# Beliefs, Information Sharing, and Mental Health Care Use Among University Students*

Alisher Batmanov†
UC San Diego

Idaliya Grigoryeva†
UC San Diego

Bruno Calderón-Hernández
ITAM

Roberto González-Téllez
Stanford University

Alejandro Guardiola
Tec de Monterrey

September 21, 2025

**Abstract**

This paper investigates the role of beliefs and stigma in shaping students' use of professional mental health services at a large private university in Mexico, where supply-side barriers are minimal and services are readily accessible. In a survey experiment with 680 students, we find that nearly 50% of students in distress do not receive professional mental health support despite a high level of awareness and perceived effectiveness, constituting a substantial treatment gap. We document stigmatized beliefs and misconceptions correlated with the treatment gap. As three-quarters of students incorrectly believe that those in distress perform worse academically and that the majority of students going to therapy are in severe distress, we implement an information intervention to correct these beliefs. We find that it increases students' sharing of on-campus mental health resources with peers and encourages them to recommend these resources when advising a friend in distress. Interestingly, we find that it lowers respondents' willingness to pay for private therapy at the end of the intervention. Yet, this effect does not translate into a long-run reduction in self-reported therapy use 6 months after the experiment, with prior therapy users showing increased off-campus take-up.

†Corresponding authors.

*We thank Prashant Bharadwaj, Gaurav Khanna, Craig McIntosh, Paul Niehaus, Frank Schilbach, participants at the NHH Field Experiments Conference, the Caltech Behavioral Economics student conference, UC San Diego Applied and Development Seminars and UC San Diego Graduate Student Research Seminars for helpful suggestions and comments. We are also very grateful to Nishith Prakash and three anonymous referees who provided very valuable comments which improved the paper. We thank Nicholas Kruus, Manuel Domínguez, and Andrea Martin Arias, and participants of the undergraduate-graduate research lab (URL) at UC San Diego for their outstanding research assistance. We are grateful for financial support from the Weiss Fund, the Institute for Humane Studies, and UC Mexico Alianza Research Fellowship. We appreciate the logistical and implementation support from Carlos Ordonez and local university representatives, without whom this project would not have been feasible. Batmanov: abatmanov@ucsd.edu, Grigoryeva: igrigoryeva@ucsd.edu, Calderón-Hernández: bruno.calderon@itam.mx, González-Téllez: rob98@stanford.edu, Guardiola: aguardiola@tec.mx.

# 1 Introduction

Student mental health and wellbeing are issues of growing concern, with suicide being the 3rd leading cause of death among 15–29 years-old's and rates of depression and anxiety continuously rising (WHO 2021). At the same time, among over 100,000 adults surveyed across 30 countries in the World Mental Health Surveys, more than 80% of those struggling with depression, anxiety, or substance use disorders report not receiving any professional support, contributing to the "treatment gap" (Patel et al. 2018). That is despite widely recognized treatments to reduce depression and anxiety such as cognitive behavioral therapy (CBT) (Cuijpers et al. 2013, 2016). This treatment gap exceeds 90% in most developing countries, with a staggering 95% of people in distress lacking professional help in countries like Mexico (Wang et al. 2007). Importantly, such a wide gap is present even in settings where treatments are available (mitigating supply-side constraints). In those cases, the low take-up of such interventions is attributed to cognitive and behavioral biases, as well as low perceived effectiveness or need (Ridley et al. 2020; Patel et al. 2018; Andrade et al. 2014; Thapar et al. 2022).

Mental distress has serious consequences for educational attainment and long-term economic outcomes (Ridley et al. 2020). Depression and anxiety — the two most prevalent mood disorders[1] — can disrupt students' educational trajectories and constrain future employment and socio-economic mobility (Cornaglia et al. 2015; Fletcher 2008). Facing the pressure to perform academically while becoming independent adults, college students stand to benefit substantially from getting timely professional support, which may help prevent mild symptoms from escalating into severe depression or anxiety during college. Yet even where university counseling is readily available, most distressed students do not seek professional help (Acampora et al. 2023)[2]. This pattern raises the possibility that demand-side factors — especially beliefs and attitudes toward therapy — contribute to the remaining treatment gap.

We conducted a survey experiment with a representative sample of 680 students from a large private university in Mexico. We document the size of the treatment gap, examine students' beliefs about mental health and therapy use, and correct potential misconceptions through an information intervention. Six months after the initial survey, we invited the participants to complete a short follow-up survey. The survey targets questions on self-reported use and recommendations of therapy

[1] In this paper, we focus specifically on depression and anxiety, for which CBT and other talk therapy treatments have been demonstrated to be effective and often are provided by the university. We will not address more severe mental illnesses, such as schizophrenia or bipolar disorder, which typically necessitate psychiatric interventions combined with medications.

[2] Studies of non-representative or not college-specific student samples limit the analysis of the demand factors without data on the supply. Acampora et al. (2023) conducted the only comparable study focusing on a single institution measuring demand for university-specific and outside mental health services, which was conducted in a large university in the Netherlands.

to peers separately for on- and off-campus, along with their willingness to share personal mental health concerns with peers, enabling us to compare the short-run treatment effects immediately after the intervention with the long-run 6-month effects on self-reported behaviors. To our knowledge, this is the first study in a developing country to leverage a broadly representative university-level sample of the student population to document the prevalence of psychological distress, examine the factors influencing support-seeking behavior, and assess the role of inaccurate beliefs in contributing to the treatment gap.[3] Our study is set in a private university with free on-campus counseling to students. Because supply constraints in accessing counseling are minimal for students, the observed "treatment gap" is likely related to demand-side frictions, such as stigma or incorrect beliefs. At other universities, especially public institutions, free psychological counseling is often scarce and the students might face financial constraints and similar demand-side hurdles in addition to a tighter supply. Hence our estimate of a roughly 50% gap could be viewed as a lower bound for the broader Mexican university student population. These conditions make our setting suitable for testing whether an intervention to correct student beliefs related to seeking mental health support is effective in reducing barriers to care.

We find that there is a significant mental health treatment gap among university students in our study, despite the availability of free on-campus counseling services. In our sample, nearly 1 in 4 students exhibit moderate to severe symptoms of depression or anxiety. Nearly half of them do not receive professional mental health support: we estimate a treatment gap of nearly 50% of students in distress not having used any professional mental health support services in the last 12 months. Notably, this gap is present even though over 90% of students in distress agree that therapy can improve their mental wellbeing substantially, and 80% of them believe the university provides a good support system for mental or emotional health. The treatment gap is significantly larger among male students, as well as among those who are not open to sharing their mental health struggles with classmates. This suggests that discomfort with vulnerability or concerns about social judgment may be contributing factors. Interestingly, while financial stress is highly positively correlated with mental distress, there is no significant association between financial stress and treatment gap.[4]

Further analysis indicates that this gap is associated with stigmatized beliefs and prevalent negative stereotypes related to mental distress and help-seeking. We identify a particularly pervasive misconception as 3 out of 4 respondents believe that students in mental distress academically

[3]Most existing studies on university students' mental health and treatment use come from developed countries. For instance, an empirical study in the Netherlands examines an intervention targeting student mental health and therapy use (Acampora et al. 2023), while survey-based studies have documented related measures in Norway (Sæther et al. 2021) and among college students in the World Mental Health Surveys across 21 countries (Auerbach et al. 2016). A systematic review by Mortier et al. (2018) provides further references to studies using college-student data.

[4]Surprisingly, we even observe that students with a stressful financial situation are marginally *more likely* to seek help when in distress, in particular by being much more likely to seek professional help on campus compared to students not reporting struggling with finances, although these differences are not statistically significant.

perform *worse or much worse* compared to students not in distress — despite no observed correlation between GPA and mental distress score across students in our sample. This highlights a prevalent stereotype of associating mental health struggles with low academic achievement, which may discourage students from sharing their mental health struggles or revealing going to therapy, as these could be construed as signals of lower performance. Among students in distress who do not seek help, 81% guess a negative correlation in an incentivized question, relative to 74% among the rest of the students. Many students underestimate how many of their peers seek professional mental health help and are open to discussing mental health struggles while overestimating the prevalence of self-stigma, resulting in a more pessimistic view of public perceptions of stigma and their peers' attitudes toward mental distress.[5] Our results broadly echo the findings from recent online and field experimental studies that identify misconceptions around willingness to discuss mental health issues and the prevalence of mental-health-related beliefs among others as potential evidence of stigma (Roth et al. 2024a; Ridley 2025; Jain & Khandelwal 2024; Acampora et al. 2023).

Having documented that the treatment gap is correlated with inaccurate beliefs and perceived stigma, we design an information intervention to correct misperceptions about mental health in three ways: (1) conveying that psychotherapy has long-term (4–5 years) benefits in reducing instances of depression, (2) normalizing therapy by noting that most students at their university who seek it do not have severe symptoms, reinforcing that therapy is not just for those in crisis, and (3) countering the misconception about the link between distress and academic performance by informing students that GPA and mental distress are uncorrelated.[6] While 97% of subjects had a correct prior on the long-term effectiveness of psychotherapy (*prior 1*), we find that nearly half held incorrect priors about the proportion of students in therapy with mild or no symptoms (*prior 2*), and 75% incorrectly believed there was a negative correlation between GPA and mental distress (*prior 3*). To evaluate the impact of this intervention, we randomly assigned participants to either a Treatment group (T), which received the bundled information intervention, or a Control group (C), in which the participants answered questions about general campus services to ensure comparable survey engagement and completion duration.

The information intervention yields three main insights. First, participants in the treatment group were more likely to engage in sharing the link to the campus psychological counseling services, with a click-through rate nearly twice that of the control group, suggesting broader and more

[5]Public/social stigma refers to societal disapproval of individuals perceived as deviating from norms. Self-stigma, in contrast, occurs when individuals internalize these negative societal views, leading to feelings of shame or diminished self-worth. Experiencing mental distress can be associated with both forms of stigma.

[6]The survey design randomized participants into three groups: *Information + Reflection* (T1), *Information Only* (T2), and Control (C). Both T1 and T2 received the same set of three infographic messages: (1), (2), (3). T1 additionally included a brief reflection prompt and a vignette depicting a peer seeking therapy, intended to evoke empathy and reduce stigma. Given that our sample of 680 valid responses is underpowered to detect differences between T1 and T2, we pool them and refer to both as the Treatment group throughout the main paper.

sustained dissemination of on-campus counseling information. When asked to provide incentivized hypothetical advice to a friend in distress, treated participants were 3.6 percentage points more likely to mention on-campus counseling services, which corresponds to a large relative effect size of 70% of the control mean. Lastly, counter to our expectations, participants in the treatment group reported a lower willingness to pay for a one-month online therapy subscription. However, this lower willingness to pay did not translate into reduced therapy take-up in the long run. If anything, treated participants were slightly more likely to seek therapy off campus and no more likely to do so on campus, relative to control participants. We also document notable heterogeneity in the long-run effects. Those who reported having used therapy at baseline were more likely to both use and recommend off-campus therapy, while those who had not used therapy showed a weaker, opposite pattern, with slightly lower therapy take-up and a greater tendency to recommend on-campus rather than off-campus services.

Finally, we find that treated participants, particularly those with lower academic performance, became less willing to discuss their own mental health issues. We attribute it to the fact that our third information component might have drawn attention to the existence of the misperception that psychological distress and academic performance are negatively linked across students, as we stated that it is a common misperception. Thus, the intervention may have inadvertently signaled that peers continue to hold stigmatized views, increasing the perceived social cost of disclosure for some students. This explanation is consistent with elevating the salience of how student peers perceive others with mental distress symptoms, and consequently driving students who were seeking health off campus — as we see a suggestive increase in the take-up of off-campus therapy in the long-run follow-up. This result also points to a methodological insight for belief-correction interventions that target misperceptions related to social norms: while correcting these misperceptions, the interventions might want to refrain from emphasizing the fact that these incorrect beliefs exist or are prevalent.

Our findings suggest that fact-based first-order belief corrections, such as those related to participants' knowledge about therapy effectiveness or therapy-goers, are more effective at spurring low-cost, low-stakes behaviors than high-cost personal actions. Providing factual information significantly improved behaviors like sharing mental health information or recommending services to others, the actions entailing minimal financial or social risk. However, similar interventions had weaker effects on more costly behaviors such as openly disclosing one's own mental health issues or initiating therapy, where entrenched barriers remain (Smith 2025). Likewise, in a refugee setting, reducing stigma concerns increased peer-to-peer communication about mental health, yet did not translate into greater therapy uptake (Smith 2025). Across students in a Dutch university, a fact-correction intervention similarly does not deliver an increase in therapy use, while suggesting an increase in the demand for information and willingness to pay for a coaching service among a subset of respondents (Acampora et al. 2023). In a lab setting, factual first-order belief correction on

therapy effectiveness did deliver a higher WTP for private therapy within the experimental setting, yet we do not observe actual therapy take-up in that study in the long run (Roth et al. 2024b).

A small set of field studies shows that larger, more persistent behavior changes come from correcting what people think *others* believe (second-order beliefs). In Indian slums, telling residents that most neighbors were willing to discuss money and mental-health issues raised sign-ups for neighborhood savings circles and listening-volunteer programs by 15–20 percentage points (pp) and increased their contributions to the groups by 29% (Jain & Khandelwal 2024). Likewise, in Saudi Arabia, informing men that peers privately favored women's work made husbands 11 p.p. more likely to help wives job-hunt and increased wives' applications or employment by 4–5 p.p. after one year (Bursztyn et al. 2020). Yet, even successfully correcting misperceptions around social norms might not translate into longer-run behavioral changes: In a field intervention in schools in Rio de Janeiro, a classroom discussion halved the misperception with students overestimating others' support for aggressive "macho" norms ("toxic masculinity") immediately after the intervention and in a follow-up, but did not have a significant effect on self-reported incidents of violence or expressing vulnerable emotions (Matavelli 2025). Hence, designing mental-health interventions that embed both elements (credible facts and clear signals of peer acceptance and support) may facilitate shifting away more stigmatized beliefs into sustained, high-stakes help-seeking and personal disclosure. And overall, updating behaviors may involve longer-term interventions and/or follow-up reinforcements to sustain behavioral changes in the long run as has been observed with longer field interventions (Dhar et al. 2022).

Our paper contributes to the literature on mental health economics, behavioral frictions in help-seeking behavior, and the role of information interventions in addressing misperceptions and treatment gaps, particularly in developing countries. We provide new evidence on demand-side constraints in a setting where professional mental health services are available on campus, allowing us to isolate attitudinal and informational barriers from structural supply-side constraints. While previous work has examined the role of affordability and availability (Patel et al. 2017; Barker et al. 2022; Haushofer et al. 2021; Bhat et al. 2022), we contribute by documenting how belief distortions and stigma inhibit take-up despite widespread recognition of therapy's benefits. This extends the literature on behavioral constraints affecting mental health decisions (Schilbach et al. 2016; Shreekumar & Vautrey 2023) and connects to broader discussions on the implication of mental health and wellbeing economic decision-making in developing countries (Schilbach et al. 2016; Rao et al. 2021).

Second, we contribute to the literature on mental health stigma and misconceptions by documenting belief distortions among students regarding therapy use and academic performance. We find that students systematically overestimate the negative relationship between mental distress and GPA — a belief that may contribute to stigma and discourage help-seeking behavior, complementing an earlier result on productivity and mental distress in a stylized online setting (Ridley

2025) with a relevant productivity measure in an academic setting. We further complement existing online experiments with US adults (Roth et al. 2024a,b) with a more real-life setting and an interpersonally connected student sample from a single university in a developing country, with additional insights capturing behaviors around promoting help-seeking among students via link sharing and giving advice to a friend. Extending on the several field experiments related to mental health, stigma and treatment take-up in Jordan, India and Nepal (Jain & Khandelwal 2024; Lacey et al. 2024; Smith 2025), we leverage the setting where supply is reasonably available to zoom in on demand-side factors and beliefs. While prior work has explored information provision as a tool for reducing stigma and increasing take-up (Osman et al. 2022; Acampora et al. 2023; Jain & Khandelwal 2024), our study provides suggestive evidence that correcting misperceptions around facts related to therapy effectiveness and use may successfully increase overall information sharing and recommendations to peers, but without substantial increase in personal therapy seeking and disclosing one's own problems.

Third, our study connects to a broader literature showing that targeting second-order misperceptions related to social norms can propel costlier behaviors across diverse settings. In a field experiment with married men in Saudi Arabia, Bursztyn et al. (2020) document that Saudi men severely underestimated other men's approval of female employment; correcting this gap raised husbands' job-search assistance and translated into measurable gains in wives' formal employment one year later. In a field intervention in schools in Rio de Janeiro, Matavelli (2025) shows that many students greatly overstate classmates' support for aggressive "macho" norms and a single classroom discussion halved the misperception, all be it with limited effects on self-reported incidents of violence and expressing vulnerable emotions. By documenting a similar hierarchy in mental-health help-seeking, we observe low-cost behaviors respond to first-order facts more, whereas high-cost actions may require shifts in perceived social norms. By implementing a facts-based correction in a university setting, we provide the first evidence from a developing-country campus on the potential of factual misperceptions for moving peer-support behavior, with some, even if limited, effects on personal help-seeking.

The rest of the paper proceeds as follows. Section 2 describes the background and setting, Section 3 introduces our conceptual framework and theory of change, Section 4 provides the details on the implementation of the survey and the experimental variations, Section 5 documents the prevalence of mental distress, professional help utilization, and identifies the treatment gap and misconceptions related to mental health and treatment seeking, Section 6 discusses the treatment effects of our information intervention, and finally, Section 7 concludes with a discussion.

# 2 Background

## 2.1 Mental Health in Mexico: Context and Setting

Mental health is an issue of rising importance and concern, with around 280 million people around the world diagnosed with some form of depression (World Health Organization 2021), accounting for about 5% of all adults suffering from this disorder. Based on several recent surveys (Healthy Minds Survey 2022), university students are experiencing even higher rates of depression and anxiety, drawing further attention to this population in research and supporting an unmet need for support (Abrams 2022). In Mexico specifically, mental health issues have gained increasing attention as a recent report by the OECD (OECD Report 2022) places Mexico among the top-3 OECD countries with the highest prevalence of depression post-pandemic, indicating a concerning rise in the prevalence of mental health conditions in recent years (See Figure B1). While there are no systematic representative surveys of college students, one of the largest student surveys on mental health and wellbeing by coverage in the US identifies 44% and 37% of students struggling with depression and anxiety, respectively (Eisenberg et al. 2022). Furthermore, while over 80% of students report needing help, only 37% receive counseling, indicating a large potential treatment gap (Eisenberg et al. 2022).

In Mexico, data on mental health and wellbeing among young people and students in particular is limited. A mental health survey conducted in 2005 of a large representative sample of adolescents (over 3,000 children, aged 12–17 years old) living in Mexico City reveals the prevalence of any anxiety disorder in the past 12 months at almost 30% and any mood disorder (including depression) at 7.2% (Benjet et al. 2009). Nationwide, there is only one nationally and regionally representative source of mental health indicators to the best of our knowledge, the Mexican Health and Nutrition Survey (ENSANUT). Based on the ENSANUT survey results in 2023, about 12% of the country's population and 10% of those aged 17–28 scored above the half-score cutoff of 10, consistent with experiencing such symptoms most or almost all of the days (Figure B2) (Bose et al. 2024).[7]

As university enrollments rise, growing attention is drawn to mental health issues among students, a demographic going through critical life transitions and often being in a vulnerable

[7]The survey includes a depression screening questionnaire CES-D-7 (*Center for Epidemiological Studies Depression Scale*), consisting of seven questions evaluating if participants had experienced symptoms of depression in the week before the survey, such as "During last week, did you feel sad/depressed?" (Bose et al. 2024). While these measures are not the same as the standardized instruments more commonly used in Economics studies, such as the PHQ-8 and GAD-7 (Kroenke et al. 2001), this survey provides an alternative continuous score measure of depressive symptoms, providing the closest comparison to the prevalence of psychological distress in Mexico. While ENSANUT is representative at the national and regional levels (regions are defined as a partition of the set of Mexican Federal States), it is not representative for population subgroups, particularly our population of interest: university students, but it is the closest estimate in the absence of other student-specific surveys and highlights our project contribution.

emotional state. In particular, there has been a growing concern over suicides in major schools, including important ones in Mexico (Salud Mental 2022; Velazquez Hernandez 2017). As the number of university enrollments in Mexico surged by almost 50% from 2008 to 2022, reaching over 4 million students (Ministry of Education, 2023), an expanded demographic may be at risk, confounded by low availability of mental health services that are both affordable and effective. The mental health crisis is then further exacerbated by existing stigma and prejudice against recognizing mental distress and seeking treatment (Lagunes-Cordoba et al. 2021; Mascayano et al. 2016; Brewer et al. 2023).

For this project, we partnered with a large private university in Mexico with approximately 20,000 students.[8] Compared to most public universities in Mexico, this institution has substantially more resources and infrastructure to support student well-being, including widely accessible on-campus mental health services. This setting allows us to isolate the demand-side determinants of therapy use in a context where supply-side barriers are minimal. Notably, 85% of respondents agree that there is a good support system on campus for students who need professional help for their mental or emotional health, suggesting that concerns about service quality are unlikely to be a primary driver of underutilization. As a result, estimates of the treatment gap and associated belief distortions in this environment likely represent a conservative lower bound relative to those at public institutions, where students often face greater financial constraints, limited access to care, and higher levels of stigma.

## 2.2 Sample Description and Treatment Gap Evidence

We begin by describing the key demographic and academic characteristics of our student sample, summarized in Table 1. The sample consists of 680 students enrolled at one private university in Mexico.[9] The average respondent is 20 years old, and the gender distribution is approximately balanced, with 51% of participants being female. Nearly all respondents are undergraduate students, and 69% report receiving some form of scholarship support. The sample also reflects a relatively advantaged socioeconomic background: over 70% of participants report that both parents hold at least a Bachelor's degree. Roughly 75% of the sample are heterosexual.

The sample is broadly representative of the university's overall student population in terms of academic fields of study. For example, STEM majors make up 46% of our sample compared to 42% in the broader university population. Other fields, such as business and creative studies, are somewhat underrepresented, while medicine and health fields are overrepresented.

We next assess the prevalence of mental health challenges in our student sample using eight diagnostic questions from shortened versions of two validated screening tools: the Patient Health

[8] In 2021, the university enrolled over 16% of all university students in the state where it is located, a figure consistent with prior years (INEGI Statistics 2000–2023).

[9] Detailed sample recruitment procedures are described in Section 4.

Table 1: Student Characteristics (N=680)

| | Mean | SD |
|---|---|---|
| Female (%) | 51 | 50 |
| Age (Years) | 20.2 | 1.9 |
| Heterosexual (%) | 74.9 | 43.4 |
| Pursuing Bachelor's (%) | 91.2 | 28.4 |
| Full scholarship (%) | 7.9 | 27.1 |
| Partial scholarship (%) | 69.1 | 46.2 |
| Both parents w/ college degree (%) | 71.3 | 45.3 |

| Fields of Study | Sample | University |
|---|---|---|
| STEM (%) | 46 | 42 |
| Business (%) | 18 | 25 |
| Medicine & Health (%) | 20 | 10 |
| Law, Econ, Government (%) | 11 | 8 |
| Creative Studies (%) | 3 | 8 |
| Architecture & Environment (%) | 2 | 7 |

Notes: The table on the left reports sample means and standard deviations of student participants' characteristics. The table on the right presents the distribution of the survey sample and the university population across fields of study.

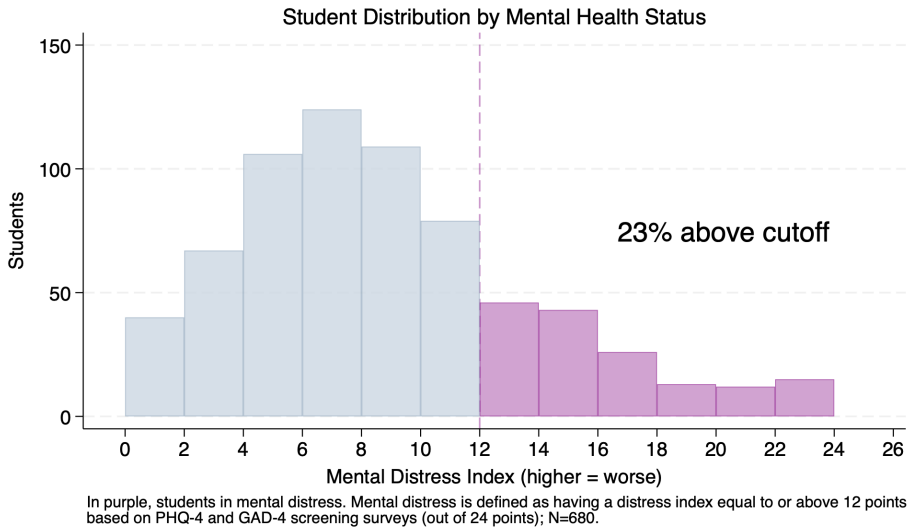
Questionnaire (PHQ-4) and the Generalized Anxiety Disorder scale (GAD-4). The PHQ captures symptoms of *major depressive disorder*, while the GAD captures symptoms of *generalized anxiety disorder*; both sets of questions ask students how frequently they experienced specific symptoms during the past two weeks (Kroenke et al. 2001; Spitzer et al. 2006). Each item is rated on a 0–3 scale and combined into a composite distress index ranging from 0 to 24. We classify students with a score of 12 or above as being in moderate to severe distress.[10] These measures are widely used in mental health screening and have been increasingly applied in economics research to study, for example, psychological distress among graduate students in the U.S. and to analyze the relationship between poverty and depression in low-income settings (Bolotnyy et al. 2022; Ridley et al. 2020).

For our sample of 680 students, Figure 1 depicts the distribution of mental distress index values with higher values indicating poorer mental health. The mean distress index is around 8.4 out of 24 possible points, which is slightly above the median value of 8. In our sample, 155 students are at or above the 12 point cutoff for distress, constituting 22.8% of all students with a 95% confidence interval of $[19.6\%, 26\%]$.[11] This shows that the prevalence of poor mental health in our sample of Mexican university students is substantial. To compare, during 2013–2016, 8.1% of American adults aged 20 and over experienced depression in a given two-week period, according to the Centers for Disease Control and Prevention (CDC) (Brody et al. 2018). In a large meta-analysis involving 44,503 participants aged 18 or older from 100 eligible studies, the prevalence of major depression was 10% (Negeri et al. 2021).

[10] The PHQ items ask how often in the past two weeks students have been bothered by: (1) little interest or pleasure in doing things, (2) feeling down, depressed, or hopeless, (3) feeling tired or having little energy, and (4) feeling bad about themselves or feeling like a failure. The GAD items ask about: (1) worrying too much about different things, (2) becoming easily annoyed or irritable, (3) being so restless that it is hard to sit still, and (4) feeling nervous, anxious, or on edge.

[11] Using a more lenient cutoff of 10 points yields a hefty 34.4% of students in distress, with a 95% confidence interval $[30.8\%, 38\%]$.

Figure 1: Mental Distress Index Distribution




Notes: This figure shows the distribution of the mental distress index across students in our sample. Blue bars represent observations for students with a mental distress score below the cutoff of 12 points, while purple bars denote observations for students with mental distress scores above the cutoff.

Several individual covariates exhibit notable differences between students in distress and those not in distress, as shown in Table B1. Students in distress are significantly less likely to identify as heterosexual (64.5% vs. 77.9%, $p < 0.001$) and are more likely to be in their third year of studies or above (63.9% vs. 50.5%, $p = 0.003$). They also report higher financial stress (70.3% vs. 51.4%, $p < 0.001$), suggesting that economic concerns may contribute to mental health disparities. Additionally, students in distress are slightly older on average (20.4 vs. 20.1 years, $p = 0.042$), and the fraction of students identifying as female is higher among those in distress (56.8% vs. 49.3%, $p = 0.104$). Other factors, such as GPA, scholarship status, and parental education, do not show statistically significant differences between the two groups.

The global treatment gap for mental health is significant, with over 80% of people with common mental health disorders — rising to more than 90% in poorer countries — not receiving treatment despite the availability of cost-effective solutions (Chisholm et al. 2016). Given a steady supply of counseling services in the university environment we are studying, it is not obvious ex-ante what the size of the treatment gap would be. The availability of and knowledge about services could, in principle, close the gap, but factors such as a lack of mental health literacy, stigma, and shame could, on the other hand, reduce demand.

We asked students in our survey about their use of professional mental health help in the last 12 months and, by splitting their responses based on whether they are in distress or not, categorized

Table 2: Professional Mental Health Help Use by Mental Distress

| | **Used Prof. Help** | **No Prof. Help** | **Total** |
|---|---|---|---|
| In Distress | 80 | 75 | 155 (23%) |
| Not in Distress | 190 | 335 | 525 (77%) |
| **Total** | 270 (40%) | 410 (60%) | 680 (100%) |

Notes: This table shows the cross-tabulation of students who have used professional mental health in the last 12 months and those who are in mental distress. We consider a student to be in distress if their mental health distress score is above or equal to 12.

them in Table 2 into one of the four groups.[12] Out of 680 respondents, 270 report using professional help either on-campus or off-campus, meaning 2 out of 5 students in our sample receive some form of support from a mental health professional. Notably, when focusing only on those in distress, we observe that 80 out of 155 students (52%) with moderate or severe symptoms of depression or anxiety have received professional treatment in the last year.[13] Therefore, the estimate of the *treatment gap* in our sample of university students in Mexico is 48%. This indicates that roughly half of students experiencing mental or emotional challenges are not receiving the psychological help they could benefit from, even though 80% of these students agree there is a good support system on campus for students who need professional help for their mental or emotional health.[14]

Still, students' largely positive views about therapy and its effectiveness do not imply an absence of stigma related to mental distress. As we show in subsequent sections, some students still report feeling disappointed in themselves when in distress and have concerns about being judged by others, including professors, parents, or peers. These perceptions may inhibit open conversations and delay help-seeking, even in a context where structural barriers to care are minimal.

While we defer a more detailed discussion of stigma in our context to a later section, here we consider whether our estimate of the treatment gap may be overstated because of the presence of potential stigma. For instance, if being in distress is stigmatized and people in distress use therapy, then sharing that one goes to therapy may signal their distress and thus be subject to stigma. If stigma about therapy-seeking is strong, perhaps some students do not truthfully report going to therapy, even though they do in reality. Although we cannot rule it out, we attempt to bound

[12]Specifically, we asked students whom they had turned to for help with mental health challenges in the past 12 months and recorded the share who selected either the '*mental health professionals at my university*' option, the '*mental health professionals outside of my university*' option, or both.

[13]When splitting the components of distress, we find that around 47% of those exhibiting symptoms of depression and 47% of those exhibiting symptoms of anxiety have received professional help.

[14]One could argue that a person in distress might not realize this, so even if they are aware that the campus provides support, they might not seek it. In our sample, 94% of those in distress report experiencing mental health challenges in the last 12 months (e.g., frequent stress, feeling anxious or down), which indicates a high level of awareness of their own mental distress.

the extent of any potential under-reporting using additional questions. Based on field focus-group discussions and several survey questions, the vast majority of students appear to hold favorable attitudes toward therapy and anticipate high levels of support from peers and family (see more on this in Subsection 5.1). Thus, we believe that under-reporting is likely modest in our context. When we only look at students who are open to sharing their mental health challenges with others ($N = 250$), we still estimate a 35% treatment gap. We believe this still represents a substantial share of students relative to the 48% treatment gap for the entire sample, which we view as a potential lower bound.

# 3 Conceptual Framework

In this section, we first use a simple dynamic model to clarify the trade-offs an individual faces when deciding whether to seek therapy. The framework builds on the foundational health-capital model of Grossman (1972), which treats health as a durable capital stock that individuals invest in over time to improve their overall utility. While Grossman's original model focused on physical health and time allocation, we adapt the structure to a mental-health setting in which therapy plays the role of investment. We then illustrate the theory of change underlying our intervention design by mapping different types of information treatments to specific components of the utility representation.[15]

Our formulation captures two key features of mental-health care: first, that therapy can improve both internal wellbeing and the quality of a person's social relationships; and second, that such improvements are uncertain and come at a cost—not only financial, but also social and psychological (stigma).

## 3.1 Treatment-Seeking Decisions

Consider an agent who seeks to maximize the present value of her lifetime well-being. Her utility depends on three components: a mental-health stock $H(t)$, a social-capital stock $S(t)$, and consumption of a general good $Z(t)$. These stocks summarize her psychological functioning and the strength of her social ties, respectively. Time is continuous and future utility is discounted at rate $\rho > 0$.

At each instant $t$, the agent decides whether to attend a therapy session, denoted by the control variable $D(t) \in \{0, 1\}$. Therapy is the only available action that can replenish either form

[15]For instance, some parts of an intervention are intended to shift beliefs about the stigma costs associated with therapy, while others may affect the perceived benefits of mental or social well-being. By clarifying how these elements enter the agent's decision problem, the framework helps interpret how various treatments may influence both therapy-seeking behavior and the willingness to disclose or discuss personal issues.

of well-being. In the absence of intervention, both $H(t)$ and $S(t)$ depreciate over time. When the agent chooses to go to therapy, she receives a discrete improvement to both stocks with probability $\pi_1 \in (0,1)$; with probability $1-\pi_1$, the session has no effect. This probability captures the likelihood that therapy produces a meaningful improvement in well-being, conditional on attending.

The agent's objective is to choose a path of therapy decisions over the time horizon $[0,T]$ to maximize lifetime utility:

$$\max_{D(\cdot)} \int_0^T \left[u_H(H(t)) + u_S(S(t)) + Z(t) - D(t)(S_s + S_p)\right] e^{-\rho t}\, dt.$$

Flow utility at time $t$ reflects the direct value of mental and social well-being, current consumption, and potential psychological costs associated with seeking therapy. Specifically, attending therapy reduces flow utility through two stigma channels: a self-stigma cost $S_s > 0$, reflecting internal feelings of shame or weakness, and a perceived stigma cost $S_p > 0$, capturing discomfort associated with how others might judge her decision to seek help. Therapy also carries a monetary cost $p_T$[16], and consumption is constrained by a constant income flow $Y$, so $Z(t) = Y - D(t)p_T$.

Both stocks evolve according to capital-accumulation equations. Without therapy, mental health and social capital depreciate exponentially at constant rates $\delta_H$ and $\delta_S$, respectively. When therapy is attended and effective, the agent receives fixed gains $G_H > 0$ and $G_S > 0$ to each stock. These gains and decay rates together determine the overall trajectory of well-being over time. More detailed discussion of the framework and the full solution of the model can be found in Appendix C.

This set-up generates a simple behavioral rule: the agent goes to therapy at time $t$ if and only if the expected benefit outweighs the total cost. Formally, the agent chooses to go to therapy when

$$\boxed{\pi_1(B_H + B_S) \;\geq\; p_T + S_s + S_p}$$

where $B_H$ and $B_S$ represent the present-value marginal benefits from an incremental improvement in mental health and social capital, respectively. These benefits reflect how much the agent values improvements in well-being—both immediately and in the future—and depend on the current state of each stock, the utility functions $u_H$ and $u_S$, and the magnitude of the therapy-induced gains.

This decision rule captures the central trade-off: therapy is undertaken when the discounted utility gain from a possible improvement in well-being exceeds the full contemporaneous cost. The left-hand side is shaped by how effective therapy is likely to be and how much the agent stands

[16] In our setting, we can also think of this price as incorporating the opportunity cost of going to therapy, including the monetary cost, the search costs to find the therapy provider, and the time cost of actually going. Hence, this price is lower but not zero even if it is the on-campus free therapy.

to gain if it succeeds. The right-hand side aggregates all costs: the monetary cost, the internal discomfort of seeking help, and the fear of being judged. In this way, the model helps explain which behavioral margins interventions may act upon. Information treatments that increase the perceived benefit of therapy, reduce self- or perceived stigma, or alter expectations about effectiveness will all shift the balance of this inequality and affect both therapy-seeking behavior and the likelihood of sharing personal issues with others.

**Recommending Therapy to a Peer.** The framework can be extended to represent an agent's decision to recommend therapy to a peer. Rather than making a choice that affects her own well-being, the agent now considers a prosocial action shaped by other-regarding preferences: she derives utility from her peer's potential improvement (Buchmann et al. 2024). We assume that the agent evaluates the peer's expected net benefit from therapy in the same way as in the individual-level decision rule, and internalizes it with weight $\alpha \in (0, 1)$, which captures the strength of her other-regarding concern.

Let $U_j = \pi_1(B_H + B_S) - p_T - S_s - S_p$ denote the peer's instantaneous net utility from attending therapy. At the same time, recommending therapy may impose a psychological cost $C_r > 0$ on the agent, reflecting anticipated discomfort, reputational concerns, or fear of being intrusive. Thus, the agent chooses to recommend therapy at time $t$ if and only if

$$\boxed{\alpha\, U_j \;\geq\; C_r}$$

This decision rule mirrors the individual's own therapy-use condition but operates on a distinct other-regarding margin. Interventions that reduce the social cost of recommending therapy or that shift beliefs about its value for others can increase peer-to-peer engagement with mental health care.

## 3.2 Theory of Change

Our intervention seeks to recalibrate potential misperceptions that students may hold about therapy and psychological distress — misperceptions that shape key components of their decisions to seek help or to recommend therapy to peers. It targets informational and psychological frictions that may contribute to underutilization of mental health services. By shifting beliefs about the effectiveness and appropriateness of therapy, as well as perceptions of how psychological distress relates to relevant academic outcomes, the intervention aims to influence both the perceived benefits and costs of seeking care. These changes, in turn, may contribute to reducing the treatment gap observed in this population.

Table 3: Predicted Effects by Intervention Component

| **Intervention Component** | **Own Mental Health** | **Peer-Directed Behaviors** |
|---|---|---|
| I1. Therapy Effectiveness | Increases perceived likelihood that therapy leads to doing better ($\uparrow \pi_1$), raising expected benefit from seeking care | Raises perceived benefit of therapy for peers ($\uparrow \pi_1$), increasing resource recommendations |
| I2. Therapy Users | Reduces internalized and perceived stigma associated with help-seeking ($\downarrow S_s$, $\downarrow S_p$) | Lowers discomfort in endorsing therapy for others ($\downarrow C_r$); reinforces norms around mental health support |
| I3. Distress and Grades | Normalizes distress and weakens link between symptoms and academic failure ($\downarrow S_p$) | Reduces stigma attached to visible symptoms, potentially increasing openness to endorse therapy |

The intervention operates through three complementary channels that correspond to three informational components embedded in the treatment:

1. **Therapy Effectiveness (I1).** Students might underestimate the probability that therapy will lead to meaningful improvement. This underestimation depresses the expected benefits of seeking care and discourages investment in mental health. By presenting clear evidence of long-term improvements in depression following therapy, the first intervention component aims to raise students' perceived likelihood of improvement ($\pi_1$). In our framework, this shifts the expected benefit term $\pi_1(B_H + B_S)$ upward, increasing the likelihood that the perceived benefit outweighs the cost of attending therapy.

2. **Therapy Users (I2).** Students often believe that therapy is only appropriate for individuals with severe mental health conditions. This perception can deter those experiencing mild or moderate distress from seeking care, as they may internalize feelings of inadequacy or anticipate negative social judgment. By emphasizing that the majority of students receiving therapy report only mild symptoms, this intervention component reframes therapy as a resource suitable for a broader population. As a result, it is expected to reduce both self-stigma ($S_s$) and perceived stigma ($S_p$), while also lowering the psychological cost of recommending therapy to others ($C_r$).

3. **Distress and Academics (I3).** Students frequently believe that psychological distress leads to academic underperformance. This belief may reinforce stigma by making mental health struggles appear socially or academically discrediting. The third component addresses this misconception by presenting data showing no meaningful relationship between distress and GPA in the student population. This information is intended to normalize distress and reduce perceived judgment associated with seeking help, thereby lowering perceived stigma ($S_p$).

Together, these intervention components are designed to operate on both sides of the decision inequality: increasing the expected benefits of therapy through belief updating about effectiveness, and decreasing the psychological and social costs of seeking care by reducing stigma. The same logic applies to peer recommendation behavior, where agents internalize others' expected utility from therapy. By shifting both beliefs about the value of therapy and the perceived cost of encouraging others to seek help, the intervention can increase treatment uptake and foster greater peer-to-peer engagement with mental health resources.

We expect our intervention to influence a set of outcomes that capture key margins of mental health care decisions, both *inward-facing* (related to one's own help-seeking behavior) and *outward-facing* (related to supporting or encouraging others). On the inward side, outcomes such as willingness to pay for therapy, self-reported therapy use, and willingness to discuss one's own mental health reflect how students' beliefs about therapy's benefits and costs evolve. As our framework predicts, these outcomes should respond positively to increased perceived likelihood of improvement ($\uparrow \pi_1$) and reductions in self- and perceived stigma ($\downarrow S_s, \downarrow S_p$). The ranking task complements these measures by capturing more implicit beliefs, specifically the extent to which distress is socially penalized, and provides an indirect proxy for shifts in stigma and normalization ($\downarrow S_p$).

Outward-facing outcomes reflect how students engage with peers around mental health. Recommending therapy and sharing resource links signal a willingness to support others' care-seeking, shaped by both concern for peer welfare ($\uparrow \alpha U_j$) and reduced reputational or interpersonal costs ($\downarrow C_r$). The donation measure captures both belief in therapy's value and altruistic preferences toward expanding access, while also indirectly validating updated beliefs about effectiveness ($\uparrow \pi_1$). These outcomes, taken together, allow us to trace the mechanisms through which belief calibration and stigma reduction translate into concrete behavioral changes across multiple domains of mental health care.

Table 4: Outcome Mapping by Type and Mechanism

| **Outcome** | **Type** | **Targeted Mechanism** |
|---|---|---|
| Willingness to Pay | Inward | Incentivized measure of private valuation of therapy. Responds to updated beliefs about benefit likelihood and reduced self-stigma; friend WTP also reflects other-regarding concern ($\uparrow \pi_1,\ \downarrow S_s$) |
| Therapy Use | Inward | Realized uptake of therapy as a function of increased perceived benefit and reduced stigma-related costs ($\uparrow \pi_1,\ \downarrow S_s,\ \downarrow S_p$) |
| Ranking Task | Inward | Measures implicit bias toward individuals experiencing distress. Reflects normalization of distress and reduced perceived academic or social consequences ($\downarrow S_p$) |
| Willingness to Discuss Own Mental Health | Inward | Captures increased comfort with self-disclosure; reflects internalized norm shifts and reduced shame or fear of judgment ($\downarrow S_s,\ \downarrow S_p$) |
| Resource Link Sharing | Outward | Reveals willingness to forward mental health resources; responds to reduced social hesitation and increased perceived value of therapy for peers ($\downarrow C_r,\ \uparrow \alpha U_j$) |
| Peer Recommendation | Outward | Captures willingness to support peers struggling with mental health. Reflects reduced stigma around recommending therapy, stronger concern for peer welfare, and updated beliefs about therapy's effectiveness ($\downarrow S_p,\ \uparrow \alpha U_j,\ \uparrow \pi_1,\ \downarrow C_r$) |
| Therapy Donation | Outward | Captures altruistic valuation of therapy and willingness to subsidize access for others; reflects belief in effectiveness and concern for peer welfare ($\uparrow \alpha U_j,\ \uparrow \pi_1$) |

# 4 Experimental Design

In this section, we describe the survey experiment, the data collection process, and the follow-up study. We then describe our treatment-randomization procedure and present evidence of successful randomization by showing that pre-determined covariates are balanced across experimental groups.

Finally, we define our pre-registered main and secondary outcomes before outlining the empirical strategies used for testing our various hypotheses.[17]

## 4.1 Intervention & Design

The survey and intervention were implemented over a short 9-day time window from November 16 to November 24, 2024, during the second half of the academic semester and prior to the exam period. We advertised the survey widely by re-sharing among specific major program coordinators and professors via e-mail, student organizations and student groups via WhatsApp and Facebook, and campus-location-targeted advertisements on Instagram. Our survey was advertised as a "Student Experience Survey" with a mix of guaranteed and raffle-based survey payouts, incentivizing completion while reducing the potential selection into the survey among students based on their prior beliefs and preferences about wellbeing and mental health (Healthy Minds Survey 2022; Acampora et al. 2023).

Participants were incentivized through a combination of guaranteed payments, random lottery draws, and performance-based bonuses for incentivized questions.[18] As a result of a wide recruitment campaign, we had over 1,000 people start the survey in just over a week's time, resulting in 680 complete responses that pass validation and attention checks. The median survey completion time was 21 minutes. The combination of recruitment channels, incentives and relatively low time cost for completing the survey allow us to get a representative sample of the student population during the academic semester, providing an informative snapshot of student mental health, beliefs, and treatment use.

In our survey, we leverage a reproducible unique respondent identifier to maintain privacy while enabling payment processing of participant performance-specific amounts and linking to the follow-up data (similar to Acampora et al. (2023)). At the beginning of the survey, participants create a Unique ID while verifying their university affiliation through institutional e-mail address in a separate form accessed by participants after completing the survey. This form is not linked to their survey responses in any way.[19] The ID section is followed by screening questions about mental health and demographics. We then gather data on therapy use, barriers and students' beliefs

[17] See Appendix E for the baseline pre-analysis plan (PAP). The baseline intervention was pre-registered in the AEA RCT Registry under RCT ID AEARCTR-0014804. We also added a pre-registration for the follow-up study.

[18] We offered a guaranteed payment of $200 MXN ($10 USD) to each of the first 100 respondents to incentivize early completion, and also offered a larger-prized raffle for which we randomly drew twenty respondents among the 680 valid-response participants, each of whom won a $2,000 MXN ($100 USD) gift card, and we give $50 MXN for correctly answering one randomly selected bonus question out of eight.

[19] The Unique ID combined the elements of each respondent mother's name, respondent's birth day, last name initials and last two phone number digits. This unique ID was then used to also link the baseline survey experiment responses to the follow-up data where we also asked the participants to re-create this unique ID. This resulted in a high match rate, with only 2 out of 350 responses not being matched to the baseline.

about therapy effectiveness and prevalence, before assessing stigma-related questions and awareness about on-campus services. After collecting baseline data, we randomly split survey respondents into three experimental groups: two treatment groups and one control group. At this stage we elicit all students' prior beliefs about the misperceptions we aim to correct in order to establish their existence. Following the priors elicitation, we show the information treatments (placebo questions) to the treatment (control) groups, and we elicit treated students' posterior beliefs to assess the extent to which the interventions managed to correct such misperceptions.

Afterwards, we collect information related to our outcomes of interest. We include a behavioral measure of sharing mental health- and therapy-related information. We do this by observing the number of clicks on a link for sharing information of on-campus services. This measure allows us to get at revealed preferences — as opposed to stated intentions — given that sharing information entails actual costs, such as having to think whom to share the information with or the risk of being perceived as intrusive. In addition to this, we implement an incentive-compatible approach to elicit students' willingness to pay for a one-month online therapy service and their willingness to donate part of their survey earnings to help cover the cost of a therapy session for a student from their university who reports that financial constraints prevent them from seeking therapy. To conclude the survey we ask respondents to provide thoughtful advice for a hypothetical friend who approaches them for emotional support (see Appendix Figure B30 for detailed survey flow).

To gauge the effectiveness of addressing attitudinal barriers to mental health care, we implemented a light-touch information intervention composed of three complementary components delivered together as a single treatment. These components were designed to target common belief-based obstacles to help-seeking: (Fact 1) perceived effectiveness of therapy, (Fact 2) the misconception that therapy is only appropriate for students in severe distress, and (Fact 3) the stereotype that higher mental distress is strongly linked to lower academic performance. The selection of these facts was informed by prior literature on psychological barriers to care (Andrade et al. 2014; Ridley et al. 2020) and early-stage fieldwork revealing the persistence of such misperceptions among students. Presenting these messages jointly allowed us to address multiple co-occurring misconceptions in a way that reflects the complexity of real-world mental health stigma and decision-making, though it also means we are not able to experimentally disentangle the separate effects of each component.[20]

We purposefully target first-order *facts* (therapy's long-run efficacy, the mild-symptom profile of most users, and the null GPA–distress correlation) rather than second-order beliefs. Fact-based corrections have been the more reliable lever for increasing help-seeking in prior mental-health

[20] After measuring prior and posterior beliefs, we find out our information treatments reduce the share of people with incorrect beliefs by 38 and 37 percentage points for facts 2 and 3 respectively. Misperceptions about Fact 1 were tiny, with just 3.1% of respondents having the incorrect prior belief, thus leaving little room for meaningful updating. We thus expect the updating of Facts 2 and 3 to be the main drivers of our results.

studies (Roth et al. 2024b; Acampora et al. 2023), while second-order corrections have yielded mixed or even negative effects on WTP (Roth et al. 2024a)[21].

### Treatment Groups

In our between-subjects design, we randomly assign the 680 students to one of three conditions:

- **Treatment 1 (T1): Information + Reflection** ($n = 227$) Students in this group were shown three different sets of information in the form of infographics. The first infographic shows that a recent study found that offering psychotherapy leads to an 11% drop in mild depression and an 8% drop in moderate depression four to five years later. The second infographic showed information disclosing that "Among [University] students who are receiving professional mental help, 2 out of 3 have only mild or no symptoms of depression and anxiety." The third and last infographic showed that "Among 53 [University] students, 3 out of 4 respondents believe that a student with mental health issues performs ***worse*** or ***much worse*** academically than a student without mental health issues. But our survey data show ***no relationship*** between students' GPA and mental distress."

  In addition to the infographics, students in this group were prompted with the following message: "*Many university students sometimes struggle with feelings of being overwhelmed, anxious, or depressed. Based on your experience, what are some effective ways students can manage these types of mental health challenges? Please explain your thoughts.*" Furthermore, we showed students in this group one of two vignettes[22] with an image of a fictitious student from their university and describing a hypothetical situation in which this student seeks help from a therapist after suffering a panic attack.

- **Treatment 2 (T2): Information Only** ($n = 221$) Students in this group were shown the same infographics as the ones shown to students in the Information + Reflection treatment with the difference that no reflection activities or vignette components were part of the treatment for this group.

- **Control (C):** ($n = 232$) Students in the control condition were not shown the infographics nor any of the vignettes. They answered additional questions about various university services to keep the overall survey time closer to that in the treatment groups.

The baseline survey allowed us to measure the behavioral information-sharing outcome and to implement the lab-in-the-field approach to measuring willingness-to-pay for mental health ser-

[21] A notable exception is Jain & Khandelwal (2024) which is specifically correcting a second-order belief, we will discuss how our findings relate to this study in the results section and discussion.

[22] The only differences across vignettes are the sex of the student appearing in the images and the name of the student. We did this to rule out treatment effects being driven by the sex of the student in the hypothetical situation.

vices. To complement these results and disentangle the personal-versus-social stigma, as well as distinguish on- versus off-campus treatment seeking, we fielded a follow-up e-mail survey inviting the 680 students in our sample to participate. The follow-up survey allows us to (i) test whether the large short-run effects on information-sharing translate into sustained on-campus services recommendations, (ii) examine the existence of a substitution effect in which students express more interest in on-campus therapy over the private online resource, and (iii) probe the internal-versus-external stigma channel through questions about willingness to talk about own mental health issues and therapy usage.

The follow-up survey was intentionally concise with nine "yes/no" questions, delivered via e-mail by different members of our project team and field assistants. In an attempt to collect a large amount of responses in a narrow time window, we incentivized responses with a raffle of 40 gift cards (valued roughly at $50 USD each). The questions focused on capturing self-reported respondent behavior over the past 6 months since the baseline survey (use of professional mental-health services on- vs. off-campus, recommendations of those same services to peers, and whether respondents had discussed their own distress or other students' therapy use with other University students). We obtained responses from 355 students, out of which 320 unique respondents provided a valid unique ID, institutional e-mail and are still attending classes at the university in 2025.[23] We were thus able to contact 47 % of our baseline sample for the follow-up survey in just 15 days time.

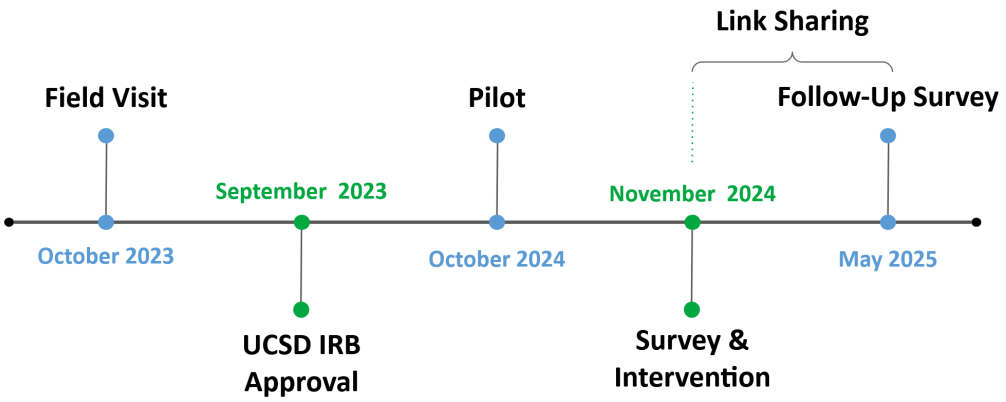
Figure 2 depicts the main stages of our intervention. We collect baseline information in November 2024, and the follow-up information six months later from April 20, 2025 to May 5, 2025. We obtained IRB (808688) approval in September 2023, prior to our pilot study, additionally on October 2024 we obtained IRB (P000882) approval from a Mexican private university. Importantly, information for our link-sharing outcome is gathered throughout the whole six months between baseline and follow-up surveys as access to the link is not constrained to accessing it during the time survey responses were being collected.

### Randomization

The intervention was implemented using Qualtrics' built-in randomizer tool. After completing baseline module — including consent to participate, mental health screening, demographics, therapy use and beliefs, on-campus service use and availability, and priors — participants were randomly assigned at the individual level to one of three groups: Treatment 1 (Information + Reflection), Treatment 2 (Information Only), or Control. The randomizer was configured to distribute respondents uniformly across treatment groups. In the T1 group, respondents were further randomized

[23] We restrict our sample to people still attending classes at the university since in our follow-up we ask about behaviors related to on-campus services. This results in losing 17 responses from people who are no longer taking classes at University in 2025.

Figure 2: Timeline of Activities




Notes: In October 2023, we conducted a pilot survey with 53 student participants. We obtained IRB approval from a private Mexican university on October 2024 with approval number P000882. Note the link-sharing outcome is measured continuously through the whole period from the baseline to the follow-up survey.

to view one of two possible vignettes describing a hypothetical scenario of a female/male student experiencing mental distress and seeking counseling. Table 5 presents evidence that treatment and control groups were not statistically different on pre-determined covariates.[24] The only exception is the share of female respondents in the treated groups is statistically larger than in the control group by 7.5 percentage points ($p$-value$<$0.1), one out of 11 covariates, which suggests this difference should not be of concern regarding bias in our estimates.

We also test and show that the covariate balance holds among the subjects who responded to the 6-month follow-up survey in Appendix Table B5. Furthermore, there is no differential attrition by treatment status: there is a 50.43% attrition rate among control and 50.45% attrition rate among treated subjects.

## 4.2 Data

### 4.2.1 Outcome Variables

In this section, we outline the set of pre-registered primary and secondary outcomes that our information intervention intends to shift.[25]

1. **On-Campus Counseling Link Sharing**. At the end of the survey, students were given an opportunity to share a link to on-campus counseling services with their peers. We tracked both the total number of human clicks and the number of unique users who clicked the link across three experimental conditions. In addition, we observe the share of clicks directly from within the survey platform (Qualtrics, presumably clicked by respondents themselves), as well

[24] See Appendix Table B6 for a balance table comparing all three experimental groups.
[25] More details about the exact processing of responses to each question can be found in Appendix A.

Table 5: Covariate Balance

| Variable | (1) Control N | (1) Control Mean/(SD) | (2) Treated N | (2) Treated Mean/(SD) | (1)-(2) Pairwise t-test N | (1)-(2) Pairwise t-test Mean difference |
|---|---|---|---|---|---|---|
| Age | 232 | 20.159 (1.848) | 448 | 20.145 (2.031) | 680 | 0.014 |
| Female | 232 | 0.461 (0.500) | 448 | 0.536 (0.499) | 680 | -0.075* |
| Financially Stressed | 232 | 0.530 (0.500) | 448 | 0.571 (0.495) | 680 | -0.041 |
| Has Scholarship | 232 | 0.651 (0.478) | 448 | 0.712 (0.453) | 680 | -0.061 |
| Receives a full scholarship | 232 | 0.082 (0.275) | 448 | 0.078 (0.269) | 680 | 0.004 |
| Moved Residence | 232 | 0.591 (0.493) | 448 | 0.621 (0.486) | 680 | -0.030 |
| GPA | 232 | 90.897 (4.659) | 448 | 91.007 (4.727) | 680 | -0.110 |
| MH Score | 232 | 8.569 (5.132) | 448 | 8.237 (5.054) | 680 | 0.332 |
| Used Therapy L12 Months | 232 | 0.233 (0.424) | 448 | 0.234 (0.424) | 680 | -0.002 |
| Open to Share MH Challenges | 232 | 0.392 (0.489) | 448 | 0.355 (0.479) | 680 | 0.037 |
| Self-stigmatize | 232 | 0.323 (0.469) | 448 | 0.286 (0.452) | 680 | 0.038 |

Notes: We pool T1 and T2 into a "Treated" group. This table shows balance on covariates across treatment groups. For each covariate we show each experimental group's sample mean and standard deviation, as well as the difference in means across both groups. Age measures the respondent's age in years, female is an indicator equal to one if the respondent is female-born, financially stressed is an indicator equal to one if the respondent described her financial situation as "Always", "Often" or "Sometimes" stressful and equal to 0 if she reported it as "Rarely" or "Never" stressful, Has scholarship is an indicator equal to one if the respondent has at least some amount of scholarship, receives a full scholarship is an indicator equal to one if the respondent's scholarship covers 100% of tuition, moved residence is an indicator equal to one if the respondent moved her residence city to pursue her current studies, GPA measures the respondent's current overall GPA on a scale from 0–100, MH score measures the student's mental health score as described in Section 2, used therapy in L12 months is an indicator equal to one if the respondent states having used therapy in the last 12 months, open to share MH challenges is an indicator equal to one if the respondent states she would be willing to share about her own personal MH challenges with others and self-stigmatize is an indicator equal to one if the respondent states she would be disappointed in herself if she suffered from mental distress. Standard errors for the difference in means test are heteroskedasticity robust. Significance levels: * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$

as those clicked via re-shares such as e-mails or SMS. We are not able to distinguish between few respondents sharing in bulk vis-à-vis many respondents sharing with few other people.

2. **Peer Advice**. Participants were asked to imagine a scenario where a friend approaches them for emotional support due to personal struggles. They were then prompted to provide open-

ended advice, which was evaluated by the length of the advice given (in words) and by whether respondents mention words such as 'therapy', 'support you', 'empathy', among others, in their response.

3. **Willingness to Pay for Therapy**. As a proxy for participants' demand for therapy, we use incentive-compatible BDM-style willingness to pay (WTP) measures (Becker et al. 1964). Specifically, we measured the maximum amount participants were willing to pay for a one-month therapy subscription from *BetterHelp*, both for themselves and for a friend (two separate incentivized questions). We normalize WTP responses by the monthly price of BetterHelp, dividing the value reported by the respondents by the equivalent price of around 6,500 Mexican pesos.[26]

4. **Donation**. Participants were asked about the share of their earnings from participating in the study they were willing to donate to help fund a therapy session for a financially constrained student at their university.[27] Participants were notified that any donation they pledged would be automatically deducted from their payment and allocated toward this funded therapy session.

5. **Ranking questions**. We asked participants to rank individuals in terms of how comfortable they would be working with them on a joint course project. We describe six hypothetical students with different traits, all of which might make it undesirable to work with a particular student. Specifically, we assess whether respondents deem it more undesirable to work with a low GPA student relative to with a student who talks about mental health issues or shows signs of having them.

6. **Therapy Use (long term)**. In the follow-up survey, we asked students whether they had used professional therapy or psychological counseling in the past six months. We ask one question for on-campus services and another one for off-campus services.

7. **Recommendations (long term)**. In the follow-up survey, we asked students whether they recommended professional therapy services to their peers. Again, we ask for both on- and off-campus services explicitly.

8. **Willingness to share/discuss issues/therapy use (long term)**. We also ask students whether they have talked about their mental health problems with other University students,

[26] In Subsection 6.2, we show that the treatment effect estimates are robust to not normalizing and/or winsorizing the WTP measure.

[27] Students were informed that their donations would be directed toward covering the cost of 1 therapy session for a fellow university student who reported that financial constraints prevent them from seeking therapy.

and whether they have talked about their or their University peers' experience with on-campus therapy or psychological counseling.

#### 4.2.2 Mental Health Care Measures and Elicited Beliefs

1. **Mental distress**. We compute a mental distress index using the PHQ-4 and GAD-4 screening questionnaires for depression and anxiety, respectively (Kroenke et al. 2001; Spitzer et al. 2006). Each question has four possible responses with values ranging from 0–3; we compute the index by summing over values across questions. Larger values imply worse mental distress and the index's support is [0, 24]. As is common practice in the health sector (Kroenke et al. 2009), we classify students as being in distress if their mental distress index is greater than or equal to the index support's midpoint of 12.

2. **Mental health care use & perceived therapy use**. We ask students whether they have/have not used professional mental health help in the last 12 months. Additionally, we asked them to guess out of every 100 University students, how many of them did they think have used professional mental health help in the last 12 months.

3. **Perceived therapy effectiveness**. We tell students that a review of 22 studies examining the effectiveness of psychotherapy for treating depression was conducted. We then ask them how many studies do they think show that therapy is an effective treatment for depression out of the 22 analyzed. Additionally, we ask them two Likert-style questions to measure the extent to which they believe therapy can improve their own (people's) mental wellbeing.

4. **Self-stigma**. To measure self-stigma we ask students how much do they agree or disagree with the statement "I would feel disappointed in myself if I had a mental health issue (e.g., anxiety or depression)." We also ask students to guess how many survey participants of the study out of every 100 responded to the aforementioned question with "Strongly Agree", "Agree", or "Somewhat Agree."

### 4.3 Study Protocols

The project received ethics approval from the University of California San Diego on September 1, 2023 and from Tec de Monterrey on October 22, 2024. We pre-registered our baseline analysis in the AEA RCT Registry under RCT ID AEARCTR-0014804. The pre-analysis plan (PAP) is publicly available on the Open Science Framework website. We additionally included a pre-analysis plan for our follow-up survey.

**Deviations from PAP:** We specified we would run regressions of outcome variables on treatment binary variables, "*controlling for key demographic and socio-economic covariates that may*

*be unbalanced at baseline [...].”* We deviate from this in two ways: (i) our main results do not include covariates, and (ii) in the robustness checks we select covariates based on a post double-selection LASSO algorithm (Belloni et al. 2013), which reduces researcher degrees of freedom.

**Hypothesized effects and mechanisms:** For our baseline analysis we pre-registered six hypotheses. In four of them we hypothesize positive treatment effects on link-sharing, WTP and donations. In the remaining two we hypothesize about heterogeneous effects by (a) size of misperception about therapy effectiveness, and (b) degree of mental health stigma. For our follow-up analysis we pre-registered that treated students would be more likely than control students to use and recommend professional mental health services. Furthermore, we hypothesized that if we observed positive effects on recommendations but not on own-usage, it would be suggestive of the information promoting peer interactions about mental health topics rather than individual demand for therapy. We also hypothesize that if we observe stronger effects for on-campus counseling use or recommendations, compared to off-campus counseling, it would be evidence of a substitution effect from off-campus options towards on-campus services. Lastly, we expect heterogeneous effects by GPA, mental distress and stigma.

## 4.4 Empirical Specification

### Information Sharing

We consider two groups: a treatment group (T) with $n_T$ individuals and a control group (C) with $n_C$ individuals. Let $k_T$ and $k_C$ be the total observed clicks from the treatment and control groups, respectively. We wish to test whether the underlying click rates in the two groups differ. Since each participant in our study could generate an unbounded number of link clicks, we modeled the click counts using a Poisson process. Denote by $\lambda_T$ the (unknown) rate of clicks per person in the treatment group and by $\lambda_C$ the (unknown) rate in the control group. The null hypothesis asserts that both groups share the same click rate, i.e. $H_0 : \lambda_T = \lambda_C$, whereas the alternative is $H_1 : \lambda_T \neq \lambda_C$. In practice, this is often expressed as testing whether the *rate ratio* $\lambda_T/\lambda_C$ equals 1.[28]

Given that our click counts are relatively small, in addition to running a test relying on large-sample approximations (Wald test, in our case), we also employed an *exact test* for two-sample Poisson comparisons (in the spirit of Fisher's exact $p$-value test on binomial data). Under $H_0$, the total number of clicks $k_T + k_C$ is fixed, and the conditional distribution of $k_T$ (the count in

[28] One could in principle model this environment as a comparison of two binomial random variables, where each observation can either result in success or failure. Thus, a binomial framework assumes a fixed upper limit on the number of "successes" each participant can contribute (e.g., at most 1 click per person). In our study, however, each participant could potentially produce multiple clicks, so there is no obvious upper bound. We, therefore, model such unbounded count data using Poisson distribution, with each group's total number of events (clicks) assumed to be Poisson($\lambda_T\, n_T$) or Poisson($\lambda_C\, n_C$), respectively.

the treatment group) is binomial with parameter

$$p = \frac{n_T}{n_T + n_C}.$$

Thus, the test assesses whether the observed $k_T$ is unreasonably large or small relative to this binomial distribution, thereby providing an exact $p$-value for the hypothesis $H_0 : \lambda_T = \lambda_C$.[29]

In addition to the joint treatment (T1&T2) vs. control comparison, we separately tested other pairwise differences (e.g., T1 vs. control, T2 vs. control, and T1 vs. T2). For each comparison, the method returns (i) a *rate ratio*, $\hat{\lambda}_T/\hat{\lambda}_C$, estimated by the ratio of observed click rates, (ii) an *exact* two-sided $p$-value, and (iii) an indicator of whether we reject $H_0$ at 5% level. Unlike approximate Poisson methods, the exact approach remains valid even when $k_T$ and $k_C$ are small. However, it does not provide a confidence interval for the rate ratio in the current implementation; we therefore focus on $p$-values and the estimated ratio to interpret group differences in click rates.

**Main Regression Specification**

To estimate treatment effects on our primary outcomes, we use a regression specification which allows us to evaluate the effectiveness of our interventions. Our specification examines the pooled effect of any intervention (T1 or T2) compared to the control group. Our estimating equation is:

$$Y_i = \alpha + \beta InfoTreatment_i + \varepsilon_i,$$

In this specification, $Y_i$ represents the outcome of interest for individual $i$, such as advice-related measures, willingness to pay for therapy, or self-reported stigma. The variable $InfoTreatment_i$ is an indicator equal to 1 if individual $i$ received any of the treatment conditions (T1 or T2), and 0 otherwise. Finally, $\varepsilon_i$ represents the error term. The coefficient $\beta$ captures the average treatment effect of the pooled intervention on the specified outcome.[30] Since we are underpowered to detect effect sizes of the magnitudes we observe for most outcomes (see Subsection B.4), we only present estimates separately by T1 and T2 for link sharing and peer advice in Subsection B.5.

# 5 Student Beliefs & Misconceptions

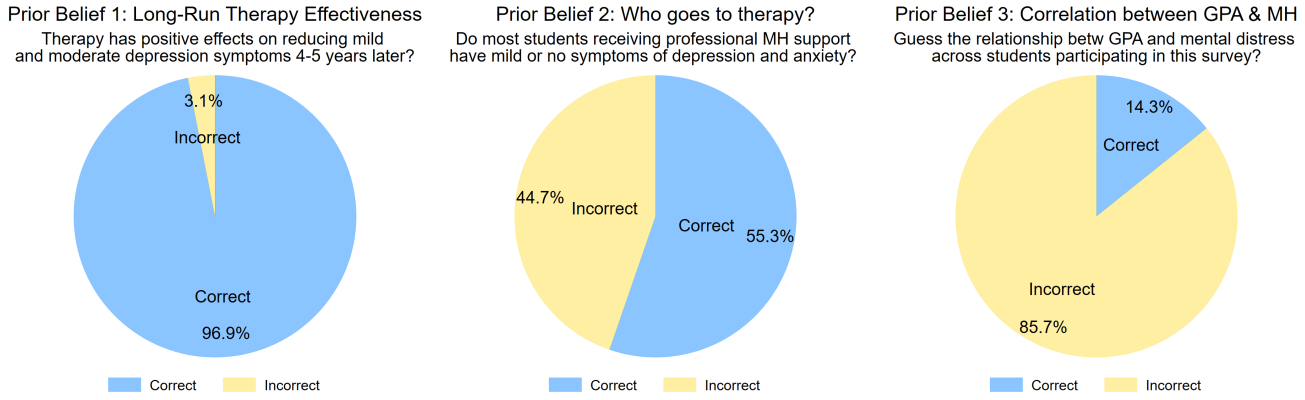
Having established the presence of a mental health treatment gap in our sample of students in Mexico, this section explores factors that may contribute to underutilization of professional support. We begin by presenting descriptive evidence on students' beliefs about the effectiveness of therapy and its prevalence among peers. We then examine observable characteristics correlated with the

[29] We carried out the exact Poisson test using `statsmodels` in Python with the `method="exact-cond"` option.

[30] We show the main results including a specification including LASSO-selected controls in Subsection B.7.

treatment gap and highlight several miscalibrated beliefs that may underlie students' decisions not to seek help. Finally, we present suggestive evidence of stigma surrounding mental distress, including students' reluctance to disclose or discuss their mental health with others.

Figure 3: Prior Beliefs about Therapy and Mental Health




Correct answer:
Yes, positive effects 4-5 years later

Correct answer:
Yes, most have mild/no symptoms

Correct answer:
No correlation b/n GPA & Distress

Notes: This figure shows the share of students ($N = 680$) who answered each of our three prior belief questions correctly or incorrectly.

## 5.1 Beliefs About Therapy Effectiveness & Peer Use

Previous research attributes the mental health treatment gap primarily to attitudinal barriers, such as low perceived need, skepticism about treatment effectiveness (Andrade et al. 2014), and stigma (Schnyder et al. 2017), despite rigorous and consistent evidence demonstrating the effectiveness of interventions (Cuijpers et al. 2013) — including in low-resource settings (Patel et al. 2017; Barker et al. 2022; Lacey et al. 2024) and specifically among college students (Cuijpers et al. 2016). We therefore begin by examining students' beliefs about the effectiveness of mental health treatments to assess whether low perceived effectiveness might contribute to the treatment gap.

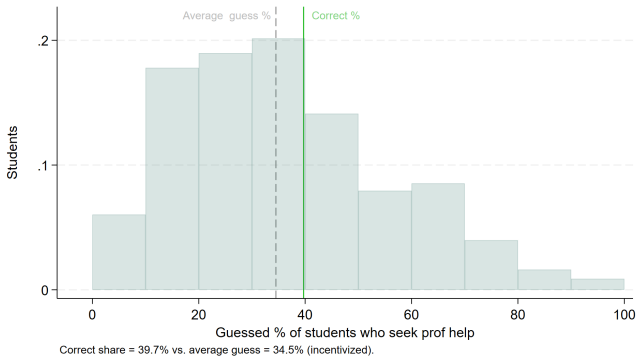
Surprisingly, we find overwhelming evidence that perceived effectiveness is high: over 90% of students agree that therapy can improve both their own and others' mental well-being. Additionally, social support for seeking therapy appears strong, with more than 91% of students believing their friends would support them in doing so, and 87% reporting the same for their parents (Appendix Table B2). These findings point to generally positive attitudes toward mental health treatment—both personally and socially—which is particularly noteworthy in a developing-country context, where stigma and more conservative views around therapy are typically more prevalent (Bhat et al. 2022; Jain & Khandelwal 2024). This pattern may reflect the relatively privileged context of our field site and student sample. Nonetheless, it highlights the importance of identifying additional contribu-

tors to the treatment gap in settings where access and perceived effectiveness are already relatively favorable.

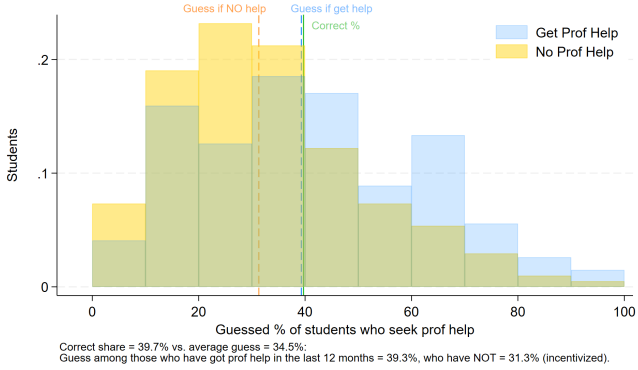
Additional incentivized questions further support the conclusion that students generally hold optimistic views about the effectiveness of therapy. When asked how many out of 22 high-quality clinical studies (as in Roth et al. (2024b)) showed that therapy is effective for treating depression, students provided a mean estimate of 17 studies. While this falls short of the correct answer (all 22 studies demonstrated effectiveness), the response indicates substantial confidence in therapy's impact. Moreover, 97% of students correctly answered the incentivized question (*prior belief #1*) about whether therapy would continue to reduce symptoms of depression four to five years after treatment (left panel of Figure 3). One of the three informational components in our intervention conveyed precisely this fact, suggesting that most students were already aware of therapy's long-term effectiveness. Together, these results imply that limited perceived effectiveness is unlikely to be a major contributor to the treatment gap in our context.

Figure 4: Student Guesses of the Prevalence of Professional Help-Seeking

(a) All Students




(b) By prior therapy use



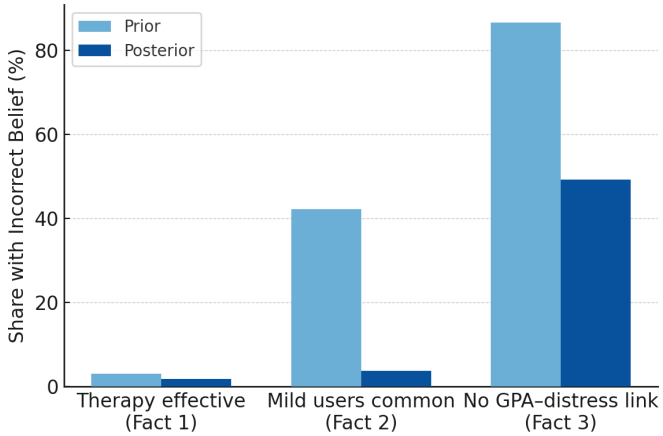

Notes: This figure shows the distribution of guesses of the percentage of students who seek professional help among University students. In Panel (a) we show all respondents while on Panel (b) we split the sample by an indicator of whether the respondent got professional help in the last 12 months (self-reported). See the CDF's of guesses by prior therapy use in Appendix Figure B7.

In our sample, 39.7% of students report having received professional mental health support in the past 12 months, and two-thirds have done so at some point in their lives. About 20% received care through on-campus services, 26% off-campus, and some used both. Additionally, 87% of students report having a friend who has received professional mental health support, highlighting the potential for peer-based information transmission (See Table B2 for full-sample means and Table 6 for means split by the respondent's level of distress). Yet, students' beliefs about therapy use among their peers further reveal important gaps. As shown in Figure 4, most students underestimate how common therapy use is, with an average guess of 34.5%. This misperception is driven largely by students who have not sought therapy themselves — their average guess is 31.3%, while those with prior therapy experience estimate 39.3%, nearly matching the true rate. These results suggest

that non-users in particular hold miscalibrated beliefs about prevailing norms, which may in turn reinforce hesitation to seek help.

The observation that many students who have not sought therapy themselves tend to underestimate how common its use is among their peers may be linked to another widespread misconception: that professional mental health support is primarily for individuals with severe symptoms. In a separate incentivized belief elicitation question, we asked whether most students receiving therapy have mild or no symptoms of depression or anxiety, or instead have moderate or severe symptoms (*prior belief #2*). While the correct answer is the former, only 55% of students answered this question correctly (central panel of Figure 3). Among treated students who were shown this fact during the intervention, the share holding the incorrect belief dropped to just 3.8% in the posterior elicitation (Figure 5), indicating substantial belief updating. This misconception may discourage help-seeking among students who feel their struggles are not "serious enough" to warrant therapy.

Figure 5: Belief Updating in the Information Intervention




Notes: This figure shows the share of students in the treatment group ($N = 448$) holding incorrect beliefs before (Prior) and after (Posterior) the information intervention across the three facts targeted in our treatment. For Fact 1 (long-term therapy benefits), 3.1% of students initially held an incorrect belief, compared to 1.8% after the intervention. For Fact 2 (most therapy users have mild or no symptoms), the incorrect belief rate fell from 42.2% to 3.8%. For Fact 3 (no GPA-distress correlation), it dropped from 86.6% to 49.3%.

While beliefs about the effectiveness and prevalence of therapy shape students' perceptions, they do not fully account for the treatment gap. Notably, perceived effectiveness of therapy does not differ significantly between students in distress and those not in distress, suggesting that skepticism about therapy's efficacy is unlikely to be a primary driver of the treatment gap (Table 6). At the same time, while students in distress are more likely to have sought professional help in the last 12 months (15 p.p. more likely) compared to their non-distressed peers, there are also more students in distress who report they would unlikely seek help when struggling with mental health (13 p.p. more) than among those currently not in distress, indicating that barriers beyond perceived effectiveness may contribute to avoid seeking help, such as stigma or other beliefs or misperceptions.

Table 6: Perceived Effectiveness & Help-Seeking by Distress

| | (1) Not in Distress Mean (SD) | (2) Distress Mean (SD) | (2)-(1) Pairwise t-test Mean difference |
|---|---|---|---|
| **A. Perceived Effectiveness & Support** | | | |
| **Perceived Effectiveness of Therapy:** | | | |
| Guess # studies ↓ depression (correct 22) | 17.02 | 17.33 | 0.31 |
| | (4.39) | (4.32) | |
| Agree: Therapy can improve my own well-being | 0.90 | 0.94 | 0.04 |
| | (0.31) | (0.25) | |
| Agree: Therapy can improve people's own well-being | 0.92 | 0.94 | 0.02 |
| | (0.27) | (0.25) | |
| **Perceived Support for Therapy:** | | | |
| Agree: Friends would support me going to therapy | 0.91 | 0.92 | 0.00 |
| | (0.28) | (0.28) | |
| Agree: Parents would support me going to therapy | 0.88 | 0.83 | –0.05* |
| | (0.32) | (0.37) | |
| **B. Use of Professional Mental Health Help** | | | |
| Sought professional mental health help **in the last 12 months** | 0.36 | 0.52 | 0.15*** |
| | (0.48) | (0.50) | |
| → professional MH help **on campus** | 0.19 | 0.26 | 0.07* |
| | (0.39) | (0.44) | |
| → professional MH help **off campus** | 0.23 | 0.37 | 0.15*** |
| | (0.42) | (0.49) | |
| Have **ever** received professional MH help | 0.63 | 0.77 | 0.14*** |
| | (0.48) | (0.42) | |
| Unlikely to seek help when struggling with mental health issues | 0.15 | 0.28 | 0.13*** |
| | (0.36) | (0.45) | |
| Have a friend who received professional MH help | 0.88 | 0.87 | –0.01 |
| | (0.33) | (0.34) | |
| Have a friend who would benefit from therapy | 0.88 | 0.95 | 0.07** |
| | (0.33) | (0.22) | |
| **Sample size** | 525 | 155 | 680 |

Notes: This table shows the difference in means across students who are/are not in distress for questions related to perceived effectiveness and support, as well as the use of professional mental health. Difference = Distress - No distress. Sample size (680). ***, **, * indicate 1, 5, 10% significance. This table shows means for questions on perceived effectiveness, support and therapy use. For items under the Perceived Effectiveness of Therapy and Perceived Support for Therapy panels we ask *How much do you agree or disagree with the following statements?* (1) *Going to therapy can improve my own mental health* (2) *In general, going to therapy can improve people's mental wellbeing* (4) *My friends would show support if I told them I am going to therapy* (5) *My parents would show support if I told them I am going to therapy*; we code as "agree" responses which state Somewhat Agree, Agree or Strongly Agree. For items under the Professional Help Received panel we ask the following Yes/No questions: (i) Have you ever received professional mental help? (ii) Do you have a friend who is currently receiving or has previously received professional mental health?, and (iii) Do you have a friend or someone you know closely who you think would benefit from therapy? Finally, we ask *If you experienced mental health challenges in the last 12 months, [...], to who did you turn for help? Select ALL that apply* for items under the (Last 12 Months) panel.

Examining predictors of help-seeking among those in distress, we find that students who are less open to discussing mental health issues with classmates exhibit a 23 percentage point higher treatment gap, suggesting that stigma or discomfort with vulnerability may serve as important barriers to care.[31] When asked whom they turned to for help with mental health challenges in the past 12 months, more than 40% of students reported relying on informal support from friends or family members (Figure B6). While these networks may offer more immediate emotional support, they are frequently not a sufficient substitute for professional care to address the root causes of students' mental distress. These patterns point to the role of stigma, social norms, and concerns about how one is perceived by others as meaningful frictions in the decision to seek mental health support, which is the topic we explore next.

## 5.2 Mental Health Misconceptions & Stigma

Stereotypes and misconceptions about mental health often shape beliefs about productivity and performance, which in turn influence individuals' willingness to disclose their mental health status. Furthermore, as seeking therapy may be perceived as a signal of poor mental well-being, some might feel discouraged from talking about their mental health struggles and accessing professional help. Prior research by Ridley (2022) found that people strongly believe workers experiencing mental distress perform worse on a communication-related task in an online experimental setting, yet his results demonstrate no actual difference in performance. Our exploratory field visits revealed similar patterns in personal anecdotes and focus group interviews, constituting a prevalent stereotype that we document below for our student sample.

We identify a particularly pervasive misconception related to mental distress and academic performance: 75% of respondents believe that students experiencing mental health issues perform worse or much worse academically than those without such issues, despite our data showing no relationship between mental distress and GPA (*prior belief #3*). As shown in the right panel of Figure 3, only 14% of students correctly report that there is no relationship, while 11% believe the relationship is actually positive. To address this misconception, we included a third informational component in our intervention, presenting students with data collected during a pilot study which demonstrates no correlation between GPA and psychological distress among their peers. As Figure 5 illustrates, this led to a 37 percentage point reduction in the share of students holding the incorrect belief.[32] This component is designed to recalibrate a stereotype particularly salient in university settings, one that may discourage students from disclosing their struggles or seeking professional help. When we plot students' cumulative GPA against their mental distress index using the full sample ($N = 680$), no meaningful relationship emerges between the two variables, as shown

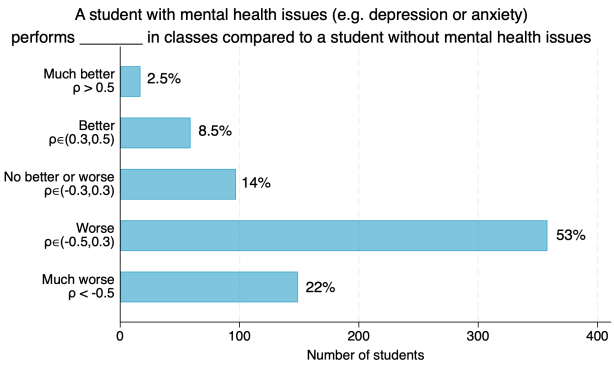
[31] Male students also show a significantly larger treatment gap than female students (12 percentage points).

[32] Following the information intervention, nearly 50% of treated respondents continued to believe in a negative relationship between GPA and mental health, despite being shown data to the contrary.

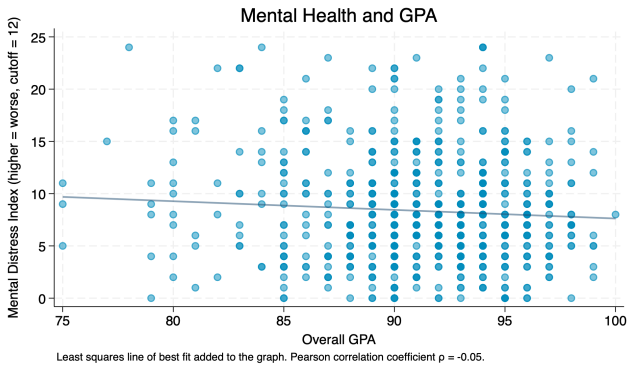
in Figure 6. Although this pattern contradicts prevailing student beliefs, the scatter plot in the right panel reveals virtually no correlation ($\rho = -0.05$)[33]. The systematic tendency by a student's peers to overestimate the negative association between mental distress and academic performance may reinforce stigma and deter some students from seeking help. Notably, the treatment gap is substantially larger among students who believe in a negative relationship (51%) compared to those who correctly perceive no relationship (33%)[34], suggesting that correcting this misconception could play a role in narrowing the gap in mental health service utilization.

Figure 6: Correlation between Mental Distress and GPA

(a) Incentivized Student Guesses




(b) Observed GPA & Mental Distress




Notes: Panel (a) shows that most students (75%) guess that the relationship between GPA and mental distress across students is negative. We elicit their beliefs in an incentivized question, clarifying that the correct answer will be calculated across the participants based on their GPA and answers to the MH questionnaire. Panel (b) shows that there is no significant relationship between mental distress and GPA, with the correlation coefficient of $\rho = -0.05$. We also test this relationship using a binary distress measure (in distress if score above 12), and equivalently find no significant relationship.

Building on this pattern of misperceptions, we also find that students tend to hold overly pessimistic beliefs about how others perceive and respond to mental health struggles. Specifically, many overestimate how common self-stigmatizing beliefs are among their peers and underestimate their peers' willingness to share mental health challenges. In our sample, 30% of respondents agree that they would feel disappointed in themselves if they had a mental health issue such as anxiety or depression — a belief we classify as self-stigmatizing. When asked to estimate how many of their peers feel the same way, those who hold this belief guess an average of 62% — more than double the actual share. Even those who do not personally endorse this belief still overestimate its prevalence, with an average guess of 45%. The 17 percentage point difference between these two groups is large and statistically significant ($p < 0.001$). Similarly, while 37% of respondents say they would be

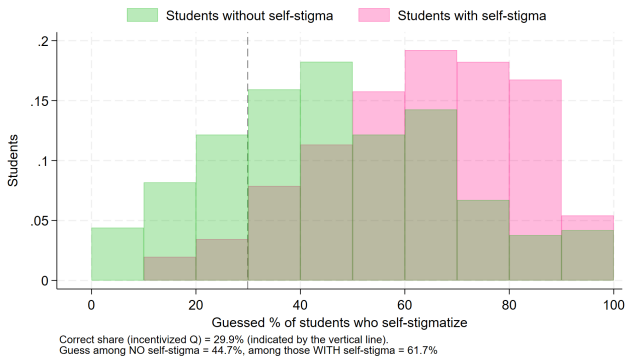
[33]Using data from Healthy Minds Survey (2022) we find a -0.107 correlation coefficient between distress and GPA for a sample of students aged 17–28 in institutions in the United States. This provides some external validity to the finding in our sample, although the data from the comparison sample comes from a developed rather than developing country.

[34]The magnitude of the difference is substantial but not statistically significant ($p = 0.13$).

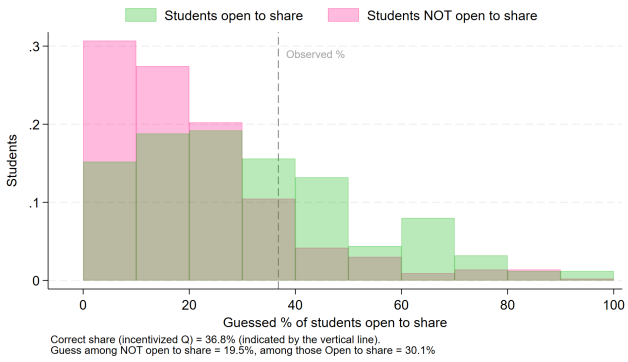
willing to share mental health challenges with classmates who are not necessarily their friends, they estimate about 30% of their peers would do the same; those unwilling to share themselves estimate an even lower peer willingness of 20%. Both perceived rates are significantly below the true value, and the 10 percentage point gap between the two groups is also statistically significant ($p < 0.001$; See Table B4 for the statistical strength of each relationship). Taken together, these results reflect a broader pattern of projection: students tend to assume that others share their own beliefs and behaviors. This tendency is evident not only in openness estimates but also in beliefs about self-stigma, where perceived prevalence tracks closely with one's own endorsement of the belief. While projection may serve as a familiar cognitive shortcut, it becomes particularly problematic when pessimistic views are projected onto others, reinforcing distorted perceptions of social norms around mental health. More broadly, these misperceptions align with recent findings from field experiments that document widespread underestimation of peers' openness and overestimation of stigma-related beliefs (Roth et al. 2024a; Ridley 2022; Jain & Khandelwal 2024; Acampora et al. 2023).

Figure 7: Perceptions: Self-Stigma and Openness to Share

(a) Perceived Self-Stigma




(b) Perceived Willingness to Share Mental Health Challenges




Notes: This figure shows, in Panel (a), the distribution of guesses of the percentage of students who would be disappointed in themselves if they had a mental health issue, and in Panel (b), the distribution of guesses of the percentage of students who would be open to share their mental health challenges with classmates who are not necessarily their friends. We show the distributions by respondents who do/do not self stigmatize, and by respondents who would/would not be open to share.

## Stigma Categorization

Stigma toward mental health is multifaceted. In our setting, direct stigma toward the use of therapy appears relatively low: most students rate therapy as highly effective, believe others benefit from it, and anticipate strong support from both friends and family. They also hold relatively accurate beliefs about how common therapy use is among peers, and qualitative responses suggest openness to discussing therapy as a tool for increasing awareness and take-up. Still, seeking therapy may carry an indirect social cost if it is viewed as a signal of psychological distress, which could lead

some students to under-report therapy use. This nuance is important: our results suggest that even when students understand the value and effectiveness of therapy, stigma related to being in distress can persist. Such under-reporting may add noise to our survey measures and suppress our ability to detect treatment effects, implying that our estimates may be conservative. Evidence from our data is consistent with this interpretation: students in psychological distress who are less open to discussing mental health issues with classmates exhibit a 23 percentage-point higher treatment gap, underscoring the role of stigma and discomfort with disclosure as barriers to care.

To better understand these dynamics, we classify mental health stigma using a 2-by-2 framework based on the target of the belief (self vs. others) and the perceived holder of that belief (oneself vs. others). As summarized in Table 7, *self-stigma* refers to internalized negative attitudes about one's own distress. In our study, this is captured through agreement with the belief that one would feel disappointed in themselves if they had a mental health issue. *Perceived stigma* reflects expectations about how others would respond to one's distress, proxied in our data by respondents' stated willingness to share mental health challenges with classmates who are not necessarily close friends.

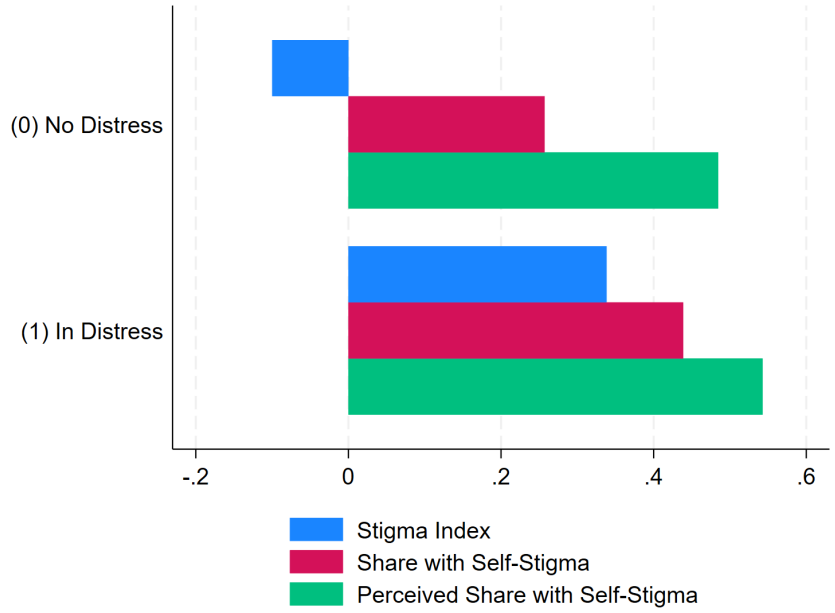
Table 7: Types of Mental Health Stigma

| | Beliefs held by me (*first-order beliefs*) | Beliefs held by others (relate to *second-order beliefs*) |
|---|---|---|
| About myself | **Self-stigma** *I believe I am weak when I am in distress* | **Perceived stigma** *I believe others think I am weak when I am in distress* |
| About others | **Personal stigma** *I believe others are weak when they are in distress* | **Perceived public stigma** *I believe others think people are weak when they are in distress* |

On the external dimension, *personal stigma* captures negative beliefs held about others who experience mental health challenges. We proxy this through ranking-based questions in which participants evaluate hypothetical classmates with traits such as low GPA, visible distress, or openness about mental health issues. *Perceived public stigma* refers to beliefs about the broader social climate — for example, what classmates, professors, or parents are thought to believe about students with mental health struggles. We capture this using participants' incentivized guesses about the share of peers who would feel disappointed in themselves if they had a mental health issue. Distinguishing these types of stigma is critical for identifying the precise barriers to help-seeking: students may view therapy as socially accepted, yet still hesitate to seek help if they internalize distress as a sign of personal failure or anticipate judgment from others — a dynamic also observed in recent research, where shifting beliefs about distress, rather than about treatment, played a key role in increasing help-seeking (Lacey et al. 2024).

To facilitate quantitative analysis of how stigma relates to mental health outcomes, we construct a composite stigma index that aggregates several dimensions of stigmatizing beliefs. While stigma is inherently multidimensional — shaped by individual attitudes, perceived social norms, and structural expectations — our goal is to summarize this variation using a reduced-form measure. Drawing on the typology in Table 7, we include three core components: (i) perceived public stigma, measured through incentivized guesses about how many peers would internalize distress as personal failure; (ii) personal stigma, captured by rankings of hypothetical peers showing signs of distress or openness about mental health; and (iii) perceived prevalence of self-stigma, reflected in beliefs about how many peers would feel disappointed in themselves if they experienced mental health problems. These beliefs not only reflect individual views but may also reinforce perceived norms and influence help-seeking behavior. We apply Principal Component Analysis (PCA) to these inputs and focus on the first principal component (PCA1), which accounts for the largest share of variation across stigma-related responses.

Figure 8: Stigma Measures By Distress




Notes: This figure shows (i) the average of the stigma index (measured in standard deviations), (ii) the share of students who have self-stigma (percent share on the axis), and (iii) the average guess of students from University who would be disappointed if themselves if they had a mental issue (percent share on the axis). We show averages and shares by an indicator of whether the student is in distress or not according to the mental distress index.

As shown in Figure 8, stigma-related beliefs are more prevalent among students who report being in psychological distress. These students are more likely to personally endorse self-stigmatizing views, and they also perceive such beliefs to be widespread among their peers. The composite index captures this clustering of internalized and projected stigma, providing a concise summary

measure that we use in subsequent heterogeneity analyses.[35] While this correlation does not by itself establish directionality — whether stigma contributes to distress, or distress shapes one's perception of stigma — it highlights the potential role of belief-based barriers in sustaining the treatment gap. By capturing variation across both individual attitudes and perceived norms, the stigma index serves as a useful empirical tool for identifying students who may be more resistant to help-seeking interventions.

# 6 Intervention Treatment Effects

In this section, we present our main results. We first describe results on information-sharing and peer advice, then shift attention to results on WTP and long-run therapy use, and finish with suggestive evidence on the mechanisms that explain our results and types of mental-health-related stigma revealed by our intervention.

## 6.1 Information-Sharing & Peer Advice

**Information-sharing:** We find that our intervention leads to the treated students being more likely to share information about on-campus counseling services with their peers (pre-registered). In the short run, as we asked the students to share the link to on-campus therapy information with peers, we observed that the link shared with the treated students was clicked 136 times, compared to only 35 clicks observed for the link shared with the control students. These figures imply click-through rates of total clicks per respondent who saw the link at the end of the survey of about 30% and 15%, respectively, given 448 total treated respondents and 232 control respondents (Table 8). We use the Poisson test to compare the click-through rates by treatment status by calculating the click-rate ratio $\lambda_T/\lambda_C$ and comparing it to 1 ($H_0$ implying no difference between the click rates). For total clicks observed one week after the initial survey, we get a ratio of 2 ($0.304/0.151 = 2.01$) with the treatment click rate being double that of the control group, which allows us to reject the null hypothesis that there is no difference in link sharing rates between the treatment and the control groups at the 0.1% level.[36]

We then provide three additional metrics to compare link-sharing between groups, which we collected at the treatment group level with the link-sharing platform, including unique link clicks, clicks from outside the Qualtrics survey platform, and additional updated numbers of long-run

[35] For details on the construction and interpretation of the stigma indices, see Appendix Table B21, which shows the correlation of PCA1 and PCA2 with the underlying components.

[36] See Appendix Table B7 for further details on our application of the Poisson test to this setting. There, we also present results by treatment arms, separating the effects for T1 vs. T2. Clicks by T1 students were lower than those observed for T2 students. Even when there is no significant difference in completion times between both treatment groups, students in T1 spent on average five more minutes completing the survey. We conjecture this extra time made T1 students less likely to engage in information sharing at the end of the survey.

Table 8: Professional Support Link Sharing and Statistical Test Summary

| | Treatment (T) | | Control (C) | | Poisson Test |
|---|---|---|---|---|---|
| | Clicks | Rate ($\lambda_T$) | Clicks | Rate ($\lambda_C$) | (approximate; exact test) |
| **A. Link Engagement** | | | | | |
| Total Clicks (Dec 2024) | 136 | 0.304 | 35 | 0.151 | *** ($p < 0.001$ ; $p < 0.001$) |
| Total Participants | 448 | | 232 | | |
| **B. Additional Metrics** | | | | | |
| Unique Clicks | 94 | 0.210 | 24 | 0.103 | *** ($p < 0.001$ ; $p = 0.001$) |
| Non-participant Share | 82% | 0.249 | 60% | 0.091 | *** ($p < 0.001$ ; $p < 0.001$) |
| Total Clicks (May 2025) | 179 | 0.400 | 58 | 0.250 | *** ($p < 0.001$ ; $p = 0.002$) |

Notes: *Total Clicks* refers to the total number of link engagements recorded within 8 days of the intervention. *Unique Clicks* counts distinct individuals who clicked the link at least once. *LR Total Clicks* refers to link clicks recorded approximately six months after the intervention. *Non-participant Share* indicates the percentage of Total Clicks originating from outside the Qualtrics platform. $\lambda_T$ and $\lambda_C$ denote the click-through rates, calculated as the number of clicks divided by the number of subjects in the Treatment group ($N = 448$) and Control group ($N = 232$), respectively. The *Poisson Test* column reports p-values from two-sided Poisson tests comparing the click-through rate ratio $\lambda_T/\lambda_C$ to 1 ($H_0$: T/C ratio = 1, or no difference between T and C). The first value in parentheses corresponds to the approximate test, and the second one corresponds to the exact test.

clicks over the 6 months after the intervention (Panel B of Table 8). First, we look at unique clicks based on unique IP addresses to alleviate the potential concern that observed clicks might stem from only a few individuals who might have clicked the link multiple times, rather than their peers who received link re-shares. Comparing *unique clicks* by the treatment group, we find a similar click-through rate ratio of about 2 (sharing frequency per respondent in the treatment group is roughly double that of the control group, 0.210/0.103), statistically different from the ratio being 1 ($p < 0.001$).

Next, we compare the shares of clicks generated outside the survey as a proxy for clicks by non-study participants via re-shares (as opposed to survey participants clicking on the link themselves). As we observe the source of the link-click (clicks from within Qualtrics vs. clicks outside Qualtrics), we find that links from the treated groups receive more clicks from outside the survey platform (82%) compared to the ones from the control group (60%), which suggests that treated students are sharing information with more presumably non-study students.

Finally, tracing clicks over the long run between our baseline survey experiment and our 6-month follow-up survey, we find that information sharing continued beyond the immediate survey completion as the long-run total clicks went up to 179 and control clicks to 58 (from 136 and 35 one week after the survey, respectively) by the 6-month follow-up. While the click-through rates

are starting to converge with the rate ratio decreasing from 2 to 1.6, we still rule out the equality of the click-through rates between the treatment and control groups at the 0.1% significance level.

**Peer Advice:** In Table 9, we show that in the short run, treated students are 3.8 percentage points more likely to mention on-campus services in their (hypothetical) advice to a friend in distress ($p$=0.06), constituting over a 50% increase over the control mean. At the same time, we find no effect on the probability of mentioning professional help more generally (*any* form of help). Interestingly, over 35.8% of the control students mention some form of professional help in their hypothetical advice.[37] Next, we compare this short-run result from the advice prompt to self-reported recommendations of suggesting therapy on- or off-campus to friends, which we collected in our follow-up survey. First, in columns (3)–(4) of Table 9, we replicate the estimates of the short-run effects on the sample of students matched from the follow-up round to the baseline participants and find that they do not differ substantially between the full ($N = 680$) and the follow-up samples ($N = 320$).

Table 9: Effects on Advice Prompt at Baseline & Recommending Therapy in 6 Months

| | SR: Advice Prompt (All) | | Advice Prompt (in Followup) | | LR: Suggested Therapy | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | On-Campus Help | Any Prof Help | On-Campus Help | Any Prof Help | ON campus | OFF campus |
| Treated | 0.038* | 0.013 | 0.039 | -0.002 | -0.019 | -0.007 |
| | (0.020) | (0.039) | (0.028) | (0.057) | (0.058) | (0.059) |
| Control Mean | 0.052 | 0.358 | 0.046 | 0.376 | 0.422 | 0.505 |
| Control SD | 0.22 | 0.48 | 0.21 | 0.49 | 0.50 | 0.50 |
| Observations | 680 | 680 | 320 | 320 | 320 | 320 |

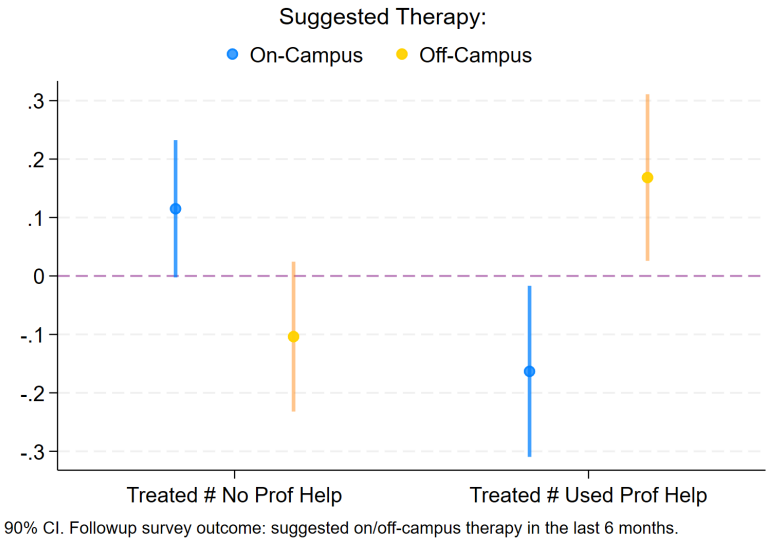
Notes: This table presents the effects of the information intervention on (i) mentions of professional help in a hypothetical advice prompt at baseline (columns 1–2), (ii) mentions in the advice prompt among the subsample tracked in the follow-up (columns 3–4), and (iii) self-reported recommendations to seek therapy on- or off-campus (columns 5–6). "On-Campus Help" refers to a specific mention of university counseling services, while "Any Prof Help" includes both on-campus and off-campus options. All outcomes are binary indicators. The treatment coefficient in column (1) represents a 3.8 percentage point increase in the likelihood of suggesting on-campus therapy relative to the control mean (5.2%), significant at the 10% level. Observations in columns 3–6 are limited to students who were matched to the follow-up survey. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SR = short-run; LR = long-run.

While the advice prompt format doesn't allow us to differentiate whether participants imply off-campus therapy specifically, we can disentangle these effects in our follow-up survey where we ask students whether they recommended on-campus and off-campus therapy services to their peers, capturing a self-reported behavioral outcome. Notably, it is different from a single-instance hypothetical advice provided in the initial survey and captures participant interactions with peers over time. On average, we do not find an effect, with coefficient estimates both for on- and off-campus recommendations close to zero and insignificant (columns (5)–(6) in Table 9). Yet, this null result masks key differential treatment effects by participants' prior use of therapy, as treated respondents

[37] In Table B10 we analyze different components of the advice. Namely, we analyze the frequency of mentioning empathetic and directive—proposing course of action, different from therapy—advice. In Table B11 we also show that the reduction in mentions of empathetic advice is compensated by increasing the probability of recommending on-campus therapy.

who had used therapy in the 12 months prior to the initial survey are 16.8 p.p. more likely to recommend off-campus therapy ($p = 0.05$), while those who had not are 11.5 p.p. more likely to recommend on-campus therapy ($p = 0.11$) (See Figure 9, point estimates in Table B9).

Figure 9: LR Therapy Recommendations to Peers




Notes: Coefficient plots show estimated treatment effects on binary outcomes measuring whether participants recommended on-campus (blue) or off-campus (yellow) therapy to peers in the past 6 months, conditional on prior therapy use. Estimates are from linear probability models with treatment indicators interacted with dummies for whether the participant reported any professional help use in the 12 months before baseline. The left group ("Treated # No Prof Help") includes treated students with no prior therapy use, while the right group includes those who did ("Treated # Used Prof Help"). Bars represent 90% confidence intervals. The outcome was measured in the follow-up survey. Heteroskedasticity-robust standard errors used.

We conjecture these effect are likely driven by the participants defaulting to recommending just one type of therapy, rather than both (as seen by the coefficients for on vs off-campus being exactly opposite for each subgroup), and for those who had not used professional help prior to the initial survey, the default recommendation option is on-campus therapy, which they learn more about as a result of the intervention.

Overall, we observe how peer interactions about sharing information and suggestions about therapy to other students are positively affected by our intervention in the short and long-run, with long-run effects shaped by participants' prior experience with therapy in a stronger way, suggesting that heterogeneity by prior experiences and beliefs might play an important role in engaging in promoting mental health services among peers (consistent with our evidence of prior beliefs being correlated with perceptions of public beliefs and behaviors discussed in Subsection 5.2). This result resonates with heterogeneity between prior therapy users and non-users in a field experiment with refugees where the treatment effects of information sharing vary by prior therapy use (Smith 2025).

Wider information sharing and mixed effects on recommendations by type are most likely attributed to updating students belief that therapy is not only for students with severe symptoms, making it more widely applicable for more peers in the eyes of the treated students. In addition, sharing information as a part of a widely publicized online study could also provide sufficient "social cover" to alleviate reputational concerns about promoting therapy.

## 6.2 Willingness to Pay for Therapy & Therapy Use

While peer interactions around therapy and counseling are up overall, we find a surprising result of a lower willingness to pay for therapy in the short run.[38]

Respondents from the treatment group show a *lower* WTP for therapy for both themselves and their friends by 3.6 p.p. ($p = 0.07$) and 3.3 p.p. ($p = 0.11$), respectively (See Figure 10a and Table B8 for point estimates, see Figure B15 for robustness of the result to outliers in the WTP measure). For donations, we find a small and insignificant effect with the treated students willing to donate 1.4 p.p. less than control students ($p = 0.45$).

One potential explanation for these puzzling negative treatment effects is that treated students may substitute away from private therapy to free on-campus services, as observed by a higher frequency of recommending them to a friend in distress in the advice prompt. Reminding students that therapy is effective and there are free on-campus resources as a part of the intervention might reduce their incentive to spend money on outside options, and thus, their willingness to pay. If this is the case, we should observe that the treated respondents use on-campus therapy more than off-campus options in our long-run follow-up.[39]

To test whether students substitute towards free on-campus therapy services, we compare the long-run treatment effects in students' self-reported use of on-campus vs. off-campus therapy from our follow-up survey. First, to support the comparability of the results on our follow-up sample, we show that the negative effects on WTP are comparable in the follow-up survey subsample: -3.6 p.p. ($p = 0.07$) for own WTP on the full sample of initial participants vs. -4.5 p.p. ($p = 0.14$) in the follow-up sample (See columns 3-4 in Table B8). Then, in Figure 10b, we provide evidence against the substitution effect explanation as we find that, if anything, treated students are 6.1 p.p. *more* likely ($p = 0.26$) to report using mental health services off-campus with a negligible insignificant effect on on-campus therapy use.

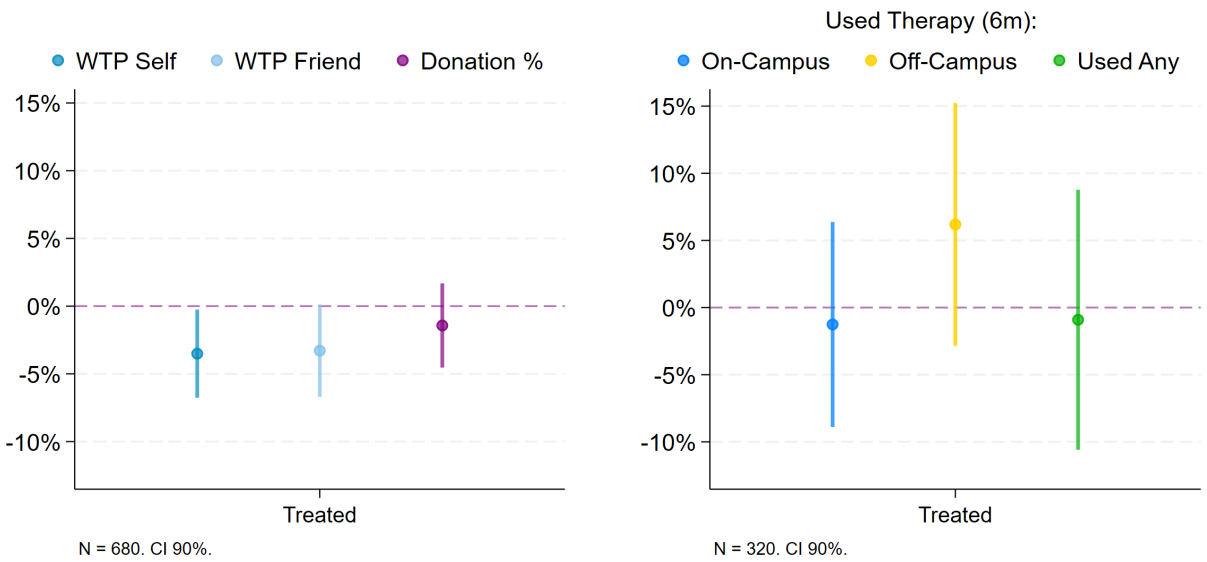
[38] In our pre-analysis plan, we hypothesized that information treatments would (i) increase students' demand for mental health support, measured by their WTP for therapy, (ii) increase perceived demand for therapy by others, measured by WTP for therapy for a friend and the share of their survey earnings they would be willing to donate to subsidize a therapy session for a fellow student.

[39] For our follow-up survey, we pre-registered this as a hypothesis and incorporated questions to specifically test it by comparing on- vs. off-campus therapy use.

Figure 10: Effects on Short-Run Willingness to Pay for Private Therapy and on Donations

(a) Short run: WTP and Donations

(b) Long run: Therapy use




Panel (a) shows point estimates and 90% confidence intervals of the treatment effects on short-run outcomes: willingness to pay (WTP) for private therapy for oneself, for a friend, and the share of survey endowment that they would donate to subsidize a financially constrained peer's therapy session. WTP outcomes are winsorized at the 1st and 99th percentiles. Results are robust to analyzing outcomes without winsorizing or using raw (level) values. We observe negative treatment effects on WTP for both self ($-3.6$ p.p., $p = 0.07$) and friend ($-3.3$ p.p., $p = 0.11$), and a small, statistically insignificant reduction in donation share ($-1.4$ p.p., $p = 0.45$). Panel (b) presents long-run treatment effects on binary indicators of self-reported therapy use (on-campus, off-campus, or any) in the follow-up survey. While the intervention may have prompted substitution toward on-campus services, we find no significant increase in on-campus use and a non-significant 6.1 p.p. increase in off-campus therapy use ($p = 0.26$), providing no support for the substitution mechanism. All estimates use heteroskedasticity-robust standard errors. Sample size: 680 (short run); 320 (long run).

In order to better understand where the negative treatment effects are coming from, we run quantile regressions on the 25th, 50th, 75th and 99th percentiles. Our results show negative effects on WTP for therapy for oneself for the three lower quartiles (significant for Q1 and Q3 with reductions of 7–6 p.p., respectively) and positive but insignificant at the top quartile. Similar results hold for WTP for therapy for a friend (See Figure B13). This suggests that negative effects are driven by "low demanders" who might be unlikely to seek therapy because their personal WTP is lower than the market price and potentially lower even than the opportunity costs associated with using free on-campus therapy. Below, we discuss several alternative interpretations of this result. While we are unable to test these explanations causally, we provide some exploratory analyses, pointing to several interesting directions for future studies.

First, it could be the case that we managed to reduce perceived stigma with our intervention (by showing that there is no correlation between GPA and mental distress scores, and a broader range of students seek therapy), yet WTP is lower because the perceived costs of being in distress are now lower. This interpretation is in line with a previous online experiment measuring the WTP

for private therapy and also finding a negative treatment effect from a de-stigmatizing intervention about public perceptions (Roth et al. 2024a). Roth et al. (2024a) propose an explanation for reduced demand for psychotherapy after lowering perceived stigma as individuals increase optimism about their social interactions, even when in distress, hence reducing the perceived need for therapy. In our setting, however, the lower WTP is not followed by a reduction in the reported use of therapy in the 6-month follow-up (Figure 10b), which is consistent with the quantile regression results pointing to "low demanders" who would not take up treatment in equilibrium driving the negative average treatment effect. This suggests that a perceived lower need for therapy is not likely to be the explanation in our lower-stigma setting. If anything, this highlights an avenue for further research to investigate reduced WTP for mental health support after information treatments and compare the experimental BDM-style WTP elicitation with observed or self-reported treatment take-up measures in the mental health context.

The puzzling results might raise questions about the validity of the BDM-style WTP questions in our study and their interpretability as truly indicating a lower demand for the service. We use a widely accepted experimental measure which has also been applied in other online experiments related to WTP for mental health services (Acampora et al. 2023; Roth et al. 2024a,b), so if implemented correctly, it should capture some meaningful variation. First, we implement it in a manner closely following the previous online experiments mentioned above, and we implement attention checks, excluding students with low attention to the survey from our study (See Subsection 4.1 for further details on the implemented attention checks). Second, we verify that the students in the treatment and control groups have no significant differences in the median completion time of the WTP questions.[40] We also document that WTP for therapy in the BDM-style question is higher for those actually using therapy, consistent with what we would expect based on revealed preferences (Figure B14).

Second, stigma might not have been reduced sufficiently to shift their own treatment-seeking behaviors, and we may have inadvertently reinforced some stereotypes related to commonly held misconceptions while trying to refute them. An insufficient reduction in stigma would explain why students are initially more in favor of on-campus services and share advice and information about them with friends in response to the information intervention. Yet, they resort to (if anything) using more off-campus services when it actually comes to seeking help themselves in the long run. Importantly, we also acknowledge that in aiming to reduce the incorrect beliefs, we highlight to students that the GPA-mental distress correlation is a common misconception, which, even if not true in data, still highlights a commonly held belief affecting their real-life interactions with peers

[40] We conduct a nonparametric equality-of-medians test for the time spent completing the WTP questions between the treatment and control students (using `medians` in STATA) and find a continuity-corrected Chi-squared $p$-value of 0.374 and 0.936 for WTP for oneself and for a friend, respectively, indicating no statistical differences in median completion time.

and others. This might explain the students' higher interest in on-campus services in the short run (at the end of the initial survey experiment) and lower WTP for a private external service in the experimental context, yet not translate into differential therapy take-up in their real-life social circles, where misconceptions persist.

Third, it could also be the case that by showing students that 2 out of 3 students who are receiving professional mental help have mild or no symptoms of depression and anxiety, in addition to showing them that GPA and mental health scores do not correlate, students might have reduced their valuation of therapy. They might have thought of therapy as a higher-value good for more severe cases requiring more trained professional providers, which should, thus, be priced higher, but updating their beliefs on who goes to therapy normalizes therapy in a way that they perceive it should not be as expensive.

Overall, we do not find one compelling explanation for the lower WTP without a negative effect on self-reported long-run therapy take-up in our study, but instead observe the potentially differential impact on either side of the underlying WTP distribution – "low demanders" experiencing a negative effect on WTP, but not affecting the equilibrium therapy take-up. Thus, comparing the average treatment effect captured by a single WTP measure to observed or self-reported therapy take-up across the WTP distribution warrants further research beyond our study. In this study, we show that the peer interactions are, thereby, easier to shift with a light-touch information intervention, encouraging students to share information about therapy services and giving advice to use it to friends, while students' own treatment-seeking is harder to shift, even when services are free and easily accessible.

## 6.3 Low Therapy Demanders & Evidence of Stigma

In this section, we explore the relevance of some of our puzzling results for different student populations and how the heterogeneity affects the interpretation and the policy takeaways from our intervention.

### Low Therapy Demanders

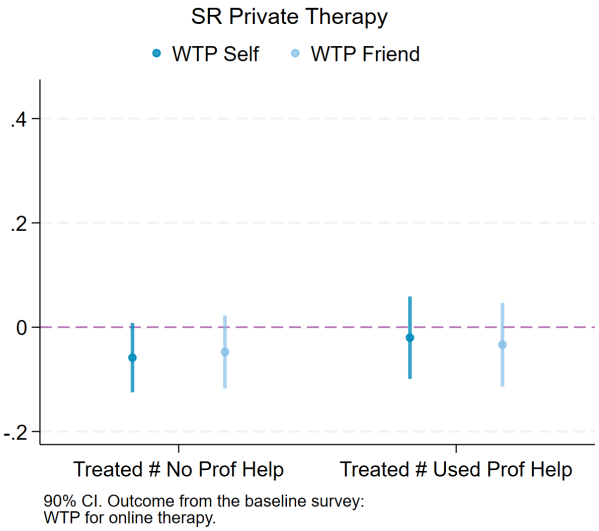
As we showed in Section 5, prior therapy use by students highly correlates with beliefs around mental health, and here we use prior use as a proxy for a student's revealed-preference valuation of therapy (higher if used previously vs. lower if not). In Figure 11a, we show suggestive evidence that negative treatment effects on WTP are driven by students with low valuation of this good (those who did not use therapy at baseline). In the short run, students who report no prior use of therapy decrease their WTP for therapy for themselves by 5.8 p.p. ($p = 0.15$) while for students with prior therapy use, the effects are closer to zero (estimate of -2 p.p., $p = 0.68$). In the long run, we similarly see that students with low valuation of therapy are less likely to use therapy while the

effect for those who used therapy before is strong and significant for off-campus therapy and weaker positive but insignificant for on-campus therapy (Figure 11b)[41].

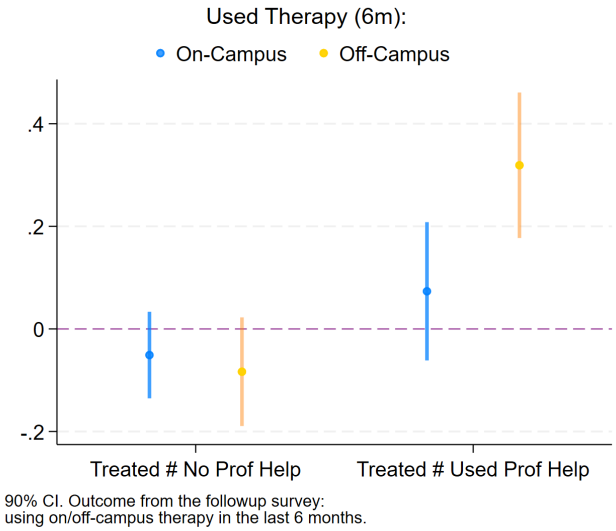
This suggests that our intervention worked as expected mostly among students who had prior exposure to professional mental health support at baseline, while the more puzzling unexpected results are driven by those with a lower baseline mental health treatment seeking. In our conceptual framework, this implies that the mechanisms we discuss in the theory of change might operate differently based on people's underlying beliefs, warranting further research into targeted correcting of misperceptions and stereotypes in different forms at different groups of interest.

Figure 11: Effects on Long-Run Using Therapy by Prior Use

(a) Effects on SR WTP by Prior Use




(b) LR Therapy Use




Notes: Coefficient plots show estimated coefficients with 90% confidence intervals from models interacting treatment status with prior therapy use (measured as self-reported use of professional mental health services in the 12 months before baseline). Panel (a) shows treatment effects on short-run willingness to pay (WTP) for private online therapy, for oneself and for a friend. WTP is measured as a percentage of the known subscription price and winsorized at the 1st and 99th percentiles. Panel (b) presents long-run treatment effects on binary indicators for self-reported use of on-campus and off-campus therapy. This panel reports effects separately for students who had previously used therapy and those who had not. Heteroskedasticity-robust standard errors used throughout. SR = short run; LR = long run.

Under the conceptual framework introduced in Table 7, we explore whether our intervention's effect was twofold. On the one hand, our intervention might have reduced *personal stigma* by showing there is no correlation between students' mental health and their GPA. On the other hand, however, we might have increased *perceived stigma*[42] by telling students that 3 in every 4 survey

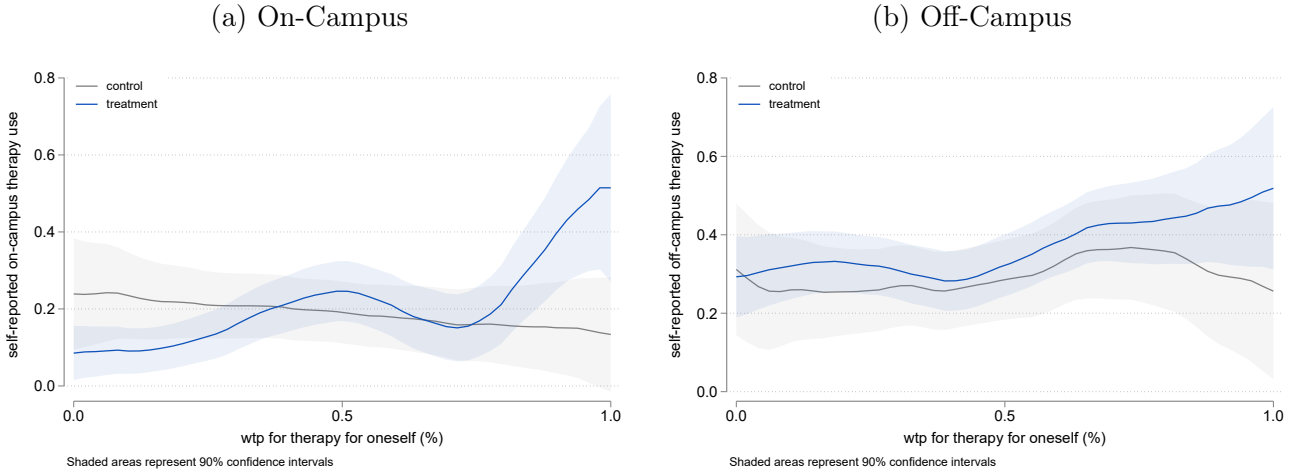
[41] Note that this is also consistent with the results on therapy recommendations by type in Figure 9

[42] Both perceived stigma about how the participant him- or herself will be viewed as well as more generally perceived public stigma of how students view other students.

respondents think a student with mental health issues performs worse than a student without mental health issues.[43]

In Figure 12, we show further suggestive evidence on the negative treatment effects coming from low demanders: we show that long-run on-campus therapy use is slightly lower among treated students than control students with *low* WTP and slightly higher for treated students than control students with *high* WTP, which results in an average zero-effect due to differences at either end of the distribution canceling out. Regarding off-campus services, we see no differences at the lower end of the distribution, but observe that among students with a higher baseline WTP, the self-reported use among treated students is higher than among comparable controls.

Figure 12: Self-reported Therapy Usage by WTP




Notes: This figure shows local polynomial estimates of long-run self-reported therapy usage (on- and off-campus) as a function of willingness to pay (WTP) for therapy for oneself, separately by treatment group; shows further suggestive evidence on the negative treatment effects coming from low demanders. Panel (a): we observe that long-run on-campus therapy use is slightly lower among treated students than control students with low WTP and slightly higher for treated students than control students with high WTP. These opposing effects cancel out on average, resulting in a net zero-effect. Panel (b) shows that regarding off-campus services, there are no differences at the lower end of the WTP distribution, but among students with a higher baseline WTP, self-reported use is higher in the treatment group than among comparable controls. Shaded areas represent 90% confidence intervals. SR = short run; LR = long run.

Finally, we note a methodological reflection on the relationship between WTP to measure potential demand for therapy and actual take-up of therapy by the participants. While eliciting willingness-to-pay (WTP) for therapy provides a useful incentive-compatible measure of potential demand, which can be implemented with corresponding study payouts and is widely used (Acampora et al. 2023; Lacey et al. 2024; Roth et al. 2024a,b), it may not perfectly predict actual therapy uptake as we observe in this study. We show that a negative treatment effect on the WTP reflects a shift among low-demanders, which might not be using therapy with or without the intervention,

[43] In another setting, for example, Arias et al. (2022) show that when voters have strong priors about politicians being malfeasant, providing them with information about malfeasance can actually *increase* malfeasant politicians' vote share as prior beliefs are further away from the truth.

yet contribute to a negative average effect. Few experiments to date have managed to track and identify sizable effects on actual therapy take-up following belief corrections. Thus, interpreting WTP gains (declines) requires caution: they may indicate latent interest (disinterest) among a particular subgroup within the WTP distribution, but future research should examine whether such interest translates into concrete help-seeking behavior, potentially by incorporating longer follow-ups or linking participants to services and monitoring enrollment.

**Revealing One's Own Mental Health State & Potential Stigma**

Our final set of results relates to personal and perceived stigma (recall Table 7): how each respondent views other students in distress and how the respondent might expect other students would view him or her if in distress. On personal stigma, we measured a ranking preference for working with students with different characteristics on a class project, allowing us to compare how respondents would rank a student with low grades vs. a student who shows symptoms of mental distress or a student who talks about mental health (these results are reported in columns (1)-(2) in Table 10). While we are underpowered (Subsection B.4) to detect a significant effect in the ranking questions, we see that the treated respondents are slightly more likely to rank a student with symptoms of mental distress above the one with a low GPA, which is in line with the information intervention ($I3$) informing subjects that distress and grades are uncorrelated. The effects are small overall (relative to the control mean), but are in the direction that we would have expected, which may signal slightly lower personal stigma (a participant's negative views of another student in distress).

Next, we turn to a measure of perceived stigma related to personal disclosure of own mental health problems, measured in the follow-up survey with self-reported questions on discussions of own mental health struggles when interacting with peers. As we showed previously, about 1 in 3 students in our sample at baseline report they would be disappointed in themselves if they had mental health issues (*'self-stigma'*) while they feel general support for going to therapy (Table 6). In our follow-up survey, to further explore another aspect of peer interactions around therapy, we asked the subjects whether in addition to giving recommendations to friends, they themselves had shared their mental health struggles with others at the university. While ex-ante we expected students to be more willing to engage in mental-health related discussions (similarly to sharing links and giving recommendations), in Table 10, we find that treated students became 7.9 p.p. ($p = 0.13$) *less* willing to talk about their own mental-health struggles (10% of the control mean) and 6.8 p.p. ($p = 0.22$) *less* willing to discuss their or their peers' experience with therapy use (18% of the control mean) — which may be related to perceived stigma.

While we can not precisely pinpoint the mechanism driving this result, we conjecture that our intervention might have reinforced perceived stigma (what students believe others think about them when in distress) as the infographic explicitly mentioned the existence of the misconception about

the negative correlation between distress and GPA across students prior to correcting it. This may have reinforced the misperception before ever correcting it, which is especially relevant to observing behaviors outside the survey experiment as respondents make decisions about sharing their struggles with others in an environment where others' beliefs were not updated (the vast majority of their peers are not survey participants) and are probably more aligned with the misperception we tried to correct rather than the truth.

Table 10: Effects on Personal & Public Stigma-Related Outcomes

| | SR: Prefer over Low-GPA Student | | LR: Discuss MH / Therapy | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | Distress Sympt | MH Talk | Own MH Issues | Therapy | Any MH |
| Treated | 0.043 | 0.026 | -0.079 | -0.068 | -0.087* |
| | (0.036) | (0.029) | (0.052) | (0.056) | (0.049) |
| Control Mean | 0.703 | 0.845 | 0.761 | 0.376 | 0.807 |
| Control SD | 0.46 | 0.36 | 0.43 | 0.49 | 0.40 |
| Observations | 680 | 680 | 320 | 320 | 320 |

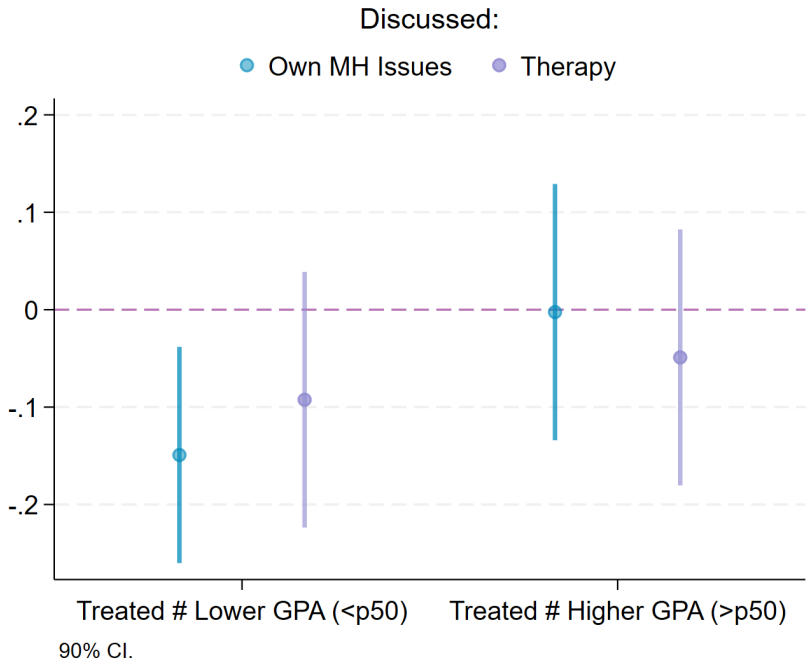
Notes: This table presents coefficient estimates examining short-run and long-run effects of the intervention on stigma-related outcomes. Columns (1)–(2) report short-run outcomes from the baseline survey experiment. Column (1) captures willingness to work with a peer exhibiting visible distress symptoms, relative to a peer with a low GPA. Column (2) measures self-reported comfort with talking about mental health more generally, relative to a peer with a low GPA. Columns (3)–(5) report long-run effects from the follow-up survey. These include whether participants report having discussed their own mental health struggles (column 3), therapy (column 4), or either topic (column 5) in the followup survey. We find that treated students became 7.9 percentage points less likely ($p = 0.13$) to report talking about their own mental health and 6.8 percentage points less likely ($p = 0.22$) to discuss therapy—corresponding to 10% and 18% of the control means, respectively. These patterns are potentially consistent with an increase in perceived public stigma. SR = short run; LR = long run. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In a setting in which students probably have a close-enough proxy of their friends' academic performance, the magnitude of the reinforcement should be given a larger weight by students with low academic performance rather than by those with high academic performance.[44] If lower-GPA students infer that many peers still hold the stereotype, the perceived social cost of admitting distress may rise. We present evidence consistent with this explanation in Figure 13. Treated students with below median GPA are 14.9 percentage points ($p = 0.03$) less likely to discuss their own mental health issues, whereas those with above median GPA are not more nor less likely to discuss their own mental health issues, relative to those in control (-0.2 p.p. estimate; $p = 0.98$). Regarding the discussion of their or their peers' experience with on-campus therapy, students with below median GPA are 9.2 percentage points ($p = 0.25$) less likely to discuss it, while students with above median GPA are 4.9 percentage points ($p = 0.54$) less likely to discuss it.[45]

[44]If students have low academic performance and their friends are aware of that, disclosing that they have mental health issues could potentially worsen their friends' perception of them. If students have high academic performance and their friends are aware of that, then disclosing mental health issues could potentially worsen their friends' perception of them, but given that the reference point is higher, the effect's magnitude is relatively smaller.

[45]Other pre-registered heterogeneity can be found in Appendix Subsection B.9.

Figure 13: TE Heterogeneity on Discussing MH Issues by GPA




Notes: This figure presents treatment effect heterogeneity on long-run mental health (MH) communication outcomes, by baseline academic performance. Estimates show coefficients with 90% confidence intervals from interacting treatment status with a binary indicator for below-median GPA. Outcomes include willingness to talk about one's own mental-health struggles (blue) and willingness to discuss one's own or peers' experience with therapy use (purple). Treated students with below-median GPA are 14.9 percentage points less willing to talk about their own mental-health struggles ($p = 0.03$), while the effect for higher-GPA students is negligible ($-0.2$ p.p.; $p = 0.98$). For therapy-related discussions, lower-GPA students are 9.2 p.p. less willing to engage ($p = 0.25$), compared to a smaller and non-significant 4.9 p.p. decrease among higher-GPA students ($p = 0.54$).

This highlights the importance of *how* we convey information in settings where generalized perceptions and truthful facts differ substantially, and potentially heterogeneously by underlying population characteristics. When addressing issues in which agent's decisions depend on other people's beliefs, it may not be sufficient to update the agent's own belief about objective facts, but also what the respondent anticipates others would think or do in response to his or her actions and choices (second-order beliefs). When designing a belief correction intervention for to correct interpersonal misconceptions, one might tell the participants that a misconception exists and/or how prevalent it is to make the corrective intervention more appealing and memorable, yet, this may reinforce the incorrect beliefs and shape how participants anticipate to be treated by others outside of the experiment.

Our heterogeneity exploration suggests that students may interpret the same belief-correcting facts through different lenses based on prior experience and beliefs. Those who had already experienced therapy at baseline integrate the new information in line with the proposed theory of change, while "low-demand" students update in ways that further lower their stated valuation of paid ser-

vices and constrain personal disclosure. Our findings echo recent work showing that belief-correction campaigns can have unintended effects when they prime interpersonal *second-order* beliefs about how others may view them in light of the presented information (Bursztyn & Yang 2022). In other words, correcting a stereotype may have the effect opposite to the intended de-stigmatization if the message first reminds recipients that the stereotype exists and/or is widely held. Future interventions on stigma-sensitive topics should therefore test how the framing of belief correction affects the effectiveness of updating with or without highlighting the prevalence of the misconception itself.

Taken together, our evidence reinforces a pattern that emerges across recent work. Fact–based, first–order belief corrections, such as updating beliefs about therapy effectiveness or typical therapy-goer profiles, successfully shift low-cost behaviors (e.g., sharing resource links or recommending services), yet leave higher-stakes actions such as self-disclosure or starting therapy unchanged overall, with stronger effects for subpopulation that were less stigmatized and/or were already using therapy at baseline (Smith 2025; Acampora et al. 2023; Roth et al. 2024b). In other settings, large and more durable changes arise when interventions target second-order misperceptions: in Indian slums, updating respondent beliefs about much higher neighbors' openness to discussing financial and mental-health stress than the majority believed increased sign-ups and contributions for savings and listening groups (Jain & Khandelwal 2024), while in Saudi Arabia, correcting men's beliefs about peer support for women's work boosted spousal job search and raised wives' labor-market activity by 4–5 p.p. after one year (Bursztyn et al. 2020). These patterns suggest that mental-health programs may need to combine credible facts *and* visible signals of peer acceptance, along with incorporate follow-up reinforcement or sustained engagement (Dhar et al. 2022), to move higher-cost, potentially more stigmatized behaviors that are harder to change with information alone, such as therapy take-up by those who have not used or considered therapy and personal disclosure of emotional distress.

# 7 Discussion

Our findings highlight the importance of misperceptions and stigma as contributors to the mental health treatment gap among students, particularly in settings where financial and structural barriers are minimal, such as when therapy is free and generally viewed as effective. While previous research has highlighted attitudinal barriers such as low perceived need or skepticism about treatment among adults, we show that students' misconceptions about who seeks therapy and how distress relates to academic performance may also contribute to underutilization of support services. The belief that therapy is only for those in severe crisis, and the perception that psychological distress is strongly associated with poor academic outcomes, may discourage students from engaging with available support. By correcting these misperceptions, our light-touch intervention increases students' willingness to share campus mental health resources and offer more proactive support to peers. While

it lowers individual willingness to pay for private therapy in the short run (immediately after the intervention), we observe no long-term reduction in therapy use six months later. In fact, among students who were already engaged in therapy at baseline, we find stronger positive effects on both off-campus therapy use and therapy recommendations.

These findings have important implications for mental health policy in university settings and beyond, particularly in developing countries where mental health stigma remains high. Our results suggest that addressing psychological frictions through belief correction can be a cost-effective way to improve engagement with available resources, especially by encouraging students to support their peers in seeking help. However, presenting students with the existing misconceptions, even while correcting them, could have lasting effects, as we find that personal disclosure of mental health problems is lower among treated subjects six months after the intervention. This highlights a dimension of belief correction that is often overlooked in information interventions: while the facts could reduce stigma and increase awareness, presenting information about the existence and prevalence of misconceptions may shift behaviors and beliefs in unintended ways.

More broadly, our results contribute to the literature on behavioral barriers to human capital investment and treatment-seeking in health-related settings. They align with recent work on how cognitive frictions influence decisions in education, labor, and health domains (Schilbach et al. 2016; Rao et al. 2021), as well as with research on the potential and limitations of correcting misperceptions to reduce stigma and shift behavioral outcomes (Bursztyn & Yang 2022). Future research could explore whether similar interventions are effective in increasing treatment take-up in contexts where financial and logistical barriers are more salient than in our setting, and whether belief correction can lead to longer-term changes in mental health norms and behaviors. As universities and policymakers expand mental health services, understanding the mechanisms that drive help-seeking decisions will be essential for designing interventions that meaningfully reduce the treatment gap.

Our mixed evidence, which shows strong gains in peer sharing but limited effects on therapy take-up and self-disclosure, suggests that correcting *factual beliefs* through information interventions may be necessary but insufficient for shifting more effortful or stigma-sensitive behaviors. This pattern is consistent with prior experimental research documenting modest effects of first-order belief corrections on mental health treatment-seeking (Acampora et al. 2023; Smith 2025; Roth et al. 2024b). These muted effects could potentially be amplified by targeting and updating second-order beliefs, particularly in cases where individuals underestimate the true norms in their communities. This approach has shown promise in other field settings (Jain & Khandelwal 2024; Bursztyn et al. 2020). For example, when slum residents in India learned that a strong majority of their neighbors were actually open to discussing financial and mental health concerns, they became substantially more likely to sign up for neighborhood savings circles and volunteer listening programs, and also contributed more to support these initiatives (Jain & Khandelwal 2024). In Saudi Arabia, men

were more likely to register their wives on a job platform, and the women were subsequently more likely to seek employment opportunities, following updated beliefs about prevailing social norms regarding women working (Bursztyn et al. 2020). Exploring such misperceptions in the context of mental health may offer a promising direction for future research. This is especially relevant given that mental health treatments like cognitive behavioral therapy (CBT) often aim to shift personal narratives and beliefs about oneself and others—the "cognitive" component of CBT—highlighting the potential of correcting interpersonal misperceptions as a mechanism for behavior change in this domain.

# References


Abrams, Z. (2022). Student mental health is in crisis. Campuses are rethinking their approach. *APA Monitor in Psychology*, 53(7), 60.

Acampora, M., Capozza, F., & Moghani, V. (2023). Mental Health Literacy, Beliefs and Demand for Mental Health Support. *SSRN Electronic Journal*.

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J. E., Al-Hamzawi, A., Borges, G., Bromet, E., Bruffaerts, R., de Girolamo, G., de Graaf, R., Florescu, S., Gureje, O., Hinkov, H. R., Hu, C., Huang, Y., Hwang, I., Jin, R., Karam, E. G., Kovess-Masfety, V., Levinson, D., Matschinger, H., O'Neill, S., Posada-Villa, J., Sagar, R., Sampson, N. A., Sasu, C., Stein, D. J., Takeshima, T., Viana, M. C., Xavier, M., & Kessler, R. C. (2014). Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine*, 44(6), 1303–17.

Arias, E., Larreguy, H., Marshall, J., & Querubín, P. (2022). Priors Rule: When Do Malfeasance Revelations Help Or Hurt Incumbent Parties? *Journal of the European Economic Association*, 20(4), 1433–1477.

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., Hwang, I., Kessler, R. C., Liu, H., Mortier, P., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., Benjet, C., Caldas-De-Almeida, J. M., Demyttenaere, K., Florescu, S., De Girolamo, G., Gureje, O., Haro, J. M., Karam, E. G., Kiejna, A., Kovess-Masfety, V., Lee, S., McGrath, J. J., O'Neill, S., Pennell, B. E., Scott, K., Ten Have, M., Torres, Y., Zaslavsky, A. M., Zarkov, Z., & Bruffaerts, R. (2016). Mental disorders among college students in the World Health Organization World Mental Health Surveys. *Psychological Medicine*, 46(14), 2955–2970.

Barker, N., Bryan, G., Karlan, D., Ofori-Atta, A., & Udry, C. (2022). Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress. *American*

*Economic Review: Insights*, 4(4), 527–45.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, 9(3).

Belloni, A., Chernozhukov, V., & Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650.

Benjet, C., Borges, G., Medina-Mora, M. E., Zambrano, J., & Aguilar-Gaxiola, S. (2009). Youth mental health in a populous city of the developing world: Results from the mexican adolescent mental health survey. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 50(4), 386–395.

Bhat, B., de Quidt, J., Haushofer, J., Patel, V., Rao, G., Schilbach, F., & Vautrey, P.-L. (2022). The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes. *SSRN Electronic Journal*.

Bolotnyy, V., Basilico, M., & Barreira, P. (2022). Graduate Student Mental Health: Lessons from American Economics Departments. *Journal of Economic Literature*, 60(4), 1188–1222.

Bose, I., Bethancourt, H. J., Shamah-Levy, T., Mundo-Rosas, V., Muñoz-Espinosa, A., Ginsberg, T., Kadiyala, S., Frongillo, E. A., Gaitán-Rossi, P., & Young, S. L. (2024). Mental health, water, and food: Relationships between water and food insecurity and probable depression amongst adults in Mexico.

Brewer, K. B., Gibson, R., Tomar, N., Washburn, M., Giraldo-Santiago, N., Hostos-Torres, L. R., & Gearing, R. E. (2023). Why Culture and Context Matters: Examining Differences in Mental Health Stigma and Social Distance Between Latino Individuals in the United States and Mexico. *Journal of Immigrant and Minority Health*.

Brody, D. J., Pratt, L. A., & Hughes, J. P. (2018). *Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016 Key findings Data from the National Health and Nutrition Examination Survey*. Technical Report 303.

Buchmann, N., Meyer, C., Sullivan, C. D., Arrieta, G., Bakhtin, M., Barnes, K., Bergstrom, K., Blattner, A., Cefala, L., Chandrasekhar, A., Ignacio Cuesta, J., De la, R. O., Dobbin, C., Freitas-Groff, Z., Einav, L., Gentzkow, M., Jalota, S., Jiménez, D., Juarez, L., Kessler, J., Layvant, M., Morten, M., Nielsen, K., Otero, S., Wittwer, M., Zohar, T., Ayan, S., Dubey, U., Al Hasib, M., Mahatab, T., Marchal, A., & Qiao, Z. (2024). *Paternalistic Discrimination **. Technical report.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review*, 110(10), 2997–3029.

Bursztyn, L. & Yang, D. Y. (2022). Misperceptions about Others. *Annual Review of Economics*, 14, 425–452.

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: A global return on investment analysis. *The Lancet Psychiatry*, 3(5).

Cornaglia, F., Crivellaro, E., & McNally, S. (2015). Mental health and education decisions. *Labour Economics*, 33, 1–12.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Canadian Journal of Psychiatry*, 58(7), 376–85.

Cuijpers, P., Cristea, I. A., Ebert, D. D., Koot, H. M., Auerbach, R. P., Bruffaerts, R., & Kessler, R. C. (2016). Psychological Treatment of Depression in College Students: A Metaanalysis. *Depression and Anxiety*, 33(5), 400–414.

Dhar, D., Jain, T., Jayachandran, S., Bangia, S., Bijker, M., Chandrasekhar, S., Chowdhuri, R. N., Favela, A., Gosselin, J., Kapoor, V., Kapur, V., Kim, L., Kovvuri, A., Mathur, S., Oh, S., Sarda, P., Seth, A., Shrestha, N., Singh, A., & Vig, R. (2022). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, 112(3), 899–927.

Eisenberg, D., Lipson, S. K., Heinze, J., & Zhou, S. (2022). *Healthy Minds Study – US National Report 2021-2022*. Technical report.

Fletcher, J. M. (2008). Adolescent depression: Diagnosis, treatment, and educational attainment. *Health Economics*, 17(11), 1215–1235.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2).

Haushofer, J., Mudida, R., & Shapiro, J. P. (2021). The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being. *SSRN Electronic Journal*.

Healthy Minds Survey (2022). Data for Research - Healthy Minds Network.

Jaadi, Z. & Whitfield, B. (2024). Principal Component Analysis (PCA): A Step-by-Step Explanation.

Jain, R. & Khandelwal, V. (2024). Silent Networks : The Role of Inaccurate Beliefs. *JMP*.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9).

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics*, 50(6).

Lacey, L., Mishra, N., Mukherjee, P., Prakash, N., Prakash, N., Quinn, D., Sabarwal, S., & Saraswat, D. (2024). Can destigmatizing mental health increase willingness to seek help? Experimental evidence from Nepal. *Journal of Policy Analysis and Management*, 44(1), 97–124.

Lagunes-Cordoba, E., Davalos, A., Fresan-Orellana, A., Jarrett, M., Gonzalez-Olvera, J., Thornicroft, G., & Henderson, C. (2021). Mental Health Service Users' Perceptions of Stigma, From the General Population and From Mental Health Professionals in Mexico: A Qualitative Study. *Community Mental Health Journal*, 57(5), 985–993.

Mascayano, F., Tapia, T., Schilling, S., Alvarado, R., Tapia, E., Lips, W., & Yang, L. H. (2016). Stigma toward mental illness in Latin America and the Caribbean: a systematic review. *Brazilian Journal of Psychiatry*, 38(1), 73–85.

Matavelli, I. (2025). We Don't Talk About Boys: Masculinity Norms Among Adolescents in Brazil. *JMP Working Paper*.

Mortier, P., Cuijpers, P., Kiekens, G., Auerbach, R. P., Demyttenaere, K., Green, J. G., Kessler, R. C., Nock, M. K., & Bruffaerts, R. (2018). The prevalence of suicidal thoughts and behaviours among college students: a meta-analysis. *Psychological medicine*, 48(4), 554–565.

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., Bhandari, P. M., Neupane, D., Brehaut, E., Benedetti, A., & Thombs, B. D. (2021). Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis.

OECD Report (2022). *Tackling the mental health impact of the COVID-19 crisis: An integrated, whole-of-society response - OECD*. Technical report.

Osman, N., Michel, C., Schimmelmann, B. G., Schilbach, L., Meisenzahl, E., & Schultze-Lutter, F. (2022). Influence of mental health literacy on help-seeking behaviour for mental health problems in the Swiss young adult community: a cohort and longitudinal case–control study. *European Archives of Psychiatry and Clinical Neuroscience*.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B. K., De Silva, M., Singh, I., Stein, D. J., Sunkel, C., & UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development.

Patel, V., Weobong, B., Weiss, H. A., Anand, A., Bhat, B., Katti, B., Dimidjian, S., Araya, R., Hollon, S. D., King, M., Vijayakumar, L., Park, A. L., McDaid, D., Wilson, T., Velleman, R., Kirkwood, B. R., & Fairburn, C. G. (2017). The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *The Lancet*, 389(10065), 176–185.

Rao, G., Redline, S., Schilbach, F., Schofield, H., & Toma, M. (2021). Informing sleep policy through field experiments. *Science*, 374(6567), 530–533.

Ridley, M. (2022). Essays on the economics of health and education.

Ridley, M. (2025). Mental Illness Discrimination. *Working Paper (Former JPM)*.

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, 370(6522).

Roth, C., Schwardmann, P., & Tripodi, E. (2024a). Depression Stigma. *CESifo Working Paper no. 11012*.

Roth, C., Schwardmann, P., & Tripodi, E. (2024b). Misperceived Effectiveness and the Demand for Psychotherapy Misperceived Effectiveness and the Demand for Psychotherapy *. *Journal of Public Economics*.

Sæther, M. H., Sivertsen, B., & Bjerkeset, O. (2021). Mental Distress, Help Seeking, and Use of Health Services Among University Students. The SHoT-Study 2018, Norway. *Frontiers in Psychiatry*, 12, 727237.

Salud Mental (2022). Que hay detras de los suicidios de estudiantes de Medicina? Esto explica la UNAM – El Financiero. *El Financiero (Mexico)*.

Schilbach, F., Schofield, H., & Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review*, 106(5), 435–440.

Schnyder, N., Panczak, R., Groth, N., & Schultze-Lutter, F. (2017). Association between mental health-related stigma and active help-seeking: Systematic review and meta-analysis. *British Journal of Psychiatry*, 210(4), 261–268.

Shreekumar, A. & Vautrey, P.-L. (2023). Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior. *Working Paper*.

Smith, E. C. (2025). Stigma and Social Cover: A Mental Health Care Experiment in Refugee Networks. *JMP Working Paper*.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10).

Thapar, A., Eyre, O., Patel, V., & Brent, D. (2022). Depression in young people. *The Lancet*, 400(10352), 617–631.

Velazquez Hernandez, A. (2017). Suicidio y depresion en estudiantes y residentes de Medicina en Mexico. *Salud Mental, Vice.com*.

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., Bruffaerts, R., de Girolamo, G., de Graaf, R., Gureje, O., Haro, J. M., Karam, E. G., Kessler, R. C., Kovess, V., Lane, M. C., Lee, S., Levinson, D., Ono, Y., Petukhova, M., Posada-Villa, J., Seedat, S., & Wells, J. E. (2007). Worldwide Use of Mental Health Services for Anxiety, Mood, and Substance Disorders: Results from 17 Countries in the WHO World Mental Health (WMH) Surveys. *Lancet*, 370(9590), 841.

WHO, T. W. H. O. (2021). Mental health of adolescents.

World Health Organization (2021). Depression.

# A Appendix: Description of Outcomes & Covariates

In this appendix we elaborate on the variables used in the analysis.

**Outcome Variables**

1. **On-Campus Counseling Link Sharing**.
   **Survey question used**: If you know that your university offers counseling services to support students' mental health? We encourage you to save and share this link to the university's counseling website with friends who might benefit from it. Spreading the word can help ensure that more students are aware of and can access these valuable resources! You can share the link directly or take a screenshot of this page and send it to your friends, we encourage you to do so. Here is the link for your convenience:

   **Variable Construction**: At the end of the survey, students were given an opportunity to share a link to on-campus counseling services with their peers.[46] We tracked both the total number of human clicks and the number of unique users who clicked the link across three experimental conditions. In addition, we observe the share of clicks directly from within the survey platform (Qualtrics, presumably clicked by respondents themselves), as well as those clicked via re-shares such as emails or SMS. We are not able to distinguish between few respondents sharing in bulk vis-à-vis many respondents sharing with few other people.

2. **Peer Advice**.
   **Survey question used**: Imagine a friend approaches you for emotional support because they are struggling with a personal or academic issue. How would you support them? What would you tell them? Take a moment to provide a thoughtful response that could genuinely help someone, which can earn you a bonus of 50 MXN. A fellow student will read your (anonymous) advice and rate it as 'Very Useful', 'Somewhat Useful' or 'Not Useful'. Responses rated as 'Very Useful' will earn a bonus of MXN 50. (One of the bonus questions will be randomly chosen for payment)

   **Variable Construction**: Participants were asked to imagine a scenario where a friend approaches them for emotional support due to personal struggles. They were then prompted to provide open-ended advice, which was evaluated by the length of the advice given (in words) and by whether respondents mention words such as 'therapy', 'support you', 'empathy', among others, on their response.

[46] See Appendix Figure B8 for the infographic containing the QR code and URL linking to the university counseling services center.

3. **Willingness to Pay for Therapy for Self**.
   **Survey question used**: In this question, we will ask you about the maximum amount you are willing to spend on 4 weeks of therapy with BetterHelp (1 session per week, 4 sessions total). Note that this service is usually priced at around 6,500 MXN for 4 weeks. We will select a few participants randomly and implement their choices - it could be you! What is your valuation, i.e. the maximum amount of money you would pay for 4 weeks of therapy from BetterHelp (1 session per week, 4 sessions total)? A computer will bid against you. The computer's bid will be a random number between 0 MXN and 7000 MXN independent of your answer. If your valuation is higher than the computer's bid, you will receive four weeks of therapy from BetterHelp for free. If your valuation is lower than the computer's bid, you will receive an amount worth the computer's bid that will be added to your Amazon gift card balance Regardless of the computer's bid, it is always in your best interest to report your true personal valuation!

   **Variable Construction**: As a proxy for participants' demand for therapy, we use incentive-compatible BDM-style willingness to pay (WTP) measures (Becker et al. 1964). Specifically, we measured the maximum amount participants were willing to pay for a one-month therapy subscription from *BetterHelp*, for themselves.

4. **Willingness to Pay for Therapy for a Friend**.
   **Survey question used**: Now, we will ask you about the maximum amount you are willing to spend on 4 weeks of therapy with BetterHelp (1 session per week, 4 sessions total) for YOUR FRIEND . Note that this service is usually priced at around 6,500 MXN for 4 weeks. We will select a few participants randomly and implement their choices - it could be you! What is your valuation, i.e. the maximum amount of money you would pay for 4 weeks of therapy from BetterHelp (1 session per week, 4 sessions total) for YOUR FRIEND? A computer will bid against you. The computer's bid will be a random number between 0 MXN and 7000 MXN independent of your answer. If your valuation is higher than the computer's bid, YOUR FRIEND will receive four weeks of therapy from BetterHelp for free If your valuation is lower than the computer's bid, YOUR FRIEND will receive an amount worth the computer's bid that will be added to their Amazon gift card balance Regardless of the computer's bid, it is always in your best interest to report your true valuation of therapy for your friend!

   **Variable Construction**: As a proxy for participants' demand for therapy, we use incentive-compatible BDM-style willingness to pay (WTP) measures (Becker et al. 1964). Specifically, we measured the maximum amount participants were willing to pay for a one-month therapy subscription from *BetterHelp*, for a friend.

5. **Donation**.

   **Survey question used**:Out of the payment you receive from participating in this survey, what percentage (%) would you like to donate to help cover the cost of a therapy session for a university student who reported that financial constraints prevent them from seeking therapy? We will automatically deduct your donation from your payment, directing it toward a funded therapy session for this student.

   **Variable Construction**: Participants were asked about the share of their earnings from participating in the study they were willing to donate to help fund a therapy session for a financially constrained student at their university.[47] Participants were notified that any donation they pledged would be automatically deducted from their payment and allocated toward this funded therapy session.

6. **Ranking questions**.

   **Survey question used**:Rank the following individuals in terms of how comfortable you would feel working closely with them on a joint course project: 1 = Most comfortable to work with 6 = Least comfortable to work with. You can drag and drop the options below. - A student who often skips classes.
   - A student who shows symptoms of anxiety or depression.
   - A student who makes inappropriate comments.
   - A student who often arrives late and leaves early.
   - A student who talks about their current mental health struggles.
   - A student who is not performing well academically.

   **Variable Construction**: We asked participants to rank individuals in terms of how comfortable they would be working with them on a joint course project. We describe six hypothetical students with different traits, all of which might make it undesirable to work with a particular student. Specifically, we assess whether respondents deem it more undesirable to work with a low GPA student relative to with a student who talks about mental health issues or shows signs of having them.

7. **Therapy Use (long term)**.

   **Survey question used**: Respond about your experience in the last six months
   - I've utilized professional mental health services (like therapy) ON CAMPUS.
   - I've utilized professional mental health services (like therapy) OFF CAMPUS.

[47] Specifically, they were informed that their donations would be directed toward covering the cost of 1 therapy session for a fellow university student who reported that financial constraints prevent them from seeking therapy.

**Variable Construction**: In the follow-up survey, we asked students whether they had used professional therapy or psychological counseling in the past six months. We ask one question for on-campus services and another one for off-campus services.

8. **Recommendations (long term)**.
   **Survey question used**: Respond about your experience in the last six months:
   - I've recommended professional mental health services (like therapy) ON CAMPUS to my peers.
   - I've recommended professional mental health services (like therapy) OFF CAMPUS to my peers.

   **Variable Construction**: In the follow-up survey, we asked students whether they recommended professional therapy services to their peers. Again, we ask for both on- and off-campus services explicitly.

9. **Willingness to share/discuss issues/therapy use (long term)**.
   **Survey questions used**:
   Respond about your experience in the last six months:
   - I've talked about my mental health problems with other students.
   - I've talked about my own or other students' experiences with ON-CAMPUS therapy or psychological counseling.

   Respond to a hypothetical situation:
   - If you had a problem, would you consider attending ON-CAMPUS therapy or psychological counseling?
   - If you had a problem, would you consider attending OFF-CAMPUS therapy or psychological counseling?

   **Variable Construction**: We also ask students whether they have talked about their mental health problems with other University students, and whether they have talked about their or their University peers' experience with on-campus therapy or psychological counseling

**Covariates: Mental Health Care Measures and Elicited Beliefs**

1. **Mental distress**.
   **Survey question used**: Over the last two weeks, how often have you been bothered by any of the following problems? (1. Not at all, 2. Several Days, 3. More than half the days, 4. Nearly Every Day)
   1. Little interest or pleasure in doing things.

2. Feeling down, depressed or hopeless.
3. Feeling tired or having little energy.
4. Feeling bad about yourself – or that you are a failure or have let yourself or your family down).
5. Worrying too much about different things.
6. Becoming easily annoyed or irritable.
7. Being so restless that it is hard to sit still.
8. Feeling nervous, anxious or on edge.

**Variable Construction**: We compute a mental distress index using the PHQ-4 and GAD-4 screening questionnaires for depression and anxiety, respectively (Kroenke et al. 2001; Spitzer et al. 2006). Each question has four possible responses with values ranging from 0–4; we compute the index by summing over values across questions. Larger values imply worse mental distress and the index's support is [0, 24]. As is common practice in the health sector (Kroenke et al. 2009), we classify students as being in distress if their mental distress index is greater than or equal to the index support's midpoint of 12.

2. **Mental health care use & perceived therapy use**.
**Survey question used**:
Have you used professional mental health help in the last 12 months? (Yes, No)
Out of every 100 students at your university, how many do you think have used professional mental health help in the last 12 months? If your guess is within 5 students of the correct answer, you will earn a bonus of 50 MXN. (The correct answer will be calculated based on responses from this survey. One of the bonus questions will be randomly chosen for payment.)

**Variable Construction**: We ask students whether they have/have not used professional mental health help in the last 12 months. Additionally, we asked them to guess out of every 100 University students, how many of them did they think have used professional mental health help in the last 12 months.

3. **Perceived therapy effectiveness**.
**Survey question used**:
Researchers have conducted many clinical studies to estimate the effectiveness of psychotherapy for treating depression. A comprehensive review looked at the 22 studies with the largest number of participants. Out of these 22 studies, how many do you think show that therapy is an effective treatment for depression? If your guess is correct, you will earn a bonus of 50

MXN. (One of the bonus questions will be randomly chosen for payment.)

How much do you agree or disagree with the following statements? ( 1. Strongly disagree, 2. Disagree, 3. Somewhat disagree, 4.Somewhat agree, 5. Agree, 6. Strongly Agree ).

1. Going to therapy can improve my own mental well-being substantially.
2. In general, going to therapy can improve people's mental well-being.
3. My friends would show support if I told them I am going to therapy.
4. My parents would show support if I told them I am going to therapy.

**Variable Construction**: We tell students that a review of 22 studies examining the effectiveness of psychotherapy for treating depression was conducted. We then ask them how many studies they think show that therapy is an effective treatment for depression out of the 22 analyzed. Additionally, we ask them a Likert-style question to measure the extent to which they believe therapy can improve their own (people's) mental wellbeing.

4. **Self-stigma**.

**Survey question used**:

How much do you agree with the following statement? "I would feel disappointed in myself if I had a mental health issue (e.g., anxiety or depression)" (1. Strongly agree ) ( 2. Agree ) (3. Somewhat agree ) (4. Somewhat disagree ) (5. Disagree ) (6. Strongly disagree ).

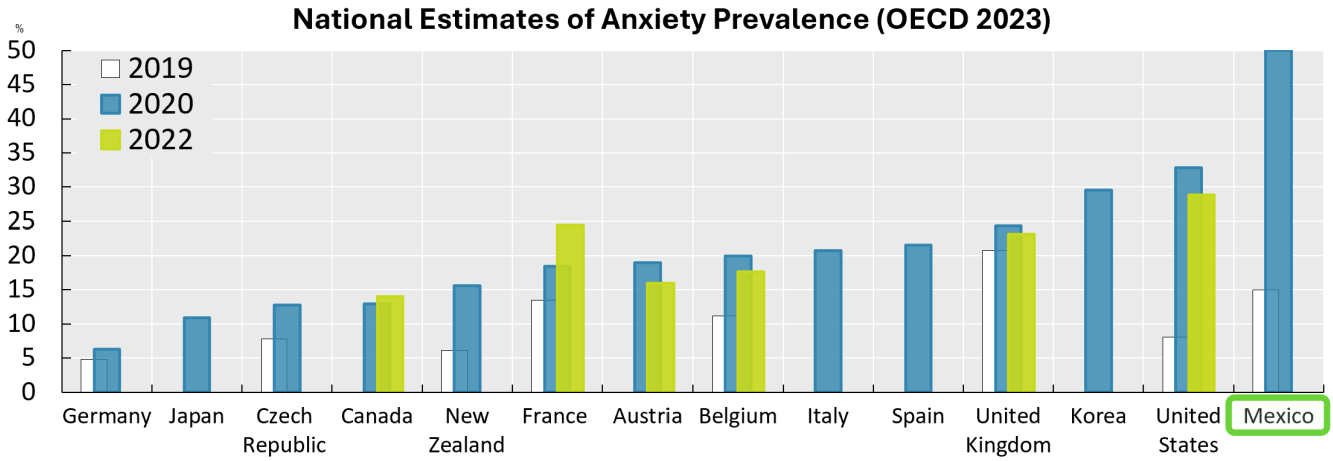
Please guess how many participants of this study, out of every 100 students, responded to the question above with "Strongly agree", "Agree", or "Somewhat agree"? In other words, out of every 100 students, how many responded Strongly Agree / Agree / Somewhat Agree that they would feel disappointed in themselves if they had a mental health issue? If your guess is within 5 students of the correct answer, you will earn a bonus of 50 MXN. (The correct estimate will be calculated based on the answers of the survey respondents. One of the bonus questions will be randomly chosen for payment.

**Variable Construction**: To measure self-stigma, we ask students how much they agree or disagree with the statement "I would feel disappointed in myself if I had a mental health issue (e.g., anxiety or depression)." We also ask students to guess how many survey participants of the study out of every 100 responded to the aforementioned question with "Strongly Agree", "Agree", or "Somewhat Agree."

# B Appendix: Figures and Tables

## B.1 National Statistics

Figure B1: National Estimates of Anxiety Prevalence (OECD 2022)

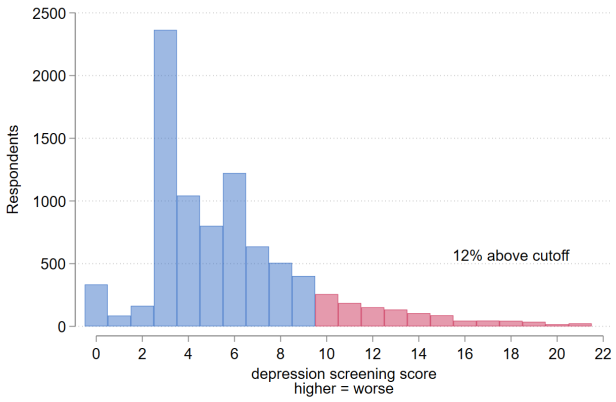



Notes: This figure shows national estimates of anxiety prevalence across OECD countries over time.

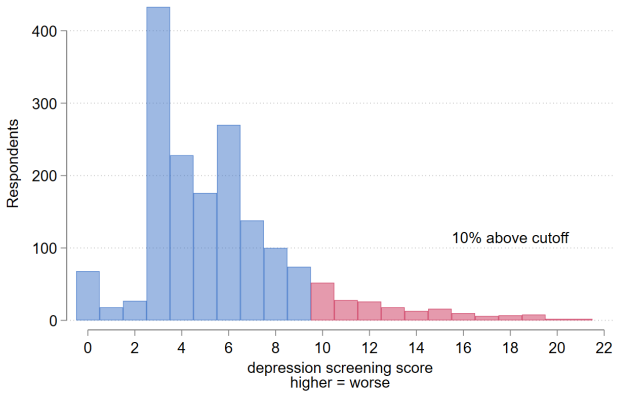
### Depression Screening Scores – ENSANUT

Figure B2: Mental Distress in Mexico - 2023

(a) National



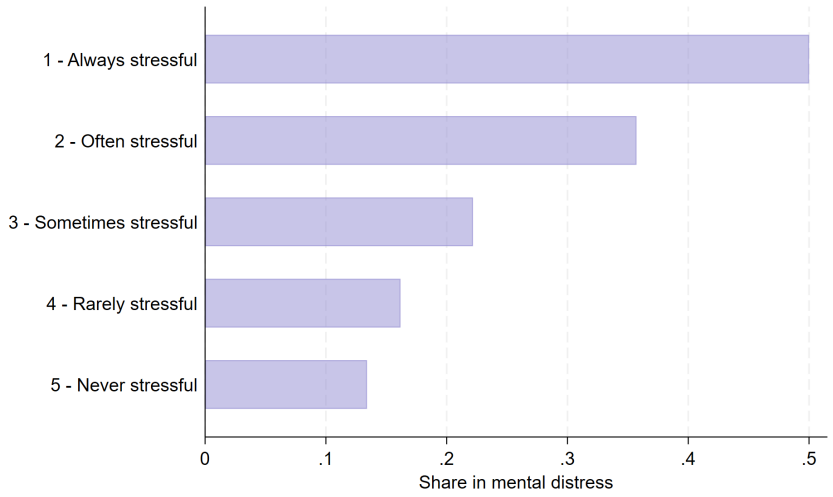

(b) National - Age ∈ [17, 28]




Notes: This figure shows the distribution of depression screening scores using data from the 2023 Mexican Health and Nutrition Survey (ENSANUT). The survey is representative of the national population. In panel (a) we show the distribution among ENSANUT respondents aged 10 years old or older (sample size = 8,696). In panel (b) we subset respondents to those between 17 and 28 years old to more closely approximate the population of university students (sample size = 1,720).

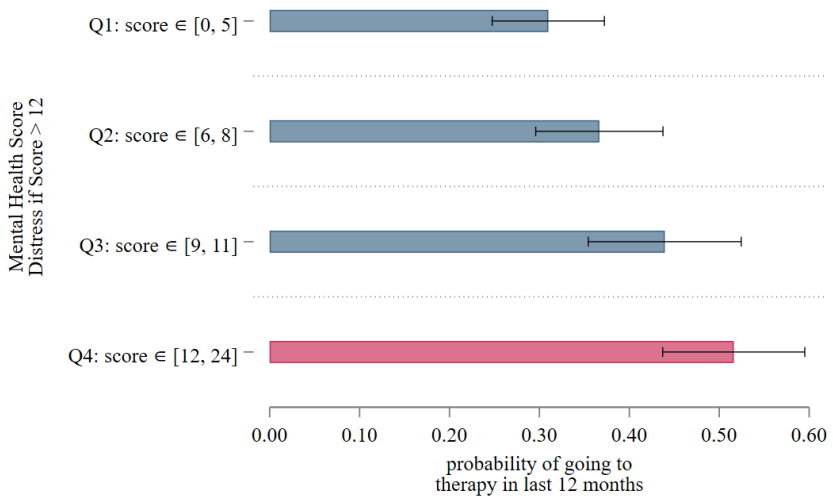
## B.2 Mental Health Index and Professional Help Use

Figure B3: Mental Distress Share by Financial Stress



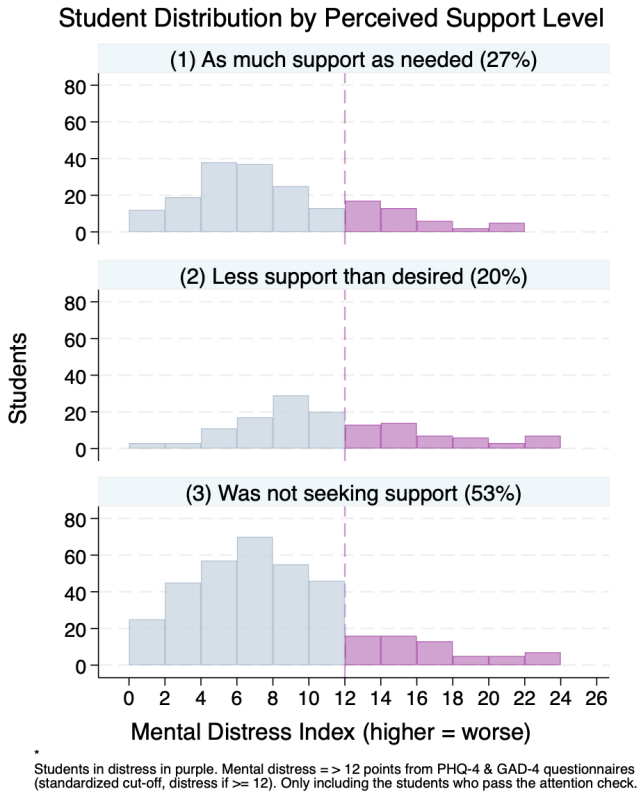

Notes: This figure shows the share of respondents in mental distress within each possible answer option to the question of *How would you describe your financial situation?*

Figure B4: Therapy Use Share by Mental Distress Quartile




Notes: This figure shows estimates of the probability of having used therapy in the last 12 months conditional on being in a given quartile of the mental health score distribution. We show 95% confidence intervals in black capped spikes.

Figure B5: Mental Distress Index Distribution by Perceived Support Level




Notes: This figure shows the distribution of mental health scores across responses to the question of whether *In the last 12 months I received...* More help than I needed/Less help than I needed/I was not seeking support.

Table B1: Comparison of Individual Covariates By Mental Distress

| | **In Distress (N=155)** | **Not in Distress (N=525)** | **p-value** |
|---|---|---|---|
| Female (%) | 56.8 | 49.3 | 0.104 |
| Age (years) | 20.4 | 20.1 | 0.042 |
| Heterosexual (%) | 64.5 | 77.9 | <0.001 |
| Year 3 or above (%) | 63.9 | 50.5 | 0.003 |
| GPA (0–100 scale) | 90.5 | 91.1 | 0.155 |
| Full scholarship (%) | 9.7 | 7.4 | 0.364 |
| Partial scholarship (%) | 67.7 | 69.5 | 0.674 |
| Financially stressed (%) | 70.3 | 51.4 | <0.001 |
| Moved from hometown (%) | 58.1 | 61.9 | 0.390 |
| Both parents with college degree (%) | 51.7 | 48.1 | 0.445 |

Notes: This table shows the means and p-value of the difference in means for covariates among students in distress and not in distress.

Table B2: Summary Statistics: Perceived Effectiveness, Support, and Therapy Use

| | **Mean** |
|---|---|
| **Perceived Effectiveness of Therapy** | |
| Agree: Therapy improves **my own** well-being | 0.904 |
| Agree: Therapy improves **people's** well-being | 0.924 |
| Agree with both | 0.897 |
| **Perceived Support for Therapy** | |
| Agree: Friends would support me going to therapy | 0.913 |
| Agree: Parents would support me going to therapy | 0.872 |
| Agree that both friends and parents would support | 0.843 |
| **Professional Help Received** | |
| Have ever received professional MH help | 0.662 |
| Have a friend who received professional MH help | 0.876 |
| Have a friend who would benefit from therapy | 0.894 |
| **(Last 12 Months)** | |
| Sought help from mental health professionals (last 12m) | 0.397 |
| → help from mental health professionals **at the university** | 0.203 |
| → help from mental health professionals **outside the university** | 0.260 |
| **Sample size** | 680 |

Notes: This table shows means for questions on perceived effectiveness, support and therapy use. For items under the Perceived Effectiveness of Therapy and Perceived Support for Therapy panels we ask *How much do you agree or disagree with the following statements?* (1) *Going to therapy can improve my own mental health* (2) *In general, going to therapy can improve people's mental wellbeing* (4) *My friends would show support if I told them I am going to therapy* (5) *My parents would show support if I told them I am going to therapy*; we code as "agree" responses which state Somewhat Agree, Agree or Strongly Agree. For items under the Professional Help Received panel we ask the following Yes/No questions: (i) Have you ever received professional mental help? (ii) Do you have a friend who is currently receiving or has previously received professional mental health?, and (iii) Do you have a friend or someone you know closely who you think would benefit from therapy? Finally, we ask *If you experienced mental health challenges in the last 12 months, [...], to who did you turn for help? Select ALL that apply* for items under the (Last 12 Months) panel.

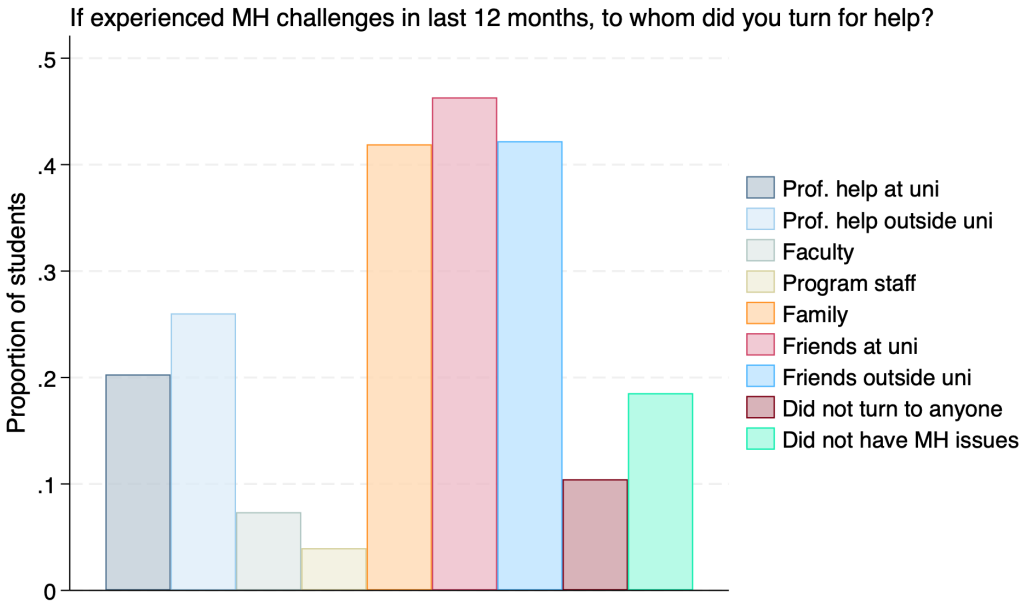
Table B3: Belief accuracy before and after the information intervention

| | | ***Incorrect* Belief %** | | |
|---|---|---|---|---|
| | **Fact from the Treatment** | **Prior** | **Posterior** | **Δ (pp)** |
| (1) | Psychotherapy yields long-term (4–5 yr) benefits[†] | 3.1 | 1.8 | –1.3 |
| (2) | Most students in therapy have mild or no symptoms | 42.2 | 3.8 | –38.3 |
| (3) | GPA is negatively correlated with mental distress | 86.6 | 49.3 | –37.3 |

*Notes*: Means are calculated among all treated respondents ($N = 448$) as posteriors were only elicited for them. "Prior" refers to beliefs elicited immediately *before* the information intervention (Fact 3 prior was elicited categorically, while posterior was elicited as a binary statement consistent with other priors and posteriors). "Posterior" is the same question asked after the information intervention . Δ is the simple difference (posterior minus prior) expressed in percentage points.

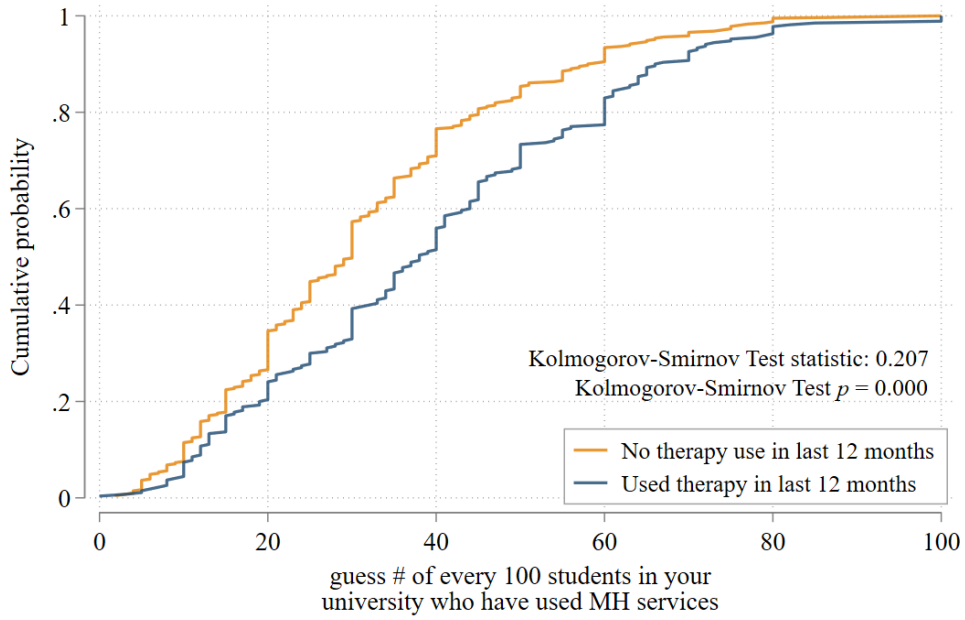
[†] Nearly all students held the correct prior on Fact 1, leaving limited scope for updating.

Figure B6: Who Did You Turn For Help?




Notes: This figure shows the share of students who chose each of the options to the question of whom did the respondent turn for help in case she experienced mental health challenges in the past 12 months.

Figure B7: CDFs of guesses of therapy use by prior own use.

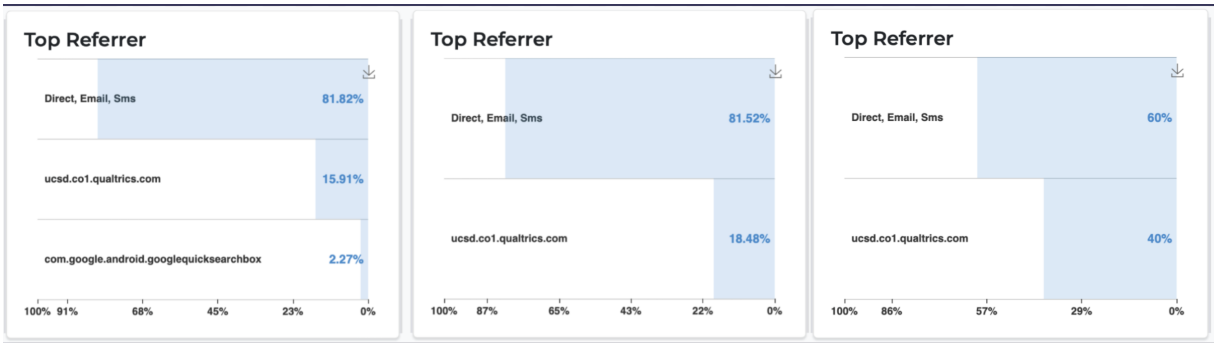



Notes: This figure shows the cumulative distribution function of the guesses of the number of students out of every 100 students from their university who have used professional mental health services in the last 12 months, splitting the sample by those who have/have not used professional mental health services in the last 12 months.

Figure B8: Resource link sharing

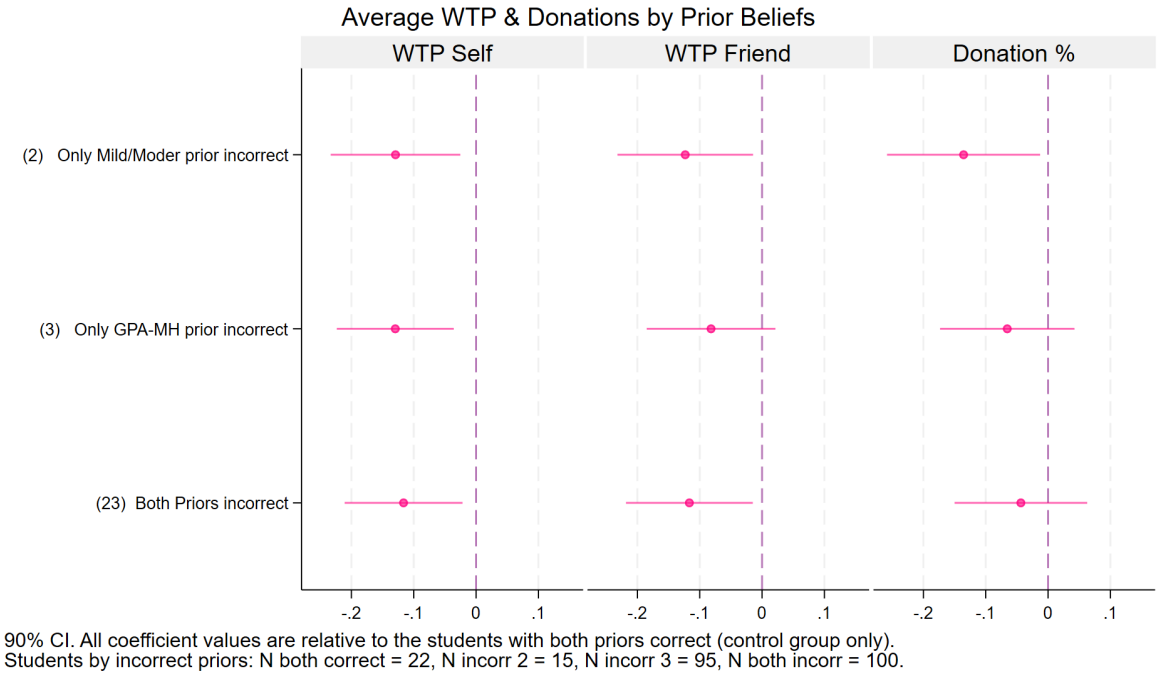

Notes: This figure shows the infographic containing the QR code for sharing campus services information.

Figure B9: Link Sharing Top Referrers by Treatment Group




Notes: Top referrer data for each experimental group: *Information+Reflection* (left), *Information* (center), *Control* (right). The figure shows the proportion of link clicks that originated from outside the Qualtrics survey interface (e.g., via direct link, email, or SMS) versus within Qualtrics. Among treated participants (T1 and T2), over 80% of clicks were external, suggesting peer sharing. In contrast, only 60% of the Control group clicks came from external sources.

Figure B10: Willingness To Pay for Private Therapy and Donation




Notes: This figure shows the difference in mean estimates on willingness to pay and donation outcomes. We estimate $Y_i = \alpha + \beta_M \texttt{OnlyMildIncorrect}_i + \beta_G \texttt{OnlyGPAIncorrect}_i + \beta_B \texttt{BothIncorrect}_i + \varepsilon_i$, where $Y_i$ is the outcome of interest, `OnlyMildIncorrect` is an indicator equal to 1 if the respondent only answered the "Mild/Moderate"-prior question incorrectly, `OnlyGPAIncorrect` is an indicator equal to 1 if the respondent only answered the "GPA-MH"-prior question incorrectly and `BothIncorrect` is an indicator equal to 1 if the respondent answered both the "Mild/Moderate"- and the "GPA-MH"-questions incorrectly. The reference group is the group of respondents who answered all priors' questions correctly. Horizontal lines represent 90% confidence intervals.

Table B4: Suggestive Correlations on Interpersonal Projection

| | (1) Guess of %: w Self Stigma | (2) Open to Share | (3) Using Therapy |
|---|---|---|---|
| Self-stigma (feels | 0.170*** | | |
| disappointed if MH issue) | (0.017) | | |
| Open to share MH challenges | | 0.106*** | |
| with classmates | | (0.015) | |
| Used therapy 12m | | | 0.081*** |
| | | | (0.016) |
| Constant | 0.447*** | 0.195*** | 0.313*** |
| | (0.010) | (0.008) | (0.009) |
| Observations | 680 | 680 | 680 |
| Mean dep var | 0.498 | 0.234 | 0.345 |
| Std dev dep var | 0.23 | 0.19 | 0.20 |

Notes: The table reports the coefficient on the belief indicator (self-stigma, openness to share) or therapy use in a regression of incentivized guesses of the prevalence of each belief/behavior among survey respondents. The coefficient captures to what extent holding a personal belief (self-stigma or being open to share problems with other students) or using therapy is associated with assuming that more students hold the same beliefs or engage in the same behavior. The constant term reflects the guess percentage among those who do not hold the belief/engage in the behavior. Robust S.E. *** 1%, ** 5%, * 10% significant.

## B.3 Balance on observables

Table B5: Covariate Balance among the Followup Respondents

| | (1) Control | | (2) Treated | | (1)-(2) Pairwise t-test | |
|---|---|---|---|---|---|---|
| Variable | N | Mean/(SD) | N | Mean/(SD) | N | Mean difference |
| Age | 109 | 19.936 | 211 | 20.118 | 320 | -0.183 |
| | | (1.577) | | (1.847) | | |
| Female | 109 | 0.523 | 211 | 0.521 | 320 | 0.002 |
| | | (0.502) | | (0.501) | | |
| Financially Stressed | 109 | 0.578 | 211 | 0.611 | 320 | -0.033 |
| | | (0.496) | | (0.489) | | |
| Has Scholarship | 109 | 0.661 | 211 | 0.768 | 320 | -0.107** |
| | | (0.476) | | (0.423) | | |
| Receives a full scholarship | 109 | 0.128 | 211 | 0.100 | 320 | 0.029 |
| | | (0.336) | | (0.300) | | |
| Moved Residence | 109 | 0.596 | 211 | 0.668 | 320 | -0.072 |
| | | (0.493) | | (0.472) | | |
| GPA | 109 | 91.798 | 211 | 91.313 | 320 | 0.485 |
| | | (4.307) | | (4.211) | | |
| MH Score | 109 | 8.596 | 211 | 8.005 | 320 | 0.592 |
| | | (5.418) | | (4.902) | | |
| Used Therapy L12 Months | 109 | 0.459 | 211 | 0.384 | 320 | 0.075 |
| | | (0.501) | | (0.487) | | |
| Open to Share MH Challenges | 109 | 0.367 | 211 | 0.336 | 320 | 0.030 |
| | | (0.484) | | (0.474) | | |
| Self-stigmatize | 109 | 0.321 | 211 | 0.294 | 320 | 0.027 |
| | | (0.469) | | (0.457) | | |

Notes: This balance table replicates the balance test just among those students who responded to the followup and took classes in the university in 2025, who comprise our followup analysis sample. We pool T1 and T2 into a "Treated" group. This table shows balance on covariates across treatment groups. For each covariate we show each experimental group's sample mean and standard deviation, as well as the difference in means across both groups. Age measures the respondent's age in years, female is an indicator equal to one if the respondent is female-born, financially stressed is an indicator equal to one if the respondent described her financial situation as "Always", "Often" or "Sometimes" stressful and equal to 0 if she reported it as "Rarely" or "Never" stressful, Has scholarship is an indicator equal to one if the respondent has at least some amount of scholarship, receives a full scholarship is an indicator equal to one if the respondent's scholarship covers 100% of tuition, moved residence is an indicator equal to one if the respondent moved her residence city to pursue her current studies, GPA measures the respondent's current overall GPA on a scale from 0–100, MH score measures the student's mental health score as described in section **??**, used therapy in L12 months is an indicator equal to one if the respondent states having used therapy in the last 12 months, open to share MH challenges is an indicator equal to one if the respondent states she would be willing to share about her own personal MH challenges with others and self-stigmatize is an indicator equal to one if the respondent states she would be disappointed in herself if she suffered from mental distress. Standard errors for the difference in means test are heteroskedasticity robust. Significance levels: * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$

Table B6: Covariate Balance across T1, T2, C

| | (1) (0) Control | | (2) (1) T1: Info + Reflection | | (3) (2) T2: Info only | | (1)-(2) | | (1)-(3) Pairwise t-test | | (2)-(3) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean/(SD) | N | Mean/(SD) | N | Mean/(SD) | N | Mean difference | N | Mean difference | N | Mean difference |
| Age | 232 | 20.159 (1.848) | 227 | 20.084 (2.218) | 221 | 20.208 (1.822) | 459 | 0.076 | 453 | -0.049 | 448 | -0.124 |
| Female | 232 | 0.461 (0.500) | 227 | 0.533 (0.500) | 221 | 0.538 (0.500) | 459 | -0.072 | 453 | -0.077 | 448 | -0.005 |
| Financially Stressed | 232 | 0.530 (0.500) | 227 | 0.599 (0.491) | 221 | 0.543 (0.499) | 459 | -0.069 | 453 | -0.013 | 448 | 0.056 |
| Has Scholarship | 232 | 0.651 (0.478) | 227 | 0.718 (0.451) | 221 | 0.706 (0.457) | 459 | -0.067 | 453 | -0.055 | 448 | 0.012 |
| Receives a full scholarship | 232 | 0.082 (0.275) | 227 | 0.084 (0.278) | 221 | 0.072 (0.260) | 459 | -0.002 | 453 | 0.009 | 448 | 0.011 |
| Moved Residence | 232 | 0.591 (0.493) | 227 | 0.626 (0.485) | 221 | 0.615 (0.488) | 459 | -0.035 | 453 | -0.025 | 448 | 0.010 |
| GPA | 232 | 90.897 (4.659) | 227 | 90.784 (5.394) | 221 | 91.235 (3.925) | 459 | 0.112 | 453 | -0.339 | 448 | -0.451 |
| MH Score | 232 | 8.569 (5.132) | 227 | 8.048 (5.003) | 221 | 8.430 (5.110) | 459 | 0.521 | 453 | 0.139 | 448 | -0.381 |
| Used Therapy L12 Months | 232 | 0.233 (0.424) | 227 | 0.181 (0.386) | 221 | 0.290 (0.455) | 459 | 0.052 | 453 | -0.057 | 448 | -0.109*** |
| Open to Share MH Challenges | 232 | 0.392 (0.489) | 227 | 0.339 (0.474) | 221 | 0.371 (0.484) | 459 | 0.053 | 453 | 0.021 | 448 | -0.032 |
| Self-stigmatize | 232 | 0.323 (0.469) | 227 | 0.295 (0.457) | 221 | 0.276 (0.448) | 459 | 0.028 | 453 | 0.047 | 448 | 0.019 |

Notes: This table shows balance on covariates across treatment groups. For each covariate we show each experimental group's sample mean and standard deviation, as well the difference in means across pairs of groups. Age measures the respondent's age in years, female is an indicator equal to one if the respondent is female-born, financially stressed is an indicator equal to one if the respondent described her financial situation as "Always", "Often" or "Sometimes" stressful and equal to 0 if she reported it as "Rarely" or "Never" stressful, Has scholarship is an indicator equal to one if the respondent has at least some amount of scholarship, receives a full scholarship is an indicator equal to one if the respondent's scholarship covers 100% of tuition, moved residence is an indicator equal to one if the respondent moved her residence city to pursue her current studies, GPA measures the respondent's current overall GPA on a scale from 0–100, MH score measures the student's mental health score as described in section **??**, used therapy in L12 months is an indicator equal to one if the respondent states having used therapy in the last 12 months, open to share MH challenges is an indicator equal to one if the respondent states she would be willing to share about her own personal MH challenges with others and self-stigmatize is an indicator equal to one if the respondent states she would be disappointed in herself if she suffered from mental distress. Standard errors for the difference in means test are heteroskedasticity robust. Significance levels: * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$

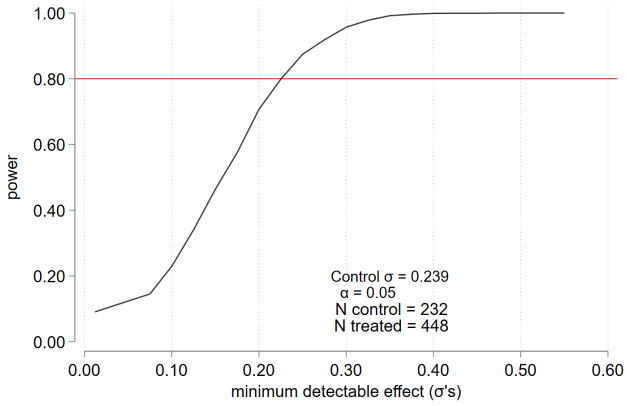
## B.4 Power Calculations

In this section we present Monte Carlo simulations for power calculations. We conduct power calculations doing the following procedure:

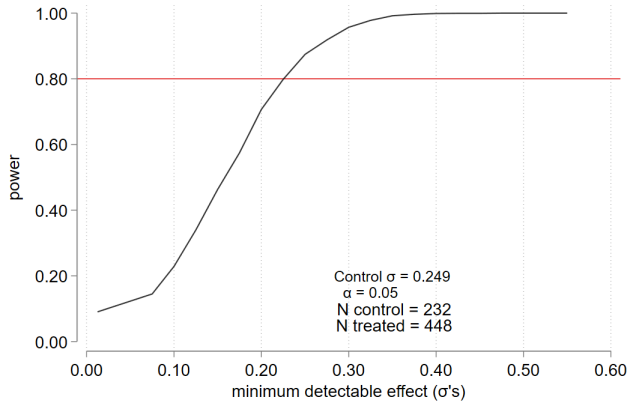
1. Fix an effect size $\tau_j$ in terms of the outcome's standard deviation.
2. Fix a significance level $\alpha = 0.05$
3. Set observations to 232 control units and 448 observations to treatment.
4. Start loop $l \in \{1, 2, \ldots, 2000\}$
5. Simulate the data generating process using the control group's observed mean and standard deviation for control units.
6. Simulate the data generating process for treated units with the control group's observed mean plus the effect size, and the standard deviation of the control group.
7. Estimate the treatment effect on the simulated data and store the p-value associated to the treatment effect in the current simulation—call it $p_l$.
8. If $l < 2000$, go back to step 4 and repeat. If $l = 2000$ continue to next step.
9. Compute the share of times for which $p_l < 0.05 = \alpha \; ; l \in \{1, 2, \ldots, 2000\}$. This is the power associated to effect size $\tau_j$, given the sample size, the treatment allocation, significance level, and outcome mean and standard deviation.

Figure B11: Power Calculation for Pre-registered Outcomes

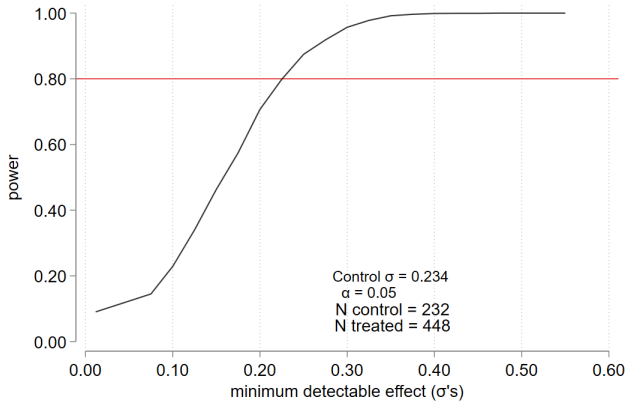
(a) WTP for Therapy for Self




(b) WTP for Therapy for Friend



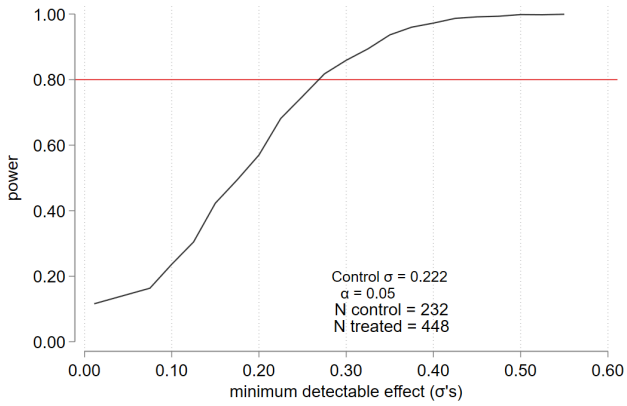

(c) Donation



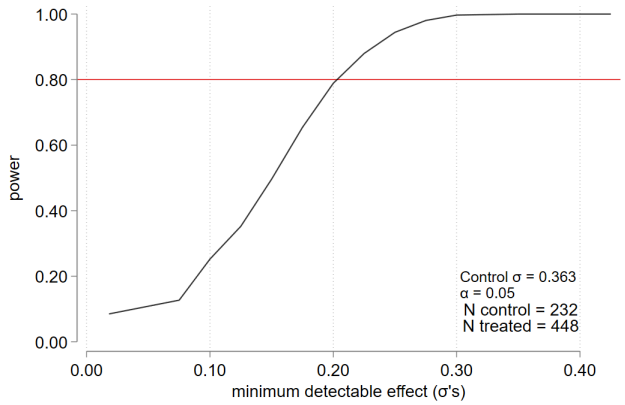

Notes: This figure shows power calculations for WTP for therapy for oneself, WTP for therapy for a friend, and donations. Standard errors used in the simulations are heteroskedasticity robust.

Figure B12: Power Calculation for Pre-registered Outcomes

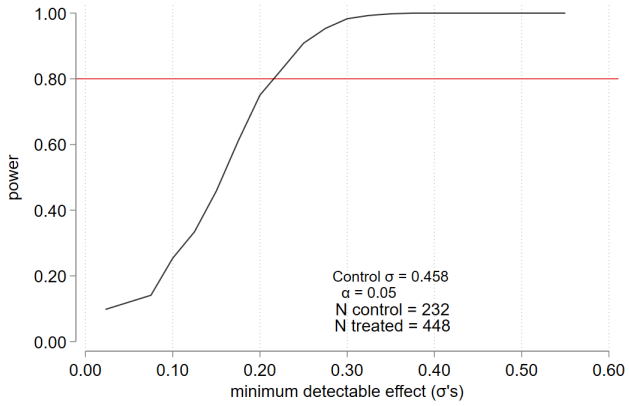
(a) Advice mentions On-Campus Prof. Help




(b) Prefer MH Talk over low GPA




(c) Donation




Notes: This figure shows power calculations for the probability of mentioning on-campus professional help in the advice, preferring a student who talks about mental health issues over one with low GPA for group work, and preferring a student who shows mental health symptoms over one with low GPA for group work. Standard errors used in the simulations are heteroskedasticity robust.

## B.5 Additional Experimental Results

Table B7: Poisson Test Results for Link Clicks

| | | Approx. Poisson | | Exact Poisson | |
|---|---|---|---|---|---|
| **Test** | **Rate Ratio** | **p-value** | **Reject $H_0$** | **p-value** | **Reject $H_0$** |
| *Panel A: Total Clicks* | | | | | |
| T1 (Info+Reflection) vs C | 1.32 | 0.221 | False | 0.260 | False |
| T2 (Info Only) vs C | 2.69 | <0.001 | True | <0.001 | True |
| T1 vs T2 | 0.49 | <0.001 | True | <0.001 | True |
| T1&T2 vs C | 2.01 | <0.001 | True | <0.001 | True |
| *Panel B: Unique Clicks* | | | | | |
| T1 vs C | 1.53 | 0.107 | False | 0.118 | False |
| T2 vs C | 2.51 | <0.001 | True | <0.001 | True |
| T1 vs T2 | 0.61 | 0.019 | True | 0.023 | True |
| T1&T2 vs C | 2.03 | <0.001 | True | 0.001 | True |

Notes: This table shows estimates for the rate ratio of treatment group link clicks to those of the control groups. In Panel A we focus on total clicks, while on Panel B we focus on "unique" clicks. For each comparison we show the p-value associated to the null hypothesis of rate ratios equal to 1, both using Wald-type and permutation-based tests. Panel A shows the estimated click-through rate ratios, p-values and test conclusions for total number of clicks, while Panel B shows the estimates for unique human clicks. It is clearly seen that regardless of the type of the test or the type of link clicks, it is always the case that we reject the null of equality of two rates between joint treatment (T1 & T2) and control groups ($p \leq 0.001$). The effect is driven by a high click-through rate in (T2) *Information Only* treatment group as the rate ratio is the highest (and p-value is the smallest) when we compare T2 and C. The click rates in (T1) *Information + Reflection* and control groups are not statistically different from each other and we cannot reject the equality of two means ($p = 0.260$ for total clicks and $p = 0.118$ for unique clicks). While the main results depicted here reflect short-run behavior within one week of the intervention, we also tracked engagement six months later. At that point, we recorded 179 total clicks in the pooled Treatment group and 58 in the Control group, indicating that the intervention effect on engagement with professional resources persisted over time and remained statistically significant.

## B.6 SR Peer Advice – Exploration by Type of Advice

The share of words or phrases related to empathetic advice is 2.9 *pp* lower for students in treatment conditions.[48] This decrease is mostly driven by students in the Treatment + Reflection group, whose share of words or phrases mentioned decreases by 3.9 *pp*, whereas that of students in the Information Only group also decreases but only by a mild—and insignificant—1.8 *pp*. Turning to directive advice we do not observe any differences between the share of words/phrases mentioned by students in the control group and those in either of the treatment groups. These results, along with the documented high level of knowledge about on-campus services, suggest students who are provided with information about therapy effectiveness change the composition of the advice they give to friends in distress. In particular, they substitute advice in which they state they, e.g., "are there for them" or "are there if they want to talk", in favor of more targeted advice where they prompt their friend about available on-campus services.

[48] See appendix section D.5 for a detailed explanation of advice processing.

Table B8: Effects on Demand for Mental Health Treatments

| | SR: WTP Therapy ($N = 680$) | | SR: WTP Therapy ($N = 320$) | | LR: Used Therapy (6m) |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | WTP % self | WTP % friend | WTP % self | WTP % friend | Any |
| Treated | -0.036* | -0.033 | -0.045 | -0.041 | -0.009 |
| | (0.020) | (0.021) | (0.031) | (0.032) | (0.059) |
| Control Mean | 0.429 | 0.426 | 0.457 | 0.460 | 0.434 |
| Control SD | 0.24 | 0.25 | 0.26 | 0.27 | 0.50 |
| Observations | 680 | 680 | 320 | 320 | 320 |

Notes: This table reports treatment effects on students' demand for mental health therapy services, measured through their stated willingness to pay (WTP) at baseline and in the follow-up survey. Columns (1)–(2) show short-run effects on the willingness to pay for therapy for themselves and for a friend, using the full sample ($N = 680$). Columns (3)–(4) restrict the sample to students who were matched in the follow-up ($N = 320$). Column (5) presents the effects on whether students reported using any form of therapy in the follow-up survey. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SR = short run (baseline); LR = long run (follow-up).

Table B9: TE by Prior Therapy Use (Follow-Up Respondents)

| | SR: WTP for Therapy | | LR: Therapy Use | | LR: Recommend Therapy | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | WTP Self | WTP Friend | Off Campus | On Campus | On Campus | Off Campus |
| Treated × No Help | -0.058 | -0.047 | -0.051 | -0.083 | 0.115 | -0.104 |
| | (0.040) | (0.042) | (0.051) | (0.064) | (0.071) | (0.078) |
| Treated × Used Help | -0.020 | -0.034 | 0.073 | 0.319*** | -0.163* | 0.168* |
| | (0.048) | (0.049) | (0.082) | (0.086) | (0.089) | (0.086) |
| Used Prof Help | 0.020 | -0.022 | 0.124 | 0.123 | 0.366*** | 0.102 |
| | (0.051) | (0.052) | (0.077) | (0.088) | (0.090) | (0.096) |
| Constant | 0.448*** | 0.470*** | 0.136*** | 0.237*** | 0.254*** | 0.458*** |
| | (0.033) | (0.035) | (0.045) | (0.056) | (0.057) | (0.065) |
| Control Mean | 0.457 | 0.460 | 0.193 | 0.294 | 0.422 | 0.505 |
| Control SD | 0.26 | 0.27 | 0.40 | 0.46 | 0.50 | 0.50 |
| Observations | 320 | 320 | 320 | 320 | 320 | 320 |

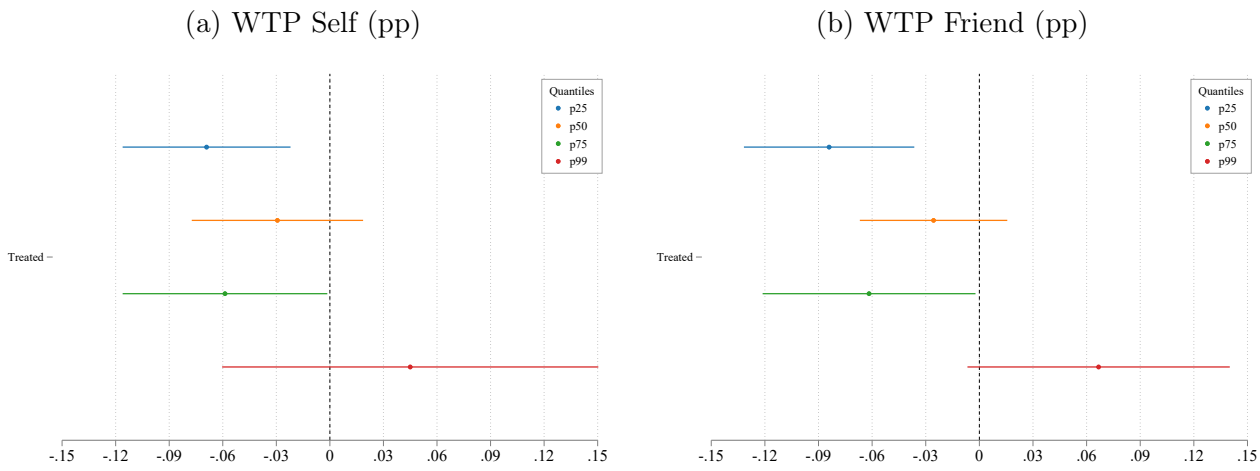
Notes: This table presents heterogeneity in treatment effects by prior use of professional mental health services among students still enrolled in 2025. The analysis interacts the treatment indicator with a binary variable for whether the respondent had used professional help before the intervention.
Columns (1)–(2) report short-run effects on willingness to pay (WTP) for therapy for self and for a friend, measured at baseline. Columns (3)–(4) show long-run effects on self-reported therapy use in the 6-month follow-up, separately for off-campus and on-campus services. Columns (5)–(6) present long-run effects on whether respondents would recommend therapy to others, again split by therapy location.Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B10: ATE on Advice

| | (1) Campus Help | (2) Empathetic Advice | (3) Directive Non-therapy | (4) Campus Help | (5) Empathetic Advice | (6) Directive Non-therapy |
|---|---|---|---|---|---|---|
| Any Treatment | 0.038* | -0.029* | -0.004 | | | |
| | (0.020) | (0.017) | (0.015) | | | |
| Info + Reflection | | | | 0.036 | -0.039** | 0.001 |
| | | | | (0.024) | (0.018) | (0.018) |
| Info Only | | | | 0.039 | -0.018 | -0.009 |
| | | | | (0.024) | (0.020) | (0.017) |
| Observations | 680 | 680 | 680 | 680 | 680 | 680 |
| R2 | 0.004 | 0.005 | 0.000 | 0.005 | 0.007 | 0.001 |
| Control Mean | 0.052 | 0.216 | 0.184 | 0.052 | 0.216 | 0.184 |

Notes: This table presents the effects of the information intervention on the content of students' responses to a hypothetical advice prompt at baseline. Each column reports the ATE, where the variable are a binary indicator coded as 1 if the student mentioned a given type of advice. Columns (1)–(3) show the pooled treatment effect for all treated students (Any Treatment), while columns (4)–(6) separate the effects by treatment arm: "Information Only" and "Information + Reflection."
"Campus Help" refers to a specific mention of on-campus mental health services, such as university counseling. "Empathetic Advice" includes responses showing emotional support, while "Directive Non-therapy" refers to concrete suggestions other than professional therapy. The intervention significantly increases the likelihood of recommending on-campus help (column 1: 3.8 percentage points, p<0.1) and significantly reduces the likelihood of offering empathetic advice (column 2: 2.9 percentage points, p<0.1). The effects are primarily driven by the "Information + Reflection" arm (columns 4–5). The estimates suggest no significant change in directive but non-therapeutic advice. All specifications include robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure B13: Quantile regressions




Notes: This figure shows treatment effect estimates and 90% confidence intervals of quantile regressions for WTP for oneself and for a friend. We denote the first quartile by p25, the median by p50, the third quartile by p75, and the 99th percentile by p99. Quantile results are not the upper quartile for donation given the bunching of chosen percentages for donations in subject responses.

Table B11: Advice Components

| | Empathetic Components Share | | | Campus Help in Advice | | | Campus Help in Advice | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) All | (2) Excl 1 | (3) Excl 0.8 | (4) All | (5) Excl 1 | (6) Excl 0.8 | (7) All | (8) Excl 1 | (9) Excl 0.8 |
| Treated | -0.029* | -0.031* | -0.027* | 0.038* | 0.036* | 0.036* | | | |
| | (0.017) | (0.017) | (0.016) | (0.020) | (0.020) | (0.020) | | | |
| Empathetic advice share | | | | | | | -0.061 | -0.091* | -0.086 |
| | | | | | | | (0.059) | (0.052) | (0.059) |
| Observations | 680 | 679 | 670 | 680 | 679 | 670 | 680 | 679 | 670 |
| Mean dep var | 0.197 | 0.196 | 0.188 | 0.076 | 0.075 | 0.076 | 0.076 | 0.075 | 0.076 |
| Std dev dep var | 0.20 | 0.20 | 0.19 | 0.27 | 0.26 | 0.27 | 0.27 | 0.26 | 0.27 |

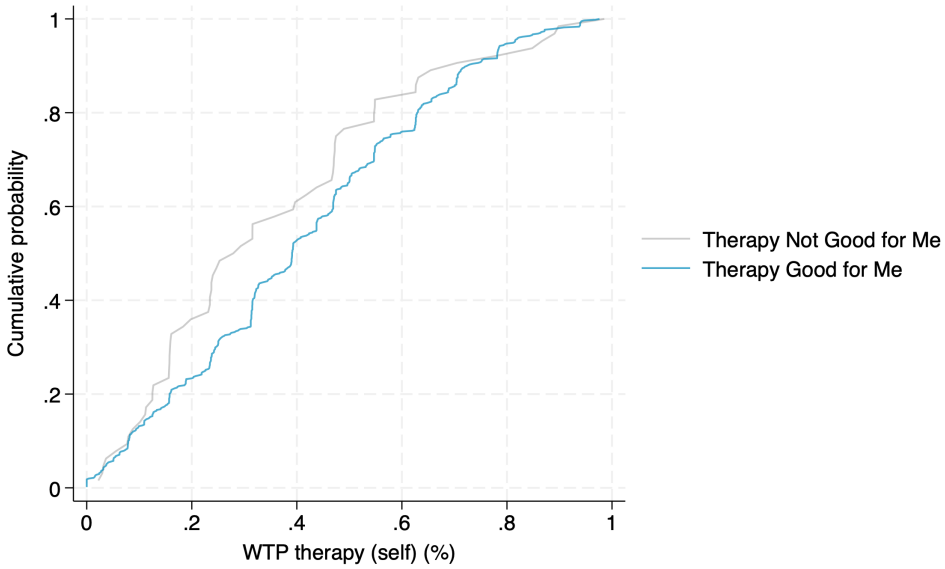
Notes: This table presents the effects of the information treatment on the composition of students' responses to a hypothetical advice prompt, measured as the share of specific components included in their message to a friend in distress. Columns (1)–(3) show effects on the Empathetic Components Share, defined as the fraction of response segments that provide emotional support. Columns (4)–(6) report the likelihood that a respondent mentions campus mental health services in the advice, controlling for treatment only. Columns (7)–(9) extend this by adding the empathetic component share. Columns (2), (5), and (8) exclude the single respondent who mentioned all tracked components, while columns (3), (6), and (9) exclude all respondents who mentioned 80% or more of the components.All regressions include robust standard errors reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B12: Effects on Personal & Public Stigma-Related Outcomes

| | SR: Prefer vs Low GPA (680) | | SR: Prefer vs Low GPA (320) | | LR: Discuss MH / Therapy | |
|---|---|---|---|---|---|---|
| | (1) Distress Sympt | (2) MH Talk | (3) Distress Sympt | (4) MH Talk | (5) Own MH Issues | (6) Therapy |
| Treated | 0.043 | 0.026 | -0.013 | 0.000 | -0.079 | -0.068 |
| | (0.036) | (0.029) | (0.052) | (0.041) | (0.052) | (0.056) |
| Control Mean | 0.703 | 0.845 | 0.743 | 0.862 | 0.761 | 0.376 |
| Control SD | 0.46 | 0.36 | 0.44 | 0.35 | 0.43 | 0.49 |
| Observations | 680 | 680 | 320 | 320 | 320 | 320 |

This table presents estimates of the short-run and long-run effects of the intervention on outcomes related to mental health stigma. Columns (1)–(2) report short-run effects measured in the baseline survey experiment using the full sample ($N = 680$). Column (1) assesses students' willingness to work with a peer exhibiting visible distress symptoms, relative to a peer with a low GPA. Column (2) measures whether students feel comfortable engaging in conversations about mental health topics more generally, again relative to the low GPA reference case. Columns (3)–(4) replicate the same short-run measures but restrict the sample to the subset of follow-up respondents ($N = 320$). Columns (5)–(6) present long-run effects based on follow-up data collected six months after the intervention. These outcomes measure whether participants discussed their own mental health issues (column 5) or the topic of therapy (column 6) with others. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SR = short run; LR = long run.

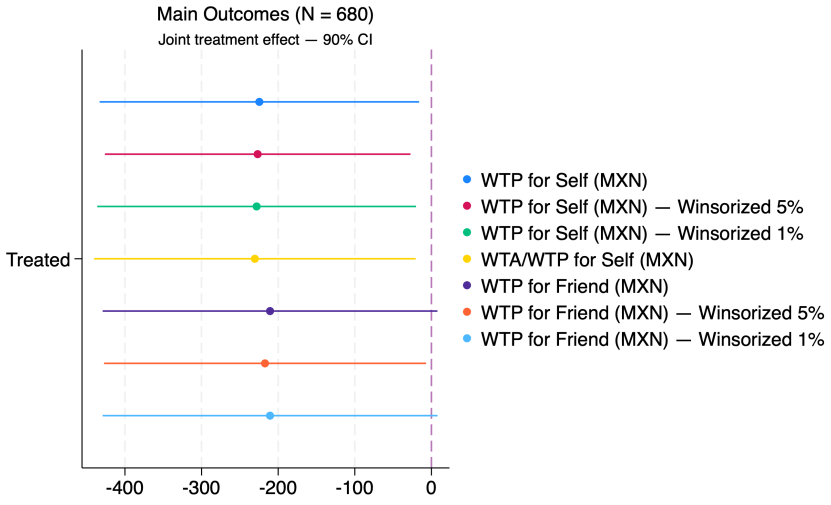
Figure B14: Validity of WTP Measures: Own WTP by Perceived Usefulness



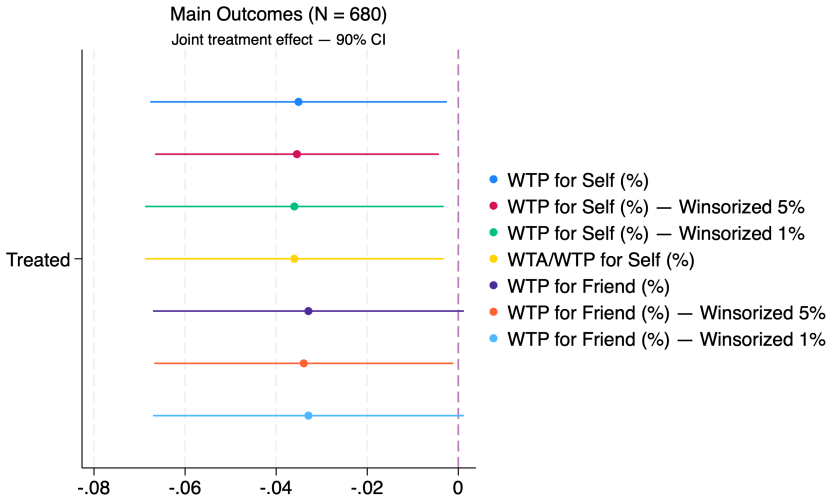

Notes: The figure plots the empirical cumulative distribution function (CDF) of the *incentivized willingness to pay for a one-month private online therapy for self*, rescaled to $[0, 1]$. The CDFs are split by *perceived usefulness of therapy for oneself* (binary indicator from the baseline Lickert-scale question asking to agree/disagree that "going to therapy can improve my own mental wellbeing"). A curve lying *below* indicates higher WTP (first-order stochastic dominance). Sample includes all respondents with non-missing WTP and the corresponding split variable.

Figure B15: Robustness of ATE Estimates to WTP Measures

(a) Measures in Mexican Pesos




(b) Percent Measures




Notes: Each panel reports coefficient estimates of the joint treatment indicator with 90% confidence intervals. *Raw MXN panel*: Outcomes are expressed in pesos. Entries labeled "WTP for Self (MXN)" and "WTP for Friend (MXN)" correspond to the unaltered willingness-to-pay (WTP) measures. Versions labeled "Winsorized 1%" and "Winsorized 5%" trim extreme values at the respective percentiles. "WTA/WTP for Self (MXN)" is a combined measure: it equals WTP when strictly positive and the negative of the willingness-to-accept (WTA) value when WTP is zero. *Percent Measures panel*: Outcomes are normalized by the monthly price of BetterHelp (6,500 MXN). Labels mirror the MXN panel, with "(%)" denoting percentage terms and "WTA/WTP for Self (%)" the normalized version of the combined self measure.

## B.7 Results with LASSO covariates

We show the results of our main regression specifications including covariates selected via post double-selection LASSO (Belloni et al. 2013).

The procedure for post double-selection LASSO is as follows:

1. Let $S$ denote the set of covariates which could be potentially included as controls in the regression specification.
2. Perform LASSO regression of the outcome of interest onto the covariates $s \in S$.
3. Denote the set of covariates which are predictive of the outcome by $S_1$.
4. Perform LASSO regression of the treatment indicator(s) of interest onto the covariates $s \in S$.
5. Denote the set of covariates which are predictive of the treatment indicator(s) by $S_2$.
6. Regress the outcome of interest onto the treatment indicators and covariates $s \in S_1 \cup S_2$

The controls selected in the specifications are respondent's age, sex, an indicator for whether they are in financial distress, an indicator for whether they have some amount of scholarship, and indicator for whether they moved to pursue their degree, their GPA, mental health score, indicator for self-reported therapy use in the last 12 months, an indicator for whether they would be open to share their mental health issues with peers who are not necessarily their friends and an indicator for whether they would be disappointed in themselves if they had mental health issues.

Our results do not change substantially with respect to specifications which do not include covariates. Standard errors remain almost unchanged and the estimates' magnitudes is slightly reduced. The reduction in estimate magnitude results in significance loss for our marginally significant outcomes at the 10% level.

Table B13: Effects on Advice Prompt at Baseline & Recommending Therapy in 6 Months

| | SR: Advice Prompt (All) | | Advice Prompt (in Followup) | | LR: Suggested Therapy | |
|---|---|---|---|---|---|---|
| | (1) On-Campus Help | (2) Any Prof. Help | (3) On-Campus Help | (4) Any Prof. Help | (5) On Campus | (6) Off Campus |
| Treated | 0.031 | 0.006 | 0.036 | -0.015 | -0.012 | 0.027 |
| | (0.020) | (0.039) | (0.030) | (0.060) | (0.056) | (0.057) |
| Control Mean | 0.052 | 0.358 | 0.046 | 0.376 | 0.422 | 0.505 |
| Control SD | 0.22 | 0.48 | 0.21 | 0.49 | 0.50 | 0.50 |
| Observations | 680 | 680 | 320 | 320 | 320 | 320 |
| LASSO Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: This table reports treatment effects on students' responses to a hypothetical advice prompt at baseline (short run, SR) and their self-reported likelihood of recommending therapy in the follow-up survey (long run, LR). All regressions include covariates selected through the post double-selection LASSO procedure (Belloni et al. 2013), indicated by the checkmarks in the "LASSO Controls" row. Columns (1)–(2) show short-run effects using the full sample ($N = 680$), while columns (3)–(4) restrict to students who were matched in the follow-up survey ($N = 320$) for comparability with long-run outcomes in columns (5)–(6). "On-Campus Help" refers to specific mentions of university counseling services, while "Any Prof. Help" includes both on- and off-campus therapy. The post double-selection LASSO algorithm selects covariates in two steps: first, based on their predictive power for the outcome, and second, based on their association with the treatment indicator. Covariates considered include age, sex, GPA, financial distress, scholarship status, relocation status, baseline mental health score, prior therapy use, willingness to share mental health issues with acquaintances, and self-stigma. All outcomes are binary. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SR = short run; LR = long run.

Table B14: Effects on Personal & Public Stigma-Related Outcomes

| | SR: Prefer over Low GPA Student | | LR: Discuss MH / Therapy | | |
|---|---|---|---|---|---|
| | (1) Distress Sympt. | (2) MH Talk | (3) Own MH Issues | (4) Therapy | (5) Any MH |
| Treated | 0.039 | 0.021 | -0.061 | -0.048 | -0.071 |
| | (0.037) | (0.029) | (0.052) | (0.055) | (0.049) |
| Control Mean | 0.703 | 0.845 | 0.761 | 0.376 | 0.807 |
| Control SD | 0.46 | 0.36 | 0.43 | 0.49 | 0.40 |
| Observations | 680 | 680 | 320 | 320 | 320 |
| LASSO Controls | ✓ | ✓ | ✓ | ✓ | ✓ |

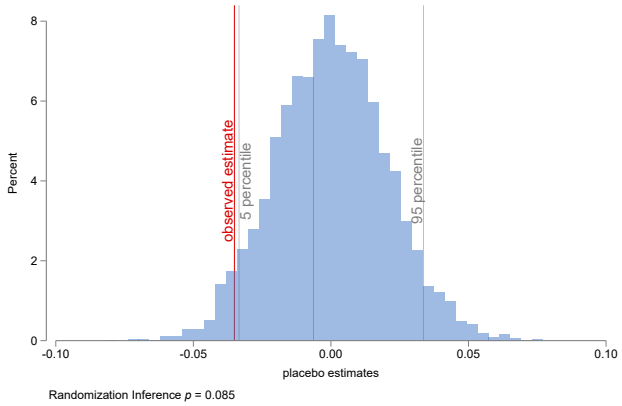
Notes: This table presents estimates of the short-run and long-run effects of the information intervention on outcomes related to personal and public stigma surrounding mental health. All regressions include covariates selected using the post double-selection LASSO procedure (Belloni et al. 2013), as indicated in the "LASSO Controls" row. Based on their predictive power for the outcome, and based on their association with the treatment indicator. Candidate controls include: age, sex, GPA, financial distress, scholarship status, relocation status, baseline mental health score, prior therapy use, openness to discussing mental health with acquaintances, and self-stigma. Columns (1)–(2) report short-run (SR) effects, measured in the baseline survey experiment using the full sample ($N = 680$). Column (1) captures whether respondents prefer to collaborate with a peer showing visible distress symptoms over one with a low GPA. Column (2) measures self-reported comfort discussing mental health in general, again relative to a low-GPA peer. Columns (3)–(5) present long-run (LR) effects using the follow-up sample of students who were still enrolled in 2025 ($N = 320$). Column (3) reports whether students discussed their own mental health struggles; column (4) captures whether they discussed therapy; and column (5) reflects whether they discussed either topic. These outcomes are designed to capture willingness to engage in potentially stigmatized conversations, thus reflecting both personal and perceived public stigma. Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SR = short run; LR = long run.

## B.8 Randomization Inference

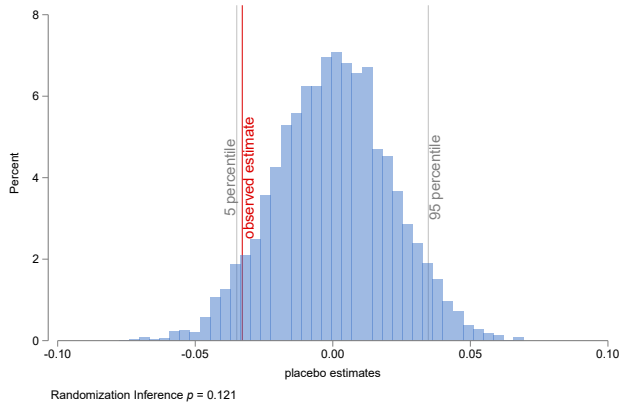
As a robustness check we perform randomization inference on our outcomes. In particular, we perform 5000 simulations in which we randomly assign a treatment indicator to different respondents and re-estimate the treatment effect on observed outcomes. If our estimate lies at either end of the distribution of 'placebo effects' (smaller then the 5th percentile or larger than the 95th percentile) we reject the null hypothesis of no treatment effect. Our p-values stemming from randomization inference are similar to the observed ones using conventional t-tests.

Figure B16: Randomization Inference on WTP and Donations

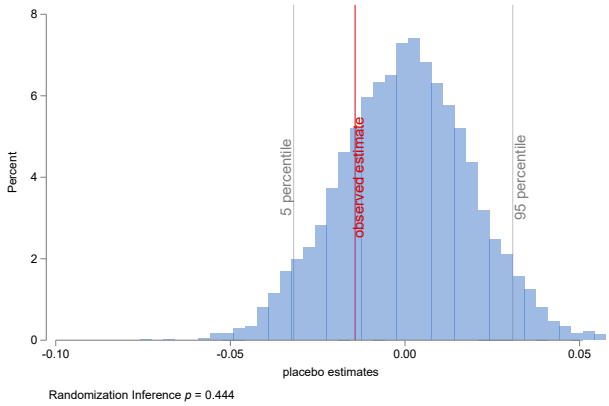
(a) WTP for therapy for self




(b) WTP for therapy for a friend



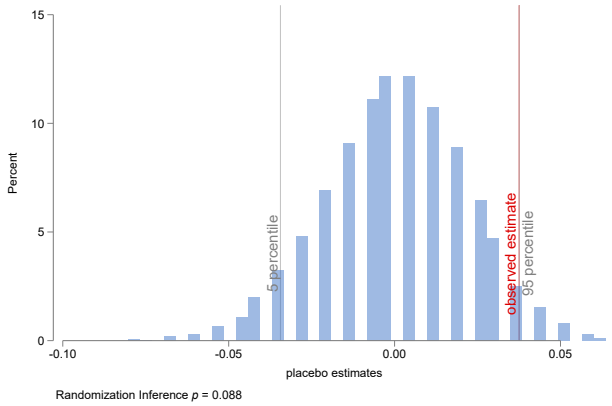

(c) Donations



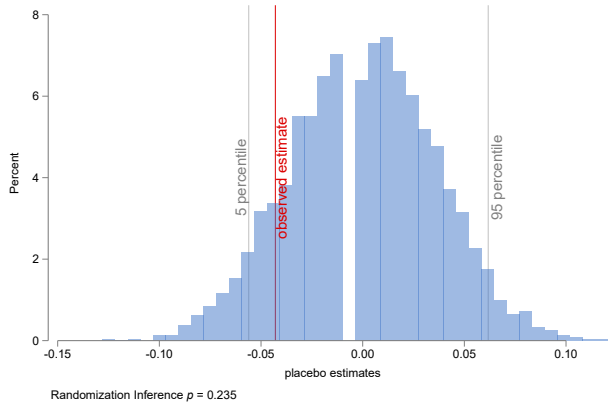

Notes: This figure shows randomization inference for the WTP and donations outcomes. Red vertical lines represent the location of the observed estimate. Gray vertical lines denote percentiles 5 and 95 of the distribution of placebo estimates.

Figure B17: Randomization Inference on Mentions of On-Campus Prfessional Help and Rankings

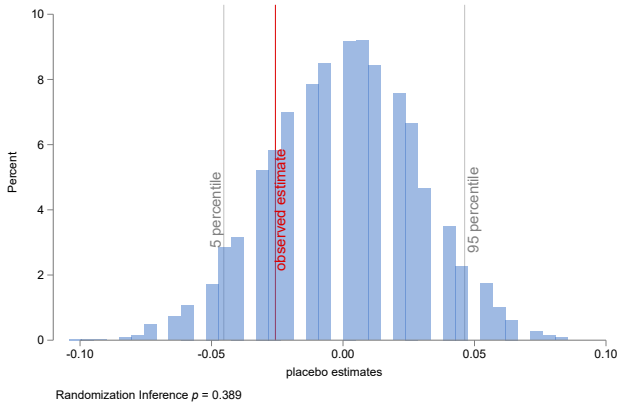
(a) Mention On-Campus Prof. Help in Advice




(b) Low GPA over MH symptoms

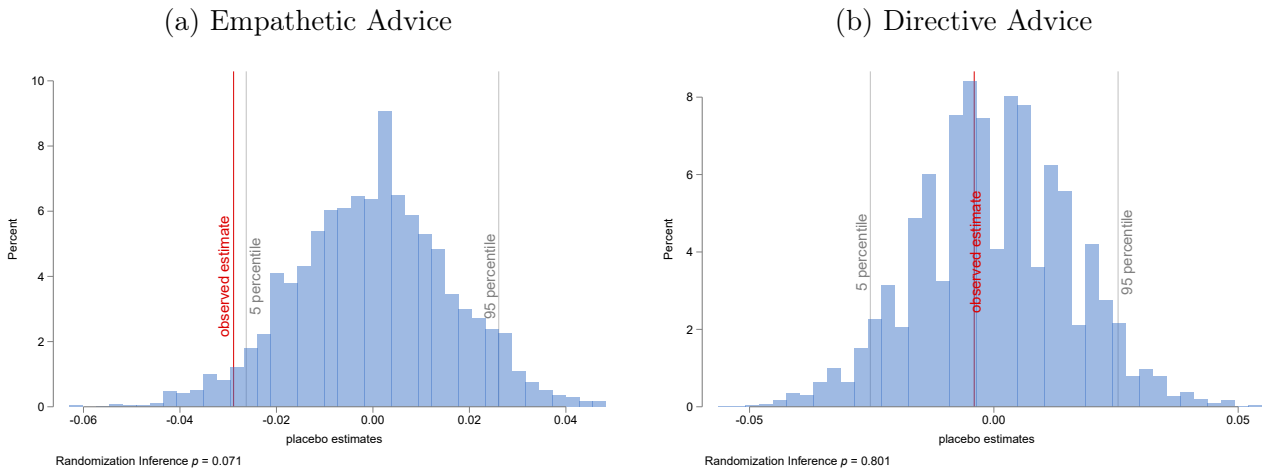



(c) Low GPA over MH Talk




Notes: This figure shows randomization inference for mentions of On-Campus professional help in advice, prefering Low GPA student over MH Talk & Low GPA student over MH Symptoms for group work. Red vertical lines represent the location of the observed estimate. Gray vertical lines denote percentiles 5 and 95 of the distribution of placebo estimates.

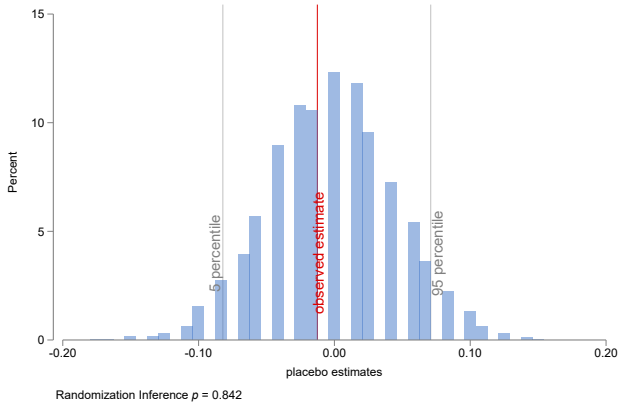
Figure B18: Randomization Inference on Empathetic and Directive Advice



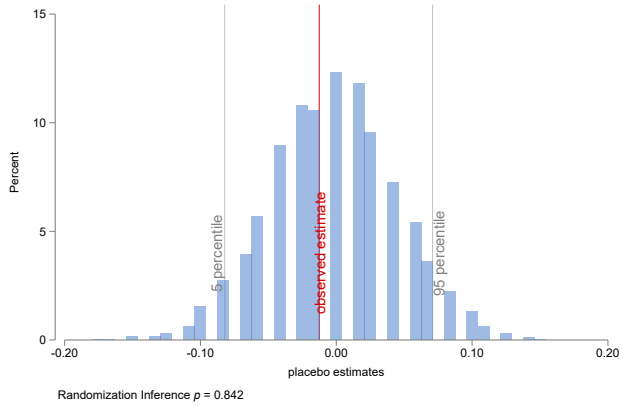

Notes: This figure shows randomization inference for the share of empathetic and directive advice components mentioned. Red vertical lines represent the location of the observed estimate. Gray vertical lines denote percentiles 5 and 95 of the distribution of placebo estimates.

Figure B19: Randomization Inference on Long Run outcomes - Usage and Hypothetical Usage

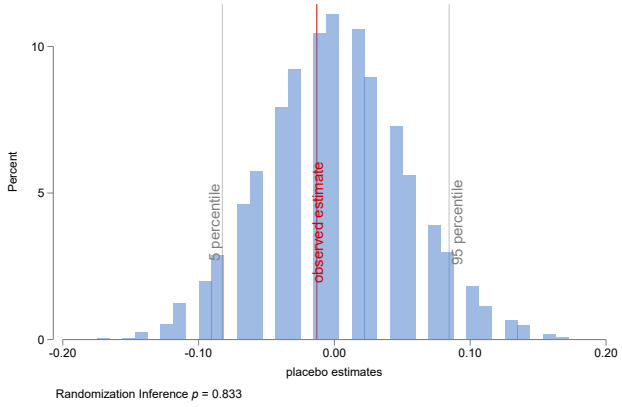
(a) Used On-Campus



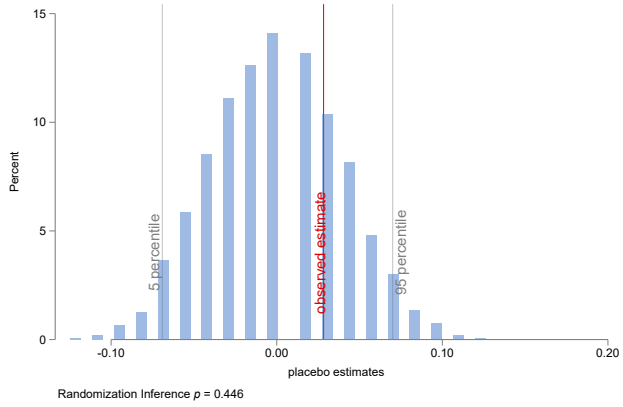

(b) Used Off-Campus




(c) Would Use On-Campus



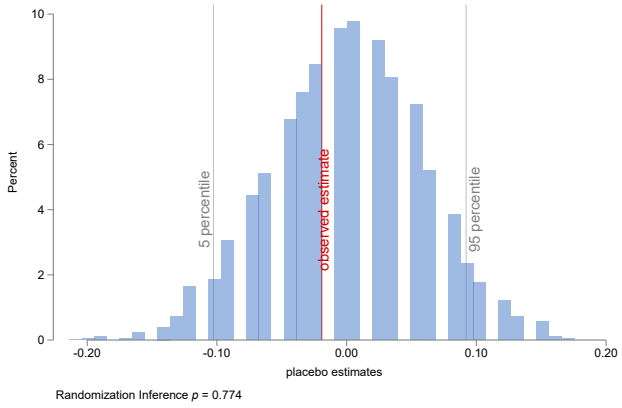

(d) Would Use Off-Campus



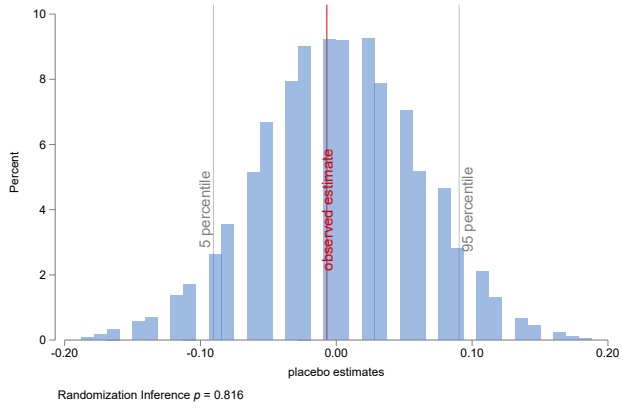

Notes: This figure shows randomization inference for self reported on- and off-campus mental health usage as well as hypothetical willingness to use these services. Red vertical lines represent the location of the observed estimate. Gray vertical lines denote percentiles 5 and 95 of the distribution of placebo estimates.

Figure B20: Randomization Inference on Long Run outcomes - Recommendations and Discussions

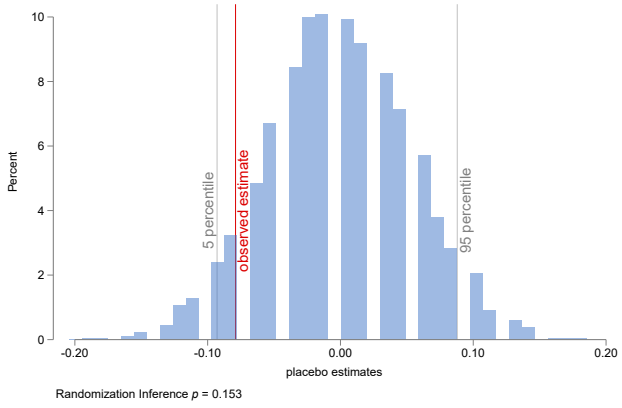
(a) Recommended On-Campus



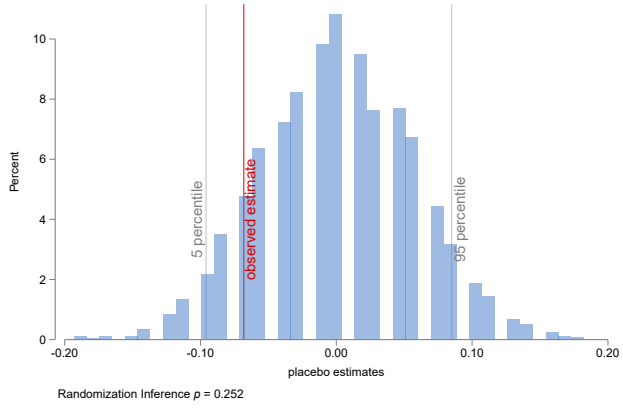

(b) Recommended Off-Campus




(c) Discussed Own Issues




(d) Discussed Experience



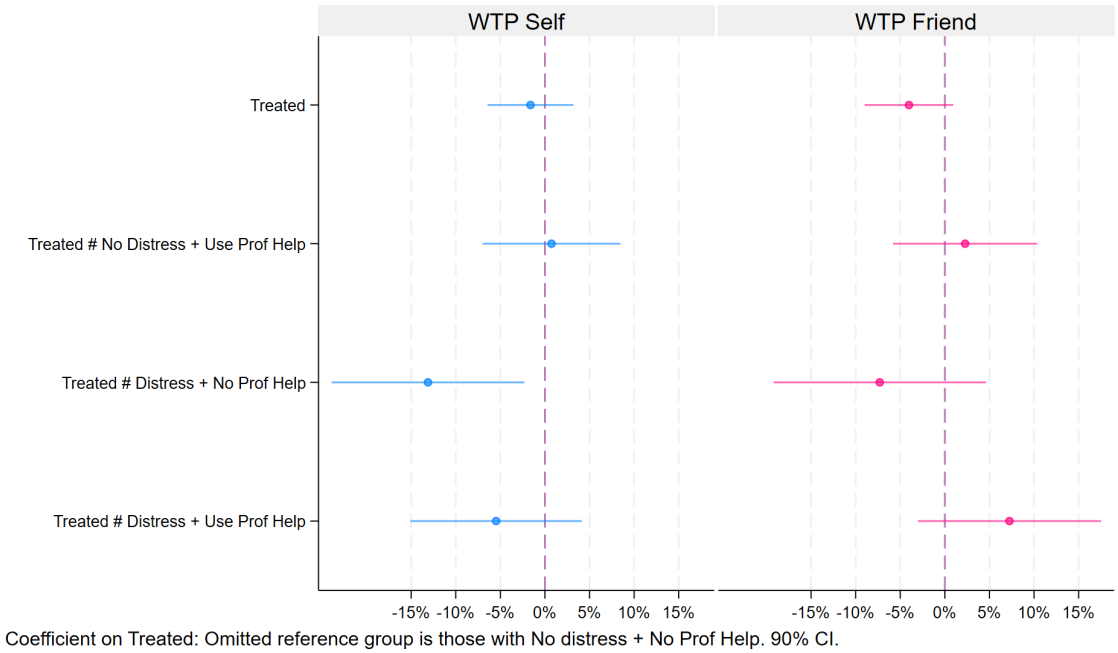

Notes: This figure shows randomization inference for self-reported recommendations of on and off-campus therapy services as well as self-reported discussions of own mental health problems and students' experience with mental health services. Red vertical lines represent the location of the observed estimate. Gray vertical lines denote percentiles 5 and 95 of the distribution of placebo estimates.

## B.9 Pre-Registered Heterogeneity

To explore heterogeneity in treatment effects, we interact the pooled treatment indicator with several key variables (focusing on the pre-registered specifications). We explore heterogeneity by groups defined by combinations of mental distress and professional help use: (0) No distress + no professional help (reference group), (1) No distress + professional help, (2) Distress + no professional help, (3) Distress + professional help. Figure B27a through Figure B27d, Figure B28a through Figure B28d, and Figure B29a through Figure B29d illustrate heterogeneous treatment effects by distress, GPA, and Stigma Index 1, respectively.

In Figure B21, we observe that for the group that might be the main target of potential interventions (Distress + No Professional Help), we observe a significant negative effect on own WTP and a smaller not significant negative effect on the WTP for a friend. While we can not robustly show the main driver of this negative updating, we conjecture that this may be aligned with the substitution effect of switching to free on-campus therapy we have discussed above. At the same time, this also highlights some limitations of using WTP measures for a private service analogue in a setting where free services are provided, limiting out discussion of this measure as an imperfect proxy for the overall demand.

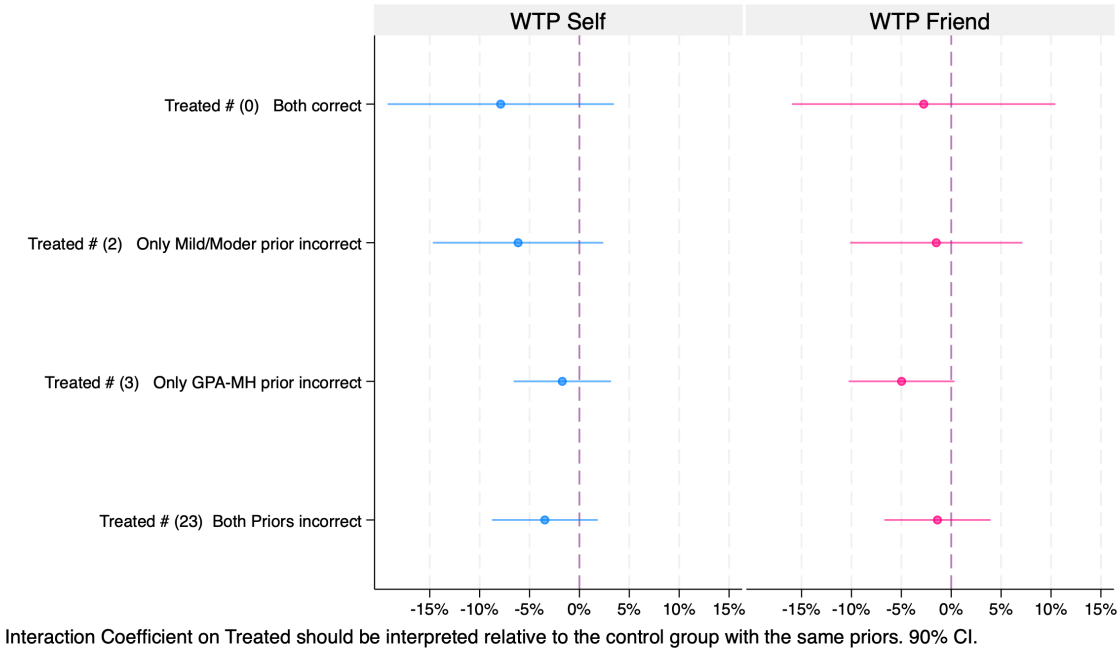
Figure B21: Heterogeneity by Distress and Professional Help Usage




Notes: This figure shows estimates and 90% confidence intervals for willingness to pay. The top-most estimates report the ATE, while the three bottom-most coefficients show heterogeneous effects across subgroups of the population of interest. In particular, we estimate heterogeneity on those who are not in distress but have used professional mental health help, those who are in distress but have not used professional mental health help and those who are in distress and have used professional mental health help.

In Figure B22, we test whether the treatment effects differ based on the accuracy of students' prior beliefs. Specifically, we examine the following groups: (0) Both priors correct (reference group), (1) Prior 2 incorrect only, (2) Prior 3 incorrect only, (3) Both priors incorrect. We observe no heterogeneity highlighting that our information intervention might have also carried the salience effect promoting subjects to think more about the shared facts beyond just updating on specific facts or statements. This is one of the limitations we face in having treated subjects with multiple facts concurrently due to constraints on power, but in future studies it might be worth exploring the differential effects of individual statements and different means of delivering them.

Figure B22: Heterogeneity by Incorrect Priors




Notes: This figure shows estimates and 90% confidence intervals for willingness to pay. In particular, we estimate $Y_i = \alpha + \beta_M$ ($\texttt{Treated}_i \times \texttt{OnlyMildIncorrect}_i$) $+ \beta_G$ ($\texttt{Treated}_i \times \texttt{OnlyGPAIncorrect}_i$) $+ \beta_B$ ($\texttt{Treated}_i \times \texttt{BothIncorrect}_i$) $+ \beta_N$ ($\texttt{Treated}_i \times \texttt{NonIncorrect}_i$) $+ \gamma_M$ $\texttt{OnlyMildIncorrect}_i + \gamma_G$ $\texttt{OnlyGPAIncorrect}_i + \gamma_B$ $\texttt{BothIncorrect}_i + \gamma_N$ $\texttt{NonIncorrect}_i + \varepsilon_i$, where $Y_i$ is the outcome of interest, `OnlyMildIncorrect` is an indicator equal to 1 if the respondent only answered the "Mild/Moderate"-prior question incorrectly, `OnlyGPAIncorrect` is an indicator equal to 1 if the respondent only answered the "GPA-MH"-prior question incorrectly, `BothIncorrect` is an indicator equal to 1 if the respondent answered both the "Mild/Moderate"- and the "GPA-MH"-prior questions incorrectly, and `NonIncorrect` is an indicator equal to one if the respondent answered both prior questions correctly. `Treated` is an indicator equal to 1 if the respondent is assigned to either of the treatment groups and 0 if they are assigned to the control group.

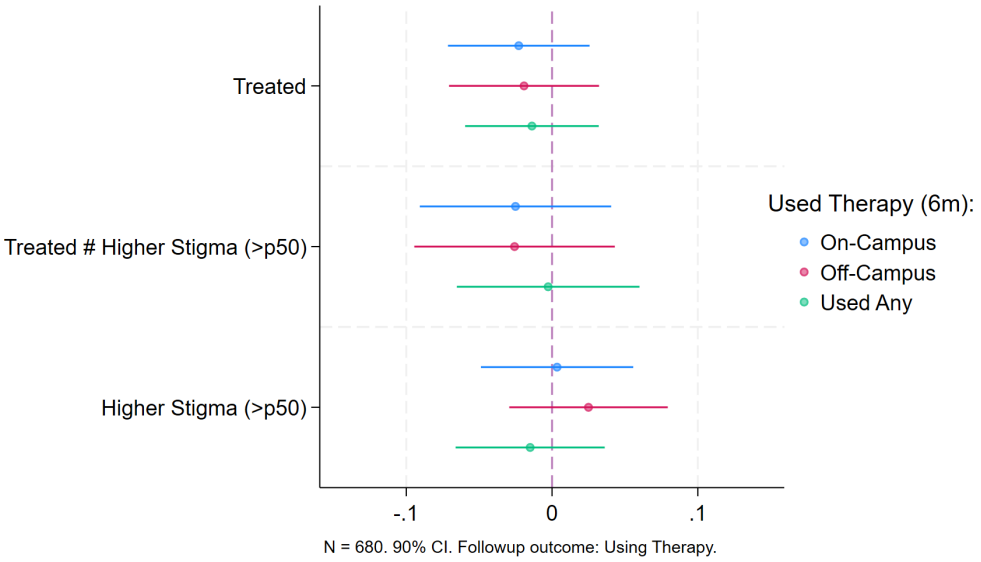
### B.9.1 Heterogeneity by Stigma

**Other notes/comments on suggestive mechanisms:**

Baseline by stigma:

We also examine how treatment effects vary by levels of the stigma index (Figure B23). Marginally insignificant negative coefficients on the interaction term provide suggestive evidence that the effects are suggestively stronger for more stigmatized individuals in the willingness-to-pay outcomes.

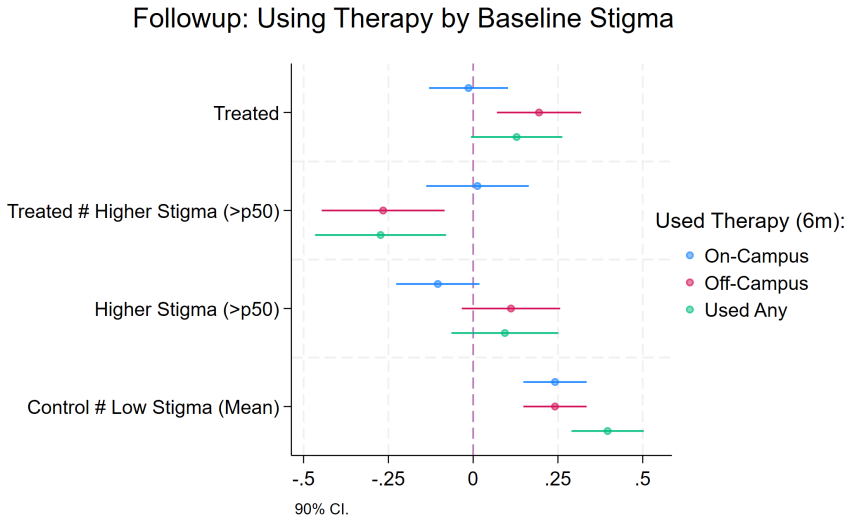
Figure B23: Treatment Effect Heterogeneity for Initial-Survey WTP by Stigma




Notes: This figure shows estimated treatment effect and their interactions with underlying stigma measure (above/below median) and 90% confidence intervals from the regressions for WTP for oneself, a friend, and the percentage of earnings donated for another participant's therapy subscription on the treatment indicator, stigma above median (using the standardized values of the PCA-derived stigma index), and their interaction.

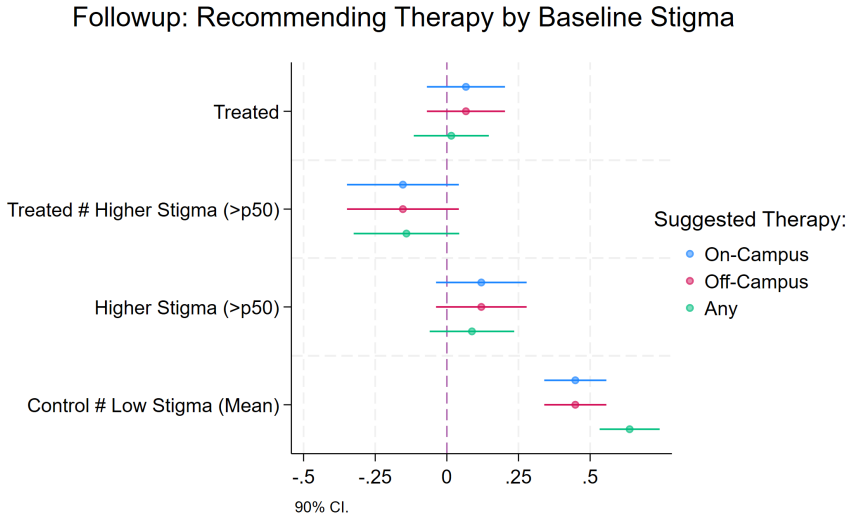
Follow-up by Stigma:

Figure B24: TE Heterogeneity on Using Therapy by Prior Therapy Use & Baseline Stigma



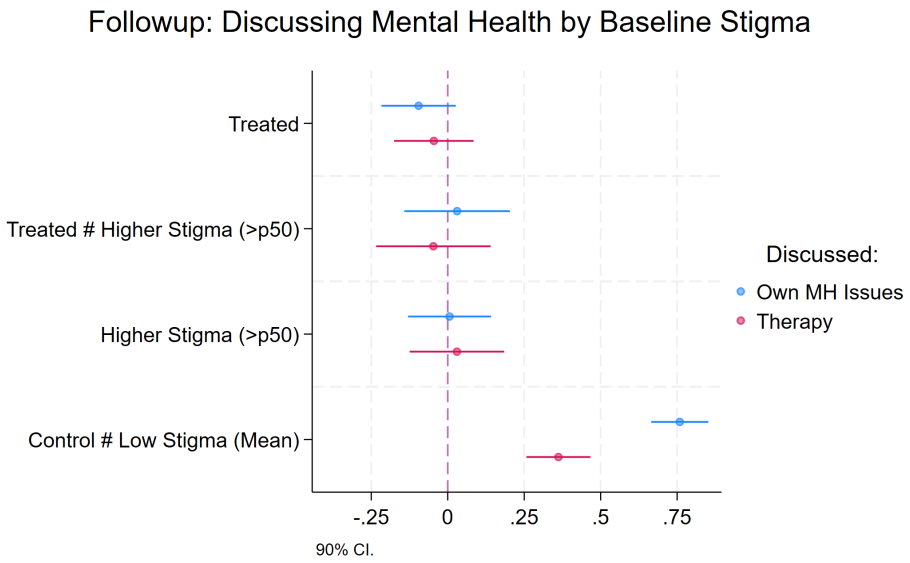

Notes: This figure shows estimated treatment effects and their interactions with baseline stigma levels (above/below median) and 90% confidence intervals from regressions on therapy use at the follow-up. Outcomes include self-reported use of on-campus, off-campus, or any type of therapy. The stigma measure is constructed using the standardized values of a PCA-derived stigma index and is dichotomized at the median. Regressions include indicators for treatment status, high stigma, and their interaction.

Figure B25: TE Heterogeneity on Recommendations by Baseline Stigma



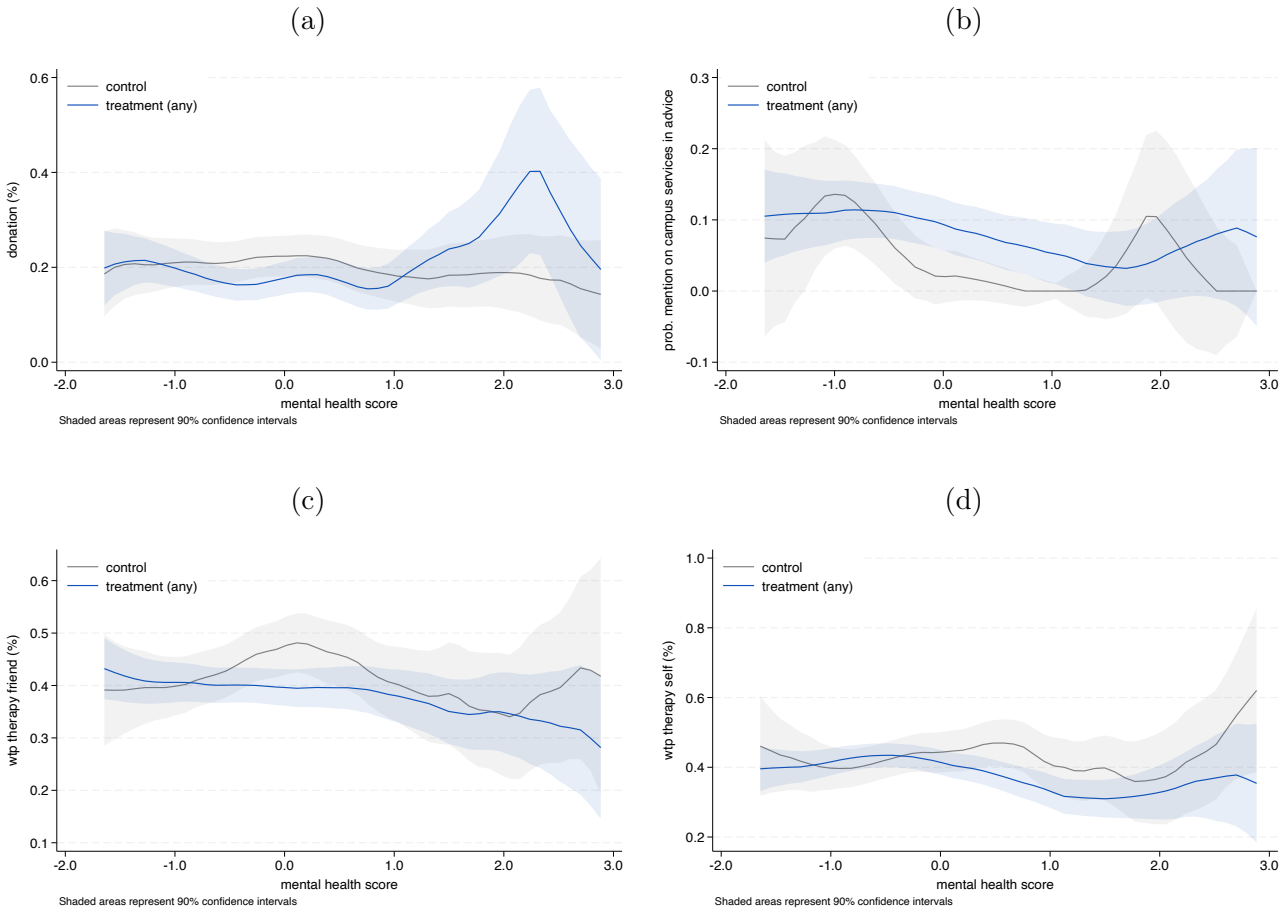

Notes: This figure shows estimated treatment effects and their interactions with baseline stigma levels (above/below median) and 90% confidence intervals on recommending therapy at the follow-up survey. Outcomes include whether respondents suggested on-campus, off-campus, or any form of therapy to others. The stigma variable is constructed from a standardized PCA-based index and dichotomized at the median. The regression includes main effects for treatment status and high stigma, as well as their interaction.

Figure B26: TE Heterogeneity on Discussing MH Issues by Prior Therapy Use & GPA




Notes: This figure shows estimated treatment effects and their interactions with baseline stigma levels (above/below median) and 90% confidence intervals from regressions on discussing mental health at the follow-up survey. The stigma measure is based on a standardized PCA-derived index and is dichotomized at the median. The regressions include main effects for treatment assignment, high stigma, and their interaction.

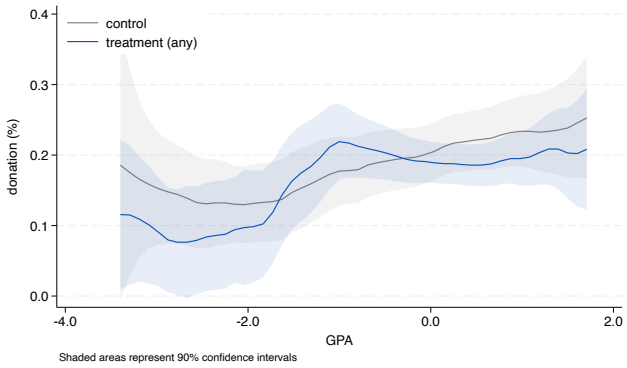
Figure B27: Heterogeneous Treatment Effects



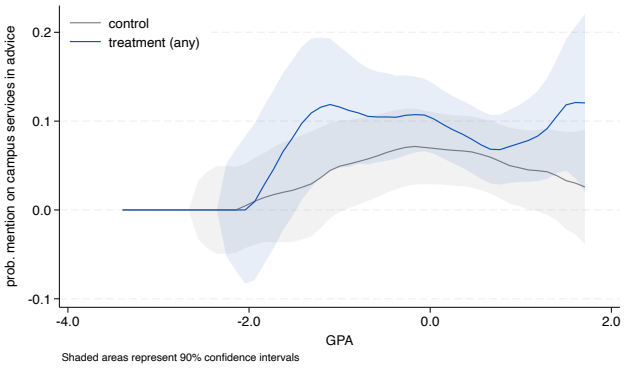

Notes: This figure displays heterogeneous effects of distress on multiple outcomes. Shaded areas represent 90% confidence intervals.

Figure B28: Heterogeneous Treatment Effects

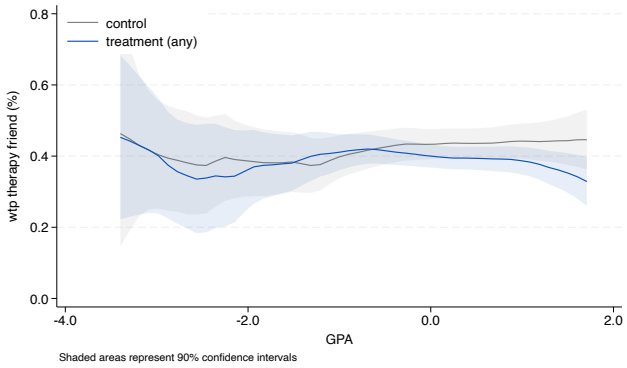
(a)



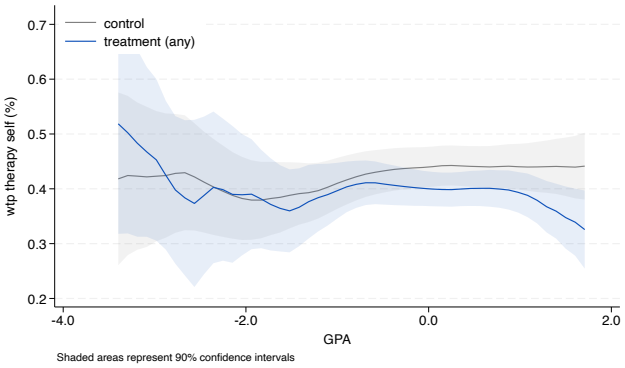

(b)




(c)



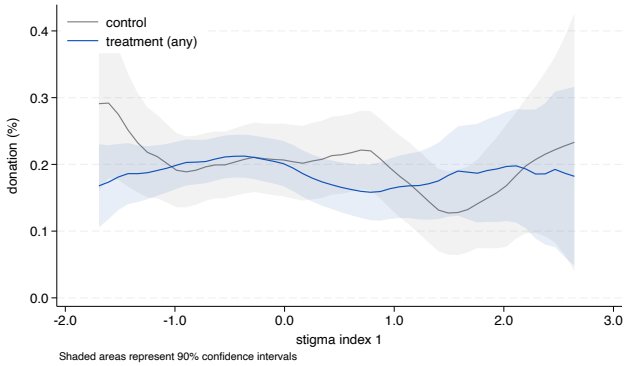

(d) Effect of Treatment on WTP for Therapy for Self



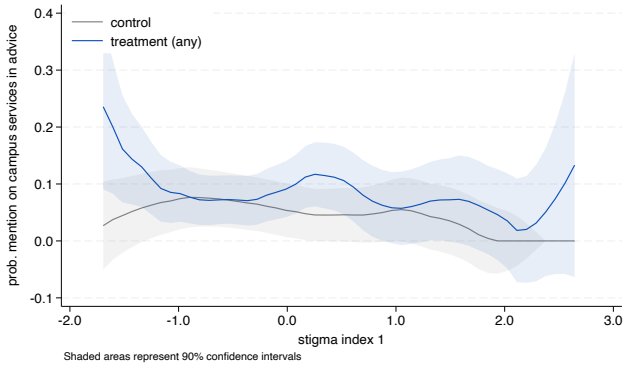

Notes: This figure displays heterogeneous effects of GPA on multiple outcomes. Shaded areas represent 90% confidence intervals.

Figure B29: Heterogeneous Treatment Effects on Stigma Index 1

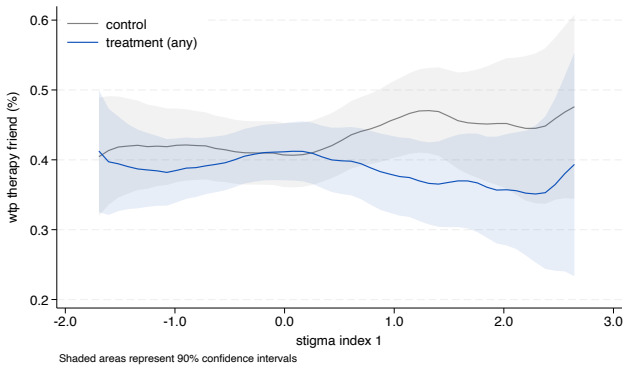
(a) Effect of Treatment on Donations



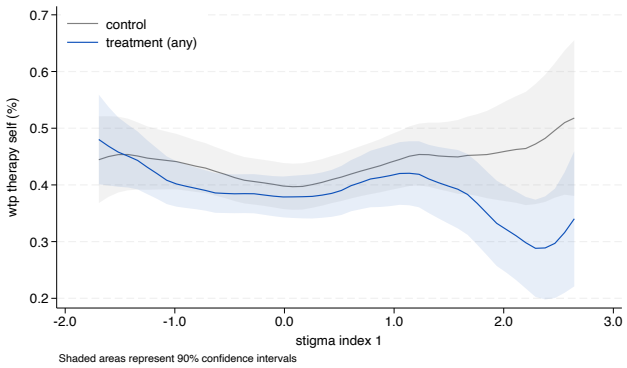

(b) Effect of Treatment on Mentioning Campus Services




(c) Effect of Treatment on WTP for Therapy for a Friend




(d) Effect of Treatment on WTP for Therapy for Self




Notes: This figure displays heterogeneous effects of distress on multiple outcomes. Shaded areas represent 90% confidence intervals.

# C Appendix: Theoretical Framework

This section extends Grossman (1972)'s "health-capital" framework to a mental-health context in which going to therapy improves both the agent's internal state and her social relationships. By embedding these dual benefit channels into the standard intertemporal optimization problem, we obtain a rule that links therapy demand to its expected gains and its full cost—including monetary fees and stigma burdens.

**Step 1. Agent's objective.** At each instant $t \in [0, T]$ the agent decides whether to go to a therapy session, $D(t) \in \{0, 1\}$. Flow utility depends on three arguments: mental-health stock $H(t)$, social-capital stock $S(t)$, and consumption of a numéraire good $Z(t)$. Going to therapy also triggers two non-monetary disutility components: the *self-stigma* term $S_s$, capturing the internal shame of recognizing a need for help, and the *perceived-stigma* term $S_p$, capturing the discomfort of anticipating negative judgment by others. Both stigma terms are incurred only when $D(t) = 1$.

The intertemporal problem is therefore

$$\max_{D(\cdot)} \int_0^T \big[u_H\big(H(t)\big) + u_S\big(S(t)\big) + Z(t) - D(t)\big(S_s + S_p\big)\big]\, e^{-\rho t}\, dt,$$

with discount rate $\rho > 0$. The utility specification is additively separable, and the linear numéraire good component $Z(t)$ implies a constant marginal utility of consumption, so every dollar—or unit of stigma—reduces flow utility one for one.

The agent earns a constant income stream $Y$. When she goes to therapy at time $t$ ($D(t) = 1$), she pays the fee $p_T$; if she doesn't ($D(t) = 0$), she pays nothing. Hence consumption of the numéraire good satisfies $Z(t) = Y - D(t)p_T$.

**Step 2. Dynamics of mental health and social capital.**

Both mental health $H(t)$ and social capital $S(t)$ are treated as capital-like stocks that erode over time in the absence of active investment. Natural psychological wear—stress, fatigue, or negative rumination—reduces $H(t)$ at a constant rate $\delta_H > 0$, while social disengagement and neglected relationships reduce $S(t)$ at rate $\delta_S > 0$. These depreciation rates are assumed constant to keep the model tractable and to mirror the original Grossman setup.

Therapy is modeled as the sole mechanism that can raise either stock above its depreciating path. When the agent decides to go to a session at time $t$ ($D(t) = 1$), her well-being may improve, but success is uncertain. Conditional on attending, therapy is effective in improving well-being with probability $\pi_1 \in (0, 1)$. Specifically, a successful session delivers discrete jumps of fixed size: $G_H > 0$ units to mental health and $G_S > 0$ units to social capital. These parameters summarize

the average psychological and relational gains from a therapy session and are taken as exogenous and time-invariant.

Let $\pi_{D(t)}$ denote the success probability that actually applies at time $t$: it equals $\pi_1$ if the agent goes to therapy ($D(t) = 1$) and zero if she does not ($D(t) = 0$).[49] In expected-value terms, the law of motion for each stock is

$$\dot{H}(t) = \pi_{D(t)} G_H - \delta_H H(t), \qquad \dot{S}(t) = \pi_{D(t)} G_S - \delta_S S(t).$$

The first term on the right captures the expected upward jump from a session, which is positive only when therapy is chosen; the second term captures continuous depreciation. This functional form preserves the capital-accumulation logic of the Grossman model while allowing therapy to provide discrete, probabilistic boosts to both mental and social well-being.

**Step 3. Solution of the dynamic problem.**

To solve the continuous-time control problem we apply the Maximum Principle, which introduces shadow prices for the two state variables. Let $\lambda_H(t)$ denote the *current-value co-state* for mental health and $\lambda_S(t)$ the co-state for social capital; each measures the marginal utility value of an extra unit of the corresponding stock. Using these shadow prices, the current-value Hamiltonian is

$$\mathcal{H} = u_H(H) + u_S(S) + Y - D\big(p_T + S_s + S_p\big) + \lambda_H\big(\pi_D G_H - \delta_H H\big) + \lambda_S\big(\pi_D G_S - \delta_S S\big).$$

The first two terms give instantaneous utility from the stocks and consumption, the middle term subtracts the monetary and stigma costs if therapy is chosen, and the last two terms add the capital gains valued at their shadow prices.

Because the control $D$ is binary, optimality at an instant reduces to comparing the Hamiltonian under $D = 1$ with that under $D = 0$. The agent will go to therapy when the difference $\Delta\mathcal{H} = \mathcal{H}(D = 1) - \mathcal{H}(D = 0)$ is non-negative. Substituting $D = 1$ and $D = 0$ and simplifying yields the *instantaneous therapy condition*

$$\pi_1\big(\lambda_H G_H + \lambda_S G_S\big) \;\geq\; p_T + S_s + S_p. \tag{IC}$$

The left-hand side is the expected marginal benefit of a session: success probability $\pi_1$ multiplied by the shadow-value gain from the discrete boosts $G_H$ and $G_S$. The right-hand side is the full cost of going to therapy—the fee plus the self- and perceived- stigma burdens. Therapy is chosen at time $t$ precisely when inequality (IC) holds.

[49] Setting $\pi_0 = 0$ means no improvement occurs when $D(t) = 0$.

**Step 4. From co-states to a decision rule.** The shadow prices evolve according to

$$\dot{\lambda}_H = \rho\lambda_H - u'_H(H), \qquad \dot{\lambda}_S = \rho\lambda_S - u'_S(S).$$

At the instant the agent weighs "go" versus "skip," we set $\dot{\lambda}_i = 0$, which yields the stationary shadow prices

$$\lambda_H = \frac{u'_H(H)}{\rho}, \qquad \lambda_S = \frac{u'_S(S)}{\rho}.$$

For clarity, define

$$B_H \equiv \frac{u'_H(H)\, G_H}{\rho}, \qquad B_S \equiv \frac{u'_S(S)\, G_S}{\rho},$$

so $B_H$ (respectively $B_S$) is the present value of the marginal utility gain from a one-unit jump in mental health (social capital) produced by a successful therapy session.

*Therapy demand condition.* Substituting these expressions into the instantaneous condition (IC) gives the compact rule

$$\boxed{\pi_1\big(B_H + B_S\big) \;\geq\; p_T + S_s + S_p}$$

**Interpretation.**

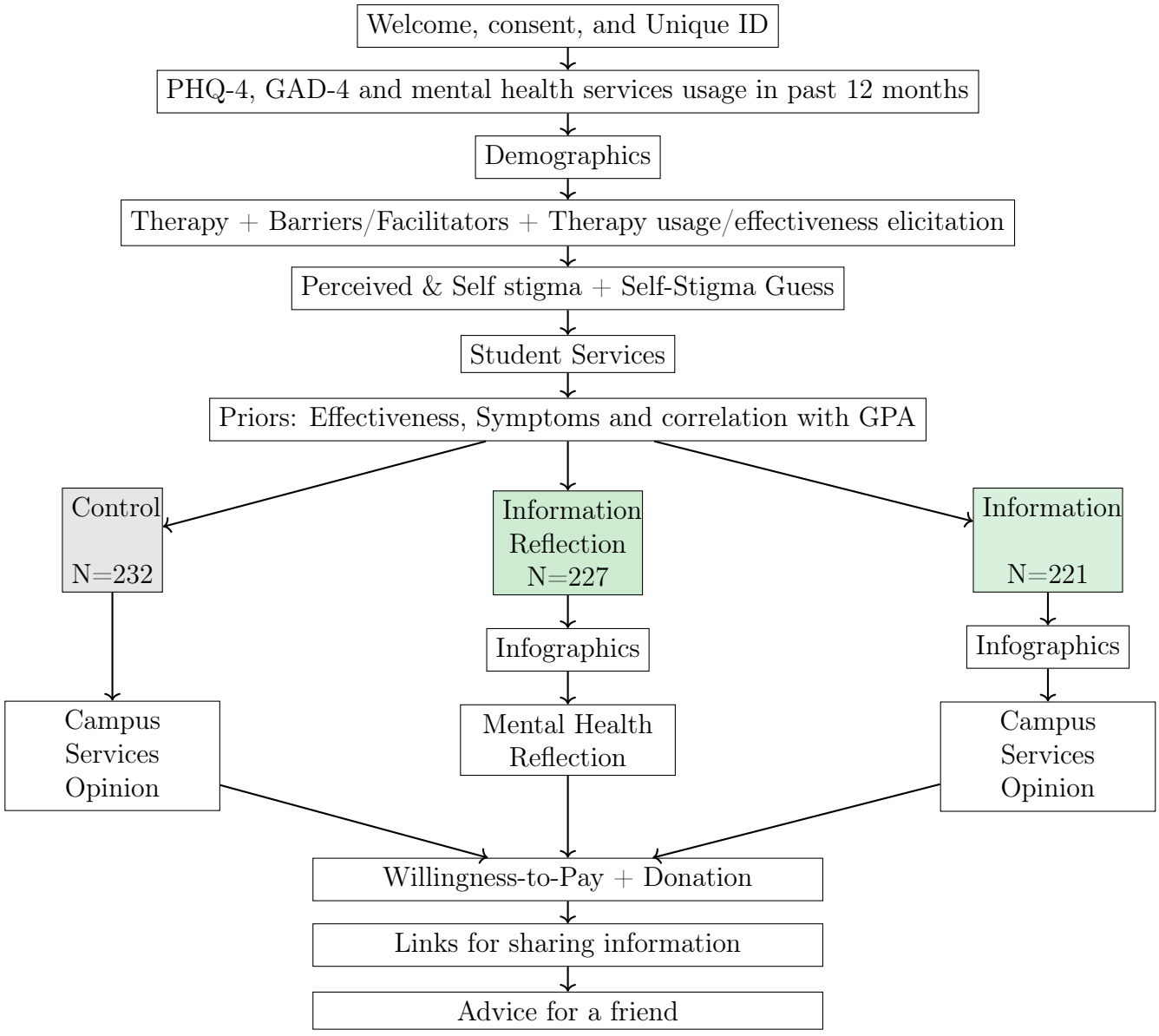
- $\pi_1$ is the likelihood that a session actually produces a meaningful improvement in the agent's well-being—higher clinical effectiveness or better therapist fit raises $\pi_1$ and thus increases demand.
- $B_H$ is the present-value *benefit* from the mental-health boost $G_H$; it grows with the marginal utility of feeling better and with the size of the therapeutic gain.
- $B_S$ is the analogous *benefit* from a stronger social network; greater relational gains or higher value placed on social connections both raise $B_S$.
- The right-hand side is the total *cost* of attending therapy in the moment: the out-of-pocket fee $p_T$ plus the self-stigma $S_s$ and perceived stigma $S_p$.

The agent chooses to go to therapy when the expected, present-valued benefits for mental and social well-being exceed the full monetary and psychological costs. Higher effectiveness, larger well-being benefits, or lower stigma tilt the balance toward treatment; higher fees or greater stigma tilt it away.

# D Appendix: Further Analyses

## D.1 Survey flowchart

Figure B30: Survey Flow




Notes: This figure depicts the survey flow.

## D.2 Heterogeneity Analysis Specifications

### Heterogeneity by Incorrect Beliefs

We test whether the treatment effects differ based on the accuracy of students' prior beliefs. Specifically, we examine the following groups: (0) Both priors correct (reference group), (1) Prior 2 incorrect only, (2) Prior 3 incorrect only, (3) Both priors incorrect.

The regression model is:

$$Y_i = \alpha + \sum_{j=1}^{3} \delta_j BeliefGroup_{ij} + \sum_{j=1}^{3} \phi_j (InfoTreatment_i \cdot BeliefGroup_{ij}) + X_i' \gamma + \varepsilon_i,$$

where $BeliefGroup_{ij}$ is an indicator variable for individual $i$ being in belief group $j$ (e.g., "Prior 2 incorrect only"), $InfoTreatment_i \cdot BeliefGroup_{ij}$ is the interaction term capturing the differential treatment effect for each belief group $j$, $\phi_j$ represents the difference in treatment effects for each group relative to the baseline (both priors correct).

**Heterogeneity by Distress and Professional Help Use**

We explore heterogeneity by groups defined by combinations of mental distress and professional help use: (0) No distress + no professional help (reference group), (1) No distress + professional help, (2) Distress + no professional help, (3) Distress + professional help.

The regression model is:

$$Y_i = \alpha + \sum_{j=1}^{3} \delta_j DistressGroup_{ij} + \sum_{j=1}^{3} \phi_j (InfoTreatment_i \cdot DistressGroup_{ij}) + X_i' \gamma + \varepsilon_i,$$

where $DistressGroup_{ij}$ is an indicator variable for individual $i$ being in distress group $j$ (e.g., "Distress + no professional help"), $InfoTreatment_i \cdot DistressGroup_{ij}$ is the interaction term capturing the differential treatment effect for each distress/help group $j$, $\phi_j$ represents the difference in treatment effects for each group relative to the baseline (no distress + no professional help).

**Heterogeneity by Stigma Index**

We also examine how treatment effects vary by levels of the stigma index. The model is specified as:

$$Y_i = \alpha + \beta_1 InfoTreatment_i + \delta StigmaIndex_i + \phi (InfoTreatment_i \cdot StigmaIndex_i) + X_i' \gamma + \varepsilon_i,$$

where $StigmaIndex_i$ is a continuous variable representing individual $i$'s stigma index score, $InfoTreatment_i \cdot StigmaIndex_i$ is the interaction term capturing how treatment effects vary with levels of stigma, $\phi$ measures the marginal change in treatment effect per unit increase in the stigma index.

Each specification allows us to analyze differential treatment effects. In the first specification, $\phi_j$ quantifies whether treatment effects vary based on prior beliefs, relative to those with both priors correct. In the second one, $\phi_j$ captures how treatment effects differ for combinations of mental

distress and professional help use, relative to the baseline group (no distress + no professional help). In the third specification, $\phi$ indicates whether treatment effects are stronger or weaker depending on the level of stigma.

These models provide insights into whether the intervention's effectiveness is moderated by key characteristics of participants.

## D.3 Incentivized bonus questions

The eight bonus questions included: (1) guessing the percentage of "Yes" responses to the question regarding therapy usage in the past 12 months which was compared to the actual calculated percentage; (2) guessing the percentage of "Yes" responses to the question on willingness to share therapy information, which was similarly compared to the actual percentage; (3) responding "22" to a specific survey question, which earned the bonus if correct; (4) guessing the percentage of "Agree" responses (including "Strongly Agree," "Agree," and "Somewhat Agree") to a question on self-stigma, validated against the computed percentage; (5) answering "Yes" to a question about therapy effectiveness, which directly earned the bonus; (6) answering "Yes" to a question about therapy effectiveness for mild-to-moderate conditions , which similarly earned the bonus; (7) categorizing the correlation between mental health scores and grade point averages into predefined categories such as "Better" or "Much Worse", with correctness determined by the computed correlation; and (8) providing open-ended advice on a specific topic, where responses deemed "Very useful" during review earned the bonus.

## D.4 Stigma Index

In the context of the study we seek to create a unified measure of stigma taking into account three distinct dimensions.

- **Perceived Public Stigma**: This dimension is defined by three variables that measure the perception fo stigma of other students, professors, and parents.
- **Self-Stigma**: This dimension corresponds to a variable that measures the number of people out of 100 that would feel disappointment for experiencing any mental health issues.
- **Personal Stigma**: This third dimension corresponds to two dummy variables measuring preference of a lower GPA over experiencing mental health symptoms and talking about mental healthy issues.

The following tables provides a comprehensive description of the variables present across the 3 dimensions. From these classifications we aim to implement not only a PCA to generated an index, but also a weighted average.

Table B15: Mental Health Stigma Variables

| | **Definition** |
|---|---|
| **Perceived Public Stigma** | |
| From students | Percentage of students that the respondent believes would view a student negatively for experiencing mental health issues like anxiety or depression. |
| From professors | Percentage of professors that the respondent believes would view a student negatively for experiencing mental health issues like anxiety or depression. |
| From parents | Percentage of student parents that the respondent believes would view a student negatively for experiencing mental health issues like anxiety or depression. |
| **Self-Stigma** | |
| Self-stigma | Respondent's estimate of how many out of 100 students would feel disappointed in themselves if they had a mental health issue. |
| **Personal Stigma** | |
| Low GPA over MH symptoms | Dummy variable where it has a value of 1 if the respondent ranked a student with a low GPA as preferred as a class project teammate rather than a student experiencing mental health distress; 0 otherwise. |
| Low GPA over MH talk | Dummy variable where it has a value of 1 if the respondent ranked a student with a low GPA as preferred as a class project teammate rather than a student openly talking about mental health issues; 0 otherwise. |

Notes: This table shows the definition of variables used as inputs for constructing our stigma index.

### D.4.1 Weighted Average

To create a unified measure of mental health stigma, we developed indices that account for three distinct dimensions of stigma: *Perceived Public Stigma*, *Self-Stigma*, and *Personal Stigma*. Each dimension was represented by relevant variables described in the table above, and the methodology for index construction is outlined below.

Table B16: Summary Statistics for Stigma Dimensions (Mean and Median Thresholds)

| Dimension | Threshold Type | Mean | Std. Dev. | Min–Max |
|---|---|---|---|---|
| Perceived Public Stigma | Mean-based | 0.428 | 0.396 | 0–1 |
| | Median-based | 0.492 | 0.397 | 0–1 |
| Self-Stigma | Mean-based | 0.513 | 0.500 | 0–1 |
| | Median-based | 0.469 | 0.499 | 0–1 |
| Personal Stigma | Mean-based | 0.204 | 0.337 | 0–1 |
| | Median-based | 0.204 | 0.337 | 0–1 |

Notes: Perceived Public Stigma reflects stigma perceptions from students, professors, and parents. Self-Stigma measures internalized stigma based on perceived social disappointment in experiencing mental health issues. Personal Stigma captures preferences for GPA trade-offs over experiencing or discussing mental health concerns.

Table B17: Summary Statistics for Composite Stigma Indices

| Index | Standardization | Mean | Std. Dev. | Min–Max |
|---|---|---|---|---|
| Composite Index (Mean-based) | Raw | -0.107 | 1.847 | -2.787 – 4.573 |
| | Standardized | -0.055 | 0.948 | -1.431 – 2.348 |
| Composite Index (Median-based) | Raw | -0.117 | 1.836 | -2.879 – 4.498 |
| | Standardized | -0.061 | 0.958 | -1.502 – 2.347 |

Notes: Composite indices classify stigma levels using mean- and median-based thresholds. Perceived public stigma is assessed separately for students, professors, and parents. Self-stigma is binarized based on a defined threshold, while personal stigma captures preferences for GPA trade-offs over experiencing or discussing mental health issues.

For each variable within the stigma dimensions, we created binary indicators based on whether the value exceeded the dimension-specific mean or median. Perceived public stigma was assessed separately for students, professors, and parents by comparing their reported percentages against predefined thresholds. Specifically, for students, values greater than 26.35 (mean) or 20 (median) indicated perceived public stigma, while for professors, the corresponding thresholds were greater than 26.61 (mean) or 20 (median). For parents, the thresholds were set at greater than 40.11 (mean) or 39.5 (median). Self-stigma was binarized using a threshold of greater than 49.79 (mean) or 50 (median). Finally, personal stigma was represented by two binary variables: the preference for a lower GPA over experiencing mental health symptoms and the preference for a lower GPA over talking about mental health issues.

**Aggregation Within Dimensions**

For each stigma dimension, aggregated measures were computed based on the proportion of satisfied binary indicators. Perceived public stigma was calculated as the mean of the three binary indicators corresponding to students, professors, and parents. Self-stigma, being a single variable,

was directly represented by its binary indicator. Personal stigma was aggregated as the mean of the two binary indicators reflecting preferences related to GPA and mental health concerns.

To account for potential variation across treatment groups, the aggregated shares for each dimension were standardized. This was achieved by centering the values around the control group's mean and dividing by the standard deviation.

Table B18: Correlations Between Stigma Dimensions and Composite Indices

| Dimension | Perceived Public Stigma | Self-Stigma | Personal Stigma | Composite Index |
|---|---|---|---|---|
| *Mean-Based Composite Index* | | | | |
| Perceived Public Stigma | 1.0000 | 0.2455 | -0.0177 | 0.6612 |
| Self-Stigma | 0.2455 | 1.0000 | 0.0255 | 0.6863 |
| Personal Stigma | -0.0177 | 0.0255 | 1.0000 | 0.5248 |
| Composite Index | 0.6612 | 0.6863 | 0.5248 | 1.0000 |
| *Median-Based Composite Index* | | | | |
| Perceived Public Stigma | 1.0000 | 0.2550 | -0.0293 | 0.6695 |
| Self-Stigma | 0.2550 | 1.0000 | -0.0041 | 0.6805 |
| Personal Stigma | -0.0293 | -0.0041 | 1.0000 | 0.5053 |
| Composite Index | 0.6695 | 0.6805 | 0.5053 | 1.0000 |

Notes: Perceived Public Stigma reflects beliefs about how students, professors, or parents view mental health issues. Self-Stigma captures internalized negative attitudes toward one's own mental health. Personal Stigma represents preferences related to GPA trade-offs over experiencing or discussing mental health concerns. The Composite Index combines these stigma dimensions into standardized measures.

**Weighted Average Stigma Index Amongst Distressed and Non-Distressed Individuals**

Table B19: Summary of Composite Stigma Indices by Mental Distress Groups

| Mental Distress Group | Mean (Mean-Based Index) | Mean (Median-Based Index) | SD (Mean-Based Index) | SD (Median-Based Index) |
|---|---|---|---|---|
| No Distress (0) | -0.1023 | -0.1075 | 0.9784 | 0.9855 |
| In Distress (1) | 0.1058 | 0.0952 | 0.8192 | 0.8414 |
| Total | -0.0549 | -0.0613 | 0.9479 | 0.9578 |

Notes: This table summarizes composite stigma indices by distress groups. The mean-based and median-based indices classify stigma levels using different statistical cutoffs. Individuals in distress show higher stigma levels compared to those without distress.

Table B20: T-Test Results for Composite Stigma Indices by Distress Groups

| Index | Group | Obs | Mean | Std. Err. | 95% CI | p-value (two-tailed) |
|---|---|---|---|---|---|---|
| Mean-Based Index | No Distress | 525 | -0.1023 | 0.0427 | [-0.1862, -0.0184] | 0.0162 |
| | In Distress | 155 | 0.1058 | 0.0658 | [-0.0242, 0.2358] | |
| | Difference | | -0.2081 | 0.0863 | [-0.3776, -0.0386] | |
| Median-Based Index | No Distress | 525 | -0.1075 | 0.0430 | [-0.1920, -0.0230] | 0.0205 |
| | In Distress | 155 | 0.0952 | 0.0676 | [-0.0383, 0.2287] | |
| | Difference | | -0.2027 | 0.0873 | [-0.3741, -0.0314] | |

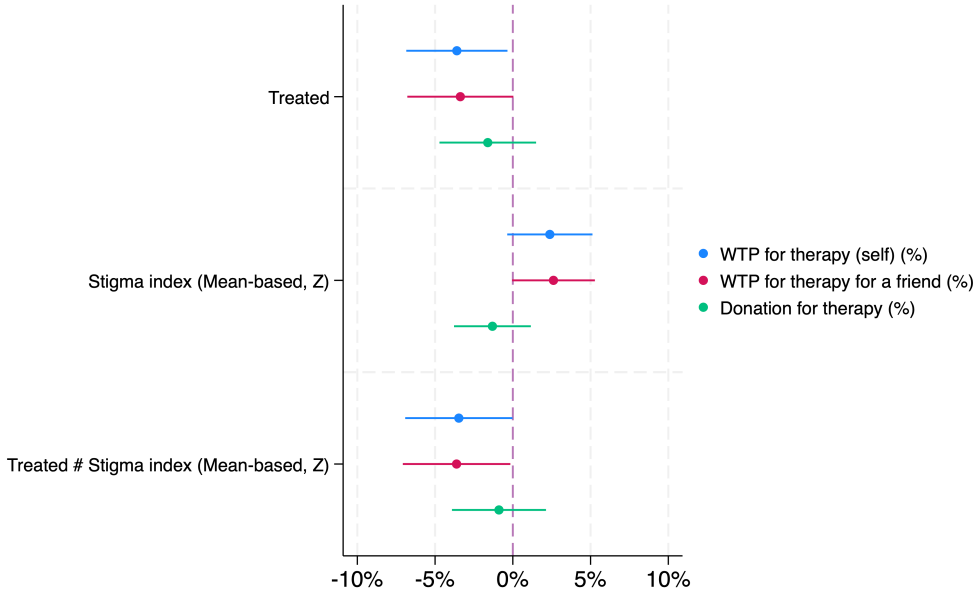
Notes: This table shows two-sample t-test results comparing composite stigma indices between individuals with and without distress. The mean-based and median-based indices classify stigma perceptions using different statistical thresholds. Results show significant differences, with distressed individuals exhibiting higher stigma levels.

The results of the two-sample t-tests indicate a significant difference in stigma indices (mean and median) between individuals with "No Distress" (No D) and those "In Distress" (In D). For the stigma index based on the mean, individuals in the "No D" group had a significantly lower stigma index (Mean = -0.102, Std. Dev = 0.978) compared to those in the "In D" group (Mean = 0.106, Std. Dev = 0.819), with a mean difference of -0.208 (95% CI: -0.378 to -0.039; $t = -2.4101$, $p = 0.0162$ for the two-tailed test). Similarly, for the stigma index based on the median, the "No D" group had a lower stigma index (Mean = -0.108, Std. Dev = 0.986) compared to the "In D" group (Mean = 0.095, Std. Dev = 0.841), with a mean difference of -0.203 (95% CI: -0.374 to -0.031; $t = -2.3228$, $p = 0.0205$ for the two-tailed test).

These findings suggest that individuals in distress experience higher levels of stigma compared to those not in distress. The statistical significance ($p < 0.05$) and confidence intervals that exclude zero provide strong evidence that these differences are unlikely due to random chance. Although the effect sizes (mean differences of -0.208 and -0.203) are relatively small, the results underscore the need for targeted interventions to address stigma among distressed individuals.

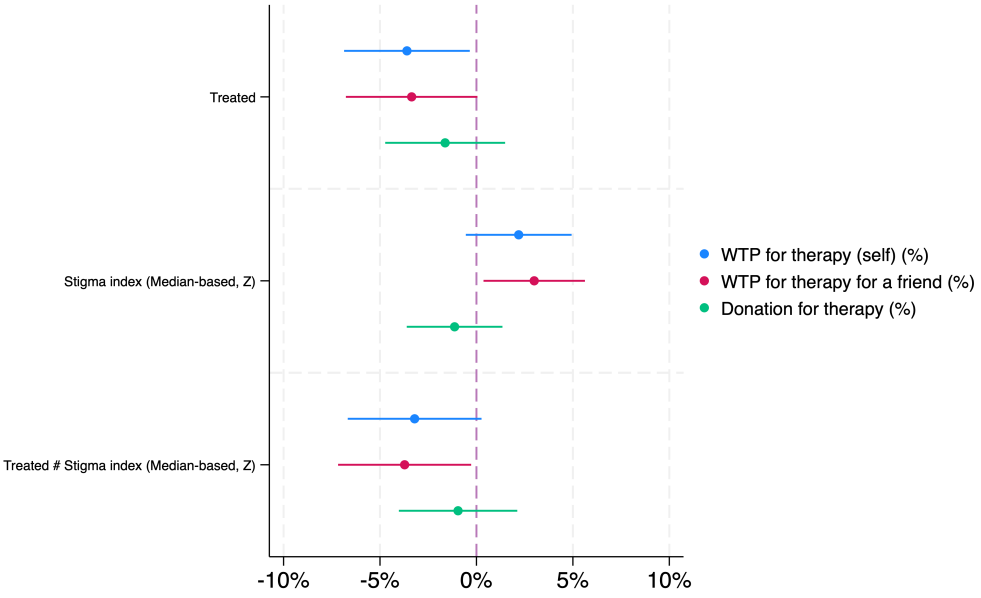
**Index from Weighted Average**

Figure B31: Main Effects by Mean Stigma Index




Notes: This figure shows treatment effects by mean stigma index, illustrating its relationship with willingness to pay (WTP) for therapy for oneself, for a friend, and donation amounts. The mean stigma index reflects perceived public stigma, self-stigma, and personal stigma, capturing overall attitudes toward mental health.

Figure B32: Main Effects by Median Stigma Index




Notes: This figure shows treatment effects by median stigma index, showing its relationship with willingness to pay (WTP) for therapy for oneself, for a friend, and donation amounts. The median stigma index captures perceived public stigma, self-stigma, and personal stigma, summarizing overall attitudes toward mental health.

### D.4.2 PCA

Principal Component Analysis (PCA) is employed as a dimensionality reduction technique to distill key insights from a dataset with multiple variables while minimizing the loss of critical information (Jaadi & Whitfield (2024)), the context of our research on mental health stigma among university students, PCA enables us to synthesize a complex set of variables—such as perceptions of therapy, barriers to seeking help, and beliefs about peer behavior—into a smaller number of components. These components capture the majority of the variance within the original dataset, providing a simplified yet meaningful representation of the underlying patterns.

In this analysis, PCA helps identify the primary dimensions of mental health stigma, which we use to construct an index reflecting the most significant factors influencing students' attitudes and behaviors. Initially, all components and loadings are considered, but subsequent iterations focus on those with the highest explained variance and loadings of 0.3 or above. This filtering ensures that we emphasize the most informative relationships between variables. Prior to applying PCA, all variables are standardized to ensure comparability and to give equal weight to each variable, regardless of its original scale.

This approach not only simplifies our data analysis but also provides a robust foundation for understanding the most influential factors shaping students' mental health perceptions and their decision-making regarding therapy.

Table B21: Correlation of Stigma Index PCA1 and PCA2 with Components

| Variable | PCA1 | PCA2 | Stigma Students | Stigma Professors | Stigma Parents | Guess Self-Stigma | Low GPA Symptoms | Low GPA Talk |
|---|---|---|---|---|---|---|---|---|
| **PCA1** | 1.00 | | | | | | | |
| **PCA2** | 0.00 | 1.00 | | | | | | |
| **Stigma Students** | 0.83*** | 0.04 | 1.00 | | | | | |
| **Stigma Professors** | 0.87*** | 0.04 | 0.64*** | 1.00 | | | | |
| **Stigma Parents** | 0.83*** | -0.03 | 0.53*** | 0.66*** | 1.00 | | | |
| **Guess Self-Stigma** | 0.50*** | 0.11** | 0.30*** | 0.27*** | 0.25*** | 1.00 | | |
| **Low GPA Symptoms** | -0.09* | 0.84*** | -0.02 | -0.03 | -0.06 | -0.00 | 1.00 | |
| **Low GPA Talk** | -0.04 | 0.85*** | -0.01 | 0.00 | -0.04 | 0.04 | 0.45*** | 1.00 |

Notes: This table shows the correlations between the two principal components (PCA1 and PCA2) and the key stigma-related variables. PCA1 primarily captures perceived public stigma from students, professors, and parents, while PCA2 reflects attitudes related to academic performance and mental health. The stigma variables represent perceived stigma from different groups, Guess Self-Stigma measures internalized stigma, and Low GPA Symptoms and Low GPA Talk capture preferences for avoiding mental health symptoms or discussions even at the cost of lower academic performance.

Table B22: Summary Statistics for Stigma Variables and PCA Indexes

| Variable | Observations | Mean | Std. Dev. | Min–Max |
|---|---|---|---|---|
| Stigma Index PCA1 | 680 | 0.000 | 1.00 | -1.93 – 3.47 |
| Stigma Index PCA2 | 680 | 0.000 | 1.00 | -0.80 – 2.66 |
| Stigma Students (Std.) | 680 | 0.000 | 1.00 | -1.22 – 3.40 |
| Stigma Professors (Std.) | 680 | 0.000 | 1.00 | -1.20 – 3.31 |
| Stigma Parents (Std.) | 680 | 0.000 | 1.00 | -1.57 – 2.34 |
| Guess Self-Stigma (Std.) | 680 | 0.000 | 1.00 | -2.20 – 2.21 |
| Low GPA Symptoms (Std.) | 680 | 0.000 | 1.00 | -0.61 – 1.65 |
| Low GPA Talk (Std.) | 680 | 0.000 | 1.00 | -0.40 – 2.50 |

Notes: PCA1 captures public stigma perceptions from students, professors, and parents, while PCA2 reflects attitudes toward academic performance and mental health. Stigma variables measure perceived stigma from different groups, self-stigma represents internalized beliefs, and low GPA variables capture preferences for academic performance over mental health concerns.

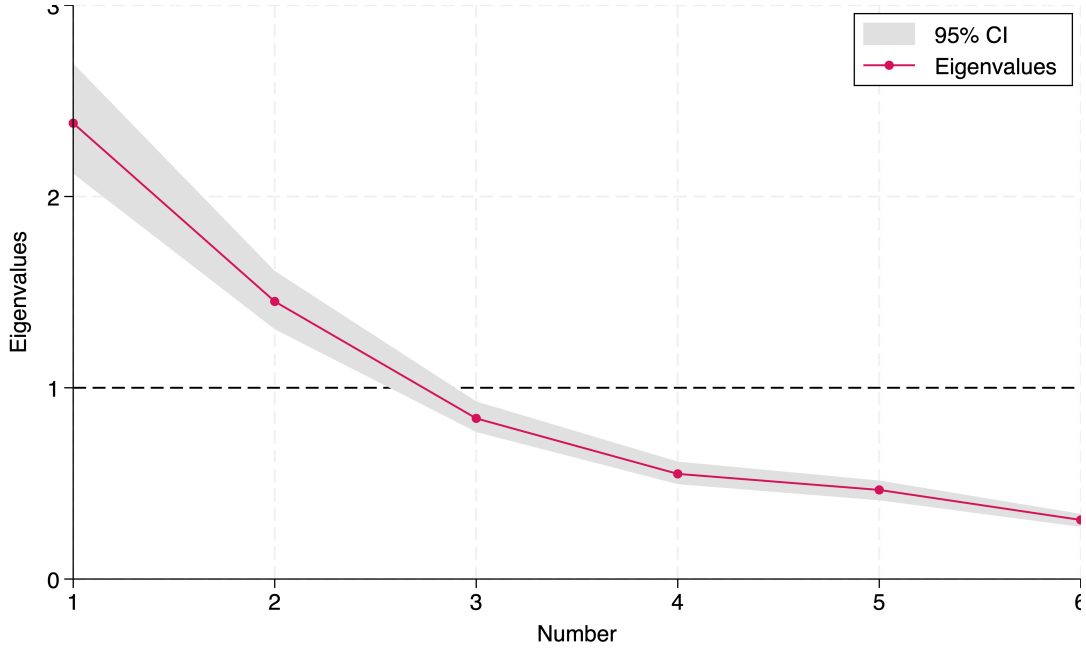
### PCA Results

Table B23: Principal Components Analysis Summary

| **Component** | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
|---|---|---|---|---|
| Component 1 | 2.3835 | 0.9320 | 0.3972 | 0.3972 |
| Component 2 | 1.4515 | 0.6114 | 0.2419 | 0.6392 |
| Component 3 | 0.8401 | 0.2902 | 0.1400 | 0.7792 |
| Component 4 | 0.5499 | 0.0839 | 0.0917 | 0.8708 |
| Component 5 | 0.4660 | 0.1569 | 0.0777 | 0.9485 |
| Component 6 | 0.3090 | – | 0.0515 | 1.0000 |
| **Summary Statistics:** | | | | |
| Number of observations | | | | 680 |
| Number of components | | | | 2 |
| Trace | | | | 2 |
| Rotation (unrotated) | | | | Principal |
| Rho | | | | 0.6392 |

Notes: The eigenvalues measure the variance explained by each principal component. According to the Kaiser criterion (Jaadi and Whitfield, 2024), only components with eigenvalues above 1 should be retained. In this case, only the first two components meet this criterion, capturing 63.92% of the total variance. These components will be used for further analysis, such as examining the loadings.

Figure B33: Screeplot




Notes: This figure shows the variance explained by each principal component. We keep components 1 and 2, which exceed the threshold of 1 for being kept in further analyses.

Eigenvalues are the measure of how much variance (information) each principal component explains in the dataset. Larger eigenvalues indicate components that explain more variance Jaadi & Whitfield (2024). From the initial PCA results in the table above and from the screeplot we can observe that only the first two components have eigenvalues of 1 and above - meaning they each explain greater variance than the rest of the components - which will be the ones we shall be keeping, and the only components we shall be considering when looking at the loadings.

Table B24: Principal Components (Eigenvectors)

| Variable | Component 1 | Component 2 | Component 3 | Component 4 | Component 5 | Component 6 |
|---|---|---|---|---|---|---|
| stigma_students | 0.5358 | 0.0351 | -0.1130 | 0.0557 | -0.7420 | 0.3810 |
| stigma_professors | 0.5649 | 0.0359 | -0.2256 | -0.0323 | 0.0352 | -0.7915 |
| stigma_parents | 0.5359 | -0.0216 | -0.2206 | -0.0214 | 0.6617 | 0.4747 |
| self_stigma | 0.3206 | 0.0934 | 0.9365 | 0.0610 | 0.0811 | -0.0327 |
| low_GPA_over_sympt | -0.0269 | 0.7059 | -0.0124 | -0.7062 | -0.0142 | 0.0446 |
| low_GPA_over_talk | -0.0551 | 0.7000 | -0.1020 | 0.7021 | 0.0593 | -0.0045 |

Notes: This table shows the loadings (coefficients) from the principal component analysis (PCA), representing how much each variable contributes to a given component. In the next step, only loadings greater than 0.3 will be considered to improve interpretability.

The loadings - the coefficients, or weights - from the Principal Component Eigenvectors table above represent the contribution of each variable to a given principal component. In the next iteration only loadings above .3 will be considered in order to better interpret components. Subsequently, we make sure the first two components we have focused on are not correlated amongst each other.

Table B25: Correlation Matrix of Principal Components

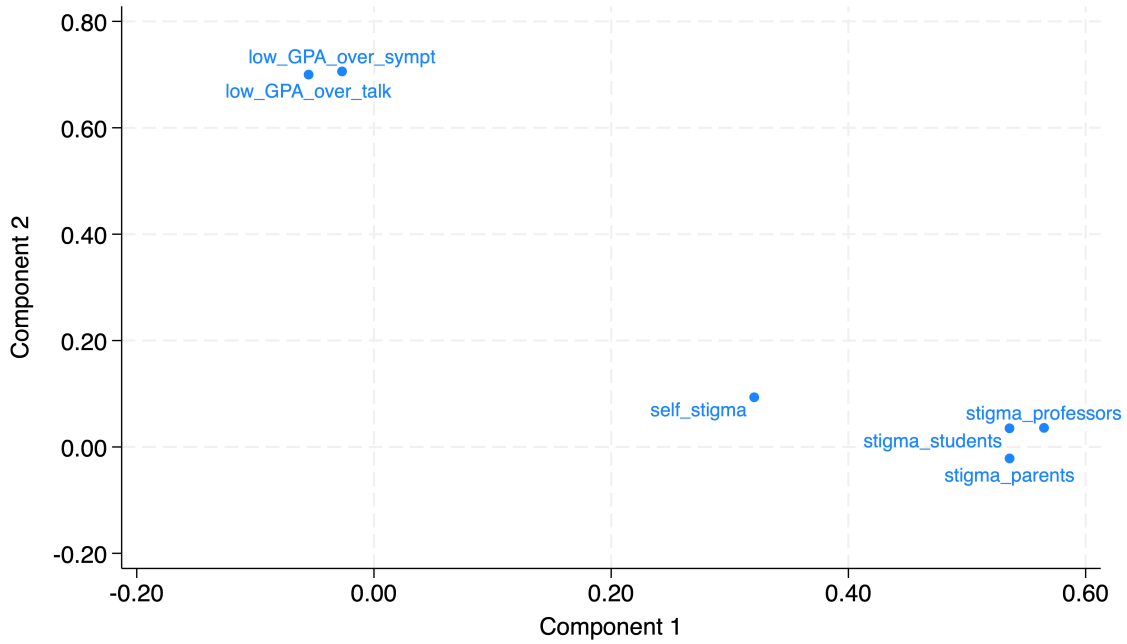
| | pc1 | pc2 |
|---|---|---|
| pc1 | 1.0 | – |
| pc2 | 0.0 | 1.0 |

Notes: This table shows the correlation matrix between both of our Principal Components. PC1 captures public stigma perceptions from students, professors, and parents, while PC2 reflects attitudes toward academic performance and mental health.

### PCA Interpretation

Component 1 primarily captures perceptions of stigma from various groups (students, professors, parents), while Component 2 reflects preferences related to mental health versus academic performance (low GPA acceptance). The following loading plot showcases the previous loadings and how related they are to each component.

Figure B34: PCA Loading Plot



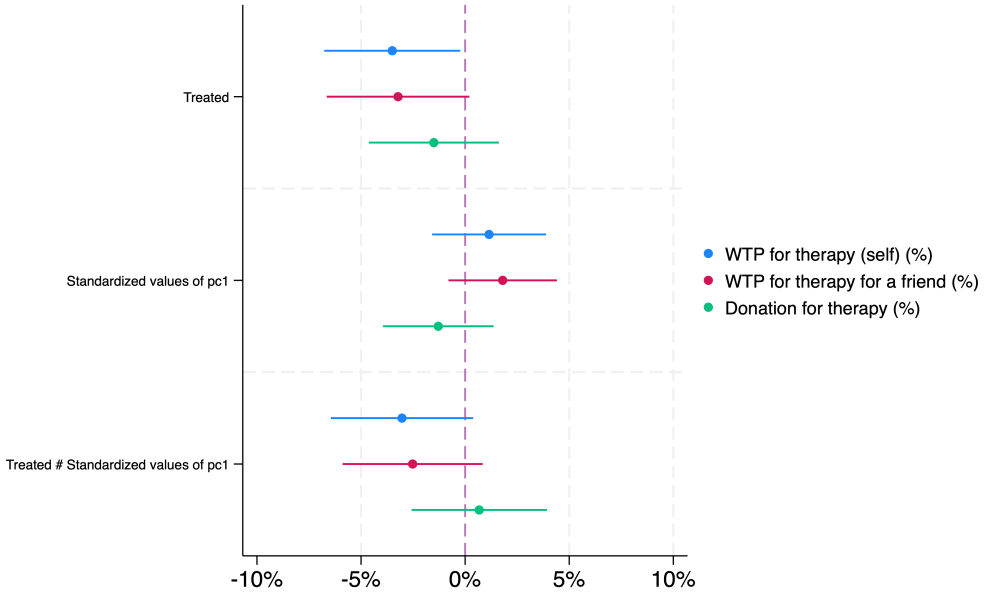

Notes: This figure shows the scatter plot of PCA loadings for each of our Principal Components. PC1 is primarily explained by perceived public stigma from students, professors, and parents, while PC2 reflects attitudes toward academic performance and mental health, particularly preferences related to GPA trade-offs.

### Index from PCA

After having done and examined the 2 components of the PCA, we proceed to construct two indexes from component 1 and component 2, we then proceed to interact of our treatment groups with the our

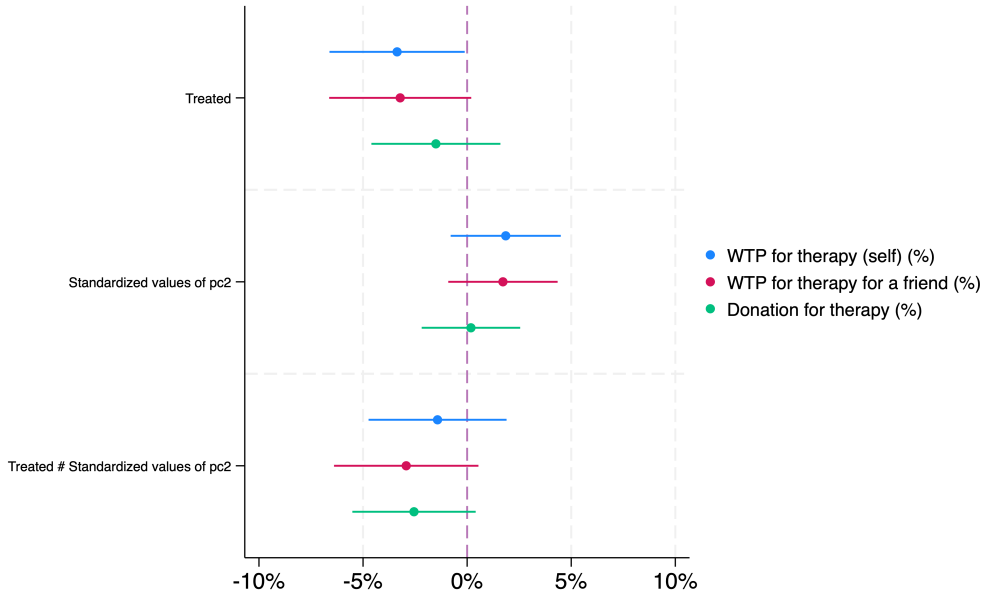
standardized stigma index with outcome variables being WTP for therapy, for self, for a friend and lastly the donation amount an individual is willing to give to help someone access mental health services.

Figure B35: Main Effects by Component 1 Stigma Index




Notes: This figure shows treatment effects by Component 1 (PC1) of the stigma index, illustrating its relationship with willingness to pay (WTP) for therapy for oneself, for a friend, and donation amounts. PC1 primarily captures perceived public stigma from students, professors, and parents, summarizing external attitudes toward mental health.

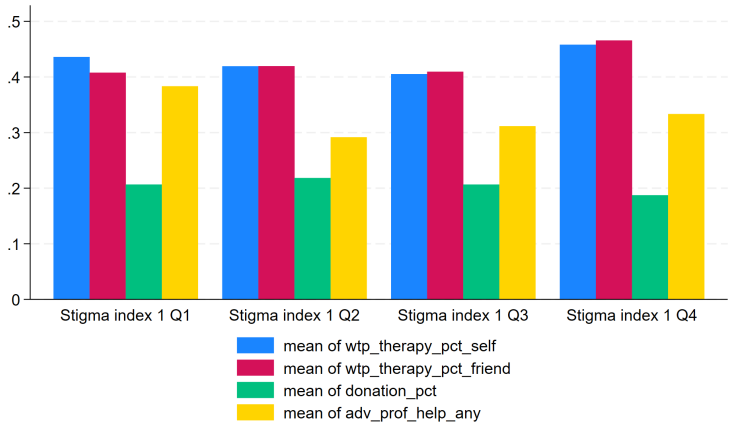
Figure B36: Main Effects by Component 2 Stigma Index




Notes: This figure shows treatment effects by Component 2 (PC2) of the stigma index, illustrating its relationship with willingness to pay (WTP) for therapy for oneself, for a friend, and donation amounts. PC2 primarily reflects attitudes toward academic performance and mental health, particularly preferences related to GPA trade-offs.

### D.4.3 Stigma & Demand for Mental Health Services

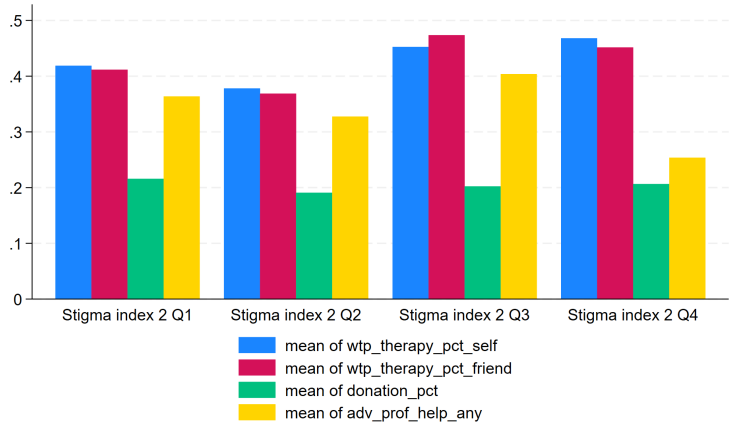
Figure B37: Outcome Means by Stigma Index 1




Notes: This figure illustrates how outcome measures vary across quartiles of stigma index 1, capturing differences in willingness to pay (WTP) for therapy, advocacy for professional mental health services, and donation behavior. Higher stigma levels are associated with increased personal investment in therapy but reduced advocacy for professional help.

Figure B37 shows how outcome measures vary across the quartiles of stigma index 1. In Q1 (lowest stigma), the WTP for therapy for oneself and a friend is high, alongside higher rates of advocating for professional mental health services. In Q4 (highest stigma) there is an increase in WTP for therapy for oneself and a friend, while a stark decrease in advocacy for professional help; donations seem relatively uniform across all quartiles. The differences across quartiles reveal that higher stigma is associated with increased personal investment in therapy but decreased engagement with broader supportive actions, such as advocating for professional help or donating.

Figure B38: Outcome Means by Stigma Index 2




Notes: This figure illustrates how outcome measures vary across quartiles of stigma index 2, capturing differences in willingness to pay (WTP) for therapy, advocacy for professional mental health services, and donation behavior. Higher stigma levels are associated with increased personal investment in therapy but reduced advocacy for professional help.

When looking at outcome means by stigma index 2, which primarily focuses on personal stigma—a dimension that mainly measures the preferences of academic performance over addressing mental health issues—one can observe that there is an increasing trend in WTP for self and friend from Q1 to Q4, and a decrease in advocacy for mental health services. Similarly to outcome means by stigma index 1, donations seem relatively stable across quartiles. These parallel trends between stigma index 1 and stigma index 2 suggest that while the two indices capture different dimensions of stigma (public and personal, respectively), their influence on behavioral outcomes, such as WTP, advocacy, and donations, is aligned. This alignment reinforces the robustness of stigma index 1 in explaining how stigma—whether public or personal—affects mental health-related decisions and highlights the consistency of stigma's negative impact on broader support for professional mental health resources.

Table B26: Correlation of Demand Variables with Stigma Indices

| Variable | WTP Therapy (Self) | WTP Therapy (Friend) | Donation (%) | Stigma Index PCA1 | Stigma Index PCA2 |
|---|---|---|---|---|---|
| Stigma Index PCA1 | -0.03 | 0.01 | -0.04 | 1.00 | |
| Stigma Index PCA2 | 0.04 | -0.00 | -0.06 | 0.00 | 1.00 |

Notes: WTP Therapy (Self) and WTP Therapy (Friend) measure willingness to pay for therapy for oneself and for a friend, respectively. Donation (%) represents the percentage of income participants are willing to donate to mental health causes. Stigma Index PCA1 captures public stigma perceptions, while Stigma Index PCA2 reflects attitudes toward academic performance and mental health. The table shows correlations between these variables.

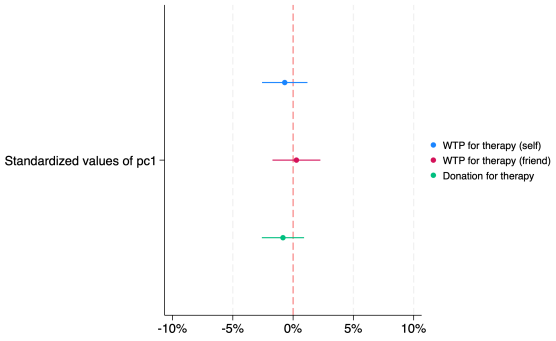
The Table B26 shows that willingness to pay (WTP) for therapy for oneself and for a friend are strongly and positively correlated, suggesting that individuals who value therapy for themselves also value it for others. Donations, while positively correlated with both WTP measures, exhibit weaker associations, indicating a different motivational factor driving altruistic behavior. Neither stigma index (PCA1 or PCA2) shows significant correlations with WTP or donations, highlighting that perceived public stigma (PCA1) and personal stigma (PCA2) are not directly linked to demand for therapy or altruistic behavior in this context.

The coefficient plots demonstrate that neither Stigma Index 1 (PCA1), representing perceived public stigma, nor Stigma Index 2 (PCA2), capturing personal stigma, has a significant impact on the demand variables. For PCA1, the effects on willingness to pay (WTP) for therapy (self and friend) and donations for therapy are minimal, indicating that public stigma perceptions do not strongly influence these behaviors. Similarly, PCA2 shows near-zero effects across the same variables, suggesting that individual attitudes and personal stigma are not major drivers of therapy demand or donation behavior.

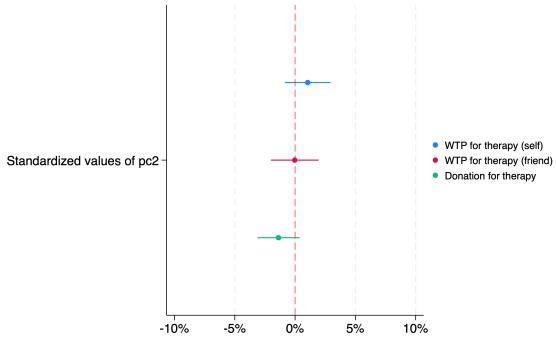
The stigma section reveals that stigma perceptions are shaped by distress levels, professional help usage, and prior beliefs about mental health. Stigma index 1 (PCA1), which captures perceived public stigma from peers, professors, and parents, provides a robust measure of how external societal attitudes influence mental health-related decisions. In contrast, stigma index 2 (PCA2) reflects personal stigma and internalized biases, such as prioritizing academic performance over mental health, but has a narrower focus and limited explanatory power.

Figure B39: Stigma Indices and WTP/Donate

(a) Impact of Stigma Index 1 on Demand




(b) Impact of Stigma Index 2 on Demand

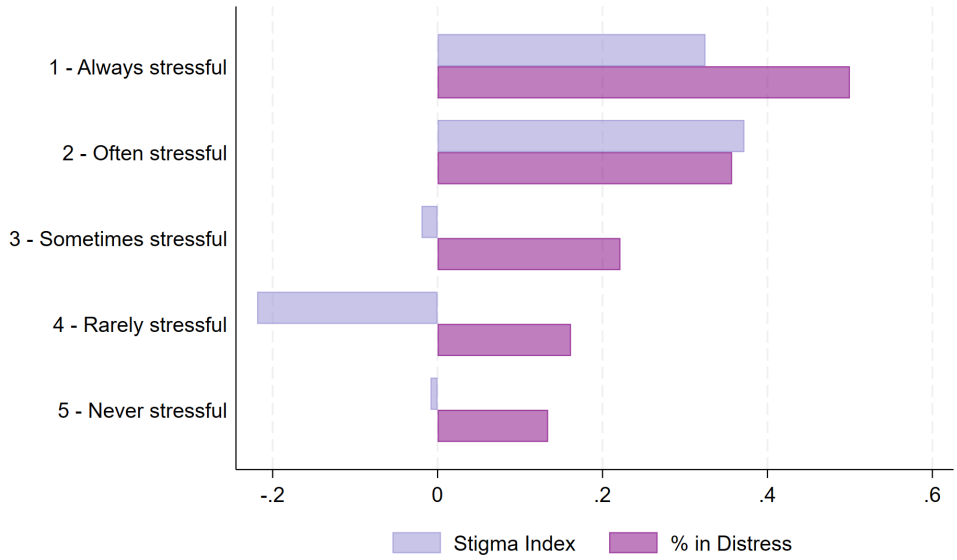



Notes: This figure shows the impact of stigma indices on demand, measured by willingness to pay (WTP) for therapy for oneself and a friend, as well as donations. Stigma Index 1 (PC1) captures public stigma perceptions from students, professors, and parents, while Stigma Index 2 (PC2) reflects personal stigma, particularly preferences for academic performance over addressing mental health issues.

The lack of significant correlations between the stigma indices and demand variables (WTP for therapy and donations) suggests that neither public nor personal stigma directly drives these behaviors. Instead, the data implies that stigma impacts broader societal norms and individual perceptions rather than immediate willingness to invest in therapy. Stigma index 1's comprehensive design offers valuable insights into the broader societal dynamics of stigma, but a more expansive experimental framework could better capture its multifaceted effects on mental health outcomes and decision-making. This underscores the need for future research to refine stigma measures and incorporate additional dimensions for a more complete understanding.

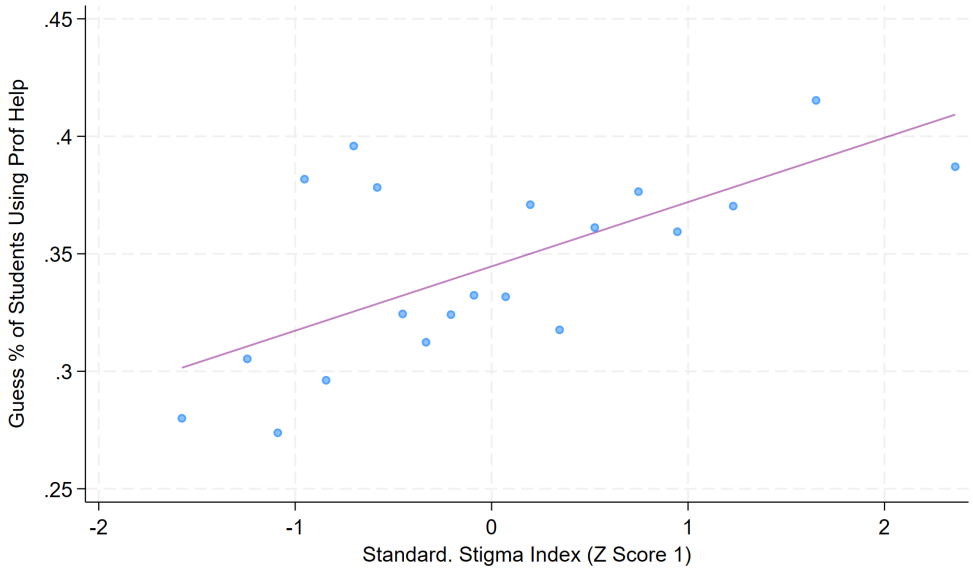
Furthermore, Figure B40 highlights a positive correlation across measures of stigma and financial stress, indicating a potential connection to demand as financial stress may act as a proxy for financial constraints.

Figure B40: Stigma & Mental Distress by Financial Stress




Notes: This figure highlights a positive correlation across measures of stigma and financial stress, indicating a potential connection to demand as financial stress may act as a proxy for financial constraints.

Figure B41: Relationship Between Stigma Index and Perceived Use of Professional Mental Health Services




Notes: This figure shows a strong positive correlation between stigma index 1 and predicted therapy use. Individuals with higher stigma levels may assume greater concealment of therapy use among peers, influencing their broader perceptions of mental health treatment.

Furthermore, in the given context, there appears to be a strong positive correlation between stigma index 1 and predicted therapy use, as shown in Figure B41. This trend aligns with the notion that individuals rationalize their assumptions about therapy use based on their own beliefs and perceived social norms. Specifically, individuals with higher stigma levels may assume higher rates of concealing therapy use among peers, which in turn influences their predictions of broader usage patterns. This rationalization mechanism underscores the role of stigma in shaping perceptions of mental health treatment, particularly through the lens of assumed societal concealment behaviors.

## D.5 Advice Indicators for mentioning words or phrases

We generate indicator variables based on the inclusion of specific words or phrases in incentivized advice provided by subjects.

We build the "Empathetic advice" variable as an indicator equal to one if any of the following variables are mentioned by the respondent: Listen is an indicator equal to one if the respondent mentions "listen." Be attentive is an indicator equal to one if the respondent mentions "I am here for you/him/her/them", "I am there for you/him/her/them." Empathy is an indicator equal to one if the respondent mentions "empathy" or "understood." Validate feelings is an indicator equal to one if the respondent mentions it is "not bad to feel bad", "it is normal not to feel ok" or "it is completely normal." Show support is an indicator equal to one if the respondent mentions "I support you", "you have my support" or "I love you."

We build the "Directive advice" variable as an indicator equal to one if any of the following variables are mentioned by the respondent: Give opinion is an indicator equal to one if the respondent mentions "opinion", "advice", "what to do", "recommend", "you/he/she should", "my experience." Seek help is an indicator equal to one if the respondent mentions "seek help/support", "refer to professional", "find resources." Mention therapy is an indicator equal to one if the respondent mentions "therapy", "psychologist" or "counseling." Do stuff you enjoy is an indicator equal to one if the respondent mentions "do something you enjoy", "activities you like" or "do stuff you like."

# E Appendix: Pre-Analysis Plan (Baseline)

## E.1 Introduction

The study is an online survey experiment with university students in Mexico as participants, collecting information on student demographics, their mental health, and their beliefs and demand for mental-health related support services. We experimentally vary subject exposure to (1) information about therapy use and effectiveness, and (2) information with a reflection activity, to test whether the light-touch intervention has an effect on the demand for mental health support services and mental health stigma[50] by students exposed to information and/or reflection.

The survey primarily consists of questions to elicit respondents' mental health state (using PHQ and GAD screeners), experience with using therapy, experience with student services, and demographic information. Experimental variation comes from a random assignment of participants to one of the three treatment conditions. In the first arm, participants are exposed to information related to mental health and support in the form of an infographic. In addition, they complete a reflection activity with an open-ended question and a vignette component. In the second arm, participants are exposed to the information component of the intervention only. In the third arm (control arm), participants are exposed to a more neutral set of questions about general campus services focused on mental health.

Our survey includes basic demographics and student-status questions, standardized mental-health related questions, including short versions of the standardized surveys for depression and anxiety, with 4 questions from each (PHQ-4 and GAD-4). This enables us to have a more concise version of the overall survey and also not prompt the participants with more sensitive questions, i.e. our short versions of PHQ-4 and GAD-4 do not include the questions on suicidality and self-harm, yet are still accepted screening protocols used in previous studies.

## E.2 Sampling and Data Collection

**Sample:** University students over 18 years old who are full-time students at a large private university in Mexico. We are going to target a representative sample across undergraduate and graduate student populations (representative by gender, faculty/department, year at university). No identifiable information will be collected, and since the participants are a representative sample of the larger student population, basic demographics would not be sufficient to identify specific individuals. We will not be collecting their names or student IDs.

**Compensation:** First 100 respondents to complete the survey will get a guaranteed payment of MXN $200 (approximately USD $10). In addition, there will be opportunities for participants to earn more depending on their performance in "bonus" questions throughout the survey (there is a total of 8 such questions, and

[50] Throughout this project, we define mental health stigma as a complex of negative attitudes, beliefs, and stereotypes that people have about those with mental health conditions

we choose one question at random for bonus payment). Furthermore, all survey participants automatically enter a raffle with big monetary prizes: all subjects have an equal chance of getting one of 20 gift cards each worth MXN $2,000 (approximately USD $100). All earnings in the survey will be paid to subjects in the form of Amazon gift card vouchers. All payments will be disbursed after the data collection for this project is finalized.

## E.3 Experiment Design

We use a between-subject design and randomly split students into three treatment conditions, with equal probability of assignment to either condition.

- **T1**: In the Information & Reflection group, subjects are exposed to (1) an information component with mental-health-related facts, including three quantitative statements about therapy effectiveness, therapy utilization and relationship between mental health and educational outcomes, and (2) a reflection component, in which they write an open-ended response to a prompt and read a vignette with a story about a student's experience using therapy.
- **T2**: Students assigned to the Information group complete the information component as described above only, followed by one open-ended question about general on-campus services (such as sports facilities or career services).
- **C**: Finally, students in Control group answer a set of multiple choice and open-ended questions which are all unrelated to mental health.

Randomization is implemented automatically through Qualtrics with a third of the subject pool in each treatment group.

### E.3.1 Outcome Variables

Our primary outcomes capture students' willingness to pay (WTP) for mental health support, including WTP for therapy sessions for themselves and for a friend, as well as their willingness to donate a portion of their survey earnings to subsidize therapy for students with financial need. Secondary outcomes capture proxies for mental-health stigma and openness to discussing mental health support and include comfort levels working with peers who have mental health issues, the likelihood of sharing mental health resources, and hypothetical support for friends experiencing personal challenges.

The list of primary outcomes:

- **Primary outcome: Willingness to pay (WTP) for therapy for self.** Subjects input their willingness to pay for a 4-week therapy service subscription for themselves (if the response is $0, we ask a hypothetical question about willingness to accept, WTA)

- **Primary outcome: Willingness to pay (WTP) for therapy for a friend.** Subjects input their willingness to pay for a 4-week therapy service subscription for a friend (subjects are asked to leave the contact details of their friend)
- **Primary outcome: Donation to cover therapy cost to a student with financial need.** Subjects input what percentage of their total earnings from the survey they would like to donate to help cover the cost of a therapy session for another student, who expressed that cost is one of the factors preventing them from seeking professional mental health help.

The list of secondary outcomes:

- **Secondary outcome: Ranking student profiles.** Subjects rank several types of (hypothetical) students in terms of how comfortable they would feel working with them on a project
- **Secondary outcome: Information link sharing.** Subjects are given a link to the university's counseling website that they can share with their friends; we will track the number of clicks on this link (3 distinct links generated for 3 treatment groups)
- **Secondary outcome: Advice.** Subjects are asked a hypothetical question about how they would support their friend who is struggling with personal issues

## E.4 Empirical Analysis Plan

In our empirical analysis, we will examine the impact of the experimental treatments on both primary and secondary outcomes related to student demand for mental health support and stigma. We will run regressions of outcome variables on treatment binary variables, controlling for key demographic and socio-economic covariates that may be unbalanced at baseline due to randomized assignment. Baseline covariates include age, gender, sexual orientation, parental education, financial aid status, college major, and year in college.

Our hypotheses focus on whether providing information and engaging students in a reflection activity increase their demand for mental health support services and their likelihood of supporting peers' access to therapy resources. We will further explore heterogeneity in treatment effects by underlying mental health status (level of mental distress), prior beliefs about mental health, and mental health stigma. Overall, we aim to understand how information and self-reflection can influence attitudes and behaviors related to mental health support among university students.

Our analysis will evaluate the following hypotheses:

**H1:** The information treatment will increase university students' demand for mental health support, measured by their willingness to pay (WTP) for therapy.

**H2:** Conditional on beliefs about therapy, engaging in a reflection activity will further increase university students' demand for therapy.

**H3:** The combination of information and reflection interventions will lead to higher perceived demand for therapy by others, measured by WTP for therapy for a friend and the fraction of survey earnings students choose to donate to a fellow student's therapy.

**H4:** The combination of information and reflection interventions will increase the likelihood of sharing mental health resources with peers, measured by the number of clicks on the shared link.

**H5:** The treatment effect will be larger for students whose prior beliefs about therapy effectiveness and utilization are further from the truth.

**H6:** The treatment effect will be larger for students who exhibit higher levels of mental health stigma.

# F Appendix: Pre-Analysis Plan (Follow-Up)

Based on the results of our survey from November 2024, we identified several patterns in the effect of the information treatment on student responses. To elicit additional information on the channels, as well as to capture long-term (persistent) effects, we decided to run a follow-up round-2 survey with the participants who had valid responses in round 1. We designed an online survey with 9 yes/no questions (incl. 8 outcome questions), with incentives to respond added with a gift card raffle based on completion. We aim to collect the responses to this survey in late April-early May.

1. Have you been taking classes in UNIVERSITY in 2025?

**Personal-Use Outcomes (on campus/off campus therapy):**

2. Have you used professional mental health services ON campus in the last 6 months?
3. Have you used professional mental health OFF campus in the last 6 months?

**Social-Use Outcomes (info sharing/discussions):**

4. Have you recommended professional mental health services ON campus to any of your peers in the last 6 months?
5. Have you recommended professional mental health services OFF campus to any of your peers in the last 6 months?
6. Have you discussed your mental health issues with other UNIVERSITY students in the last 6 months?
7. Have you discussed other UNIVERSITY students' use of professional mental health services with other students in the last 6 months?

**Hypothetical Personal Use Outcomes:**

8. Would you consider going to therapy ON campus if you had issues?
9. Would you consider going to therapy OFF campus if you had issues?

**Hypotheses**

1. **Personal Use (actual & hypothetical):**
   Treated subjects are more likely than control subjects to have used professional mental health services in the past 6 months or consider seeking therapy if they had issues.
2. **Social Use (peer recommendations):**
   Treated subjects are more likely than control subjects to recommend professional mental health services to their peers and to discuss their own mental health and/or others' use of mental health services with peers.
3. **Social information sharing vs. personal demand:**
   If we observe positive effects on recommendation and discussion outcomes, but not on subjects' own use of or willingness to seek therapy, it would suggest that the information intervention primarily promoted peer interactions around mental health topics rather than directly increasing individual demand for therapy.
4. **On-campus free counseling vs. off-campus private counseling:**
   If we observe stronger positive effects for on-campus counseling use and recommendations compared to off-campus counseling, it would support the substitution hypothesis that the information intervention encouraged students to prefer and promote on-campus services over off-campus options.

**Heterogeneity analysis**

We expect that the information treatment differentially affected respondents in the long run depending on their **GPA** (related info mentioned in fact 3), **level of mental distress** (related info mentioned in fact 2), and **stigma** (implicit channel), we will explore heterogeneous effects by baseline GPA, level of mental distress, and stigma measures.