# Explaining the Productivity Paradox

## Experimental Evidence from Educational Technology

RISE Annual Conference 2023

Andreas de Barros

University of California, Irvine
School of Education & Department of Economics (by courtesy)

de Barros, Andreas. (2023). Explaining the Productivity Paradox: Experimental Evidence from Educational Technology. (EdWorkingPaper: 23-853). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/mf5h-r010

andydebarros.bsky.social

To Avanti Fellows,
the Government of Haryana and Haryana's school administrators, teachers, and students,
Harvard University, J-PAL, MIT, MSDF, USAID, and my advisors:

**Thank you.**

# 1/ Introduction

# The productivity paradox remains unresolved—

> *You can see the computer age everywhere but in the productivity statistics.*
>
> Solow (1987)

# —nowadays, we just call it the *modern* productivity paradox.

> *We thus appear to be facing a redux of the Solow (1987) paradox:*
>
> *We see transformative new technologies everywhere but in the productivity statistics.*
>
> Brynjolfsson et al. (2017)

# EdTech: A prime example of why it is so hard to isolate the effect of technology

Does Educational Technology (EdTech) indeed serve a purpose of either **factor-augmenting** or **complementary technology?**

EdTech investments often lead to an **increase in instructional inputs**, instructional time.

Muralidharan et al. (2019); Angrist et al. (2022)

# EdTech: A prime example of why it is so hard to isolate the effect of technology

Does Educational Technology (EdTech) indeed serve a purpose of either **factor-augmenting** or **complementary technology?**

EdTech investments often lead to an **increase in instructional inputs**, instructional time.

Muralidharan et al. (2019); Angrist et al. (2022)

EdTech investments often lead to **substitution of instructional inputs.**

Bettinger et al. (2022); Ma et al. (2020)

Fabregas (2018); Jamison et al. (1981); Johnston and Ksoll (2022); Naik et al. (2020); Navarro-Sola (2019); Seo (2017)

# EdTech: A prime example of why it is so hard to isolate the effect of technology

Does Educational Technology (EdTech) indeed serve a purpose of either **factor-augmenting** or **complementary technology?**

EdTech investments often lead to an **increase in instructional inputs**, instructional time.

Muralidharan et al. (2019); Angrist et al. (2022)

EdTech investments often lead to **substitution of instructional inputs.**

Bettinger et al. (2022); Ma et al. (2020)

Fabregas (2018); Jamison et al. (1981); Johnston and Ksoll (2022); Naik et al. (2020); Navarro-Sola (2019); Seo (2017)

Most commonly: A **mixture of changes** in study time, substitution of teacher-led instruction, and adjustments to instructional technology.

Araya et al. (2019); Banerjee et al. (2007); Carrillo et al. (2010); Lai et al. (2015); Linden (2008); Taylor (2018)

# This study: What I do and what I find

An RCT to measure the causal effects of an EdTech program that encourages Indian teachers to **blend their instruction** with high-quality video materials.

- Largest cluster-randomized trial studying EdTech as a potential **complement to teaching**, largely shutting down other channels
- In comparison to a non-tech program; detailed obs. of mechanisms (1,500+ classroom obs.)
- In partnership with a state government, across 240 schools, 25k+ students

# This study: What I do and what I find

An RCT to measure the causal effects of an EdTech program that encourages Indian teachers to **blend their instruction** with high-quality video materials.

- Largest cluster-randomized trial studying EdTech as a potential **complement to teaching**, largely shutting down other channels
- In comparison to a non-tech program; detailed obs. of mechanisms (1,500+ classroom obs.)
- In partnership with a state government, across 240 schools, 25k+ students

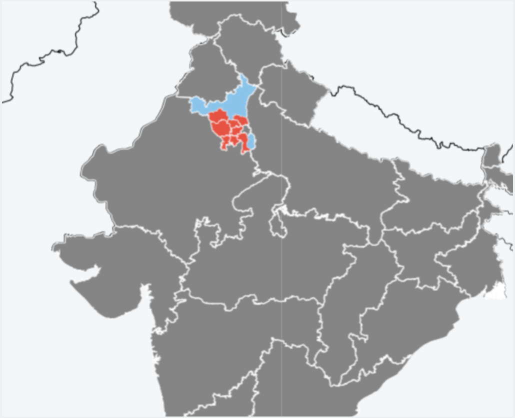
I find **negative effects** of the EdTech intervention on math, no effects on science learning, no effects of non-tech intervention, after 11 months.

- ITT effects of -0.15 SD in math, for the intervention promoting blended instruction
- Implementation failure ruled out; **negative effects on instruction, student attitudes**
- Consistent with adjustment costs as an explanation for the productivity paradox

# 2/ Context and intervention

# Our setting is the scale-up of a public-private partnership in Haryana's public secondary schools

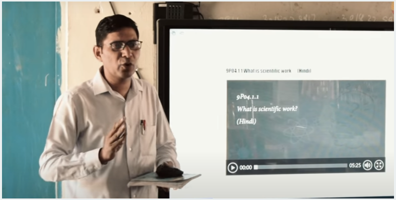
Figure 1: Location of the study



*Notes.* This figure depicts the state of Haryana (in light blue) and the eight districts selected for the study (highlighted in red).

# The intervention provides monthly, on-site coaching to promote blended instruction in math, science

Figure 2: Intervention components



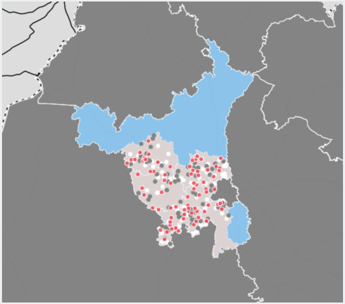

(a) Information and Communication Technology (ICT) infrastructure, video materials


(b) Workbooks, pedagogy

# **3/** Research design and analytical strategy

# We conduct a 3-arm, cluster-RCT, comparing the EdTech program against the non-tech alternative and a business-as-usual control group

Figure 3: Study schools by treatment status



*Notes.* This figure depicts the 240 study schools by experimental status. Red dots indicate ICT schools (T1), dark gray indicates workbook-only schools (T2), and white indicates control schools. Stratified randomization within triplets of matched schools. 10 re-randomizations to increase balance across T1, T2, and C, following a “min-max” strategy (cf. Banerjee et al., 2020; Bruhn and McKenzie, 2009).

▸ Sampling

# Our main outcome of interest is student learning in mathematics and science

**Main outcome**

- *Written assessments*: Group-administered; paper-based; ~60 multiple choice questions (two hours).
- Linked with a two-parameter Item Response Theory (IRT) model.

# Our main outcome of interest is student learning in mathematics and science

**Main outcome**

- *Written assessments*: Group-administered; paper-based; ~60 multiple choice questions (two hours).
- Linked with a two-parameter Item Response Theory (IRT) model.

**Secondary outcomes**

- *Sub-competencies in content domains*: 4 domains in math (e.g., algebra vs. geometry), 3 domains in science (biology, chemistry, physics).
- *Sub-competencies in cognitive domains:* Higher-order vs. lower-order thinking skills (HOTS/LOTS).

▸ 2PL IRT ▸ Distribution ▸ TIF

# Our main outcome of interest is student learning in mathematics and science

**Main outcome**

- *Written assessments*: Group-administered; paper-based; ~60 multiple choice questions (two hours).
- Linked with a two-parameter Item Response Theory (IRT) model.

**Secondary outcomes**

- *Sub-competencies in content domains*: 4 domains in math (e.g., algebra vs. geometry), 3 domains in science (biology, chemistry, physics).
- *Sub-competencies in cognitive domains:* Higher-order vs. lower-order thinking skills (HOTS/LOTS).

▸ 2PL IRT ▸ Distribution ▸ TIF

# We collect additional data guided by the program's Theory of Change

**Implementation fidelity**

- *Teaching and learning materials:* (1) Classroom observations; (2) student surveys; (3) backend data from the software.
- *Teacher training (offsite), monitoring and coaching (on-site):* (1) Attendance records; (2) an application documenting all work done by NGO staff.

# We collect additional data guided by the program's Theory of Change

**Implementation fidelity**

- *Teaching and learning materials:* (1) Classroom observations; (2) student surveys; (3) backend data from the software.
- *Teacher training (offsite), monitoring and coaching (on-site):* (1) Attendance records; (2) an application documenting all work done by NGO staff.

**Intermediate outcomes**

- *Instructional behaviors and quality*:
  (1) Classroom obs. (*Stallings, QUIP*); (2) teacher, student surveys
- *Student attitudes towards mathematics, science:*
  Student survey

▸ *QUIP* ▸ Student survey

# We collect additional data guided by the program's Theory of Change

**Implementation fidelity**

- *Teaching and learning materials:* (1) Classroom observations; (2) student surveys; (3) backend data from the software.
- *Teacher training (offsite), monitoring and coaching (on-site):* (1) Attendance records; (2) an application documenting all work done by NGO staff.

**Intermediate outcomes**

- *Instructional behaviors and quality*:
  (1) Classroom obs. (*Stallings, QUIP*); (2) teacher, student surveys
- *Student attitudes towards mathematics, science:*
  Student survey

▸ *QUIP* ▸ Student survey

# We estimate intent-to-treat effects with OLS regressions

$$Y_{isvr} = \alpha + \sum_{k=1}^{2} \beta_k T_{ksvr} + \boldsymbol{X}_{isvr} + \phi_r + \epsilon_{isvr} \tag{1}$$

- $\beta_k$ captures the intent-to-treat (ITT) effect of each program variant $k$, for follow-up round $t$, for student $i$ in school $s$, in village $v$, and randomization stratum $r$.

# We estimate intent-to-treat effects with OLS regressions

$$Y_{isvr} = \alpha + \sum_{k=1}^{2} \beta_k T_{ksvr} + \boldsymbol{X}_{isvr} + \phi_r + \epsilon_{isvr} \tag{1}$$

- $\beta_k$ captures the intent-to-treat (ITT) effect of each program variant $k$, for follow-up round $t$, for student $i$ in school $s$, in village $v$, and randomization stratum $r$.
- $t$ refers to a baseline ($t = 1$), and an endline after 11 months ($t = 2$).
- To increase precision: $\boldsymbol{X}_{isvr}^{t=1}$ as covariates.
- $\boldsymbol{X}_{isvr}^{t=1}$: vector of baseline controls selected through a Lasso procedure, from $Y_{isvr}^{t=1}$, student age, gender, school-level admin. data ("DISE"), and village-level census data.
- Randomization strata fixed effects $\phi_r$ included. Standard errors clustered at the school level.

▸ Covariates

# The sample is similar to India's population of gov. schools; randomization led to three similar groups

**Representativeness**

- I compare study schools with the **universe of India's government secondary schools** (on census village characteristics, school characteristics, National Achievement Survey, State board exams).
- Study schools are positively selected within the state, but **fairly representative of India**.

# The sample is similar to India's population of gov. schools; randomization led to three similar groups

**Representativeness**

- I compare study schools with the **universe of India's government secondary schools** (on census village characteristics, school characteristics, National Achievement Survey, State board exams).
- Study schools are positively selected within the state, but **fairly representative of India**.

**Balance**

- **No differences in attrition**: 24 percent to endline; balanced across the three groups.
- **Baseline balance**, on time-invariant observables (1/24 tests of village-, 2/27 tests of school-level characteristics, no difference in board exams or student demographics).
- **Baseline imbalance for test scores for one group comparison** (Workbook vs control); included as control, does not affect the results in robustness checks.

▸ Balance

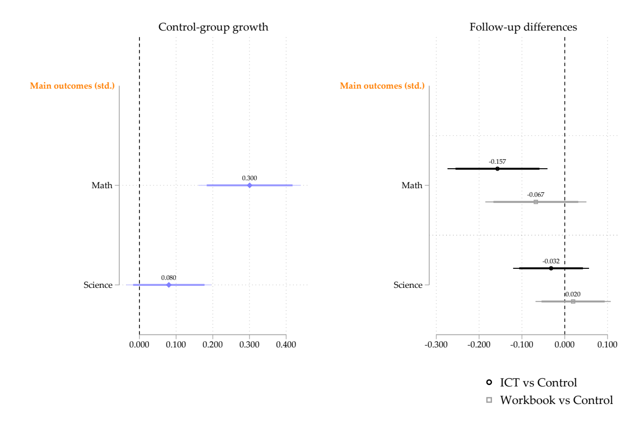
# Key components of the intervention were implemented as intended and taken up well

1. ICT infrastructure upgrades implemented as planned. ▸ Figure
2. Videos installed, used (slightly less in math). ▸ Figure
3. Workbooks distributed and used, outside the classroom. ▸ Figure
4. Limited use of in-class exercises, peer learning, in both groups. ▸ Figure
5. Limited on-site coaching (anecdotal).

# 4/ Results

# After 11 months of the ICT intervention, I find neg. effects on math, no effects on science test scores

Figure 4: Intent-to-treat (ITT) effects on student learning




*Notes.* All estimations include randomization strata fixed effects and a vector of school- and village-level covariates, selected via LASSO. Horizontal bars show confidence intervals (s.e.s clustered at the school level).
*Sample.* 18,562 grade-9 and grade-10 students.

# Following a registered Pre-analyis Plan, I report additional results for effects on test scores

1. Results are similar for higher- vs lower-order skills. ▸ Figure
2. Grade-9 students and weaker students are affected *only slightly* more negatively (in math). ▸ Math ▸ Science
3. Math effects are *only slightly* larger for below-level material; otherwise uniform across domains. ▸ Figure

# Following a registered Pre-analyis Plan, I report additional results for effects on test scores

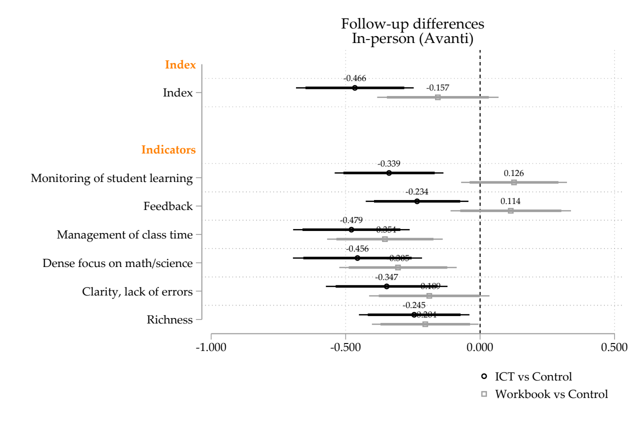
1. Results are similar for higher- vs lower-order skills. ▸ Figure
2. Grade-9 students and weaker students are affected *only slightly* more negatively (in math). ▸ Math ▸ Science
3. Math effects are *only slightly* larger for below-level material; otherwise uniform across domains. ▸ Figure
4. There is *limited* heterogeneity across districts. ▸ Math ▸ Science

# Following a registered Pre-analyis Plan, I report additional results for effects on test scores

1. Results are similar for higher- vs lower-order skills. ▸ Figure
2. Grade-9 students and weaker students are affected *only slightly* more negatively (in math). ▸ Math ▸ Science
3. Math effects are *only slightly* larger for below-level material; otherwise uniform across domains. ▸ Figure
4. There is *limited* heterogeneity across districts. ▸ Math ▸ Science

# I find negative effects of the ICT intervention on the quality of instruction students receive

Figure 5: ITT effects on instructional quality

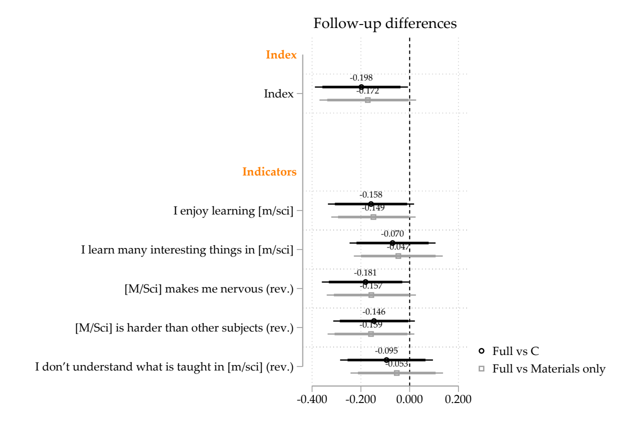



*Notes.* All estimations include randomization strata fixed effects and a vector of school- and village-level covariates, selected via LASSO. Horizontal bars show confidence intervals (s.e.s clustered at the school level).
*Sample.* 1,343 classroom observations in mathematics and science.

▸ Adjusted ▸ Practices (Stallings)

# For both treatment arms, I find negative effects on student perceptions, attitudes

Figure 6: ITT effects on student perceptions, attitudes




*Notes.* All estimations include randomization strata fixed effects and a vector of school- and village-level covariates, selected via LASSO. Horizontal bars show confidence intervals (s.e.s clustered at the school level).
*Sample.* 1,214 student interviews.

▸ Unstd.

# 5/ Conclusion

# Conclusion

- I observed **negative effects** on instructional quality, student attitudes; **negative to null effects** on student learning.

# Conclusion

- I observed **negative effects** on instructional quality, student attitudes; **negative to null effects** on student learning.
- My findings suggest **introducing new Information and Communication Technologies may disrupt**, lead to substantial adjustment costs—at least in the short run.

# Conclusion

- I observed **negative effects** on instructional quality, student attitudes; **negative to null effects** on student learning.
- My findings suggest **introducing new Information and Communication Technologies may disrupt**, lead to substantial adjustment costs—at least in the short run.
- I also observed negative effects on student attitudes for the non-tech intervention, suggesting *any* reform may explain at least part of the story.

# Conclusion

- I observed **negative effects** on instructional quality, student attitudes; **negative to null effects** on student learning.
- My findings suggest **introducing new Information and Communication Technologies may disrupt**, lead to substantial adjustment costs—at least in the short run.
- I also observed negative effects on student attitudes for the non-tech intervention, suggesting *any* reform may explain at least part of the story.
- It will be worthwhile observing long-term effects—are there positive returns on these costly adjustments?

# Conclusion

- I observed **negative effects** on instructional quality, student attitudes; **negative to null effects** on student learning.
- My findings suggest **introducing new Information and Communication Technologies may disrupt**, lead to substantial adjustment costs—at least in the short run.
- I also observed negative effects on student attitudes for the non-tech intervention, suggesting *any* reform may explain at least part of the story.
- It will be worthwhile observing long-term effects—are there positive returns on these costly adjustments?
- Currently, (post) Covid-19 studies on EdTech should invest in additional efforts to isolate mechanisms and be cautious about the potential disruption of instruction.

# Impact and next steps

- **The NGO pivoted** and turned the math intervention into a remote, after-school program. It tries to improve the science intervention over time.

# Impact and next steps

- **The NGO pivoted** and turned the math intervention into a remote, after-school program. It tries to improve the science intervention over time.
- With **funding from USAID and MSDF**, I study the effects of both interventions in a new cohort of children.

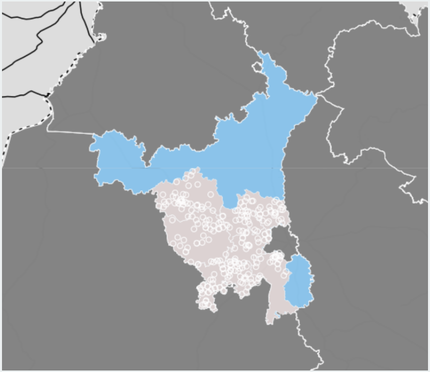
# Impact and next steps

- **The NGO pivoted** and turned the math intervention into a remote, after-school program. It tries to improve the science intervention over time.
- With **funding from USAID and MSDF**, I study the effects of both interventions in a new cohort of children.
- The Haryana government invited me to **join an expert panel** that consulted on the state's provision of a tablet-based EdTech program to 1m+ public school children.

# 6/ Appendix

# We purposely sampled 240 secondary schools, following an infrastructure survey

Figure A1: Sample of 240 government senior secondary schools

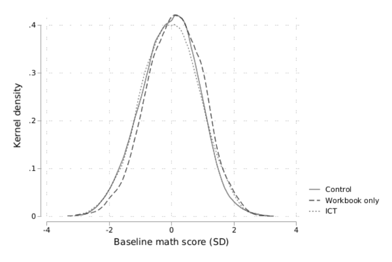


*Notes.* This figure depicts the 240 senior secondary government schools purposely selected for the study.

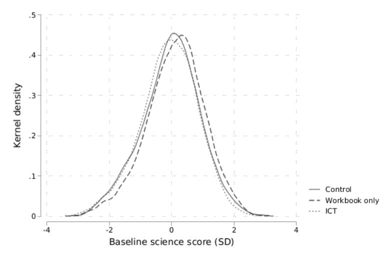
◂ Back

# The written tests do not suffer from ceiling or floor effects

Figure A2: Empirical distribution of test scores, by subject




(a) Mathematics



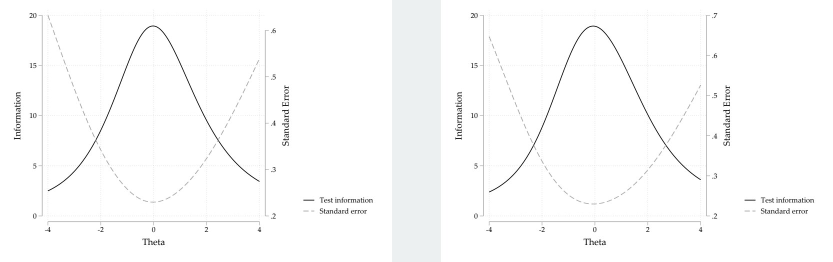

(b) Science

*Notes.* This figure provides the empirical distribution of test scores, as per 2PL IRT models, for students. Each panel shows kernel density plots, by experimental group, at baseline.

Back

# The written tests measure with high levels of precision, across the ability distribution

Figure A3: Test information functions (TIFs)




(a) Mathematics

(b) Science

*Notes.* This figure provides the test information functions, and corresponding standard errors of measurement, for the mathematics and science tests, as per 2PL IRT models.

◂ Back

# The two-parameter logistic IRT model

The two-parameter logistic (2PL) model predicts the probability of correctly answering each test question (or “item”) $i$, given student’s $j$ ability $\theta$ and two item parameters: item discrimination $a$ and item difficulty $b$.

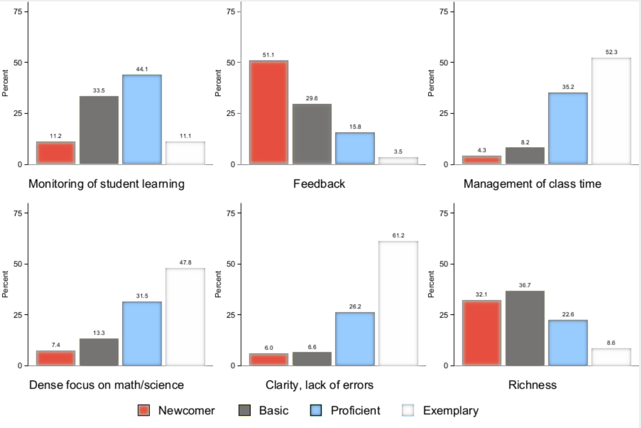
$$P_{ij}\left(\theta_j, b_i, a_i\right) = \frac{\exp\left[a_i\left(\theta_j - b_i\right)\right]}{1 + \exp\left[a_i\left(\theta_j - b_j\right)\right]} \tag{2}$$

- Assigns different weights to individual test questions.
- Allows for linking of ability estimates onto a common scale, across assessment rounds and grades.
- Provides information on measurement error across the ability distribution.
- Estimated via marginal max. likelihood (MML).

Back

# We measure instructional behaviors, quality with the *QUIP* classroom observation instrument

Figure A4: *QUIP* framework

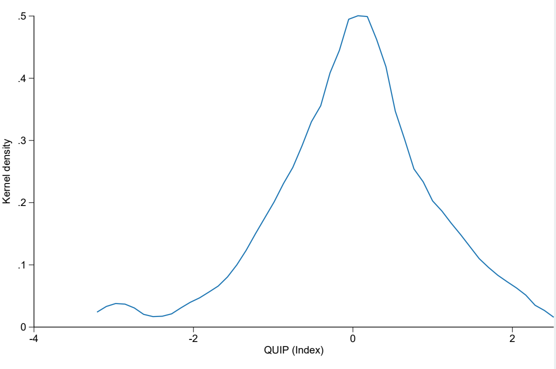



*Notes.* This figure reports on the distribution of QUIP scores, showing histograms for each QUIP element.

Back

# We measure instructional behaviors, quality with the *QUIP* classroom observation instrument (ctd.)

Figure A5: *QUIP* framework




*Notes.* This figure provides a kernel density plot of the QUIP index score. “Index” refers to the inverse covariance matrix-weighted aggregate across the six elements.

Back

# We measure student attitudes towards mathematics with one-on-one interviews

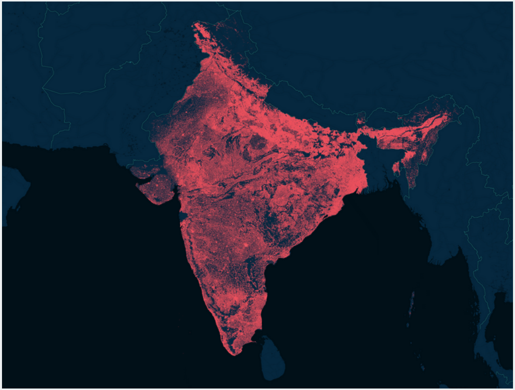
We calculate an inverse covariance matrix-weighted index, for the following items

- Student enjoys learning math
- Student finds math easy to understand
- Math makes the student nervous (reversed)
- Student finds math harder than other subjects (reversed)

Back

# Through GIS, we link India’s EMIS to village-level censuses and satellite-recorded nightlights data

Figure A6: India’s schools

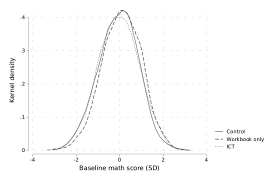


*Notes.* This figure shows all registered schools as per India’s education management information system (as of 2017), and their locations (as of 2020). Village-level census tracts are for 2011. Satellite data comes from SHRUG.

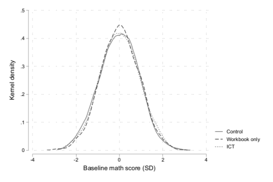
Back

# Results are robust to dropping the seven most severely imbalanced randomization triplets

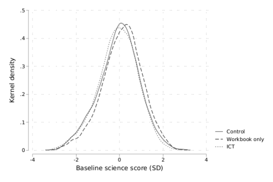
Figure A7: Balance on baseline test scores before/after dropping imbalanced strata.



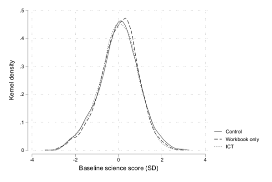

(a) Mathematics




(b) Math (adj.)



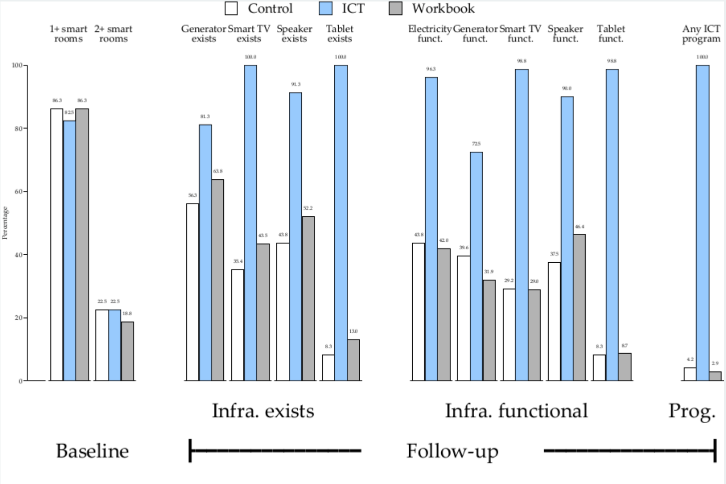

(c) Science




(d) Science (adj.)

Back

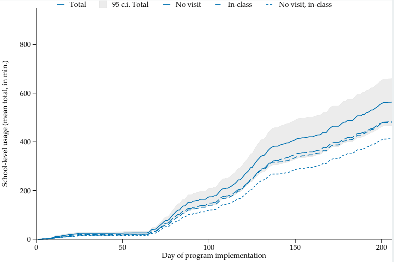
# ICT infrastructure installed and functional—the other schools do not have ICT programs



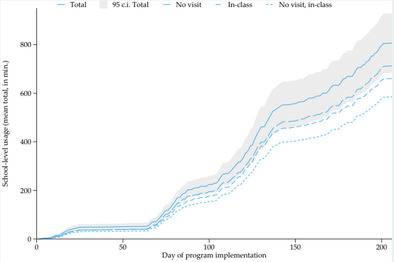

Back

# Teachers use the video materials—including on days without school visits

Figure A9: Video usage, by subject (in minutes)




(a) Math

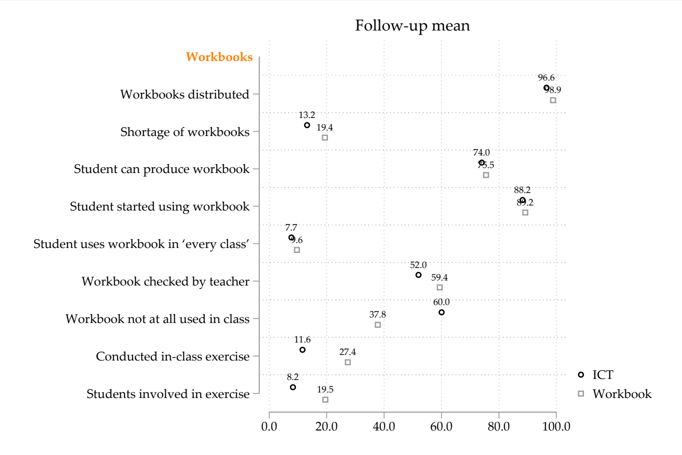



(b) Science

Back

# Workbooks were distributed and they are used (outside the classroom)

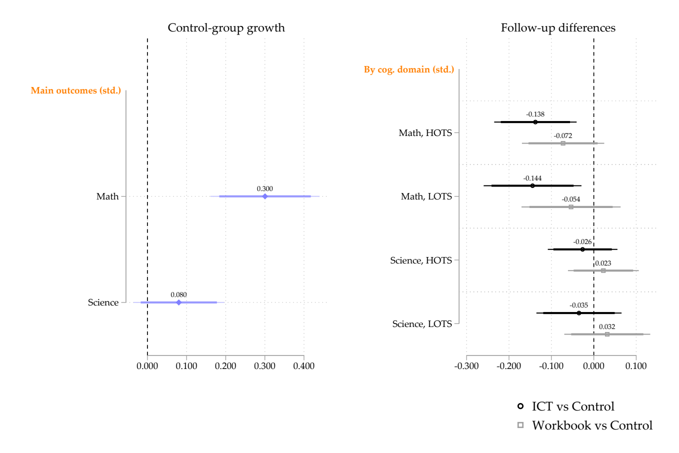
Figure A10: Availability and use of workbooks, in-class exercises, by treatment group




Back

# Effects are similar across higher-order and lower-order skills

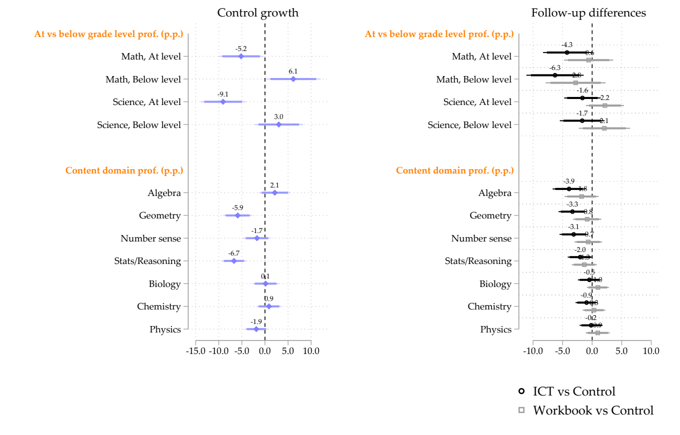
Figure A11: ITT effects on student learning by cognitive domain




Back

# Effects are similar across curricular grade-levels and across content domains

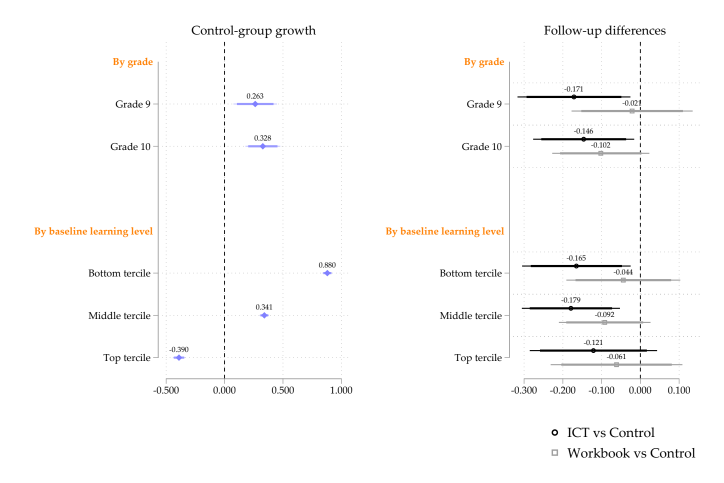
Figure A12: ITT effects on student learning by content domain




Back

# Effects for student subgroups, in math

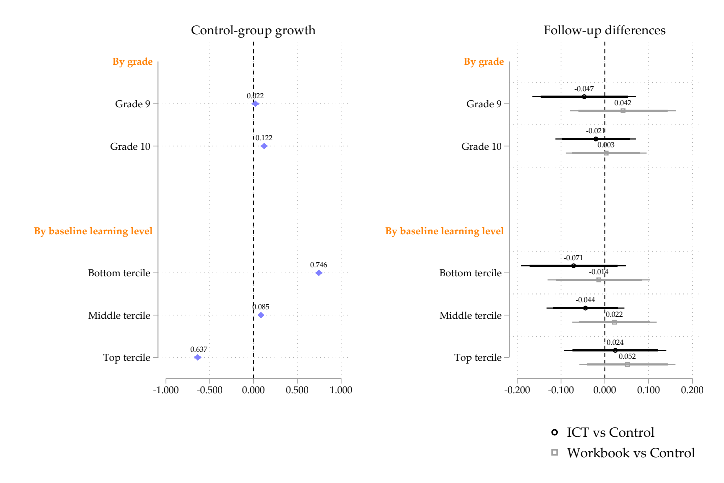
Figure A13: Heterogeneous ITT effects on math learning




Back

# Effects for student subgroups, in science

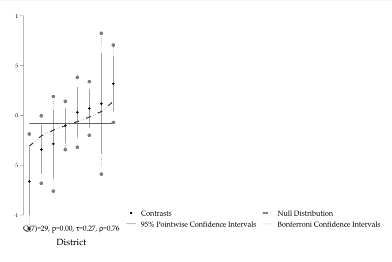
Figure A14: Heterogeneous ITT effects on science learning



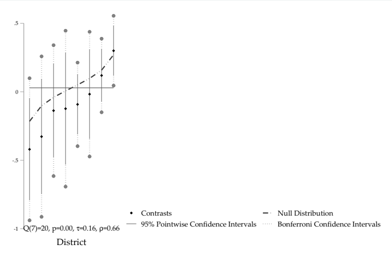

Back

# Test scores: Effects by district, in math

Figure A15: Heterogeneous ITT effects on math learning




(a) ICT vs Control

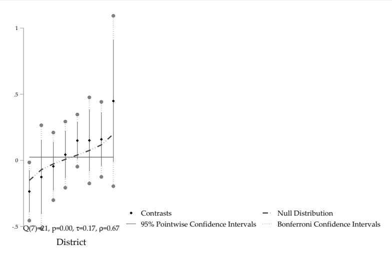



(b) Workbook vs Control

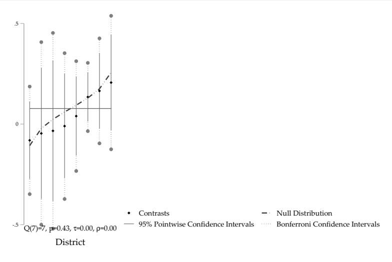
Back

# Test scores: Effects by district, in science

Figure A16: Heterogeneous ITT effects on science learning




(a) ICT vs Control

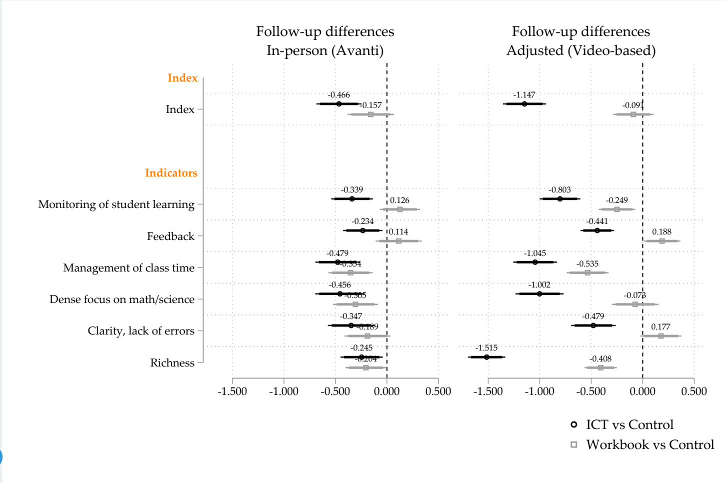



(b) Workbook vs Control

◂ Back

# NGO-administered ratings provide an upper bound for effects on instructional quality

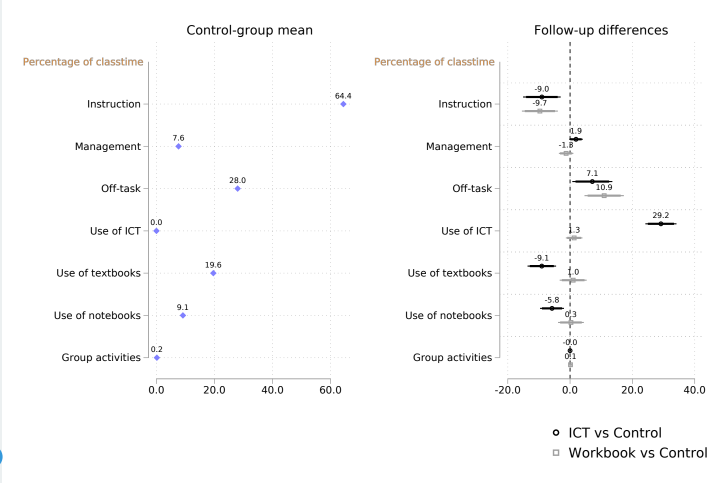
Figure A17: ITT effects on instructional quality (adjusted)




Back

# The ICT intervention increased EdTech usage; both interventions reduced time-on-task

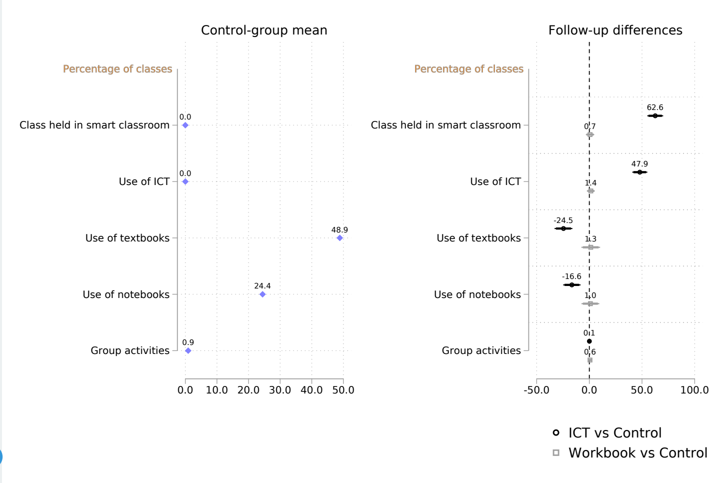
Figure A18: ITT effects on teaching behaviors




Back

# The ICT intervention increased EdTech usage; both interventions reduced time-on-task

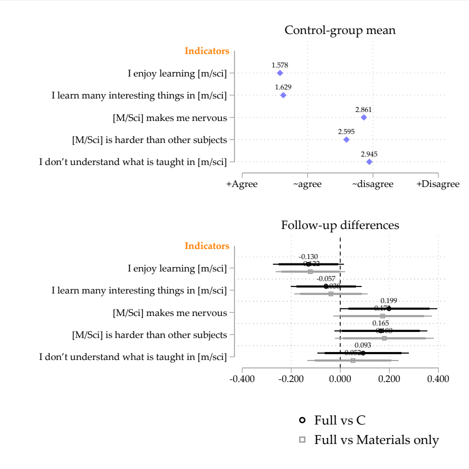
Figure A19: ITT effects on teaching behaviors




Back

# Effects on student perceptions, attitudes (unstandardized)

Figure A20: ITT effects on student perceptions, attitudes




Back

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

Araya, R., Arias Ortiz, E., Bottan, N.L., Cristia, J.P., 2019. Does Gamification in Education Work? Experimental Evidence from Chile. Working Paper IDB-WP-982. Inter-American Development Bank. Washington, D.C.

Banerjee, A.V., Chassang, S., Montero, S., Snowberg, E., 2020. A Theory of Experimenters: Robustness, Randomization, and Balance. American Economic Review 110, 1206–1230.

Banerjee, A.V., Cole, S., Duflo, E., Linden, L., 2007. Remedying Education: Evidence from Two Randomized Experiments in India. The Quarterly Journal of Economics 122, 1235–1264.

Bettinger, E., Fairlie, R., Kapuza, A., Kardanova, E., Loyalka, P., Zakharov, A., 2022. Diminishing Marginal Returns to Computer-Assisted Learning. Journal of Policy Analysis and Management , pam.22442.

Bruhn, M., McKenzie, D., 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. American Economic Journal: Applied Economics 1, 200–232.

Brynjolfsson, E., Rock, D., Syverson, C., 2017. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. Technical Report w24001. National Bureau of Economic Research. Cambridge, MA.

Carrillo, P.E., Onofa, M., Ponce, J., 2010. Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador. Technical Report IDB-WP-223. Inter-American Development Bank. Washington, D.C.

Fabregas, R., 2018. Broadcasting Human Capital? The Long Term Effects of Mexico's Telesecundarias.

Jamison, D.T., Searle, B., Galda, K., Heyneman, S.P., 1981. Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement. Journal of Educational Psychology 73, 556–567.

Johnston, J., Ksoll, C., 2022. Effectiveness of interactive satellite-transmitted instruction: Experimental evidence from Ghanaian primary schools. Economics of Education Review 91, 102315.

Lai, F., Luo, R., Zhang, L., Huang, X., Rozelle, S., 2015. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. Economics of Education Review 47, 34–48.

Linden, L.L., 2008. Complement or substitute? The effect of technology on student achievement in India. Working Paper 17. The World Bank. Washington, D.C.

Ma, Y., Fairlie, R., Loyalka, P., Rozelle, S., 2020. Isolating the "Tech" from EdTech: Experimental Evidence on Computer Assisted Learning in China. Technical Report w26953. National Bureau of Economic Research. Cambridge, MA.

Muralidharan, K., Singh, A., Ganimian, A.J., 2019. Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. American Economic Review 109, 1426–1460.

Naik, G., Chitre, C., Bhalla, M., Rajan, J., 2020. Impact of use of technology on student learning outcomes: Evidence from a large-scale experiment in India. World Development 127, 104736.

Navarro-Sola, L., 2019. Secondary School Expansion through Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria.

Seo, H.K., 2017. Do School Electrification and Provision of Digital Media Deliver Educational Benefits? First-year Evidence from 164 Tanzanian Secondary Schools. Working Paper E-40308-TZA-2. International Growth Centre. London.

Solow, R.M., 1987. We'd better watch out. New York Times Book Review .

Taylor, E.S., 2018. New Technology and Teacher Productivity.