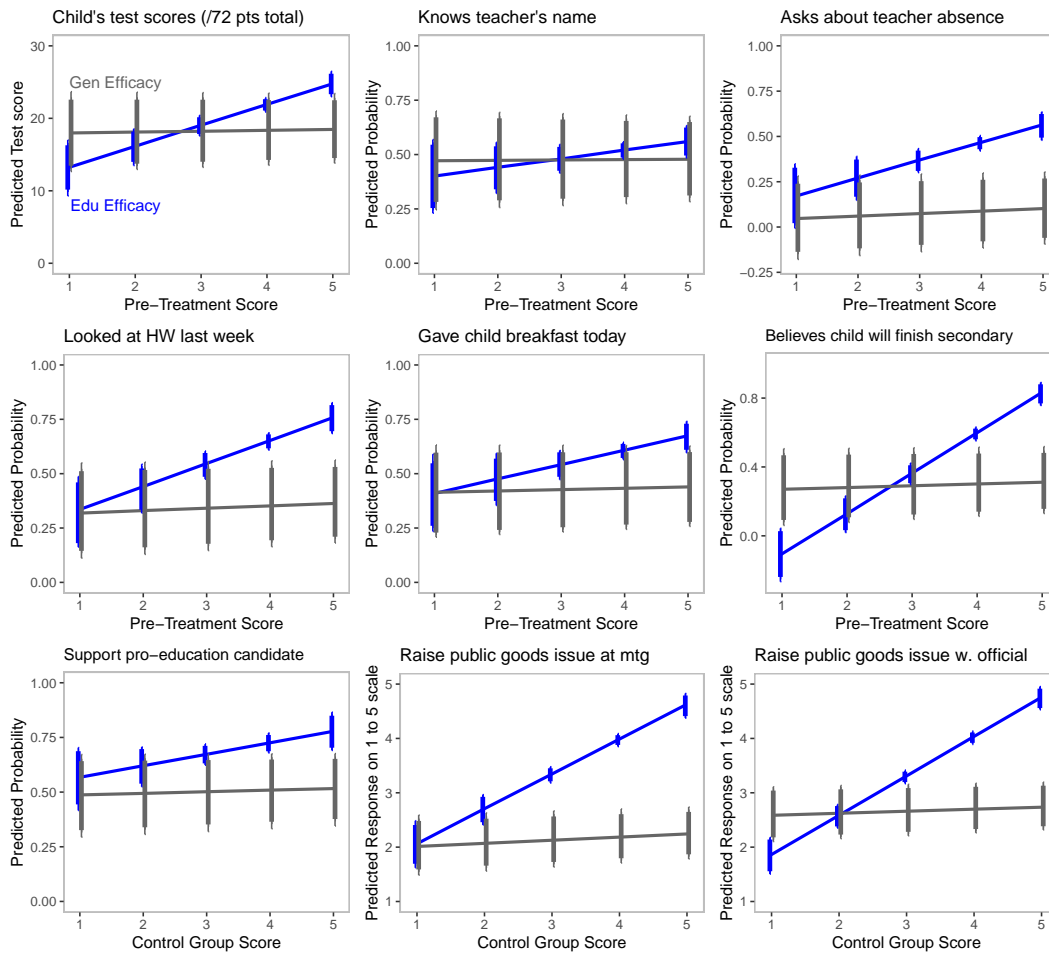


Figure S10: Positive linear relationships between our educational self-efficacy score (x-axis, blue) and pro-development outcomes (y-axis) from both studies, compared with the general self-efficacy score (x-axis, gray). From Study 1 (rows 1-2), these plots show the predicted probabilities for children’s test scores and parental attitudinal and behavioral outcomes at baseline. From Study 2 (row 3), these plots show the predicted probabilities for pro-education political behaviors for the SO (control) parents (with 95 % and 90% CIs). OLS models, controlling for all demographic covariates, with school cluster robust standard errors.

Figure S10 compares the predictive power of our educational self-efficacy score with the general self-efficacy score from Jerusalem and Schwarzer (1995). After controlling for other socio-demographic covariates, the associations between pro-development outcomes and the general self-efficacy score shown in gray are much weaker compared to those with our educational self-efficacy score shown in blue.

9 Bukoba School-Level Interviews and Outcomes

In addition to our quantitative analyses, our qualitative research captured important post-treatment dynamics, indicating some potentially important differences in outcomes between treatment arms in the manner that we predicted theoretically (i.e. increased pro-development behaviors with boosts in self-efficacy under VP).¹³ Specifically, we collected interview data 6 months and 2 years after the experimental intervention in Bukoba for Study 1. We discuss both types of data in more detail below.

We hired two research assistants to follow up with parents, classroom teachers and head teachers approximately 6 months after the conclusion of the endline survey. We planned for this research to cross-validate our survey findings, to capture any long term effects, and to investigate any unexpected or unanticipated behaviors that were not captured on the survey. More than anything, we set out to identify whether in slightly more conversational and reflective settings, individuals might provide clues or trace evidence about how they were affected by the treatment and as such, we oversampled on communities from our main treatment arm. Specifically, we identified the top 4 and bottom 2 school communities in terms of measured average gains in ESES from within the VP treatment group; the top 2 and bottom 2 from the IW group; and top 1 and bottom 1 from the SO group. Within each of these school communities, our research assistants interviewed the head teacher and one classroom teacher (total of 24 teachers) and two households among the study participants. We first selected the top 5 households in terms of measured educational efficacy gains from the VP and IW groups, and the best matching households from the SO group based on wealth, household size, religion, and number of children. We then randomly ordered these 5 households, and our research assistants interviewed the study parent of 2 of these households, going down the list of who was available (total of 24 parents).

First, we summarize the key observations made by the facilitators during the IW and VP meetings, respectively, in Table S4. We cannot include results from the SO group, as that arm did not include a meeting-based intervention. Of particular interest, during the IW workshops, the facilitators noted that parents were curious about what the sponsoring NGO, Twaweza, would do to help solve school problems. It is important to note that we had not discussed this possible outcome with our facilitators, nor was it part of the meeting scripts. By contrast, in the VP meet-

¹³Reporting of qualitative data and analysis collected associated with RCT's is a disappointingly rare practice. One important exception is Rao, Ananthpur and Malik (2017), which demonstrates the key role qualitative data analysis can play in the interpretation of study findings.

ings, facilitators noted that parents expressed interest in volunteering and forming peer networks. Unprompted, these parents coordinated to donate money to buy materials for their respective schools.

In other words, VP parents indicated an inclination to take responsibility for influencing the local governance of education on their own, whereas IW parents were still looking to external authorities to “do the work.” This is despite the fact that in both sets of interventions, parents had been delivered the same message – that they have a role to play in their children’s education.

Second, for the follow up research conducted by our research assistants, we provided structured interview guides for interviewing parents, classroom teachers, and head teachers. We designed the qualitative research to focus on evaluating mechanisms with respect to very positive (and null) cases of efficacy change in order to detect behavioral effects. Our goal here was to try to identify evidence of actual post-treatment change measured through narrative accounts. The imbalance in our sampling strategy reflected a desire to uncover any evidence of change in the VP arm in particular – because if we could not find any, that would indicate that we should not reject the null hypothesis of no effect.

All interviews were transcribed. Subsequently, while blind to treatment assignment, we coded the responses to all questions with respect to whether they provided a clear answer to questions about observed changes in relevant pro-development behaviors; and if so, whether there was positive change (i.e., more pro-education actions by parents observed in the past few months), no change, or negative change. Because we did not find any responses that indicated negative change, essentially, all responses were coded in a binary manner with respect to the presence or absence of positive change. Examples of statements by teachers and parents and how we coded them are provided in table S5.

Perhaps not surprising, given our sampling strategy, parents across treatment arms overwhelmingly made comments suggesting positive change in pro-development behaviors (80.8%, 88.9%, 83.3% of comments were positive for VP, IW and SO parents respectively).¹⁴ As it was beyond our capacity for this study to have our research assistants actually embed themselves in the respective communities, to observe parental behaviors on their own, we were still forced to rely upon self-reporting, and again, in this context, parents (with self-reported positive change in self-efficacy) almost universally reported positive change in pro-development. Thus, we did not observe

¹⁴The p-values from simple two-sided T-tests comparing the proportion of positive VP and IW parent comments and the proportion of positive VP and SO parent comments are 0.46 and 0.89, respectively.

Information Workshop

- Parents seemed to think that Twaweza would be involved with teacher training after the study was over.
- Parents asked if Twaweza would buy books for the school teaching parents how to assist their children in their schoolwork at home.
- Parents asked how Twaweza would help solve the school's problems.

Validated Participation

- One parent volunteered to donate 2 footballs to the school. After the school tour, Parents collectively decided to donate 500 shillings each to buy tree seedlings for the school. Parents set up a network for themselves, with leadership to follow up on the promises that they had made during the meetings.
- Over the course of the meetings, parents seemed to interact and listen to each other more.
- Parents decided to initiate their own network with leadership roles regarding parental engagement for their children.
- The parents organized into a network, and all agreed to donate 1000 shillings to renovate the boys' toilets.

Table S4: This table summarizes the qualitative evidence of facilitator notes taken during the IW and VP intervention meetings

<u>Coded as:</u>	Classroom and Head Teachers	Parents and Caregivers
No Change	<p>“I have been teaching this class for a month now but throughout this period I have not have any interactions with parents.” – Classroom Teacher</p>	<p>Do you feel you can help your child with his schoolwork? “No, maybe his brother can help him. He has time for studying with his brothers,when I am busy in the kitchen..” – Parent</p>
Positive Change	<p>“Yes. There was a public meeting around here, in which the District Executive Director (DED) was asked very sensible question by some parents on the school issues; the parents asked the DED how comes that the school is facing a shortage of teachers ... How can the school performance be improved without enough teachers? I find this as very new experience! This had never happened before.” – Classroom Teacher</p>	<p>“What moved me most is to realize that even the government wants to see us, common people working with the community in addressing the challenges of educating our children. I have also realized that the teachers find us making follow-ups to the education of our children and so they happen to be more responsible in their dealing with our children. There was nothing that I did not like in the study.” – Parent</p>

Table S5: This table shows examples of coded responses to structured questionnaire, blinded to treatment arm.

differences across treatment arms.

Yet, comments from the classroom and head teachers were much more revealing. Teachers were not only unaware that the study involved multiple treatment arms across schools, but they also did not know the full content of the interventions at their own schools. Additionally, they were asked about parents' behaviors not their own or their fellow teachers. For these reasons, we believe they were less susceptible to social desirability bias, especially any that would privilege one treatment arm over the others. 81.8% of teacher comments indicated positive change in the VP communities, but for IW and SO communities, the proportion of positive teacher comments were much less at 53.6% and 11.1%, respectively.¹⁵¹⁶

But more telling than simply the quantity of the comments from the respective arms is what was actually said (all comments are reported in SI section 9). In the VP schools, teachers reported on quite meaningful engagement and interest, as well as visible active behaviors on the part of parents.

The comparison with the IW responses is striking, because the number and duration of the meetings was extremely similar to the VP schools, and it seems implausible that social desirability bias would have varied across treatment arms. And yet, the nature of the responses from the IW schools portrays decidedly less engaged parents, with few concrete examples of any change since the time of the end of the intervention. We think it is reasonable to conclude that the Validated Participation intervention may well have provided the efficacy boost needed to encourage parents to participate in a more meaningful and sustained manner as compared with simply receiving information on its own.

Thus, we learn several lessons from this research. First, the validity of parent self-reports may be adversely affected by social desirability bias. Second, classroom teachers and head teachers seem less susceptible to these biases and appeared more willing to provide a range of (potentially more accurate) responses about parents' pro-development behaviors. Third, based on those accounts, it appears quite plausible that parents from school communities treated with VP emerge as more

¹⁵The p-values from simple two-sided T-tests comparing the proportion of positive VP and IW teacher comments and the proportion of positive VP and SO teacher comments are 0.02 and 10^{-4} , respectively.

¹⁶Section 9 in the SI provides additional details on the distribution of number of comments made by parents and teachers across treatment arms. Since we are simply reporting the proportion of positive comments, we may be concerned that there are a few individuals making many comments and that those who give more comments are also more likely to make positive comments. Figures S11 and S12 show that the distributions of number of comments made for VP and IW parents (ranging from 1 to 3) and teachers (ranging from 2 to 7) are comparable. And figures S13 and S14 show that for both parents and teachers, the positive comments were not concentrated in those who gave the most comments.

active citizens than those who received IW or SO.

Figures S11 and S12 show the distribution of comments made (x-axis) across treatment arms for parents and teachers, respectively. While SO parents and teachers made the least number of comments, the distributions for VP and IW parents and teachers are comparable.

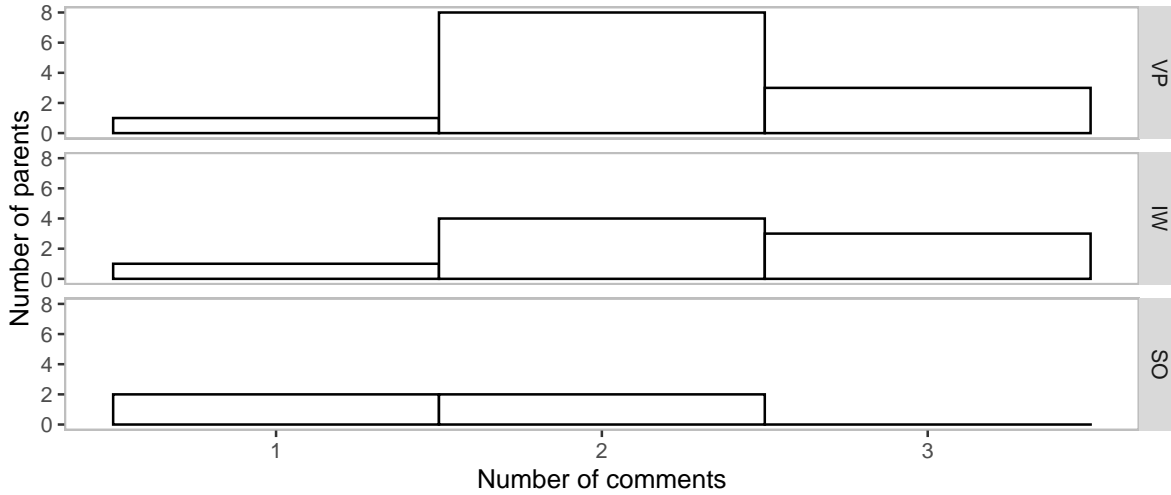


Figure S11: This figure shows the distribution of comments made by parents across treatment arms.

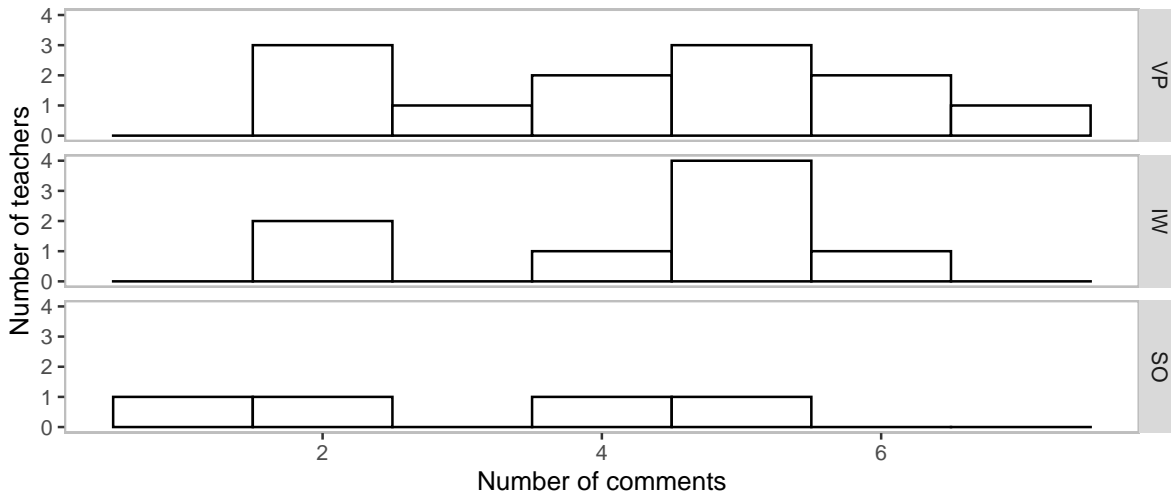


Figure S12: This figure shows the distribution of comments made by teachers across treatment arms.

Figures S13 and S14 show the number of comments and whether they indicate positive change (blue) or no change (gray) for each parent and teacher (x-axis), sorted from least to most comments, across treatment arms. These figures confirm that parents generally had positive comments, but

across both parents and teachers, the positive comments were not concentrated in those who gave the most comments.

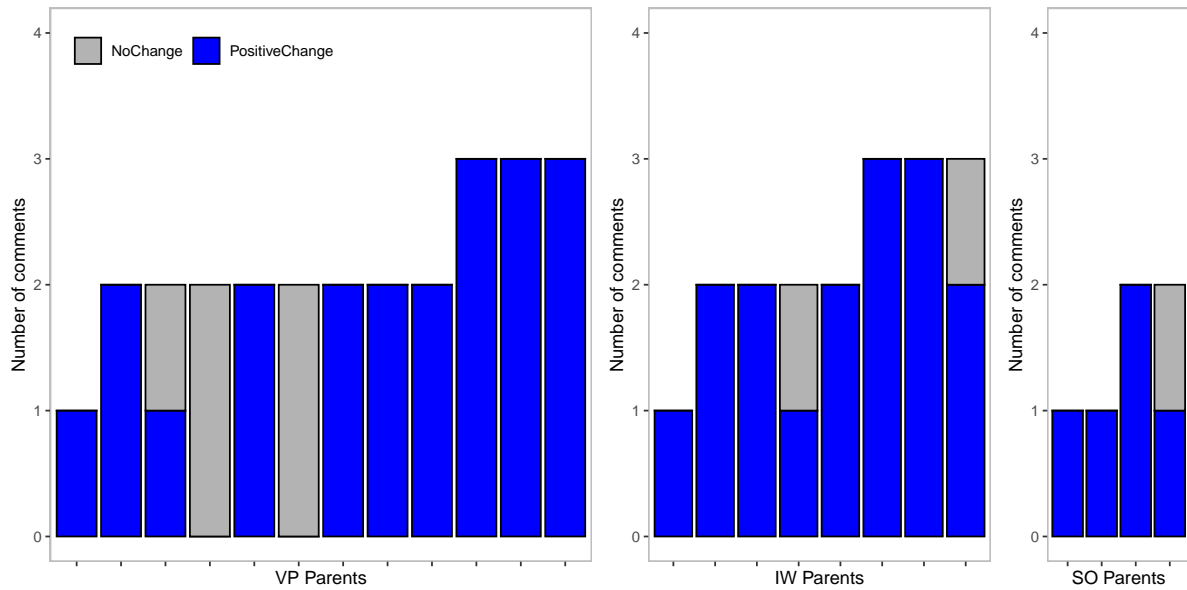


Figure S13: This figure shows the number of comments (positive or not) made by each parent, sorted from least to most comments, across treatment arms.

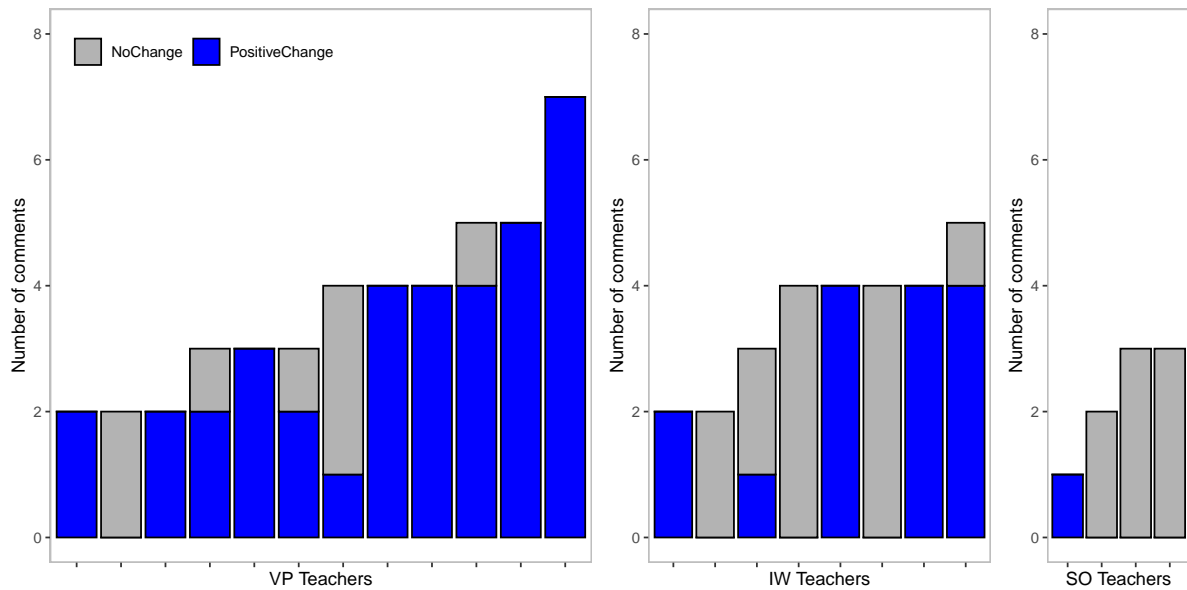


Figure S14: This figure shows the number of comments (positive or not) made by each teacher, sorted from least to most comments, across treatment arms.

Finally, below are examples of responses by teachers across the three treatment arms:

Responses from Validated Participation Schools

1. They came to school to see the teacher. This was observed to both parents of standard three and other classes Moreover, now we have phone numbers of parents which have enhanced communication between teachers and parents Qn. Is the rate of communication different or similar to how it was done before the recent meetings? There are differences, now communication and cooperation between parents and teachers has increased.
2. During our times when we were studying our parents were checking exercise books, even if they didn't know to read and write they still wanted to know if we are writing or doing something at school. Parents of nowadays they just want to use their children in doing farm and other domestic activities. Like here in Fridays, parents can just tell their children to go to the street market and miss studies. You know that at least now the environment is very supportive, there is solar electricity but the morale for education is very low.
3. Sometimes they do assist; when we provide homework to pupils, we sometimes ask children to ask for help from parents; some pupils report to have been assisted by their parents but some do not.
4. For example, when you left parents came to read the poster and now I can see some children have exercise books for their home exercises. They have more books than previous time. About whether they provide time and space I am not sure, I did not make follow-up of what is going on at home. In most cases we encourage parents to come at school and do their responsibilities. Parents are involved in different village meetings where the ask leaders about school developments.
5. We had a meeting with parents, about forty parents turned up. In that meeting we discussed issues of truancy, cleanliness and children's food. From that day some parents are implementing what we agreed like now students are clean, they have uniforms and shoes. Some parents give their children food but others don't. Another problem was the time children were coming to school, at least now they come a bit early, particularly those children who are from (redacted).
6. Yes. Some of them. There are some parents who came to ask for their children's progress. Some others came to report their children that they were sick.
7. We have been communicating with parents since you left. First of all those parents who

participated in the study continued to come and ask how their children are progressing academically. For example when they saw that their children's exercise books did not contain dates or is not marked, they came to ask to the respective subject teacher

8. There is one neighbor who also participated; he told me that if this assessment could be done frequently it could raise education morale. I see the changes to his child; at least he knows what is going on. In general parents were happy and I think that facilitation you did made them enjoy the training. You know when you see people are talking about it know that they are happy! I suggest that if there is another possibility you can do it in all schools and may be you bring together parents and teachers.
9. When you see their exercises done in the classroom, they are doing it well. So I hope they will do better than the previous class
10. They will do better because their parents participate more in the education of their children after the series of meetings held by Twaweza.
11. This class has thirty three students and all can read and write. I'm expecting that they are going to do well though they might have little weaknesses. Qn. Can you explain those weaknesses? I mean you cannot say for sure everything will be alright but these are human beings they may fall sick during examination
12. Yes! They were engaged but very minimally, you know to change some ones behavior is not easy. Since you came now when you call them or ask them to provide something to their children they respond and do it
13. There was some parental engagement but it was very small compared to the situation after the study began; prior to the study, parents would only come if called but now some parents come on their own and ask teachers about their children's progress. This parental engagement has not reached 100%.
14. Before you came, we were observing and asking ourselves why children are not attending school? When we compare to this time after your meetings, I think parents were lacking some knowledge about why their children should come to school. I think you encouraged them, parents are now very close to school and they continue coming. For example, some

of them have volunteered to bring some glasses and doors for construction of pre-primary children classroom.

15. First of all other teachers their views was that the meetings could involve all teachers. Those scores for students, some teachers were not happy. Now they are putting more efforts so as to improve their performances.
16. I think their participation in school will result in good results to the great extent. Possible results would include reducing absenteeism, and the performance in academics will improve.
17. It will be better; the teachers and the parents are now working as a team. We only need the government to furnish the school with enough teachers and the rest will be very okay.
18. As I said it was on 27th September, 2016. The parents spoke a lot and they encouraged each one to put into practice the agreed five objectives.
19. No. Everything has been said already.
20. I remember after earthquake, some parents contributed tree poles and money which was sent to the central pool.
21. There was a public meeting around here in which the District Executive Director (DED) was asked very sensible question by some parents on the school issues; the parents asked the DED how comes that the school is facing a shortage of teachers whereas you have this report on your table from the head teacher? How can the school performance be improved without enough teachers? I find this as very new experience! This had never happened before.
22. No. There have never been any interactions. If a child has a problem it is the responsibility of the class teacher to call the parent and discuss the problem. Problems that we face include students' learning materials; if the class teacher notices any problem, he informs me and in turn I give him permission to call and discuss the matter with the parent; but since November there has not been such requests.
23. Parents brought a letter asking for a meeting with head teacher, academic teacher and class three teacher. Unfortunately, the head teacher had an emergency so I represented him on that meeting. Parents also comes to see classroom teachers and there is good communication.

24. Yes, on 27th September, 2016 after this study, we had a parents' meeting intended to include parents of all the children of this school. This meeting was initiated by the standard three parents because after the study the grade 3 parents set five objectives to implement. We called this meeting share these objectives with other parents. The attendance was comparatively very good since 328 out of 528 parents attended the meeting. This best attendance ever recorded was directly connected with influence of the study. The five objectives were put forward by the parents were discussed in his meeting.
25. There is no cooperation. You call the parents for the meeting but only 40 attend while they were expected to be 90 through 100. The number remains the same; it does not increase.
26. Prior to the study parents would only come if you force the students to go home and come with their parents but now just announce the meeting, you will be surprised, the attendance will be supper.

Responses from Information Workshop Schools

1. I have not seen any changes, may be they got that seminar and went with it home. If it is exercise books, children absenteeism is just the same! Considering now the situation is more difficult, people are focusing on what they will eat. May be only those who are supported by TASAF, you can see them with all school requirements. May be if they could be seeing us who went to school have better life, now here people who are doing business have better life than us, now the parent asks I have to encourage my child to study so what?
2. Homework! You are wasting your time. One day I asked if there is anyone who can provide a lamp for his child to study. They said no. but there are very few parents now, who have bought solar, and those children are clearly seen. You cannot plant a farm without weeding then expecting to harvest Most of these children do not have time to study at home, we are seeing their performances. May be their parents can buy exercise books, pen and school uniforms. Sometimes even these, they do after being forced to do so, when you look at students' exercise books you will be surprised the way they are worn out
3. I think I should say no. this is because I have not seen any parent coming to make follow-up of their children's academic progress may be if there is meeting. Most of these parents are not coming

4. In this year, I think they will do much better than last year. We are working hard to ensure that they do better, teachers are more committed! I think they will do better.
5. According to what I can see, the way they were doing before is the same to what is done now. Although I don't know what you told them, I don't see any changes
6. I had a meeting with parents in September. The meeting aimed at electing members for the school committee. The new directives require the chairperson and the deputy chairperson of the committee to have attained at least a Form Four level of education. Some other issues discussed in the meeting were the problem of students' absenteeism. There were some hot issues raised but they were connected to the previous leadership; for example parents questioned the school leadership on the contributions made for the toilets that were put under the control of the village leaders.
7. Regarding the parents' participation, I will not say much because things that make them participate is beyond their capacity; for example, the distance from home to school, and school lunch. We will keep on encouraging them to participate in our meetings.
8. Some specific actions that were observed included some parents who came to school to inquire from teachers on how to fill in the family tree. We have not observed specific action made by parents as a group. There are individual parents that come to school to report if their children will not attend school.
9. Yes. There is communication because parents thought that the exercise is sustainable and so they were asking the possibility for them to share their learning with the rest of the parents. The challenge was they were few.
10. Prior to the start of the study parents lacked awareness of their roles in education but now there are some observable changes; for example, to some extent parents show some awareness such that in the past parents would hardly come to school even if called but now they come sometimes on their own. Prior to the study, the interaction between parents and teachers was very little but now we see a difference; in the past a parent came to school only if they were being called but now they come freely even when they are not called.

Responses from Survey Only/Control Schools

1. Very small percent; we usually instruct children to tell their parents to check their exercise books but when we ask children whether the parents did as asked, their response becomes NO. In addition, when we instruct children to ask their parents to buy them exercise books, it takes a long time before we find children having the appropriate exercise books. This gives us the impressions that parents do not participate in the education of their children. In fact, I teach English and the majority of parents do not understand English, so they can do nothing to assist their children. What I know is that parents don't give their children time for studying instead they assign them other domestic roles like fetching water and firewood. Even when a child struggles on his own to find time for studying in the evening, some parents may be reluctant because they feel that the child is increasing the kerosene budget; this is partly contributed by the economic hardships affecting most rural families.
2. I have been teaching this class for a month now but throughout this period I have not have any interactions with parents.
3. As I can see, 60% of the students will perform well based on their day to day assessments. I only speak for the grade 3 students of this year because I did not teach the last year's grade 3 students.
4. Parents' participation is very small; when children leave home to school, parents feel less concerned about the child. But if a child has a problem, such as absenteeism, teachers might call the parent, and together they can discuss the problem; the parent will implement the set strategy for the solution.
5. After your study some parents came, they expected to get results of the examination you gave to their children. Qn. How many parents came to ask for results? Almost ten parents showed up but we couldn't give those results as up to now the school has not received those results from Twaweza. Any other interactions apart from results? In most cases these parents are coming to make follow-up of how their children are progressing or else they came when we call them if there are problems.

10 Binary Coding of Study 1 Follow-up with Teachers on School-Level Outcomes

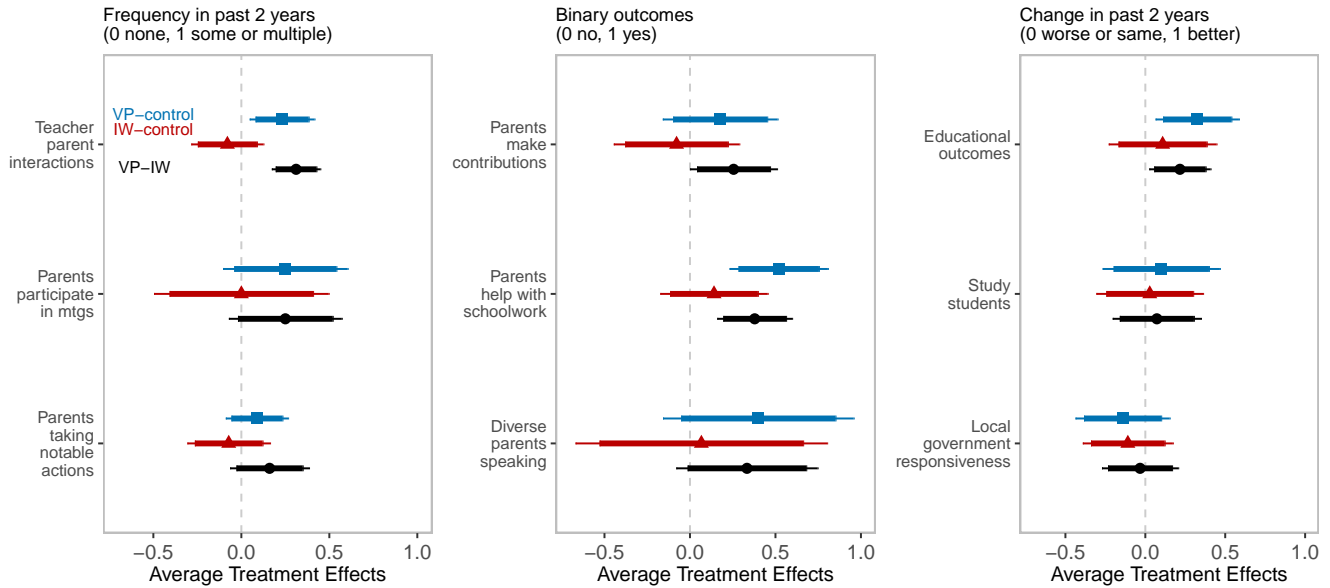


Figure S15: Downstream Follow-Up in Study 1 Schools, two years later (N=117 teachers; 24 school clusters): Average treatment effects of VP (blue) and IW (red) compared to control on parental active citizenship outcomes 2 years later as reported on structured assessments with teachers (with 95% and 90% CIs). OLS models with respondent type FE and school cluster robust standard errors. VP-IW difference (black).

For the Study 1 Follow-up with Teachers in Bukoba, we coded the outcomes about frequency (0 none, 1 some, 2 multiple) and change (-1 worse, 0 same, 1 better) on three-point scales. As a robustness check, we code these outcomes on a binary scale. For frequency outcomes, 0 none and 1 some or multiple. For change outcomes, 0 worse or same and 1 better. As Figure S15 shows, results are not substantively from Figure 5 in the manuscript.

11 Regression Adjustment using Lin’s Regression

As a robustness test for our experimental results, we use the OLS interacted adjustment method proposed by Lin (2013). We do this to address concerns about how our main regression specifications may be biased due to the baseline imbalance of efficacy scores in Study 1 and if there are heterogeneous treatment effects that are strongly related to covariates. Specifically, the adjusted specifications shown in this section include the interactions between demeaned covariates and the treatment indicators in order to improve precision.

11.1 Study 1: Effects on Parent and Student Outcomes in Bukoba Pilot

For Study 1, we use OLS, regressing endline outcomes (Y_{i2}) on Validated Participation (VP) or Information Workshop (IW) treatment indicators, controlling for de-meaned baseline outcome measures ($\tilde{Y}_{i1} = Y_{i1} - \bar{Y}_1$) as well as de-meaned demographic covariates ($\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$) along with their interaction with the treatment indicators, with cluster-robust standard errors at the school level:

$$Y_{i2} = \alpha + \beta_1 VP_i + \beta_2 IW_i + \beta_3 \tilde{Y}_{i1} + \beta_4 \tilde{\mathbf{x}}_i + \beta_5 VP_i \times \tilde{Y}_{i1} + \beta_6 VP_i \times \tilde{\mathbf{x}}_i + \beta_7 IW_i \times \tilde{Y}_{i1} + \beta_8 IW_i \times \tilde{\mathbf{x}}_i + \epsilon_i \quad (S1)$$

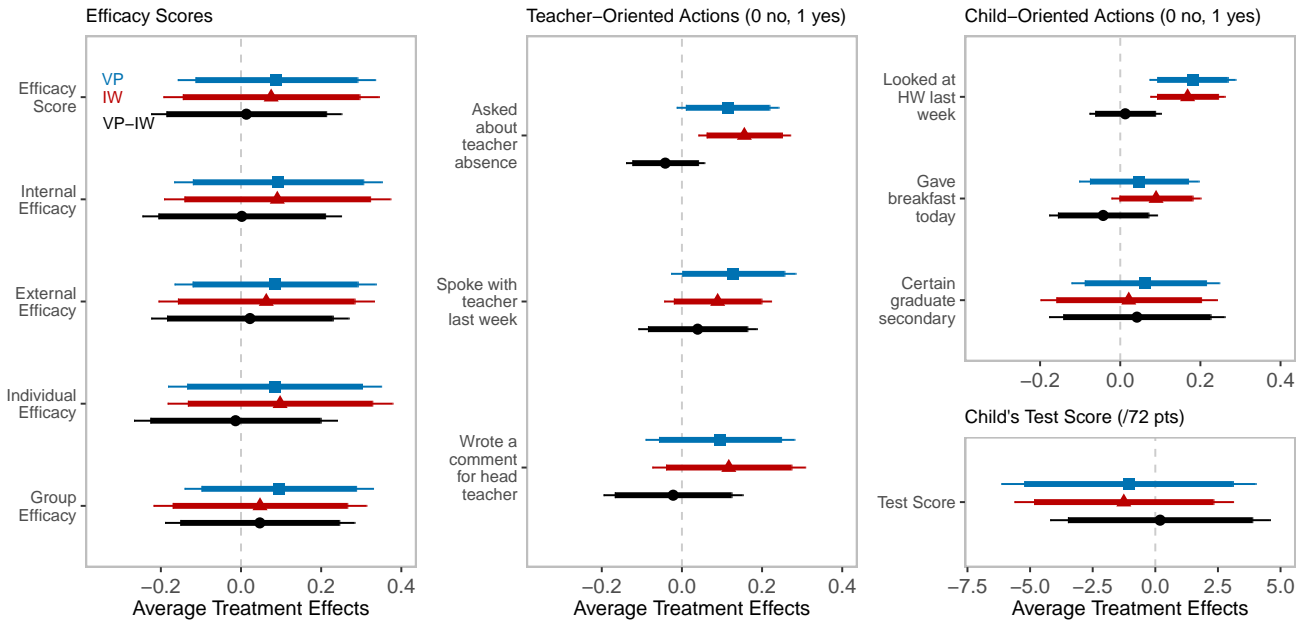


Figure S16: Study 1 (N=598 parents; 24 school clusters): Average treatment effects of VP (blue) and IW (red) on endline efficacy and active citizenship outcomes (with 95 % and 90% CIs). OLS models using treatment-by-covariate interactions. VP–IW difference (black).

Figure S16 is not substantively different from Figure 4 in the manuscript.

We do not perform the same test for the results from Study 1 Follow-up with Teachers on School-Level Outcomes, because we do not perform covariate adjustment for those models.

11.2 Study 2: Effects on Parent Outcomes in Kilosa

For Study 2, as the design of Study 2 was cross-sectional, we use OLS regressing outcomes (Y_i) on Validated Participation (VP) or Information Workshop (IW) treatment indicators, controlling for de-meaned demographic covariates ($\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$) along with their interaction with the treatment indicators, with school fixed effects:

$$Y_i = \alpha_s + \beta_1 VP_i + \beta_2 IW_i + \beta_3 \tilde{\mathbf{x}}_i + \beta_4 VP_i \times \tilde{\mathbf{x}}_i + \beta_5 IW_i \times \tilde{\mathbf{x}}_i + \epsilon_i \quad (\text{S2})$$

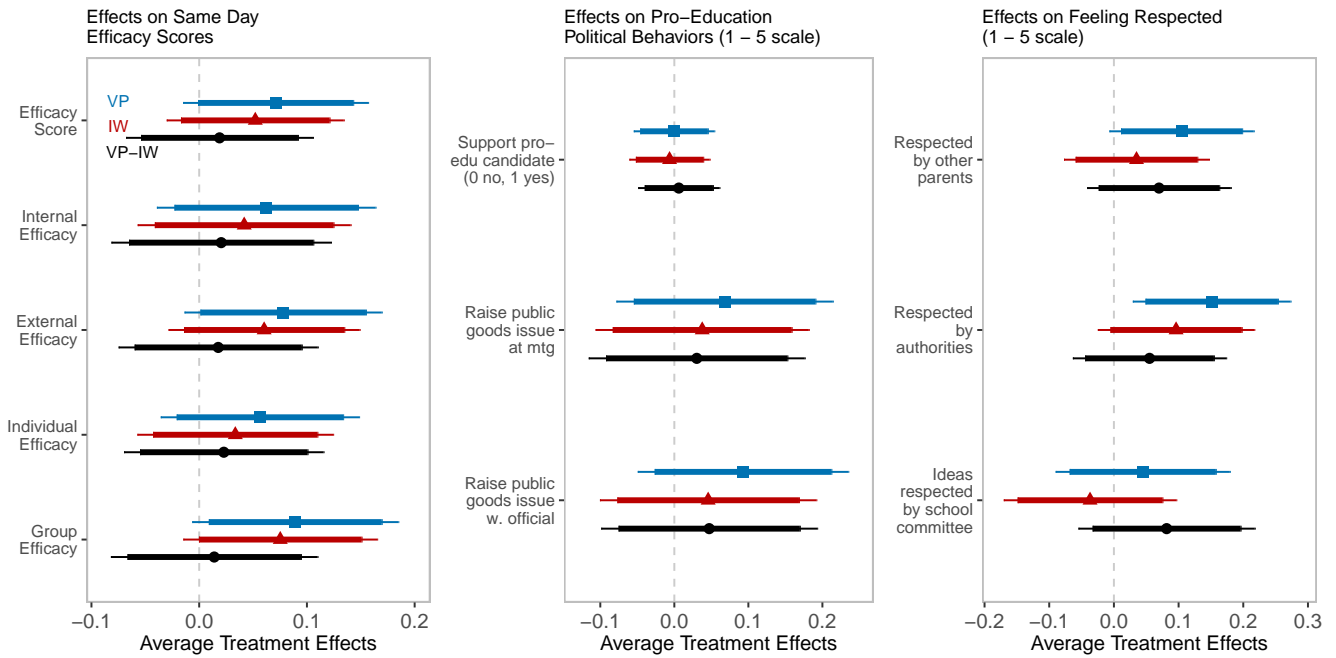


Figure S17: Study 2 (N=1633 parents; individual-level treatment assignment): Average treatment effects of VP (blue) and IW (red) on same-day, post-meeting efficacy, pro-education political behaviors, and feeling respected (with 95% and 90% CIs). OLS models using treatment-by-covariate interactions. VP-IW difference (black).

Figure S17 is not substantively different from Figure 6 in the manuscript.

12 Multiple Hypothesis Testing using Benjamini-Hochberg Procedure

In this section, we address concerns about multiple hypothesis testing by adjusting for the false discovery rate (FDR). We show the Benjamini-Hochberg (BH) adjusted p-values for each test of our findings shown in Figures 6 and 5 in the paper. Since these are two separate studies, we adjust within each study. Furthermore, for the sets of outcomes, we show the adjusted p-values within estimates (VP, IW) across outcomes, and then within outcomes across estimates. With these more stringent adjusted p-values, all the statistically significant effects of VP retain their significance.

Outcome	Treatment	Effect size	SE	p-value	adjusted p-value
Efficacy Score	VP	0.07	0.04	0.05	0.10
Efficacy Score	IW	0.05	0.04	0.12	0.12
Internal Efficacy	VP	0.06	0.05	0.11	0.21
Internal Efficacy	IW	0.04	0.05	0.21	0.21
External Efficacy	VP	0.08	0.05	0.05	0.10
External Efficacy	IW	0.06	0.05	0.10	0.10
Individual Efficacy	VP	0.06	0.05	0.11	0.22
Individual Efficacy	IW	0.03	0.05	0.25	0.25
Group Efficacy	VP	0.09	0.05	0.03	0.06
Group Efficacy	IW	0.07	0.05	0.06	0.06
Support pro-edu candidate (0 no, 1 yes)	VP	0.00	0.03	0.50	0.63
Support pro-edu candidate (0 no, 1 yes)	IW	-0.01	0.03	0.63	0.63
Raise public goods issue at mtg	VP	0.07	0.07	0.17	0.34
Raise public goods issue at mtg	IW	0.03	0.07	0.34	0.34
Raise public goods issue w. official	VP	0.10	0.07	0.09	0.19
Raise public goods issue w. official	IW	0.05	0.07	0.26	0.26
Respected by other parents	VP	0.10	0.06	0.03	0.07
Respected by other parents	IW	0.04	0.06	0.26	0.26
Respected by authorities	VP	0.15	0.06	0.01	0.01
Respected by authorities	IW	0.10	0.06	0.06	0.06
Ideas respected by school committee	VP	0.04	0.07	0.26	0.52
Ideas respected by school committee	IW	-0.03	0.07	0.68	0.68

Table S6: Adjusted p-values using Benjamini-Hochberg for Kilosa study results (Fig. 6 in the paper): adjusting within each outcome, across treatment groups. Under these more stringent adjusted p-values, the efficacy outcomes lose their statistical significance.

Outcome	Treatment	Effect size	SE	p-value	adjusted p-value
Efficacy Score	VP	0.07	0.04	0.05	0.11
Efficacy Score	IW	0.05	0.04	0.12	0.33
Internal Efficacy	VP	0.06	0.05	0.11	0.15
Internal Efficacy	IW	0.04	0.05	0.21	0.36
External Efficacy	VP	0.08	0.05	0.05	0.11
External Efficacy	IW	0.06	0.05	0.10	0.33
Individual Efficacy	VP	0.06	0.05	0.11	0.15
Individual Efficacy	IW	0.03	0.05	0.25	0.36
Group Efficacy	VP	0.09	0.05	0.03	0.11
Group Efficacy	IW	0.07	0.05	0.06	0.33
Support pro-edu candidate (0 no, 1 yes)	VP	0.00	0.03	0.50	0.50
Support pro-edu candidate (0 no, 1 yes)	IW	-0.01	0.03	0.63	0.68
Raise public goods issue at mtg	VP	0.07	0.07	0.17	0.21
Raise public goods issue at mtg	IW	0.03	0.07	0.34	0.41
Raise public goods issue w. official	VP	0.10	0.07	0.09	0.15
Raise public goods issue w. official	IW	0.05	0.07	0.26	0.36
Respected by other parents	VP	0.10	0.06	0.03	0.11
Respected by other parents	IW	0.04	0.06	0.26	0.36
Respected by authorities	VP	0.15	0.06	0.01	0.08
Respected by authorities	IW	0.10	0.06	0.06	0.33
Ideas respected by school committee	VP	0.04	0.07	0.26	0.28
Ideas respected by school committee	IW	-0.03	0.07	0.68	0.68

Table S7: Adjusted p-values using Benjamini-Hochberg for Kilosa study results (Fig. 6 in the paper): adjusting within each treatment group, across outcomes. Under these more stringent adjusted p-values, the efficacy outcomes lose their statistical significance.

Outcome	Treatment	Effect size	SE	p-value	adjusted p-value
Educational outcomes	VP	0.38	0.15	0.01	0.02
Educational outcomes	IW	0.12	0.19	0.27	0.27
Study students	VP	0.10	0.16	0.28	0.56
Study students	IW	-0.03	0.17	0.56	0.56
Local government responsiveness	VP	-0.08	0.19	0.66	0.74
Local government responsiveness	IW	-0.13	0.20	0.74	0.74
Parents taking notable actions	VP	0.21	0.17	0.13	0.26
Parents taking notable actions	IW	-0.18	0.22	0.79	0.79
Teacher parent interactions	VP	0.64	0.13	0.00	0.00
Teacher parent interactions	IW	-0.08	0.17	0.68	0.68
Parents participate in meetings	VP	0.25	0.16	0.08	0.15
Parents participate in meetings	IW	0.00	0.23	0.50	0.50
Parents make contributions	VP	0.18	0.15	0.14	0.27
Parents make contributions	IW	-0.08	0.17	0.67	0.67
Parents help with schoolwork	VP	0.52	0.13	0.00	0.00
Parents help with schoolwork	IW	0.14	0.14	0.17	0.17
Diverse parents speaking	VP	0.40	0.24	0.07	0.14
Diverse parents speaking	IW	0.07	0.32	0.42	0.42

Table S8: Adjusted p-values using Benjamini-Hochberg for Bukoba study results (Fig. 5 in the paper): adjusting within each outcome, across treatment groups. All the statistically significant effects still retain their significance.

Outcome	Treatment	Effect size	SE	p-value	adjusted p-value
Educational outcomes	VP	0.38	0.15	0.01	0.03
Educational outcomes	IW	0.12	0.19	0.27	0.79
Study students	VP	0.10	0.16	0.28	0.31
Study students	IW	-0.03	0.17	0.56	0.79
Local government responsiveness	VP	-0.08	0.19	0.66	0.66
Local government responsiveness	IW	-0.13	0.20	0.74	0.79
Parents taking notable actions	VP	0.21	0.17	0.13	0.18
Parents taking notable actions	IW	-0.18	0.22	0.79	0.79
Teacher parent interactions	VP	0.64	0.13	0.00	0.00
Teacher parent interactions	IW	-0.08	0.17	0.68	0.79
Parents participate in meetings	VP	0.25	0.16	0.08	0.14
Parents participate in meetings	IW	0.00	0.23	0.50	0.79
Parents make contributions	VP	0.18	0.15	0.14	0.18
Parents make contributions	IW	-0.08	0.17	0.67	0.79
Parents help with schoolwork	VP	0.52	0.13	0.00	0.00
Parents help with schoolwork	IW	0.14	0.14	0.17	0.79
Diverse parents speaking	VP	0.40	0.24	0.07	0.14
Diverse parents speaking	IW	0.07	0.32	0.42	0.79

Table S9: Adjusted p-values using Benjamini-Hochberg for Bukoba study results (Fig. 5 in the paper): adjusting within each treatment group, across outcomes. All the statistically significant effects still retain their significance.

13 Regression Tables

Regression Tables for Figures 4, 5, and 6 in the manuscript.

	Efficacy	Internal Eff	External Eff	Individual Eff	Group Eff
VP	0.10 (0.10)	0.09 (0.10)	0.10 (0.12)	0.08 (0.11)	0.12 (0.10)
IW	0.08 (0.10)	0.09 (0.11)	0.08 (0.11)	0.09 (0.11)	0.08 (0.11)
Base Efficacy	0.24 (0.05)***				
Base Internal Eff		0.23 (0.05)***			
Base External Eff			0.18 (0.05)***		
Base Individual Eff				0.23 (0.05)***	
Base Group Eff					0.20 (0.05)***
Political	-0.01 (0.04)	-0.14 (0.05)***	0.09 (0.06)	-0.03 (0.04)	0.00 (0.05)
Parent (vs. Caregiver)	0.10 (0.08)	0.02 (0.08)	0.16 (0.10)	0.13 (0.08)	0.05 (0.09)
Muslim	-0.00 (0.08)	0.03 (0.08)	-0.04 (0.10)	0.00 (0.08)	-0.02 (0.08)
Literate	0.15 (0.09)	0.21 (0.13)	0.12 (0.10)	0.13 (0.09)	0.18 (0.11)
Household Head	-0.09 (0.08)	-0.05 (0.10)	-0.10 (0.08)	-0.10 (0.09)	-0.05 (0.09)
Household Size	-0.01 (0.01)	0.01 (0.02)	-0.03 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Female	-0.00 (0.05)	-0.02 (0.07)	0.02 (0.05)	0.04 (0.06)	-0.06 (0.06)
SES Index	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)
No. of Children	0.03 (0.02)	0.02 (0.02)	0.04 (0.02)	0.04 (0.02)	0.02 (0.02)
Education Level	0.07 (0.03)**	0.07 (0.03)*	0.07 (0.03)*	0.09 (0.03)**	0.04 (0.03)
Age	0.04 (0.03)	-0.01 (0.04)	0.08 (0.03)***	0.04 (0.03)	0.04 (0.03)
Intercept	2.54 (0.27)***	2.81 (0.29)***	2.54 (0.27)***	2.44 (0.27)***	2.84 (0.28)***
Base Mean	3.89	4.01	3.79	3.90	3.88
Base SD	0.71	0.82	0.76	0.77	0.76
VP - IW	0.02 (0.1)	0 (0.1)	0.03 (0.11)	-0.01 (0.11)	0.04 (0.11)
R ²	0.13	0.14	0.08	0.13	0.09
Adj. R ²	0.11	0.11	0.06	0.11	0.07
Num. obs.	574	574	574	574	574
RMSE	0.04	0.05	0.05	0.05	0.05
N Clusters	24	24	24	24	24

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S10: Study 1 Bukoba Efficacy Outcomes: Educational Efficacy score, Internal, External, Individual, and Group scores on VP and IW interventions (as opposed to SO) using OLS models with demographic covariates and school cluster robust standard errors. Below each model are the means and standard deviations for baseline.

	Teacher Absence Ask	Spoke w Teacher	Comment for HT	Looked at HW	Gave Breakfast	Secondary	Test Score
VP	0.11 (0.05)*	0.13 (0.07)*	0.09 (0.09)	0.16 (0.05)***	0.09 (0.05)*	0.07 (0.08)	-1.06 (1.77)
IW	0.15 (0.04)***	0.10 (0.07)	0.12 (0.09)	0.15 (0.04)***	0.10 (0.05)**	0.02 (0.10)	-1.15 (1.67)
Base Teacher Absence Ask	0.16 (0.04)***						
Base Looked at HW				0.11 (0.05)**			
Base Gave Breakfast					0.35 (0.04)***		
Base Secondary						0.09 (0.04)**	
Base Test Score							0.49 (0.06)***
Base Efficacy	0.06 (0.03)**	0.11 (0.03)***	-0.01 (0.04)	0.06 (0.03)**	0.00 (0.02)	0.07 (0.03)**	0.09 (0.77)
Political	0.00 (0.04)	-0.02 (0.04)	0.01 (0.05)	-0.04 (0.03)	0.00 (0.04)	-0.02 (0.07)	0.87 (1.42)
Parent (vs. Caregiver)	0.04 (0.05)	0.12 (0.06)*	-0.07 (0.05)	0.05 (0.05)	0.02 (0.05)	0.08 (0.08)	0.17 (1.96)
Muslim	0.11 (0.04)**	0.07 (0.06)	0.18 (0.08)**	0.06 (0.04)	-0.08 (0.04)*	-0.04 (0.06)	0.86 (1.10)
Literate	0.12 (0.08)	0.05 (0.08)	0.09 (0.08)	0.27 (0.07)***	0.14 (0.08)	0.17 (0.10)*	2.12 (2.25)
Household Head	-0.02 (0.06)	-0.05 (0.07)	0.05 (0.07)	-0.00 (0.05)	0.02 (0.04)	0.05 (0.07)	0.78 (1.29)
Household Size	0.00 (0.02)	-0.00 (0.01)	0.02 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.22 (0.21)
Female	0.09 (0.05)	-0.00 (0.06)	-0.12 (0.05)**	0.03 (0.05)	0.10 (0.03)***	-0.01 (0.06)	-1.25 (1.31)
SES Index	-0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.02 (0.01)	0.01 (0.02)	0.64 (0.30)**
No. of Children	0.00 (0.02)	0.02 (0.02)	-0.02 (0.02)	0.00 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.43 (0.36)
Education Level	0.00 (0.02)	0.02 (0.03)	0.00 (0.03)	-0.01 (0.02)	-0.01 (0.03)	-0.02 (0.03)	-0.75 (0.66)
Age	-0.02 (0.02)	0.04 (0.02)	-0.07 (0.02)***	0.01 (0.01)	0.01 (0.02)	-0.03 (0.02)	-0.57 (0.63)
Intercept	0.17 (0.18)	-0.54 (0.18)***	0.47 (0.22)**	0.02 (0.18)	0.18 (0.18)	0.12 (0.24)	21.92 (5.52)***
Base Mean	0.45	Endline only	Endline only	0.64	0.6	4.31	21.59
Base SD	0.5	Endline only	Endline only	0.48	0.49	0.94	11.42
VP - IW	-0.05 (0.04)	0.04 (0.08)	-0.03 (0.09)	0.01 (0.03)	-0.01 (0.06)	0.05 (0.1)	0.1 (1.66)
R ²	0.10	0.06	0.06	0.10	0.21	0.06	0.23
Adj. R ²	0.07	0.04	0.03	0.08	0.19	0.03	0.21
Num. obs.	574	574	574	572	571	574	594
RMSE	0.03	0.03	0.03	0.03	0.03	0.03	0.82
N Clusters	24	24	24	24	24	24	24

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S11: Study 1 Bukoba Behavior and Test Score Outcomes on VP and IW interventions (as opposed to SO) using OLS models with demographic covariates and school cluster robust standard errors. Below each model are the means and standard deviations for baseline.

	Teacher Interactions	Participate in Mtgs	Notable Actions	Edu Outcomes	Study Students	Local Gov Response
VP	0.64 (0.13)***	0.25 (0.16)	0.21 (0.17)	0.38 (0.15)**	0.10 (0.16)	-0.08 (0.19)
IW	-0.08 (0.17)	0.00 (0.23)	-0.18 (0.22)	0.12 (0.19)	-0.03 (0.17)	-0.13 (0.20)
Resp Type FE	Yes	Yes	Yes	Yes	Yes	Yes
VP - IW	0.73 (0.14)	0.25 (0.16)	0.39 (0.22)	0.26 (0.13)	0.13 (0.17)	0.05 (0.14)
R ²	0.18	0.09	0.04	0.11	0.02	0.01
Adj. R ²	0.16	0.00	0.02	0.08	-0.02	-0.02
Num. obs.	116	23	115	115	54	112
RMSE	0.74	0.39	0.77	0.47	0.47	0.63
N Clusters	24	23	24	24	21	24

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S12: Study 1 Bukoba, 3 point scaled School-Level Outcomes on VP and IW interventions (as opposed to SO) using OLS models with respondent type fixed effects and school cluster robust standard errors.

	Contributions	Help w. Schoolwork	Diverse Speaking
VP	0.18 (0.15)	0.52 (0.13)***	0.40 (0.24)
IW	-0.08 (0.17)	0.14 (0.14)	0.07 (0.32)
Resp Type FE	Yes	Yes	Yes
VP - IW	0.26 (0.13)	0.38 (0.11)	0.33 (0.21)
R ²	0.06	0.18	0.20
Adj. R ²	0.04	0.16	0.10
Num. obs.	114	89	19
RMSE	0.48	0.46	0.40
N Clusters	24	24	19

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S13: Study 1 Bukoba Binary School-Level Outcomes on VP and IW interventions (as opposed to SO) using OLS models with respondent type fixed effects and school cluster robust standard errors.

	Efficacy	Internal Eff	External Eff	Individual Eff	Group Eff
VP	0.07 (0.03)*	0.06 (0.04)	0.08 (0.04)*	0.06 (0.04)	0.09 (0.04)*
IW	0.05 (0.04)	0.04 (0.03)	0.06 (0.05)	0.03 (0.04)	0.07 (0.05)
Morning session	0.04 (0.03)	0.02 (0.03)	0.06 (0.04)	0.05 (0.03)	0.03 (0.03)
Female	-0.19 (0.05)***	-0.20 (0.07)***	-0.19 (0.04)***	-0.14 (0.05)**	-0.26 (0.05)***
Kisw Language	0.26 (0.08)**	0.35 (0.08)***	0.19 (0.08)*	0.35 (0.09)***	0.15 (0.08)*
Parent	-0.07 (0.05)	-0.04 (0.05)	-0.10 (0.05)*	-0.05 (0.05)	-0.09 (0.05)*
Literate	0.16 (0.07)**	0.19 (0.07)**	0.14 (0.08)	0.14 (0.07)*	0.19 (0.07)**
Muslim	-0.04 (0.05)	-0.06 (0.05)	-0.02 (0.06)	-0.06 (0.06)	-0.02 (0.05)
Political	0.08 (0.05)	0.09 (0.05)	0.07 (0.05)	0.04 (0.05)	0.12 (0.05)**
Maj Ethnic Group	-0.08 (0.08)	-0.09 (0.08)	-0.06 (0.10)	-0.06 (0.09)	-0.10 (0.08)
Household Size	-0.01 (0.02)	-0.00 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Age (binned)	0.03 (0.02)	0.02 (0.02)	0.03 (0.02)	0.03 (0.02)	0.02 (0.02)
No. of Children	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Child Grade	0.00 (0.03)	0.02 (0.03)	-0.01 (0.03)	-0.00 (0.03)	0.01 (0.03)
SES	0.02 (0.01)*	0.02 (0.01)*	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)*
Education	0.11 (0.03)***	0.14 (0.03)***	0.09 (0.03)***	0.11 (0.03)***	0.11 (0.03)***
Intercept	3.03 (0.16)***	2.81 (0.16)***	3.21 (0.19)***	2.99 (0.17)***	3.08 (0.19)***
School FE	Yes	Yes	Yes	Yes	Yes
Control Mean	3.50	3.59	3.43	3.54	3.45
Control SD	0.73	0.88	0.78	0.79	0.82
VP - IW	0.02 (0.03)	0.02 (0.04)	0.02 (0.03)	0.03 (0.03)	0.02 (0.04)
R ²	0.14	0.13	0.11	0.10	0.16
Adj. R ²	0.12	0.12	0.09	0.09	0.14
Num. obs.	1630	1630	1630	1630	1630
RMSE	0.71	0.84	0.76	0.76	0.78
N Clusters	16	16	16	16	16

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S14: Study 2 Kilosa Efficacy Outcomes: Educational Efficacy score, Internal, External, Individual, and Group scores on VP and IW interventions (as opposed to SO) using OLS models with demographic covariates and school fixed effects. Below each model are the means and standard deviations for control groups.

	Pro-Ed Candidate	Pub Goods Mtg	Pub Goods Official	Respected Parents	Respected Auth	Ideas Respected
VP	0.00 (0.03)	0.07 (0.06)	0.10 (0.06)	0.10 (0.03)***	0.15 (0.04)***	0.04 (0.07)
IW	-0.01 (0.02)	0.03 (0.07)	0.05 (0.05)	0.04 (0.05)	0.10 (0.06)	-0.03 (0.05)
Morning session	-0.03 (0.02)	-0.03 (0.05)	0.06 (0.06)	0.00 (0.03)	-0.02 (0.04)	0.00 (0.05)
Female	0.03 (0.03)	-0.36 (0.10)***	-0.27 (0.07)***	-0.11 (0.09)	-0.09 (0.07)	-0.18 (0.07)**
Kisw Language	0.07 (0.04)*	0.06 (0.09)	0.38 (0.12)**	0.38 (0.08)***	0.33 (0.06)***	0.20 (0.08)**
Parent	0.02 (0.04)	-0.12 (0.06)*	-0.07 (0.11)	-0.16 (0.06)**	-0.08 (0.06)	-0.17 (0.08)*
Literate	0.03 (0.06)	0.37 (0.09)***	0.24 (0.12)*	-0.12 (0.10)	-0.08 (0.10)	0.11 (0.12)
Muslim	0.00 (0.04)	-0.00 (0.07)	-0.11 (0.05)**	-0.03 (0.06)	-0.05 (0.05)	-0.12 (0.06)*
Political	-0.05 (0.03)*	0.27 (0.07)***	0.06 (0.07)	0.25 (0.05)***	0.17 (0.06)**	0.12 (0.06)*
Maj Ethnic Group	0.03 (0.04)	-0.27 (0.14)*	-0.05 (0.11)	-0.24 (0.09)**	-0.11 (0.08)	-0.19 (0.11)
Household Size	-0.01 (0.01)	0.01 (0.02)	-0.00 (0.02)	0.01 (0.01)	0.04 (0.02)*	0.00 (0.01)
Age (binned)	0.03 (0.01)*	0.10 (0.03)***	0.12 (0.04)**	0.08 (0.03)**	0.14 (0.04)***	0.09 (0.03)**
No. of Children	-0.00 (0.01)	-0.00 (0.04)	0.01 (0.04)	0.02 (0.03)	-0.04 (0.04)	0.01 (0.03)
Child Grade	0.00 (0.02)	0.06 (0.04)	0.04 (0.03)	0.04 (0.02)	0.05 (0.02)**	-0.05 (0.03)
SES	0.02 (0.01)**	0.06 (0.02)***	0.02 (0.02)	0.01 (0.01)	-0.00 (0.02)	0.04 (0.02)**
Education	0.04 (0.02)*	-0.02 (0.03)	0.08 (0.05)	0.04 (0.04)	0.10 (0.05)*	0.13 (0.06)**
Intercept	0.51 (0.10)***	3.16 (0.36)***	2.75 (0.27)***	3.50 (0.22)***	2.84 (0.14)***	2.60 (0.31)***
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	0.70	3.66	3.65	4.08	3.87	3.00
Control SD	0.46	1.29	1.26	0.98	1.09	1.16
VP - IW	0.01 (0.03)	0.04 (0.05)	0.05 (0.06)	0.07 (0.05)	0.06 (0.07)	0.08 (0.06)
R ²	0.05	0.12	0.09	0.07	0.07	0.09
Adj. R ²	0.03	0.11	0.07	0.05	0.06	0.07
Num. obs.	1630	1630	1630	1630	1630	1630
RMSE	0.45	1.21	1.21	0.93	1.01	1.13
N Clusters	16	16	16	16	16	16

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table S15: Study 2 Kilosa Behavioral and Respect Outcomes on VP and IW interventions (as opposed to SO) using OLS models with demographic covariates and school fixed effects. Below each model are the means and standard deviations for control groups.

References

- Abadie, Alberto. 2005. "Semiparametric difference-in-differences estimators." *The Review of Economic Studies* 72(1):1–19.
- Finkel, Steven E. 1985. "Reciprocal effects of participation and political efficacy: A panel analysis." *American Journal of political science* pp. 891–913.
- Fong, Christian, Marc Ratkovic and Kosuke Imai. 2014. "CBPS: R package for covariate balancing propensity score." *Comprehensive R Archive Network (CRAN)* .
- Gallagher, Mary E. 2006. "Mobilizing the law in China: "informed disenchantment" and the development of legal consciousness." *Law & Society Review* 40(4):783–816.
- Hartman, Erin and F Daniel Hidalgo. 2018. "An equivalence approach to balance and placebo tests." *American Journal of Political Science* 62(4):1000–1013.
- Jerusalem, Matthias and Ralf Schwarzer. 1995. "Generalized self-efficacy scale." *Measures in health psychology: A user's portfolio. Causal and control beliefs*. pp. 35–37.
- Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *The Annals of Applied Statistics* 7(1):295–318.
- Moore, Ryan T. 2016. *blockTools: Blocking, Assignment, and Diagnosing Interference in Randomized Experiments*. R package version 0.6-3.
- Morrell, Michael E. 2005. "Deliberation, democratic decision-making and internal political efficacy." *Political Behavior* 27(1):49–69.
- Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Armida Alishjabana, Arya Gaduh and Rima Prama Artha. 2011. "Improving educational quality through enhancing community participation: results from a randomized field experiment in Indonesia." *World Bank Policy Research Working Paper Series, Vol .*
- Przeworski, Adam. 1998. "Deliberation and ideological domination." *Deliberative democracy* 1:140.
- Rao, Vijayendra, Kripa Ananthpur and Kabir Malik. 2017. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." *World Development* .
- Soboleva, Irina. 2019. "No Politics, Please! When Efficacy Suppresses Civic and Political Engagement."
- Stokes, Susan. 1998. "Pathologies of deliberation." *Deliberative democracy* 1:123.
- Zhou, Yang-Yang and Evan Lieberman. 2020. "Replication Data for: Self-efficacy and Citizen Engagement in Development: Experimental Evidence from Tanzania."
URL: <https://doi.org/10.7910/DVN/SLYEIJ>