POPULATED PRE-ANALYSIS PLAN
for
Learning about COVID-19: Improving Knowledge via Incentives and Feedback*

James Allen IV[1,2,6], Arlete Mahumane[3], James Riddell IV[4], Tanya Rosenblat[5], Dean Yang[1,2,6], and Hang Yu[7,8]

[1]Department of Economics, University of Michigan.
[2]Ford School of Public Policy, University of Michigan.
[3]Beira Operational Research Center, National Institute of Health, Mozambique.
[4]Division of Infectious Diseases, University of Michigan Medical School.
[5]School of Information and Department of Economics, University of Michigan.
[6]Population Studies Center, University of Michigan.
[7]National School of Development, Peking University.
[8]Institute of South-South Cooperation and Development, Peking University

Version: May 26, 2023

*Contacts: alleniv@umich.edu; deanyang@umich.edu. Acknowledgements: Faustino Lessitala provided top-notch leadership and field management. Patricia Freitag, Ryan McWay, and Maggie Barnard provided excellent research assistance. Julie Esch, Laura Kaminski, and Lauren Tingwall's grant management was world-class. We appreciate feedback from Hoyt Bleakley, Brian Jacob, Laston Manja, Kwasi Tabiri, and participants in Michigan's Health, History, Development, and Demography (H2D2) Seminar and the Conference on Experimental Insights from Behavioral Economics on COVID-19 (JHU-LSE). This work is supported by the Abdul Latif Jameel Poverty Action Lab via the Innovation in Government Initiative at the Massachusetts Institute of Technology (grant number IGI-1366), Innovations for Poverty Action via the Peace and Recovery Program at Yale University (grant number MIT0019-X9), the Michigan Institute for Teaching and Research in Economics via the Ulmer Fund (grant number G024289), and the National Institute on Aging of the National Institutes of Health (award number T32AG000221). Our study protocols were approved by Institutional Review Boards (IRBs) at the University of Michigan (Health Sciences and Social and Behavioral Sciences IRB, approval number HUM00113011) and the Mozambique Ministry of Health National Committee on Bioethics for Health (Portuguese acronym CNBS, reference number 302/CNBS/20). The study was submitted to the American Economic Association's RCT Registry on August 25, 2020, registration ID number AEARCTR-0005862: 10.1257/rct.5862-3.0

1 Introduction

On August 25, 2020, prior to baseline data collection, we uploaded our pre-analysis plan (PAP) "Learning about COVID-19: Improving Knowledge via Incentives and Feedback" to the American Economic Association's RCT Registry, registration ID number AEARCTR-0005862: https://doi.org/10.1257/rct.5862-3.0.We follow Duflo et al. (2020), assembling the full set of pre-specified analyses in a Populated PAP document. This full Populated PAP can be accessed on our research website:
https://fordschool.umich.edu/mozambique-research/combating-covid-19.

Note that we adhere to the nomenclature used in the main text of Allen IV et al. (2021) "Teaching and Incentives: Substitutes or Complements?" to refer to the treatment conditions (i.e., "Incentive" and "Teaching") rather than that used in the PAP (i.e., "Knowledge Incentive" and "Tailored Feedback", respectively). Additionally, while the PAP refers to the primary outcome variables as 1) the Knowledge Index (based on 40 questions), and 2) the Feedback-Eligible Knowledge Index (based on 20 questions), the main text of the paper only focuses on the second of these two referring to it as the "COVID-19 Knowledge Test Score". Throughout this Populated PAP, we refer to these outcomes, respectively, as 1) the Overall Test Score (or simply Test Score), and 2) the Feedback-Eligible Test Score.

2 Primary Analyses

We estimate intent-to-treat (ITT) effects using the following ordinary-least-squares (OLS) regression specifications. To estimate the causal effect of the Incentive treatment, we run:

$$Y^{all}_{i,j,t=3} = \alpha_0 + \alpha_1 Incentive_{ij} + \alpha_2 Teaching_{ij} + \alpha_3 Joint_{ij} + \eta \mathbf{B}_{ijt} + \gamma_i + \varepsilon_{ij} \tag{2.1}$$

where $Y^{all}_{i,j,t=3}$ is the Overall Test Score for respondent i in community j, measured in Round 3 survey; $Incentive_{ij}$, $Teaching_{ij}$, and $Joint_{ij}$ are indicators for inclusion in the respective treatment groups; $\mathbf{B}_{ijt}$ is a vector representing the share of correct answers to questions asked in Round 1 and Round 2, respectively [1]; γ_i are community fixed effects; and ε_{ij} is a mean-zero error term. We report robust standard errors.

To estimate the causal effect of the Teaching and Joint treatments, we run:

$$Y^{feedback}_{i,j,t=3} = \beta_0 + \beta_1 Incentive_{ij} + \beta_2 Teaching_{ij} + \beta_3 Joint_{ij} + \eta \mathbf{B}_{ijt} + \gamma_i + \varepsilon_{ij} \tag{2.2}$$

where $Y^{feedback}_{i,j,t=3}$ is the Feedback-Eligible Test Score for respondent i in community j, measured in Round 3 (endline survey), and other right-hand side variables are as specified in Equation 2.1.

Results from estimating these equations are in Table 1. Overall, the coefficient signs, magnitudes, and statistical significance levels are very similar in Column 1 (for the Overall Test Score) and Column 2 (for the Feedback-Eligible Test Score). Each of the treatments has positive effects on the outcomes that are statistically significant at conventional levels even after pre-specified multiple hypothesis testing adjustment across three coefficients in the two regressions (p-values in square brackets, <0.001 in each case). The estimate, $\hat{\lambda}$, of the complementarity parameter is nearly identical across the two regressions. Coefficient estimates in Column 1 (for the Overall Test Score) and Column 2 (for the Feedback-Eligible Test Score) are very similar for the Incentive treatment effect (first row).

[1] The average respondent correctly answered 72.1% and 77.3% of the 20 knowledge questions in Rounds 1 and 2, respectively.

Table 1: **Regression of Test Score (TS) on Treatments**

VARIABLES	(1) Overall Test Score (TS)	(2) Teaching-Eligible TS
Incentive	0.0200	0.0156
	(0.0054)	(0.0060)
	[0.0003]	
Teaching	0.0160	0.0288
	(0.0055)	(0.0064)
		[0.0003]
Incentive plus Teaching (Joint)	0.0496	0.0581
	(0.0055)	(0.0060)
		[0.0003]
$\hat{\lambda}$	0.0136	0.0137
	(0.0084)	(0.0095)
Observations	2,117	2,117
R-squared	0.319	0.333
Control Mean DV	0.781	0.784
Control SD DV	0.108	0.123
p-value: $\lambda = 0$	0.1048	0.1462
p-value: λ = -0.0265	0.0000	0.0000
p-value: Incentive = Teaching	0.5292	0.0713
p-value: Incentive = Joint	0.0000	0.0000
p-value: Teaching = Joint	0.0000	0.0001

Notes: The Overall Test Score (TS) is the share of correct answers to all 40 knowledge questions in Round 3: 12 on general knowledge, 16 on preventive actions, and 12 on government actions. The Feedback-Eligible TS is the share of correct answers to the 20 knowledge questions in Round 3 that were eligible for the Teaching treatment (i.e., also asked in Round 2): 6 on general knowledge, 8 on preventive actions, and 6 on government actions. λ is the complementarity parameter (see Section 2 of main text). $\hat{\lambda}$ is coefficient on "Incentive plus Teaching" (Joint) minus sum of coefficients on "Incentive" and "Teaching". P-values adjusted for pre-specified multiple hypothesis testing are in square brackets. All regressions also include community fixed effects and controls for pre-treatment (Rounds 1 and 2) Test Scores. Robust standard errors in parentheses.

We also pre-specified other secondary analyses. First, we pool the Incentive, Teaching, and Joint treatments together to examine the effect of any treatment on the primary outcomes. Results in Table 2 for the coefficient on the indicator for receiving any treatment, "Pooled Treatment", is statistically significantly positive at conventional levels in each regression. Second, we analyze impacts of the treatments on test scores based on topical subcategories: general knowledge, protection methods, and government policies. Regressions are as described above but replacing the outcomes with the test scores for each indicated subcategories. Results, in Table 3, are broadly similar to the estimates in Table 1. The estimated complementarity parameter

$\hat{\lambda}$ appears largest (most positive) for the preventive actions subcategory (Columns 2 and 5).

Table 2: **Regression of Test Score (TS) on Pooled Treatment**

VARIABLES	(1) Overall Test Score (TS)	(2) Teaching-Eligible TS
Pooled Treatments	0.0289	0.0346
	(0.0041)	(0.0045)
Observations	2,117	2,117
R-squared	0.308	0.320
Control Mean DV	0.781	0.784
Control SD DV	0.108	0.123

Notes: Column 1: the Overall Test Score (TS) is the share of correct answers to all 40 knowledge questions in Round 3: 12 on general knowledge, 16 on preventive actions, and 12 on government actions. Column 2: the Feedback-Eligible TS is the share of correct answers to the 20 knowledge questions in Round 3 that were eligible for the tailored feedback treatment (i.e., also asked in Round 2): 6 on general knowledge, 8 on preventive actions, and 6 on government actions. All regressions also include community fixed effects and controls for pre-treatment (Rounds 1 and 2) Test Scores. Robust standard errors in parentheses.

Third, we analyze impacts of the treatments on self-reported COVID-19 preventive behaviors. Outcomes include respondents' support for social distancing, following government social distancing recommendations, and the number of preventive actions taken by the household to prevent the spread of COVID-19. All outcomes are socially desirable and advocated by the government, so positive coefficients would be considered "good". Results in Table 4 are mixed and inconclusive. Six out of nine coefficients in the table are positive, and three are negative. Only the negative coefficient on Teaching in Column 1 and the positive coefficient on Incentive in Column 2 are statistically significantly different from zero at conventional levels.

Fourth, we run a regression with indicators for knowledge treatments, the cross-randomized social distancing treatments and their interaction terms to test for significant interactions between the treatments implemented for two separate experiments in the same population. Results are in Table 5, Columns 1 and 2. There are six interaction terms in each regression. In Column 1, one coefficient (Teaching x LE) is statistically significant at the 10% level. In Column 2, that same coefficient is statistically significant at the 5% level, and another in that column (Incentive x LE) is significant at the 10% level. Looking at the patterns of coefficients overall, these appear to be chance occurrences. There is no corresponding effect of the LE (leader endorsement) treatment on the "Incentive plus Teaching" (Joint) treatment, which we should expect to also appear if the LE treatment truly interacted with the knowledge treatments. In Columns 3 and 4, we also verify that our primary treatment effect estimates are very similar when the Test Score outcome measure excludes social distancing knowledge questions, which are most susceptible to being affected by the social distancing treatments. Overall, there does not appear to be substantial evidence of interactions between the set of knowledge treatments and the set of social distancing treatments.[2].

[2]Note these are separate experiments with different pre-specified outcomes of interest. As our primary interest was never to examine interactions between these treatments sets, we do not believe it would be accurate to characterize our results as focusing on the "short model" (a weighted average of effects across different cross-randomized treatment groups), along the lines of Muralidharan et al. (2019)

Table 3: **Regression of Test Score (TS) Subcategories on Treatments**

VARIABLES	(1) General TS	(2) Preventive TS	(3) Government TS	(4) Teaching-Eligible General TS	(5) Teaching-Eligible Preventive TS	(6) Teaching-Eligible Government TS
Incentive	0.0094	0.0184	0.0421	0.0018	0.0118	0.0419
	(0.0084)	(0.0065)	(0.0083)	(0.0099)	(0.0088)	(0.0099)
Teaching	0.0154	0.0125	0.0223	0.0265	0.0234	0.0299
	(0.0085)	(0.0067)	(0.0087)	(0.0102)	(0.0092)	(0.0109)
Incentive plus Teaching (Joint)	0.0374	0.0487	0.0644	0.0415	0.0535	0.0749
	(0.0087)	(0.0065)	(0.0084)	(0.0103)	(0.0087)	(0.0100)
$\hat{\lambda}$	0.0126	0.0178	0.0001	0.0133	0.0183	0.0031
	(0.0131)	(0.0100)	(0.0127)	(0.0157)	(0.0136)	(0.0154)
Observations	2,117	2,117	2,117	2,117	2,117	2,117
R-squared	0.199	0.204	0.211	0.206	0.257	0.189
Control Mean DV	0.790	0.768	0.790	0.797	0.827	0.789
Control SD DV	0.159	0.116	0.165	0.189	0.170	0.202
p-value: $\lambda = 0$	0.3333	0.0759	0.9955	0.3985	0.1774	0.8410
p-value: Incentive = Teaching	0.5361	0.4486	0.0410	0.0354	0.2756	0.3090
p-value: Incentive = Joint	0.0048	0.0001	0.0170	0.0008	0.0000	0.0025
p-value: Teaching = Joint	0.0278	0.0000	0.0000	0.2130	0.0036	0.0001

Notes: The Overall Test Score (TS) subcategories (Columns 1-3) are the share of correct answers in Round 3 to the 12 questions on general knowledge, 16 questions on preventive actions, and 12 questions on government actions, respectively. The Feedback-Eligible TS subcategories (Columns 4-6) are the share of correct answers to the questions in Round 3 that were eligible for the tailored feedback treatment (i.e., also asked in Round 2): 6 on general knowledge, 8 on preventive actions, and 6 on government actions, respectively. λ is the complementarity parameter (see Section 2 of main text). $\hat{\lambda}$ is coefficient on "Incentive plus Teaching" (Joint) minus sum of coefficients on "Incentive" and "Teaching". All regressions also include community fixed effects and controls for pre-treatment (Rounds 1 and 2) Test Scores. Robust standard errors in parentheses.

Table 4: **Regressions of Behavior on Treatments**

VARIABLES	(1) Supports Social Distancing	(2) Followed Government Recommendation in past 14 days	(3) Preventive Action Practice in Past 14 Days
Incentive	0.0067	0.0278	0.0130
	(0.0040)	(0.0110)	(0.0072)
Teaching	-0.0175	0.0121	-0.0007
	(0.0085)	(0.0123)	(0.0075)
Incentive plus Teaching (Joint)	-0.0017	0.0104	0.0076
	(0.0058)	(0.0127)	(0.0072)
Observations	2,117	2,117	2,117
R-squared	0.067	0.065	0.278
Control Mean DV	0.992	0.945	0.764
Control SD DV	0.091	0.229	0.138
p-value: Incentive = Teaching	0.0051	0.2019	0.1122
p-value: Incentive = Joint	0.1398	0.1697	0.5232
p-value: Teaching = Joint	0.1053	0.9049	0.3361

Notes: Column 1: indicator equal to one if respondent answers "yes" to supporting "the practice of social distancing (SD) to prevent the spread of coronavirus" and zero otherwise. Column 2: indicator for SD according to self if respondent answered "yes" to observing the government's recommendations on SD in the last 14 days, and zero otherwise. Column 3: share of eight social distancing behaviors (Column 4) and five household prevention behaviors (Column 5) that the respondents report doing in the last 14 days. All regressions also include community fixed effects and controls for pre-treatment (Rounds 1 and 2) Test Scores. Robust standard errors in parentheses.

Table 5: **Regressions of Interactions of Knowledge Treatments and Social Distancing Treatments**

VARIABLES	(1) Overall Test Score (TS)	(2) Teaching-eligible TS	(3) Overall TS without SD Index	(4) Teaching-eligible TS without SD Index
Incentive	0.0159	0.00236	0.0205	0.0169
	(0.00862)	(0.00977)	(0.00619)	(0.00694)
Teaching	0.00318	0.0120	0.0199	0.0350
	(0.00882)	(0.0102)	(0.00620)	(0.00727)
Incentive plus Teaching	0.0477	0.0528	0.0581	0.0688
	(0.00842)	(0.00895)	(0.00636)	(0.00704)
Social Norm Correction (SNC)	-0.0101	-0.0151		
	(0.00764)	(0.00833)		
Leader Endorsement (LE)	-0.00797	-0.0169		
	(0.00728)	(0.00790)		
Incentive × SNC	0.00654	0.0159		
	(0.0128)	(0.0143)		
Incentive × LE	0.00677	0.0279		
	(0.0133)	(0.0147)		
Teaching × SNC	0.0181	0.0229		
	(0.0134)	(0.0152)		
Teaching × LE	0.0242	0.0323		
	(0.0136)	(0.0157)		
Incentive plus Teaching × SNC	-0.00304	0.000286		
	(0.0138)	(0.0151)		
Incentive plus Teaching × LE	0.00840	0.0161		
	(0.0130)	(0.0138)		
Observations	2,117	2,117	2,117	2,117
R-squared	0.322	0.336	0.291	0.311
Control Mean DV	0.781	0.784	0.748	0.751
Control SD DV	0.108	0.123	0.121	0.141

Notes: Dependent variable in Columns 1 and 2 defined in Table 1. Dependent variable in Column 3: Overall TS calculated without the 8 knowledge questions on social distancing actions – that is, the share of correct answers to 32 knowledge questions in Round 3: 12 on general knowledge, 8 on household preventive actions, and 12 on government actions. Dependent variable in Column 4: Feedback-Eligible TS calculated without the 4 Feedback-Eligible knowledge questions on social distancing actions. All regressions also include community fixed effects and controls for pre-treatment (Rounds 1 and 2) Test Scores. Robust standard errors in parentheses.

3 Further Analysis of Incentive Treatment

Table 6 presents analyses specified in the "Further Analysis for Knowledge Incentive (K1) Treatment" section of the PAP. These analyses were designed to explore possible mechanisms behind the treatments, particularly the Incentive treatment.

Table 6: **Analyses of Mechanisms for Incentive Treatment**

VARIABLES	(1) Previously-asked test score	(2) Newly-asked test score	(3) Count of info sources	(4) Count of official info sources	(5) Count of unofficial info sources	(6) Dummy: Most trusted source is official	(7) Confidence (1-4) that attentive during feedback	(8) Dummy: Asked to repeat feedback
Incentive	0.0189	0.0208	0.0095	-0.0128	0.0219	0.0020		
	(0.0056)	(0.0081)	(0.1022)	(0.0391)	(0.0862)	(0.0174)		
Teaching	0.0231	0.0018	-0.1124	-0.0348	-0.0771	0.0144		
	(0.0060)	(0.0078)	(0.0992)	(0.0399)	(0.0846)	(0.0170)		
Incentive plus Teaching (Joint)	0.0540	0.0414	0.0869	0.0569	0.0271	0.0308	-0.0126	-0.0012
	(0.0056)	(0.0080)	(0.1028)	(0.0393)	(0.0875)	(0.0162)	(0.0216)	(0.0166)
Observations	2,117	2,117	2,117	2,117	2,117	2,117	856	856
R-squared	0.344	0.150	0.420	0.201	0.417	0.180	0.094	0.151
Control Mean DV	0.784	0.777	3.243	1.796	1.447	0.888	3.936	0.068
Control SD DV	0.116	0.144	2.206	0.748	1.861	0.316	0.305	0.252
p-value: $\lambda = 0$	0.1725	0.1176	0.2156	0.0799	0.5271	0.5706		
p-value : Incentive = Teaching	0.5376	0.0351	0.2849	0.6229	0.3069	0.5300		
p-value: Incentive = Joint	0.0000	0.0239	0.5098	0.1160	0.9584	0.1287		
p-value: Teaching = Joint	0.0000	0.0000	0.0802	0.0415	0.2835	0.3752		

Notes: Column 1: the "Previously-asked test score" is the share of correct answers to the 20 or more knowledge questions in Round 3 that were also randomly asked of the respondent in Round 1 or Round 2: at least 6 on general knowledge, 8 on preventive actions, and 6 on government actions. Column 2: the "Newly-asked test score" is the share of correct answers to the 20 or fewer knowledge questions in Round 3 that were randomly not asked of the respondent in Round 1 or Round 2: at most 6 on general knowledge, 8 on preventive actions, and 6 on government actions. Columns 3: Count of possible official information sources including radio, TV, ATM screen messages, SMS messages from telecom companies (Column 4), and possible unofficial information sources WhatsApp, Facebook, family members, friends, health workers, community nonprofit/NGO, community leaders, religious leaders, traditional healers or midwives (Column 5). Column 6: Indicator equal to one if responded that most trusted information source is an official source, and zero otherwise. Columns 7-8: Questions on attentiveness and repeating feedback were asked to respondents in the Teaching and Joint treatment groups immediately after feedback was given. Attentiveness (column 7) measured with the question "How confident are you that you were able to pay attention to the feedback I just provided?" (1=Not Confident At All, 2=A Little Confident, 3=Mostly Confident, 4=Completely Confident, 0=refuse to answer). Repeating feedback (column 8) is indicator equal to one if respondent requested to repeat feedback; this regression also controls for the outcome in column 7. All regressions also include community fixed effects. Robust standard errors in parentheses.

First, Columns 1 and 2 show the effect of treatment on the share of correct responses to previously-asked and newly-asked questions, respectively. Because question selection was randomized within each question subcategory in Round 1 and Round 2, determination of which knowledge questions belong to each respondents previously-asked and newly-asked questions vary in ways uncorrelated to respondent characteristics.

This analysis finds that the Incentive treatment has significantly positive effects (and very similar point estimates) on both previously-asked and newly-asked questions. This suggests that, when incentivized to do so, respondents were just as successful at seeking out correct information on known details as they were at seeking out correct information on other details on known topics. The results also suggest that respondents do not suffer from an information constraint, as it shows they are able to seek out information on known topics simply when given more incentive to do so. The results in Columns 1 and 2 also highlight the major difference in the Incentive versus the Teaching treatment. Whereas Incentive has significantly positive effects (and very similar point estimates) on both previously-asked and newly-asked questions, Teaching only has a significantly positive effects on previously-asked questions, most of which were the Feedback-Eligible questions from Round 2. This difference can be explained by the mechanisms motivating the design of each treatment: the Incentive treatment increases the marginal benefit of all knowledge acquisition on the specified topics, while the Teaching treatment decreases the marginal cost of informational search for only those questions which were eligible for teaching.

Second, to see if treatment changed how respondents gathered information, the survey instrument also asked respondents if they heard about COVID-19 from a list of 13 possible sources (e.g., TV, radio, family members, community leaders, etc.), and then asked which source they trusted most. Columns 3-6 reveal that the treatments had little effect through this mechanism. Treatment did not change a respondent's total number of information sources, or (controlling for the total number of information sources) the number of official sources (e.g., TV, radio, government-sponsored ATM and SMS messages) or unofficial sources (all other). There is weak evidence that the Joint treatment led to increased trust of official sources.

Third, to see if the offer of incentives changed how respondents in the Joint treatment undertook the Teaching treatment, we asked respondents after receiving Teaching if they paid attention during teaching (on a 1-4 Likert scale) and if they would like the teaching repeated. We hypothesized that, because the marginal benefit of receiving and then repeating teaching is higher for respondents in the Joint treatment (due to the Incentive), respondents in the Joint treatment would report higher attentiveness and request repeated teaching more often (conditional on attentiveness) than those in the standalone Teaching treatment. This is one mechanism through which the interventions could be complements to each other. However, columns 7-8 reveal that this is not the case: in the subsample of respondents receiving any teaching, those in the Joint treatment (with the Incentive) do not report higher attentiveness (column 7) nor, after controlling for attentiveness, request repeated teaching more often (column 8). Thus, we find no evidence that the interventions are complements through this possible mechanism. It is worth noting that the dependent variable means suggest that nearly all respondents in the subsample reported maximum attentiveness and only 7% requested repeated teaching.

4 Further Analysis of Teaching Treatment

Table 7 presents analyses specified in the "Further Analysis for Tailored Feedback (K2) Treatment" section of the PAP. These analyses were designed to explore possible mechanisms behind the knowledge treatments, particularly the teaching intervention.

Table 7: **Analyses of Confidence in Feedback**

VARIABLES	(1) Confidence in R2 correct beliefs	(2) Confidence in R2 incorrect beliefs	(3) Share of incorrect clues	(4) Share of correct clues	(5) Share of correct clues
Incentive	0.8337	-0.0350	-0.0166	0.0113	0.0113
	(0.3645)	(0.3375)	(0.0133)	(0.0150)	(0.0150)
Teaching	-0.1420	-1.9793	-0.0061	0.0137	
	(0.4003)	(0.3419)	(0.0140)	(0.0159)	
Incentive plus Teaching (Joint)	0.6488	-2.9611	-0.0446	0.0603	
	(0.3645)	(0.3277)	(0.0122)	(0.0140)	
Correct-strong incorrect-weak feedback					0.0433
					(0.0200)
Correct-weak incorrect-weak feedback					0.0343
					(0.0185)
Correct-strong incorrect-strong feedback					0.0486
					(0.0183)
Correct-weak incorrect-strong feedback					0.0237
					(0.0201)
Observations	2,114	2,114	2,112	2,112	2,112
R-squared	0.314	0.085	0.101	0.320	0.318
Control Mean DV	19.311	2.155	0.119	0.829	0.829
Control SD DV	7.242	5.535	0.244	0.311	0.311
p-value: $\lambda = 0$	0.9406	0.0662	0.2640	0.1181	
p-value: Incentive = Teaching	0.0303	0.0000	0.5009	0.8962	
p-value: Incentive = Joint	0.6582	0.0000	0.0444	0.0026	
p-value: Teaching = Joint	0.0758	0.0114	0.0081	0.0060	

Notes: Column 1: Measured on a 9-point scale from -4 (incorrect, completely confident), -3 (incorrect, somewhat confident), -2 (incorrect, a little confident), -1 (incorrect, not confident), 0 (neutral, responded don't know or refuse to answer), 1 (correct, not confident), 2 (correct, a little confident, 3 (correct, somewhat confident), 4 (correct, completely confident) and summed across all 8 Feedback-Eligible preventive action questions. Column 2: the opposite of Column 1, calculated by multiplying Confidence in R2 correct beliefs by -1. Column 3: percent of clues nominated (out of a possible 2) that contained incorrect information. Columns 4 & 5: percent of clues nominated (out of a possible 2) that contained correct information. The independent variables "correct-strong incorrect-weak" through "correct-weak incorrect-strong" correspond to for randomized sub-treatments that determined which preventive action beliefs were eligible for feedback in the teaching intervention; see text for more details. All regressions also include community fixed effects. Robust standard errors in parentheses.

For the eight questions on preventive actions in Round 2, we elicit the subject's confidence in their answers on 0-4 Likert scale. We use the Likert scale to rank the subjects' confidence in their correct beliefs and in their incorrect beliefs. We then define strong (weak) correct beliefs as correct responses where confidence ranks in the upper (lower) half of the ranking. For example, if one has 6 correct responses, correct responses ranked 1, 2, and 3 are considered strong correct beliefs and responses ranked 4, 5, and 6 are considered weak correct beliefs. If the number of correct responses is odd, then the cutoff for strong vs. weak beliefs is N/2 +/- .5 where the +/- is randomly determined with equal probability. If there is a tie in the rank of the subjects' confidence in their correct beliefs, rank is determined arbitrarily so that rank is always unique. We use an identical procedure to define strong (weak) incorrect beliefs.

We then provided feedback for a subset of the responses, depending on the respondent's random assignment to one of four sub-treatments:

1. Correct-strong, incorrect-weak
2. Correct-weak, incorrect-weak
3. Correct-strong, incorrect-strong
4. Correct-weak, incorrect-strong

Thus, feedback may be provided for both correct and incorrect responses; for example, if selected to receive feedback on "drinking hot tea" as a preventive action, feedback will state: "For 'Drinking hot tea', you chose YES/NO/DON'T KNOW/REFUSE TO ANSWER. Your answer is CORRECT/INCORRECT. This action will NOT prevent spreading coronavirus to yourself and others."

Additionally, directly after answering the eight preventive action questions in Round 2 and before the feedback, we ask subjects to nominate two of their answers as "clues" to other people in the community who are also study participants (among questions they answered with yes or no). Respondents are told that "clues" may be chosen to be shared anonymously with other study participants in their community on a future survey. Respondents are then reminded of their answers to the preventive action questions and asked to select two or choose from the following other options: "I do not wish to share any actions", "Don't know", "Refuse to answer".

In Round 3 we ask again the answers to the same binary knowledge questions and confidence in their answers on the 0-4 Likert scale.

We use these data to test the following hypotheses: a) Teaching strengthens correct beliefs and weakens incorrect beliefs as measured by Likert-level confidence. b) Teaching weakens the propensity to suggest incorrect clues in Round 3 and increases the propensity to suggest correct clues. c) The most effective treatment to increase the share of positive clues is the (Correct-weak, incorrect-strong) sub-treatment which strengthens weakly held correct beliefs (and hence increases propensity to suggest these clues) and weakens strongly held incorrect beliefs (and hence decreases propensity to suggest these clues).

Results are in Table 7. First, Columns 1 and 2 show that Incentives and Teaching jointly have a strong effect on confidence: they raise confidence in correct beliefs and strongly lower confidence in incorrect beliefs. Teaching appears to be most effective in making people less confident when their beliefs are incorrect but has no effect when people hold correct beliefs. This likely follows from the fact that respondents' confidence is highly clustered around -4 (incorrect, completely confident) and $+4$ (correct, completely confident). Hence, positive feedback cannot move confidence any higher for most respondents and negative feedback moves most them all the way from -4 to $+4$. Columns 3 and 4 show that the Incentive plus Teaching (Joint) treatment has a significantly negative effect on the share of incorrect clues (minus 4%) and a significantly positive effect on the share of correct clues (plus 6%). However, an analysis of the effect of the four feedback sub-treatments on the share of correct clues does not confirm our initial hypothesis: the point estimate on the correct-weak/incorrect-strong sub-treatment is lower than for the other three sub-treatments. The two sub-treatments where we provide feedback on strongly held correct beliefs increase the share of correct clues the most.

5 Additional Figures

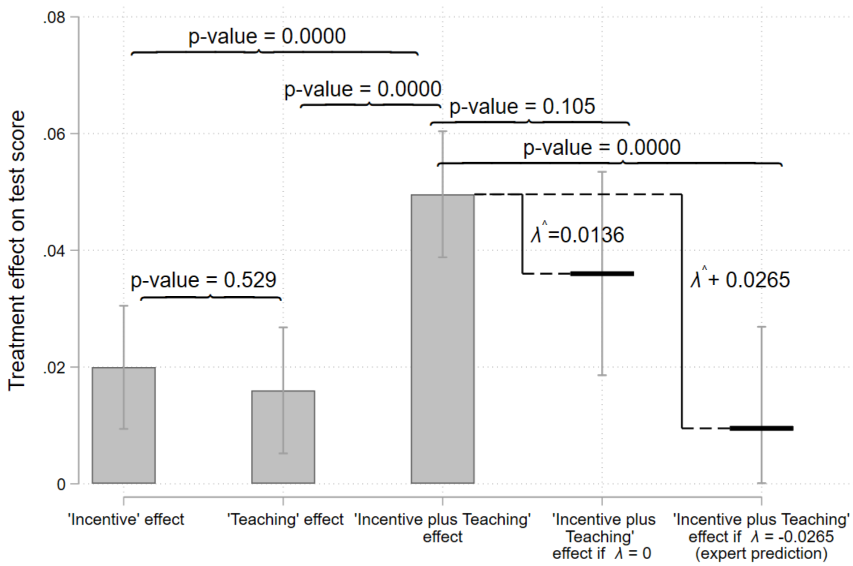
We show here additional figures that correspond to those in the main text, but that relate to the other pre-specified primary outcome (the Overall Test Score based on 40 COVID-19 knowledge questions). We show these to emphasize that key findings and conclusions are robust to examination of either of the two pre-specified primary outcome variables.

In Figure 1, we display in Panel (a) treatment effects and the complementarity parameter from analyses of the Overall Test Score based on 40 COVID-19 knowledge questions. The corresponding main text Figure 2 is replicated in Panel (b) for comparison. They key conclusion is stable across the two figures: the test that $\lambda = 0$ is rejected at marginal levels of statistical significance (in fact, in Panel (a) the p-value is a bit closer to conventional levels of statistical significance, at 0.105).

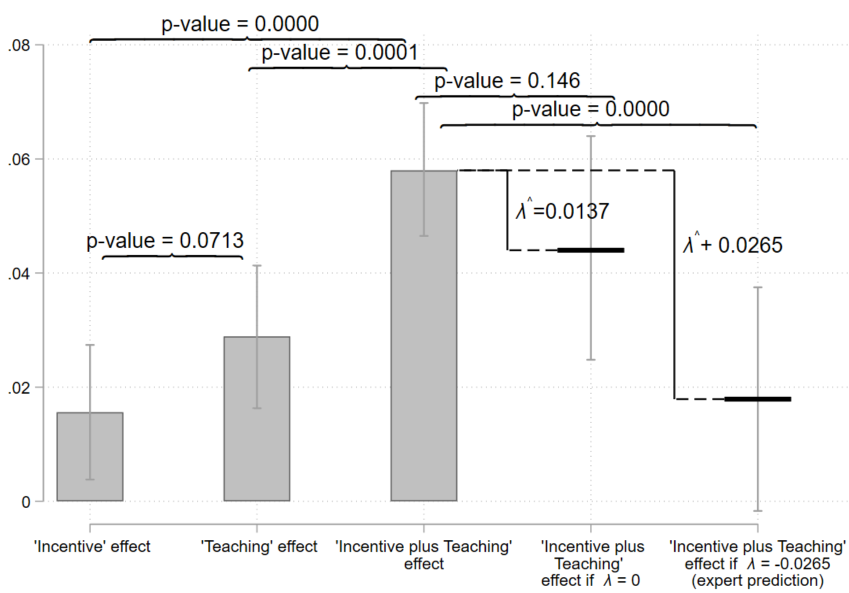
In Figure 2, we display in Panel (a) CDFs of the Overall Test Score based on 40 COVID-19 knowledge questions. The corresponding main text Figure 3 is replicated in Panel (b) for comparison. Both figures show that the Joint treatment is the most effective, shifting the CDFs of test scores furthest to the right.

Figure 1: **Treatment Effects and Test of Complementarity Parameter λ Against Benchmark Values**

(a) Overall Test Score



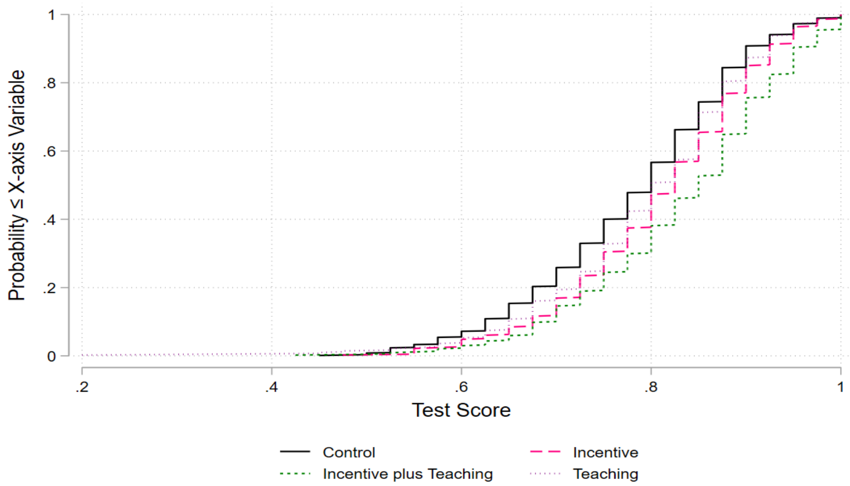
(b) Feedback-Eligible Test Score



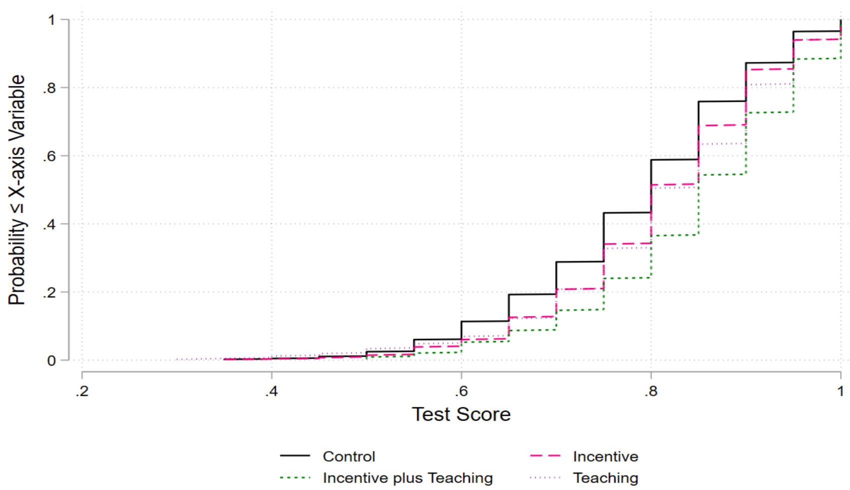
Notes: Overall Test Score is fraction of correct responses on COVID-19 knowledge out of 40 questions. Feedback-Eligible Test Score is a fraction of correct responses on COVID-19 knowledge out of 20 questions previous asked in the Round 2 (baseline) survey. In each panel of figure, bars in first three columns display regression coefficients representing treatment effects (and 95% confidence intervals) for "Incentive", "Teaching", and "Incentive plus Teaching" ("Joint") treatments. Floating solid horizontal lines in fourth and fifth columns display "Incentive plus Teaching" ("Joint") treatment effects that would be implied by different benchmark values of complementarity parameter λ. Difference between values in 3rd and 4th columns is actual estimated complementarity parameter, $\hat{\lambda}$; the test that this difference is equal to zero tests the null that $\lambda = 0$. Difference between values in 3rd and 5th columns is difference between $\hat{\lambda}$ and mean expert prediction, -0.0265; the test that this difference is equal to zero tests the null that $\lambda = -0.0265$.

Figure 2: **Cumulative Distribution Functions of Test Score by Treatment Group**

(a) Overall Test Score



(b) Feedback-Eligible Test Score



Notes: Overall Test Score is fraction of correct responses on COVID-19 knowledge out of 40 questions. Feedback-Eligible Test Score is a fraction of correct responses on COVID-19 knowledge out of 20 questions previous asked in the Round 2 (baseline) survey. Figure depicts the cumulative distribution function of this variable for the "Control" group, the "Incentive" treatment arm, the "Teaching" treatment arm, and the "Incentive plus Teaching" ("Joint") treatment arm.

References

Allen IV, J., A. Mahumane, J. Riddell IV, T. Rosenblat, D. Yang, and H. Yu (2021). Teaching and Incentives: Substitutes or Complements? *NBER Working Paper* (28976).

Duflo, E., A. Banerjee, A. Finkelstein, L. Katz, B. Olken, and A. Sautmann (2020). In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plan for RCTs in Economics. *NBER Working Paper Series W26993*.

Muralidharan, K., M. Romero, and K. Wuthrich (2019). Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments. *NBER Working Paper*.