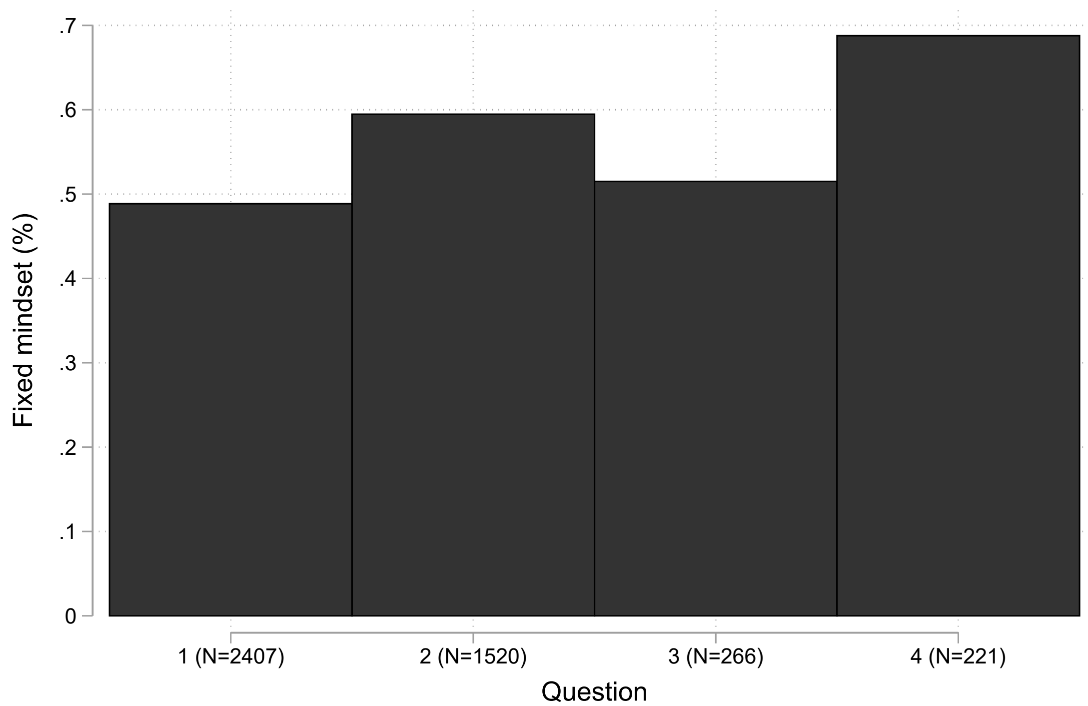
Supplementary Information for 'Student Mindsets Matter: Experimental Evidence from a Growth Mindset Intervention in Brazil'

Appendix A. Additional details on growth mindset surveys and messages

This Appendix provides additional details on growth mindset surveys and sample SMS sequences by treatment arm.

Figure A1 displays the fraction of fixed mindset students based on each survey question. We can see that the fraction of answers was relatively homogeneous across questions. The differences in sample sizes across them reflects fast declining response rates in these SMS surveys (see Lichand et al., 2024a).

Figure A.1. Fraction of fixed mindset answers for each survey question



Note: This Figure shows the fraction of answers sorted as fixed mindset by question sent to students. In the x-axis we also report the number of messages answered.

Next, we show sample sequences of messages sent to students and examples of their answers.

Figure A.2. Sample SMS sequence for dynamic complementarities

Figure A.3. Sample SMS sequence for high returns to effort

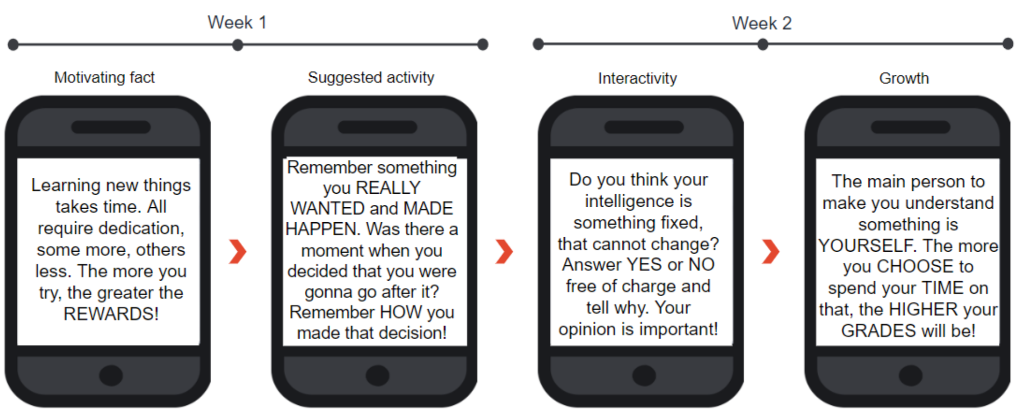


Figure A.4. Sample SMS sequence for low effort costs

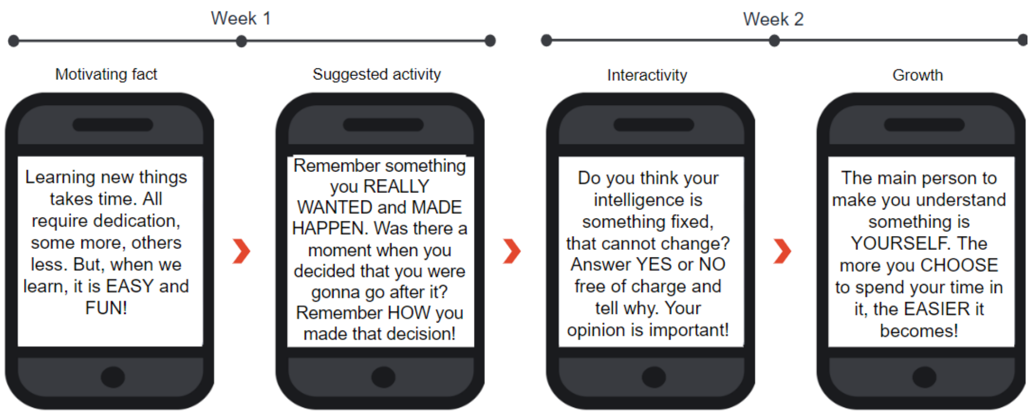


Figure A.5. Sample SMS sequence for risk-taking

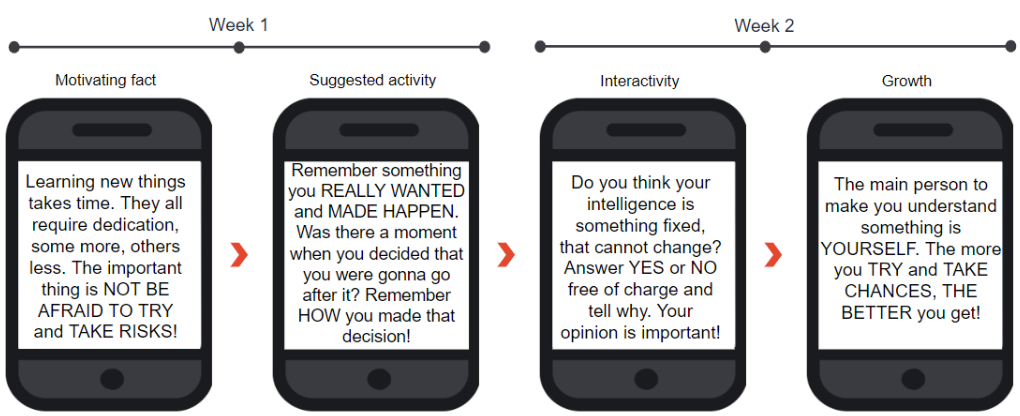


Figure A.6. Sample SMS sequence for future orientation

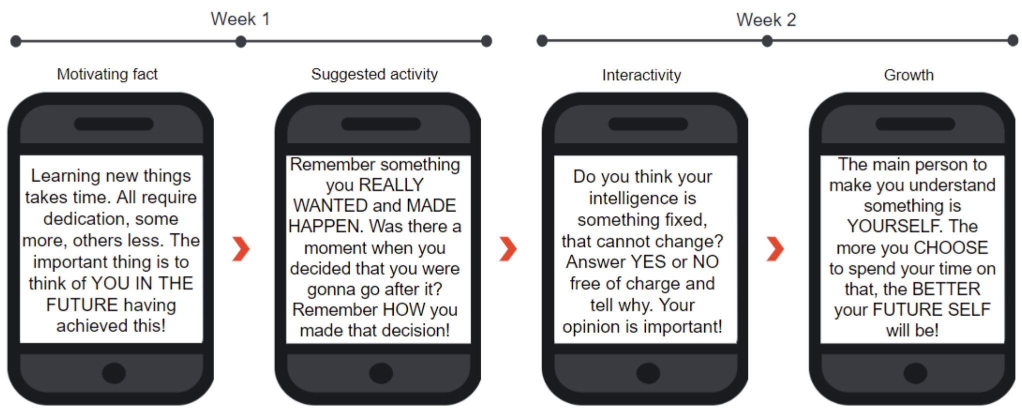
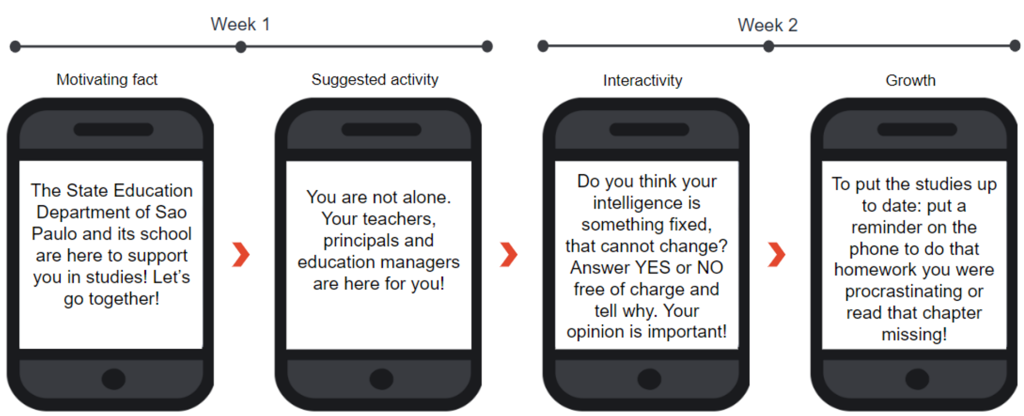


Figure A.7. Sample SMS sequence for placebo



Appendix B. Results for the additional experiment

Concurrently to the experiment described in the main text, we conducted an additional experiment that mirrored the former in terms of its cluster-randomized assignment and of the specifics of each treatment arm; the only difference being that, in that experiment, messages about each sub-component of growth mindset were *embedded in practical suggestions* to help students navigate the challenges of being back to school after a prolonged period of remote learning. This choice was motivated by concerns of the Education Secretariat and of the implementing partner that, at the time, students might be struggling so much with sizeable cumulative learning losses that such practical suggestions might be key to engage them with the content of the text messages. Because the content of the SMS messages of this additional experiment was not pre-registered, we analyze the results of the two experiments separately.

For completeness, we include all results for this additional experiment in this Appendix. Table B1 shows balance tests for the additional experiment. Table B2 shows estimates of the treatment effects on effort. Table B3 shows treatment effects for students' learning. Generally, we find small and non-significant effects of the treatment on the outcomes of interest. Table B4 shows that results do not change much controlling for school size. Results are that embedding growth mindset content amidst practical suggestions ended up making all treatment arms look more similar, muddling the specific mechanisms of interest. Worse, these messages ended up being overall less engaging to students – messages with embedded content were found to be significantly less effective and, if anything, slightly detrimental to learning outcomes.

Table B.1—Descriptive statistics and balance tests for the additional experiment

	Averages							Pooled vs.	= across
	T1	T2	T3	T4	T5	T6	Control	Control (p-value)	all groups (p-value)
Panel A: Availability of information									
Missed baseline std. test	0.70	0.69	0.70	0.69	0.69	0.69	0.71	0.18	0.22
Panel B: Students' characteristics									
Male	0.51	0.52	0.51	0.50	0.51	0.51	0.51	0.45	0.15
Non-white students	0.27	0.27	0.29	0.28	0.29	0.29	0.29	0.88	0.12
<1 MW	0.32	0.31	0.33	0.31	0.33	0.33	0.32	0.55	0.43
1-4 MW	0.34	0.32	0.33	0.32	0.32	0.34	0.34	0.43	0.57
4-7 MW	0.12	0.13	0.12	0.13	0.12	0.14	0.13	0.25	0.32
8-11 MW	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.88	0.95
Schools reopened	0.51	0.49	0.50	0.48	0.48	0.51	0.53	0.75	0.91
Panel C: Baseline outcomes									
Avg. Math report card grade (2020)	5.85	5.95	5.90	5.87	5.92	5.87	5.88	0.40	0.08
Avg. Portuguese report card grade (2020)	5.97	6.09	6.00	6.02	6.02	6.07	5.98	0.12	0.15
Avg. Math attendance (2020)	0.92	0.92	0.92	0.92	0.92	0.92	0.91	0.18	0.97
Avg. Portuguese attendance (2020)	0.91	0.92	0.92	0.91	0.92	0.91	0.90	0.45	0.35
All means equal zero (p-value)								0.32	0.58
N							49,228		
Municipality fixed-effects								yes	yes

Note: This Table show means of variables of interest for each treatment arm and the control group . P-values computed using randomization strata fixed effects and with standard errors clustered at the classroom level. P-value for the joint hypothesis that all differences equal zero based on a chi-squared statistic on a multinomial logit model. Data on students' gender, race, Q4/2020 report card grades and attendance comes from administrative records. Income brackets based on the Brazilian 2015 minimum wage, which was R$ 788 at that time, equivalent to approximately US$ 209 back then. Income data is not routinely collected by the Education Secretariat, but was collected by Lichand et al. (2022) for a similar sample in 2016. We impute income brackets for each student using a Poisson model, joint with multiple-imputation methods, based on student characteristics, trained in the 2016 dataset.

Table B.2—Treatment effects of the additional experiment on student effort

	(1) Cumulative minutes by week 20	(2) Math attendance	(3) Portuguese attendance
Panel A: Pooled treatment x control			
Mindset treatment	6.89	-0.001	0.002
	(12.29)	(0.002)	(0.002)
Panel B: Separate treatment arms x control			
Dynamic complementarities	19.04	0.001	0.004
	(14.31)	(0.003)	(0.003)
High returns to effort	20.22	-0.001	-0.001
	(17.12)	(0.002)	(0.002)
Low effort costs	18.55	-0.001	0.002
	(18.02)	(0.002)	(0.002)
Future orientation	20.22	-0.001	0.003
	(18.17)	(0.002)	(0.002)
Risk-taking	12.11	-0.001	-0.004
	(16.54)	(0.003)	(0.003)
Placebo	18.75	-0.001	-0.002
	(14.52)	(0.003)	(0.003)
N	191,340	128,323	
Control mean	385.14	0.90	0.90
Municipality fixed-effects	yes	yes	yes
Quarter fixed-effects	yes	yes	yes

Note: This Table shows ITT estimates pooled treatment effects and placebo estimates on effort measures. We show results for three different variables: the cumulative number of minutes spent on the Education Secretary online platform, and the recorded attendance for Portuguese and math classes during the first and second quarter of 2020. Treatment effects are measured in minutes and as a difference of proportions, respectively. We also report the control mean and the r-squared for regressions. All estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if $p<0.1$, ** $p<0.05$ and *** if $p<0.01$.

Table B.3—Treatment effects of the additional experiment on learning outcomes

	(1) Math report card grade	(2) Portuguese report card grade	(3) Standardized scores
Panel A: Pooled treatment x control			
Mindset treatment	0.021*	0.014	-0.014
	(0.009)	(0.009)	(0.012)
Panel B: Separate treatment arms x control			
Dynamic complementarities	0.015	0.005	-0.017
	(0.011)	(0.011)	(0.012)
High returns to effort	0.026*	0.011	0.026
	(0.010)	(0.010)	(0.010)
Low effort costs	-0.008	0.001	-0.008
	(0.013)	(0.009)	(0.013)
Future orientation	0.026*	0.014	-0.026*
	(0.012)	(0.011)	(0.012)
Risk-taking	0.006	0.011	0.006
	(0.004)	(0.012)	(0.004)
Placebo	0.040***	0.035***	-0.013
	(0.011)	(0.011)	(0.015)
N	128,323		78,976
Δ std. scores Q1	0.19	0.17	0.13
Municipality fixed-effects	yes	yes	yes
Quarter fixed-effects	yes	yes	yes
Inverse probability weighting	no	no	yes

Note: This Table shows ITT estimates pooled treatment effects, placebo estimates, and separate estimates for each treatment arm on learning outcomes. We show results for three different variables: math and Portuguese report card grades and standardized scores. For the latter, we average math and Portuguese scores. All treatment effects are measured in control group standard-deviaitons.. We also report the control mean and the r-squared for regressions.. All estimates include municipality and quarter fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if p<0.1, ** p<0.05 and *** if p<0.01.

Appendix C. Additional results

This Appendix compiles several additional results. Table C1 analyzes selection among students who took up standardized tests, documenting that the treatments did *not* induce selective non-response. Next, Table C2 presents descriptive statistics for the sub-sample of students who took all AAP tests, documenting that the experiment remains balanced based on this sample restriction.

Table C.1—Selective non-response tests

	Participated in any test	Participated in all tests
Panel A: Pooled treatment x control		
Pooled treatment	0.019	0.007
	(0.017)	(0.013)
Placebo	0.004	-0.0001
	(0.023)	(0.017)
p-value diff. [pooled - placebo]	0.40	0.54
Panel B: Separate treatment arms x control		
Dynamic complementarities	0.010	0.009
	(0.024)	(0.018)
High returns to effort	0.018	0.001
	(0.022)	(0.016)
Low effort costs	0.004	-0.001
	(0.023)	(0.017)
Future orientation	0.035	0.017
	(0.022)	(0.017)
Risk-taking	0.016	0.004
	(0.021)	(0.016)
Control Average	0.31	0.17
p-value (F test)	0.66	0.51
Municipality fixed-effects	yes	yes
N	108,345	
R^2	0.06	0.06

Note: This Table shows ITT estimates pooled treatment effects, placebo estimates, and separate estimates for each treatment arm on participation in standardized test scores. All treatment effects are represented in percentage points. In column (1), the dependent variable is a dummy for participating in any of the 2021 AAP tests. In colunm (2), the dependent variaable is a dummy for students that participated in all 2021 AAP tests. We also report the p-value for the F test congruent with the joint null hypothesis that all coefficients are equal to zero and the r-squared for regressions. All estimates include municipality fixed-effects and standard-errors are clustered at the school-level. * if p<0.1, ** p<0.05 and *** if p<0.01.

Next, Figure C1 documents the prevalence of growth mindset accounting for all survey responses until that week, along with the cumulative treatment effect of the pooled treatment relative to the control group at each week. The gap between the pooled treatment and control groups appears immediately and persists throughout the experiment. Missing weeks are those for which there were no valid SMS survey responses on record (see Lichand et al., 2024a).

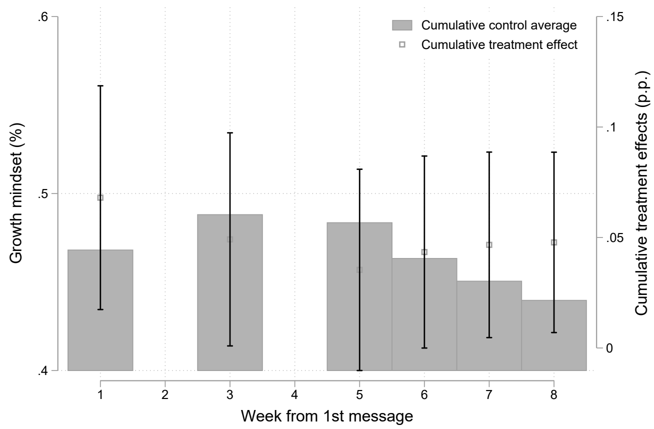
Table C3 documents treatment effects on report card grades. Relative to standardized scores, these outcomes have the disadvantage of not being centrally graded and not necessarily being comparable over time. On the other hand, we have data for these outcomes for nearly all students, since these are mandatory and high stakes. Table C3 includes treatment effects on high dropout risk (our proxy for actual dropouts). This proxy is necessary because the State Secretary

Table C.2—Balance tests for the sub-sample of students who took all AAP tests

	Averages							Pooled vs.	= across
	T1	T2	T3	T4	T5	T6	Control	Control (p-value)	all groups (p-value)
Male	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.21	0.15
Non-white students	0.28	0.27	0.27	0.26	0.26	0.27	0.27	0.72	0.80
< 1 MW	0.31	0.30	0.32	0.30	0.31	0.33	0.32	0.44	0.56
1-4 MW	0.32	0.35	0.33	0.32	0.34	0.34	0.34	0.52	0.45
4-7 MW	0.17	0.18	0.15	0.13	0.16	0.17	0.15	0.84	0.93
8-11 MW	0.07	0.08	0.08	0.07	0.08	0.08	0.08	0.62	0.47
Middle school	0.70	0.68	0.67	0.68	0.68	0.71	0.66	0.87	0.87
Avg. Math report card grade (2020)	6.50	6.52	6.67	6.52	6.51	6.51	6.50	0.12	0.09
Avg. Portuguese report card grade (2020)	6.67	6.69	6.78	6.72	6.70	6.65	6.66	0.15	0.13
Avg. Math attendance (2020)	0.93	0.95	0.94	0.94	0.95	0.95	0.94	0.75	0.64
Avg. Portuguese attendance	0.94	0.95	0.94	0.95	0.95	0.95	0.94	0.71	0.79
All means equal zero (p-value)								0.45	0.32
N							34,830		
Municipality fixed-effects								yes	yes

Note: This Table show means of variables of interest for each treatment arm and the control group. . P-values computed using randomization strata fixed effects and with standard errors clustered at the classroom level. P-value for the joint hypothesis that all differences equal zero based on a chi-squared statistic on a multinomial logit model. Data on students' gender, race, Q4/2020 report card grades and attendance comes from administrative records. Income brackets based on the Brazilian 2015 minimum wage, which was R$ 788 at that time, equivalent to approximately US$ 209 back then. Income data is not routinely collected by the Education Secretariat, but was collected by Lichand et al. (2022) for a similar sample in 2016. We impute income brackets for each student using a Poisson model, joint with multiple-imputation methods, based on student characteristics, trained in the 2016 dataset.We restrict the sample to students that took all AAP tests.

FIGURE C.1. CUMULATIVE TREATMENT EFFECTS OF TREATMENT ON STUDENT MINDSETS



Note: This Figure shows the cumulative control average (gray bars) and cumulative ITT estimates pooled treatment effects (gray squares) on the probability that students present a growth mindset, separately by weeks since the first message. Weeks when no messages were left empty. The thin black lines represent 95% confidence intervals. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the SMS survey.

automatic re-enrolled students from one year to another until 2022, masking true dropout rates. To circumvent this problem, we proxy dropout rates with students not taking tests for Math and Portuguese (Lichand and Doria, 2024). Since these tests are required for students' grade progression, not taking them is a clear sign of student disengagement, which is highly correlated with dropout risk.

The table documents that the pooled mindset intervention also significantly improved report card grades and dropout risk. Similarly to our main results, the pooled treatment increased report card grades by 0.09 to 0.1 SD. The treatment also slightly reduced dropout risk. Unlike standardized scores, however, the placebo intervention also significantly impacted report card outcomes, backing up our choice to focus on standardized test scores – which centrally graded – in the main text.

Figure C2 then documents dynamic treatment effects on standardized test scores, by treatment arm. Similarly to the pooled effect, we observe a sharp decay after the treatment ends for all arms. All estimated effects are smaller and no longer statistically significant after the second school quarter.

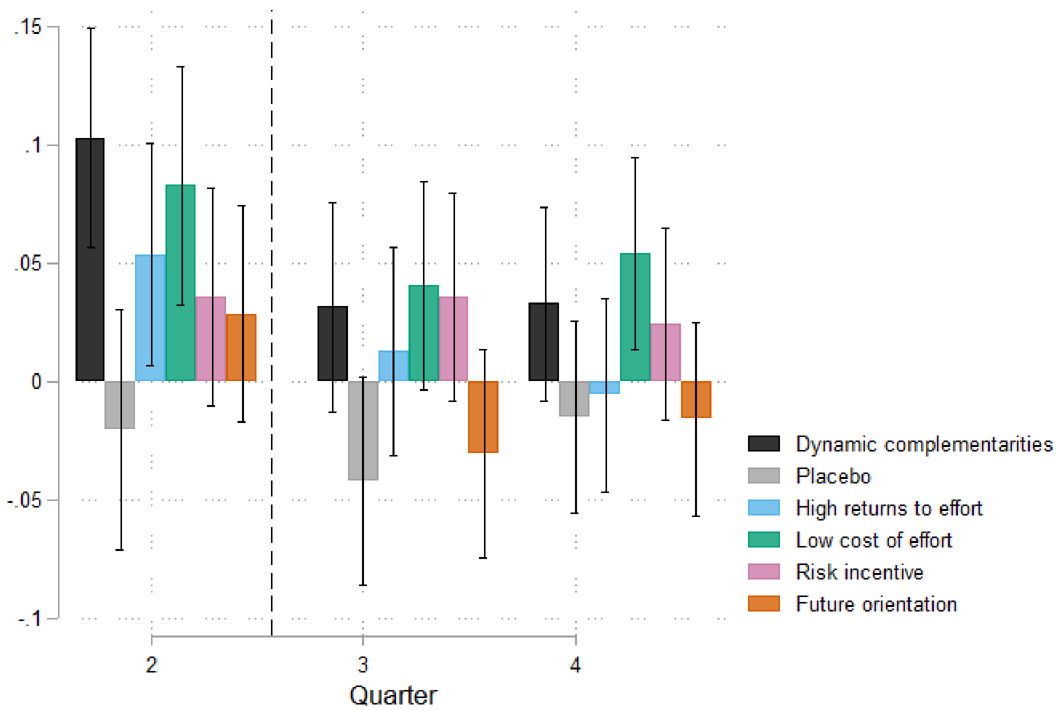
One concern is that the treatment itself might have affect participation in standardized tests, inducing differential non-response and biasing our estimates of treatment effects on learning outcomes. We can directly correct for differential participation across standardized tests by implementing bounds (Lee, 2009), which account for treatment effects on exam take-up (if any). Table C4 estimates lower bounds for treatment effects on learning. Results discussed in the main text

Table C.3—Treatment effects on report card grades and dropout risk

	(1) Math report card grade	(2) Portuguese report card grade	(3) High dropout risk
Panel A: Pooled treatment x control			
Pooled treatment	0.099***	0.091***	-0.003*
	(0.011)	(0.011)	(0.002)
Placebo	0.083***	0.072***	-0.008***
	(0.014)	(0.014)	(0.002)
p-value diff. [pooled - placebo]	0.07	0.09	0.00
Panel B: Separate treatment arms x control			
Dynamic complementarities	0.069***	0.047***	-0.002
	(0.014)	(0.014)	(0.002)
High returns to effort	0.123***	0.128***	-0.001
	(0.014)	(0.014)	(0.002)
Low effort costs	0.088***	0.088***	-0.006***
	(0.016)	(0.014)	(0.002)
Future orientation	0.166***	0.130***	0.001
	(0.014)	(0.014)	(0.002)
Risk-taking	0.046***	0.064***	-0.004*
	(0.014)	(0.014)	(0.002)
N	181,332		181,332
Control mean	0.00	0.00	0.03
Δ std. scores Q1	0.19	0.17	
Municipality fixed-effects	yes	yes	yes

Note: This Table shows two-stage ITT estimates of pooled treatment effects, placebo estimates, and separate treatment effects for each treatment arm on additional outcomes. We show results for average of math and Portuguese report card grades treatment effects for a proxy for high dropout risk, which consists in a dummy for students that had no math or Portuguese report card grades recorded. All estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if p¡0.1, ** p¡0.05 and *** if p¡0.01.

Figure C.2. Dynamic treatment effects on standardized scores, by treatment arm



Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on standardized tests, separately by quarter and treatment arm. We average standardized scores for Portuguese and math. Each bar from a different color represents a different treatment arm. The thin dashed line marks the end of the intervention. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

are robust to selection concerns.

TABLE C.4—LOWER BOUNDS FOR TREATMENT EFFECTS ON STD. TEST SCORES

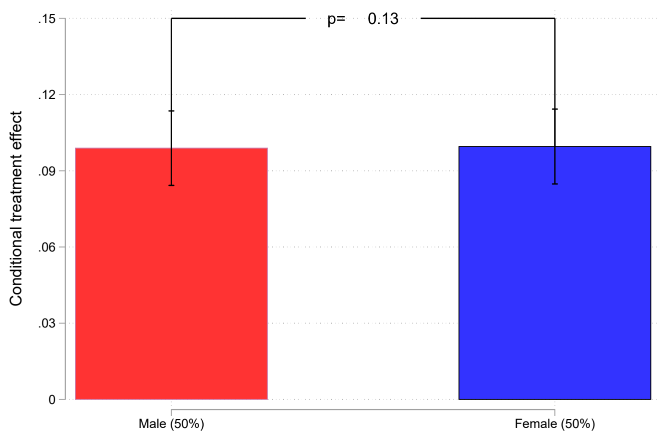
	(1) Raw sample	(2) IPW	(3) First-difference model	(4) Constant sample
Pooled treatment	0.025***	0.031***	0.075***	0.034***
	(0.018)	(0.020)	(0.025)	(0.021)
Placebo	-0.022	0.004	-0.004	-0.038
	(0.023)	(0.027)	(0.035)	(0.030)
p-value diff. [pooled - placebo]	0.00	0.00	0.00	0.00
N	181,332		56,028	
Δ Q1 avg. std. test scores (control)	0.13			
Municipality fixed-effects	yes	yes	yes	yes

Note: This Table shows ITT estimates pooled treatment effects and placebo estimates on Q1/2021 and Q2/2021 standardizes scores. We average standardized scores for Portuguese and math. Treatment effects are measured in standard deviations of the control group. We also report the control mean and the r-squared for regressions. Column (1) shows OLS regressions without any correction for selection. In column (2) we show results weighing observations by the inverse of their predicted probability of participating in the AAP test. In column (3), we restrict attention to students that took the last AAP of 2020 and the first AAP of 2021 and report first-difference estimates on standardized scores. Finally, in column (4) we report treatment effects for the constant sample of students that took at least one these two tests. All estimates include municipality fixed-effects and standard-errors are clustered at the school-level. * if p<0.1, ** p<0.05 and *** if p<0.01.

When it comes to conditional average treatment effects (CATE), we do not find significantly heterogeneous results by student gender or race, or by the timing of school reopening amidst the pandemic, by phone ownership, or by education level.

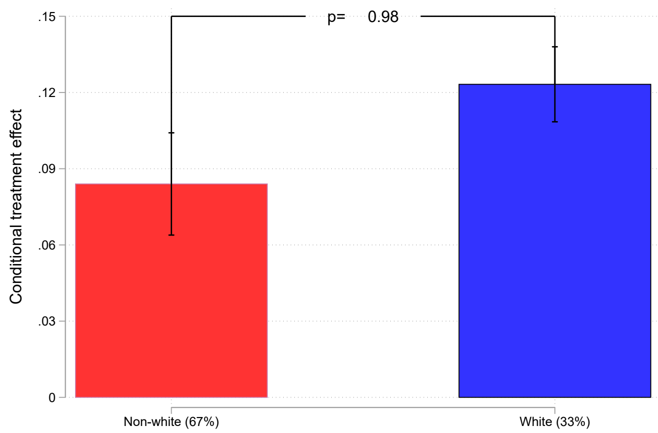
Table C4 shows that our main results are robust to controlling semi-parametrically for school size. We add linear, quadratic and cubic controls for the number of students in each school. Results remain nearly unaltered.

Figure C.3. Heterogeneous treatment effects of the pooled intervention on std. test scores, by gender



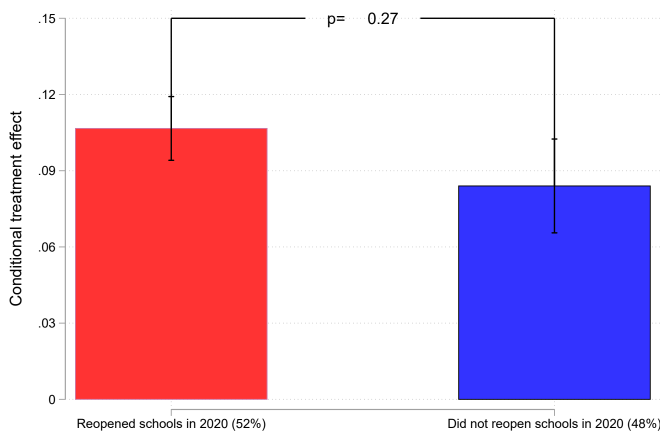
Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on Q2/2021 standardized tests, separately for males and females. We average scores for Portuguese and Math standardized tests. We also report the p-values for the two-sided null hypothesis that treatment effects are equal across groups. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Below each bar, we show the fraction of students belonging to each group. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

Figure C.4. Heterogeneous treatment effects of the pooled intervention on std. test scores, by race



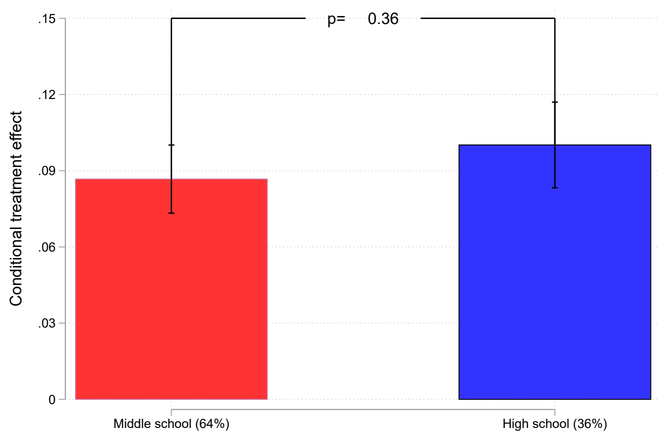
Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on Q2/2021 standardized tests, separately for white and non-white students. We average scores for Portuguese and Math standardized tests. We also report the p-values for the two-sided null hypothesis that treatment effects are equal across groups. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Below each bar, we show the fraction of students belonging to each group. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

Figure C.5. Heterogeneous treatment effects of the pooled intervention on std. test scores, by municipality decision to reopen schools



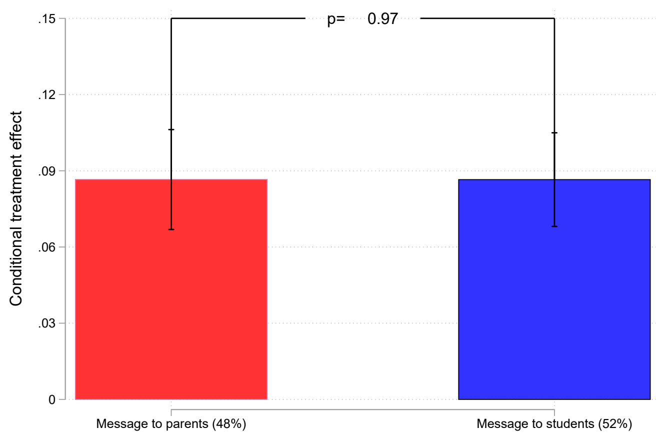
Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on Q2/2021 standardized tests, separately for municipalities that decided to reopen schools in 2020. We average scores for Portuguese and Math standardized tests. We also report the p-values for the two-sided null hypothesis that treatment effects are equal across groups. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Below each bar, we show the fraction of students belonging to each group. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

Figure C.6. Heterogeneous treatment effects of the pooled intervention on std. test scores, by grade



Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on Q2/2021 standardized tests, separately for high school and middle school. We average scores for Portuguese and Math standardized tests. We also report the p-values for the two-sided null hypothesis that treatment effects are equal across groups. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Below each bar, we show the fraction of students belonging to each group. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

FIGURE C.7. HETEROGENEOUS TREATMENT EFFECTS OF THE POOLED INTERVENTION ON STD. TEST SCORES, BY PHONE OWNERSHIP



Note: This Figure shows ITT estimates pooled treatment effects (excluding the placebo) on Q2/2021 standardized tests, separately by phone ownership. We average scores for Portuguese and Math standardized tests. We also report the p-values for the two-sided null hypothesis that treatment effects are equal across groups. Treatment effects are measured in standard deviations of the control group. The thin black lines represent 95% confidence intervals. Below each bar, we show the fraction of students belonging to each group. Estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test.

Table C.5—Robustness to controlling for school size

	(1) Raw sample	(2) IPW	(3) First-difference model	(4) Constant sample
Panel A: Pooled treatment x control				
Pooled treatment	0.045***	0.051***	0.087***	0.055***
	(0.018)	(0.020)	(0.025)	(0.021)
Placebo	-0.020	0.004	-0.004	-0.033
	(0.023)	(0.027)	(0.035)	(0.030)
p-value diff. [pooled - placebo]	0.00	0.00	0.00	0.00
Panel B: Separate treatment arms x control				
Dynamic complementarities	0.083***	0.100***	0.101***	0.082***
	(0.022)	(0.025)	(0.033)	(0.028)
High returns to effort	0.040*	0.037	0.082***	0.052**
	(0.022)	(0.024)	(0.032)	(0.027)
Low effort costs	0.070***	0.079***	0.152***	0.098***
	(0.023)	(0.027)	(0.033)	(0.029)
Risk-taking	0.027	0.033	0.095***	0.047*
	(0.022)	(0.025)	(0.033)	(0.027)
Future orientation	0.021	0.022	0.037	0.010
	(0.022)	(0.014)	(0.031)	(0.026)
N	181,332		56,028	
Δ std. scores Q1		0.13		
Municipality fixed-effects	yes	yes	yes	yes

Note: This Table shows ITT estimates pooled treatment effects, placebo estimates, and separate estimates for each treatment arm on learning outcomes. We show results for three different variables: math and Portuguese report card grades and standardized scores. For the latter, we average math and Portuguese scores.All treatment effects are measured in control group standard-deviaitons.. We also report the control mean and the r-squared for regressions.. All estimates include municipality and quarter fixed-effects and standard-errors are clustered at the school-level. We also control for a third-order polynomial of school size. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if p<0.1, ** p<0.05 and *** if p<0.01.

Appendix D. Two-sample Two-stage Least Squares (TS2SLS) Estimation

This Appendix estimates a direct relationship between student mindsets, on the one hand, and student effort and learning outcomes, on the other, through a Two-Sample Two-Stage Least Squares (TS2SLS) estimator, similar to Angrist and Krueger (1992) and Inoue and Solon (2009).

Concretely, we first predict causal changes in student mindsets using SMS survey data, based on the heterogeneous treatment effects of the dynamic complementarities arm – the one with the largest first-stage effect size. Based on these estimates, we can impute student mindsets in the administrative data based on student characteristics, which allows us to estimate the direct relationship. As long as the mindset intervention only affects student effort and learning outcomes through its effect on the probability of expressing a growth mindset (and not through additional mechanisms) – the identification assumption –, then the TS2SLS estimator captures the causal effect of growth mindset on the former. Naturally, this is a strong assumption, and this is why we relegate those results to the Appendix.

The main challenge for computing the estimator is the limited SMS survey sample size, constraining statistical power to precisely detect first stage estimates. To deal with this challenge, we adjust TS2SLS inference following Choi, Gu, and Shen (2018).

Table D.1—TS2SLS estimates using dynamic complementarities treatment as instrument

	(1)	(2)	(3)	(4)
	Effort		Outcomes	
	Attendance	Cumulative minutes by week 20	Avg. report card grades	Avg. standardized test scores
Growth mindset	0.628***	928.43***	0.210***	0.799**
	(0.123)	(369.45)	(0.285)	(0.225)
Control mean	0.90	582.369	0.00	0.00
First-stage F statistic		1.88		
First-stage N		3,249		
Second-stage N	181,332	87,306	181,332	145,748

Note: This Table shows two-stage 2sls of pooled treatment effects on effort and learning outcomes. We measure effort by averaging math and Portuguese attendance and by tracking the total time spent by students by week twenty. We also show results for average of math and Portuguese report card grades and standardized scores Finally, we show results treatment effects for a proxy for high dropout risk, which consists in a dummy for students that had no math or Portuguese report card grades recorded.In order to compute the TS2SLS estimator, we use survey data to compute the first-stage estimates and the remaining administrative data All estimates include municipality fixed-effects and standard-errors are clustered at the school-level. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if p<0.1, ** p<0.05 and *** if p<0.01.

For robustness, Table D2 computes treatment effects relative to the placebo intervention, instead of the control group. Table D3 re-weights the sample in the first stage of the estimator to match the characteristics of the universe of students.

Last, Table D4 re-weights the sample accordingly, but in the second stage of the estimator. All results are similar to the main TS2SLS specification.

Table D.2—TS2SLS estimates of mindset on all outcomes of interest using "dynamic complementarities" as instrument and placebo as control

	(1)	(2)	(3)	(4)
	Effort		Outcomes	
	Attendance	Cumulative minutes by week 20	Avg. report card grades	Avg. standardized test scores
Growth mindset	0.628***	928.43***	0.210***	0.799**
	(0.123)	(369.45)	(0.285)	(0.225)
Control mean	0.90	582.369	0.00	0.00
First-stage F statistic	1.88			
First-stage N	3,249			
Second-stage N	181,332	87,306	181,332	145,748

Note: This Table shows two-stage 2sls of pooled treatment effects on effort and learning outcomes. We restrict the sample to the dynamic complementarities treatment arm and the placebo as a control. We measure effort by averaging math and Portuguese attendance and by tracking the total time spent by students by week twenty. We also show results for average of math and Portuguese report card grades and standardized scores. Finally, we show treatment effects for a proxy for high dropout risk, which consists in a dummy for students that had no math or Portuguese report card grades recorded. In order to compute the TS2SLS estimator, we use survey data to compute the first-stage estimates and the remaining administrative data All estimates include municipality fixed-effects and standard-errors are obtained using a robust to weak IV estimator. * if p<0.1, ** p<0.05 and *** if p<0.01.

Table D.3—TS2SLS estimates of mindset on all outcomes of interest using "dynamic complementarities" as instrument re-weighting the first-stage

	(1)	(2)	(3)	(4)	(5)
	Effort			Outcomes	
	Attendance	Cumulative minutes by week 20	Avg. report card grades	Avg. standardized test scores	High dropout risk
Growth mindset	0.435***	742.47***	0.755**	0.813**	-0.069
	(0.086)	(271.85)	(0.344)	(0.325)	(0.071)
Control mean	0.90	582.369	0.00	0.00	0.13
First-stage F statistic			1.85		
First-stage N			3,249		
Second-stage N	181,332	87,306	181,332	145,748	181,332

Note: This Table shows two-stage 2sls of pooled treatment effects on effort and learning outcomes. We measure effort by averaging math and Portuguese attendance and by tracking the total time spent by students by week twenty. We also show results for average of math and Portuguese report card grades and standardized scores Finally, we show results treatment effects for a proxy for high dropout risk, which consists in a dummy for students that had no math or Portuguese report card grades recorded.In order to compute the TS2SLS estimator, we use survey data to compute the first-stage estimates and the remaining administrative data. All estimates include municipality fixed-effects and standard-errors are obtained using a robust to weak IV estimator. We weight observations by the inverse of their predicted probability of participating in the SMS survey. * if p<0.1, ** p<0.05 and *** if p<0.01.

Table D.4—TS2SLS estimates of mindset on all outcomes of interest using "dynamic complementarities" as instrument re-weighting the second-stage

	(1)	(2)	(3)	(4)	(5)
	Effort			Outcomes	
	Attendance	Cumulative minutes by week 20	Avg. report card grades	Avg. standardized test scores	High dropout risk
Growth mindset	0.444***	748.12***	0.766**	0.813**	-0.070
	(0.085)	(271.42)	(0.286)	(0.227)	(0.073)
Control mean	0.90	582.369	0.00	0.00	0.13
1-year learning rate (2019)				0.44	
1-year learning rate (2020)				0.18	
First-stage F statistic			1.85		
First-stage N			3,249		
Second-stage N	181,332	87,306	181,332	145,748	181,332

Note: This Table shows two-stage 2sls of pooled treatment effects on effort and learning outcomes. We measure effort by averaging math and Portuguese attendance and by tracking the total time spent by students by week twenty. We also show results for average of math and Portuguese report card grades and standardized scores Finally, we show results treatment effects for a proxy for high dropout risk, which consists in a dummy for students that had no math or Portuguese report card grades recorded.In order to compute the TS2SLS estimator, we use survey data to compute the first-stage estimates and the remaining administrative data. All estimates include municipality fixed-effects and standard-errors are obtained using a robust to weak IV estimator. We weight observations by the inverse of their predicted probability of participating in the AAP test. * if p<0.1, ** p<0.05 and *** if p<0.01.

Appendix E. Pre-analysis plan

Appendix E.1. Introduction

Recent evidence documents that low-cost interventions aimed at giving teenagers a growth mindset (versions of the message "your brain is like a muscle") have very sizable and positive effects on their test scores and on high school drop-out rates. In particular, the Growth Mindset intervention developed for and described in Yeager et al. (2016) has been shown to be effective in promoting a growth mindset for 9th grade students both in the US (Yeager et al., 2019) and Norway (Bettinger et al., 2018). However, this intervention requires the use of computers and internet, resources that are not so widely available in many developing countries. On the other hand, mobile phone(s) can be found in almost every kind of household. Specifically, in Brazil it is estimated that more than 90% of the households have at least one mobile phone (IBGE, 2018). This investigation is particularly relevant for the COVID-19 pandemic period, as public school students can no longer attend school in person and have limited access to homeschooling resources.

For this study, we partnered with São Paulo State's Secretary of Education (SEDUC-SP), in Brazil. SEDUC-SP provides access to 1,415,290 mobile phone contacts either from students themselves (grades 10 to 12) or a students' parent or legal guardian (grades 6 to 9) to evaluate different forms of the growth mindset intervention.

We translated and adapted the computer-based growth mindset intervention from English to Portuguese, following the methodology in Bettinger et al. (2018), that translated this intervention from English to Norwegian. Due to limitations on school operations during the COVID-19 pandemic, we deliver the computer-based intervention remotely to each student by means of a text message link, instead of the original in-person delivery at the school. To access this content, students will need a smartphone, tablet or computer with internet access. We also adapted this content to be delivered exclusively through text messages, making it more accessible to vulnerable populations and scalable in both developed and developing countries. This adaptation process considered the methodology for SMS interventions used in Bettinger et al. (2020) that developed a parent engagement SMS intervention called EDUQ+, namely a platform powered by Movva, a Brazilian social impact startup, which allows schools to send messages to parents with information about children's attendance and grades, and which nudges them with motivating facts and suggested activities to engage them in their children's school life. The SMS-based version of EDUQ+ has been shown to be effective in Brazil, where communication with parents had large impacts on attendance, test scores and promotion rates (Bettinger et al., 2020). We will evaluate the effectiveness of this intervention in Brazil in its computer- and SMS-based versions separately on outcomes such as likelihood of dropping out, engagement with the remote schooling app provided by SEDUC-SP, and test scores.

Moreover, rather than just comparing average treatment effects of computer vs. SMS messages, this paper investigates why the growth mindset interventions works, unbundling mindsets into its underlying mechanisms. We do so by randomly assigning 800,000 Brazilian students from grades 6 to 12 to different versions of the mindset intervention, varying whether students are exposed to specific components of the mindset message: (i) whether or not messages make school activities salient (without conveying any message related to mindsets), whether or not messages emphasize (ii) high returns to effort or (iii) Low effort costs, and whether or not messages highlight (iv) the value of risk-taking (risk preferences) or (v) that of assigning higher weights to future outcomes (time preferences). This pre-analysis plan summarizes the design of a phone-based experiment designed to test the following hypotheses:

1) Can the growth mindset intervention improve academic and behavioral learning outcomes in Brazil?
    - Hypothesis: The mindset intervention increases the likelihood of self-reported growth mindset, improves performance in Math, increases engagement with content while homeschooling, improves grades and promotion rates, and decreases dropout rates.
        - In Brazil, different from the US and Norway, public schools provide education mostly to students from low-income families. Thus, it is not clear that children's learning outcomes will improve with this intervention, as the resources for developing a growth mindset may be scarcer.
2) To what extent can an SMS-based intervention replicate the impacts of the computer-based intervention?
    - Hypothesis: The effect of the SMS mindset intervention is the same as that of the computer- based intervention delivered remotely.
3) Are the effects of the mindset intervention primarily driven by inference about lower costs of effort, inference about higher returns to effort, inference about risk attitudes, or inference about future- orientation?
    - · Hypothesis: Emphasizing Low effort costs, high returns to effort, risk-taking or future- orientation leads to differential impacts of the SMS intervention on student outcomes.
4) Are the effects of the intervention at least partly driven by its salience effects?
    - Hypothesis: Making school activities salient, even in the absence of the mindset message, impacts student outcomes.
5) Are the effects of the intervention persistent?

- Hypothesis: The intervention affects student outcomes such as their mindset and school attendance even in the months after communication has taken place.

6) Do additional rounds of the intervention induce larger impacts?
   - Hypothesis: Extending the SMS intervention for additional months increases its impacts on student outcomes.
7) Does the intensity of the mindset intervention's effect differ by grade?
   - Hypothesis: The mindset intervention has an inverted U-shaped effect across grades 6 to 12, with its peak effect happening during the transition from middle to high school (from grade 9 to 10 in Brazil).
8) Does a combination of computer-based and text-message intervention yield larger effects?
   - Hypothesis: The computer-based and text message delivery formats have complementary features that increase the impact of the mindset treatment on students' outcomes.

Appendix E.2. Intervention - Experimental design

The intervention will be evaluated through a phone-based randomized control trial with 866,666 students from grades 6 to 12 of São Paulo's state schools. Students will be randomly assigned to either a pure control group (who will not receive any kind of growth mindset content), to an SMS-based content only treatment group, or to the computer-based content treatment group.

We undertake 3 experiments to answer the research questions of interest. Experiment 1 evaluates the impact of the computer-based intervention, adapted to the Brazilian context, to test hypothesis 1. In September, 400,000 students will receive a text message with a link to access the online content typically used in the computer-based intervention, made available through an online platform (Qualtrics). Upon accessing the link, students will be randomized on the spot to one out of 2 different groups, with equal probability:

1) A placebo group, which will receive an online intervention about the brain; or
2) A treatment group, which will receive an online growth mindset intervention.

The placebo intervention will instruct students in this group about the brain's features and functionalities but will not provide any information about mindset. The treatment intervention will instruct students in the treatment group about growth mindset and how to develop it. Both treatments are based on text, illustrated with a few images (keeping fidelity to prior studies, but adapting to local context). Audio recording of the texts are also made available for students who prefer that option (in case they have enough airtime credit and bandwidth to access it).

Based on other studies with similar settings, we expect 2%-12% of the target sample to access the online content. We would be able assign and observe outcomes immediately after the intervention only for that sub-sample. We estimate a conservative minimum detectable effect (MDE) of 6.3% standard deviation for comparing the treatment group to the placebo group considering a 2% take-up rate of the total sample for this experiment, and reaching 2.6% of a standard deviation at a take-up rate of 12%.

Additionally, we will observe administrative outcomes (attendance, grades, and access to the distance learning platform) for all students, which will allow us to estimate selection into Experiment 1, and estimate treatment effects re-weighting observations by the inverse of the selection probability. Experiment 2 aims at testing hypotheses 2, 3 and 4. Also in September, 400,000 students (with no overlap with Experiment 1) will receive 1 SMS, assigned to 1 out of 6 different groups. There is also a control group, of identical size as each of the other cells, assigned to receive no SMS, as follows:

1) Typical growth mindset intervention group ("your brain is like a muscle, ...");
2) Salience group ("we want you to stay in school");
3) High returns to effort group ("if you put more effort in, you can always improve relative to yourself");
4) Low effort costs group ("studying is actually more fun than you might think");
5) Risk-taking group ("trying is always worth it, failure does not mean anything about your potential");
6) Future-orientation group ("when thinking about your studies, it is important to think about your future!"); or
7) A pure control group, which does not receive any text message.

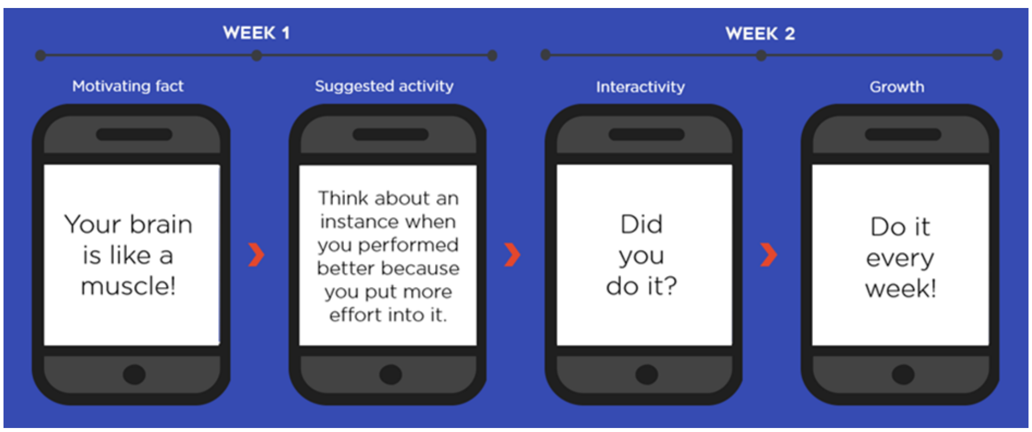
For this experiment, we estimate an MDE of 1.7% of a standard deviation on outcomes obtained from administrative data when comparing each treatment cell against the pure control group. For a comparison between computer- and SMS-based interventions, we expect an MDE of 39% of a standard deviation at a 2% take-up rate, and 6.6% of a standard deviation at a 12% take-up rate.

Experiment 3 aims at testing hypotheses 5 and 6. From October through December, the plan is to stick to the assignment for a sub-sample of students, potentially with more messages per student per month. We plan on updating the pre-analysis plan before the end of September to specify the final sample and number of messages. That decision depends on other tests being undertaken by SEDUC together with Movva to determine the optimal trade-off between number of students impacts by the nudges and number of nudges per student per month.

If SEDUC decides to stick to 1 message per month, the plan is to repeat the messages with slight variations in framing but keeping both the assignment and the core concept within each group. Alternatively, if SEDUC decides to go with more than 1 message per month (with a smaller sample size), we will randomly draw users within each treatment arm to receive additional messages. That experiment would allow us to estimate persistence and fade-out of treatment effects, as well as saturation effects (contrasting those assigned to receive additional messages to the ones who received messages only in September, and those to the pure control group).

In case Movva sends more than 1 SMS per month, communication will be based on sequences of four messages: a motivating fact, a suggested activity, an interactive message, and a growth message (see an example below).

Figure E.1. Example of an SMS sequence for the Typical Growth Mindset Intervention Group



Across the different treatment arms, each message tries to stick closely to the concept within that arm. In particular, the distinction between high returns to effort and Low effort costs will be done by appealing the concept of willpower in the Low effort costs Group (Job, Dweck and Walton, 2010). Additionally, we intend to embed a 4th experiment into experiment 3 to evaluate hypothesis 8, or how the computer-based intervention format interacts with the text message format. To do so, we will select a random sub-sample of subjects in Experiment 3, to be determined and updated here after September, to receive a text message containing the link to the computer-based treatment. If the two formats yield the same results, we should not expect different or improved outcomes from their combination. However, if there are complementarities between the two formats (e.g., the computer-based treatment is better at engaging students, and the text message treatment's salience allows to fixate the concepts over time), we might find outcomes to improve with respect to each intervention separately.

Appendix E.3. Outcomes

We will assess how the different versions of the Mindset intervention impact learning efforts and outcomes for students enrolled in grades 6 to 12 (ranging 10-18 years old students). To do so, we will use administrative records to evaluate treatment effects on:

- Daily access to the distance learning platform;
- Daily time online on the distance learning platform;
- Weekly attendance (by school subject) when the time regular classes resume;
- Quarterly grades (by school subject) when regular classes resume;
- Student dropouts.

We will measure baseline levels of growth mindset, motivation to return to school when regular classes resume, students' experiences with distance learning, and feelings of isolation during the pandemic (school-grade averages, as of early September, before the intervention starts). We will do so by assigning each student one of four questions randomly, i.e., 25% of students in a given grade-school pair will receive one text message with one of the following questions:

1) On a scale from 1 to 6, how much do you disagree (0) or agree (6) with the following: Your intelligence is something about you that you can't change very much.
2) Do you plan on going back to school once in-person classes resume?
3) Tell us about your experience of studying from home during this period without in-person classes.
4) Share with us how you've been feeling during the past few months without seeing your teachers and classmates.

These questions will be rephrased accordingly when the message recipient is the student's caregiver and not the student.

In addition, we also have information on various other baseline variables: the history of access to the distance learning platform since May; attendance and grades for the first quarter of the year, provided by administrative records; predicted risk of dropouts at the student-level, on a 0-100 scale.

We are interested in estimating heterogeneous treatment effects of the interventions according to those baseline measures. Additionally, between September and December, students who take-up the computer-based intervention and rotating sub-samples of 1,000 students within each SMS treatment arm and the pure control group (total 7,000 students per week) will be drawn to receive a question via SMS, to capture treatment effects on mindset. Questions are variations of Q1 above, following Bettinger et al. (2018).

For experiment 1, we expect to send the SMS question to all students who take-up the computer-based intervention up to a 4.75% of the total sample. As a conservative estimate with a take-up rate of 2% and 2% of response rate to SMS questions, we estimate an MDEs of 44% within a single week, and as low as 11% when accumulated over 15 weeks. At a 4.75% take-up rate and 2% response rate for the SMS question, we estimate the MDEs for a single week to be 28%, and as low as 7.4% when accumulated.

For experiment 2, at a weekly response rate of 2%, we estimate an MDEs of 88% within a single week, and as low as 22% accumulated over 15 weeks. At a weekly response rate of 12%, we estimate an MDEs of 36% within a single week, and as low as 5.8% accumulated over 15 weeks.

Appendix E.4. Estimation

We will document the effects of the treatments on the following outcome categories for the 6th to 12th graders in our sample (aged 11 and 18 years old respectively):

- Students' self-reported mindset (growth or fixed), weekly;
- Student-level access to the distance learning platform, measured by the daily access and the daily time on the platform;
- When actual classes return: school attendance, grades, grade retention and dropout rates, measured by administrative records.

Whenever we have access to multiple outcome variables mapped into a single outcome category (e.g. grades for several school subjects), we will build summary measures: following Kling, Liebman and Katz (2007), we will normalize all outcomes to z-scores, and run seemingly unrelated regressions (SUR) to compute effect sizes for each outcome category.

Appendix E.5. Deviations from the pre-analysis plan

This paper focuses on Experiment 2 outlined in the pre-analysis plan. Our attempt at implementing Experiment 1 and the results of Experiment 3 are described in companion papers (Lichand et al. 2024a, 2024b). As such, our focus in this paper is on testing hypotheses 2, 3, and 4.

While we follow the pre-analysis plan as closely as possible, there were two significant deviations from had been pre-specified. First, changes in the sample design. Since we could only roll out the experiment in the aftermath of the pandemic, the São Paulo State Education Secretariat mandated a reduction in the sample size from the original target of 400,000 students. Moreover, we further subdivided the available sample into two – Experiment A, whereby treatment content focused narrowly on the mechanisms of interest, and Experiment B, which embedded such content into practical suggestions to help students get back to their school routine as in-person classes resumed. We did so following advice of the implementing partner, which was concerned that students might not pay as much attention to conceptual messages as to practical suggestions. As we discuss in Appendix B, conceptual messages actually worked much better, presumably because the mindset message got washed out in embedded content. With such a large sample size, we still had enough statistical power to detect relevant effect sizes of each of the different treatment arms in each of the experiments.

The second significant deviation from the pre-analysis plan refers to the presentation of outcomes. At the time at which we pre-registered the analysis plan, there was still uncertainty about what student data the research team would be able to access. Student-level data on standardized tests was granted only at a later stage. We prioritize standardized test results as our primary learning outcome because, in contrast to report card grades, they are centrally graded. In fact, while we do see placebo effects on the latter, they are absent from the former. In any case, results on report card grades are presented in full in Appendix C. Treatment effects for the pooled intervention are very robust to the outcome choice. There are, however, some differences in the ranking of the different treatment arms depending on which outcome is used.