Moral Education and Child Development: A Randomized Controlled Trial in Japan*

Hideo Akabayashi † Kenju Kamei ‡ Tim Ruberg § Mirka Zvedelikova ¶

January 29, 2026

Abstract

Children in Japan face numerous challenges in school, which may hinder their skill development. Although parental involvement could help mitigate these difficulties, limited parent-child interaction may worsen these adverse effects. We conduct a randomized controlled trial (RCT) designed to mitigate these adverse effects by encouraging parents to discuss school-related difficulties with their children. Our six-week, home-based program utilizes a dual approach: educating parents via weekly emails describing common challenges faced by children, while parents watch age-appropriate videos on moral education with their children and discuss their contents. The videos are presented by either a representative teacher or a celebrity, allowing us to test whether narrator choice enhances intervention effectiveness. We evaluate this low-cost and highly scalable intervention using behavioral outcomes measured immediately after program completion, as well as a broad set of cognitive and non-cognitive outcomes assessed approximately six months later.

JEL Classification: D83; D91; I21; I24; M37.
Keywords: Moral education; Child development; Early childhood intervention; Celebrity endorsement; Randomized controlled trial (RCT).

*The authors would like to thank Rajeev Dehejia, Mai Seki, and seminar participants at Doshisha University, Keio University, and the University of Tokyo as well as participants at the EMSR Conference, the Asia-Pacific Meeting of the ESA, the AFE Conference, the European Meeting of the ESA, and the EAEA International Conference for their helpful comments. We gratefully acknowledge financial support from the Faculty of Economics, Keio University, and JSPS KAKENHI Grant Number 21H04982. We are also grateful for the technical support provided by the CREOC, Keio University. The views expressed in this paper are those of the authors and do not reflect the positions or policies of the organizations mentioned above. Declarations of interest: none. This research was approved by the Research Ethics Review Committee of Keio University on July 12, 2023. AEA registry trial ID number: AEARCTR-0013723.

†Hideo Akabayashi, Keio University, Faculty of Economics, 2-15-45, Mita, Minato-ku Tokyo 108-8345, Japan; Email: hakab@econ.keio.ac.jp

‡Kenju Kamei, Keio University, Faculty of Economics, 2-15-45, Mita, Minato-ku Tokyo 108-8345, Japan; Email: kenju.kamei@keio.jp

§Tim Ruberg, Market Design Center, University of Tokyo, Graduate School of Economics, 7-3-1 Hongo, Bunkyo-ku, 113-0033 Tokyo, Japan; Email: ruberg@g.ecc.u-tokyo.ac.jp

¶Mirka Zvedelikova, Faculty of Economics, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan; Email: zvedelikova@cc.kyoto-su.ac.jp.

1 Introduction

The child mental health crisis has been described as "the defining public health crisis of our time," having a long-lasting impact on well-being and economic circumstances beyond the affected individual (Currie, 2025). Japan is not an exception, with over 40% of adolescents in Tokyo affected by behavioral problems, including anxiety and depression (Nagaoka et al., 2024). Schools can be a major source of stress, with confirmed cases of bullying and chronic absenteeism reaching an all-time high and around 400 child suicides in 2023 (MEXT, 2024).[1] Pressure to perform, course overload, peer pressure, bullying, and anxiety can have a negative impact on academic achievement, in turn affecting future career paths, but also on the development of socioemotional skills. While the home environment including parent-child interaction is a strong determinant of academic achievement and child well-being and therefore could help mitigate these negative effects (Cunha and Heckman, 2007; Fiorini and Keane, 2014; Zumbuehl et al., 2021), children in Japan spend most of their time outside the home or studying (Nakamura et al., 2020; Hertog and Zhou, 2021), leaving little time for interactions with parents to form close trusting relationships, thus rarely disclosing and discussing personal issues (Nucci et al., 2014).

This paper describes short-run findings from a low-cost, easily scalable intervention, conducted in 2023, using online educational tools in a home-based environment to address social issues that children may be experiencing, targeting Japanese children between the ages of 9 to 14. We employ a dual approach: raising awareness of social issues in parents, equipping them with information necessary to discuss the issues with their children, and providing easily approachable educational videos to children, showcasing a related dilemma. The weekly videos are based on moral education textbooks with text and illustrations, narrated by either a female teacher or a female celebrity, with varying degrees of identity revealed. The topics we cover are (i) school life and friends, (ii) bullying, (iii) self-confidence, (iv) self-control, and (v) happiness and well-being. This treatment was conducted in 6 consecutive weekly cycles, with weekly short recall and compliance surveys, and final short and long-term surveys.

We recruit over a thousand households to participate in this randomized controlled trial (RCT) to test the impact of the online intervention in both a nationally representative sample of the child age group, as well as in an at-risk population. The videos are positive in nature and tailored to the issues Japanese children tend to struggle with. We provide evidence that treatment assignment was random and therefore treatment and control group are balanced

[1] Figure A1 in the Appendix presents the development of bullying, absenteeism, and suicide cases among children in Japan over time.

with respect to observable characteristics. Attrition is low and overall participation measured by watch statistics, recall questions, and individual feedback suggests that the delivery of the intervention was successful.

Our results suggest that the intervention has little impact on short-term behvaioral outcomes in the nationally representative sample, but has a positive effect on children identified as at-risk by baseline behavioral issues (i.e., SDQ), low academic scores, or family background. In particular, we find persistent improvements for children from high-SES households and mothers who spend more time at home, predominantly in moral behavior and self-control. Although one goal of this intervention is to improve the parent-child relationship, it seems that it requires a parent (often the mother) that has the cognitive capacity to understand the materials and is able to properly convey the message. While further analysis is needed to clarify the mechanism behind these results and how children attributes interact with our intervention, it is clear that a general implementation of such a program may not lead to the intended goals, while a targeted one could help children in need.

We further examined the role of narrator type and identity disclosure. Celebrity narration helped mitigate negative effects observed under teacher narration, particularly for well-being and self-esteem, but otherwise had limited direct impact on behavioral outcomes. The most robust positive effects were observed in participation, engagement, and recall, particularly when the celebrity's identity was disclosed without showing their picture. Adding a picture negated these gains, indicating that added visual may distract from substantive content. These findings highlight both the potential and the shortcomings of edutainment strategies in education: professional narration can improve engagement and avoid harm, but excessive emphasis on the celebrity may crowd out the intended educational message.

This research contributes to several strands of the literature. First, we add to the growing literature on RCTs targeting the home environment and parental involvement to improve children's cognitive or non-cognitive outcomes (Hastings and Weinstein, 2008; Avvisati et al., 2014; Doss et al., 2019; Mayer et al., 2019; York et al., 2019; Attanasio et al., 2020; Bergman, 2021; Angrist et al., 2025). These studies are typically conducted in cooperation with schools, whereas we reach households directly, avoiding potential issues by going through schools. Second, we contribute to the literature on RCTs using online tools to deliver their intervention (Figlio et al., 2013; Bergman and Chan, 2021; Cacault et al., 2021; Angrist et al., 2022; Berlinski et al., 2022; Phillips and Reber, 2022). These interventions are quite attractive as they are of low cost and easily scalable. To the best of our knowledge, we are the first to implement a large-scale intervention targeting the home environment using online tools in Japan. Finally, we provide evidence as to the effectiveness of "edutainment," combining education with elements from the entertainment industry, such as shows or celebrities, to

present educational content (Gentzkow and Shapiro, 2008; Veer et al., 2010; Nownes, 2012; Bergkvist and Zhou, 2016; Kearney and Levine, 2019; Bjorvatn et al., 2020; Schartel Dunn and Nisbett, 2020).

The remainder of this paper proceeds as follows. In Section 2, we describe the institutional context and moral education in Japan. In Section 3, we outline the experimental design, describe the intervention, explain our outcome measures, and provide balancing and attrition tables. In Section 4, we discuss our pre-registered hypotheses. In Section 5, we describe our empirical strategy. In Section 6, we present our main results. In Section 7, we discuss differential effects by treatment arm (i.e., celebrity narration). Finally, Section 8 concludes.

2 Moral Education in Japan

Compulsory education in Japan consists of six years of elementary school starting from age 6 and three years of junior high school from age 12 with a common curriculum set by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). This is followed by largely universal three-year high school education from age 15, with an attendance rate of 98.8% (MEXT, 2021). The school year starts in April, and children enjoy about six weeks of summer break from mid-July to late August before the beginning of the next term. We implemented our intervention from mid-June to mid-July 2024 (6 weeks), to avoid conflicts with long breaks, entrance exams and other potentially interfering major events, to children in grades 4 through 9 (9 to 14 years of age).

The nine years of compulsory schooling include a weekly Moral Education class, covering topics such as (i) bullying, friendship, compassion, and trust, (ii) tolerance and understanding, (iii) happiness, (iv) independence, ambition, and responsibility, and (v) hope and courage, among others. It comprises around 35 lesson hours per year and thus 3–4% of all hours. This class is delivered using stories which present a dilemma and pose questions to students regarding their own behavior in these hypothetical situations without providing a clear answer. These stories are meant to be discussed in class with the teacher facilitating the discussion.

We base the intervention videos on moral education textbooks to encourage parental involvement in the program using MEXT certified materials to avoid any potential harm to children. This part of the intervention is, by design, mild. Eight certified moral education textbooks are available in the Japanese market, with local boards of education typically making the selection for the schools in their jurisdiction. We selected stories from textbooks with low market shares (2–9%) not repeated in other textbooks. The low market share is caused by small textbook portfolios of the publishing houses and corresponding weak

marketing efforts, not by low quality.

The textbooks classify each story by topic corresponding to the curriculum. We chose stories from the following categories to target the issues Japanese children face: (i) school life and friends, (ii) bullying, (iii) self-confidence, (iv) self-control, (v) happiness and well-being. The bullying topic is discussed in two weeks to present different perspectives, i.e., the different forms bullying can take and its long-term consequences for the victim. To ensure the content is age appropriate and children have not been previously exposed to the stories, we create three groups based on school grade (4–5, 6–7, 8–9) and select the appropriate stories from a textbook for one higher grade or from a textbook for the same grade but discussed later in the school year, in case no other option is available.

However, as moral education is part of the compulsory schooling curriculum, students will eventually be exposed to the same topics. Therefore, it is important to emphasize that our intervention is not designed to test the effectiveness of moral education itself or the teaching mode (e.g., online as opposed to in-classroom). Instead, our intervention aims to use moral education as a catalyst to increase parent-child interaction, allowing for a better, more trustful relationship. It is possible that in the absence of a similar curriculum, the video content itself may have an effect on non-cognitive outcomes. However, the Japanese setting does not allow us to test this hypothesis.

3 Experimental Design

3.1 Study Design

Our intervention targeted children and their parents across Japan, who participated in a total intervention period of six weeks. Parents in the treatment group received weekly emails about the prevalence and severity of a specific social issue in Japan, and were asked to have their children watch a video in which a teacher or a celebrity reads a story from a moral education textbook related to the social issue. Both treatment and control groups were invited to participate in short and long-run post-intervention surveys with treated children and parents receiving additional weekly surveys. An illustration of the study design is shown in Figure 1.

3.2 Weekly Schedule

The design of each intervention week is as follows. First, on Thursday at 4 PM, treated parents received an email informing them about the schedule for the given week. Starting in week 2, this email also contained a link to a short video content recall survey (e.g., main

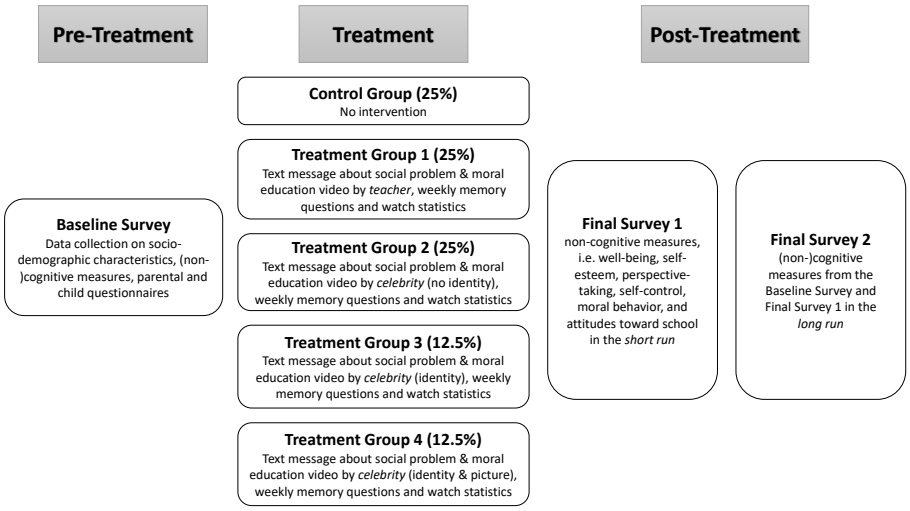


Figure 1: Illustration of Study Design

Notes: The figure shows an illustration of the study design.

character name) for the child to answer. On Friday at 4PM, parents received a second email with information about the social issue of the week, a short parental survey regarding the previous video (e.g., them having watched it with the child) and a link to the week's video to be shown to the child. The video is immediately available, with the watch period starting on Friday at 4 PM and closing on Wednesday 4 PM. The children thus answer the recall survey starting 24 hours after the video is no longer available. An optional third weekly email is sent on Wednesday, 8 AM, a reminder to parents of non-responding children to the previous week's recall survey. The control group parents receive a weekly message informing them there is no video to watch in this week. Figure 2 depicts this weekly schedule

3.3 Intervention Content

The first week's topic is *School Life and Friends* and informs parents about the problems their children may face in school. A message to the parents emphasizes the growing problem of absenteeism in Japan, which is often the result of problems with friends, a poor child-teacher relationship, or a lack of understanding of the school materials (nippon.com, 2023; Child Development Science Research, 2024). The second week focuses on *Bullying*, with an emphasis on exclusion and gossip. Parents are informed about the prevalence of bullying in Japanese schools and the different forms that bullying can take (e.g., physical violence, hiding things, gossiping) (Ministry of Education, Culture, Sports, Science and Technology, 2019; Children and Families Agency, 2024). The third week focuses on the negative (long-term)

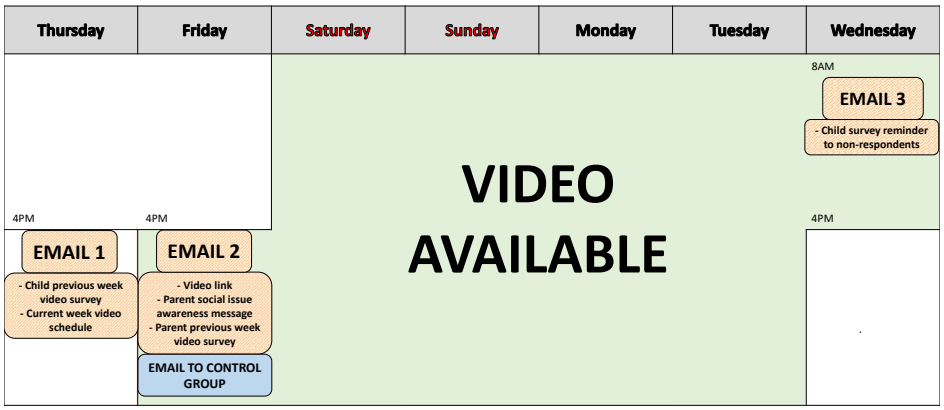


Figure 2: Illustration of Weekly Schedule

Notes: The figure shows the weekly schedule of the intervention.

consequences that bullying can have on a child (Oshio et al., 2013; Takizawa et al., 2014; Arseneault, 2018). The fourth week targets *Self-Confidence* with a focus on individuality, goals and dreams. The message explains to parents the lack of self-confidence of Japanese children in international comparison, based on the results of the Programme for International Student Assessment (PISA) (National Institute for Educational Policy Research, 2016; The Nippon Foundation, 2019). It also provides examples of how parents can help improve their child's self-confidence. The fifth week's topic is *Self-Control*. The message explains to parents the concept of self-control as it relates to smartphone use and provides suggestions for parents on how to reduce the amount of time their child spends using a smartphone. The sixth week focuses on *Happiness and Well-Being*, informing parents about Japan's comparatively low average levels of life satisfaction and happiness (UNICEF, 2021; Economic and Social Research Institute, 2024), and providing examples of how children can achieve happiness based on a popular book of Maeno (2017). To prevent the order and the timing of the video provision from affecting the children's motivation and thus the effectiveness of the intervention (i.e., overlapping with national holidays or important exams), we further divide each treatment group and reverse the order of the weekly topics for one group.[2]

While all parents of treated children received identical messages, we prepare four different types of videos and thus distinguish four different treatment groups. The videos show a text on a white background with appropriate phonetic aid for Japanese characters, accompanied by the original textbook illustrations. The narrator reads the presented text and the videos are, on average, 6 minutes long, adjusted to identical length for both narrators. The first

[2]After the final survey concluded, we allowed all families in the control group to watch the age-appropriate videos as well.

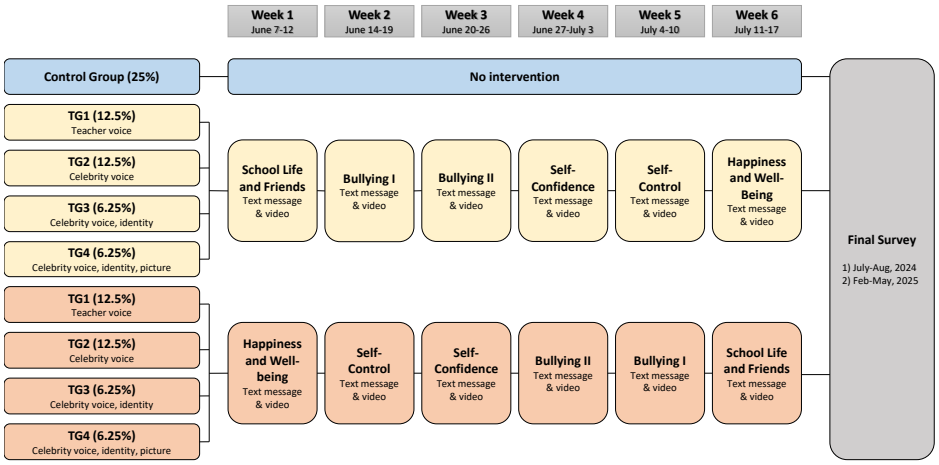


Figure 3: Intervention Timeline and Content

Notes: The figure shows the intervention timeline and and the weekly content.

treatment group (TG1) watched videos presented by a representative female teacher without revealing her identity or face. The second treatment group (TG2) was shown identical videos presented by a female celebrity without revealing her identity or face. The third treatment group (TG3) watched the same video as TG2, but the name of the celebrity was revealed. Finally, the fourth treatment group (TG4) was provided with the name of the celebrity and her picture at the beginning and end of the video. Combined with three age-based groups, 12 video versions per week and a total of 72 videos were created. Figure 3 presents the details of the intervention timeline and content.[3]

The celebrity is close in age to the teacher (around 30 years old at the time of voice recording). She is a former member of a major pop group, singer, actress, TV personality and regularly appears on TV. While not all children in the program may be familiar with her, given the age range, parents are expected to have at least a passing knowledge of her and the pop group she belonged to. This affiliation is mentioned alongside her name. Using a professional with acting experience trained in voice work in the celebrity narration treatment arms creates a clear contrast with the teacher narration. The professional narration is very clear, differentiates characters by voice, fully conveys emotions and varies speed and volume as appropriate. While the teacher was not coached as to how to read the text, compared to the celebrity, the narration gives a one-dimensional impression.

[3]In the Online Appendix, we share all emails sent to the parents in Japanese along with their English translation.

3.4 Participants

This intervention is a part of a larger project studying the disruptive effects of the COVID-19 pandemic on the home and learning environment of children, searching for ways to mitigate these effects. Families invited to participate in the larger project, with baseline survey conducted in early 2023, were drawn from the residential registry ensuring a regionally balanced distribution. For the current intervention, we invited all households with at least one child in grade 4 of elementary school to grade 3 of junior high school (ages 9 to 14) for a total of 2,684 families, covering all of Japan.[4] For households with more than one eligible child, we chose the child to participate randomly. 1,091 families accepted our invitation. Table A1 in the Appendix presents results from linear probability models of program participation to test whether families selectively decide to participate in the intervention or not. Although we find that families with more children are more likely to participate, this is not surprising as it is easier for these families to have a child that wants to participate. Similarly, a small negative correlation with mother's fulltime employment status suggests that mothers who have more time are more likely to participate. However, it is important to note that the gender of the child and parental education seems to be unrelated to participation.[5]

Participating children were then randomly assigned into control and treatment groups, stratified by grade level (age) and gender. 25% were assigned to the control group, 25% to TG1 (teacher voice), 25% to TG2 (celebrity voice), 12.5% to TG3 (celebrity voice and identity) and finally, 12.5% to TG4 (celebrity voice, identity, and picture). For each treatment group, half of the sample received the treatment in standard order and half in reversed order. Figure 4 shows the number of participants and treatment assignment in detail.[6]

3.5 Data Collection

We collect data from the baseline survey, data on engagement with the materials through the video platform, weekly recall questions and parental surveys during the intervention, a short-term final survey two weeks after the program concluded, and a long-term survey six months after the conclusion of the program.

The baseline survey, as part of a larger project establishing a new household panel survey, was conducted in January and February 2023 and asked a battery of questions about the household, including socioeconomic background, daily activities and attitudes, child char-

[4] In Figure A2 in the Appendix, we show the regional distribution of the final participants across Japan.

[5] Figure A3 in the Appendix compares our sample with population characteristics. On average, families in our intervention have slightly higher-educated parents and mothers who are less likely to work fulltime. However, their household income is very similar to the population mean.

[6] Figure A4 in the Appendix presents the stratification of treatment assignment by grade level.

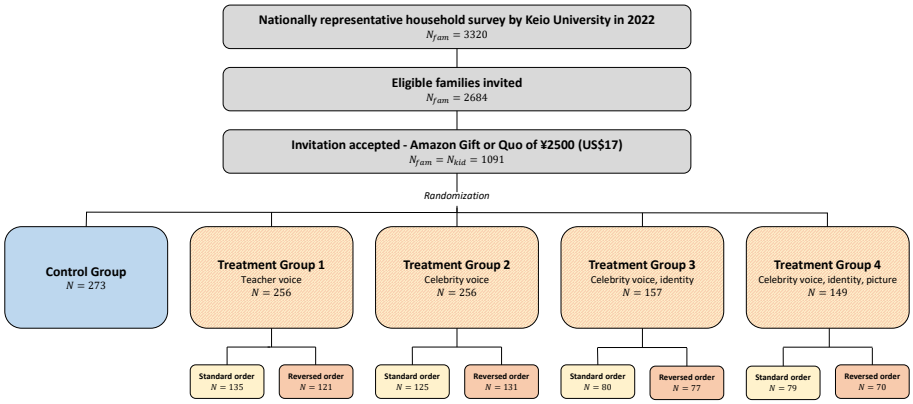


Figure 4: Participants and Group Assignment

Notes: The figure shows the number of participants and treatment assignment.

acteristics, as well as having children take a cognitive test in math and Japanese.[7] While the Japanese assessment consists of vocabulary as well as reading and writing of Kanji characters, the math assessment consists of calculations of word problems concerning numbers and the manipulation of figures (Akabayashi et al., 2023). Every item is vertically equated across grades using item response theory, thus allowing us to measure the individual latent math and Japanese theta scores (Yamaguchi et al., 2019). Further, we use responses to the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) to create a measure of behavioral problems. We rescale the SDQ score such that higher values imply "better behavior" and thus less behavioral problems. The test scores and the SDQ score are standardized by grade with a mean of zero and a unity standard deviation.[8] Finally, we use information from this baseline survey to control in our baseline specification for the stratification variables female and grade level, as well a the number of siblings, a dummy for being born in the first quarter of a year, dummies for each parent having graduated from a college, and a dummy for mother's employment status.

The platform we used to deliver the videos provided us with detailed watch statistics such as the number of times the video was watched, the total time watched, and a heatmap representing repeated or skipped parts. Additionally, we sent two surveys to the participants each week. The first survey was addressed to the child (through the parent) asking simple recall questions and comments regarding the video, and the second one to parents inquiring about whether they watched the videos, how many times, and whether they watched together with their child. We use these watch statistics and video engagement-related questions to

[7]The final long-term survey is a second wave of this household panel study.
[8]In the heterogeneity section, we sum both test scores and use this comprehensive measure instead.

measure participation and engagement with the materials, as well as use them in a complier analysis. In particular, we focus on six measures: (i) number of videos watched by the child, (ii) number of correct (recall) answers, (iii) number of (voluntary) comments under the video, (iv) percentage finished, (v) number of videos watched by the parent, (vi) number of videos watched by the parent with the child.

To measure children's behavioral outcomes in the short-term post survey, we used the following validated questionnaires. First, we use the Japanese version of the WHO-5 well-being index introduced by WHO (1998) and translated and validated by Awata et al. (2007) and Topp et al. (2015). This five-item index measures positive well-being during the past two weeks on a six-point Likert scale. The scores vary from 0 to 5 and thus the final index ranges from 0 and 25. Second, we use the Japanese version of the Rosenberg self-esteem scale introduced by Rosenberg (1965) and translated and validated by Mimura and Griffiths (2007). This index measures self-esteem as a sense of worth or value and consists of ten items on a four-point Likert scale, thus ranging from 10 to 40. Third, we use a measure of perspective-taking or empathy based on the Japanese version of the Interpersonal Reactivity Index (IRI) (Davis, 1983; Hasegawa et al., 2009), which assesses empathy-related emotions directed toward others, such as sympathy and compassion. This index consists of nine items on a five-point Likert scale, thus ranging from 9 to 45. Fourth, we create an index based on the School Liking and Avoidance Questionnaire (SLAQ) (Ladd and Price, 1987; Ladd, 1990; Otsui et al., 2006), which captures children's emotional attitudes toward school. This index is constructed as the mean of eleven three-point Likert scale items, thus ranging from 1 to 3. Fifth, we create an index based on moral behavior introduced and validated by Saito and Ikeda (2016). This index consists of the two dimensions personal and peer relationships. While the former consists of five five-point Likert scale items, the latter one is based on four items. We combine both subscales, which leads to an overall scale of moral behavior ranging from 9 to 45. Finally, we use the Japanese version of the Self-Control Rating Scale (SCRS) introduced by Kendall and Wilcox (1979) and translated and validated by Tsukamoto (1997, 2010). This index is based on nineteen four-point Likert scale items measuring an individual's capacity for self-control, particularly in everyday situations that require regulation of attention, emotion, and behavior. To allow for comparisons of effect sized, we standardized all outcome variables by grade to mean zero and unity standard deviation.

Data on the long-term survey is not available at this point and thus will not be discussed in this paper.

Table 1: Covariate Balancing Including Attriters

	Control group	Treatment group				
		TG1	TG2	TG3	TG4	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
= 1 if female	0.50	0.52	0.49	0.49	0.48	0.92
# of siblings	1.27	1.20	1.20	1.15	1.19	0.66
= 1 if young	0.22	0.20	0.21	0.29	0.26	0.30
= 1 if father college degree	0.59	0.60	0.61	0.58	0.62	0.95
= 1 if mother college degree	0.45	0.40	0.40	0.43	0.46	0.68
= 1 if mother fulltime empl.	0.38	0.35	0.41	0.36	0.41	0.74
Observations	273	256	256	157	149	

Notes: This table shows mean values for child and parental characteristics measured pre-treatment. p-values are based on F-tests of joint significance. The math and language scores are standardized by grade. $N = 1091$.

3.6 Balancing

In this subsection, we test whether there are any statistically significant differences in observable characteristics, i.e., stratification variables, between the treatment and control groups. Table 1 presents sample means for the control group and the four treatment groups alongside p-values from F-tests of joint significance of differences of all treatment groups relative to the control group.[9]

For all baseline characteristics, we fail to reject the null hypothesis that the coefficients are jointly equal to zero and conclude that the randomization is well balanced.[10]

3.7 Attrition

We further test whether take-up and attrition might be different between the control group and the four treatment groups. In Table 2, we present the number of families that agreed to participate, the ones who actually answered the final short-run survey, and the percentage of drop outs.[11]

Our results suggest that while only 7.6% of families in the control group dropped out of

[9]While balancing of observed variables across treatment groups supports our identification strategy, any remaining imbalances do not imply that the our treatment effects are biased since the observed variables examined in the balancing exercise are included as controls.

[10]Table A2 in the Appendix presents balancing checks without attriters. The observation of a successfully balanced treatment assignment holds also for the attrited sample.

[11]This attrition table considers families to have successfully participated in the intervention if we can measure the WHO-5 well-being index. There could be, however, still missings in other outcome variables.

Table 2: Take-Up and Attrition Rate

	Control group	Treatment group				Total
		TG1	TG2	TG3	TG4	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment assignment	273	256	256	157	149	1091
Final survey, short run	252	213	222	137	126	950
Attrition rate	.076	.167	.133	.127	.153	.128
Child # videos watched		3.139	3.041	3.622	3.410	
Child # correct answers		4.516	4.210	5.297	4.784	
Child # comments		2.057	1.893	2.473	2.151	
Child percentage finished		0.887	0.879	0.912	0.884	
Parent # videos watched		2.679	2.823	3.063	3.025	
Parent # videos watched w/ child		2.060	2.094	2.258	2.186	

Notes: This table shows the number of participants in the intervention and the final survey, respectively, and the corresponding attrition rate. It also shows summary statistics of take-up measures for TG1 to TG4.

the intervention, between 12.7 and 16.7% of families in the treatment groups dropped out of the intervention, leaving us with a successful participation rate of 87.2%. This is quite high compared to similar interventions and the differences between the groups are small, suggesting that any attrition bias is minimal at best. Further, the small differences in take-up between the treatment groups might be explained by the type of treatment. For instance, families in TG3 have the lowest drop-out rate and show the strongest engagement with the materials.

4 Hypotheses

The main goal of this intervention is to improve child development through an improved parent-child relationship using videos based on moral education textbook stories and an information treatment in the home environment. Therefore, the first of our preregistered hypotheses is as follows:

Hypothesis 1 (H1)**.** *Compared to children in the control group, treated children will show improvements in non-cognitive and cognitive skills.*

Because we target specific social issues and encourage children to reflect on the dilemma presented in each week's story and discuss the implications with their parents, we expect to see improvements in socioemotional skills. Therefore, the second hypothesis is as follows:

Hypothesis 2 (H2)**.** *Compared to children in the control group, treated children will show improvements in well-being, self-esteem, perspective-taking, self-control, moral behavior, and attitudes toward school.*

The videos are presented by either a teacher or a celebrity. Because the celebrity is trained and more experienced in entertaining her audience and keeping her listeners attentive, we expect stronger improvements in non-cognitive and cognitive skills for children who watched videos presented by a celebrity. Therefore, the third hypothesis is as follows:

Hypothesis 3 (H3)**.** *Compared to children who watched videos presented by a representative teacher (TG1), children who watched videos presented by a celebrity (TG2 & TG3 & TG4) will show greater improvements in non-cognitive and cognitive skills, greater video engagement, and better recall.*

Although we expect the celebrity to be able to engage children more, it is a priori unclear whether knowing their identity leads to greater engagement. Nevertheless, the fourth hypothesis is as follows:

Hypothesis 4 (H4)**.** *Compared to children who watched videos presented by a celebrity but did not know the identity of the celebrity (TG2), children who knew about the celebrity's identity (TG3 & TG4) will show greater improvements in non-cognitive and cognitive skills, greater video engagement, and better recall.*

Many educational materials presented by a celebrity not only disclose their identity, but also have the celebrity appear in person (e.g., TV shows). To test the impact of the visual aspect, we present a picture of the celebrity at the beginning and the end for TG4. We hypothesize that the additional disclosure of the celebrity's face is likely to distract children from the actual content, we thus expect a weaker treatment effect for TG4 compared to TG3. Therefore, our fifth hypothesis is as follows:

Hypothesis 5 (H5)**.** *Compared to children who watched videos presented by a celebrity and knew the identity of the celebrity but did not see the celebrity's picture (TG3), children who saw the celebrity's picture (TG4) will show weaker improvements in non-cognitive and cognitive skills, less video engagement, and worse recall.*

To summarize, we expect the following ranking of the different groups after the intervention with respect to non-cognitive and cognitive skills: control group $<$ TG1 $<$ TG2, TG4 $<$ TG3. Given our hypotheses, it is a priori unclear whether children in TG2 or TG4 will perform better.

5 Empirical Framework

We test Hypotheses 1 and 2 using the following specification:

$$y_i = \alpha + \beta \text{TG}_i + \gamma r_i + \mathbf{x}_i' \boldsymbol{\delta} + \epsilon_i, \tag{1}$$

where y_i is an outcome variable of child i (or its parent). TG is an indicator for being in any treatment group. The coefficient of interest β is the intent-to-treat (ITT) effect of being in any treatment group. r_i indicates whether child i watched the videos in reversed order, $\mathbf{x}_i$ is a vector of control variables (i.e., gender, number of siblings, relative age, parental education, and mother's employment status), $\boldsymbol{\delta}$ is a coefficient vector, and ϵ_i is an idiosyncratic error term.

Equation (1) can be adjusted to allow for differences between the treatment groups:

$$y_i = \alpha + \beta_1 \text{TG1}_i + \beta_2 \text{TG2}_i + \beta_3 \text{TG3}_i + \beta_4 \text{TG4}_i + \gamma r_i + \mathbf{x}_i' \boldsymbol{\delta} + \epsilon_i. \tag{2}$$

We can then test for differences in the effectiveness of the different treatment groups. To test Hypothesis 3, we aggregate TG2, TG3, and TG4 into one group, adjust the Equation 2 accordingly, and test $\beta_1 = \beta_{2,3,4}$. Similarly, we can test Hypothesis 4 by aggregating TG3 and TG4 in Equation 2 and test $\beta_2 = \beta_{3,4}$. Finally, we can test Hypothesis 5 by using Equation 2 and test $\beta_3 = \beta_4$.

Although the ITT allows important insights into the overall effect of treatment assignment, we know which children actually engaged with the materials. Therefore, we also estimate two-stage least squares regressions to obtain estimates on the effectiveness of the program on those children that actually participated in the program (i.e., compliers). In particular, we assume that children who watched two or more videos (out of six) properly engaged with the materials, and thus define the indicator $\text{Watch}(2)_i$ as treatment. Our first stage thus estimates:[12]

$$\text{Watch}(2)_i = \alpha + \pi \text{TG}_i + \gamma r_i + \mathbf{x}_i' \boldsymbol{\delta} + \epsilon_i. \tag{3}$$

The estimated coefficient of π captures the predicted increase in the likelihood of watching two or more videos in response to being randomly assigned to any treatment group. The

[12] Although not reported the results do not change if we instead use other measures of participation such as (i) parents watched two videos or more, (ii) more than 50% of total movie length watched, and (iii) more than two correctly answered recall questions.

second stage is then:

$$y_i = \alpha + \beta \widehat{\text{Watch}}(2)_i + \gamma r_i + \mathbf{x}_i' \boldsymbol{\delta} + \epsilon_i. \tag{4}$$

where $\widehat{\text{Watch}}(2)_i$ holds the predicted values from the first stage.

Because we test the effectiveness of our intervention on a wide range of outcomes, we will report p-values adjusted for multiple hypothesis testing at the table level based on the approach of Romano and Wolf (2005) using 4,999 bootstrap replications. Finally, all regressions use robust standard errors to allow for possible heteroscedasticity.

6 Results

6.1 Participation and Engagement

In a first step, we want to make sure that children and parents in our treatment groups actually engaged with the materials. Using detailed watch statistics from our video platform and weekly surveys asking about self-reported participation, we obtain measures for each video and treatment group on their participation. Table 2 suggests that on average children have watched more than 3 (out of 6) videos. They have answered more than 4 questions correctly,[13] and they wrote open-end comments for on average 2 movies.[14] On average, they finished around 90% of the movies. Finally, parents watched around 3 movies and 2 of them with their child.

Figures A5, A6, A7, and A8 in the Appendix present watch statistics for each movie by treatment group and grade level. While the overall watch behavior does not differ between the 6 movies, the engagement with the materials measured by children's answers and comments dropped in the middle of the intervention. This pattern can be explained by children engaging more in the first two weeks, but due to our reversed order, a decline in week 5 and 6 is being outweighed by presenting the movies in reversed order to half of the treated children. However, in our main analysis suggests no difference in treatment effectiveness by the order in which children watched the videos.

[13] There were in total 12 questions. Children who did not respond to the survey are indicated with a 0. These question typically take the following form: (i) "What were the two points the teacher mentioned at the beginning of the talk?" or (ii) "In the story, what did the protagonist know about ...?" with three options to choose from.

[14] We ask for comments in the following way: "If you have any thoughts or feelings after hearing the story, please write them here." While most comments are quite short and simply state whether children liked the movie or not, some children engaged with the topic and reflected on the movie in a differentiated way.

Table 3: Overall ITT Effect

	Well-being (1)	Self-esteem (2)	Perspective-taking (3)	School attitude (4)	Moral behavior (5)	Self-control (6)
= 1 if treated	−0.156*	−0.090	0.006	−0.016	0.064	0.084
	(0.087)	(0.088)	(0.087)	(0.090)	(0.083)	(0.093)
	[.343]	[.782]	[.979]	[.979]	[.797]	[.797]
Reverse dummy	✓	✓	✓	✓	✓	✓
Child char.	✓	✓	✓	✓	✓	✓
Parent char.	✓	✓	✓	✓	✓	✓
Observations	836	817	824	809	828	810

Notes: This table shows results from OLS regressions of the outcome variables on treatment assignment and a set of control variables. Robust standard errors are given in parentheses. Step-down adjusted p-values robust to multiple hypothesis testing based on Romano and Wolf (2005) are given in brackets. *, **, and *** denote significance at the 10%-, 5%-, and 1%-level, respectively.

6.2 Short-Term Impacts on Children

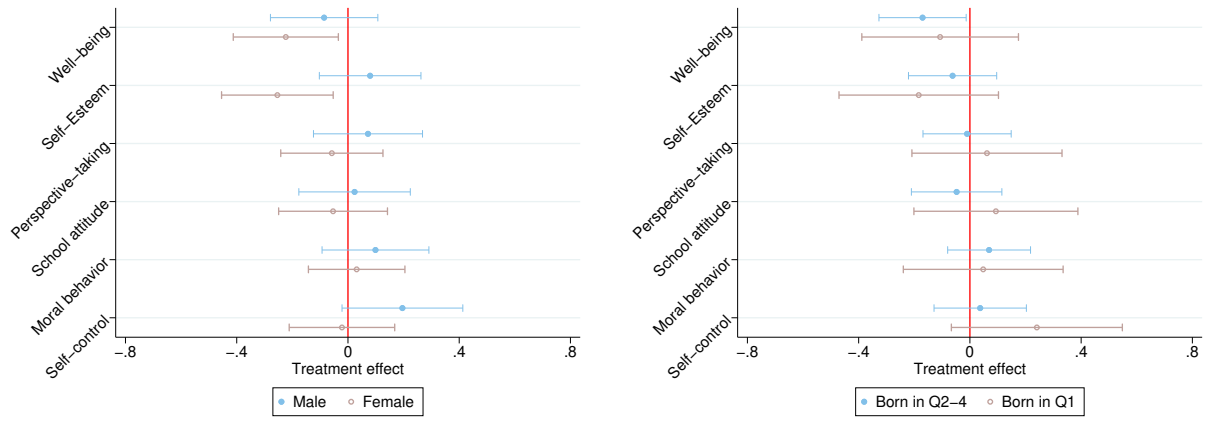
Our main analysis in the short run tests Hypotheses 1 and 2. In particular, we evaluate the effect of our intervention on a set of behavioral outcomes described in Subsection 3.5. We first present results from estimating Equation (1) in Table 3.

Our results suggest no statistically significant effect on any behavioral outcome measure in the short run besides a small negative effect on well-being. Although this coefficient does not survive a correction for multiple hypothesis testing, we should not ignore this finding. It could be that, at least in the short run, our information treatment and raising awareness of the problems that children in Japan have (e.g., the negative consequences of bullying) might raise concerns and thus decrease overall well-being.

In Table A3 in the Appendix, we present estimates from two-stage least squares regressions. These results suggest no significant difference between the ITTs and effects on compliers. Therefore, for the rest of the paper, we report ITT effects unless otherwise mentioned.

In the next step, we allow the intervention to have a differential effect on behavioral outcomes based on observable characteristics. In particular, we test whether the gender or relative age affects the impact of the intervention in Figure 5.

It seems that the small negative effect on well-being and self-esteem is purely driven by females (in red), whereas we find a positive effect on self-control for males (in blue).



(a) Sex

(b) Relative age

Figure 5: Heterogeneity by Sex and Relative Age

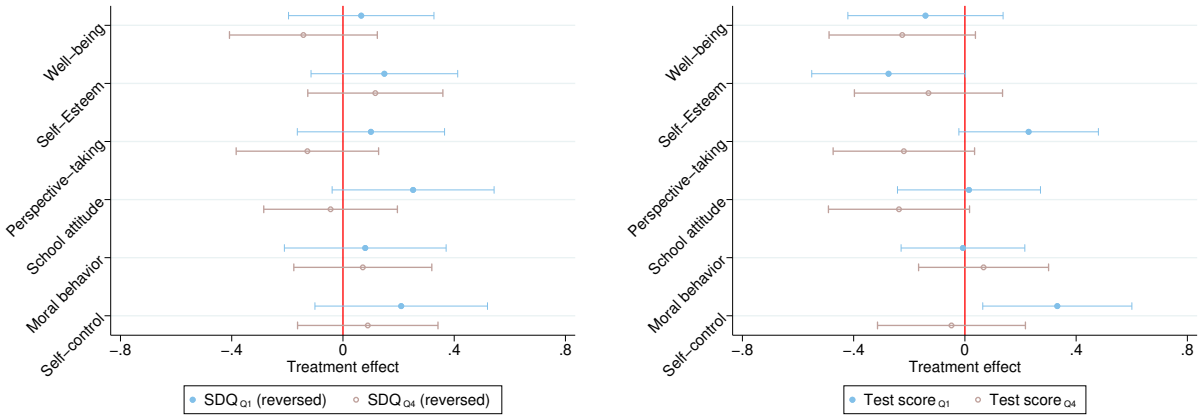
Notes: This graph shows results from OLS regressions of the outcome variables on treatment assignment and a set of control variables. 90% confidence intervals are based on robust standard errors.

However, this effect is not statistically significant.[15] Further, we find no clear difference between children being born in the first quarter of the year and thus being relatively young in their class (in red) and children being born in the second to fourth quarter and thus being relatively old in their class (in blue).

In the next step, we evaluate the effectiveness of the intervention for children that are at highest risk. It is possible that we do not find an overall effect because our intervention targets a representative sample in a developed country, which, despite increasing problems in child mental health issues, have a low likelihood of suffering from these problems. Educational interventions typically focus on children in need (e.g., Angrist et al., 2002; Banerjee et al., 2007). This approach is also reasonable from a policy perspective - interventions cost resources and may be a burden on the family and other involved institutions. One feature of our RCT is that the design allows us to compare the impact on at-risk children as well as on children without behavioral problems. We identify these children along several dimensions. First, we observe behavioral problems as measured by the Strength and Difficulties Questionnaire (SDQ) scale in the baseline survey. Second, we identify low achievers based on a cognitive test included in the baseline survey. Figure 6 presents results along these heterogeneity dimensions.

Although none of the presented coefficients is statistically significant, there is a tendency that children with behavioral problems (in blue) gain from the intervention compared to chil-

[15]It could be that females react stronger than males to the videos as they are presented by a female narrator. Unfortunately, we cannot test for this and thus it is up to future modifications of this interventions with a male narrator.



(a) Baseline SDQ

(b) Baseline test scores

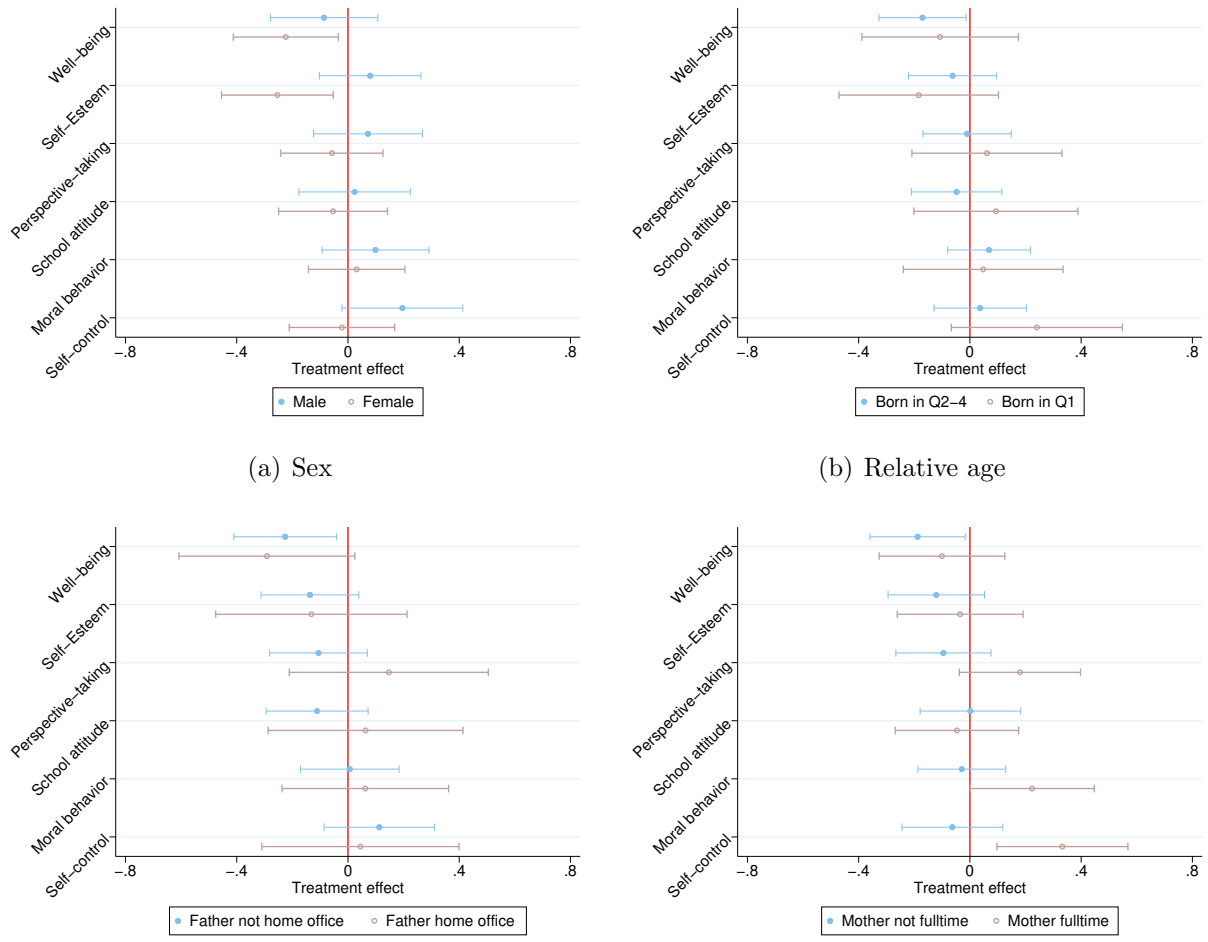
Figure 6: Heterogeneity by SDQ and Test Scores

Notes: This graph shows results from OLS regressions of the outcome variables on treatment assignment and a set of control variables. 90% confidence intervals are based on robust standard errors.

dren without behavioral problems (in red). However, there is strong evidence that children who performed lowest in the cognitive tests gain the most from the intervention in terms of self-control and perspective-taking, even thogh the latter coefficient is not statistically significant. As self-control skills, such as the ability to concentrate and see tasks to the end, are included in the SDQ scale and are also directly related to academic achievement, our intervention seems to positively affect the aspects the children in the lowest quartile for both measures likely struggle with.

Next, focusing on the heterogeneous impact of our intervention based on household socioeconomic status, we present the results in Figure 7. The variables of interest are parental educational attainment (university degree for mother and father), father's home office opportunity, and mother's employment status (working full time or not; not category includes both part time work and stay at home mothers). While many of these variables are correlated, we believe examining them separately allows us to better uncover the mechanisms behind the results and further identify which children can benefit from our intervention.

Our results suggest a consistent statistically significant improvement in behavioral outcomes, specifically moral behavior and self-control, for children from high SES families. These results hold for all SES definitions, but are most pronounced for children with highly educated mothers and mothers working full time. One potential mechanism behind these results could be low baseline time spent with the child for busy mothers. Another possible mechanism may be deeper engagement with the material and corresponding discussion with the child for highly educated parents. The latter interpretation is supported by the



(a) Sex

(b) Relative age

(c) Father home office

(d) Mother empl. status

Figure 7: Heterogeneity by Parental Characteristics

Notes: This graph shows results from OLS regressions of the outcome variables on treatment assignment and a set of control variables. 90% confidence intervals are based on robust standard errors.

decreased well-being and self-esteem in children with mothers without college education. It is the parent who opts in to the program. If the child is forced to participate and does not enjoy the program, the intervention could potentially backfire, resulting in lower well-being.

While we cannot precisely identify which parent participated which week, given the Japanese context, we assume it is predominantly mothers. Our baseline survey suggests that mothers spend most of their time with their children while fathers engage less with them (see Figure A9 in the Appendix).

To summarize our results related to Hypotheses 1 and 2, we find no effect of our intervention on any of the measured behavioral outcomes in the representative sample. However, this is an expected result in a scaled-up sample. As children with no issues stand to benefit

little from the intervention, we test the impact on at-risk children based on pre-intervention behavioral issues and low test scores. We find a positive impact on self-control, a predictor of good behavior and academic achievement, and partially perspective-taking. Further examining which children may benefit from the program based on their characteristics and household socioeconomic status, we find persistent improvements for children from high-SES households and mothers who spend more time at home, predominantly in moral behavior and self-control. This intervention seems to require a parent (often the mother) that has the cognitive capacity to understand the materials and is able to properly convey the message. This could suggest that our materials are too complicated and must be adjusted for low-SES households. While further analysis is needed to clarify the mechanism behind these results and how children attributes interact with our intervention, it is clear that a general implementation of such a program may not lead to the intended goals, while a targeted one could help children in need.

7 Celebrity Endorsement and Edutainment

In this section, we investigate the effectiveness of the narrator and their identity reveal as stated in Hypotheses 3 to 5. In Table 4, we first present estimates on participation and engagement with the materials, as this can help to explain any differential effect. Then, in Table 5, we present estimates with respect to behavioral outcomes.

First, Hypothesis 3 states that children who watched videos presented by a celebrity (TG2 & TG3 & TG4) will show greater improvements in non-cognitive and cognitive skills, greater video engagement, and better recall compared to children who watched videos presented by a representative teacher (TG1). Starting with the latter, results from an analysis restricted to the treated children in Table 4 suggest that we cannot find any statistically significant difference between TG1 and TG2–4 in terms of participation and engagement. In particular, there is no statistically significant difference between TG1 and TG2. However, results in Table 5 suggest a statistically significant improvement in terms of self-esteem ($p = 0.034$). This, however, only compensates the seemingly negative impact for TG1 and thus we conclude that we can reject Hypothesis 3.

Second, Hypothesis 4 states that children who knew about the celebrity's identity (TG3 & TG4) will show greater improvements in non-cognitive and cognitive skills, greater video engagement, and better recall compared to children who watched videos presented by a celebrity but did not know the identity of the celebrity (TG2). Results from Table 4 suggest strong positive effects on participation and engagement. These effects are statistically significant for the number of videos watched by a child ($p = 0.001$), the number of correctly

Table 4: ITT Effect Celebrity Voice vs. Identity & Picture

	Child				Parent	
	# videos watched	# correct answers	# comments	Percentage finished	# videos watched	# videos watched w/ child
	(1)	(2)	(3)	(4)	(5)	(6)
= 1 if celebrity (TG2)	−0.095	−0.381	−0.216	−0.005	0.119	0.009
	(0.152)	(0.288)	(0.164)	(0.020)	(0.215)	(0.213)
	[.868]	[.632]	[.867]	[.868]	[.873]	[.873]
= 1 if celebrity (TG3)	0.610***	0.977***	0.494**	0.038*	0.439*	0.202
	(0.169)	(0.335)	(0.201)	(0.021)	(0.255)	(0.253)
	[.006]	[.033]	[.043]	[.273]	[.273]	[.614]
= 1 if celebrity (TG4)	0.246	0.176	0.022	0.009	0.448*	0.166
	(0.177)	(0.332)	(0.206)	(0.023)	(0.256)	(0.269)
	[.472]	[.964]	[.964]	[.964]	[.346]	[.964]
Reverse dummy	✓	✓	✓	✓	✓	✓
Child char.	✓	✓	✓	✓	✓	✓
Parent char.	✓	✓	✓	✓	✓	✓
TG1 = TG2–4 (p-value)	.180	.606	.601	.736	.170	.774
TG2 = TG3–4 (p-value)	.001	.002	.016	.090	.136	.444
TG3 = TG4 (p-value)	.070	.039	.048	.233	.976	.906
Observations	644	644	644	650	546	546

Notes: This table shows results from OLS regressions of the outcome variables on treatment assignment (TG1 vs. TG2 vs. TG3 vs. TG4) and a set of control variables. TG1: teacher voice. TG2: celebrity voice. TG3: celebrity voice and identity. TG4: celebrity voice, identity, and picture. Robust standard errors are given in parentheses. Step-down adjusted p-values robust to multiple hypothesis testing based on Romano and Wolf (2005) are given in brackets. *, **, and *** denote significance at the 10%-, 5%-, and 1%-level, respectively.

answered recall questions ($p = 0.002$), the number of comments a child wrote in the weekly post-video surveys regarding the videos ($p = 0.016$), and the percentage of the watch time the child saw ($p = 0.090$). In particular, children in TG3 compared to TG1 watched on average 0.6 more videos, answered around 1 more question correctly, and wrote 0.5 comments more. They also finished around 3.8 percentage points more of the videos, and watched 0.4 more videos with their parent. While these effects are still mostly positive for children in TG4, they are only statistically significantly different from TG1 for the number of videos watched with their parent. This, however, did not translate in any statistically significant improvements in behavioral outcomes (see Table 5), and thus we can accept this Hypothesis 4 for participation and engagement but not for behavioral outcomes.

Third, Hypothesis 5 states that children who saw the celebrity's picture (TG4) will show weaker improvements in non-cognitive and cognitive skills, less video engagement, and worse recall compared to children who watched videos presented by a celebrity and knew the identity of the celebrity but did not see the celebrity's picture (TG3). Results from Table 4 suggest that we can accept this hypothesis with respect to participation and engagement. It seems that TG4 has a statistically significantly lower number of videos watched by a child ($p = 0.070$), the number of correctly answered recall questions ($p = 0.039$), and the number of comments a child wrote in the weekly post-video surveys regarding the videos ($p = 0.048$) compared to TG3. Compared to TG1, including the celebrity's picture negates almost all of the gains realized by revealing the celebrity's identity. However, this does not translate in lower behavioral outcomes (see Table 5) and thus we must reject Hypothesis 5 for behavioral outcomes.

In summary, we tested Hypotheses 3 to 5 for participation, engagement and recall outcomes, comparing teacher's voice to celebrity's voice, celebrity's voice with identity revealed, and celebrity's voice with identity and picture revealed. We find no difference based on voice only, indicating that while celebrity narrator mitigated some negative effects of teacher narration on short-term behavioral outcomes, the mechanism behind those effects is not increased engagement. We also find that revealing a celebrity's identity strongly increased participation, engagement, and recall outcomes among most measured dimensions. However, this result is almost entirely driven by the treatment arm disclosing identity but not revealing the celebrity's face (TG3). This finding suggests that attention-grabbing strategies targeting children need to be carefully considered, as including more information about the narrator can have unintended negative effects. Visual elements can distract rather than enhance learning. Nevertheless, none of the heightened participation and engagement translates to improvements in behavioral outcomes measured in the short-term.

8 Conclusion

The child mental health crisis has been described as "the defining public health crisis of our time," having a long-lasting impact on well-being and economic circumstances beyond the affected individual (Currie, 2025), with Japan similarly experiencing rapidly increasing cases of bullying, absenteeism, and suicide among adolescents (see Figure A1 in the Appendix).

This paper describes short-run findings from a low-cost, easily scalable intervention, conducted in 2023, using online educational tools in a home-based environment to address social issues that children may be experiencing, targeting Japanese children between the ages of 9 to 14. We employ a dual approach: raising awareness of social issues in parents, equipping

Table 5: ITT Effect by Treatment Arms

	Well-being (1)	Self-esteem (2)	Perspective-taking (3)	School attitude (4)	Moral behavior (5)	Self-control (6)
= 1 if teacher (TG1)	−0.228**	−0.215**	−0.025	−0.069	0.061	0.078
	(0.106)	(0.108)	(0.108)	(0.107)	(0.100)	(0.119)
	[.177]	[.177]	[.909]	[.909]	[.909]	[.909]
= 1 if celebrity (TG2)	−0.146	−0.031	0.036	−0.055	0.104	0.122
	(0.105)	(0.108)	(0.103)	(0.115)	(0.101)	(0.114)
	[.626]	[.940]	[.940]	[.940]	[.765]	[.765]
= 1 if celebrity (TG3)	−0.052	0.015	0.000	0.069	0.040	0.036
	(0.124)	(0.117)	(0.123)	(0.124)	(0.112)	(0.132)
	[.991]	[.991]	[.999]	[.989]	[.991]	[.991]
= 1 if celebrity (TG4)	−0.150	−0.065	0.015	0.052	0.030	0.078
	(0.134)	(0.121)	(0.128)	(0.128)	(0.110)	(0.122)
	[.708]	[.965]	[.965]	[.965]	[.965]	[.965]
Reverse dummy	✓	✓	✓	✓	✓	✓
Child char.	✓	✓	✓	✓	✓	✓
Parent char.	✓	✓	✓	✓	✓	✓
TG1 = TG2–4 (p-value)	.246	.034	.620	.395	.940	.933
TG2 = TG3–4 (p-value)	.651	.939	.777	.277	.451	.537
TG3 = TG4 (p-value)	.509	.531	.912	.904	.934	.760
Observations	836	817	824	809	828	810

Notes: This table shows results from OLS regressions of the outcome variable on treatment assignment (TG1 vs. TG2 vs. TG3 vs. TG4) and a set of control variables. TG1: teacher voice. TG2: celebrity voice. TG3: celebrity voice and identity. TG4: celebrity voice, identity, and picture. Robust standard errors are given in parentheses. Step-down adjusted p-values robust to multiple hypothesis testing based on Romano and Wolf (2005) are given in brackets. *, **, and *** denote significance at the 10%-, 5%-, and 1%-level, respectively.

them with information necessary to discuss the issues with their children, and providing easily approachable educational videos to children, showcasing a related dilemma. The weekly videos are based on moral education textbooks with text and illustrations, narrated by either a female teacher or a female celebrity, with varying degrees of identity revealed.

Our results suggest that the intervention has little impact on short-term behavioral outcomes in the nationally representative sample, but has a positive effect on children identified as at-risk by baseline behavioral issues (i.e., SDQ), low academic scores, or family background. In particular, we find persistent improvements for children from high-SES households and mothers who spend more time at home, predominantly in moral behavior and

self-control. We further examined the role of narrator type and identity disclosure. Celebrity narration helped mitigate negative effects observed under teacher narration, particularly for well-being and self-esteem, but otherwise had limited direct impact on behavioral outcomes. The most robust positive effects were observed in participation, engagement, and recall, particularly when the celebrity's identity was disclosed without showing their picture. Adding a picture negated these gains, indicating that added visual may distract from substantive content. These findings highlight both the potential and the shortcomings of edutainment strategies in education: professional narration can improve engagement and avoid harm, but excessive emphasis on the celebrity may crowd out the intended educational message.

Taken together, our findings demonstrate that scalable, low-cost online interventions can be implemented successfully at the national level and can reach families directly without relying on schools. Yet, their short-term effects on behavioral outcomes are limited and uneven, with greater benefits for at-risk children. The long-term follow-up survey data, not yet available at the time of writing this manuscript, will help us assess whether the observed engagement gains and heterogeneous short-term impacts translate into sustained improvements in cognitive and non-cognitive skills. More broadly, our results contribute to the literature on parental involvement, the use of ICT in education, and the design of scalable RCTs, while offering policy lessons on when and for whom such interventions are most effective.

References

Akabayashi, H., Ruberg, T., Shikishima, C., and Yamashita, J. (2023). Education-oriented and care-oriented preschools: Implications on child development. *Labour Economics*, 84:102410.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558.

Angrist, N., Bergman, P., and Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*, 6(7):941–950.

Angrist, N., Kabay, S., Karlan, D., Lau, L., and Wong, K. (2025). Human capital at home: Evidence from a randomized evaluation in the Philippines. NBER Working Paper No. 33574, Cambridge, MA.

Arseneault, L. (2018). Annual research review: The persistent and pervasive impact of being bullied in childhood and adolescence: Implications for policy and practice. *Journal of Child Psychology and Psychiatry*, 59(4):405–421.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E. (2014). Getting parents involved: A field experiment in deprived schools. *Review of Economic Studies*, 81(1):57–83.

Awata, S., Bech, P., Yoshida, S., Hirai, M., Suzuki, S., Yamashita, M., Ohara, A., Hinokio, Y., Matsuoka, H., and Oka, Y. (2007). Reliability and validity of the Japanese version of the world health organization-five well-being index in the context of detecting depression in diabetic patients. *Psychiatry and Clinical Neurosciences*, 61(1):112–119.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3):1235–1264.

Bergkvist, L. and Zhou, K. Q. (2016). Celebrity endorsements: A literature review and research agenda. *International Journal of Advertising*, 35(4):642–663.

Bergman, P. (2021). Parent-child information frictions and human capital investment: Evidence from a field experiment. *Journal of Political Economy*, 129(1):286–322.

Bergman, P. and Chan, E. W. (2021). Leveraging parents through low-cost technology the impact of high-frequency information on student achievement. *Journal of Human Resources*, 56(1):125–158.

Berlinski, S., Busso, M., Dinkelman, T., and Martínez, C. (2022). Reducing parent-school information gaps and improving education outcomes: Evidence from high-frequency text messages. *Journal of Human Resources*, 60(1):1439–1464.

Bjorvatn, K., Cappelen, A. W., Sekei, L. H., Sørensen, E. Ø., and Tungodden, B. (2020). Teaching through television: Experimental evidence on entrepreneurship education in Tanzania. *Management Science*, 66(6):2308–2325.

Cacault, M. P., Hildebrand, C., Laurent-Lucchetti, J., and Pellizzari, M. (2021). Distance learning in higher education: Evidence from a randomized experiment. *Journal of the European Economic Association*, 19(4):2322–2372.

Child Development Science Research (2024). Ministry of Education, Culture, Sports, Science and Technology-commissioned research and study on analysis of the causes of absenteeism (in Japanese). https://kohatsu.org/20240325research-report/, retrieved May 22, 2024.

Children and Families Agency (2024). Anti-bullying measures from the Children and Families Agency (in Japanese). https://www.cfa.go.jp/policies/ijime-boushi, retrieved May 22, 2024.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Currie, J. (2025). Presidential address: Investing in children to address the child mental health crisis. *American Economic Review*, 115(5):1369–1399.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113.

Doss, C., Fahle, E. M., Loeb, S., and York, B. N. (2019). More than just a nudge supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources*, 54(3):567–603.

Economic and Social Research Institute (2024). Trends in international discussions on well-being "beyond GDP" and Japan's efforts (in Japanese). https://www.esri.cao.go.jp/jp/esri/archive/e_rnote/e_rnote090/e_rnote082.pdf, retrieved May 22, 2024.

Figlio, D., Rush, M., and Yin, L. (2013). Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, 31(4):763–784.

Fiorini, M. and Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, 32(4):787–836.

Gentzkow, M. and Shapiro, J. M. (2008). Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study. *Quarterly Journal of Economics*, 123(1):279–323.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5):581–586.

Hasegawa, M., Horiuchi, Y., Suzuki, K., Sado, M., and Sakamoto, A. (2009). Jidou-you tajigen kyoukansei shakudo no shinraisei: Datousei no kentou (in Japanese). *Paasonariti Kenkyuu (Japanese Journal of Personality)*, 17(3):307–310.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123(4):1373–1414.

Hertog, E. and Zhou, M. (2021). Japanese adolescents' time use: The role of household income and parental education. *Demographic Research*, 44:225.

Kearney, M. S. and Levine, P. B. (2019). Early childhood education by television: Lessons from Sesame Street. *American Economic Journal: Applied Economics*, 11(1):318–350.

Kendall, P. C. and Wilcox, L. E. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology*, 47(6):1020.

Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment. *Child Development*, 61(4):1081–1100.

Ladd, G. W. and Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development*, 58(5):1168–1189.

Maeno, T. (2017). Four factors of happiness as design parameters of a product/service. In Fukuda, S., editor, *Emotional engineering*, volume 5, pages 183–189. Springer.

Mayer, S. E., Kalil, A., Oreopoulos, P., and Gallegos, S. (2019). Using behavioral insights to increase parental engagement: The parents and children together intervention. *Journal of Human Resources*, 54(4):900–925.

MEXT (2021). 高等学校等への進学率［推移］. https://www.mext.go.jp/content/20211207-mxt_koukou02-000019354_01.pdf, retrieved September 2, 2025.

MEXT (2024). Regarding Results of Survey on Problematic Behavior, Chronic School Absences and Other Issues Relating to Student Guidance and Counseling Academic Year R5 (in Japanese). https://www.mext.go.jp/content/20241031-mxt_jidou02-100002753_1_2.pdf, retrieved March 7, 2025.

Mimura, C. and Griffiths, P. (2007). A Japanese version of the Rosenberg Self-Esteem Scale: Translation and equivalence assessment. *Journal of Psychosomatic Research*, 62(5):589–594.

Ministry of Education, Culture, Sports, Science and Technology (2019). Changes in the definition of bullying (in Japanese). https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2019/06/26/1400030_003.pdf, retrieved May 22, 2024.

Nagaoka, D., Uno, A., Usami, S., Tanaka, R., Minami, R., Sawai, Y., Okuma, A., Yamasaki, S., Miyashita, M., Nihsida, A., Kasaia, K., and Ando, S. (2024). Identify adolescents' help-seeking intention on suicide through self- and caregiver's assessments of psychobehavioral problems: Deep clustering of the Tokyo TEEN Cohort study. *The Lancet Regional Health – Western Pacific*, 43.

Nakamura, R., Yamashita, J., Akabayashi, H., Tamura, T., and Zhou, Y. (2020). A comparative analysis of children's time use and educational achievement: Assessing evidence from China, Japan and the United States. *Chinese Journal of Sociology*, 6(2):257–285.

National Institute for Educational Policy Research (2016). OECD Programme for International Student Assessment: Summary of international results from the 2015 survey (in Japanese). https://www.nier.go.jp/kokusai/pisa/pdf/2015/03_result.pdf, retrieved May 22, 2024.

nippon.com (2023). Pandemic disruption leads to record school absenteeism in Japan in 2022 (in Japanese). https://www.nippon.com/ja/japan-data/h01812/, retrieved May 22, 2024.

Nownes, A. J. (2012). An experimental investigation of the effects of celebrity support for political parties in the United States. *American Politics Research*, 40(3):476–500.

Nucci, L., Smetana, J., Araki, N., Nakaue, M., and Comer, J. (2014). Japanese adolescents' disclosure and information management with parents. *Child Development*, 85(3):901–907.

Oshio, T., Umeda, M., and Kawakami, N. (2013). Childhood adversity and adulthood subjective well-being: Evidence from Japan. *Journal of Happiness Studies*, 14:843–860.

Otsui, K., H Otta, M., Takeshima, K., Matsumi, J., and Ladd, G. W. (2006). Development of a Japanese version of the School Liking and Avoidance Questionnaire (in Japanese). *The 47th Annual Meeting of the Japanese Society of Social Psychology*, pages 796–797.

Phillips, M. and Reber, S. (2022). Does virtual advising increase college enrollment? Evidence from a random-assignment college access field experiment. *American Economic Journal: Economic Policy*, 14(3):198–234.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press, New Jersey.

Saito, S. and Ikeda, S. (2016). Suku-ru enge-jimento to jidenteki kioki toshite no doutokusei no kanren - doutoku kyouiku he no katsuyou wo mezashite (in Japanese). *Gakkou Kyouiku Soudan Kenkyuu (School Education Consultation Research)*, 26:26–34.

Schartel Dunn, S. and Nisbett, G. (2020). If Childish Gambino cares, I care: Celebrity endorsements and psychological reactance to social marketing messages. *Social Marketing Quarterly*, 26(2):80–92.

Takizawa, R., Maughan, B., and Arseneault, L. (2014). Adult health outcomes of childhood bullying victimization: Evidence from a five-decade longitudinal British birth cohort. *American Journal of Psychiatry*, 171(7):777–784.

The Nippon Foundation (2019). The Nippon Foundation's "Survey of the attitudes of 18-year-olds," vol. 20. Theme: "Attitudes toward country and society" (9 countries surveyed) (in Japanese). https://www.nippon-foundation.or.jp/who/news/pr/2019/20191130-38555.html, retrieved May 22, 2024.

Topp, C. W., Østergaard, S. D., Søndergaard, S., and Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3):167–176.

Tsukamoto, S. (1997). Jidou hyoutei jiko tousei shakudo no kousei oyobi shinraisei (in Japanese). *Kyouiku Keiei Kenkyuu (Research on Educational Management)*, 3(3):76–86.

Tsukamoto, S. (2010). Jikoseigyo ga chuugakusei no gakkou tekiyou ni oyobosu eikyou - jikoseigyo koudou shakudo ni yoru kentou (in Japanese). *Gakkou Hoken Kenkyuu (Japanese Journal of School Health)*, 52(2):143–150.

UNICEF (2021). A world affecting children: What shapes child well-being in developed countries? (in Japanese). https://www.unicef.or.jp/library/pdf/labo_rc16j.pdf, retrieved May 22, 2024.

Veer, E., Becirovic, I., and Martin, B. A. (2010). If Kate voted conservative, would you? The role of celebrity endorsements in political party advertising. *European Journal of Marketing*, 44(3/4):436–450.

WHO (1998). Wellbeing measures in primary health care/The DEPCARE project: Report on a WHO meeting. WHO Regional Office for Europe, Copenhagen.

Yamaguchi, K., Shikishima, C., Hoshino, T., Shigemasu, K., and Akabayashi, H. (2019). Vertical scaling of academic ability tests for elementary school first year students through junior high school third year students (in Japanese). *Japanese Journal of Psychology*, 90(4):408–418.

York, B. N., Loeb, S., and Doss, C. (2019). One step at a time the effects of an early literacy text-messaging program for parents of preschoolers. *Journal of Human Resources*, 54(3):537–566.

Zumbuehl, M., Dohmen, T., and Pfann, G. (2021). Parental involvement and the intergenerational transmission of economic preferences, attitudes and personality traits. *Economic Journal*, 131(638):2642–2670.

A Appendix – Additional Tables and Figures

Table A1: LPM of Program Participation

	Program participation		
	(1)	(2)	(3)
= 1 if female	0.003		0.009
	(0.013)		(0.015)
# of siblings	0.039***		0.045***
	(0.007)		(0.008)
= 1 if young	−0.008		−0.014
	(0.016)		(0.017)
= 1 if father college degree		0.013	0.013
		(0.016)	(0.016)
= 1 if mother college degree		0.019	0.019
		(0.016)	(0.016)
= 1 if mother fulltime empl.		−0.030**	−0.027*
		(0.015)	(0.015)
Observations	4310	3535	3535

Notes: This table shows linear probability models of program participation vs. invitation. The math and language scores are standardized by grade. *, **, and *** denote significance at the 10%-, 5%-, and 1%-level, respectively.

Table A2: Covariate Balancing Excluding Attriters

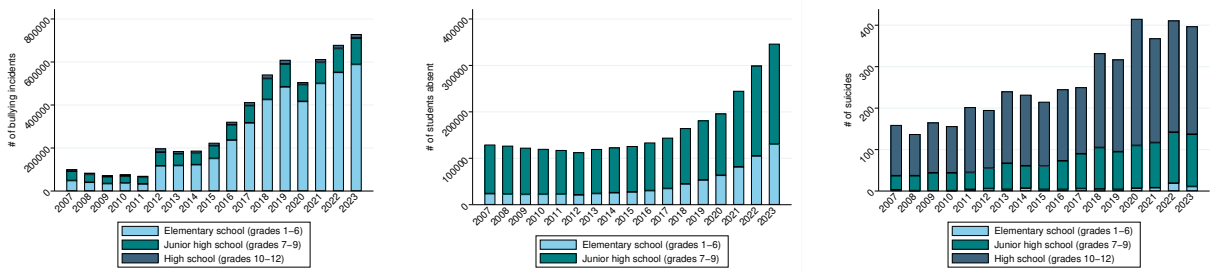
	Control group	Treatment group				
		TG1	TG2	TG3	TG4	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
= 1 if female	0.50	0.53	0.49	0.50	0.48	0.91
# of siblings	1.28	1.23	1.24	1.16	1.21	0.74
= 1 if young	0.23	0.20	0.22	0.29	0.25	0.36
= 1 if father college degree	0.59	0.59	0.64	0.58	0.65	0.66
= 1 if mother college degree	0.43	0.39	0.40	0.42	0.49	0.54
= 1 if mother fulltime empl.	0.38	0.33	0.42	0.34	0.38	0.38
Observations	252	213	221	137	126	

Notes: This table shows mean values for child and parental characteristics measured pre-treatment. p-values are based on F-tests of joint significance. The math and language scores are standardized by grade. Data is restricted to nonmissing observations in the well-being measure.

Table A3: Effect for Compliers

	Well-being (1)	Self-esteem (2)	Perspective-taking (3)	School attitude (4)	Moral behavior (5)	Self-control (6)
= 1 if participated	−0.175*	−0.100	0.007	−0.017	0.072	0.094
	(0.097)	(0.098)	(0.097)	(0.100)	(0.092)	(0.104)
	[.341]	[.781]	[.979]	[.979]	[.797]	[.797]
Reverse dummy	✓	✓	✓	✓	✓	✓
Child char.	✓	✓	✓	✓	✓	✓
Parent char.	✓	✓	✓	✓	✓	✓
F-stat	2520	2493	2556	2556	2549	2451
Observations	836	817	824	809	828	810

Notes: This table shows results from 2SLS regressions of the outcome variables on program participation and a set of control variables. Program participation (measured by a dummy indicating more than two movies watched) is instrumented by treatment assignment. Robust standard errors are given in parentheses. Step-down adjusted p-values robust to multiple hypothesis testing based on Romano and Wolf (2005) are given in brackets. *, **, and *** denote significance at the 10%-, 5%-, and 1%-level, respectively.



(a) Bullying (b) Absenteeism (c) Suicide

Figure A1: Prevalence of Social Issues in Japan

Source: Survey on Problematic Behavior, Chronic School Absences, and Other Issues Relating to Student Guidance and Counseling. *Notes*: Panel (a) shows the development of recognized cases of bullying across schools in Japan. Panel (b) shows the development of recognized cases of absenteeism (30 days or more) across schools in Japan. Data for high school students are not available. Panel (c) shows the development of suicides across schools in Japan.

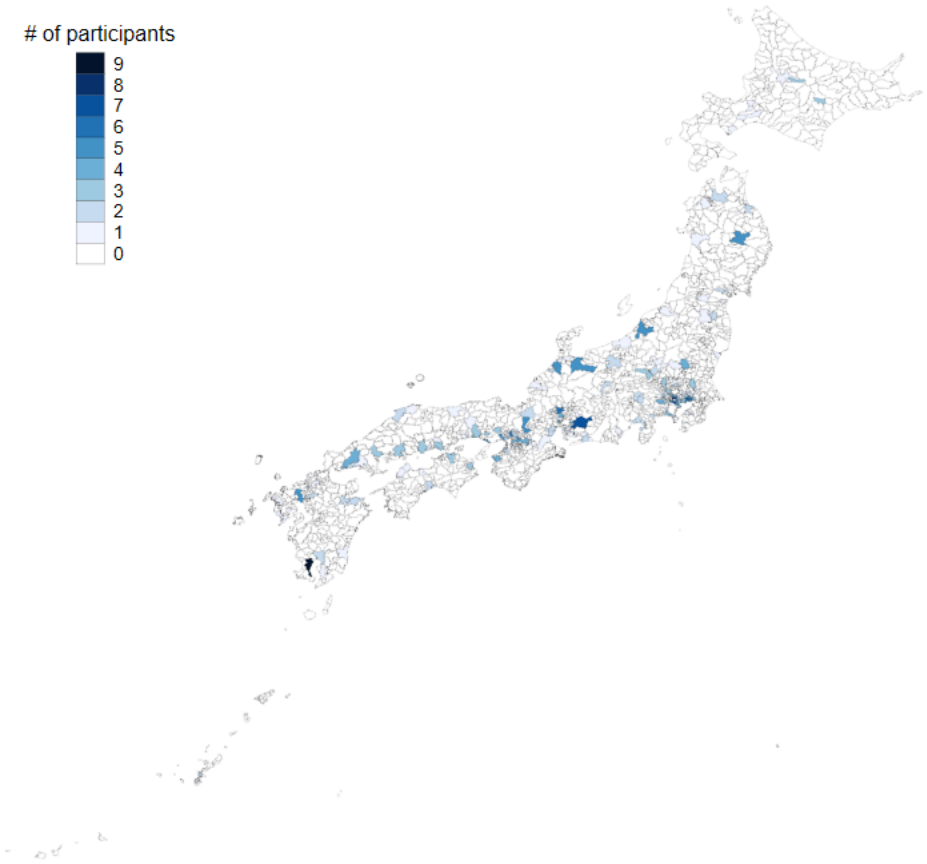


Figure A2: Regional Distribution of Final Participants

Notes: The figure shows the regional distribution of the final participants across Japanese municipalities.

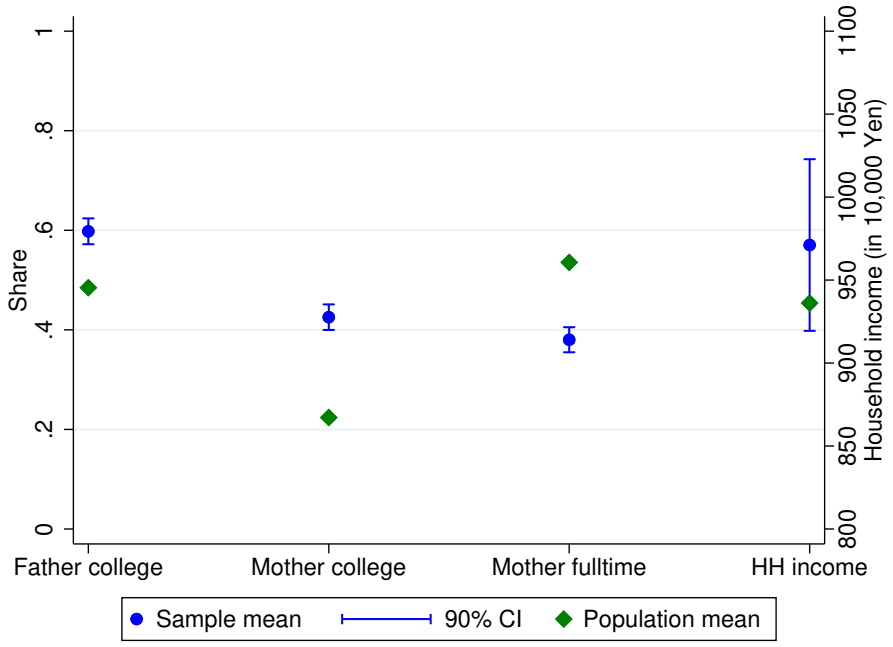


Figure A3: Representativeness of Participants

Notes: This graph compares sample means and 90% confidence intervals (in blue) with representative population means based on the JHSP/KHPS (in green) for selected variables.

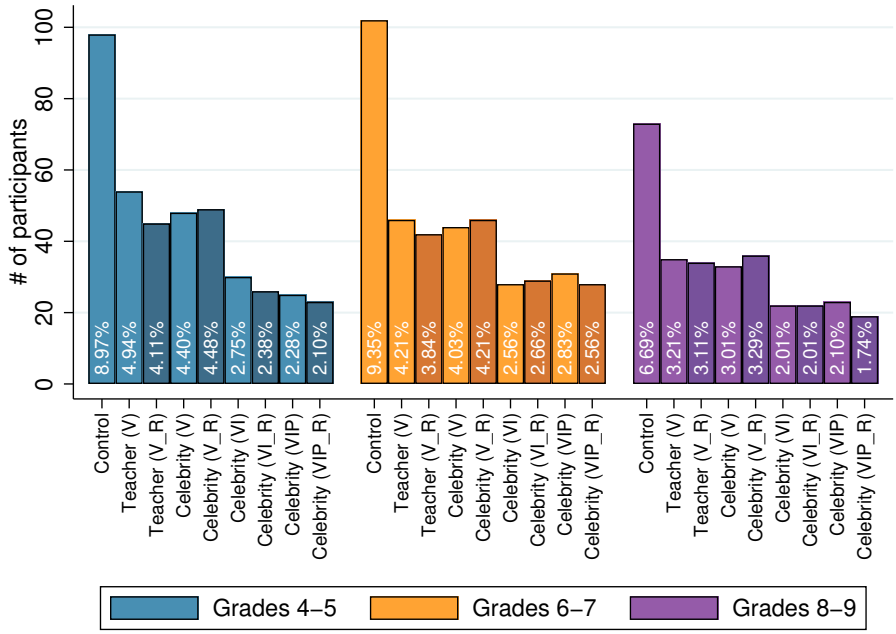
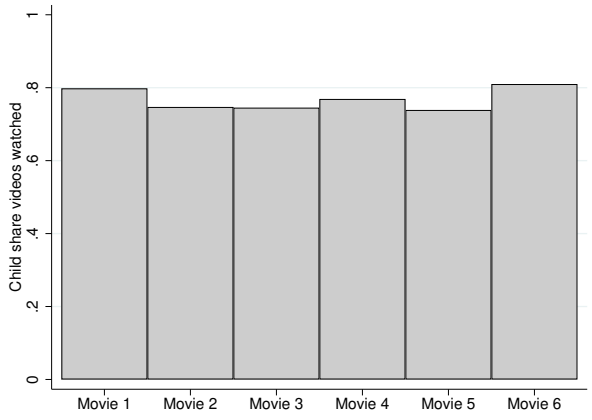
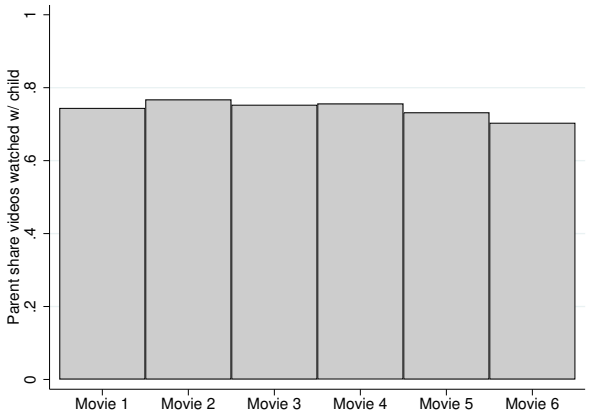


Figure A4: Stratification by Grade Level

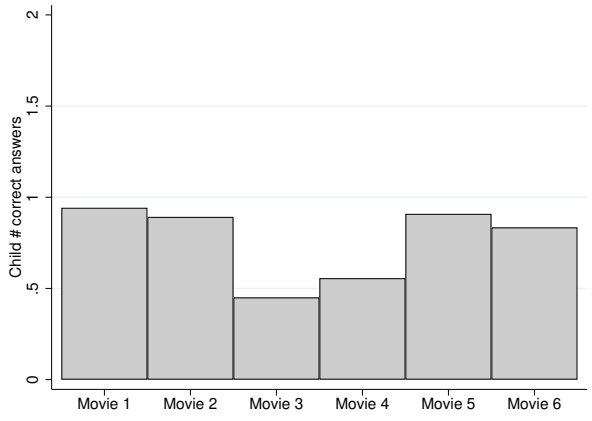
Notes: This graph shows the distribution of participants (including attriters) across grade levels and treatment groups. $N = 1091$.



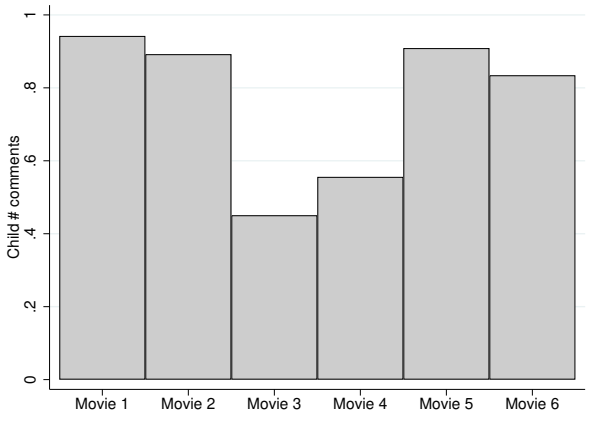
(a) Child watched



(b) Parent watched w/ child



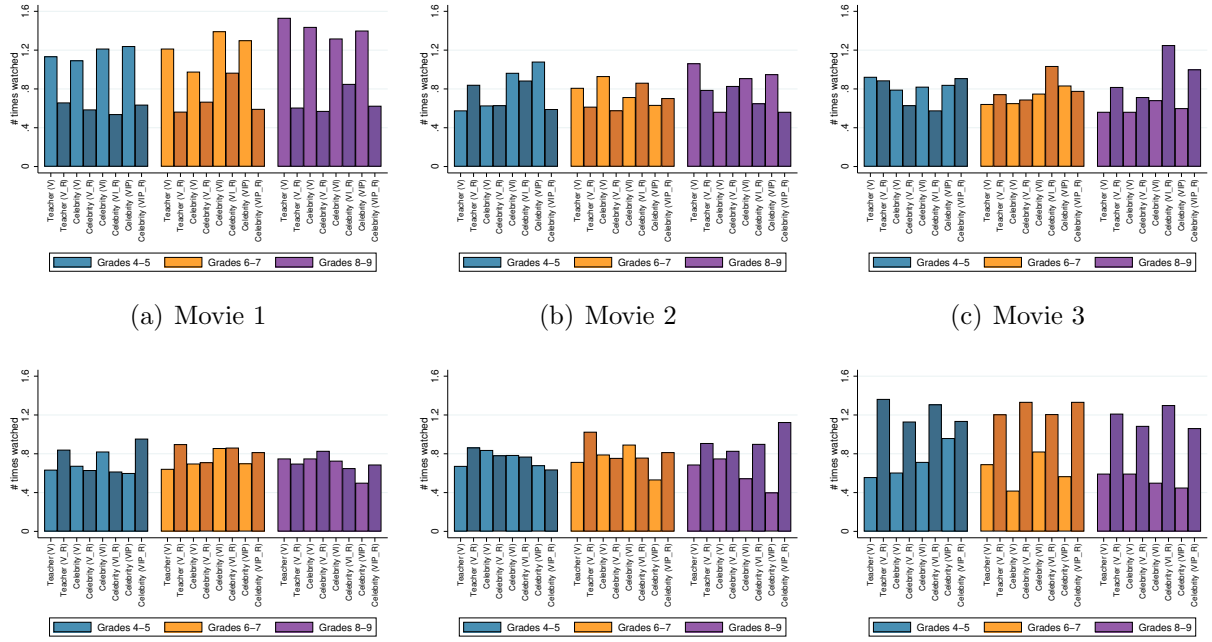
(c) Child # correct answers



(d) Child # comments

Figure A5: Watch Statistics

Notes: This figure shows watch statistics for each movie.

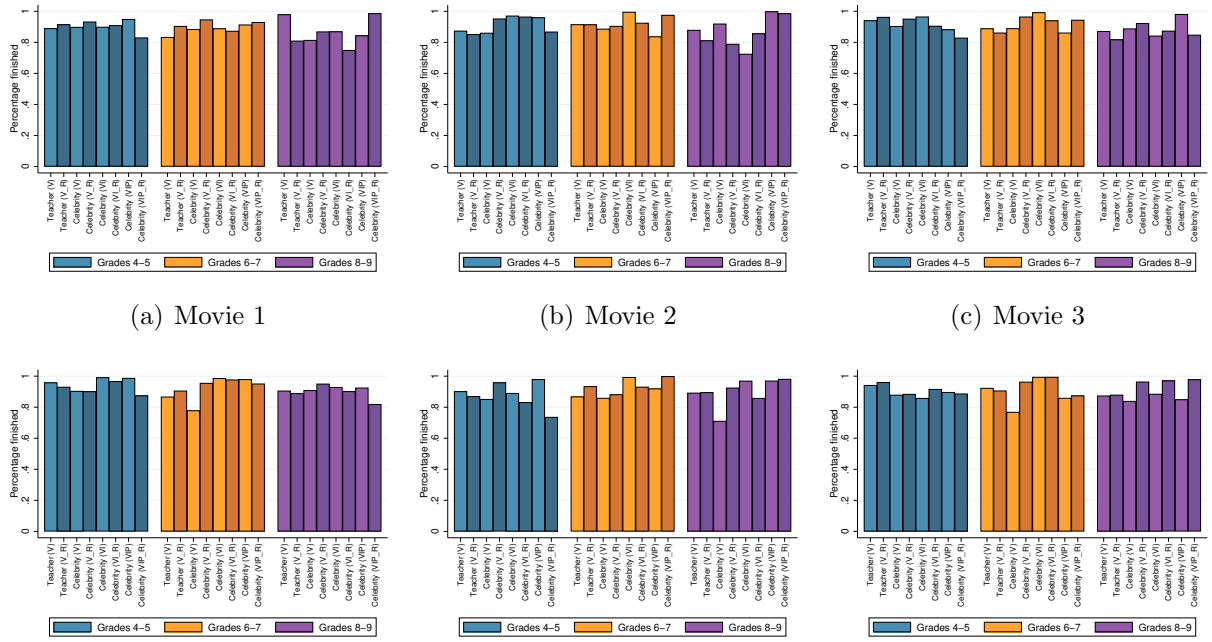


(a) Movie 1 (b) Movie 2 (c) Movie 3

(d) Movie 4 (e) Movie 5 (f) Movie 6

Figure A6: Watch Statistics – # Times Watched

Notes: This figure shows watch statistics (i.e., the mean # of times watched) for each movie for the different treatment groups.

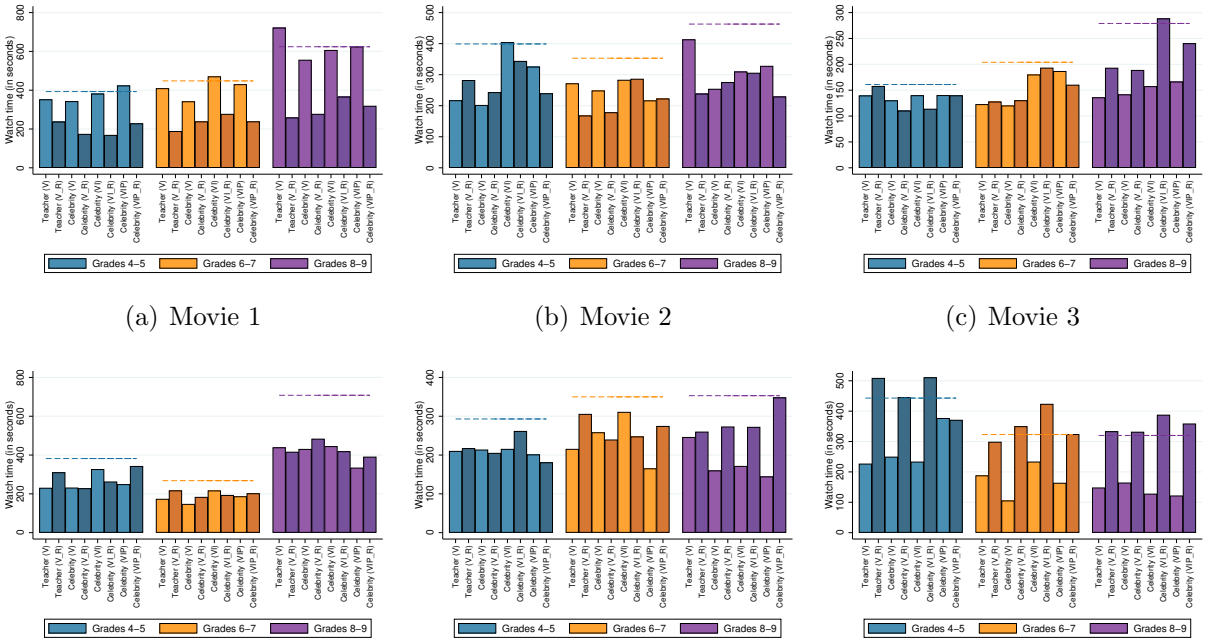


(a) Movie 1 (b) Movie 2 (c) Movie 3

(d) Movie 4 (e) Movie 5 (f) Movie 6

Figure A7: Watch Statistics – Percentage Finished

Notes: This figure shows watch statistics (i.e., the mean percentage finished) for each movie for the different treatment groups.



(a) Movie 1 (b) Movie 2 (c) Movie 3

(d) Movie 4 (e) Movie 5 (f) Movie 6

Figure A8: Watch Statistics – Watch Time

Notes: This figure shows watch statistics (i.e., the mean time watched) for each movie for the different treatment groups. Dashed lines indicate length of the video.

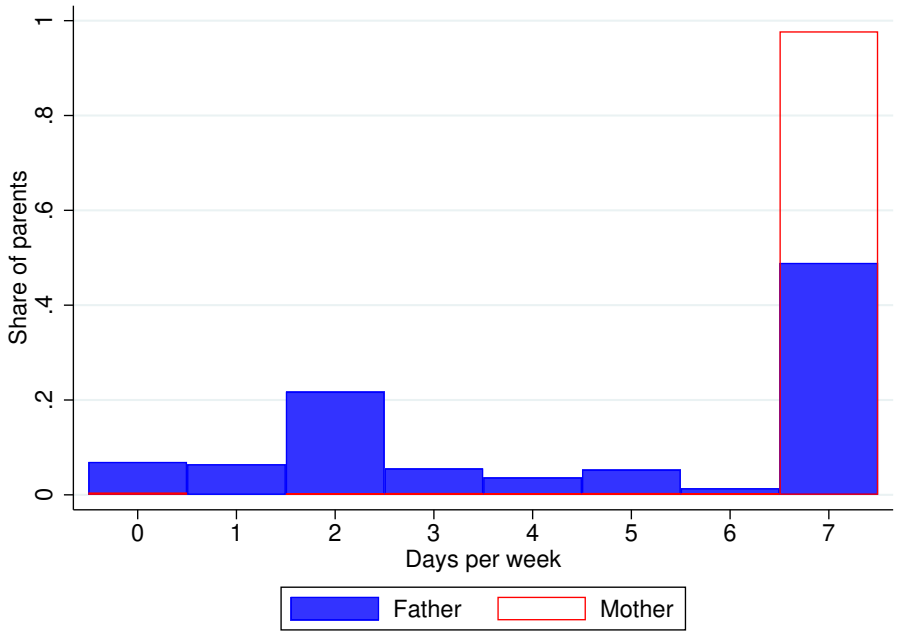


Figure A9: Parental Involvement

Notes: This graph shows the distribution of days per week the parent spends with her child.