Evaluating C-SEF for university students and staff

Pre-analysis plan

Simon Finster* Michelle González Amador† Edwin Lock‡
Francisco Marmolejo-Cossío§,¶

13th May, 2022

Abstract

We introduce the *COVID-19 safe education framework* (C-SEF), a novel mechanism for monitoring and safeguarding a population in educational institutions. The mechanism aims to mitigate the effect of COVID-19 restrictions on economic, psychological and educational well-being, a phenomenon that has been referred to as 'learning loss' in other similar instances.

We propose a utility-weighted algorithmic approach to pool testing individuals in a scarce-resource setting. The mechanism uses pooled testing to identify healthy individuals, who are permitted to return on-site, while guaranteeing that infected individuals continue to work off-site. The utility-based optimisation approach underlying our solution takes into account factors such as psychological well-being, need for academic resources and socio-economic status in order to optimally allocate testing.

To evaluate the efficacy and robustness of our proposed solution, we conduct a randomised controlled trial at the Potosinian Institute of Scientific and Technological Research (IPICYT). The outcome of our experiment will be measured in terms of performance and productivity, mental health and infection probabilities. Moreover, the evaluation results will allow us to make policy recommendations for better governance of educational institutions during pandemic times.

1 Introduction

We present an adaptable mechanism – the COVID-19 safe education framework (C-SEF) – for algorithmic group testing and institutional reopening, and evaluate it with students and staff at IPICYT (Potosinian Institute of Scientific and Technological Research), in San Luis Potosí, Mexico. We study the effect of this protocol on health, learning/productivity and psychological well-being with a two-group randomised controlled trial.

After two years of an evolving COVID-19 pandemic, educational institutions around the world face the pressing issue of fully reopening their facilities in a safe way. With approximately 79% of its population vaccinated with at least two doses (Bhatia et al., 2022), and with a nation-wide policy and effort to prioritise vaccination of the population in the education sector, Mexican universities are preparing to fully reopen for the 2022 autumn term. Our partner institute, the Potosinian Institute of Scientific and Technological Research, plans to reopen their doors to the students and faculty from June 2022 onward. In order to do so safely, we propose a principled reopening with a pilot of C-SEF, a utility-based protocol for group-testing.

*Nuffield College & Department of Economics, University of Oxford and University of Hagen, simon.finster@fernuni-hagen.de
†UNU-MERIT & Maastricht University, School of Business and Economics, mgonzalez@merit.unu.edu
‡Nuffield College & Department of Economics, University of Oxford, edwin.lock@economics.ox.ac.uk
§Harvard School of Engineering and Applied Sciences, fjmarmol@seas.harvard.edu
¶Authors ordered alphabetically.

At present (May 2022), IPICYT's institutional policy is for all community members to work remotely. Students, researchers, and some administrative staff are working from home and enter university premises only in exceptional circumstances; exceptions are made at the discretion of the department heads. While evidence suggests that remote schooling, teaching, and research activities lead to suboptimal performance and other negative outcomes (Marshall et al., 2020; OECD, 2020), the high infection rates in San Luis Potosí throughout the pandemic and an extremely limited number of tests left no other viable option in order to ensure the safety of its community. Many educational institutions with budget constraints face similar problems, and especially so in low- and middle-income countries (LMICs) where testing supplies are scarce.

To help educational institutions in their reopening stage for in-person learning safely and effectively, we have devised C-SEF, a monitoring protocol with an algorithmic base. The protocol's main objective is to make optimal use of scarce testing resources and provide a systematic approach to using the physical facilities of the institution. C-SEF uses a utility-based approach to group and test the relevant population in pools such that testing resources are used efficiently while addressing educational inequality and fairness considerations. The utility-based approach of the algorithm behind the C-SEF protocol also addresses equity concerns that may arise from random or human-based decision-making. We evaluate the efficacy of C-SEF on the full population of students, researchers, and staff at IPICYT in a two-group randomised controlled trial. The treatment group will follow the protocol, and the control group will continue with the current institutional policy. We are interested in learning the effects of the C-SEF protocol on learning, performance and productivity, infection probabilities, and mental health outcomes. The results of our trial should allow us to determine the causal effect of the proposed mechanism and inform policymakers of alternative approaches to keeping a workforce, students, or other populations safe in their work and learning environment during a pandemic. Importantly, and in contrast with IPICYT's current plans for a June 2022 reopening, our approach to testing and reopening would guarantee the safety of the student and staff population, while simultaneously making sure that scarce testing resources are used optimally. Each participating individual's right to health and safety, as well as personal liberties in the context of their work and study place will be maintained.

2 Experimental design

We implement a mechanism to ensuring the safety of an educational institution (C-SEF) on the academic population at IPICYT in San Luis Potosí, Mexico. We randomly assign individuals into a treatment and a control group. Subjects in the treatment group follow a pool-testing protocol and are allowed to freely use the facilities of the university so long as they test negative. Individuals in the control group are expected to work remotely and only use the facilities for emergencies[1].

2.1 Preliminaries and notation

We briefly define some notation to formalise our setup, the algorithms used, and the methodology to evaluate our study. A subject in our pool of participants is denoted by $i \in [n] = \{1, \ldots, n\}$. Each subject belongs to one of the three disjoint sub-populations of students $\mathcal{S}$, research staff $\mathcal{R}$, or administrative staff $\mathcal{A}$. Our treatment variable is denoted by $\tau = 0$ (control) and $\tau = 1$ (treated). Each subject is assigned to either the treatment or control group, and accordingly we set $\tau_i \in \{0, 1\}$. Subject i's assigned treatment is equivalent to their treatment status[2].

[1] The definition of an emergency is at the discretion of the head of department to which the student/staff belongs.

[2] Our proposed analysis strategy is therefore a conservative Intention to Treat (ITT), under the assumption of some non-compliance. We further propose to estimate the Average Treatment Effect on the Treated (ATET) with a fully complying subset of the population.

2.2 Randomisation

The physical isolation of treatment group and control groups is essential for our protocol to work; both to avoid health spillovers and disentangle psychological dynamics. The IPICYT campus lends itself, for the most part, to a two-group randomisation approach. The campus consists of two similar buildings for research in natural sciences, two similar buildings for research in computer science and mathematics, one building for classes, and one building for administration. The individuals participating in our trial are randomly assigned to the treatment and control group. Note however that some individuals in the control group will still work on-site based on exceptions. Therefore, we use the following strategy: one of the natural science buildings is randomly assigned to be treated and one of the CS and math buildings is randomly assigned to be treated. Crucial for this approach to be effective is that individuals in both buildings are comparable. Given IPICYT's reports about their staff, we know that both research students and researchers are assigned to work in each of the buildings contingent only on their academic discipline, based on no individual characteristics important for the evaluation of our framework. Hence, we may consider the assignment as pseudo-random. Nevertheless, we further collect a number of covariates to conduct a balance analysis[3].

Within the staff building, we also need to physically separate the space allocated to treatment and control group. Our strategy here is to separate the four-storied building by floors. We randomly select either the top two or the bottom two floors for treatment, and the remaining two floors for control. Similar to the other two buildings, we accept that the assignment is pseudo-random.

Students can be assigned to classrooms depending on availability and suitability for the course taught. Moreover, students in the control group will exclusively participate in online learning. Therefore, each individual in the student population is assigned randomly to either treatment or control.

2.3 Treatment group

Individuals assigned to the treatment group are invited to participate in the C-SEF protocol. All information about the institute-wide study is shared with the IPICYT population via an online communication campaign co-produced by the principal investigators and the communications department of the Institute, and hosted at the 'C-SEF at IPICYT website'. Clear and direct institutional communication to students and the community at large is key to guarantee protocol engagement and successful reintroduction of face-to-face activities (cf. Lorenc et al. (2021); Mukherjee et al. (2021); Sheikh et al. (2020)). Consent is obtained from each individual via a form, which precedes a survey that collects baseline data. There exists no natural disincentive to participate in the study, and on the contrary, participation may indirectly give participants more freedom of movement, so we confidently estimate study compliance to be at nearly 100%.

The C-SEF protocol uses a utility-weighted selection algorithm that groups individuals for pool testing. Individuals who test negative are allowed to attend their work or study activities on campus. Individuals who test positive are asked to isolate and work remotely until their next test.

2.3.1 Testing mechanism

We propose a COVID-19 testing and reopening mechanism that uses pooled testing to systematically monitor and reopen research and educational institutes. Our data-driven test allocation algorithm decides who to test, and how to pool them, in order to maximise overall expected social welfare. In the pooled testing regime, saliva samples from multiple individuals (the 'pool') are pooled and tested as one. A positive result indicates that at least one individual is infected, and a negative result implies that the entire pool is healthy. Our containment mechanism allows all individuals in negative pooled tests to enter the institute, as they are healthy, while individuals in positive tests may not, as one or more individuals in the pool are infected. Aside from guaranteeing that only healthy individuals enter the institute, our use of pooled testing also allows us to monitor the prevalence of COVID-19 among the institute's population with a limited number of tests.

[3]Should there be a clear imbalance, we will subscribe to a post-experimental matching approach as proposed by Bruhn and McKenzie (2009)

The welfare objective we propose addresses existing inequalities by prioritising individuals with lower socio-economic status as well as mental health or limited access to academic or other essential resources. To construct our objective, we assume that each individual derives a utility gain from working or studying on-site, as opposed to remotely. This utility gain is larger if the student is disadvantaged in any way. For example, a student who shares a room with other family members may not be able to work effectively from home, and would benefit more from returning to the institute than an affluent student who has their own study room at home. We further assume that each individual is healthy with a certain probability. These utilities and probabilities of infection are constructed in a separate procedure from data collected from each individual (cf. Section 2.3.3). Note that utility gains of individuals are realised only if they are permitted to return to the institute, and this is the case only if they are contained in a pooled test that is negative. Our algorithm allocates tests in such a way that the total expected utility of individuals who test negative, and may thus return to the institute, is maximised.[4]

The algorithm described below determines a 'fair' allocation of tests between all individuals entering the algorithm. However, the algorithm is run daily, and therefore we consider fairness also in this dynamic context. For this, we enact the following rules:

(i) We restrict an individual's window of access to campus facilities to two days (48 hours). This ensures that a negatively tested individual remains healthy throughout the period of time they are permitted to enter the premises, with high probability.

(ii) We allow individuals in a positive group test to be eligible for re-testing the next day. Individuals in testing pools with a positive result may be prohibited from entering the institute even if they are healthy, as only one positive sample in the pool would be sufficient to restrict five people's access. However, individuals in such positive pools may be re-considered by the testing algorithm on subsequent days, providing a 'second chance' at gaining access to the institute.

(iii) We restrict negatively tested individuals who have been given on-site access for four consecutive days (96h) from entering the testing pool for the following two days (48h), so that everyone gets a chance of being selected for testing at some point.[5]

2.3.2 Testing algorithm

We formulate our problem as a non-linear optimisation problem in order to maximise welfare as follows. We have n individuals $[n] := \{1, \dots, n\}$, and every person $i \in [n]$ has a probability q_i of being healthy and a utility gain u_i from working in the institute (as opposed to online). A *pooled test* is represented by a vector $\boldsymbol{x} \in \{0,1\}^n$, where $x_i \in \{0,1\}$ denotes whether person i is contained in the test. In order to minimise the possibility of false negatives, pools are typically limited to sizes of 5 or 10 (depending on the testing protocols and testing sensitivity). We impose a pool size limit of $G = 5$, which is captured by $\sum_{i\in[n]} x_i \leq G$. Moreover, the number of tests available (our testing budget) is B. We index tests with a superscript $t \in [B]$, and call $(\boldsymbol{x}^t)_{t\in[B]}$ a *testing allocation* with B tests. We require that tests are non-overlapping, i.e. every person is contained in at most one test. This is captured by the constraints $\sum_{t\in[B]} x_i^t \leq 1$ for all individuals $i \in [n]$.

For each test $\boldsymbol{x}$, the probability that it is negative is given by $\prod_{i\in[n]} q_i^{x_i}$.[6] The test contributes a utility of $\boldsymbol{x}\cdot\boldsymbol{u}$ if negative and 0 otherwise. Hence the expected utility of test $\boldsymbol{x}$ is given by $(\boldsymbol{x}\cdot\boldsymbol{u})\prod_{i\in[n]} q_i^{x_i}$. Moreover, as tests are non-overlapping, the expected welfare of a testing allocation is the sum of the expected utilities

[4] Our formulation of the problem as a non-linear optimisation problem allows for addition of further fairness constraints, such as gender balance or proportional test allocation between students and researchers, or different departments. Such constraints are not used in this trial.

[5] A researcher with 48 hours of on-site access may still be restricted how much performance they can gain, as on-site work may often not be concluded within this time frame. For this reason, we allow individuals to be eligible for testing a second time per week.

[6] Here we assume that each individual's probability of being healthy is independent. This simplification is commonly assumed in epidemiological models (Vinitsky et al., 2021).

for each test in the allocation. Our objective is to maximise overall welfare, which leads to the following non-linear optimisation problem.

$$\max_{\boldsymbol{x}} \quad \sum_{t=1}^{B} \left[(\boldsymbol{x}^t \cdot \boldsymbol{u}) \prod_{i=1}^{n} q_i^{x_i^t} \right] \tag{1a}$$

$$\text{s.t.} \quad \sum_{t=1}^{B} x_i^t \leq 1, \quad \forall i \in [n], \tag{1b}$$

$$\sum_{i=1}^{n} x_i^t \leq G, \quad \forall t \in [B], \tag{1c}$$

$$x_i^t \in \mathbb{N}_0, \quad \forall i \in [n], \forall t \in [B] \tag{1d}$$

This optimisation problem can be solved in a variety of ways. For small problem instances, a non-linear solver such as SCIP can be used. For larger instances, such as the ones we encounter in the pilot study, integer linear programming reformulation techniques can be applied to formulate integer linear programs that either exactly or approximately solve the optimisation problem. These programs can then be solved with commercial solvers such as CPLEX or Gurobi. For this pilot, we work with Gurobi.

We note that we could introduce population *clusters* to improve computational tractability. Clusters capture the notion of grouping people who have identical utility and probability of being healthy. In particular, if everyone in the population has a unique pair of utilities and health probabilities, each person forms their own cluster. However, as we expect clusters greater than 1 to occur, this reduces the size of the optimisation problem. After a testing allocation for clusters is computed, individuals from clusters are selected uniformly at random for inclusion in testing pools.

2.3.3 Constructing utilities and probabilities of being healthy

Utilities. We estimate utilities using a weighted scoring system that partially emulates Noothigattu et al. (2021). Each participant i in the IPICYT population $[n]$ answers a series of questions, listed in the baseline survey in Appendix A, that capture their perceived socioeconomic status, mental health, and access to and need of learning and research resources. The answers to some of these questions result in different utility increments. For each relevant question, the utility increments are given in parentheses (but are not shown to the subject). All questions are such that only one answer can be selected. These utility increments are aggregated in a weighted sum to obtain each individual's utility score, denoted by u_i for all subjects $i \in [n]$. In the following, we describe how u_i decomposes down to each individual answer in the survey.

The utility score consists of three sub-scores: u_i^{pr} captures the utility that subject i is hypothesised to gain from increased productivity when studying or working on-site; u_i^{psy} captures the benefit on subject i's mental health from attending in-person; u_i^{se} is an additional utility bonus for socio-economically disadvantaged individuals, reflecting the fact that those individuals are disproportionately affected by quarantining or remote work. The overall utility score is given by

$$u_i = \sum_{k \in \{pr, psy, se\}} w^k u_i^k$$

The weights w^{pr}, w^{psy}, and w^{se} are determined as a convex combination of $(1/3, 1/3, 1/3)$ and $(\hat{w}^{pr}, \hat{w}^{psy}, \hat{w}^{se})$, which represent the weighting preferences of IPICYT.[7]

[7] These weights are determined as the average of private, anonymous submissions of decision makers and department heads, as in Noothigattu et al. (2021).

We further define the composition of u_i^{pr}, u_i^{psy}, and u_i^{se}. Let $P_{i,z}^k$ denote the number of points "achieved" by the answer of subject i to question z in category $k \in \{pr, psy, se\}$. Let Z^k denote the number of questions relevant in category k. Relevance of a question to a specific category is marked in the survey in Appendix A by the corresponding abbreviation just after the question numbering.[8] Then, for each category, the score is computed as

$$u_i^k = \frac{1}{Z^k} \sum_z P_{i,z}^k$$

In our trial, the survey is taken only once before the start of the trial, but in a longer application of C-SEF it would be reasonable to update utilities in regular intervals.

Health probabilities. For the purpose of the trial, the probability of an individual being healthy is identical for every subject in a given age group. The baseline probability for a given age group is determined using local public health data. In contrast with utilities, health probabilities are immediately updated for those subjects who were tested, based on the results of pooled tests. As a consequence, subjects who have not yet been allocated a test yet in the trial are, ceteris paribus (utility component), weakly prioritised to individuals who have been allocated a test already. Details on the posterior probabilities of infection are given in Section 2.5.1.

2.4 Control group

Individuals who are randomly assigned to the control group are invited to participate in the study via the same online communication strategy as the treatment group. Consent is obtained through an online form alongside a survey with baseline individual-level data. They are asked to continue respecting the current institutional COVID-19 policy: remote work is the norm and access to the building is only granted as an exemption for urgent matters. Since this is the current status quo, we expect full compliance of the group.

2.4.1 Testing the control group

We cannot hope to test the entire control group. Evidently, individuals need to be notified of their test results. This means that for individuals who tested negatively, the willingness *not* to work on-site despite a negative test (but due to being part of the control group) will be low. Therefore, we only invite a random sub-sample of the control group for regular tests.

2.5 Outcomes

We are interested in three main outcomes: the average probability of infection among participants, the subjects' stress level and subjective well-being, as well the subjects' self-assessed performance and productivity.

2.5.1 Infection rates

Re-introducing face-to-face learning and working is a challenge that has been approached from different perspectives, especially in developed countries. It has been shown that there exists a negative correlation between institutional testing and infection rates (Mukherjee et al., 2021). Therefore, we aim to detect the presence of a similar *causal* relationship in our trial. Because we use pooled tests, we need to be careful about our measure of infection. We choose posterior probabilities of infection (or being healthy) as they accurately represent the heterogeneous prevalence of pooled tests across and within treatment and control groups.

[8] Not all questions are relevant for the construction of utilities.

Posterior probabilities of infection For each individual, we hypothesise a uniform prior probability of infection p_i and corresponding probability of being healthy $q_i = 1 - p_i$ for all $i \in [n]$. After each test, we update the probability of infection according to the test result. Let $t(i)$ denote the test individual i is allocated to, let $R_{t(i)} \in \{0, 1\}$ (where 0 means negative and 1 positive) denote the outcome of test $t(i)$, and let $I_{t(i)}$ denote the pool of individuals tested in test $t(i)$. Then subject i's posterior probabilities of infection are given by[9]

$$p_i \mid [R_{t(i)} = 1] = \frac{p_i}{1 - \prod_{q \in I_{t(i)}} q_j}$$
$$p_i \mid [R_{t(i)} = 0] = 0$$

2.5.2 Mental health outcomes

Mental health problems related to social isolation as a consequence of the COVID-19 pandemic have been documented for students and the general population (Martinez Arriaga et al., 2021). We conjecture that putting in place a safe education protocol decreases stress levels among students, researchers, and staff, by increasing sociability (Becchetti et al., 2017) and modulating the perception of health risk in the institute (Shan et al., 2022). Consequently, subjective well-being may also be positively affected. Stress is measured via the validated 4-item Perceived Stress Scale by Sheldon Cohen (Cohen et al., 1994) and we use a variation of the European Quality of Life Survey measure of subjective well-being, using a life/subject evaluation approach (OECD., 2013)[10]

2.5.3 Performance, productivity, and learning.

The pandemic has disrupted learning processes and decreased productivity of Mexican students (Limón-Vázquez et al., 2020; Martinez Arriaga et al., 2021) and female researchers (King and Frederickson, 2021). A significant portion of this downfall in productivity may be due to remote work with limited access to the necessary resources for work, research, and learning. We conjecture that our testing protocol improves (self-assessed) productivity and performance (in learning environments).

We use a composite score for the evaluation of performance and productivity. Let $P_{i,z}^{ppa}$ denote the number of points "achieved" by the answer of subject i to question z pertaining to 'Performance, productivity, and sense of achievement'. Let Z^{ppa} denote the number of relevant questions. Then the score is computed as

$$p_i = \frac{1}{Z^{ppa}} \sum_z P_{i,z}^{ppa}$$

3 Power and sample size determination

We estimate detectable effect sizes and statistical power given the three possible outcomes outlined in Section 2.5. For reduced infections, we assume the reported posterior probabilities is a continuous variable and estimate Cohen's d with a two-sample t-test. For mental health (stress score) and improved performance (self-assessment), we estimate and report achievable power given a set of conservative probabilities per category, i.e. we perform an ordinal response power calculation. Table 1 presents the results of our computations.

With a standard power score of 0.8 for number of infections at $\alpha = 0.01$ and $\alpha = 0.05$ and a conservative total $N = 500$ (from a population of 570), we are within the small to moderate effect size detection zone (Cohen,

[9] The posterior probabilities of being healthy could be used equivalently and are given by $q_i \mid [R_{t(i)} = 1] = \frac{\left(1 - \prod_{j \in I_{t(i)} j \neq i} q_j\right) q_i}{1 - \prod_{q \in I_{t(i)}} q_j}$ and $q_i \mid [R_{t(i)} = 0] = 1$.

[10] Note that we use baseline Stress and Subjective Well-being in our utilites' computation. On the other hand, we use endline Stress and Subjective Well-being to analyse between-group intervention effects.

	Power $(1-\beta)$	**Cohen's** d	α	**Total** N
Infections	0.80	0.251	0.05	500
Mental health	1	0.5	0.05	500
Productivity	1	0.5	0.05	500

Table 1: Power analysis, performed using the R packages pwr & Hmisc, under the assumption of no clustering

1992). For mental health outcomes, in particular the 5 point stress score, we fall within the moderate to large effect size detection zone. For conservative $N = 500$ and $\alpha = 0.05$, we obtain a power score of 1 for medium effects size detection, far above the conventional 0.8. With no theoretical benchmark[11]. Finally, for $N = 500$ and $\alpha = 0.05$, the estimated power score for our 5 point productivity measure is well above the benchmark 0.8[12]. Our conservative estimates show that the sample size of the experiment will allow us to confidently report moderate to large effect sizes.

Moreover, we calculate power under the assumption of intra-class correlation between academic subdisciplines, as per our randomisation strategy based on building divisions in Section 2.2.

	Power $(1-\beta)$	**Cohen's** d	α	**Total** N
Infections	0.80	0.47	0.05	500
	Power $(1-\beta)$	**Odds ratio**	α	**Total** N
Mental health	0.80	2	0.05	500
Productivity	0.80	2	0.05	500

Table 2: Power analysis, performed using the R packages ClusterPower, WebPower and an ad-hoc simulation-approach with 1000 replications, under the assumption of clustering

Our results show that in the testable presence of intra-class correlation, we can confidently report medium to large effect sizes for our three outcomes of interest. To perform the clustered-randomised power calculation we use an ICC of 0.22 and a cluster standard deviation of 0.5, as per the literature on educational institutions (Hedges and Hedberg, 2007). Moreover, we assess power on the basis of odds ratio for the likert-type outcomes mental health and productivity, which can be thought of as follows: an odds ratio of 1.48 can be considered a small effect size or the equivalent to a Cohen's d of 0.2; an odds ratio of 3.45 can be considered a moderate effect size or the equivalent to a Cohen's d of 0.5; finally, an odds ratio of 9 can be considered a large effect size or the equivalent to a Cohen's d of 0.8.

4 Metrics and methods

We propose a two-group experimental design where $[n] = 570$ subjects are randomly assigned to either a treatment or a control group. Group balance in observed and unobserved heterogeneity is a direct result of random assignment, allowing for treatment status to be the only source of exogenous variation. As such, the mean group difference in the outcomes of interest can be presented as the causal effect of C-SEF along those dimensions.

We test each hypothesis in Section 2.5 for the entire sample as well as separately on each sub-population $\mathcal{S}, \mathcal{R}$, and $\mathcal{A}$ (students, researchers, administrative staff). We estimate the average treatment effect[13] on the

[11] We estimate these scores using a vector of conservative probabilities for each ordered level in the 5 point factor scale, $p = \{.2, .2, .3, .2, .1\}$, where 1 means no stress and 5 is the highest level of stress.

[12] Again, we use a vector of conservative probabilities for each ordered level such that $p = \{.1, .1, .3, .4, .1\}$, where 1 means poor, 2 means below average, and 3 means average, 4 is above average and 5 is high

[13] We previously explain that, while we hope to deliver an ATE, we are likely to deliver ITT results, and further complement these with ATET for compliers and relevant k subsets of the population.

four, previously introduced, outcome variables.We denote our outcomes as $y_i \in \{p_i, s_i, w_i, \pi_i\}$:

- The average posterior probability of infection of all individuals that were tested at least once during the trial[14]
- The average stress level and subjective well-being, measured by each individual's stress score s_i and well-being score w_i
- Subjects' self-assessed learning [15] and performance score π_i

The average treatment effect is estimated with a simple linear model. Let Y denote the stacked vector of outcomes $(y_1, \ldots, y_n)$ and similarly for p_i, s_i, w_i, π, and treatment dummy τ_i. Let β denote the vector of parameters to be estimated. The independent variables are subsumed in $X = (1^n, \mathrm{T}, C)$, where 1^n is an $n-$vector of ones, and C is a matrix of covariates. Let ϵ denote the vector of error terms ϵ_i. We estimate the model

$$Y = X\beta + \epsilon$$

for $Y \in \{S, W, \Pi\}$ and test the hypothesis $\beta_1 = 0$. Let $\widetilde{P}$, $\widetilde{X}$, and $\widetilde{\epsilon}$ denote the vectors and matrices where we only stack individuals that were tested at least once in the trial. We also estimate

$$\widetilde{P} = \widetilde{X}\beta + \widetilde{\epsilon}$$

and again test the hypothesis $\beta_1 = 0$. Note that the well-being score, stress score, and performance and productivity score are ordinal variables, i.e. we estimate an ordinal logistic model with bootstrapped H2C standard errors.

4.1 Covariates

As part of the project, we collect additional socioeconomic and psychosocial data of participants. These data will be used twofold. First, some of these features enter into the utility estimations needed for the testing algorithm, as explained in Section 2.3.2. Second, we use relevant features/variables to check for group balance and, as needed, as covariates in our proposed in linear and non-linear models.

- Socio-economic attributes: gender, age, ethnicity, educational affiliation, perceived socio-economic status, financial dependants.
- Academic or job resources: access to internet, access to job materials, need to collaborate in person, access to a dedicated working space outside of the office.
- Psychosocial attributes: Sociability, Fear of the virus, subjective well-being (generalized).

All covariates and their measurement strategy can be found in the baseline survey in Appendix A.

4.2 Data sources and attrition

The experiment will be carried out over the course of a month, in collaboration with IPICYT (Instituto Potosino de Investigación Científica y Tecnológica), in San Luis Potosí, Mexico. For the duration of the experiment, community members will be reporting to classes and to work under two different regimes: C-SEF and remote work policy. Participation in the study will require little effort from community members/subjects, and we therefore expect attrition to be close to zero. In the case of attrition, we will conduct an analysis of the data to understand whether missingness, otherwise known as missing data due to attriters or nonresponse, is independent of potential outcomes or systematic, and how it may bias our results and/or estimations. All findings will be reported in the final study.

[14] We assume a normal distribution but estimate a beta-regression model as a robustness check.
[15] We adapt a measure based on the fit of 5 point likert scales, as per Versteeg and Steendijk (2019)

5 Ethical considerations

Carrying out a trial in the midst of a global pandemic is not without risk and, therefore, ethical concerns. We adhere to international and local safety standards, and guarantee to provide health security and anonymity, as well as the freedom to withdraw from the experiment at any point, to anyone that may need or want it, irrespective of their gender, socioeconomic status, department affiliation, position or rank within the institute, or experimental condition assignment. An essential component of our mechanism is guaranteeing that all resources are efficiently and optimally used to preserve the well-being of our target communities, while simultaneously addressing fairness concerns that arise from unstructured or non-systematic approaches to viral monitoring and safekeeping; the control group is therefore a safe status quo. We emphasise that IPICYT plans to reopen entirely, for students, faculty, and staff, from June 2022 onward. Without the C-SEF protocol, the risk of contagion for individuals on-site would be significantly higher. Indeed, our approach only allows people who tested negatively on campus. Moreover, access is restricted to 48 hours from submitting the test sample, which minimises the risk of individuals becoming infectious within the access window.

Our pooling protocol approaches the problem of optimising the use of scarce testing resource by maximising efficiency while incorporating aspects of fairness and minimising the risk of contagion. We see each individual as having certain immutable and equal rights: (a) the right of not being exposed to undue health risks, (b) the right to needs-based priority access to testing, and (c) the right to personal liberties related to their IPICYT position while not endangering the health of other individuals and the university as a whole. Our mechanism aims to provide the greatest benefits to the least advantaged individuals in our population of participants. At the same time, we are concerned with practicability and algorithmic feasibility. Therefore, we maximise an aggregate welfare function in the static framework that is executed daily. Moreover, we rely on additional constraints and relaxations to introduce fairness and ensure minimal risk in the dynamic framework. To do so, we (i) restrict an individual's window of access to campus facilities, (ii) allow individuals in a positive group test to be eligible for re-testing the next day, and (iii) restrict negatively tested individuals with current access from entering the testing pool more than twice in a given week. In the following, we give additional details and justification for choosing our constraints:

(i) By setting the access window to two days (48 hours), we ensure that a negatively tested individual remains healthy throughout the period of time they are permitted to enter the premises, with high probability.

(ii) Individuals in testing pools with a positive result may be probihibited from entering the institute even if they are healthy, as only one positive sample in the pool would be sufficient to restrict five people's access. However, individuals in such positive pools may be re-considered by the testing algorithm on subsequent days, providing a 'second chance' at gaining access to the institute.

(iii) A researcher with 48 hours of on-site access may still be restricted how much performance they can gain, as on-site work may often not be concluded within this timeframe. For this reason, we allow individuals to be eligible for testing a second time per week. In order for everyone to get a chance of being selected for testing, however, we impose an upper bound of four days (96 hours) on on-site access: any individual who has has reached the maximum four days of consecutive access to the institute will not be considered for testing in the following two days (48 hours).

By enacting a trial phase, we provide scientific evidence for the impact of our framework. If it evaluates positively, IPICYT may use this approach for any duration they deem necessary to keep their institute safe, while promoting greater work and study performance as well as mental well-being.

References

L. Becchetti, L. Corrado, and P. Conzo. Sociability, altruism and well-being. *Cambridge Journal of Economics*, 41(2):441–486, 2017.

G. Bhatia, P. K. Dutta, and J. McClure. Mexico: The latest coronavirus counts, charts and maps, May 2022. URL https://graphics.reuters.com/world-coronavirus-tracker-and-maps/countries-and-territories/mexico/.

M. Bruhn and D. McKenzie. In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4):200–232, 2009.

J. Cohen. Statistical power analysis. *Current directions in psychological science*, 1(3):98–101, 1992.

S. Cohen, T. Kamarck, R. Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10(2):1–2, 1994.

L. V. Hedges and E. C. Hedberg. Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1):60–87, 2007.

M. M. King and M. E. Frederickson. The pandemic penalty: The gendered effects of covid-19 on scientific productivity. *Socius*, 7:23780231211006977, 2021.

A. K. Limón-Vázquez, G. Guillén-Ruiz, and E. V. Herrera-Huerta. The social isolation triggered by covid-19: Effects on mental health and education in mexico. *Health and Academic Achievement-New Findings*, 2020.

A. Lorenc, J. M. Kesten, J. Kidger, R. Langford, and J. Horwood. Reducing covid-19 risk in schools: a qualitative examination of secondary school staff and family views and concerns in the south west of england. *BMJ paediatrics open*, 5(1), 2021.

D. T. Marshall, D. M. Shannon, and S. M. Love. How teachers experienced the covid-19 transition to remote instruction. *Phi Delta Kappan*, 102(3):46–50, 2020.

R. J. Martinez Arriaga, L. P. Gonzalez Ramirez, J. M. de la Roca-Chiapas, and M. Hernández-González. Psychological distress of covid-19 pandemic and associated psychosocial factors among mexican students: An exploratory study. *Psychology in the Schools*, 58(9):1844–1857, 2021.

U. K. Mukherjee, S. Bose, A. Ivanov, S. Souyris, S. Seshadri, P. Sridhar, R. Watkins, and Y. Xu. Evaluation of reopening strategies for educational institutions during covid-19 through agent based simulation. *Scientific Reports*, 11(1):1–24, 2021.

R. Noothigattu, N. Shah, and A. Procaccia. Loss functions, axioms, and peer review. *Journal of Artificial Intelligence Research*, 70:1481–1515, 2021.

OECD. *OECD guidelines on measuring subjective well-being*. OECD, 2013.

OECD. *Education and COVID-19: Focusing on the long-term impact of school closures*. Orginzation for Economic Cooperation and Development Publishing, 2020.

B. Shan, X. Liu, A. Gu, and R. Zhao. The effect of occupational health risk perception on job satisfaction. *International Journal of Environmental Research and Public Health*, 19(4):2111, 2022.

A. Sheikh, A. Sheikh, Z. Sheikh, and S. Dhami. Reopening schools after the covid-19 lockdown. *Journal of global health*, 10(1), 2020.

M. Versteeg and P. Steendijk. Putting post-decision wagering to the test: a measure of self-perceived knowledge in basic sciences? *Perspectives on Medical Education*, 8(1):9–16, 2019.

S. Vinitsky, A. Gusev, V. Derbov, P. Krassovitskiy, F. Pen'kov, and G. Chuluunbaatar. Reduced sir model of covid-19 pandemic. *Computational Mathematics and Mathematical Physics*, 61(3):376–387, 2021.

A Survey

1. **Socio-demographic attributes**

1.1 Identifier

Please write down your IPICYT ID number: _____

1.2 Role

What is your role at the university?

□ Taught student

□ Research student

□ Researcher

□ Staff (Administration, maintenance, other employees of IPICYT) _____

1.3 Affiliation [Only for students and researchers]

Which department are you affiliated with?

□ Maths and Computer Science

□ Natural Sciences

□ Other: _____

1.4 Gender

Which gender do you identify yourself with?

□ Female

□ Male

□ Other: _____

□ Prefer not to say

1.5 Age

Please indicate your age in two digits: ____

1.6 Ethnicity

Which ethnic group do you identify most with?

□ White

□ Indigenous

□ Mestizo

□ Afrolatino

□ Other: _____

2. **Family, work, and socio-economics features**

2.1 (se) How many dependants do you have?

This could be children, children and partner, other relatives, etc.

□ Answer: _____

□ 0-1 [1pt.]

□ 2 [2pt.]

□ 3 [3pt.]

□ 4 [4pt.]

□ 5+ [5pt.]

2.2 (pr) How many people live in the same household as you?

This could be children, children and partner, siblings, other relatives, housemates etc.

□ Answer: _____

□ 0-1 [1pt.]

□ 2 [2pt.]

□ 3 [3pt.]

□ 4 [4pt.]

□ 5+ [5pt.]

2.3 (pr) How much of your time during a normal work day do you spend working on a computer?

□ 0-10% [5pt.]

□ 11-30% [4pt.]

□ 31-50% [3pt.]

□ 51-70% [2pt.]

□ 70-100% [1pt.]

2.4 (pr) How much of your time during a normal work day do you spend on communication with colleagues?

□ 0-10% [1pt.]

□ 11-30% [2pt.]

□ 31-50% [3pt.]

□ 51-70% [4pt.]

□ 70-100% [5pt.]

2.5 (pr) How much of your time during a normal work day do you spend working in a team?

□ 0-10% [1pt.]

□ 11-30% [2pt.]

□ 31-50% [3pt.]

□ 51-70% [4pt.]

□ 70-100% [5pt.]

2.6 Socio-economic class

(se) Look at the image of the ladder below. Imagine this ladder pictures how Mexican society is set up:

- At the top of the ladder are the people that are best off - they have the most money, the highest amount of schooling, and the jobs that bring the most respect.
- At the bottom are the people who are the worst off - they have the least money, little or no education, no job or jobs that no one wants or respects.

Now think of your family, please tell us where you think your family would be on this ladder: _____

□ 10-9 [1pt.]

□ 8-7 [2pt.]

□ 6-5 [3pt.]

□ 4-3 [4pt.]

□ 2-1 [5pt.]

2.7 (se) Perceived socio-economic status

People sometimes describe themselves as belonging to the working class, the middle class, or the upper or lower class. Would you describe yourself as belonging to the

□ Upper class [1pt.]

□ Upper middle class [2pt.]

□ Lower middle class [3pt.]

□ Working class [2pt.]

□ Lower class [1pt.]

□ Prefer not to answer [0pt.]

3. **Using digital media**

3.1 (pr) How much of your work time do you spend using the internet?

□ 0-10% [5pt.]

□ 11-30% [4pt.]

□ 31-50% [3pt.]

□ 51-70% [2pt.]

□ 70-100% [1pt.]

3.2 (pr) How much of your leisure time do you spend using the internet?

□ 0-10% [5pt.]

□ 11-30% [4pt.]

□ 31-50% [3pt.]

□ 51-70% [2pt.]

□ 70-100% [1pt.]

3.3 (pr) How do you usually access the internet from home? Please tick all appropriate boxes.

□ Through laptop + wifi [1pt.]

□ Through laptop + mobile connection [2pt.]

□ Through phone + wifi [3pt.]

□ Through phone + mobile connection [4pt.]

□ Other: [if none] [5pt.]

4. **Psychosocial features**

4.1 (psy) Sociability

Please write down the percentage of individuals (in your social circle) who would agree with the following statement about yourself: 'I spend a lot of time visiting friends' _____

□ 0-10% [5pt.]

□ 11-30% [4pt.]

□ 31-50% [3pt.]

□ 51-70% [2pt.]

□ 70-100% [1pt.]

4.2 Fear

Please rate the extent to which you experience the following feelings at this moment: Fear because of the COVID-19 disease/ the SARS COV-2 virus.

□ Not at all

□ Not really

□ Neutral

□ Somewhat

□ Very much

4.3 (psy) Perceived Stress Scale

1	**2**	**3**	**4**	**5**
Never	Almost never	Sometimes	Fairly often	Very Often

Based on the scale above, where zero indicates never experiencing that situation and four indicates experiencing that situation very often, please rate the following statements:

- In the last month, how often have you felt that you were unable to control the important things in your life?_____
- In the last month, how often have you felt confident about your ability to handle your personal problems?_____
- In the last month, how often have you felt that things were going your way?_____
- In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?_____

[Each score translates into the identical number of points. Then the average of the four sub-questions is computed.]

4.4 (psy) Subjective well-being

All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied: _____

□ 10-9 [1pt.]

□ 8-7 [2pt.]

□ 6-5 [3pt.]

□ 4-3 [4pt.]

□ 2-1 [5pt.]

4.5 Subjective well-being

Taking all things together on a scale of 1 to 10, how satisfied are you about IPICYT's efforts to keep you safe in the institute throughout the pandemic? _____

4. **Performance self-assessment**

5.1 (ppa) Self-assessment of performance

How would you rate your overall performance for your job or degree in the past 4 weeks?

□ Poor [1pt.]

□ Below average [2pt.]

□ Average [3pt.]

□ Above average [4pt.]

□ High [5pt.]

5.2 Self-assessment of learning

After the COVID-19 pandemic began, the way we learn and interact with our peers drastically changed. How would you say your learning experience has been in the past 4 weeks?

Please rate your learning process and experience between 1 and 10, where 1 is poor and 10 is excellent: _____

5.3 (ppa) Self-assessment of productivity

How would you rate your day-to-day productivity in your work in the past 4 weeks?

□ Poor [1pt.]

□ Below average [2pt.]

□ Average [3pt.]

□ Above average [4pt.]

□ High [5pt.]

5.4 (ppa) Self-assessment of achievement (supervisor goals)

Considering again the work for your job or degree during the past 4 weeks, please select the statement that fits your situation best.

□ I have struggled to achieve the goals set by my supervisor/employer/course teachers [1pt.]

□ I have managed to achieve some of the goals set by my supervisor/employer/course teachers [2pt.]

□ I have achieved many of the goals set by my supervisor/employer/course teachers [3pt.]

□ I have achieved most of the goals set by my supervisor/employer/course teachers [4pt.]

□ I have achieved all or exceeded the goals set by my supervisor/employer/course teachers [5pt.]

5.5 (ppa) Self-assessment of achievement (own goals)

Considering again the work for your job or degree during the past 4 weeks, please select the statement that fits your situation best.

□ I have struggled to achieve the goals I set for myself [1pt.]

□ I have managed to achieve some of the goals I set for myself [2pt.]

□ I have achieved many of the goals I set for myself [3pt.]

□ I have achieved most of the goals I set for myself [4pt.]

□ I have achieved all or exceeded the goals I set for myself [5pt.]

B Consent form

You are invited to take part in a research project conducted by researchers from the University of Oxford, Harvard University, and the United Nations University in conjunction with IPICYT. In concordance with international standards in the practice of randomised studies, this project has received ethical approval from the Research Ethics Committee at IPICYT and The Central University Research Ethics Committee at Oxford University.

We ask that you read this form carefully prior to deciding to participate in the study. If you decide you do not want to participate, you may leave at any time without providing a reason and without penalty.

Purpose: The purpose of this study is to understand how the implementation of an algorithmic-base safe education protocol influences students and staff well-being and productivity during a pandemic.

What happens during the study: This study requires you to follow one of two protocols: a group-testing protocol wherein you are required to participate in COVID-19 group-testing activities (provided by the LANBAMA laboratory at IPICYT) as instructed by the research team, or a remote-work protocol - otherwise known as the current institutional work policy during the pandemic. We also ask all participants to respond to a short survey at the beginning and at the end of the trial -within a month's time- where you will be asked sociodemographic questions, alongside a set of psychological questions. While you may feel uncomfortable with some of the questions asked, you do not need to answer questions that you do not want to answer. Furthermore, for the purpose of COVID-19 testing, you will be asked to give a saliva sample to the technicians at LANBAMA. The sample will be used directly on the day of reception and will be destroyed after being processed for a qPCR test. The sample(s) will not be stored.

Participation: The trial is expected to run for a month, throughout the month of June 2022, during which participants who are randomly selected to follow the C-SEF protocol will receive free COVID-19 testing. Throughout the course of the month the principal investigators will link health data (i.e. COVID-19 test results) to survey data (collected at the beginning and end of the trial). However, at the end of the trial all gathered data will be anonymised. If you wish to withdraw consent on the use of your data, please contact mail@c-sef.com. You always have the option of stopping your participation in the study and you may leave at any time without providing a reason and without penalty.

Potential risks: Participation in this study poses minimal risk. Participating in the protocol is likely to decrease the risk of COVID-19 contagion given the detailed level of monitoring, especially compared to the alternative, coming back to the institute without a safekeeping mechanism. However, while the probability of infection can be minimised and contained, it is not guaranteed to be zero. There is always a very small chance to get infected when participating in social activities, and COVID-19 comes with small and major consequences; among which, fever, cough, loss of taste and smell, respiratory problems and, in extreme cases, death.

Moreover, your survey responses are strictly confidential and will only be accessible to the researchers. Below, we describe the steps we are taking to protect your privacy. In addition, your decision on whether to participate will not adversely affect your relationship with IPICYT or any other institution to which the researchers are affiliated.

Benefits: Participating in this study means that you are aiding further development of science. Additionally, a successful trial would allow IPICYT to reformulate the institutional policy regarding work and study during the current and future waves of the pandemic into one that gives you more social interactions and flexibility with a minimised risk of contagion.

Data protection and privacy: The information collected during the study will be kept private. In concordance with the Data Protection Act of 1998 & 2018, the University of Oxford is the data controller. Responsible members of the University of Oxford and IPICYT may be given access to data for monitoring and or audit of the study to ensure we are complying with the guidelines or as otherwise required by law. Moreover, in concordance with the signed Memorandum of Understanding, the Instituto Potosino de Ciencia y Tecnología (IPICYT) will store and anonymise the original data in a secure server. No one other than members of IPICYT Supercomputing Centre and responsible members of the University of Oxford will have access to any records of this trial, and none of the records will identify you using any personal information. The data will be stored in electronic form, password protected. A copy of the anonymised data will be provided to the primary investigators listed at the top of this form. The data that we collect from you may be transferred to, and stored or processed at a destination outside Mexico. By submitting your personal data, you agree to this transfer, storing, or processing. After completion of the study, you cannot withdraw your personal information. Your individual privacy will be maintained in all publications or presentations resulting from this study. No information about you provided by you during this research will be disclosed to others without your written permission, except:

- if necessary to protect your rights or welfare (for example, if you are injured and need emergency care); or
- if required by law.

Additional information: If you are interested in receiving additional information about the results of the study, please contact the study authors.

Concerns: If you have any questions or concerns about any aspect of this project, you can contact the study authors at mail@c-sef.com, who will do their best to answer your query. The researcher(s) should acknowledge reception of your concern within 10 working days and give you an indication of how they intend to address it. If you fail to receive a response, are dissatisfied with the response you receive, or desire to report an aspect of how the study is being conducted, please contact the relevant Chair of Research Ethics Committee at the University of Oxford:

Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee;
Email: ethics@socsci.ox.ac.uk
Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

The Chair will seek to resolve the matter in a reasonably expeditious manner.

Please confirm the following by marking each of the boxes next to the statements.

Please Mark Each Box

- I confirm that I have read and understand the information for the above study and have had the opportunity to properly consider the information provided. ☐
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without any adverse consequences. ☐

- I understand the risks associated with participating in this study as explained in the information sheet. □
- I understand that a saliva sample will be taken during the study and that this sample will be tested for COVID-19. I understand that the sample will be destroyed after completion of this test or if I withdraw my consent. □
- I consider these samples a gift to the University of Oxford and the LANBAMA laboratory and I understand I will not gain any direct personal benefit from this. □
- I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford and IPICYT where it is relevant to my taking part in this study. I give permission for these individuals to access my data. I give permission for anonymised data to be made publicly available at the end of the research. □
- I understand that this project has been reviewed by, and received ethics clearance through, the Research Ethics Committee at IPICYT and the Central University Research Ethics Committee at Oxford University. □
- I understand who will have access to the personal data provided, how the data will be stored, and what will happen to the data at the end of the project. □
- I understand how this research will be written up and published. □
- I understand how to raise a concern or make a complaint. □
- I agree to take part in the study. □

By selecting "Yes, I agree to participate" below you are signifying that you have read and understood the above information and are agreeing to have the data that you provide during the course of the study to be processed accordingly.

□ Yes, I agree to participate

□ No, I do not agree to participate