Pre-analysis plan:
The Impact of Digital Mental Health Technology on Mental Health, Education, and Economic Outlook of Young Adults

Armando N. Meier[1+], Nathalie Monnet[2+], Clémence Kieny[2], Andrea Carolina Bello-Tocancipá[3], Laura Torres-Cuéllar[3], Naomi Koerner[4], Juan Vargas-Nieto[3], Gustavo Perdomo-Patino[5,6], Ricardo Araya[7], Omar Yesid Vargas Rodriguez[8], Fabio Idrobo[3,6,9], Martha Escobar Lux[3,5,6], Mauricio Avendano*[2,10]

[1]Faculty of Business and Economics, University of Basel, Switzerland

[2]Centre for Primary Care and Public Health (Unisanté), University of Lausanne, Switzerland

[3]Population Health Division, Fundación Santa Fe de Bogotá, Colombia

[4]Toronto Metropolitan University, Toronto, Canada

[5]Mental Health Department, Fundación Santa Fe de Bogotá, Colombia

[6]Faculty of Medicine, University of Los Andes, Bogota, Colombia

[7]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[8]Administrative Department of Social Prosperity, Colombia

[9]Departement of Psychological and Brain Sciences, Boston University, USA

[10]Department of Social and Behavioral Sciences, Harvard School of Public Health, USA

*Corresponding author: mauricio.avendano@unil.ch

+ These authors contributed equally

Motivation

We study the impact of digital mental health technology on mental health, educational, and economic outcomes of low-income students in Colombia from a large-scale randomized controlled trial. We are collaborating with the national government to embed the trial in the *Youth Income* (*Renta Joven*) scheme, a nationwide conditional cash-transfer program that provides regular educational subsidies to over 360,000 young people from low-income backgrounds. This unique government-academia partnership enables unparalleled access to all program participants and their administrative data on educational attainment, including enrolment and academic grades. Combining administrative and survey data will enable us to test the impacts of deploying mental health technologies alongside existing governmental programs.

Our aim is to investigate to what extent access to digital mental health technology deployed at scale (1) improve the mental health of low-income students incl. across different baseline levels of mental health, (2) improve educational performance and retention in higher education, (3) influence low-income students' decision-making closely linked to human capital investments and future orientation, namely self-control and patience incl. an incentivized measure of time preferences (4) and affect the labor market and educational (grades) aspirations of students. These research questions will be addressed by measuring key outcomes based on state-of-the-art scales from the psychology, behavioral science, and economics literature. Additionally, the trial design will enable us to explore a range of further important questions such as its cost-effectiveness and whether supplementing the digital intervention with peer support yields different effects compared to the digital intervention alone. This document presents the initial pre-analysis plan (PAP), additional projects will be detailed in separate PAPs.

Design

To address our research questions, all eligible *Renta Joven* participants will be invited to take a screening survey to assess consent, ability and means to participate. Next, we will conduct a baseline survey to assess their mental health symptoms (depression, anxiety, and PTSD), and baseline educational, decision-making, and labor market variables. Eligible participants will be randomly assigned to two main comparison groups: the treatment group, which will have access to digital mental health support through our platform; and the control group, which will receive general information about standard care. Finally, we assess mental health, education, economic aspirations, and future orientation at zero and three months after the end of the intervention. To capture longer-term effects, we will conduct a six-month follow-up, which will also serve to evaluate cost-effectiveness and implementation. A timeline is available in SI (figure 1).

We intend to additionally preregister long-run outcomes and corresponding analytical specifications in a separate pre-analysis plan, to be developed in time before a comprehensive additional match with administrative data in 2026, for the analysis of administrative data focusing on longer-run follow-up measures of educational, labor market, and health outcomes.

Digital Platform. The intervention consists of a digital, transdiagnostic, and adapted program, accessed by participants through an online platform. The intervention is specifically targeted to young beneficiaries of the *Renta Joven* program, and it was co-designed by the University Hospital Fundación Santa Fe de Bogotá (FSFB), the Colombian Government's Department of Social Prosperity (Prosperidad Social), the Centre for Primary Care and Public Health at the University of Lausanne (Switzerland), and staff and young participants in the *Renta Joven* program. To create the digital platform, we followed state-of-the-art approaches in transdiagnostic therapy and developed a dedicated platform together with a highly experienced IT developer. The platform provides access to different transdiagnostic tools

and activities based on Cognitive Behavioral Therapy to help young people overcome symptoms of depression, anxiety, and PTSD. It includes six modules: motivational interviewing, behavioral activation, problem-solving, interpersonal effectiveness, sleep strategies, and PTSD symptom management (including cognitive reappraisal and mindfulness strategies). Additionally, an emotional regulation toolbox is available for flexible use as needed. All modules and tools rely on and are inspired by components of protocols tested in comparable populations (Castro-Camacho et al. 2019; Bonilla-Escobar et al. 2018).

Sample. Participants must meet specific criteria to be included in our study, which we assess through an initial screening survey. Participants must provide informed consent, be 18 years or older, and be registered as college or vocational training students in the *Renta Joven* program. Participants are also required to have access to a computer, tablet, or smartphone with internet connectivity. They must also answer positively to a question about their willingness to use a digital platform for mental health and well-being. Additionally, to ensure response reliability, participants who complete the baseline survey too quickly—defined as those falling below four minutes—will be excluded, as this suggests limited engagement (pretesting by the team showed that the baseline survey takes around 15 to 25 minutes to complete at a reasonable pace). Moreover, we excluded the participants that also signed up to be a peer; participants who answered multiple times (we applied the following criteria in the corresponding order: first, we excluded anyone with at least one response classified as high-risk; second, we removed all incomplete responses; third, we retained only the first completed baseline survey),; and we excluded participants who could not be found in the *Renta Joven* database.

To be eligible, participants must report mild, moderate, or severe symptoms of depression, anxiety, and/or PTSD based on screening scales as defined below. Individuals who are asymptomatic will not be invited to participate in the trial but will receive referrals to mental health education resources and informational materials aimed at maintaining their mental health. These participants will also be provided with guidance on accessing resources and healthcare routes should they need attention in the future. Finally, participants reporting a high risk of suicide will also be excluded from the trial, because they may require specialized emergency treatment. These are classified as high risk if they answer positively to the question “For the past two weeks, have you thought about hurting yourself or taking your own life?”. Such individuals will be contacted by a project psychologist within 24 hours of completing the survey and offered psychological first aid, crisis management, and assistance in identifying and accessing local mental health services appropriate for their needs, which have been collated and approved by the Ministry of Health and Social Protection. The psychologist will track their progress and verify their access to the identified resources.

Mental health assessments. Participants in the trial will undergo an assessment of mental health symptoms with three internationally recognized standard scales, all of which have also been validated in the Colombian context. Higher scores of all scales indicate more severe symptoms. The first scale is the Patient Health Questionnaire-9 (PHQ-9), designed to measure the severity of **depressive symptoms** (Kroenke, Spitzer, and Williams 2001; Cassiani-Miranda et al. 2021). This scale comprises nine items, each rated on a scale from 0 ("not at all") to 3 ("nearly every day"), resulting in a total score ranging from 0 to 27. The second scale, the Generalized Anxiety Disorder-7 (GAD-7), evaluates symptoms of **anxiety** (Spitzer et al. 2006; Camargo et al. 2023). It includes seven items, also rated from 0 ("not at all") to 3 ("nearly every day"), yielding a total score between 0 and 21. The final instrument, the Primary Care Posttraumatic Stress Disorder (PTSD) for DSM-5 (PC-PTSD-5), is used to assess **symptoms of PTSD**. This scale consists of 5 yes/no items, with a maximum total score of 5 (Gómez-Restrepo et al. 2016; Prins et al. 2016).

Based on the results from the scales detailed above, participants will be classified into mild, moderate or severe levels of symptoms. We classify participants with symptoms in three groups of **overall severity** of symptoms across the three scales: **mild** (defined as PHQ-9 scores between 6 and 8, GAD-7

scores between 3 and 8, or PC-PTSD-5 score of 2), **moderate** (PHQ-9 scores between 9 and 14, GAD-7 scores between 9 and 14, or PC-PTSD-5 score of 3), and **severe** (PHQ-9 scores of 15 or higher, GAD-7 scores of 15 or higher, or PC-PTSD-5 scores of 4 or higher), based on a high severity score in any of the three scales. Participants falling below all the defined thresholds are categorized as asymptomatic and not enrolled in the trial.

Randomization. We will implement stratified randomization by overall severity level as defined above (mild, moderate, severe). Fifty percent of participants within each severity group will be randomly allocated to the control group (usual care) and fifty percent to the treatment group (digital platform), with comparisons between these groups used to test our primary hypotheses. To optimize resource allocation and enhance efficacy, the level of support provided will be tailored to the severity of participants' mental health symptoms. Participants with mild symptoms in the treatment group will access platform resources and a virtual forum for self-guided support. Those with moderate and severe symptoms in the treatment group will be further randomized into two equal subgroups: 50% to platform-only and 50% to platform-plus-peer support. The peer's role is exclusively to encourage participants to use the platform through a text-based message system within the platform. Peers do not engage in any kind of direct support or psychotherapy, and their contact with participants is anonymized, i.e., the identify of participants is not revealed to the peer. Each peer is supervised by a psychologist. To prioritize peer support based on severity, moderate cases are allowed up to two series of exchanges with their assigned peer, while those with severe symptoms have no limit on their exchanges with their peer. Our main analysis will compare the treatment group (comprising platform-only and platform-plus-peer) with the control group across all severity groups to maximize statistical power.

Surveys. We will measure participants' mental health, education, labor market and future orientation outcomes as well as beliefs about educational attainment and labor market outcomes both in baseline and follow-up assessments. Below we list the most important raw measures across dimensions (some of those will be combined for hypotheses tests as outlined below).

<u>Mental health measures</u>

- **Depression:** Patient Health Questionnaire-9 (PHQ-9) (Kroenke, Spitzer, and Williams 2001; Cassiani-Miranda et al. 2021).
- **Anxiety**: Generalized Anxiety Disorder-7 (GAD-7) (Spitzer et al. 2006; Camargo et al. 2023).
- **PTSD:** Primary Care Posttraumatic Stress Disorder (PTSD) for DSM-5 (PC-PTSD-5) (Gómez-Restrepo et al. 2016; Prins et al. 2016).
- **Life satisfaction:** On a scale of 0 to 10, where 0 means you are not at all satisfied and 10 means you are completely satisfied, how satisfied are you with your life in general these days?
- **Sleep quality:** How satisfied or dissatisfied do you feel with your sleep habits? To what extent do you think your sleep problem affects your daily life?
- **Mental health stigma**: The 3-item Self-Stigma of Seeking Help (SSOSH) Scale uses a Likert-type scale to measure the perception that seeking help from a psychologist or other mental health professional threatens one's self-regard, satisfaction with oneself, self-confidence, and overall worth as a person (Vogel, Wade, and Haake 2006; Larrahondo et al. 2021; Brenner et al. 2021).
- **Behavioral Activation for Depression Scale**, 9-item version: assesses how much individuals engage in activities that align with their values and goals, as well as how much they avoid activities due to depression-related feelings like sadness or lack of motivation (Kanter et al. 2007; García Lis 2019).

<u>Education</u>

- **Academic enrolment continuity**: the question is "Are you still enrolled in the same academic program as in the last survey?". We further ask "Why are you no longer enrolled in the

academic program?" with the options: I graduated from the program, I temporarily interrupted my studies, I dropped out, another reason asking for which one.

- **Grade Point Average**: We ask questions about current and expected GPA in the survey, as well as their cumulated GPA. For university students, we also measure cumulated GPA based on administrative data.
- **Efforts**: how many hours do you currently spend studying during a typical day? How frequently did the following occur? (1) I skipped a whole week of classes, (2) I skipped some classes, (3) I arrived late for classes. On a 5-point Likert scale from never to always.
- **Courses**: how many courses or modules are you currently taking? How many are you expected to take? Were you failing any? How many were you failing?
- **Work**: in the last month, what was the main activity you spend most of your time? Last month, did you engage in any paid activity? In that activity, what was your role? Did your employment contract include social benefits? Last month, on average, how many hours per week did you dedicate to this job? (if graduated) what was your labor income in the past month?

Future orientation

- **Patience** is assessed through a 10-point Likert scale asking: "How willing are you to give up something beneficial to you now in order to obtain greater benefits in the future?".
- **Time-Discounting** is measured through a series of two hypothetical binary choices between immediate and delayed financial rewards (Falk et al. 2018). In our context, participants are asked, "Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you two situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e., future prices are the same as today's prices. Please consider the following: (1) Would you prefer 28,000 Pesos today or 43,100 Pesos in 12 months?". If the answer is "today", we further ask (2) "Would you prefer 28,000 Pesos today or 56,400 Pesos in 12 months?". These questions are incentivized in the second follow-up. At follow-up 2, the measure will be incentivized with the following phrasing: "You are given the choice between receiving a payment today or a payment in 12 months. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. The computer will randomly select one of the two choices of several individuals participating in this survey and implement it. If you are chosen, you will actually be contacted soon after the survey or in 12 months, the amount will be paid out as a food voucher according to your choice. For each scenario, please indicate your choice: (1) Would you prefer 28,000 Pesos today or 43,100 Pesos in 12 months?". If the answer is "today", we further ask (2) "Would you prefer 28,000 Pesos today or 56,400 Pesos in 12 months?".
- **Brief Self-Control Scale**: The BSCS is a short 13-item self-report format measure of the general self-control construct (Tangney, Baumeister, and Boone 2004; Campo-Arias, Oviedo, and Herazo 2014). The BSCS total score is obtained, after the appropriate items have been reversed, by adding the answer given to each item on a 5-point ordinal format (from 1=Totally false, to 5=Totally true). Higher scores indicate higher levels of self-control. It ranges between 13 and 65.

Educational and labor market expectations

- **Education outlook**: In addition to asking expected GPA and cumulated GPA, we ask the following questions: What are your plans for the next academic period? On a scale from 0 to 100, how confident are you that you will continue studying in the next academic period? How likely is it that you will complete the program in which you are enrolled? If you could study as much as you wanted, what would be the highest level of education you would like to achieve?

- **Employment outlook**: Six months after graduation, how much do you think you could earn per month? On a scale of 0 to 100, how confidence are you that you will find a job within 6 months of graduation? In the next 10 years, what job would you like to be doing (your dream job)? On a scale of 0 to 100, how confident are you that you will achieve your preferred/dream job?

Administrative data. Since our randomized controlled trial (RCT) is integrated into a governmental program, we will have access to administrative data for participants who have consented to data linkage. In a pilot study, over 98% of participants who agreed to join the study have consented to link their data with administrative records. The national ID number, collected across all surveys, will be used as the primary identifier to merge survey responses with administrative data. This access will also enable us to examine educational outcomes, including cumulated GPA among university students, which universities automatically report to the government.

Data collection and Sample size

Data collection. The online surveys are collected through REDCap. The intervention will take place in the Spring 2025. We aim to collect data for 11,250 eligible participants (baseline survey). We expect attrition rates of 40% and 50%, at the 0-, and 3- month follow-ups, respectively, retaining approximately 6,750 participants immediately after intervention ends, 5,250 at 3 months. To minimize attrition, both treatment and control group participants will receive weekly reminders to complete follow-up assessments. They will be given 6 weeks for each follow-up survey.

To motivate participation at the start of the trial, prior to sending individualized emails and WhatsApp messages, we will use social media, calls, and mental health awareness events to encourage participation. To further encourage participation, we will offer incentives through raffles based on participants completing all follow-up surveys to encourage survey completion (up to USD 50). For completing each follow-up, participants will also receive a fixed small payment of around COP 5,000 or USD 1.2, corresponding roughly to a minimum wage for 30 minutes of work. Finally, we will follow-up unresponsive participants via phone (we will send reminders for survey completion and participation based on phone numbers collected in the screening survey via WhatsApp and follow-up with phone calls).

Note that we will have virtually no attrition for educational outcomes among university students, as we measure those based on linked administrative data for all trial participants.

Statistical power. We will invite all eligible *Renta Joven* participants (>360,000), ensuring a sufficiently large sample size to detect even small effects with high statistical power and to do highly powered equivalence testing comparing our effect sizes to previous literature. The required sample size was calculated using the following assumptions: an equal share of participants in control and treatment groups, 80% power, and a two-tailed t-test with an alpha level of 0.05 for each main hypothesis.

For effect size estimates, where previous effect sizes were available, we rely on Lund et al. (2024) looking at impacts of mental health interventions on adult outcomes (Lund et al. 2024). We use their meta-analytically identified average effects of interventions for two outcomes: mental health (0.23 SD) and subjective poverty measures, as the latter is similar to our measure of labor market expectations and is also based on survey data (0.16 SD).

When previous research firmly established a plausible range of effect sizes, researchers can rely on these previously estimated effect sizes for power analysis. Given the absence of field evidence

in LMICs on future orientation (self-control and patience) as well as educational attainment, however, it is not possible to assess power based on previous effect sizes. There are also no precise theoretical predictions on the size of the effects (there are only predictions about the sign of the effects). To compute power for future orientation and educational outcomes, we therefore first rely on the standard, conservative definition of a small absolute effect size of 0.2 standard deviations based on Cohen's d (Altay et al. 2022). Note that this is more conservative than estimates from related literature in high-income countries: meta-analytic estimate based on 6 trials including a total of 2,428 participants report positive effects of digital mental health interventions on academic performance of, on average, 0.26 SD, which is not statistically significant (Bolinski et al. 2020).

To detect the effect sizes in Lund et al. (2024) for mental health and labor market expectations and an effect size of 0.2 SD for educational outcomes and future orientation, the required sample size ranges between 296 and 613 participants (combined treatment and control groups). These values are all well below our expected sample sizes of 6,750 participants immediately post-intervention, and 11,250 participants when using administrative data (Lund et al. 2024).

When focusing on detecting the smallest effect size of interest (SESOI), which we derived from the lower bound of the confidence interval in Lund et al. (2024) for mental health and labor market expectations (0.13 SD and 0.05 SD, respectively), the required sample sizes increase to between 927 and 6,272 participants, which might exceed our expected sample size. Nevertheless, as the average effects reported by Lund et al. (2024) and in the broader literature (Bolinski et al. 2020; Lund et al. 2024; Conley, Durlak, and Kirsch 2015; Kim et al. 2023) are at least twice as large as the SESOI, we are very well-powered to detect effect sizes of around 0.1–0.15 SD which are generally regarded as small and are smaller than the 0.2 SD effect sizes usually taken as the reference effect size for a small effect. With effect sizes of 0.10 to 0.15, we estimate needing between 696 and 1,568 participants for null hypothesis testing across all outcomes, including future orientation and educational attainment where we do not have reliable comparison effect sizes based on previous evidence. We will use administrative data for the educational outcomes, which demand the largest sample sizes, providing us with the power to also detect very small impacts on educational outcomes.

Reassuringly, our expected sample size at follow-up points allows us to detect minimum effect sizes of down to 0.05 SD, which is very small. Our strong statistical power allows for both null hypothesis and equivalence testing, enabling us to compare our results with previous average effect sizes even for small differences. In addition to testing our primary hypotheses, we will report how our effects diverge from established findings in the literature.

Analysis

We will assess the impact of the intervention on mental health, education, future orientation and educational as well as employment outlook using a combination of survey and administrative data. The national ID number, collected in all surveys, will be used as the key identifier for merging the different surveys and the administrative data. Administrative data will also allow us to examine educational outcomes, such as cumulative GPA, which is automatically reported by universities to the government.

The initial evaluation will focus on two primary outcomes: mental health symptoms and education, with decision-making and labor market outlook as secondary outcomes. Importantly, these outcomes will be measured at baseline and in follow-up surveys. For the analysis, we will use survey data from the first and second follow-up to measure mental health, future orientation, and labor market expectations. We will only analyze the data once the follow-up period for survey measuring the respective outcome of has expired. For educational outcomes, we will rely on the earliest administrative data available. Specifically, we will use the earliest available dataset to include the

Spring term 2025 grades. Details of the main initial research questions, outcome measures and expected effects are provided in Table 1. As mentioned earlier, we intend to examine long-run effects of the intervention. We may also preregister additional analyses for data not yet collected or for better understanding heterogeneity in treatment effects across the mental health spectrum.

For the initial evaluation, we focus on the first measurement points of all outcomes with a similar distance to the intervention as previous trials testing impacts on mental health. There are three main reasons: First, we want to ensure minimal influence of attrition on survey outcomes. Second, we want to ensure comparability of our mental health estimates with previous smaller scale trials and also allow for tests against previous trial effects. Third, it seems reasonable to expect the mental health improvements to affect grades and enrolment immediately when students have access to the digital mental health technology. This is the primary effect we initially aim to capture. Finally, to allow for timely distribution of our findings given that many governments in LMICs are currently considering rolling out initiatives to improve mental health.

Primary Outcomes. Initially, we will use as our main outcome of mental health the PHQ-ADS index (Kroenke et al. 2016), a composite measure of depression and anxiety derived from the summed scores of the PHQ-9 and GAD-7 scales. PHQ-ADS scores range from 0 to 48, with higher scores reflecting greater severity of symptoms. To ensure comparability of effect sizes, we will standardize this measure. The primary analysis of the intervention's impact on mental health will be based on PHQ-ADS scores collected during the first follow-up survey. We will also separately explore effects on standardized scores for depression (PHQ-9), anxiety (GAD-7), and PTSD (PC-PTSD-5), as well as changes in a binary variable indicating severe symptoms of depression, anxiety, and PTSD.

For education, we will evaluate two main outcomes: academic enrolment continuity and academic success. We measure enrolment continuity by whether individuals have dropped out by the second follow-up survey, conducted 6 months after intervention start. We define dropout as 1 if participants interrupted their program or dropped out, otherwise we define it as 0. We assess academic success of university students (ca. 70% of the sample) by their accumulated grade point average (GPA) between the beginning of their studies and the end of the Spring term 2025, based on administrative data.

Secondary Outcomes. As our secondary outcomes, we will focus on future orientation and labor market as well as educational attainment expectations. Future orientation will be based on a combination of two key economic and psychological constructs measuring present orientation: (1) time preferences—measured by self-reported patience and a quantitative time discounting measure and (2) the Brief Self-Control Scale (BSCS). Time preference measures will be the standardized average of the patience and time discounting standardized measures following the validated scale (Falk et al. 2018). Future orientation will be constructed by averaging the time preference measure and the BSCS measure and then standardizing. We will use the measurements from the first follow-up.

Employment expectations which will be assessed through participants' aspirations for future earnings with an outlook of 6 months post-graduation measured in the first follow-up survey. We will standardize and truncate income at the 95^{th} percentile to avoid outliers affecting our estimates. We also measure the probability of having a job within 6-months of graduation. Finally, we also measure people's expectations about their cumulative grades and continued enrolment. When we have the corresponding administrative data available (likely in 2026), we will compare these expectations to realized outcomes and describe these analyses in a separate pre-analysis plan.

Additional Outcomes. We will provide additional analyses on outcomes that are believed to be on the causal pathways between a mental health intervention and our main outcomes, namely mental health and education. We will assess educational investment, a behavioral marker of long-term orientation,

by measuring how often people have missed, skipped, or arrived late to class based on the first follow-up, and the number of hours spent studying. We will also explore whether participants in the treatment group will be less likely to experience severe symptoms of anxiety and depression (an indicator of PHQ-ADS >=30 measured in the first follow-up survey) and how one of the key targets of the intervention, behavioral activation (measured in follow-up 1 by the short version of the Behavioral Activation for Depression scale), changes. Using causal machine learning, we will carefully study how mental health effects and platform use vary across the population to understand the potential for targeting of digital mental health interventions.

Analysis. We will compare outcomes between participants who were randomly allocated to the group offered a digital mental health intervention (the treatment group) with the group of participants who were not offered the digital intervention (the control group). We will use an intention-to-treat analysis including all participants that are randomized with complete follow-up data. The main analysis compares the outcome across groups using linear regression with robust standard errors and *p*-values from two-sided tests. Our model specification is as follows:

$$Y_{i1} = \beta_1 Treatment_{i0} + \beta_2 Severity_{i0} + \beta_3 phq9_{i0} + \beta_4 gad7_{i0} + \beta_5 ptsd_{i0} + _{i1} + \beta_6 Y_{i0} + \mathbf{X'_{i0}}\alpha + region_{i0} + wkbl_{i0} + sena_{i0} + \varepsilon_{i1}$$

Where Y_{i1} indicates the outcome from the follow-up. Across all specifications, we control for the baseline measure of severity levels ($Severity_{i0}$) which are our strata fixed effects. We also include baseline continuous measures of mental health ($phq9_{i0}$, $gad7_{i0}$, and $ptsd_{i0}$, all standardized based on the baseline measurements). Moreover, we control for the baseline measurement of the outcome, Y_{i0}, when available. X'_{i0} includes a vector of controls, namely: sex, being as a victim of the armed conflict, age, expecting a child, and year of program start. In addition, we include indicators for region ($region_{i0}$), week of baseline completion ($wkbl_{i0}$) and a technical program ($sena_{i0}$, as opposed to university programs). These account for unobserved differences in regional healthcare policies or mental health issues, timing differences in mental health assessments and type of studies. All indicators are included as dummy variables. Standard errors will be heteroscedasticity robust at the individual level. For the analysis, we will use the *reghdfe* package in Stata 18 to absorb the main fixed effects.

The treatment effect will be measured by β_1 where we use two-sided hypothesis testing. Apart from heteroskedasticity robust standard errors, we will also report heteroskedasticity robust standard errors adjusted for multiple hypotheses testing accounting for correlations in outcome variables and adjusting separately considering only our primary or only our secondary outcomes (using the main regression specification).

Additional analyses. We will do several additional analyses offering complementary information and sensitivity assessments for our primary outcomes.

In addition to the intent-to-treat estimates, we will estimate the treatment effect of taking-up the intervention using instrumental variable analysis. We report instrumental variable estimates with group assignment as the instrument, actual participation as the endogenous variable, and symptoms of depression and anxiety or educational performance variables as the outcomes. The estimates show the treatment effects among people who took up the intervention under the assumption that outcomes were only affected by taking up the intervention.

In a different analysis, we will explore the link between changes in mental health and changes in educational performance induced by assignment to the treatment group, using an instrumental variable approach. Symptoms of depression and anxiety will be the endogenous variable and educational performance the outcome. This estimate should be interpreted carefully, as it only has a causal interpretation if the effect of treatment assignment on educational performance operates exclusively through changes in mental health.

Using causal forests, we will carefully study how mental health effects and platform use vary across the population to understand the potential for targeting of digital mental health interventions and whether people who might benefit most based on the control group do benefit most of being assigned to treatment.

Through the platform, routine data are collected on several indicators of platform use and engagement, such as number of logins over the intervention period, time spent using the platform, contacts with the peers for those in the peer support group, and number of modules completed. These data will also be used in order to assess treatment non-compliance, and to obtain estimates of the impact of treatment assignment on mental health and educational outcomes using instrumental variable (IV) estimation to adjust for the proportion of non-compliant participants (Ye et al. 2014).

Separately, we will further explore questions around non-compliance and how it affected treatment effects. We will register these analyses separately under the trial registration. First, we will also routine data collected by the platform in descriptive analysis that aim to identify individual predictors of treatment compliance, and the impact of the peer support group on treatment compliance. Second, in follow-up 3, we will apply a scale that assesses treatment group participants' views on and experiences with the platform. These exploratory analyses aim to assess the impact of the peer support on engagement, and to identify individual characteristics that predict engagement with the platform.

Table 1. Synthesis of the research questions and outcomes measurement.

Question	Hypothesis	Outcome	Interpretation
Primary outcomes			
To what extent does access to digital technology improve mental health symptoms of low-income university and vocational training students?	Compared to participants in the control group, participants who will access the digital mental health intervention will show less symptoms of depression and anxiety.	We will use the PHQ-ADS index, summing the scores from the PHQ-9 and GAD-7 scales, at follow-up 1 (immediately after the intervention).	A negative and statistically significant ($p<0.05$) coefficient for the PHQ-ADS index will show a beneficial effect of the treatment on mental health symptoms compared to individuals who did not have access to digital mental health technology.

To what extent does access to digital mental health technology improve educational performance and retention in higher education?	Compared to participants in the control group, participants who will access the digital mental health intervention will have a higher average cumulative GPA.	We will use cumulative GPA from the administrative data available in the Fall 2025 (standardized).	A positive and statistically significant ($p<0.05$) coefficient for GPA will show a beneficial effect of the treatment compared to individuals who did not have access to the digital mental health platform.
	Compared to participants in the control group, participants who will access the digital mental health intervention will have higher retention rates in their study program.	To assess retention at follow-up 2, participants are asked, "Are you still enrolled in the same academic or vocational program as you were in the last survey?". For those no longer enrolled, a follow-up multiple-choice question inquires, "Why are you no longer enrolled?" with response options including: "I changed program", "I graduated from the program," "I temporarily interrupted my program," and "I dropped out.". We define dropout as 1 if participants interrupted their program or dropped out, otherwise we define it as 0.	A negative and statistically significant ($p<0.05$) coefficient for the enrolment continuity dummy will show a beneficial effect of the treatment compared to individuals who did not have access to the digital mental health platform.
Secondary outcomes			
What is the impact of a digital mental health intervention on future orientation, closely linked to human capital formation?	Compared to participants in the control group, participants who will access the digital mental health intervention will show more future orientation.	The future orientation measure will combine two outcomes measured at follow-up 1: patience (self-reported patience and a quantitative time discounting measure) and the Brief Self-Control Scale (BSCS). - Patience is assessed through a 10-point Likert scale asking, "How willing are you to give up something beneficial to you now in order to obtain greater benefits in the future?". The measure will be standardized. - Time-Discounting is measured through	A positive and statistically significant ($p<0.05$) coefficient for future orientation will show a beneficial effect of the treatment compared to individuals

		a series of two hypothetical binary choices between immediate and delayed financial rewards. In our context, participants are asked, "Suppose you were given the choice between receiving a payment today or a payment in 12 months. The payment today is the same in each situation, but the payment in 12 months varies. For each scenario, please indicate your choice, assuming no inflation: (1) Would you prefer 28,000 Pesos today or 43,100 Pesos in 12 months?"; and (2) "Would you prefer 28,000 Pesos today or 56,400 Pesos in 12 months?" - The BSCS is a short 13-item self-report format measure of the general self-control construct. The BSCS total score is obtained, after the appropriate items have been reversed, by adding the answer given to each item on a 5-point ordinal format (from 1=Totally false, to 5=Totally true). Higher scores indicate higher levels of self-control. It ranges between 13 and 65 and will be standardized. Future orientation will be constructed by averaging a combined index of standardized time-discounting and patience with the BSCS measure and then again standardizing.	who did not have access to the digital mental health platform.
Does a digital mental health intervention improve labor market and educational expectations?	Compared to participants in the control group, participants who will access the digital mental health intervention will show more optimistic labor market and educational expectations.	For labor market, the question, measured at follow-up 1, evaluates labor market expectations with, "Six months after graduation, how much do you think you could earn per month?". We will standardize and truncate the measure at the 95th percentile. For educational, we ask expected cumulative GPA (What do you think your CUMULATIVE grade point average will be at the end of the current semester?). The measure will be winsorized (at the 5th percentile) and standardized.	A positive and statistically significant ($p<0.05$) coefficient for expected earnings and cumulative GPA will show a positive effect of the treatment compared to individuals who did not have access to the digital mental health platform.

Additional outcomes			
To what extent does access to digital technology impact severe symptoms of anxiety and depression?	Compared to participants in the control group, participants who will access the digital mental health intervention will be less likely to show severe symptoms of depression and anxiety.	We will use the PHQ-ADS index, summing the scores from the PHQ-9 and GAD-7 scales, measured at follow-up 1 (immediately after the intervention). We define severe symptoms as 1 if individuals report a score of >= 30 and 0 otherwise.	A negative and statistically significant ($p<0.05$) coefficient for the severe symptoms dummy will show a beneficial effect of the treatment on severe depression and anxiety symptoms compared to individuals who did not have access to digital mental health technology.
To what extent does access to digital technology increase behavioral activation?	Compared to participants in the control group, participants who will access the digital mental health intervention will have higher scores of behavioral activation scale.	The BADS is a short 9-item self-report format measure of behavioral activation, measured at follow-up 1 (immediately after the intervention). The BADS total score is obtained, after the appropriate items have been reversed, by adding the answer given to each item on a 7-point Likert scale from “Not at all” to “Extremely”. Higher scores indicate higher levels of activation level and lower levels of rumination. It ranges between 0 and 54 and will be standardized.	A positive and statistically significant ($p<0.05$) coefficient for the BADS scale will show a beneficial effect of the treatment on behavioral activation compared to individuals who did not have access to digital mental health technology.

What is the impact of a digital mental health intervention on educational investment?	Compared to participants in the control group, participants who will access the digital mental health intervention will show more regular class attendance.	Educational investment will be composed of three items reported in response to the following question, measured at follow-up 1 (immediately after the intervention): "In the last academic period, how frequently did the following things occur? 1) I missed classes for a full week, 2) I skipped some classes, 3) I arrived late for classes. Respondents answer on a 4-point scale from never to always to each item (# 160). We average the responses across the three items and standardize the result.	A positive and statistically significant ($p<0.05$) coefficient for class attendance will show a beneficial effect of the treatment compared to individuals who did not have access to the digital mental health platform.

References

Altay, Sacha, Marlène Schwartz, Anne-Sophie Hacquin, Aurélien Allard, Stefaan Blancke, and Hugo Mercier. 2022. "Scaling up Interactive Argumentation by Providing Counterarguments with a Chatbot." *Nature Human Behaviour* 6 (4): 579–92. https://doi.org/10.1038/s41562-021-01271-w.

Bolinski, F., N. Boumparis, A. Kleiboer, P. Cuijpers, D. D. Ebert, and H. Riper. 2020. "The Effect of E-Mental Health Interventions on Academic Performance in University and College Students: A Meta-Analysis of Randomized Controlled Trials." *Internet Interventions* 20 (April):100321. https://doi.org/10.1016/j.invent.2020.100321.

Bonilla-Escobar, Francisco J., Andrés Fandiño-Losada, Diana M. Martínez-Buitrago, Julián Santaella-Tenorio, Daniel Tobón-García, Edgar J. Muñoz-Morales, Ivan D. Escobar-Roldán, et al. 2018. "A Randomized Controlled Trial of a Transdiagnostic Cognitive-Behavioral Intervention for Afro-Descendants' Survivors of Systemic Violence in Colombia." *PloS One* 13 (12): e0208483. https://doi.org/10.1371/journal.pone.0208483.

Brenner, Rachel E., Kimberly F. Colvin, Joseph H. Hammer, and David L. Vogel. 2021. "Using Item Response Theory to Develop Revised (SSOSH-7) and Ultra-Brief (SSOSH-3) Self-Stigma of Seeking Help Scales." *Assessment* 28 (5): 1488–99. https://doi.org/10.1177/1073191120958496.

Camargo, Loida, Jorge Herrera-Pino, Salomon Shelach, Marcio Soto-Añari, Maria F. Porto, Monica Alonso, Miguel González, et al. 2023. "GAD-7 Generalised Anxiety Disorder Scale in Colombian Medical Professionals during the COVID-19 Pandemic: Construct Validity and Reliability." *Revista Colombiana de Psiquiatría (English Ed.)* 52 (3): 245–50. https://doi.org/10.1016/j.rcpeng.2021.06.011.

Campo-Arias, Adalberto, Heidi Celina Oviedo, and Edwin Herazo. 2014. "Estigma: barrera de acceso a servicios en salud mental." *Revista Colombiana de Psiquiatría* 43 (3): 162–67. https://doi.org/10.1016/j.rcp.2014.07.001.

Cassiani-Miranda, Carlos Arturo, Angy Karina Cuadros-Cruz, Harold Torres-Pinzón, Orlando Scoppetta, Jhon Henrry Pinzón-Tarrazona, Wendy Yulieth López-Fuentes, Andrea Paez, Diego Fernando Cabanzo-Arenas, Sergio Ribero-Marulanda, and Elkin René Llanes-Amaya. 2021. "Validity of the Patient Health Questionnaire-9 (PHQ-9) for Depression Screening in Adult

Primary Care Users in Bucaramanga, Colombia." *Revista Colombiana De Psiquiatria (English Ed.)* 50 (1): 11–21. https://doi.org/10.1016/j.rcp.2019.09.001.

Castro-Camacho, Leonidas, Michel Rattner, Diana Melisa Quant, Laura González, Julián David Moreno, and Amantia Ametaj. 2019. "A Contextual Adaptation of the Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders in Victims of the Armed Conflict in Colombia." *Cognitive and Behavioral Practice* 26 (2): 351–65. https://doi.org/10.1016/j.cbpra.2018.08.002.

Conley, Colleen S., Joseph A. Durlak, and Alexandra C. Kirsch. 2015. "A Meta-Analysis of Universal Mental Health Prevention Programs for Higher Education Students." *Prevention Science* 16 (4): 487–507. https://doi.org/10.1007/s11121-015-0543-1.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–92. https://doi.org/10.1093/qje/qjy013.

García Lis, Derly Rocío. 2019. "Propiedades psicométricas de la escala Bads Short-Form (Bads-Sf) en población Colombiana." https://repositorio.konradlorenz.edu.co/handle/001/2187.

Gómez-Restrepo, Carlos, Cecilia De Santacruz, María Nelcy Rodriguez, Viviana Rodriguez, Nathalie Tamayo Martínez, Diana Matallana, and Lina M. Gonzalez. 2016. "Encuesta Nacional de Salud Mental Colombia 2015. Protocolo del estudio." *Revista Colombiana de Psiquiatría* 45 (December):2–8. https://doi.org/10.1016/j.rcp.2016.04.007.

Kanter, Jonathan W., Patrick S. Mulick, Andrew M. Busch, Kristoffer S. Berlin, and Christopher R. Martell. 2007. "The Behavioral Activation for Depression Scale (BADS): Psychometric Properties and Factor Structure." *Journal of Psychopathology and Behavioral Assessment* 29 (3): 191–202. https://doi.org/10.1007/s10862-006-9038-5.

Kim, Jiyeong, Lois M. D. Aryee, Heejung Bang, Steffi Prajogo, Yong K. Choi, Jeffrey S. Hoch, and Elizabeth L. Prado. 2023. "Effectiveness of Digital Mental Health Tools to Reduce Depressive and Anxiety Symptoms in Low- and Middle-Income Countries: Systematic Review and Meta-Analysis." *JMIR Mental Health* 10 (March):e43066. https://doi.org/10.2196/43066.

Kroenke, Kurt, Robert L. Spitzer, and Janet B. W. Williams. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine* 16 (9): 606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kroenke, Kurt, Jingwei Wu, Zhangsheng Yu, Matthew J. Bair, Jacob Kean, Timothy Stump, and Patrick O. Monahan. 2016. "Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials." *Psychosomatic Medicine* 78 (6): 716. https://doi.org/10.1097/PSY.0000000000000322.

Larrahondo, Brayan F., Jenny García Valencia, Andrés Mauricio Rangel Martínez-Villalba, Juan Pablo Zapata Ospina, and Daniel C. Aguirre-Acevedo. 2021. "Validation of the Self Stigma of Seeking Help (SSOSH) Scale in a Population of Colombian Medical Students." *Revista Colombiana De Psiquiatria (English Ed.)* 50 (2): 82–91. https://doi.org/10.1016/j.rcpeng.2020.01.002.

Lund, Crick, Kate Orkin, Marc Witte, John H. Walker, Thandi Davies, Johannes Haushofer, Sarah Murray, et al. 2024. "The Effects of Mental Health Interventions on Labor Market Outcomes in Low- and Middle-Income Countries." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w32423.

Prins, Annabel, Michelle J. Bovin, Derek J. Smolenski, Brian P. Marx, Rachel Kimerling, Michael A. Jenkins-Guarnieri, Danny G. Kaloupek, et al. 2016. "The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample." *Journal of General Internal Medicine* 31 (10): 1206–11. https://doi.org/10.1007/s11606-016-3703-5.

Spitzer, Robert L., Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. 2006. "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7." *Archives of Internal Medicine* 166 (10): 1092. https://doi.org/10.1001/archinte.166.10.1092.

Tangney, June P., Roy F. Baumeister, and Angie Luzio Boone. 2004. “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success.” *Journal of Personality* 72 (2): 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x.

Vogel, David L., Nathaniel G. Wade, and Shawn Haake. 2006. “Self-Stigma of Seeking Help Scale.” https://doi.org/10.1037/t00524-000.

Ye, Chenglin, Joseph Beyene, Gina Browne, and Lehana Thabane. 2014. “Estimating Treatment Effects in Randomised Controlled Trials with Non-Compliance: A Simulation Study,” June. https://doi.org/10.1136/bmjopen-2014-005362.

Supplementary Information

Month	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Timeline study
Jan 2025	6	7	8	9	10	11	12	
	13	14	15	16	17	18	19	
	20	21	22	23	24	25	26	
	27	28	29	30	31	1	2	
Feb 2025	3	4	5	6	7	8	9	Screening survey: 03.02.2025 - 02.03.2025 (4 weeks) Baseline survey: 1 day after screening completion 04.02.2025 - 09.03.2025 (5 weeks)
	10	11	12	13	14	15	16	Reminder 1 screening survey Deadline Bti
	17	18	19	20	21	22	23	Reminder 2 screening survey
	24	25	26	27	28	1	2	Reminder 3 screening survey Screening survey closes (02.03.2025)
Mar 2025	3	4	5	6	7	8	9	Baseline survey closes (09.03.2025)
	10	11	12	13	14	15	16	Randomization (11-12.03.2025) BTI Randomization platform setup (13-14.03.2025)
	17	18	19	20	21	22	23	
	24	25	26	27	28	29	30	Treatment week 1 (start of treatment)-llamadas a un grupo
	31	1	2	3	4	5	6	Treatment week 2-recordatorio semanal
Apr 2025	7	8	9	10	11	12	13	Treatment week 3-recordatorio semanal
	14	15	16	17	18	19	20	Treatment week 4-recordatorio semanal
	21	22	23	24	25	26	27	Treatment week 5-recordatorio semanal
	28	29	30	1	2	3	4	Treatment week 6-recordatorio semanal
May 2025	5	6	7	8	9	10	11	Treatment week 7-recordatorio semanal
	12	13	14	15	16	17	18	Treatment week 8-recordatorio semanal
	19	20	21	22	23	24	25	Treatment week 9
	26	27	28	29	30	31	1	Treatment week 10 (end of treatment)
Jun 2025	2	3	4	5	6	7	8	Treatment week (extra)
	9	10	11	12	13	14	14	First follow-up sendout (3 months)
	16	17	18	19	20	21	21	Reminder 1
	23	24	25	26	27	28	28	Reminder 2
	30	1	2	3	4	5	6	Reminder 3
Jul 2025	7	8	9	10	11	12	12	Reminder 4
	14	15	16	17	18	19	19	Reminder 5
	21	22	23	24	25	26	27	
	28	29	30	31	1	2	3	
Aug 2025	4	5	6	7	8	9	10	
	11	12	13	14	15	16	17	
	18	19	20	21	22	23	24	
	25	26	27	28	29	30	31	
	1	2	3	4	5	6	7	Second follow-up (6 months)
Sep 2025	8	9	10	11	12	13	14	Reminder 1
	15	16	17	18	19	20	21	Reminder 2
	22	23	24	25	26	27	28	Reminder 3
	29	30	1	2	3	4	5	Reminder 4
Oct 2025	6	7	8	9	10	11	12	Reminder 5 Follow-up closes (11.10.2025)
	13	14	15	16	17	18	18	
	20	21	22	23	24	25	25	
	27	28	29	30	31	1	2	
Nov 2025	3	4	5	6	7	8	9	
	10	11	12	13	14	15	16	
	17	18	19	20	21	22	23	
	24	25	26	27	28	29	30	
	1	2	3	4	5	6	7	Third follow-up (9 months)
Dec 2025	8	9	10	11	12	13	14	Reminder 1
	15	16	17	18	19	20	21	Reminder 2
	22	23	24	25	26	27	28	Reminder 3
	29	30	31	1	2	3	4	Reminder 4
Jan 2026	6	7	8	9	10	11	12	Reminder 5 Follow-up closes (12.01.2026)
	13	14	15	16	17	18	19	

Figure 1: Study Timeline