DISCUSSION PAPER SERIES

DP17601

NURTURING CHILDHOOD CURIOSITY TO ENHANCE LEARNING: EVIDENCE FROM A RANDOMIZED PEDAGOGICAL INTERVENTION

Sule Alan and Ipek Mumcu

DEVELOPMENT ECONOMICS

CEPR

ISSN 0265-8003

NURTURING CHILDHOOD CURIOSITY TO ENHANCE LEARNING: EVIDENCE FROM A RANDOMIZED PEDAGOGICAL INTERVENTION

Sule Alan and Ipek Mumcu

Discussion Paper DP17601
Published 24 October 2022
Submitted 16 October 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

NURTURING CHILDHOOD CURIOSITY TO ENHANCE LEARNING: EVIDENCE FROM A RANDOMIZED PEDAGOGICAL INTERVENTION

Abstract

We evaluate a pedagogical intervention that aims to improve the learning quality of elementary school children by nurturing their curiosity. The pedagogy, aimed primarily at science teaching, was practiced by children's teachers for an entire academic year. We test the effectiveness of this pedagogy using objective test scores and a novel measure of curiosity. Our curiosity measure involves first creating a sense of information deprivation, then quantifying the urge to acquire information and the ability to retain information. We find that the intervention increases curiosity, the ability to retain knowledge, and science test scores. The intervention also makes friendship networks a potent tool to disseminate knowledge within classrooms. Our research design establishes the causal link between the urge to know and deep learning. The evidence can help design better pedagogical tools to increase pupil and teacher engagement and the quality of learning worldwide.

JEL Classification: N/A

Keywords: Deep learning, Curiosity

Sule Alan - salancrossley@gmail.com
European University Institute, Bilkent University and CEPR

Ipek Mumcu - I.Mumcu@exeter.ac.uk
University of Exeter

Acknowledgements
We are grateful to J-PAL Post-Primary Education Initiative and ING Bank Turkey for funding this study. We thank seminar participants at EUI, LSE, Bologna ESA 2022 for their valuable comments. We thank Enes Duysak, Mert Gumren, Elif Kubilay, Ozge Seyrek and Melek Celik for wonderful field assistance. Study 1 AEA Registry: AEARCTR-0003957. Study 2 AEA Registry: AEARCTR-0008629.

Nurturing Childhood Curiosity to Enhance Learning:

Evidence from a Randomized Pedagogical Intervention *

Sule Alan† and Ipek Mumcu‡

October 19, 2022

Abstract

We evaluate a pedagogical intervention that aims to improve the learning quality of elementary school children by nurturing their curiosity. The pedagogy, aimed primarily at science teaching, was practiced by children's teachers for an entire academic year. We test the effectiveness of this pedagogy using objective test scores and a novel measure of curiosity. Our curiosity measure involves first creating a sense of information deprivation, then quantifying the urge to acquire information and the ability to retain information. We find that the intervention increases curiosity, the ability to retain knowledge, and science test scores. The intervention also makes friendship networks a potent tool to disseminate knowledge within classrooms. Our research design establishes the causal link between the urge to know and deep learning. The evidence can help design better pedagogical tools to increase pupil and teacher engagement and the quality of learning worldwide.

Keywords: Curiosity; Deep Learning, Pedagogy, Achievement;

---

*We are grateful to J-PAL Post-Primary Education Initiative and ING Bank Turkey for funding this study. We thank seminar participants at EUI, LSE, Bologna ESA 2022 for their valuable comments. We thank Enes Duysak, Mert Gumren, Elif Kubilay, Ozge Seyrek and Melek Celik for wonderful field assistance. Study 1 AEA Registry: AEARCTR-0003957. Study 2 AEA Registry: AEARCTR-0008629.

†Corresponding Author: European University Institute, Florence, Italy and J-PAL

‡University of Exeter, United Kingdom

1 Introduction

Today, more children than ever enroll in primary and post-primary education in the developing world. Despite this progress, the quality of education remains low. Millions of children in developing countries leave school without the necessary foundational skills to help them achieve their potential and lead productive lives.[1] Low teacher quality, overcrowded classrooms, and inadequate levels of school inputs such as poorly designed curricula and insufficient teaching materials are among the many factors contributing to low learning outcomes (Glewwe, Lambert and Chen (2020)). Recent research highlights the role of pedagogy as a potentially effective policy tool to combat the issue of poor education quality. While there is no consensus on what constitutes a good pedagogy, teaching practices that respond to the needs of students at all levels, build on their individual strengths, and encourage them to learn through experimentation are likely to be effective.[2] Unfortunately, most traditional instruction techniques lack these features. They ignore heterogeneous learning paths, compel students to be passive listeners, and prevent the development of an active and inquisitive mind (Blanchard, Southerland and Granger (2009); Granger et al. (2012); Terrenghi et al. (2019); Ashraf, Banerjee and Nourani (2021)).

In this paper, we evaluate the effectiveness of a pedagogy that aims to nurture children's curiosity and improve learning outcomes. Motivated by the recent evidence on the neural mechanisms of human curiosity and its connection to deep learning, the pedagogy strives to cultivate children's natural urge to learn and explore novel phenomena. The pedagogy primarily targets scientific curiosity and is practiced by teachers throughout the academic year. Treated teachers receive extensive training on the importance of curiosity as a critical driver of success and practical ways to promote it in the classroom by leveraging children's natural love of mystery, humor, and excitement. Teachers are provided a pedagogical toolkit containing materials to support the prescribed teaching practices. These materials are various visual and reading materials that champion inquisitive minds and encourage children to question the conventional. While this pedagogy is relevant for any curricular topic, the

[1] According to 2017 Annual Status of Education Report for India, about 25% aged 14-18 fail to read basic text fluently in their language, 57% struggle with division (3 digits by 1 digit) (ASER (2018)). Results from similar tests in Pakistan and East Africa paint a similar picture. PISA and TIMMS results highlight large learning gaps between the developing and the developed world (Gust, Hanushek and Woessmann (2022)).

[2] For example, tailoring the level of teaching to children's ability has been shown to be effective in helping those who lag behind to catch up. (Banerjee et al. (2007); Banerjee et al. (2016); Banerji and Chavan (2016)).

toolkit predominantly targets scientific curiosity and contains activities that promote scientific inquiry in a joyful manner.

Curiosity, a fundamental component of human cognition, is considered one of the most critical drivers of success in most aspects of life. Berlyne (1954) and Loewenstein (1994) provide a theoretical framework for epistemic curiosity, described as "desire for knowledge" [3]. Cognitive psychology links curiosity to achievement in many domains ranging from education to health and overall life satisfaction; see (Chamorro-Premuzic and Furnham, 2006; Kashdan and Silvia, 2009; von Stumm, Hell and Chamorro-Premuzic, 2011; Gottfried et al., 2016; Shah et al., 2018). Recent advances in neuroscience shed light on the neural mechanisms of curiosity and how it is linked to learning. Gruber, Gelman and Ranganath (2014) show via functional magnetic resonance imaging that the brain's reward system is evoked when people are curious about a phenomenon, facilitating more enjoyable learning and knowledge retention (deep learning) through memory consolidation[4]. Moreover, they show that once sparked, curiosity creates deep learning moments and enhances the learning of any topic. While recognized as a powerful engine of learning, curiosity has not been studied within the context of education and education policy. Our limited knowledge of how to cultivate such a context-dependent trait and the difficulty of measuring it in its complexity are obvious reasons for the lack of policy-relevant studies. We know of no large-scale study that measures curiosity in school children nor a study that shows how it can be enhanced in the school environment. This paper advances the literature on both of these fronts.[5]

The pedagogical program we evaluate was implemented as two independent randomized controlled trials in two large provinces of Turkey. The first trial, implemented in the 2018-2019 academic year in the province of Mersin, included 50 primary and 27 post-primary schools. The post-primary schools were later dropped out of the program as it was decided that the proposed pedagogy was more suitable for the primary school level. To improve the power of the first study, we re-implemented the program in the neighboring province, Adana,

[3] Loewenstein (1994) also references therein describe curiosity in an elementary school student as reacting positively to new or mysterious events by showing urge to explore them.

[4] Memory consolidation is a process by which acquired information or experiences are poured into the long-term memory. This is more likely to happen when stimuli spark curiosity; see Gruber and Ranganath (2019)

[5] Recently, psychologists have shown interest in the relationship between what they refer to as "epistemic emotions" and learning. Epistemic emotions include intellectual courage, astonishment, curiosity, interest, wonder, surprise, the joy of verification, and the satisfaction of knowing; These studies are correlational in nature; See Vogl et al. (2019*b*) and Vogl et al. (2019*a*).

recruiting 84 additional primary schools. This second study took place in the 2021-2022 academic year. Our combined sample has 134 primary schools with about 11,000 primary school students and 425 teachers. After collecting detailed baseline data from all the children and teachers in Fall 2018 (Study 1) and Fall 2021 (Study 2), we randomly assigned 78 schools to treatment (25 in Study 1, 43 in Study 2). Teachers from the selected schools received training on the prescribed pedagogy. Teachers were given the entire academic year to practice the pedagogy in everyday teaching of the curricular topics, with a greater emphasis on science lessons. We test the effectiveness of the pedagogy using objective test scores, educational aspirations, and a novel measure of curiosity. When we implemented the second study in the Fall of 2021, we also collected longer-term data from the first study subjects (about three years after the implementation in 2018).

Curiosity is challenging to measure due to its context-dependent nature. Psychologists use survey tools to elicit different types of curiosity in adults (Litman and Spielberger (2003); Collins, Litman and Spielberger (2004); Litman, Collins and Spielberger (2005); Kashdan et al. (2020)). Behavioral tasks are used for very young children (Jirout and Klahr (2012)). While self-reported item-set questions can be useful to measure curiosity in adults, such questions usually have low reliability when implemented on children and adolescents. The reliability issue is all the more serious in our context, where some children are exposed to a treatment that encourages all behavioral expressions of curiosity. The lack of a reliable measure of curiosity for school children motivated us to develop an incentivized task-based tool. Our tool benefits from the theoretical framework developed by Loewenstein (1994) and draws insights from neuroscientific research on curiosity. Our curiosity measure involves first creating a sense of information deprivation, then quantifying the urge to acquire information, and finally, assessing the degree of knowledge retention after satisfying the urge.

To develop the task, we first conducted extensive pilot surveys to determine the topics of interest for our target age group. We identified eight broad interest categories representing about 95% of all reported topics by children. These are, "science", "animals", "history", "human anatomy', "vehicles", "cartoons", "space", and "sports". We then prepared eight booklets with these titles and placed ten surprising facts that are unlikely to be known to a child (or to an adult) in each booklet. The implementation of the task begins by showing children each booklet and telling them that they contain facts that most people do not know. This step aims to create the urge to know. After recording their preferred booklets, we elicit the children's willingness to pay for them. For this, we first endow children with experimental tokens that can be converted into small gifts of value. We then ask them to state the highest number of tokens they are willing to sacrifice for their preferred booklet, including the option of zero tokens. Our measure of the urge to acquire information, i.e., curiosity, is the child's willingness to pay for his/her preferred booklet.[6] The novelty of our task lies in its temporal component. After eliciting the urge to acquire information and distributing booklets, we revisit all classrooms exactly one week later, unannounced. During this surprise visit, we give the students and the teachers a 40-question test that contains questions whose answers are in the booklets. The performance in this test is our measure of knowledge retention (deep learning).

To identify the effect of the treatment on knowledge retention, we implement the curiosity task during the first visit using two regimes. In classrooms that are randomly assigned to the first regime, children receive their preferred booklets based on a randomly determined market price. In classrooms that are assigned to the second regime, only half of the children in a given classroom receive booklets. In these classrooms, the booklet distribution is purely random, regardless of children's willingness to pay and their choice of booklets. By ensuring that the number of booklets and the composition of topics are balanced across treatment status, the second regime allows us to estimate the treatment effect on knowledge retention. This regime also allows us to document, for the first time, the causal effect of curiosity on deep learning.

We first document that our measure of curiosity, the willingness to pay for a booklet, correlates well with fluid IQ, performance on the retention test, and actual test scores (crystallized IQ). We then estimate the effect of the program on curiosity, knowledge retention, test scores, and educational aspirations. We find that the program significantly increases children's curiosity by about 0.11 standard deviations (0.35 extra tokens forgone). The effect of the program on science curiosity is similar in size (0.10 standard deviations). While untreated children choose to give up 3.2 tokens for a science-related booklet on average, this amount is 3.6 tokens for treated children, implying a 12.5% increase in the willingness to pay for information. The effect of the intervention on deep learning is striking. Treated children score about 0.12 standard deviations higher in the unannounced test. What is even more striking is that after about 3 years and 1.5 years of school closure due to the recent pandemic, treated students score 0.14 standard deviations higher on the booklet test than untreated children, revealing remarkable knowledge retention. Our design also allows us to estimate the

[6] Willingness to pay elicitation is a standard method in economics research. In the context of information as a good, Hjort et al. (2021) uses this method to elicit policy-makers' willingness to pay for evidence in Brazil

degree of deep learning when curiosity is sparked (memory consolidation): A one standard deviation increase in curiosity leads to a 1.28 standard deviation higher knowledge retention.

The positive effects of the program on curiosity extend to actual learning outcomes. The program significantly improves children's objective test scores in science with no statistically significant impact on math and verbal scores. The estimated effect size on science test scores is about 0.08 standard deviations in the short term. We find that the positive effect on science test scores persists into middle school years, even after a long school closure. Treated students score 0.07 standard deviations higher than untreated students in a science test covering the middle school curriculum. Finally, we show that the intervention significantly raises children's aspirations to go to university and study science, although these effects are not as precisely estimated in the long-run.

Our results suggest that the program's success likely stems from its ability to unleash curiosity in both teachers and students. Treated teachers report a significant increase in their own curiosity level. They also report having adopted a growth mindset. We show that the program also increases children's tolerance for uncertainty and makes them more critical in their thinking process. Finally, we provide evidence that the program made friendship networks much more effective information dissemination tools, likely contributing to the positive results we estimate for knowledge retention and test scores. Treated students who did not receive a booklet but had someone in their friendship network who received a booklet scored 0.18 standard deviation higher booklet knowledge than untreated students in the same condition. These results strongly suggest more efficient information dissemination in treated classrooms where students are more curious and passionate about pursuing knowledge.

Our contribution is threefold. First, we evaluate a pedagogical intervention that targets a component of human cognition, curiosity, that has not been studied at a large scale and in a policy-relevant context before. Second, leveraging the neuroscientific evidence on curiosity and memory consolidation, we offer a novel approach to measuring curiosity in primary school children. Finally, combining the two, we establish, for the first time, the causal link between childhood curiosity and deep learning. We show that once sparked, curiosity creates deep learning moments and leads to enhanced knowledge retention. Therefore, the results of the paper are of high policy relevance. They can help us design better pedagogical tools to increase pupil and teacher engagement and the quality of learning globally. The results are particularly relevant for the developing world, where the learning outcomes have been alarmingly low and have deteriorated even further due to the Covid-19 pandemic (Goldhaber et al. (2022)).

Our paper relates to several strands of the economics literature. First, by showing the effectiveness of a particular pedagogical approach, it contributes to the literature that strives to improve learning outcomes in developing countries. This literature establishes that school-based inputs have very little effectiveness when not complemented by correct teaching practices (Glewwe et al. (2004); Kremer, Glewwe and Moulin (2009); Kremer, Brannen and Glennerster (2013)). Related literature explores the ways to improve teacher motivation and engagement and shows that extrinsic motivations have limited effectiveness in improving learning outcomes (de Ree et al. (2018)). Second, the paper also relates to a growing literature that shows that social and emotional skills are likely malleable and can be fostered at young ages (Alan and Ertac, 2018; Alan, Boneva and Ertac, 2019; Alan et al., 2021). By showing that an important trait can be cultivated in the classroom through a change in teaching practices, we advance this literature. Third, by testing a pedagogy that focuses mainly on science teaching, the paper speaks to the literature that aims to increase the STEM participation of girls (Buser, Peter and Wolter (2017); Fischer (2017); Kahn and Ginther (2017)). Our paper shows that teaching practices may play an important role in raising girls' aspirations for higher education and STEM majors. Finally, we contribute to neuroscience and psychology literature in their efforts to understand the cognitive underpinnings of human curiosity by causally linking the desire for knowledge with deep learning in children.

The rest of the paper is organized as follows. Section 2 summarizes the key features of the program and the context in which it was implemented. Section 3 details the evaluation design and gives a detailed account of our outcome measures, including our task-based curiosity measure. Section 4 describes the data and presents our main results. In Section 5, we present the causal link between curiosity and learning and explore mechanisms through which the program improved knowledge retention and achievement outcomes. We conclude in Section 6.

2 Evaluation Context and The Nature of The Pedagogical Program

The program we evaluate has been developed by a team of pedagogy experts and curricula developers in an innovation center connected to a private university. The program's overarching objective is to promote scientifically informed teaching practices to improve learning outcomes at all grade levels. It aims to do so by replacing traditional teaching with techniques that can stimulate children's interest in academic matters. Given the global emphasis on promoting STEM education and improving learning outcomes in science, the program puts a greater emphasis on science teaching.

The Turkish primary school system is such that a centrally appointed teacher is assigned to a classroom in grade 1. The teacher is expected to teach the same pupils until the end of grade 4, after which he/she goes back to teaching at grade 1.[7] Students move on to middle school for grades 5 to 8, where each subject is taught by a different (branch) teacher. The program has been developed exclusively for primary school teachers. It is thought that the ideal context to implement the prescribed pedagogy would be where a single teacher has a full day of contact with his/her pupils and when science concepts are formally introduced. Such a context is grade 4 of primary school in Turkey.

The intervention was an intensive teacher training program. In training seminars, teachers were first introduced to the concept of curiosity as a fundamental driver of academic achievement. Then, teachers were introduced to various pedagogical practices to cultivate curiosity in the classroom environment. These practices include ways to allow students to question and challenge the conventional, seek evidence, and ways to encourage them to express their interests openly in the classroom. During seminars, teachers interactively practiced ways of tapping into children's natural love of mystery, surprise, and humor to lock in their attention and create teachable moments before teaching important and difficult-to-teach concepts.

Teachers also received a toolkit to help them practice the pedagogy. They worked on this toolkit during the training seminars with the guidance of education consultants. The toolkit contains various visual and reading materials that emphasize the importance of having an inquisitive and open mind and the benefits of asking questions. These materials are designed to help practice the prescribed pedagogy and are not meant to be a set of materials to be covered in a specified period of time. Rather, they are designed to help the teacher create teachable moments before she introduces a new and complex topic. For example, before introducing a science topic on the solar system, which is an official curricular item to be covered, students see a short video on the mysteries of space. The video is designed to capture

[7] While this is a general practice, there are many exceptions to this rule. Firstly, the headteacher can decide which grade level the newly appointed teacher should begin teaching based on the needs of the school. Secondly, the Ministry can re-appoint a teacher, voluntarily or involuntarily, to another school at any grade level. These rotations tend to occur frequently for early career teachers

students' attention, tapping into their love of mystery to create a teachable moment. As another example of creating a teachable moment, this time, using humor, the teacher reads a funny story about a girl who gets excited about exploding liquids before introducing a topic on chemical reactions. While most activities are related to science, the toolkit contains some non-science activities as well. For example, in one of the activities, students read about a fictional student with a deep interest in painting using unconventional tools (finding making a mess with raw eggs liberating).

The overall feedback from the teachers regarding the program content was extremely positive. The majority of teachers reported that the program made everyday teaching, not just science teaching, much more enjoyable for children and for themselves. See the Online Appendix D for examples of activities provided in the toolkit and some implementation photographs.

3 Evaluation Design and Outcomes of Interest

The program was implemented as two independent randomized trials three years apart. The first trial took place in the 2018-2019 academic year covering 50 primary and 27 post-primary schools in the province of Mersin (Study 1). Despite the program's target grade of 4, with the recommendation of the local authorities in Mersin (Study 1), we decided to test the program also in the first year of the post-primary context by including a sample of 5th graders and their science teachers. However, it became clear during the training phase that the prescribed pedagogy would be very hard to implement in a middle school setting. Invited middle school teachers expressed their concerns regarding the larger number of pupils per classroom and the more demanding nature of the national curriculum relative to primary schools. We then removed middle schools from the study. Because we lost 27 schools after middle schools dropped out of the program, to improve the power of our design, we ran a second trial in the 2021-2022 academic year. The second trial covers 84 primary schools in the neighboring province, Adana.[8]

In both trials, local authorities provided us with a list of schools located in socioeconomically deprived neighborhoods in their provinces. Teachers from these schools were offered participation in the program. The program participation was voluntary on the part of teach-

[8] We first launched the second trial in 2019-2020 but failed to implement and evaluate it due to the Covid-19 related school closures, which lasted about 1.5 years in Turkey. We launched the second trial again as soon as schools opened in Fall 2021

ers. The program was oversubscribed in both provinces. Because Turkish state schools are generally big, containing multiple classrooms for each grade level, we were compelled to select 2 to 7 classrooms randomly from each school for evaluation purposes.[9] Two trials pooled together provide us with about 11,000 students and 425 teachers from 134 socioeconomically deprived state primary schools in two large provinces of Turkey. The majority of our sample is composed of 4th graders. We also have some third-grade students in our first study sample.[10]

The timeline of each trial is as follows: We collected baseline data for Study 1 in October 2018. We then conducted the randomization at the school level. We stratified our randomization by district and grade level. The probability of treatment is 50%, assigning 25 schools to treatment and 25 to control in Study 1. Teacher training seminars for Study 1 took place in November 2018. Short-term endline data were collected in May 2019. We collected baseline data for Study 2 in October 2021 and conducted the randomization at the school level, in the same manner, stratifying by district[11]. The probability of treatment is 50%, assigning 43 schools to treatment and 41 to control in Study 2. Teacher training seminars for 43 treatment schools took place in October 2021. Short-term endline data were collected in May 2022 for this study.

Both baseline and endline data collection were carried out by the research team, assisted by locally recruited and trained field assistants. We made sure that teachers were not present in classrooms during data collection. At baseline, we spent about three lecture hours in each classroom to conduct incentivized games, achievement and psychometric tests, and surveys. We implement our behavioral curiosity task at endline only. Because of the temporal nature of the task, we organized two visits for each classroom at endline, one week apart. On the first visit, we spent about two lecture hours conducting the curiosity task and collecting other relevant data subject tests, surveys, and friendship network information). Our second visit was unannounced. Upon arrival at the school, we kindly asked the teacher to spare us one lecture hour to implement a couple of tests on students and themselves. We will explain the nature of these tests in the next subsections.

[9]Primary school sizes vary significantly in our sample, ranging from remote village schools with a single 4th-grade class to overcrowded urban schools with over 15 classrooms for each grade level.

[10]We admitted a small number of grade 3 classrooms in the first study, comprising about 16% of the sample in this study. This is because we received an overwhelming interest from these teachers and admitted them to the program

[11]We managed to limit our sample to 4th graders in the second study.

In October 2021, almost three years after the first implementation of the program in Mersin (Study 1), we managed to conduct another round of data collection. Locating the original subjects of the first study was challenging as while most were scattered around various middle schools in the same province, some had left the province or left the education system altogether. We eventually located 86% of our original participants with the help of the provincial authority's database. Among those, 84% were formally registered in a state middle school in the province, and 7% moved to the private system or moved out of the province. In the end, we managed to collect long-term data from about 64% of the original participants. The attrition is more likely for girls and refugees, exacerbated by the long school closures due to the Covid-19 pandemic, but balanced across treatment status (p-value=0.66). Figure 1 depicts the timeline of the trials. Both trials were registered at the AEA Registry before their respective endline dates. The first trial was registered on March 8, 2019, along with a pre-analysis plan. The second trial was registered on November 30, 2021, referring to the first registry for the PAP.

Next, we will explain our curiosity task and the way we implement it in the classroom at endline, followed by descriptions of other outcomes of interest.

3.1 A Task-based Approach to Measuring Childhood Curiosity

We offer a novel incentivized task to measure curiosity and use it as our primary behavioral outcome. We designed this task to capture two prominent aspects of human curiosity: an urge to acquire knowledge and the retention of the acquired knowledge. For the first component, we benefit from the conceptual framework developed by Loewenstein (1994). Based on this framework, we first create a sense of information deprivation in children and then quantify the degree of the urge to acquire information. The second component of our task is informed by the neural mechanisms of curiosity documented in Gruber, Gelman and Ranganath (2014). That is, the higher the urge to know, the stronger the knowledge retention upon satisfying the urge (memory consolidation).

To develop the task, we first conducted extensive pilot surveys in several out-of-sample schools to determine the interests of the target age group. Compiling all our survey responses, we identified eight interest categories representing about 95% of all topics of interest. These are, "science", "animals", "history", "human anatomy', "vehicles", "cartoons", "space", and "sports". We then prepared eight small booklets for each topic with a cover that clearly shows the above titles. For example, the cover of the space booklet reads "The mysteries of SPACE," with eye-catching space illustrations to create information deprivation. Figure 2 shows the covers of all eight booklets. We placed in each booklet exactly ten pieces of information that are surprising and highly unlikely to be known by children (or by adults). Examples include, "the color of dawn on Mars is blue" in the space booklet, "the actual color of the black box in planes is orange" in the vehicles booklet, or "the shortest battle in history took 38 minutes" in the history booklet.

The implementation of the task in a classroom follows the following steps: We arrive at the classroom with booklets and a basket full of small gift items. The latter are small stationery items that are of value to children of the socioeconomic group we target in this study. We present the booklets to the children one by one, showing the title cover. We tell them that each booklet contains some incredible facts that are mostly unknown to people. We then ask children to rank these booklets according to their interest in the topic, 1 being the most interesting and 8 being the least interesting.

After obtaining their ranking, we inform children that everyone has an endowment of 10 tokens, and each token can be converted into a gift from our gift basket. We show children these gift items one by one. We then tell them they can also use their tokens to purchase a booklet if they want to. For this, they first need to state the booklet they would like to purchase by ticking the relevant box. We emphasize that they do not have to buy a booklet if they do not want to.[12] Then, we begin explaining how this purchase will be made in practice. We first emphasize that all booklets have the same price, and each student can only buy one booklet. We tell them no one knows the price of a booklet yet, but they need to state their willingness to pay for their preferred booklet, using the options ranging from zero to 10. Then we explain to children that one of two things may happen in their classroom:

- Market price implementation regime: In this regime, we randomly choose a booklet price (between 1 and 10) for the classroom. Students whose willingness to pay falls under the revealed market price does not receive their desired booklet. They, therefore, convert all their tokens into gift items. Those whose willingness to pay is at or above the revealed market price receive their desired booklets at the market price and convert their remaining tokens into gift items.
- Half-half implementation regime: In this regime, we do not choose a market price for

[12] Children see 9 options here, 8 topics, and an option of "no booklet." The layout of the task screen is provided in the Online Appendix.

the classroom. Instead, a random half of the classroom receives booklets and all 10 tokens worth of gift items, regardless of their stated willingness to pay and the type of booklet they prefer. The other half of the classroom receives 10 tokens worth of gift items but no booklet. We explain the rationale behind this implementation regime below.

The elicited willingness to pay, ranging from zero to 10, is our measure of "urge to know," i.e., curiosity. This measure is theoretically independent of the implementation regime, and our data corroborates this: Mean willingness to pay across regimes is statistically not different from each other (p-value=0.48). We conjecture that the treatment will increase children's willingness to pay for information on their preferred topic. Given the program's heavy focus on science, we expect this effect to be particularly prominent in the willingness to pay for science-related booklets, which we refer to as "scientific curiosity." These booklets are science, space, human body, animals and vehicles.

In addition to measuring the urge to acquire knowledge, we measure knowledge retention using the temporal component of our task. Specifically, we re-visit all classrooms, unannounced, precisely one week later. In this surprise visit, we give children a 40-question multiple-choice test containing 5 questions from each booklet. The score from this test is our measure of knowledge retention. However, when elicited under the market price regime, this outcome has two confounds. If the program increases children's overall curiosity, measured as the willingness to pay, we expect more students in treatment classrooms to have access to booklets under the random market price regime. This differential availability of knowledge is the first confound in measuring retention, as more availability likely leads to more knowledge mechanically. Similarly, if, say, science topics are more popular in treatment classes, more science booklets will be available in treated classrooms rendering differential availability of science-related knowledge, the second confound. In the half-half regime, a random half of the children in each class receive a randomly chosen booklet, regardless of their willingness to pay and their preferred booklet. Therefore we eliminate both confounds under this regime. By making the amount and the type of information available independent of the treatment status, this regime allows us to identify the program's impact on knowledge retention. It also allows us to estimate the causal effect of curiosity on deep learning. In Study 1, a given classroom had a 50% chance of being subject to either regime, and children were informed accordingly. Because the causal effect of the treatment on information retention can be estimated only in the half-half regime, to improve the power of the experimental design, we implemented the half-half regime in most classrooms (95%) in the second study, and children were informed accordingly. The Online Appendix E gives full instructions for the task and its implementation.

3.2 Learning Outcomes and Educational Aspirations

If the program successfully stimulates students' curiosity, we expect deeper learning of curricular topics as well. In particular, given the program's heavy emphasis on science teaching, we expect treated students to achieve higher test scores in science. To assess the impact of the program on actual learning outcomes, we implemented tests on math, Turkish (in visit 1), and science (in visit 2) in all classrooms. Because there is no standardized testing system in Turkey for the grade levels we work with, we designed a testing inventory based on the national curriculum[13] All tests were implemented in classrooms in the absence of teachers.

In addition to learning outcomes, we assess whether the program affected children's educational aspirations and their plans for study majors. For this, we ask children i) whether they would like to go to university, and if so, ii) about their aspired topic of study. We acknowledge that this is not a reliable measure of major choice considering the age of our subjects. Nevertheless, we believe that it gives us a signal regarding the program's success in raising educational aspirations in children.

We also collected data on various cognitive and non-cognitive skills, beliefs, and preferences at baseline and endline. Children's fluid was measured using Raven's progressive matrices (Raven and Court (1998)) at baseline. We conducted standardized achievement tests, both baseline, and endline. We elicited risk and ambiguity attitudes using Gneezy and Potters (1997) risky investment task, both at baseline and endline. Using surveys, we construct measures of curiosity, grit, impulsivity, critical thinking, and gender role beliefs. The motivation to collect these outcomes is to establish the validity of our task-based curiosity measure and explore potential channels through which the program might impact learning outcomes.

We also collected rich information from teachers. At baseline, we collected their fluid IQ via Raven's test and their emotional intelligence through reading the mind in the eyes test (Baron-Cohen et al. (1997)). In addition to demographic information collected at baseline, we collected information on their teaching styles, beliefs, behaviors, and, importantly, their

[13]We benefited from the Ministry's question bank in preparing these questions. We extensively piloted the tests to ensure the appropriateness of the difficulty level

curiosity at baseline and endline. Finally, we tested teachers' curricular knowledge in science to establish whether the intervention increased their content knowledge. We conducted this test in the second (surprise) visit along with the 40-question booklet test[14]. Full measurement inventory for students and teachers is presented in the Online Appendix C.

4 Data and Results

Before randomizing schools into treatment arms for each study sample, we visited all classrooms in person and spent three lecture hours collecting detailed baseline data from children and teachers. While we were collecting data from children, teachers were out in a quiet room working on their own surveys and tests.

Table 1 presents the balance of baseline variables across treatment status for the pooled sample. Panel 1 presents the balance of the baseline student characteristics, and Panel 2 the balance of classroom and teacher characteristics. Balance for each study separately is presented in Table A1 and A2 in the Online Appendix. We detect no significant imbalance in any of the variables in either study and conclude that randomization was successful.

We estimate the average treatment effects of the program on outcomes of interest by conditioning on baseline covariates and strata fixed effects:

$$y_{ics} = \alpha_0 + \alpha_1 T_s + X'_{ics} + \beta + W'_{cs}\gamma + \delta_d + \varepsilon_{ics} \tag{1}$$

where y_{ics} is the outcome of interest for child i in classroom c, school s. T_s is the binary treatment indicator, which equals one if school s is in the treatment group and zero otherwise, and X'_{ics} is a vector of student-level observables, W'_{cs} is a vector of classroom and teacher level observables measured at baseline. The former includes student gender, age in months, standardized fluid IQ score, risk aversion, and baseline achievement test scores. The latter include class size, the share of refugees in the classroom, teacher experience, teacher IQ, and teacher gender. δ_b are district fixed effects. The estimated $\hat{\alpha}_1$ is the average treatment effect. Standard errors are clustered at the school level. Throughout the text, we present the results from the pooled sample. The summary of the results for each province separately is given in Figure A1 in the Online Appendix. We present our full results corrected for multiple hypotheses testing (sharpened q-values and Romano Wolf p-values) in Table A3 in the Online Appendix. Most of our results survive the adjustments.

[14]Both science and booklet tests for teachers were implemented in the second study only

All treated teachers were expected to practice the proposed pedagogy upon receiving training. Recall that participation in the program was voluntary, and the program was oversubscribed. However, we acknowledge that compliance in terms of the actual implementation may not be perfect. To assess compliance, we asked treated teachers to report their estimated degree of program implementation at endline. Specifically, we asked them to mark their estimated degree of implementation using an unmarked 10cm line. The elicited distance gives us a continuous measure of program implementation intensity ranging anywhere between zero and 100%. Note that because this is a pedagogical intervention that aims to influence the way teachers teach, the reported implementation intensity is purely subjective. Nevertheless, we believe that it gives us an idea of teacher compliance, albeit noisy. Figure 3 depicts the distribution of the reported implementation intensity for the pooled sample. Overall, treated teachers report to have accomplished 81% program coverage. Given this high but imperfect compliance, the estimated $\hat{\alpha}_1$ should be interpreted as the average intent to treat effect (ITT).

4.1 The Predictive Validity of the Curiosity Task

Before presenting the program effects, we show that our curiosity measure (willingness to pay for a booklet) has predictive validity, i.e., correlates well with knowledge retention and test scores. To do this, we use our control sample. Figure 4 depicts the distribution of forgone tokens for the control sample. Children, on average, forgone 6.3 tokens to receive their desired booklet, with the minimum WTP being zero (6% of the sample) and a maximum of 10 (22.7% of the control sample).

Curiosity is known to be associated with higher cognitive ability in individuals, and our data corroborates this evidence. Table 2 presents the predictive power of overall curiosity, scientific curiosity, and non-science curiosity on science, math, and verbal test scores, as well as knowledge retention (performance on the respective questions in the booklet test). Panel 1 presents raw associations, and panel 2 presents the associations controlling for IQ. The results in this table confirm that our measure of curiosity has reasonable validity in predicting crystallized cognitive ability. Correlations are particularly strong for scientific curiosity, that is, the willingness to pay for a science-related booklet. One standard deviation increase in the willingness to pay for a science-related booklet is associated with 0.083 standard deviations higher science test scores, 0.02 standard deviation higher math scores, and 0.09 standard deviation higher verbal test scores.

Another important question is whether curiosity is associated with knowledge retention, as suggested by neuroscientists. Panel 1 also shows that a higher willingness to pay for a booklet is associated positively with the retention of information provided in that booklet. Specifically, a one standard deviation increase in the willingness to pay for a science booklet is associated with 0.08 standard deviations higher knowledge retention in science. This association remains strong, controlling for IQ (0.08 standard deviations). Finally, note that non-science curiosity, which is the willingness to pay for either history, sports, or cartoons booklet, is negatively associated with crystallized IQ but still positively associated with higher retention in non-science knowledge. The way we implement our curiosity task in the classroom and the fact that we have a randomly implemented program that enhances curiosity allows us to go beyond these correlations. In Section 5 we will show how curiosity causally leads to higher knowledge retention, i.e., deep learning in children.

4.2 Treatment Effects on Curiosity

We first explore whether the program affects children's interests and in particular, whether it increases their interest in science. Table 3 Panel 1 presents the estimated average marginal effects of the program on topic choice. The first column shows the treatment effect on the probability of choosing to purchase a science-related booklet (science, animals, space, vehicles, human anatomy). The second column presents the treatment effect on choosing a non-science booklet (history, sports, and cartoons). The last column gives the estimated effect of the treatment on "no interest," i.e., the probability of choosing not to purchase a booklet. Notice that about 50% of the children in the control group stated their willingness to purchase a science-related booklet. This value goes up to 54% in the treatment group, and this difference is statistically significant at the 1% level. It appears that the program shifted children's interest to science topics but not much at the expense of non-science topics (see column 2). As shown in column 3 of the table, the program lowered the probability of "no interest," i.e., stating zero willingness to pay, by 2.9 percentage points, representing about a remarkable 50% effect. The program effect on interest in science can also be seen in visual clarity in Figure 5. Treated children are significantly more likely to rank science, animals, and space booklets as their top 3.

Table 3 Panel 2 presents the estimated treatment effects on the willingness to pay for the desired booklet. Note that the measure is standardized to have zero mean for the control group so that the coefficient estimates are standard deviation effects. Column 1 presents the overall willingness to pay for any preferred booklet, column 2 presents the effect on the willingness to pay for a science-related booklet, and the last column for a non-science booklet. We estimate a significant 0.11 standard deviation effect on overall curiosity. In terms of tokens, this corresponds to forgoing about 0.35 extra tokens for a booklet. Given that children forgo 6.1 of their tokens on average for their preferred booklets in the control group, this effect implies a 6% treatment effect. The effect on science curiosity is similar with about 0.10 standard deviation treatment effect, again precisely estimated. Similarly, treated students exhibit higher curiosity for non-science subjects than untreated students. These results show that the program is successful in stimulating children's interest and curiosity. Our next question is whether this stimulated curiosity translates into actual learning. The temporal component of our task, along with the half-half implementation of booklet distribution, allows us to answer this question.

4.3 Treatment Effects on Learning

The estimated treatment effects on the willingness to pay suggest that in the market price regime, where the price of a booklet is determined randomly, treated classrooms necessarily end up with a proportionally higher number of booklets. This means that treated classrooms have more information (booklets) available for all, making it more likely to acquire and retain the knowledge provided in booklets. A clean identification of the effect of the program on knowledge retention requires the amount and the content of information to be balanced across treatment status. The half-half implementation regime delivers this by design. Recall that in classrooms subject to this regime, we distributed the booklets randomly to half of the students regardless of their willingness to pay and their choice of booklets. Panel 1 Table 4 presents the estimated treatment effects on booklet test scores using the full sample for comparison purposes. The first 3 columns give short-term, and the last 3 give long-term effects (only Study 1). Panel 2 presents the results using only the classes that were subject to the half-half regime. The effect of the program on the ability to retain knowledge is striking. Treated students perform significantly better than untreated students in the 40-question booklet test. Considering the half-half regime, where we have clean identification, treated students perform about 0.12 standard deviations higher than untreated students overall, and the performance difference is similar for performance in science topics (0.10 sd). Note that treated students perform better even in non-science topics of the test, consistent with the effects we estimate on non-science curiosity in Panel 2 of Table 3. What is truly remarkable is that after 3 years and a devastating pandemic, treated students still exhibit much higher booklet knowledge than untreated students, supporting the claims of neuroscientists that enhanced curiosity is associated with memory consolidation, i.e., deep learning. Treated students perform 0.14 standard deviations higher in the booklet test given after 3 years. The retention of science-related topics after 3 years is about 0.16 standard deviations.

Our next question is whether these positive learning effects extend to actual learning outcomes. Panel 3 in Table 4 presents the treatment effects on math, verbal (Turkish), and science test performance. While we do not estimate statistically significant effects on math and Turkish, we find that treated students perform significantly better than untreated students in the science test. The effect is about 0.08 standard deviations and significant at the 1% level. The positive effect on science test scores also persists into middle school years. We find that treated students still perform better than untreated students in science (0.07 standard deviation) 3 years after the implementation of the program. The near-zero effect sizes for math and verbal scores, while a significant and persistent effect on science, are not surprising given the program's heavy emphasis on science.

4.4 Treatment Effects on Educational Aspirations

To measure educational aspirations, we asked children two questions. First, we asked whether they intended to go to university when they grew up. Second, if they did, we asked what study major they wanted to pursue. For the latter, we gave them a full list of study majors to choose from. The first column of Table 5 presents the estimated treatment effect (average marginal effect) on the willingness to go to university. The following columns present the estimated average marginal effects on planned study majors. These are science, engineering, medicine, and Non-STEM (social sciences and humanities). Note first that almost all (95%) children in the control group stated that they plan to go to university when they grow up. Nevertheless, we still estimate a significant treatment effect on this high base, albeit small in size (1 percentage point). More importantly, only 12% of the children in the control group state their plan to major in science at university. This value is 2.4 percentage points higher for the treatment group, implying a 17% treatment effect. We estimate null effects for engineering and medicine. The estimated negative effect on non-STEM majors suggests that the positive effect we estimate for science comes at the expense of non-STEM majors. While 61% of the students express a preference toward a social science topic in the control group, treated students are 0.02 percentage points less likely to state such a preference. Note, however, while estimated sizes remain similar in the long run, the effects are estimated imprecisely.

4.5 Heterogeneity in Treatment Effects

As stated in our PAP, we explore heterogeneity in treatment effects with respect to two characteristics. First, we check whether the estimated effects are different across gender. Second, we investigate whether the program has a differential impact on children with different levels of cognitive ability. The first panel in Table 6 shows that the program's effect on the shift toward science topics mainly comes from girls. Treated girls are 7.8 percentage points more likely to choose a science booklet relative to untreated girls. The corresponding estimate is statistically zero for boys. As for choosing no booklet (no interest), we estimate no gender heterogeneity. Both boys and girls in the treatment group are significantly less likely to choose "no booklet" than those in the control group, suggesting that the program stimulated the overall interest of both boys and girls.

Similarly, we detect a significant gender heterogeneity in the treatment effect on curiosity. The estimates in Panel 2 indicate that while the program is effective in increasing curiosity for both genders, the results seem stronger for girls. Treated girls have 0.17 standard deviations higher scientific curiosity than untreated girls. We reject the equality of effects for overall curiosity as well as science and non-science curiosity. However, we estimate no significant gender heterogeneity in retention and test scores (Table 7). Finally, we do not detect any noteworthy gender heterogeneity in aspirations; see Table 8.

Table A4 in the Online Appendix presents treatment effect heterogeneity with respect to fluid cognitive ability. Here, we use our measure of fluid IQ (Raven score) and estimate treatment effects separately for high (above median) and low IQ (below median) levels. Overall, the estimated effects seem stronger for students with higher cognitive ability, although we fail to reject the equality of the estimated effects in most cases. The exception is the treatment effect on curiosity. As can be seen in Panel 2, while the program seems effective in increasing curiosity for all cognitive levels, its effect is stronger for students with high cognitive ability. This is reflected in the retention results (Table A5, Panel 1), but we fail to reject the equality of the estimates (see the borderline p-values). Finally, We do not estimate any treatment effect heterogeneity in test scores or aspirations with respect to IQ.

Taken together, our results suggest that the program was highly successful in increasing children's interest in science and stimulating their curiosity. In addition, it was highly effective in enhancing children's ability to retain the acquired knowledge and improving science test scores. In the next section, we will explore possible mechanisms through which the program achieves these positive results.

5 Potential Mechanisms

5.1 The Effect of Curiosity on Learning

We conjecture that curiosity is causally linked to deep learning. Therefore teaching practices that nurture children's curiosity lead to higher quality (deep) learning and better achievement outcomes though sparking children's curiosity. Our experimental design allows us to test this conjecture using the following simple structural model of memory consolidation (deep knowledge) though enhanced curiosity:

$$y_{ics} = \beta_0 + \beta_1 Curiosity_{ics} + X'_{ics}\beta_2 + W'_{cs}\beta 3 + \varepsilon_{ics} \tag{2}$$

where y_{ics} is the level of knowledge student i in classroom c, school s possess in a given time. $Curiosity_{ics}$ is the student's urge to know. X'_{ics} is a vector of student-level observables, which include demographics and all non-cognitive skills measured pre-treatment. W'_{cs} is a vector of classroom and teacher-level observables measured at baseline. Naturally, one's curiosity is naturally endogenous to her knowledge. However, our design allows us to estimate this structural equation using the treatment assignment as an instrument. We can conduct this estimation using the classrooms where we had the half-half regime. We also estimate this equation for test scores in the same way, using the full sample but the exclusion restriction for the latter is less likely to hold. Figure 7 depicts the margin plots of the second stage estimates for knowledge retention (performance on the booklet test) and for achievement scores. The positive causal effect of curiosity on retention is clear in Panel 1. The estimated coefficients are presented in Table A7. The results suggest that a one standard deviation increase in curiosity leads to a 1.28 to 1.13 standard deviation increase in knowledge. The effect on curricular knowledge in science is also large and significant. A one standard deviation increase in science curiosity increases science test scores by about 0.80 standard deviations, depicted in visual clarity in Figure 7 Panel 2. We estimate economically significant effects of curiosity on math and verbal ability, but the estimates do not reach statistical significance. The results for the long-term (Study 1) are too imprecise due to the small sample.

The IV estimator has the exclusion assumption that the effect on outcomes is coming only through the effect of the program on curiosity. While we conjecture that this is the primary mechanism, especially for the knowledge retention results, we explore other channels. Specifically, we test whether the program also had any impact on some key non-cognitive skills in children. We also explore the role of teachers, in particular, the impact of the program on their styles, practices, beliefs, and behaviors.

5.2 Effects on Students' Socioemotional Skills

While the program has a specific focus on curiosity, it is possible that it affects some other attributes in children. To explore this, we consider preferences, behaviors, and attitudes that are likely to be associated with curiosity. Specifically, we consider risk and ambiguity attitudes, impulsivity (Sleddens et al. (2013)), grit (Duckworth et al. (2007)) and critical thinking (Sosu (2013)). First, we show the extent to which these attributes are related to curiosity as we measure it. Here, we also check whether our task-based measure correlates with the survey measure of curiosity proposed by Kashdan and Silvia (2009). Then we explore whether the program has any effect on them.

Table 9 presents associations of curiosity, measured as the willingness to pay, with various attributes measured using surveys and incentivized tasks. Panel 1 presents raw associations, whereas Panel 2 shows associations with the respective attribute controlling for fluid IQ. The latter is to assess the extent to which our curiosity measure predicts individual preferences and socio-emotional skills over and above what is predicted by fluid IQ. As can be seen in both panels, while there is a negative correlation between science curiosity and impulsivity, curiosity, especially science curiosity, correlates positively with grit, risk and ambiguity tolerance, and survey-based curiosity measure. The correlations and their precision remain after we control for fluid IQ (Panel 2). Given these correlations, it is plausible that the program that enhances curiosity might also influence these attributes.

Figure 8 depicts the estimated effects on the aforementioned attributes. Note that for self-reported curiosity and grit, we also have long-term results. Consistent with the effects we estimate on the behavioral task, we estimate a large and precise treatment effect on self-reported curiosity. We find that this effect persists into adolescence (Study 1). We also find that treated children have become more tolerant of risk and ambiguity and better critical thinkers relative to untreated children. The latter are short-term results as we do not have long-term data on them.

5.3 Effects on Teachers' Practices and Beliefs

Given its success in affecting student outcomes, it is likely that the program was also successful in influencing teaching practices and teachers' beliefs. Recall that we collected detailed information from teachers regarding their everyday teaching practices and pedagogical beliefs both at baseline and endline. For the former, we adapted some of the item questions from the Teaching and Learning International Survey (TALIS) questionnaire (OECD (2013)), and constructed the following styles: Modern teaching, inquiry-based pedagogy, warmth, extrinsic motivator, competence. For the latter, we construct an index for growth mindset using Dweck (2008), attachment to the profession, and gender stereotyping. Finally, we also measure teachers' curiosity using Kashdan and Silvia (2009). We standardized all measures, so all coefficient estimates are standard deviation effects.

Figure 9 plots the estimated treatment effect on teaching styles, pedagogical beliefs, science, and booklet knowledge. All estimates are obtained by controlling for the baseline value of relevant outcomes, teacher demographics, and strata fixed effects. Estimates for science test scores and booklet test scores are available only for Study 2. What emerges from the figure is that the program made a positive impact on the teaching styles, teachers' curiosity, and mindset. Treated teachers report 0.23 (0.27) standard deviations and higher curiosity (growth mindset) than untreated teachers. In terms of teaching styles, the effects on warmth and the tendency to practice modern and inquiry-based teaching emerge as the most prominent. The treatment effects for these style items seem sizeable but are estimated imprecisely. These findings are consistent with the positive feedback we received from teachers regarding the program throughout the implementation period. It is clear that teachers who embrace the prescribed pedagogy become more curious themselves and adopt a more growth mindset. Note that we safely rule out a mechanism whereby teachers learn more curricular material themselves and therefore improve students' science scores. We estimate precise null effects on teachers' curricular knowledge in science. We also find no evidence of higher booklet knowledge in treated teachers. The latter rules out a mechanism whereby teachers learn the information provided in booklets and teach the students.

5.4 Information Dissemination in the Classroom

One of the explanations for these positive results may be that the pedagogy also made the classroom share information more readily and efficiently. It is shown that curiosity also leads to passionate information sharing (Hartung and Renner (2013); Litman and Pezzo

(2007)). Our design allows us to investigate whether the program made the classroom a denser learning environment where students enthusiastically share what they learn with their peers. We collected friendship networks at baseline and endline by asking each student to nominate at most 3 peers in their classrooms as their friends. With these nominations and the fact that we know who received which booklet, we can gain a deeper understanding of how the information provided to a subset of students in classrooms is disseminated and how treatment interacts with the way information is disseminated.

Table 10 shows the treatment effect on information retention (booklet test scores) for students who received a booklet (Panel 1) and those who did not (Panel 2). The effect sizes are larger for those who have the books in their possession (0.15 standard deviations) even though booklet ownership status was randomly assigned. What is striking is the evidence that the information was shared within treated classrooms much more effectively. Treated students who ended up not receiving a book performed 0.07 standard deviations better than their untreated counterparts. In Panel 3, we present the retention results for students who did not receive a booklet but had at least one student in their friendship network who received a booklet they wished to receive. The results are remarkable: We estimate about 0.19 standard deviation higher booklet knowledge for these students overall, indicating a pursuit of information. Treatment effects on science and non-science knowledge retention are 0.16 and 0.14 standard deviations for these students, respectively. These results strongly indicate more efficient information dissemination in treated classrooms where students are more curious and passionate about pursuing knowledge.

Overall, while we believe that the pedagogical program improves deep learning, especially in science, by increasing children's curiosity, we find that the program's impact on critical thinking, tolerance for uncertainty on the children's part, teaching practices, and heightened curiosity on the teachers' part may also play a role. Finally, as an additional channel, we show that the program made the classroom friendship networks more conducive to sharing information and learning from peers.

6 Conclusion

We test the effectiveness of a pedagogical program that aims to cultivate children's curiosity in the classroom. We evaluate the program with respect to children's urge to acquire knowledge and their ability to retain it. The pedagogy is informed by recent research on the neural mechanisms of human curiosity and mainly targets science teaching in elementary schools.

It involves offering teachers practices that encourage children to question the conventional and allow the behavioral expression of curiosity. In addition, the program teaches teachers how to utilize children's natural love of mystery, surprise, and humor to create teachable moments. The program was implemented as two independent clustered randomized controlled trials in two large provinces of Turkey, involving 134 primary schools, 425 teachers, and over 10,000 children of age 9 to 11.

To evaluate the program's effect on children's curiosity, we develop a behavioral measure that quantifies children's urge to acquire knowledge and their ability to retain knowledge for an extended period. We find that the intervention increases children's curiosity, measured by their willingness to pay for information and their ability to retain knowledge, i.e., learn deeply. We also show that the pedagogy significantly improves children's objective test scores in science and raises their educational aspirations for science. Moreover, the effects we estimate on knowledge retention and test scores persist well into the adolescent years.

The results are promising and likely to have high external validity. While the participation was voluntary, in practice, the program was oversubscribed. In all participating schools, the majority of the teachers were eager to join the program. Considering the policy issue of motivating teachers, the program's positive effects on teachers are particularly encouraging. The program was also highly cost-effective. The toolkit for teachers and other written materials are now available free of charge. The remaining program costs include printing hard copy materials, distributing the materials to schools, and conducting teacher training. Total printing costs were about 30,000 USD, the distribution costs were 9,000 USD, and teacher training costs were about 6,000 USD. These values imply a minimal (4 USD) program cost per child.

Global learning poverty is at its worse in the wake of a devastating pandemic. While the learning crisis predates the pandemic, the pandemic-related school closures made matters disproportionately worse for underprivileged children. They further widened the already sizeable socioeconomic achievement gaps to an alarming level in both developed and developing countries. The crisis now calls for evidence-informed and scalable actions more urgently than ever. One action may be to equip teachers with effective teaching practices that have a high chance of increasing teacher and pupil engagement, resulting in quality learning. We provide rigorous evidence on the effectiveness of one such scalable and cost-effective action. There are a couple of ways this program can be scaled up. One way is through incorporating the training in regular professional development seminars given to teachers at the beginning of the academic year. Another way can be to offer seminar courses for teacher candidates in universities. It is not clear which medium of delivery would be more effective and may be a topic of future research.

References

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy*, 126(5): 1865–1911.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking*." *The Quarterly Journal of Economics*, 136(4): 2147–2194.

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *The Quarterly Journal of Economics*, 134(3): 1121–1162.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

ASER. 2018. "Annual Status of Education Report 'Beyond Basics' (Rural) 2017." ASER Centre.

Ashraf, Nava, Abhijit Banerjee, and Vesall Nourani. 2021. "Learning to Teach by Learning to Learn." 115.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India." National Bureau of Economic Research Working Paper 22746.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India*." *The Quarterly Journal of Economics*, 122(3): 1235–1264.

Banerji, Rukmini, and Madhav Chavan. 2016. "Improving Literacy and Math Instruction at Scale in India's Primary Schools: The Case of Pratham's Read India Program." *Journal of Educational Change*, 17(4): 453–475.

Baron-Cohen, Simon, Therese Jolliffe, Catherine Mortimore, and Mary Robertson. 1997. "Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome." *Journal of Child Psychology and Psychiatry*, 38(7): 813–822.

Berlyne, D. E. 1954. "A Theory of Human Curiosity." *British Journal of Psychology. General Section*, 45(3): 180–191.

Blanchard, Margaret R., Sherry A. Southerland, and Ellen M. Granger. 2009. "No Silver Bullet for Inquiry: Making Sense of Teacher Change Following an Inquiry-Based Research Experience for Teachers." *Science Education*, 93(2): 322–360.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review*, 107(5): 125–130.

Chamorro-Premuzic, Tomas, and Adrian Furnham. 2006. "Intellectual Competence and the Intelligent Personality: A Third Way in Differential Psychology:." *Review of General Psychology.*

Collins, Robert P, Jordan A Litman, and Charles D Spielberger. 2004. "The Measurement of Perceptual Curiosity." *Personality and Individual Differences*, 36(5): 1127–1141.

de Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers. 2018. "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia." *The Quarterly Journal of Economics*, 133(2): 993–1039.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology*, 92: 1087–1101.

Dweck, Carol S. 2008. *Mindset: The New Psychology of Success.* Random House Digital, Inc.

Fischer, Stefanie. 2017. "The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence." *Labour Economics*, 46: 211–226.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics*, 74(1): 251–268.

Glewwe, Paul, Sylvie Lambert, and Qihui Chen. 2020. "Chapter 15 - Education Production Functions: Updated Evidence from Developing Countries." In *The Economics of Education (Second Edition).* , ed. Steve Bradley and Colin Green, 183–215. Academic Press.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods*." *The Quarterly Journal of Economics*, 112(2): 631–645.

Goldhaber, Dan, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger. 2022. "The Consequences of Remote and Hybrid Instruction During the Pandemic." National Bureau of Economic Research, Inc 30010.

Gottfried, Adele Eskeles, Kathleen Suzanne Johnson Preston, Allen W. Gottfried, Pamella H. Oliver, Danielle E. Delany, and Sirena M. Ibrahim. 2016. "Pathways from Parental Stimulation of Children's Curiosity to High School Science Course Accomplishments and Science Career Interest and Skill." *International Journal of Science Education*, 38(12): 1972–1995.

Granger, E. M., T. H. Bevis, Y. Saka, S. A. Southerland, V. Sampson, and R. L. Tate. 2012. "The Efficacy of Student-Centered Instruction in Supporting Science Learning." *Science*, 338(6103): 105–108.

Gruber, Matthias J., and Charan Ranganath. 2019. "How Curiosity Enhances Hippocampus-Dependent Memory: The Prediction, Appraisal, Curiosity, and Exploration (PACE) Framework." *Trends in Cognitive Sciences*, 23(12): 1014–1025.

Gruber, Matthias J., Bernard D. Gelman, and Charan Ranganath. 2014. "States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit." *Neuron*, 84(2): 486–496.

Gust, Sarah, Eric A. Hanushek, and Ludger Woessmann. 2022. "Global Universal Basic Skills: Current Deficits and Implications for World Development."

Hartung, Freda-Marie, and Britta Renner. 2013. "Social Curiosity and Gossip: Related but Different Drives of Social Functioning." *PLOS ONE*, 8(7): e69996.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2021. "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities." *American Economic Review*, 111(5): 1442–1480.

Jirout, Jamie, and David Klahr. 2012. "Children's Scientific Curiosity: In Search of an Operational Definition of an Elusive Concept." *Developmental Review*, 32(2): 125–160.

Kahn, Shulamit, and Donna Ginther. 2017. "Women and STEM."

Kashdan, Todd B., and Paul J. Silvia. 2009. "Curiosity and Interest: The Benefits of Thriving on Novelty and Challenge." In *Oxford Handbook of Positive Psychology, 2nd Ed. Oxford Library of Psychology*, 367–374. New York, NY, US:Oxford University Press.

Kashdan, Todd B., David J. Disabato, Fallon R. Goodman, and Patrick E. McKnight. 2020. "The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer Subscales While Separating Overt and Covert Social Curiosity." *Personality and Individual Differences*, 157: 109836.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science*, 340(6130): 297–300.

Kremer, Michael, Paul Glewwe, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics*, 1(1 (January 2009)): 112–135.

Litman, Jordan A., and Charles D. Spielberger. 2003. "Measuring Epistemic Curiosity and Its Diversive and Specific Components." *Journal of Personality Assessment*, 80(1): 75–86.

Litman, Jordan A., and Mark V. Pezzo. 2007. "Dimensionality of Interpersonal Curiosity." *Personality and Individual Differences*, 43(6): 1448–1459.

Litman, Jordan A., Robert P. Collins, and Charles D. Spielberger. 2005. "The Nature and Measurement of Sensory Curiosity." *Personality and Individual Differences*, 39(6): 1123–1133.

Loewenstein, George. 1994. "The Psychology of Curiosity: A Review and Reinterpretation." *Psychological Bulletin*, 116(1): 75–98.

OECD. 2013. "TALIS User Guide for the International Database."

Raven, John C, and John Hugh Court. 1998. *Raven's Progressive Matrices and Vocabulary Scales.* Vol. 759, Oxford pyschologists Press Oxford.

Romano, Joseph P, and Michael Wolf. 2005. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." *Journal of the American Statistical Association*, 100(469): 94–108.

Shah, Prachi E., Heidi M. Weeks, Blair Richards, and Niko Kaciroti. 2018. "Early Childhood Curiosity and Kindergarten Reading and Math Academic Achievement." *Pediatric Research*, 84(3): 380–386.

Sleddens, Ester F. C., Stef P. J. Kremers, Nanne K. De Vries, and Carel Thijs. 2013. "Measuring Child Temperament: Validation of a 3-item Temperament Measure and 13-item Impulsivity Scale." *European Journal of Developmental Psychology*, 10(3): 392–401.

Sosu, Edward M. 2013. "The Development and Psychometric Validation of a Critical Thinking Disposition Scale." *Thinking Skills and Creativity*, 9: 107–119.

Terrenghi, Ilaria, Barbara Diana, Valentino Zurloni, Pier Cesare Rivoltella, Massimiliano Elia, Marta Castañer, Oleguer Camerino, and M. Teresa Anguera. 2019. "Episode of Situated Learning to Enhance Student Engagement and Promote Deep Learning: Preliminary Results in a High School Classroom." *Frontiers in Psychology*, 10.

Vogl, Elisabeth, Reinhard Pekrun, Kou Murayama, and Kristina Loderer. 2019*a*. "Surprised–Curious–Confused: Epistemic Emotions and Knowledge Exploration." *Emotion*, No Pagination Specified–No Pagination Specified.

Vogl, Elisabeth, Reinhard Pekrun, Kou Murayama, Kristina Loderer, and Sandra Schubert. 2019*b*. "Surprise, Curiosity, and Confusion Promote Knowledge Exploration: Evidence for Robust Effects of Epistemic Emotions." *Frontiers in Psychology*, 10.

von Stumm, Sophie, Benedikt Hell, and Tomas Chamorro-Premuzic. 2011. "The Hungry Mind: Intellectual Curiosity Is the Third Pillar of Academic Performance." *Perspectives on Psychological Science*, 6(6): 574–588.

Tables

Table 1: Balance at Baseline

	N	Control Mean	Treatment Mean	Diff pvalue
Student Characteristics				
Male	13039	0.51	0.51	0.96
Age in Months	13039	112.43	112.74	0.25
Fluid IQ Score	10912	-0.08	-0.05	0.82
Math Score	10922	-0.07	-0.04	0.90
Verbal Score	10922	-0.05	-0.02	0.96
Curiosity	13039	-0.05	-0.03	0.59
Risk Attitude	13039	2.61	2.58	0.77
Ambiguity Attitude	10409	2.47	2.42	0.84
Gender Roles	10613	0.03	0.02	0.36
Home - Computer	10758	0.50	0.52	0.51
Home - Internet	10738	0.80	0.80	0.72
Siblingship Size	10814	2.74	2.72	0.90
Birth Order	10814	2.61	2.59	0.99
Teacher Characteristics				
Male	425	0.27	0.29	0.68
Age	425	45.49	44.66	0.25
Fluid IQ Score	425	17.76	17.70	0.81
Cognitive Empathy Score	425	23.05	22.94	0.87
Married	425	0.83	0.85	0.62
Number of children	425	1.81	1.71	0.18
Teaching experience in Years	425	21.01	20.50	0.41
University Graduate	425	0.94	0.95	0.50
Curiosity	425	-0.05	0.09	0.15
Gender Styping Beliefs	425	-0.05	-0.05	0.88
Growth Mindset	425	0.01	0.06	0.56
Professional Attachment	425	0.01	-0.00	0.85
Competence Beliefs	424	0.01	0.07	0.50
Modern Teaching	425	0.05	0.07	0.69
Extrinsic Motivator	425	-0.03	-0.11	0.20
Warmth	425	-0.08	-0.03	0.51
Inquiry-based Pedagogy	424	-0.04	0.03	0.40
Classroom Characteristics				
Classroom size	425	31.14	30.78	0.60
Refugee Share	425	0.07	0.07	1.00

The table presents the balance at baseline for the pooled sample. The p-values from the test of equality between control and treatment are shown in the last column. Test scores and survey items are standardized to have a mean zero and a standard deviation of 1.

Table 2: Predictive Power of Curiosity Task

Panel 1: Raw associations

	Science	Maths	Verbal	Retention
Overall Curiosity	0.041**	0.027	0.035*	0.047***
	(0.02)	(0.02)	(0.02)	(0.02)
Science Curiosity	0.083***	0.024*	0.086***	0.084***
	(0.02)	(0.01)	(0.01)	(0.02)
Non-Science Curiosity	-0.049***	-0.002	-0.056***	0.087***
	(0.02)	(0.01)	(0.01)	(0.02)
Observations	4558	4558	4675	4558

Panel 2: Raw associations controlling for IQ Score

	Science	Maths	Verbal	Retention
Overall Curiosity	0.021	0.014	0.011	0.040***
	(0.02)	(0.02)	(0.02)	(0.01)
Science Curiosity	0.063***	0.012	0.061***	0.076***
	(0.02)	(0.01)	(0.01)	(0.02)
Non-Science Curiosity	-0.046***	0.000	-0.052***	0.088***
	(0.02)	(0.01)	(0.01)	(0.02)
Observations	4558	4558	4675	4558

The table presents OLS coefficients of the regression of test scores (science, verbal, math and booklet test) on task-based curiosity measure (WTP). The analysis uses only the control sample. Standard errors are clustered at the school level and are reported in parentheses. Coefficient estimates are standard deviation effects, controlling district fixed effects in Panel 1, district fixed effects and fluid IQ (Raven's score) in Panel 2.

Table 3: Treatment Effect on the Choice of Booklet and Level of Curiosity

Panel 1: Choice of Booklet

	Science Related	Non-Science Related	No booklet
Treatment	0.040***	-0.011	-0.029***
	(0.01)	(0.01)	(0.01)
Control Mean	0.50	0.44	0.06
Observations	10858	10858	10858

Panel 2: Level of Curiosity

	Curiosity	Science Curiosity	Non-Science Curiosity
Treatment	0.111***	0.100***	-0.011
	(0.04)	(0.03)	(0.02)
Control Mean	-0.01	-0.00	-0.00
Observations	10852	10851	10851

Estimates are obtained via OLS. The dependent variables are binary indicators of choosing a science-related booklet (science, space, vehicles, human body, and animals) in column 1, choosing a nonscience-related booklet (history, sports, and cartoons) in column 2, and choosing no booklet option in column 3. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 4: Treatment Effect on Learning

Panel 1: Knowledge Retention - Full Sample

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment	0.119***	0.107***	0.089**	0.083*	0.121***	0.005
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Control Mean	-0.01	-0.01	-0.01	-0.00	0.00	0.00
Observations	10580	10580	10580	2417	2417	2417

Panel 2: Knowledge Retention - Half Half

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment	0.115**	0.103**	0.086*	0.143**	0.157**	0.071
	(0.05)	(0.05)	(0.04)	(0.07)	(0.06)	(0.06)
Control Mean	-0.04	-0.03	-0.03	-0.04	-0.02	-0.04
Observations	9055	9055	9055	1327	1327	1327

Panel 3: Test Scores

	Short Term			Long Term		
	Science	Maths	Verbal	Science	Maths	Verbal
Treatment	0.078***	0.017	0.033	0.074*	-0.018	-0.019
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Control Mean	-0.01	0.00	-0.00	0.00	0.00	0.00
Observations	9939	10390	10668	2417	2417	2417

Estimates are obtained via OLS. The dependent variables are standardized booklet test scores (knowledge retention) in Panels 1 and 2, and standardized subject test scores in Panel 3. The first 3 columns give short-term results using the pooled sample, and the last 3 provide the long-term results of Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 5: Treatment Effect on Aspirations

Panel 1: Short Term

	University	Science	Engineering	Medical	Non-STEM
Treatment	0.008*	0.024***	0.001	-0.003	-0.022**
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Control Mean	0.95	0.12	0.12	0.16	0.61
Observations	10683	10176	10176	10176	10176

Panel 2: Long Term

	University	Science	Engineering	Medical	Non-STEM
Treatment	0.010	0.020	0.010	-0.018	-0.012
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Control Mean	0.95	0.13	0.12	0.22	0.54
Observations	2312	2176	2176	2176	2176

Estimates are obtained via OLS. The dependent variables are binary choice variables of intention to go to university, intention to choose a science major, engineering major, medicine, and non-STEM major. Panel 1 presents short-term results from the pooled sample, and Panel 2 long-term results from Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 6: Heterogeneous Treatment Effects - Gender

Panel 1: Choice of Booklet

	Science Related	Non-Science Related	No booklet
Treatment = Girls	0.078***	-0.044**	-0.035***
	(0.02)	(0.02)	(0.01)
Treatment = Boys	0.002	0.022	-0.024***
	(0.02)	(0.02)	(0.01)
P-Value : Girls=Boys	0.003	0.010	0.218
Control Mean - Girls	0.50	0.44	0.07
Control Mean - Boys	0.49	0.44	0.06
Observations	10858	10858	10858

Panel 2: Level of Curiosity

	Curiosity	Science Curiosity	Non-Science Curiosity
Treatment = Girls	0.148***	0.174***	-0.056*
	(0.05)	(0.03)	(0.03)
Treatment = Boys	0.075*	0.028	0.033
	(0.04)	(0.03)	(0.04)
P-Value : Girls=Boys	0.059	0.001	0.066
Control Mean - Girls	-0.07	-0.01	-0.04
Control Mean - Boys	0.05	0.01	0.04
Observations	10852	10851	10851

Estimates are obtained via OLS. The dependent variables are binary indicators of choosing a science-related booklet (science, space, vehicles, human body, and animals) in column 1, choosing a nonscience-related booklet (history, sports, and cartoons) in column 2, and choosing no booklet option in column 3. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 7: Heterogeneous Treatment Effects - Gender

Panel 1: Knowledge Retention - Full Sample

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment = Girls	0.119**	0.107**	0.088*	0.031	0.054	-0.009
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
Treatment = Boys	0.119**	0.107**	0.089**	0.130*	0.163***	0.040
	(0.05)	(0.04)	(0.04)	(0.07)	(0.06)	(0.07)
P-Value : Girls=Boys	0.985	0.994	0.981	0.213	0.192	0.561
Control Mean - Girls	-0.09	-0.09	-0.06	-0.05	0.02	-0.12
Control Mean - Boys	0.07	0.07	0.04	0.05	-0.02	0.12
Observations	10580	10580	10580	2417	2417	2417

Panel 2: Knowledge Retention - Half Half

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment = Girls	0.113**	0.101**	0.085*	0.087	0.084	0.057
	(0.06)	(0.05)	(0.05)	(0.08)	(0.08)	(0.09)
Treatment = Boys	0.117**	0.104**	0.088*	0.183*	0.196**	0.097
	(0.05)	(0.05)	(0.05)	(0.09)	(0.08)	(0.09)
P-Value : Girls=Boys	0.930	0.932	0.952	0.434	0.322	0.762
Control Mean - Girls	-0.12	-0.11	-0.09	-0.07	0.03	-0.16
Control Mean - Boys	0.04	0.05	0.02	-0.01	-0.07	0.07
Observations	9055	9055	9055	1327	1327	1327

Panel 3: Test Scores

	Short Term			Long Term		
	Science	Maths	Verbal	Science	Maths	Verbal
Treatment = Girls	0.059	0.018	0.023	0.105**	-0.038	-0.006
	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
Treatment = Boys	0.098***	0.015	0.043	0.039	-0.003	-0.028
	(0.03)	(0.03)	(0.03)	(0.06)	(0.05)	(0.06)
P-Value : Girls=Boys	0.335	0.931	0.503	0.381	0.567	0.741
Control Mean - Girls	-0.01	-0.02	0.11	-0.02	0.02	0.12
Control Mean - Boys	-0.01	0.03	-0.12	0.02	-0.02	-0.12
Observations	9939	10390	10668	2417	2417	2417

Estimates are obtained via OLS. The dependent variables are standardized booklet test scores (knowledge retention) in Panels 1 and 2, and standardized subject test scores in Panel 3. The first 3 columns give short-term results using the pooled sample, and the last 3 provide the long-term results of Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 8: Heterogeneous Treatment Effects - Gender

Panel 1: Short Term

	University	Science	Engineering	Medical	Non-STEM
Treatment = Girls	0.008	0.031***	0.003	-0.010	-0.024
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Treatment = Boys	0.008	0.016	-0.001	0.005	-0.020
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
P-Value : Girls=Boys	0.998	0.305	0.751	0.341	0.829
Control Mean - Girls	0.96	0.08	0.06	0.23	0.63
Control Mean - Boys	0.94	0.15	0.17	0.09	0.58
Observations	10683	10176	10176	10176	10176

Panel 2: Long Term

	University	Science	Engineering	Medical	Non-STEM
Treatment = Girls	0.004	0.027	0.005	-0.044	0.013
	(0.01)	(0.02)	(0.02)	(0.03)	(0.04)
Treatment = Boys	0.013	0.001	0.021	0.016	-0.038
	(0.01)	(0.03)	(0.03)	(0.02)	(0.04)
P-Value : Girls=Boys	0.603	0.440	0.680	0.086	0.376
Control Mean - Girls	0.96	0.10	0.05	0.30	0.56
Control Mean - Boys	0.94	0.16	0.19	0.13	0.52
Observations	2312	2176	2176	2176	2176

Estimates are obtained via OLS. The dependent variables are binary choice variables of intention to go to university, intention to choose a science major, engineering major, medicine, and non-STEM major. Panel 1 presents short-term results from the pooled sample, and Panel 2 long-term results from Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table 9: Associations Between Curiosity Task (WTP) and Socio-emotional Skills

Panel 1: Raw associations

	Grit	Impulsivity	Risk	Ambiguity	Critical Thinking	Curiosity Survey	Science Curiosity
Overall Curiosity	0.051***	-0.014	0.228***	0.189***	0.050***	0.037**	0.042**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Science Curiosity	0.049***	-0.056***	0.086***	0.075***	0.067***	0.047***	0.053***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Non-Science Curiosity	-0.006	0.044***	0.108***	0.086***	-0.027*	-0.015	-0.017
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Observations	4524	4650	5070	5066	3635	4954	4954

Panel 2: Raw associations controlling for IQ Score

	Grit	Impulsivity	Risk	Ambiguity	Critical Thinking	Curiosity Survey	Science Curiosity
Overall Curiosity	0.044**	-0.007	0.232***	0.193***	0.040**	0.028*	0.031*
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Science Curiosity	0.042***	-0.048***	0.090***	0.079***	0.058***	0.038**	0.042***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Non-Science Curiosity	-0.005	0.043***	0.107***	0.086***	-0.025	-0.014	-0.015
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Observations	4524	4650	5070	5066	3635	4954	4954

The table presents OLS coefficients of the regression of socio-emotional skills, survey measure of curiosity and risk/ambiguity preferences on task-based curiosity measure (WTP). Risk and ambiguity preferences are measured via incentivized tasks. The analysis uses only the control sample. Standard errors are clustered at the school level and are reported in parentheses. Coefficient estimates are standard deviation effects, controlling district fixed effects in Panel 1, district fixed effects and fluid IQ (Raven's score) in Panel 2.

Table 10: Knowledge Retention Through Dissemination

Panel 1: Booklet Received

	Retention	Science Retention	Non-Science Retention
Treatment	0.151***	0.131**	0.120**
	(0.06)	(0.05)	(0.05)
Control Mean	0.00	-0.00	-0.00
Observations	4197	4197	4197

Panel 2: No Booklet Received

	Retention	Science Retention	Non-Science Retention
Treatment	0.079	0.072*	0.058
	(0.05)	(0.04)	(0.04)
Control Mean	0.00	0.00	0.00
Observations	5307	5307	5307

Panel 3: Network Effect

	Retention	Science Retention	Non-Science Retention
Treatment	0.185**	0.163**	0.143*
	(0.08)	(0.07)	(0.07)
Control Mean	0.00	0.00	0.00
Observations	1068	1068	1068

Estimates are obtained via OLS. The dependent variables are standardized booklet test scores (knowledge retention). Panel 1 uses the sample of booklet recipients only in the Half-Half regime. Panel 2 uses the sample of students who did not receive any booklet. Panel 3 uses the sample of students who did not receive any booklet but have at least one person in their network who has received the booklet of their choice. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Panel 3 specification include total number of friendship ties the student have in the classroom. Grade and district fixed effects are also included.

Figures

Figure 1: Timeline of the Two Trials

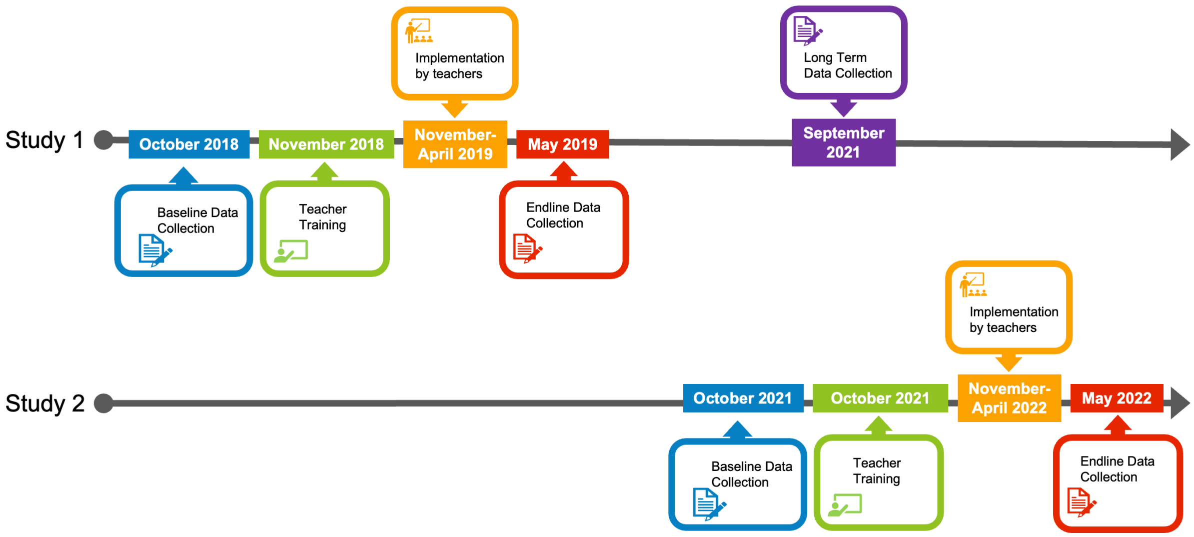
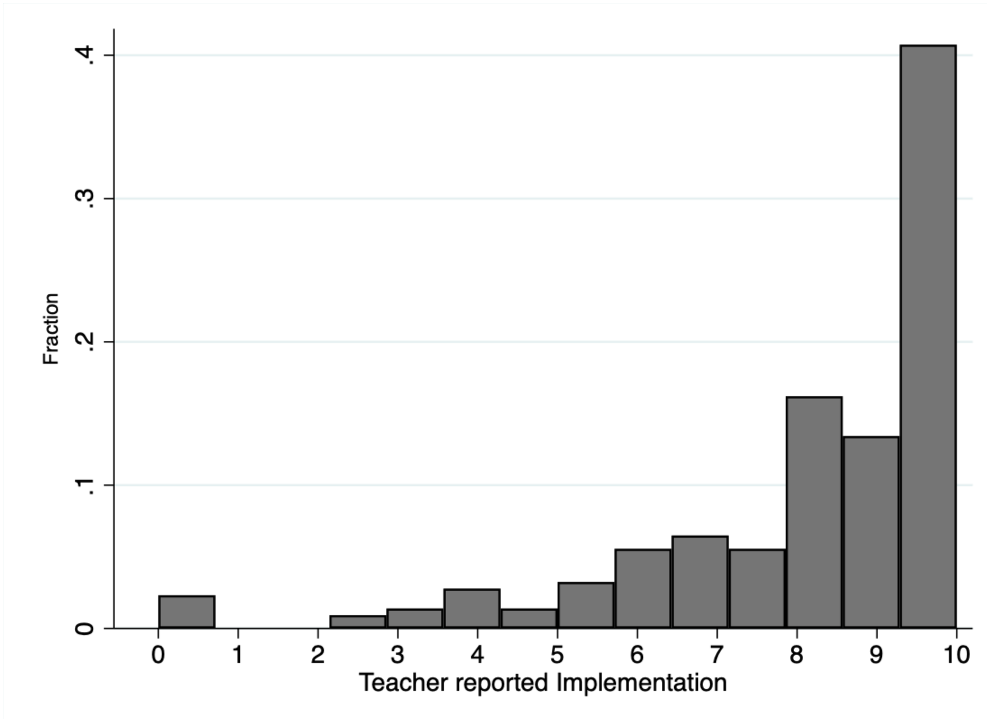


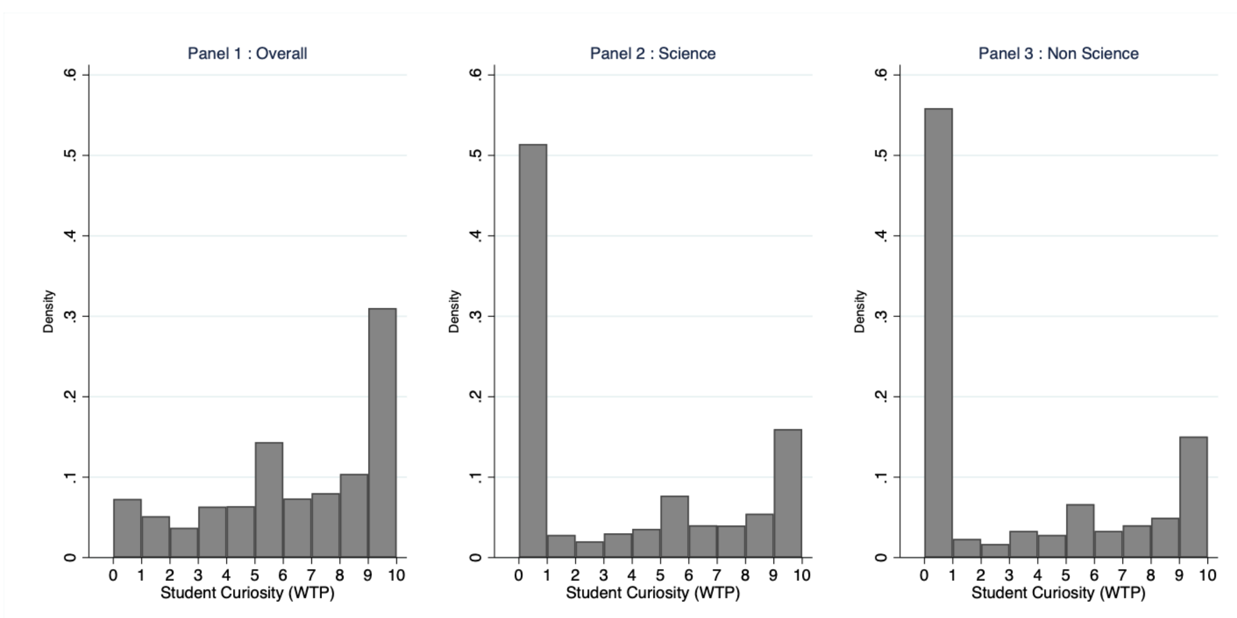
Figure 2: Covers of the Booklets

Figure 3: Implementation Intensity



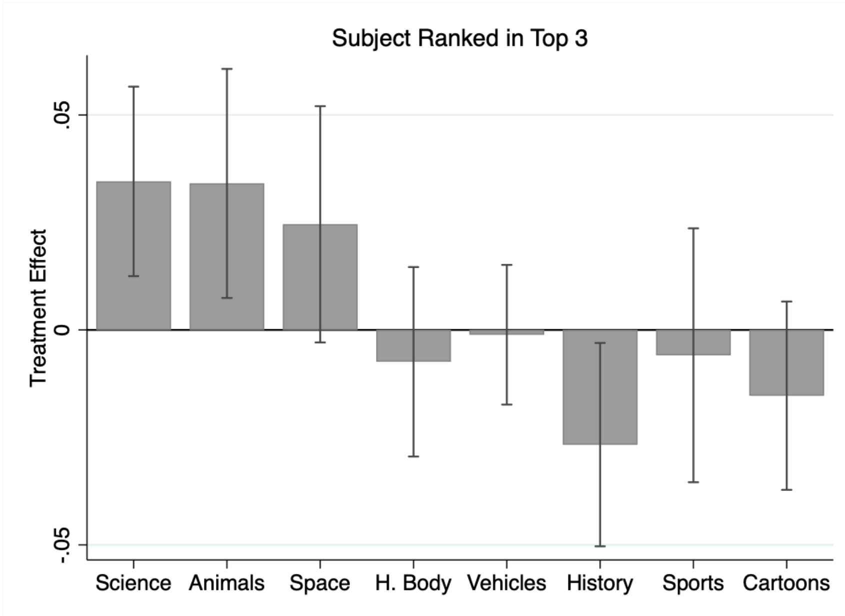
The figure depicts the program implementation intensity reported by treated teachers at endline. Teachers were given a 10cm line that has a moving cursor to report the level they believe represents their implementation intensity, zero representing no implementation, and 10 a 100% implementation.

Figure 4: Student Curiosity Distribution (WTP)



Figures depict the distribution of the number of tokens forgone for a booklet (Panel 1), for a science-related booklet (Panel 2), and for a non-science related booklet (Panel 3).

Figure 5: Treatment effect on the Ranking of Booklets



The figure depicts the marginal treatment effects obtained from logistic regressions on subject interest. The dependent variables are binary indicators of one if the respective booklet is ranked as one of the top 3 interests by the student. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Figure 6: Theory of Change-Potential Mechanisms

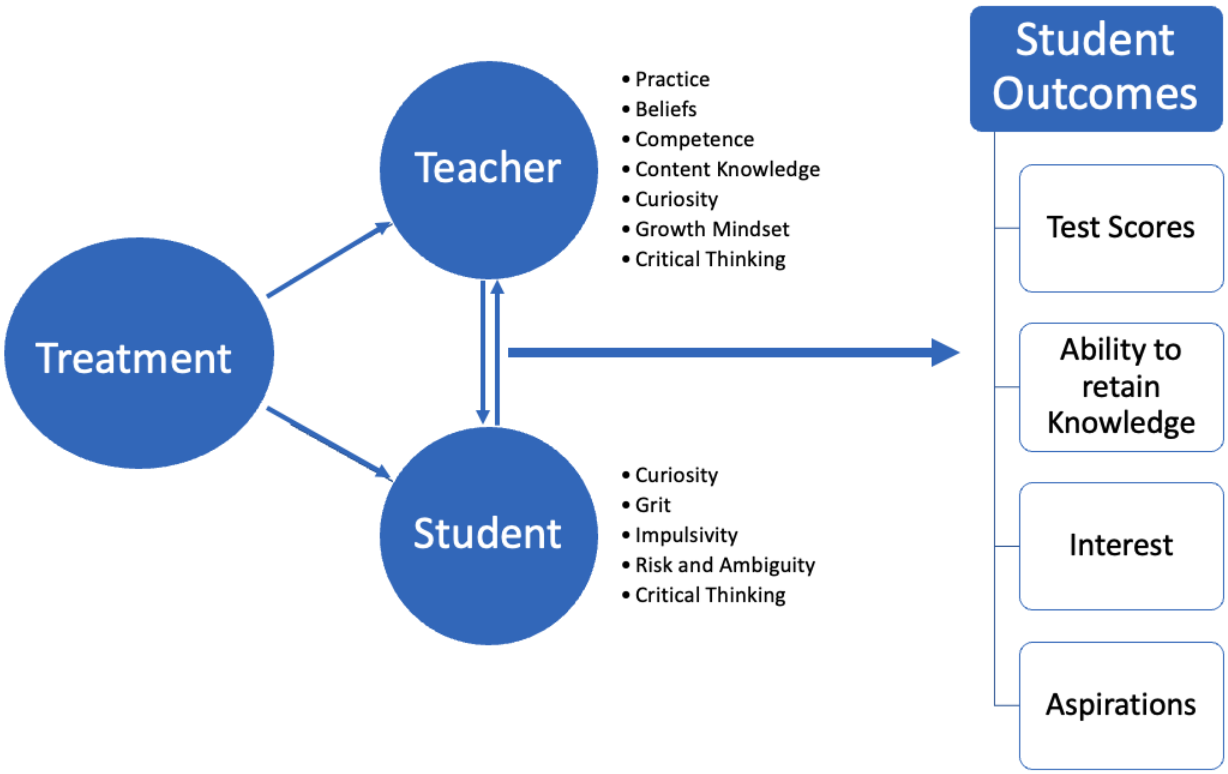
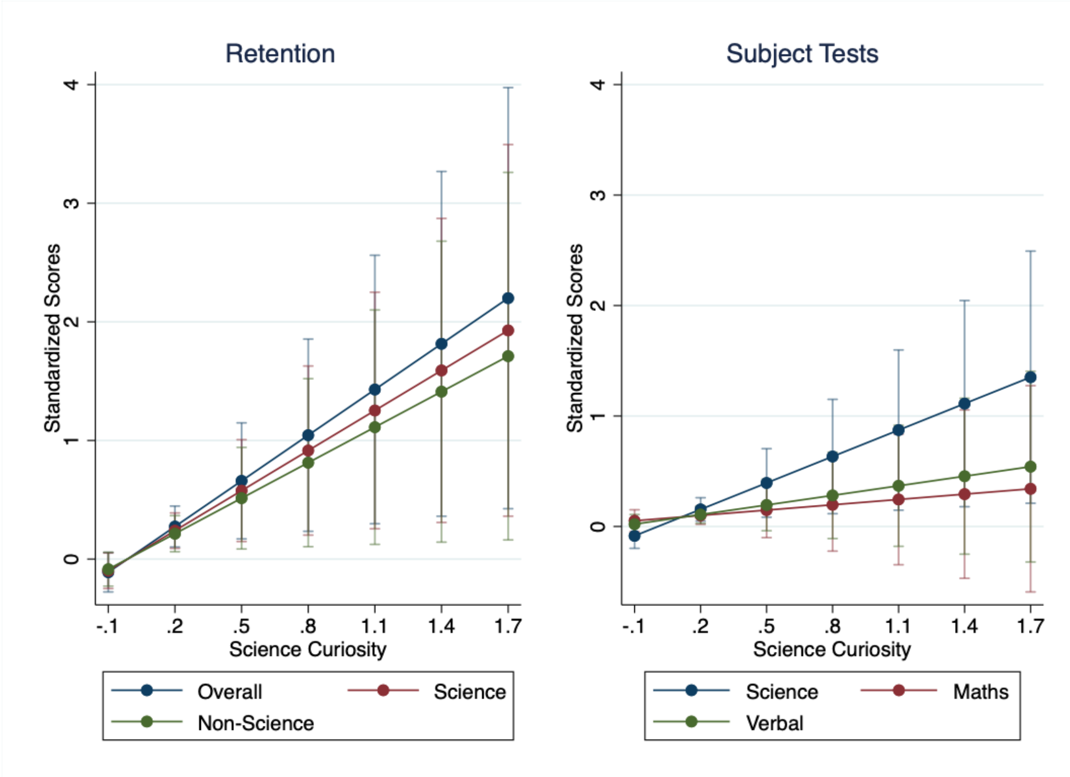
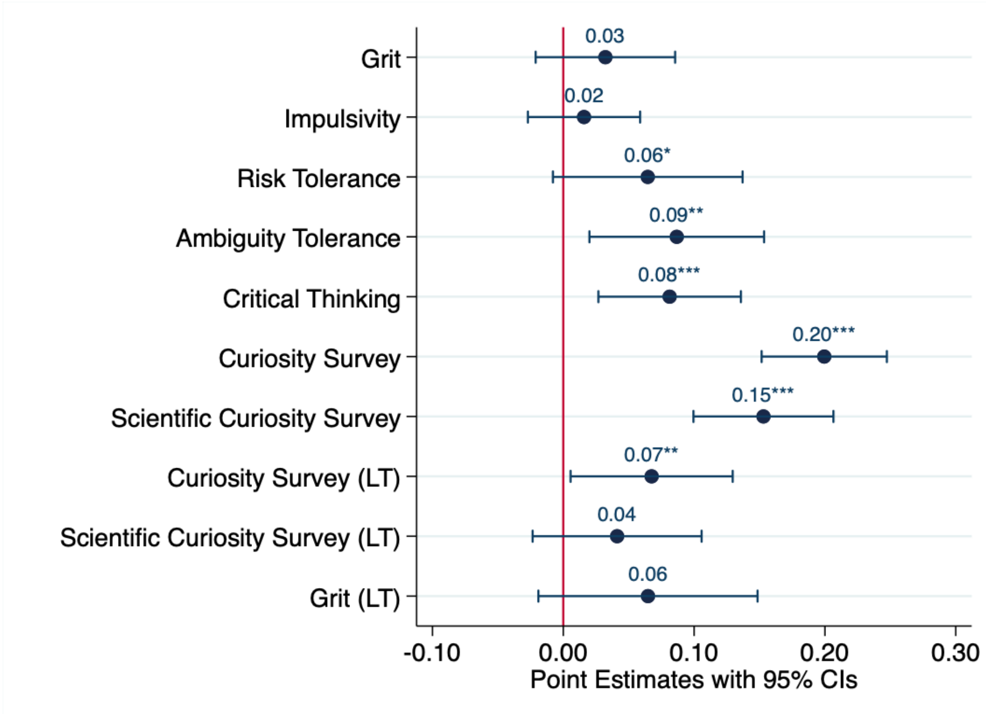


Figure 7: Effect of Curiosity on Learning - IV results



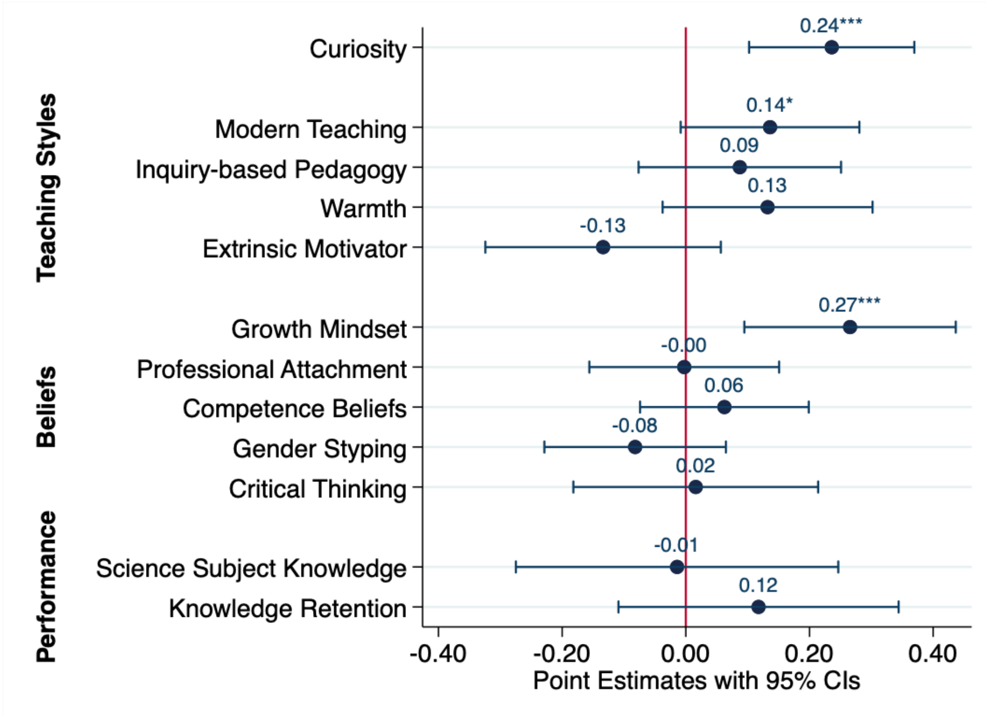
The figures depict the margin plots obtained from the two-stage least squares estimation of the effect of science curiosity on booklet retention (Panel 1), and test scores (Panel 2), where science curiosity is instrumented with the binary treatment assignment. Standard errors are clustered at the school level. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Figure 8: Treatment Effects on Students' Beliefs and Attitudes



The figure depicts the estimated treatment effects on socio-emotional skills, beliefs and attitudes. Standard errors are clustered at the school level. Covariates include individual baseline characteristics: gender, age, baseline fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores, baseline classroom and teacher characteristics: class size, the share of refugees, teacher gender, experience, and fluid IQ. Grade and district fixed effects are also included.

Figure 9: Treatment Effects on Teacher's Pedagogical Beliefs and Teaching Styles



The figure depicts the estimated treatment effects on teachers skills, beliefs, attitudes and teaching styles. Standard errors are clustered at the school level. Covariates include teacher gender, experience, fluid IQ, class size, the share of refugees. Grade and district fixed effects are also included.

Online Appendix- Not For Publication

A Additional Tables and Figures

Table A1: Balance at Baseline: Study 1

	N	Control Mean	Treatment Mean	Diff pvalue
Student Characteristics				
Male	3786	0.51	0.51	0.53
Age in Months	3786	110.89	111.56	0.47
Fluid IQ Score	3376	-0.01	0.05	0.82
Math Score	3386	-0.01	0.08	0.73
Verbal Score	3386	0.05	0.12	0.92
Curiosity	3786	0.06	0.07	0.90
Risk Attitude	3786	2.13	2.08	0.61
Ambiguity Attitude	2873	1.94	1.97	0.66
Gender Roles	3254	0.01	0.01	0.64
Home - Computer	3286	0.53	0.54	0.83
Home - Internet	3273	0.67	0.65	0.30
Siblingship Size	3324	2.71	2.67	0.68
Birth Order	3324	2.63	2.57	0.99
Teacher Characteristics				
Male	129	0.38	0.32	0.55
Age	129	43.05	42.19	0.39
Raven Score	129	19.13	19.17	1.00
Cognitive Empathy Score	129	22.65	22.91	0.57
Married	129	0.83	0.83	0.84
Number of children	129	1.58	1.51	0.46
Teaching experience in Years	129	18.95	18.17	0.37
University Graduate	129	0.92	0.93	0.66
Curiosity	129	-0.08	0.12	0.22
Gender Styping Beliefs	129	-0.03	-0.24	0.12
Growth Mindset	129	0.06	0.08	0.86
Professional Attachment	129	-0.08	0.15	0.18
Competence Beliefs	128	-0.00	0.21	0.18
Modern Teaching	129	0.04	0.15	0.33
Extrinsic Motivator	129	-0.01	-0.20	0.12
Warmth	129	0.02	0.13	0.28
Inquiry-based Pedagogy	128	-0.16	0.07	0.08*
Classroom Characteristics				
Classroom size	129	28.20	30.35	0.33
Refugee Share	129	0.14	0.13	0.76

The table presents the balance at baseline for Study 1 sample. The p-values from the test of equality between control and treatment are shown in the last column. Test scores and survey items are standardized to have a mean zero and a standard deviation of 1.

Table A2: Balance at Baseline: Study 2

	N	Control Mean	Treatment Mean	Diff pvalue
Student Characteristics				
Male	9253	0.51	0.51	0.68
Age in Months	9253	113.02	113.26	0.37
Fluid IQ Score	7536	-0.11	-0.10	0.90
Math Score	7536	-0.10	-0.10	0.99
Verbal Score	7536	-0.09	-0.08	0.99
Curiosity	9253	-0.09	-0.07	0.49
Risk Attitude	9253	2.80	2.80	0.97
Ambiguity Attitude	7536	2.65	2.61	0.65
Gender Roles	7359	0.04	0.02	0.24
Home - Computer	7472	0.49	0.52	0.42
Home - Internet	7465	0.85	0.87	0.29
Siblingship Size	7490	2.75	2.74	1.00
Birth Order	7490	2.61	2.60	0.99
Teacher Characteristics				
Male	296	0.22	0.27	0.33
Age	296	46.54	45.75	0.42
Raven Score	296	17.17	17.06	0.79
Cognitive Empathy Score	296	23.22	22.96	0.58
Married	296	0.83	0.86	0.46
Number of children	296	1.91	1.80	0.27
Teaching experience in Years	296	21.89	21.53	0.73
University Graduate	296	0.95	0.96	0.62
Curiosity	296	-0.03	0.07	0.34
Gender Styping Beliefs	296	-0.05	0.03	0.51
Growth Mindset	296	-0.01	0.06	0.57
Professional Attachment	296	0.05	-0.07	0.35
Competence Beliefs	296	0.02	0.01	0.93
Modern Teaching	296	0.05	0.04	0.87
Extrinsic Motivator	296	-0.03	-0.07	0.63
Warmth	296	-0.12	-0.11	0.94
Inquiry-based Pedagogy	296	0.02	0.01	0.93
Classroom Characteristics				
Classroom size	296	32.41	30.97	0.23
Refugee Share	296	0.04	0.04	0.74

The table presents the balance at baseline for Study 2 sample. The p-values from the test of equality between control and treatment are shown in the last column. Test scores and survey items are standardized to have a mean zero and a standard deviation of 1.

Table A3: Multiple Hypothesis Testing

	Original P-Value	Sharpened Q-Value	Romano Wolf P-Value
Panel 1: Student Outcomes			
Experimental Task			
Science Related Booklet	0.000	0.002	0.012
Non-Science Booklet	0.323	0.158	0.509
No Booklet	0.000	0.002	0.012
Overall Curiosity	0.006	0.013	0.046
Science Curiosity	0.000	0.001	0.008
Non-Science Curiosity	0.666	0.320	0.691
Retention	0.009	0.015	0.050
Science Retention	0.009	0.015	0.050
Non-Science Retention	0.027	0.026	0.090
Achievement & Aspirations			
Science	0.007	0.014	0.084
Maths	0.554	0.268	0.926
Verbal	0.202	0.135	0.631
University Aspiration	0.082	0.060	0.383
Science Aspiration	0.000	0.002	0.012
Engineering Aspiration	0.896	0.399	0.938
Medical Aspiration	0.724	0.337	0.938
Non-STEM Aspiration	0.049	0.045	0.293
Students' Beliefs & Attitudes			
Grit	0.234	0.140	0.457
Impulsivity	0.466	0.230	0.539
Risk	0.080	0.060	0.269
Ambiguity	0.011	0.017	0.048
Critical Thinking	0.004	0.011	0.024
Curiosity Survey	0.000	0.001	0.002
Science Curiosity	0.000	0.001	0.002
Panel 2: Teacher Outcomes			
Curiosity	0.001	0.008	0.012
Modern Teaching	0.064	0.274	0.513
Inquiry-based Pedagogy	0.293	0.522	0.858
Warmth	0.126	0.395	0.699
Extrinsic Motivator	0.167	0.430	0.762
Growth Mindset	0.003	0.015	0.036
Professional Attachment	0.974	0.842	0.998
Competence Beliefs	0.367	0.580	0.858
Gender Styping	0.272	0.522	0.858
Critical Thinking	0.872	0.842	0.998
Science Subject Knowledge	0.915	0.842	0.998
Knowledge Retention	0.305	0.522	0.858

The table presents estimation results for sharpened False Discovery Rate (FDR) q-values (Anderson (2008)) and adjusted p-values via Romano and Wolf (2005) multiple hypothesis correction. To accommodate Romano-Wolf correction to control for family wise error rate (FWER), we group our outcome variables into three, namely (i) experimental outcomes, (ii) achievement and aspiration related outcomes, (iii) beliefs and attitudes.

Table A4: Heterogeneous Treatment Effects - IQ

Panel 1: Choice of Booklet

	Science Related	Non-Science Related	No booklet
Treatment = Low IQ	0.028	0.007	-0.034***
	(0.02)	(0.01)	(0.01)
Treatment = High IQ	0.047***	-0.021	-0.026***
	(0.01)	(0.01)	(0.01)
P-Value : Low = High	0.348	0.143	0.402
Control Mean - Low IQ	0.48	0.45	0.07
Control Mean - High IQ	0.50	0.43	0.06
Observations	10858	10858	10858

Panel 2: Level of Curiosity

	Curiosity	Science Curiosity	Non-Science Curiosity
Treatment = Low IQ	0.062	0.048	0.003
	(0.05)	(0.04)	(0.03)
Treatment = High IQ	0.142***	0.134***	-0.019
	(0.04)	(0.03)	(0.03)
P-Value : Low = High	0.060	0.058	0.572
Control Mean - Low IQ	-0.05	-0.04	0.00
Control Mean - High IQ	0.02	0.02	-0.00
Observations	10852	10851	10851

Estimates are obtained via OLS. The dependent variables are binary indicators of choosing a science-related booklet (science, space, vehicles, human body, and animals) in column 1, choosing a nonscience-related booklet (history, sports, and cartoons) in column 2, and choosing no booklet option in column 3. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table A5: Heterogeneous Treatment Effects - IQ

Panel 1: Knowledge Retention - Full Sample

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment = Low IQ	0.065	0.055	0.054	0.012	0.022	-0.006
	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)
Treatment = High IQ	0.153**	0.140**	0.111**	0.124**	0.164***	0.027
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
P-Value : Low = High	0.111	0.118	0.248	0.253	0.143	0.694
Control Mean - Low IQ	-0.13	-0.14	-0.06	-0.16	-0.04	-0.23
Control Mean - High IQ	0.07	0.08	0.03	0.10	0.03	0.15
Observations	10580	10580	10580	2417	2417	2417

Panel 2: Knowledge Retention - Half Half

	Short Term			Long Term		
	Retention	Science Retention	Non-Science Retention	Retention	Science Retention	Non-Science Retention
Treatment = Low IQ	0.071	0.045	0.079*	0.049	0.052	0.026
	(0.05)	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)
Treatment = High IQ	0.145**	0.141**	0.092	0.197**	0.202**	0.115
	(0.07)	(0.06)	(0.06)	(0.09)	(0.09)	(0.08)
P-Value : Low = High	0.239	0.112	0.808	0.187	0.207	0.327
Control Mean - Low IQ	-0.16	-0.16	-0.10	-0.14	-0.02	-0.24
Control Mean - High IQ	0.04	0.05	0.01	0.04	-0.03	0.11
Observations	9055	9055	9055	1327	1327	1327

Panel 3: Test Scores

	Short Term			Long Term		
	Science	Maths	Verbal	Science	Maths	Verbal
Treatment = Low IQ	0.064*	0.012	0.055*	0.036	-0.129**	0.008
	(0.03)	(0.04)	(0.03)	(0.06)	(0.06)	(0.08)
Treatment = High IQ	0.088**	0.021	0.023	0.094*	0.047	-0.033
	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.05)
P-Value : Low = High	0.652	0.888	0.402	0.440	0.057	0.658
Control Mean - Low IQ	-0.36	-0.20	-0.43	-0.32	-0.34	-0.21
Control Mean - High IQ	0.21	0.13	0.26	0.21	0.22	0.14
Observations	9939	10390	10668	2417	2417	2417

Estimates are obtained via OLS. The dependent variables are standardized booklet tests scores (knowledge retention) in Panel 1 and 2, standardized subject test scores in Panel 3. The first 3 columns give short-term results using the pooled sample, the last 3 give the long-term results of Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

Table A6: Heterogeneous Treatment Effects - IQ

Panel 1: Short Term

	University	Science	Engineering	Medical	Non-STEM
Treatment = Low IQ	0.007	0.017	0.007	0.006	-0.029*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Treatment = High IQ	0.009*	0.028***	-0.002	-0.008	-0.018
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
P-Value : Low = High	0.813	0.438	0.506	0.322	0.576
Control Mean - Low IQ	0.94	0.12	0.10	0.14	0.64
Control Mean - High IQ	0.96	0.11	0.13	0.18	0.59
Observations	10683	10176	10176	10176	10176

Panel 2: Long Term

	University	Science	Engineering	Medical	Non-STEM
Treatment = Low IQ	0.016	0.001	0.002	-0.021	0.018
	(0.01)	(0.03)	(0.02)	(0.03)	(0.04)
Treatment = High IQ	0.004	0.021	0.019	-0.011	-0.029
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
P-Value : Low = High	0.480	0.571	0.568	0.788	0.325
Control Mean - Low IQ	0.92	0.13	0.08	0.22	0.57
Control Mean - High IQ	0.97	0.12	0.14	0.21	0.52
Observations	2312	2176	2176	2176	2176

Estimates are obtained via OLS. The dependent variables are binary choice variables of intention to go to university, choose science major, engineering major, medicine and non-STEM major. Panel 1 presents short term results from the pooled sample, Panel 2 long-term results from Study 1. Standard errors are clustered at the school level and are reported in parentheses. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

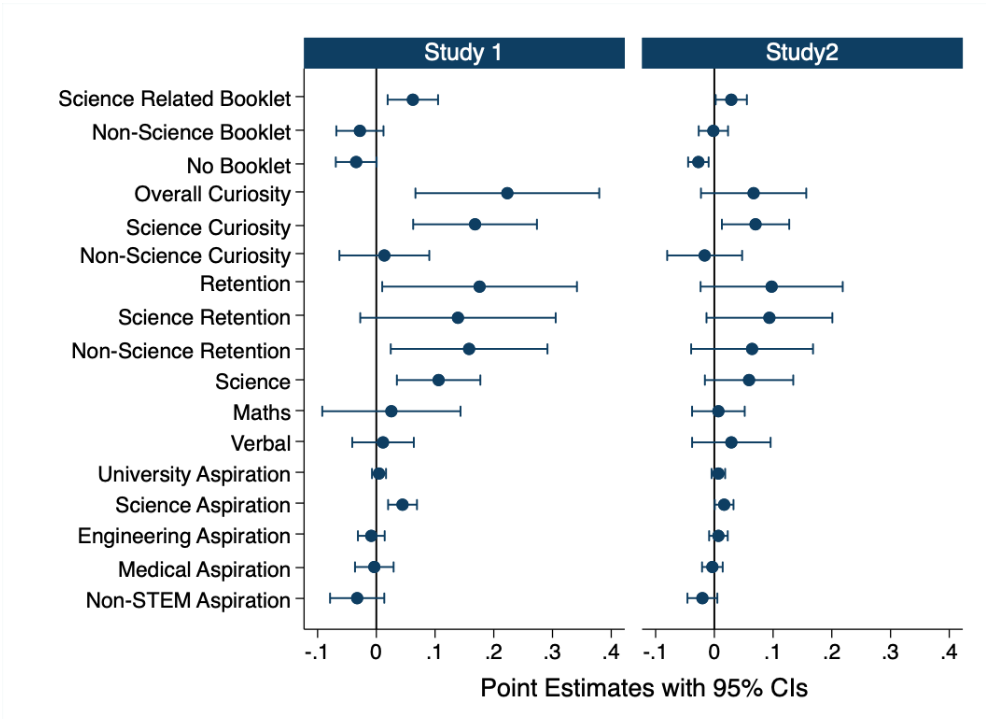
Table A7: Effect of Curiosity - IV Results

	Retention	Science Retention	Non-Science Retention	Science	Maths	Verbal
Science Curiosity	1.284**	1.125**	0.998**	0.798**	0.161	0.289
	(0.55)	(0.48)	(0.48)	(0.35)	(0.29)	(0.27)
Observations	9733	9733	9733	9733	9733	10002

The table presents the 2SLS results depicted in Figure 7 in the main text. The estimates are obtained from the two-stage least squares estimation of the effect of science curiosity on booklet retention and test scores, where science curiosity is instrumented with the binary treatment assignment. Standard errors are clustered at the school level. Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.Standard errors clustered at school level and are reported in parentheses.Outcome variable is the standardized score from second week test on the curiosity booklet knowledge. Control variables include dummy variable for male, dummy variable for middle school, age in months, IQ, baseline risk behavior, baseline survey measure of curiosity, baseline math and verbal scores. Models include strata fixed effects.

B Figures

Figure A1: Treatment Effects -by Study Sites



The figure depicts the estimated treatment effects and their 95% confidence intervals for all outcomes considered in the study. Standard errors are clustered at the school level. The vertical line indicates a treatment effect of 0. The first three outcomes are the choice of booklets, the following three are curiosity levels based on the experimental task, followed by the booklet performance (7-9), subject test scores (10-12), and educational aspirations(13-17). Covariates include gender, age, fluid IQ, risk tolerance, survey measure of curiosity, math and verbal scores as individual baseline characteristics, class size, the share of refugees, teacher gender, experience, and fluid IQ as baseline classroom and teacher characteristics. Grade and district fixed effects are also included.

C Survey Inventories

We provide some example questions from our student and teacher surveys below. The full inventory for both is available upon request.

Table A8: Student Survey Inventories

Inventory	Exemplary Items
4-point likert scale: completely agree, agree, disagree, completely disagree	
Curiosity	There are always questions on my mind.
	When I hear a word that I do not know, I am eager to learn it.
Scientific Curiosity	It is fun to break things into pieces to see what is inside.
	I never hesitate to ask questions.
Grit	Obstacles or setbacks may discourage me.
	I often set a goal but later choose to pursue a different one.
Impulsivity	I tend to say the first thing that comes to mind, without thinking about it.
	I interrupt people when they are talking.
Critical Thinking	It's important to understand other people's viewpoint on an issue.
	I usually check the credibility of the source of information before making judgements.

Table A9: Teacher Survey Inventories

Inventory	Exemplary Items
4-point likert scale: completely agree, agree, disagree, completely disagree	
Teaching Styles	I encourage my students to do research on topics they are interested in and discuss these topics with me. (Inquiry-based Pedagogy)
	It does not matter if there is noise in the classroom as long as the students are busy with something productive. (Modern Teaching)
	Punishment is necessary to create a disciplined class. (Extrinsic Motivation)
	Teachers should be serious and authoritative in their relationships with students. (Warmth)
Professional Satisfaction	I am very pleased to have chosen teaching as a profession.
Competence	It is difficult for me to communicate effectively with students.
Growth Mindset	Your intelligence is something that you can't change very much.
Critical Thinking	I sometimes find a good argument that challenges some of my firmly held beliefs.
Gender Stereotyping	Men have better judgment compared to women; hence they are better leaders.

D Implementation Items and Moments

Figure A2: Curious Classroom Toolkit

Figure A3: Various Classroom Activities

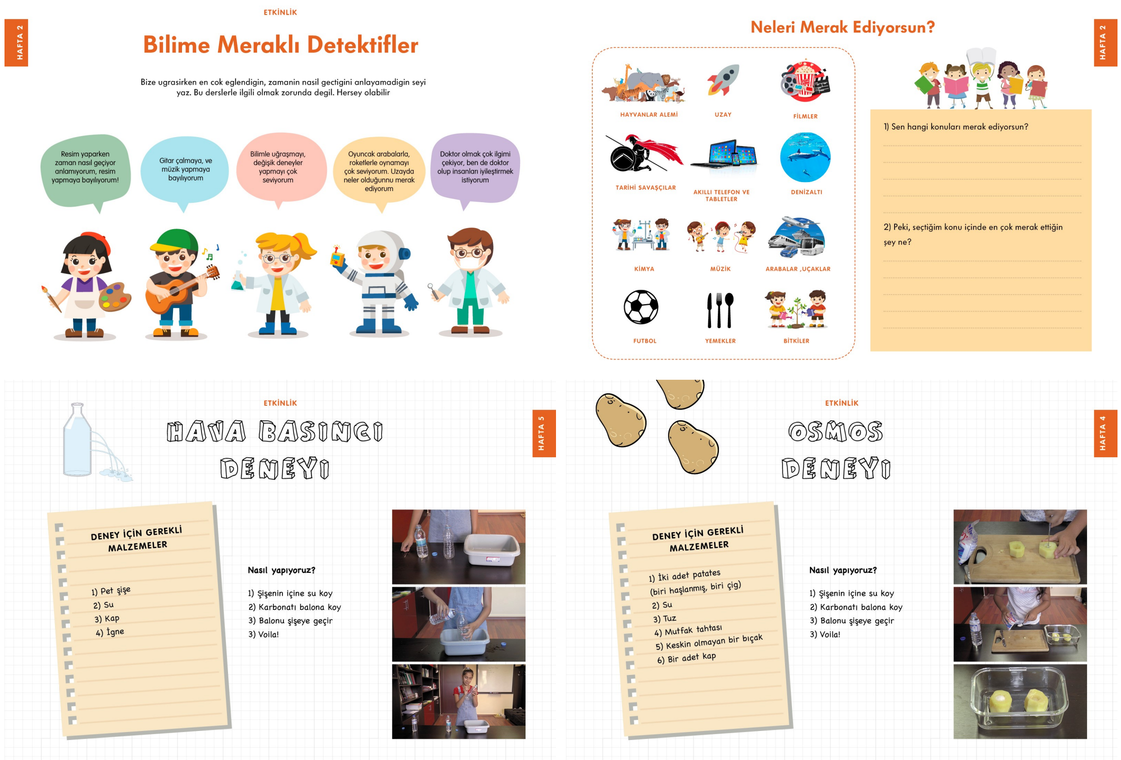


Figure A4: Creating Teachable Moments

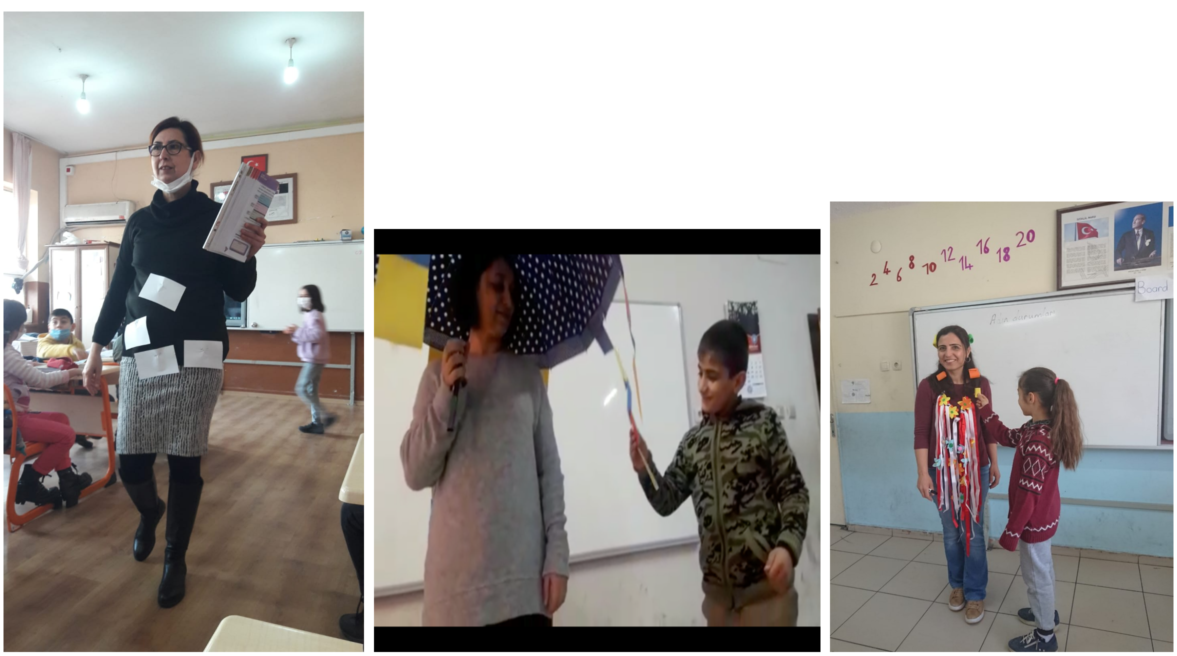


Figure A5: Examples of Children's Activities

E Instructions for Incentivized Games

A Curiosity Task

Hi everybody. We will play some fun games with you today. By playing these games, you will have a chance to earn gift tokens from us, with which you can get any gift you want from our gift bag [show the items in the gift bag]. The number of gifts you will receive will depend on your choices in these games. To get the gifts, you need to collect tokens, as each gift in our basket has a different token value. The more tokens you have, the more gifts you will be able to get at the end of our visit.

Each game has its own rules, and we will slowly explain all of them. But our main rule is discretion. You will need to make all our choices discretely, without showing anyone. Do you understand this rule? Excellent!

Now, see that we brought 8 booklets to you today. These booklets contain some incredible facts that most people do not know. [Start introducing them one by one]. This is the Space booklet. It has incredible facts in it. [Show animals], this is a booklet that contains astonishing facts about animals. [Go though each booklet in the same manner and always in the same order].

Now, we would like you to rank the booklets from most attractive to the least according to your own taste. Please type 1 beside the picture of the booklet that interests you the most, 2 for the second most interesting you find and keep going until 8, which would be the booklet least interesting to you. [Make sure everyone finishes their ranking and press continue before the next step].

Now, if you want, you can purchase one (and only one) of these booklets from us. How? Well, first, know that we give all of you 10 tokens. All of you have 10 tokens. You can use these tokens to get some of these nice stationery items from us. You can also get one booklet if you want. You don't have to get a booklet. You can convert all your tokens to gift items if you wish to. [Make sure children understand they do not have to purchase a booklet]. But if you do want a booklet, you need to first indicate which booklet you want to purchase on your tablet. Then you need to indicate how many of those 10 tokes you would be willing to give us back to purchase this booklet. You can say zero, meaning you don't want a booklet and want to convert all your tokens into gifts. Or you can say any number from 1 to 10.

But how do you really purchase a booklet? One of two things can happen in your classroom. You can be classroom type A or B. Let's see what happens in type A classrooms: Let's say student A decides to forgo 3 tokens, student B 5, and student C 7 tokens. Here is what we will do. We will pick a number from this bag. The bag contains folded little papers. In each paper, a number between 1 to 10 is written. [Show the black bag and show the little paper pieces]. The number we pull from this bag will be the price of a booklet for this classroom.

[Now, start giving the examples based on the 3 students above]. Let's say we picked number 8. Then we will look at everyone's decision of willingness to pay for their preferred booklet. Student A marks 3. She can't get the booklet she wants because the price is 8.

Instead, we will convert all her 10 tokens into gifts. The same goes for students B and C because their willingness to pay fell under the price of the booklet in this classroom.

But let's say we pick the number 5 instead of 8, so the price is 5. Student A still won't get a booklet and will receive 10 tokens worth of gifts. Student B, however, will give us her 5 tokens, get the booklet she wants and convert her remaining 5 tokens into gifts. What about student C? Well, she says she is willing to forgo 7 tokens but does she need to? NO. The price is 5, why should she? So we will get 5 tokens from her, give her the booklet she wants, and she will convert the remaining 5 tokens to gifts, just like student B.

What about a student who states zero willingness to pay? Well, she will not receive a booklet at any price. What about a student who states 10? She will certainly receive a booklet in the classroom type A.

What if your classroom is type B, which is much more likely as most classrooms will be type B. If your classroom is a type B, no matter how much you are willing to pay for a booklet, and no matter which booklet you prefer, a random half of the classroom will receive booklets, and the other half will not. We will pick half the students randomly from your class list.

Now, time to make decisions. First, tap the booklet you want to purchase. Don't forget there is an option that says "I do not want a booklet". You can tap that if you don't want a booklet. After making your choice, please tap the number of tokens you are willing to forgo to get the booklet you choose. [Make sure everyone makes their decisions and press continue.]

- Implementation, Type A (Market Price): Please pick a number from the black bag. Distribute the booklets accordingly.
- Implementation Type B (Half-Half): Please select the random half of the classroom using the class list and distribute the booklets only to them. Make sure every classroom has all 8 booklets.

B Risk and Ambiguity Games

Now we will play two games. [Type Game 1 and Game 2 on the board]. These two games are almost identical to each other. You will earn some gifts from these games. But you will collect the gifts from only one of the games, i.e., the gifts will not accumulate. We will pick one of these two games randomly for this classroom at the end of the visit, and you will get your gifts based on the decisions you make for that game. Now, let me explain the games Game 1: We will give you 5 tokens for this game. You can convert these tokens into small gifts in our bag [show all the gifts]. Now, think about a bucket [draw a bucket on the board]. You can put some of your tokens in this bucket if you want. You don't have to. If you don't, you have your 5 tokens, no problem. But what happens if you put some of your tokens in the bucket? Then, you draw a ball from this black bag [show the black bag]. There are two balls in this bag. One is yellow, and one is purple [show the balls]. The tokens you put in the bucket triple if you draw the yellow ball. You lose all the tokens you put in the bucket if you draw the purple ball. But not the ones you didn't put in the bucket. Tokens you don't put in the bucket are always safe.

Let's see some examples now: If you put none of your tokens in the bucket. What happens? NOTHING. You have 5 tokens. Let's say you put 1 token in the bucket. You have 4 safe ones left. Nothing happens to them. Then you draw a ball from the bag. If you draw the yellow ball, your 1 token becomes 3 tokens. Add to that your 4 safe ones. You now have 7 tokens. But what if your pick the purple ball. Then you lose that 1 token you put in the bucket, and you have 4 tokens. Now, let's say you put 2 tokens in the bucket. [Go on until you give the example of 5 tokens].

Now, decide how many tokens you want to put in the bucket. Please tap the number on your tablet and press continue.

Game 2: Now, we will play the second game. The second game is the same as the first game. You have 5 tokens, there is a bucket, and the tokens you put in the bucket triple if the yellow ball is drawn. They disappear if the purple ball is drawn. All the same. Except now, you don't know the colors of the balls in this new bag [pick the other bag, so children see this is not the same bag as in game 1]. Both balls can be yellow. In that case, you certainly win. Both balls can be purple, in which case you certainly lose. Or, one of them may be yellow, the other purple as in Game 1. The fact is, you do not know.

Now, please decide how many tokens you will put in the bucket. Please tap the number on your tablet and press continue.