The Alongside Digital Mental Health Program for Teens: A Pilot Evaluation

Katherine Cohen, Andrew Rapoport, Elsa Friis, Shannon Hill, Sergey Feldman, Jessica Schleider

Submitted to: JMIR Formative Research
on: February 26, 2025

Table of Contents

Original Manuscript..... 5

The Alongside Digital Mental Health Program for Teens: A Pilot Evaluation

Katherine Cohen[1, 1*] MA; Andrew Rapoport[1*] MPH; Elsa Friis[2] PhD; Shannon Hill[1] MPH; Sergey Feldman[2] PhD; Jessica Schleider[1] PhD

[1]Department of Medical Social Sciences Feinberg School of Medicine Northwestern University Chicago US
[2] Alongside Care Seattle US
[*]these authors contributed equally

Corresponding Author:
Katherine Cohen MA
Department of Medical Social Sciences
Feinberg School of Medicine
Northwestern University
625 N Michigan Ave
21st Floor
Chicago
US

Abstract

Background: Youth are increasingly experiencing psychological distress. Schools are ideal settings for disseminating mental health support, but they are often insufficiently resourced to do so. Digital mental health tools represent a unique avenue to address this gap. The Alongside digital program is one such tool, intended as universal prevention and early intervention. The platform includes social-emotional learning and self-help wellness features as well as an AI-powered chatbot designed to build coping skills.

Objective: This evaluation aimed to examine the near-term impact of Alongside app usage on students' self-reported mental health outcomes.

Methods: We conducted a non-randomized pilot pragmatic evaluation leveraging anonymized user data. All data came from current Alongside users attending public middle and high schools in Texas and New Mexico, between 10-18 years old. Schools were actively engaged in partnership with Alongside and approved all procedures. Users were asked to complete mental health questionnaires upon app registration and at one-month and three-months post-registration. We conducted pre-registered analyses as well as exploratory analyses to determine how symptoms changed over time and what factors (e.g., demographic, app usage) predicted changes in distress.

Results: Analyses revealed statistically significant within-person decreases in overall distress (YP-CORE; primary outcome) from baseline to one-month with a small effect size ($t(42) = 2.21$, $p = 0.03$, $r = 0.34$); however, there was no evidence that scores significantly decreased from baseline to three-months ($W = 1821$, $n = 85$, $p = 0.16$). We found that at three-months, identifying as part of the LGBTQ+ community predicted greater decreases in distress; otherwise no demographic factors were significant predictors. In a non-registered exploratory analysis of a subsample of users who reported elevated distress at baseline, decreases in distress were seen at both one-month ($W = 128$, $n = 20$, $p = 0.02$, $r = 0.52$) and three-months ($W = 682$, $n = 42$, $p = 0.004$, $r = 0.45$).

Conclusions: There may be short-term benefits associated with using the Alongside digital program. Further studies are required to determine potential preventative effects. Clinical Trial: https://osf.io/m8t6k

(JMIR Preprints 26/02/2025:73180)
DOI: https://doi.org/10.2196/preprints.73180

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

The Alongside Digital Mental Health Program for Teens: A Pilot Evaluation

Katherine Cohen, MA*[a], Andy Rapoport, MPH*[a], Elsa Friis, PhD[b], Shannon Hill, MPH[a], Sergey Feldman, PhD [bc] Jessica Schleider, PhD[a]
[a]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University; 625 N Michigan Ave 21st floor, Chicago, IL, 60608
[b] Alongside
[c] Allen Institute for Artificial Intelligence

*Indicates equal co-authorship

Abstract

Background: Youth are increasingly experiencing psychological distress. Schools are ideal settings for disseminating mental health support, but they are often insufficiently resourced to do so. Digital mental health tools represent a unique avenue to address this gap. The Alongside digital program is one such tool, intended as universal prevention and early intervention. The platform includes social-emotional learning and self-help wellness features as well as an AI-powered chatbot designed to build coping skills.

Objective: This evaluation aimed to examine the near-term impact of Alongside app usage on students' self-reported mental health outcomes.

Methods: We conducted a non-randomized pilot pragmatic evaluation leveraging anonymized user data. All data came from current Alongside users attending public middle and high schools in Texas and New Mexico, between 10-18 years old. Schools were actively engaged in partnership with Alongside and approved all procedures. Users were asked to complete mental health questionnaires upon app registration and at one-month and three-months post-registration. We conducted pre-registered analyses (https://osf.io/m8t6k) as well as exploratory analyses to determine how symptoms changed over time and what factors (e.g., demographic, app usage) predicted changes in distress.

Results: Analyses revealed statistically significant within-person decreases in overall distress (YP-CORE; primary outcome) from baseline to one-month with a small effect size ($t(42) = 2.21$, $p = 0.03$, $r = 0.34$); however, there was no evidence that scores significantly decreased from baseline to three-months ($W = 1821$, $n = 85$, $p = 0.16$). We found that at three-months, identifying as part of the LGBTQ community predicted greater decreases in distress; otherwise, no demographic factors were significant predictors. In a non-registered exploratory analysis of a subsample of users who reported elevated distress at baseline, decreases in distress were seen at both one-month ($W = 128$, $n = 20$, $p = 0.02$, $r = 0.52$) and three-months ($W = 682$, $n = 42$, $p = 0.004$, $r = 0.45$).

Conclusions: There may be short-term benefits associated with using the Alongside digital program. Further studies are required to determine potential preventative effects.

Keywords: Digital mental health; Schools; Youth

Introduction

Children and adolescents are increasingly experiencing psychological distress and mental health concerns; from 2009 to 2019, the proportion of youth in the United States who reported persistent feelings of sadness or hopelessness rose from 26% to 37% [1]. The COVID-19 pandemic exacerbated already-concerning mental health outcomes in youth, with reports suggesting that minoritized youth (e.g., homeless youth, LGBTQ youth, youth of Color) experienced a disproportionately negative effect of pandemic stressors [2, 3]. Unfortunately, access to needed mental health care for youth is limited by both structural (e.g., financial) and attitudinal barriers (e.g., stigma) [4]. Further, there are increasing calls to approach the youth mental health crisis from a population health perspective by teaching positive coping skills to all youth, which may serve to prevent sub-clinical distress from progressing to the point of needing intervention [5, 6]. To address the rising rates of distress in youth, multilevel solutions are needed that can universally promote wellness while also linking students with additional needs to supports that can be accessed with minimal barriers or delays.

One of the most common settings where youth attempt to access mental health care is schools [7], likely due to schools being a unique setting in which certain barriers to access (e.g., cost, location) are comparatively minimal. Additionally, schools are an ideal setting for universal prevention and early intervention efforts, as they are one of the only settings where the majority of a population (i.e., youth) are found in one place, and the already-educational nature of schools is consistent with providing psychoeducation surrounding mental wellbeing [8, 9]. Students, particularly those that are traditionally underserved, are more likely to receive mental health support and wellness resources in schools independent of structural and policy challenges they may face externally, in their homes and communities [10]. Schools are overwhelmed by mental health needs among their students, however, and are often incapable of meeting the needs of all students who seek services or could benefit from preventative efforts [11, 12]. Although some supports and funding were made available during the pandemic to schools, those are likely not continuing [13]. To prevent worsening of outcomes and provide support as quickly as possible, schools need cost-effective solutions that do not strain an already overstretched system, while filling gaps that existing support cannot.

Digital mental health tools and interventions may be viable options for support that can add to an array of school-based services. Digital technology is widely used by young adults [14]. Many youth already use digital tools in efforts to aid their mental health and everyday distress, with more than 80% searching online, including on social media, for mental health information [15]. Digital resources offer the opportunity for 24/7 support, meaning they can be accessed during acute moments of distress. Further, many students may be hesitant to ask for in-person support due to concerns around stigma; alternatively, they may wish to address issues themselves, as youth often report the desire for self-autonomy [4]. Critically, digital tools cannot and should not replace other school-based mental health supports, such as services delivered by a licensed school-based provider. Rather, they can be cost-effective solutions to deliver psychoeducational and skill-based content for universal or selective prevention. Considering the Multi-Tiered Systems of Support (MTSS) model, a widely used framework for categorizing and implementing school-based mental health supports, digital tools may be best used as “Tier 1” supports for all students, or “Tier 2” supports for students identified as at-risk for developing mental health disorders. Such tools need to be non-stigmatizing and promote student enrollment and participation in a way that is developmentally appropriate, flexible to reach them at the times and in the ways in which they need support, in a student’s native

language, and fundamentally facilitate their sense of autonomy in promoting wellbeing.

The Alongside digital program was designed by experts in digital mental health, ed-tech, and education to leverage the benefits of digital mental health solutions and provide schools with an efficacious and cost-effective method for bolstering student wellbeing without overburdening staff. Alongside provides personalized social-emotional learning and self-help wellness support to students in 4th through 12th grades across the country and identifies students in need of higher levels of support. Alongside includes features such as psychoeducational and mindfulness resources, a 24/7 chatbot that allows students to engage in personalized skill building, and student-initiated onramps to seek additional support. Our objective during this pilot pragmatic evaluation was to leverage anonymized user data to examine the near-term impact of Alongside app usage on student self-reported clinical and wellbeing outcomes. We tested the following pre-registered hypotheses:

Hypothesis 1. We hypothesized that completion of at least one chat within the app (outside an initial onboarding session) would be associated with significant decreases in overall distress, depression & anxiety symptoms, as well as hopelessness and loneliness, and significant increases in mental health treatment expectancies at one-month and three-months post-app registration. Overall distress was our primary outcome while other measures were pre-registered as secondary outcomes.

Hypothesis 2: We hypothesized that users would show similar decreases in overall distress across sociodemographic groups.

Hypothesis 3: We hypothesized that greater engagement in the app (i.e., multiple chat sessions) would predict larger improvements in overall distress among users.

We also tested hypotheses that we did not pre-register as exploratory: 1) We tested each hypothesis separately for users with elevated symptoms at baseline; 2) We tested whether users who showed clinically significant decreases in distress differed in their app usage from users who did not show clinically significant decreases in distress; and 3) We tested whether LGBTQ users differed in their app usage or baseline distress from non-LGBTQ users. Exploratory analyses are described in more detail subsequently in the methods section.

Methods

Sample

This paper reports results from a non-randomized, open, pragmatic evaluation using de-identified data. All participants were current Alongside users enrolled in public middle and high schools in Texas and New Mexico. Users were between the ages of 10 and 18 years old. Upon app registration, users completed a set of questionnaires that asked about their mental health symptoms and help-seeking expectancies (additional details provided below). Approximately one month and three months after app registration, users were asked to complete the questionnaires again via an in-app notification. All questionnaires were available in Spanish and English. Participants were given a month to complete the questionnaires. Teachers and counselors at schools were allowed to answer student questions about the questionnaires or define words. There were no upper limits on sample size. Additionally, Alongside collected app usage data (e.g., which features of the app users engaged in, and how frequently) throughout the three-month period. All data (i.e., questionnaire data and usage data) were collected internally by Alongside as part of their ongoing efforts to evaluate and improve their program. The de-identified, user-level data

was provided to researchers at Northwestern University for users who met the following criteria: 1) completed at least one chat within the app outside of the initial rollout session (i.e., onboarding session) and 2) consented to allow their data to be used for research purposes. As all data were part of a completely anonymous evaluation, this analysis was deemed as nonhuman subjects research in consultation with the institutional review board at Northwestern University.

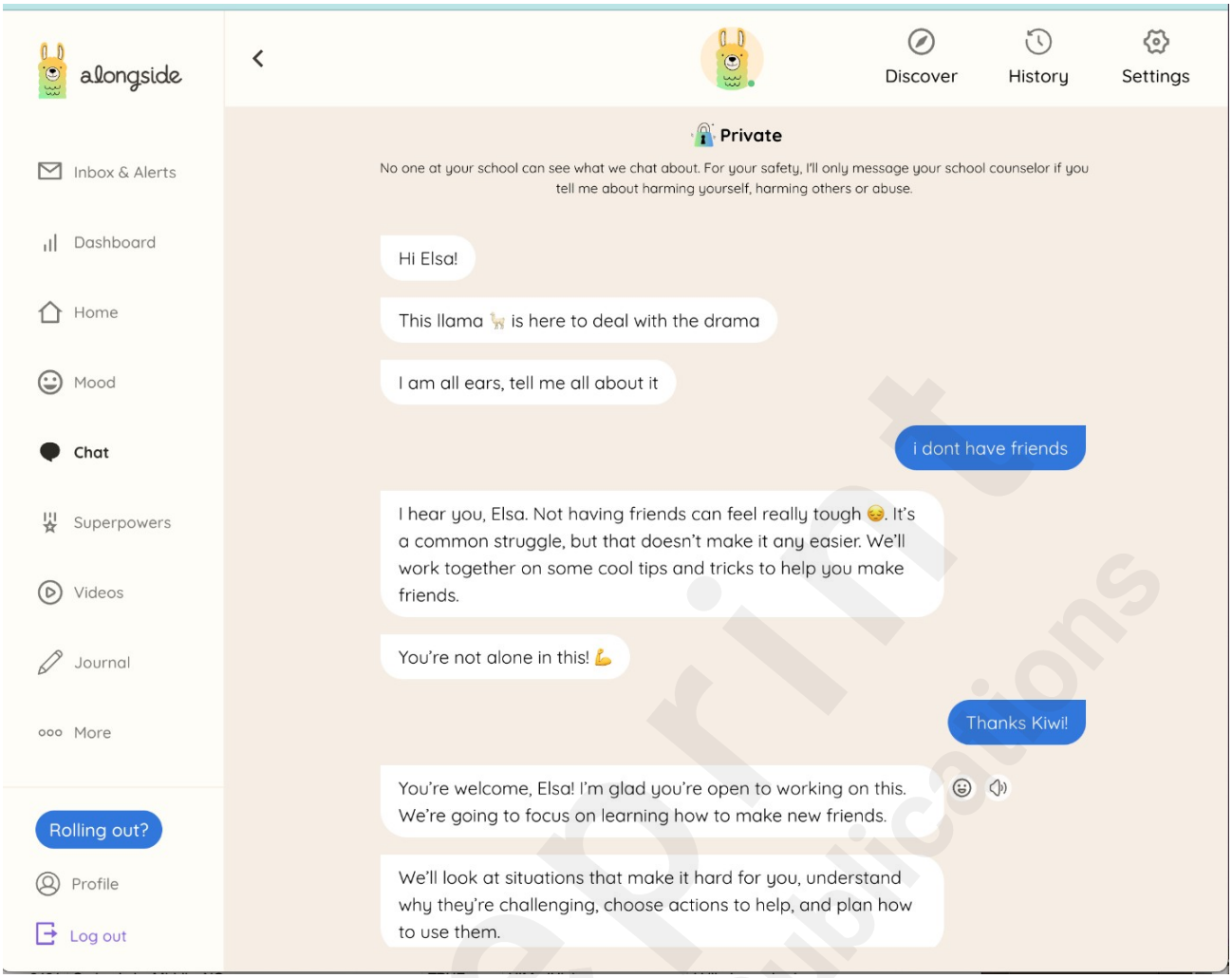
Alongside

The Alongside app contains multiple features designed to provide personalized support for mental health. An initial prototype was developed in June 2022 based on over 50 in-depth interviews with students and school stakeholders. The first version of Alongside was piloted across the country during the 2022-2023 school year. Data from pilots and a 3-month collaborative co-development process with teens refined Alongside's theory of change and features.

At the time of this evaluation, Alongside provided personalized skill building through matching a mix of rule-based and highly regulated generative AI chats grounded in Alongside's skill-building framework called EMPOWER: Engage (provide empathy and validation), Motivate (help the user understand how a skill helps them), Practical examples (learn through practical examples), Operationalize (apply the skill to their own life), Work on it (put the skill into action in real life), Evaluate (follow-up on how it goes), and Reinforce (celebrate progress and effort, not just the outcome). The EMPOWER framework was developed using concepts from motivational interviewing, common factors theory, self-determination theory, and was informed by qualitative review and coding of over 5,000 chats. Skills encompassed multiple theoretical approaches including 3rd wave CBTs including DBT, ACT, Mindfulness-based approaches, narrative therapy, social skills training, positive psychology, and solution focused therapy. Content covered within chats includes mental health coping, academic success, relationship skills, self-awareness, social awareness, responsible decision-making, and self-management. Several of these skill domains map onto social-emotional learning domains identified by CASEL. Chats are meant to be stand-alone supports, such that any one chat session is intended as a self-contained coping resource. Chats could be conceptualized as single-session interventions, which are theoretically grounded in the idea that meaningful insights can occur at any moment and change is possible within one encounter [16]. Therefore, chats are not only available once but are available one at a time.

In the version of the app that was implemented when the data from this evaluation was collected (between January and June 2024), the chatbot included over 120 rule-based and generative AI chat "modules" that aligned with over 50 common challenges students reported in pilot data (i.e., building healthy friendships, boosting self-esteem, reducing anxiety). A screenshot with an example chat is presented in Figure 1. Alongside is a member of the EDSAFE AI Industry council and takes a multifactor approach to ensuring the safe use of generative AI. At the time of this evaluation, generative AI was utilized to provide validating statements to students such as "I understand it must be hard to feel so alone" to add personalization. Users could engage with chats in Alongside in over thirty languages. Students are aware all engagement on Alongside is confidential unless they express a risk of harm to themselves, someone else, or report abuse or harassment. Below we include additional information on Alongside's procedures in these instances. Outside of severe issues, students have autonomy to send a message or share a summary of their chat with their school counselor. Other activities available in the app include psychoeducational videos (e.g., "The Power of Speaking Up: How Asking for Help Changed My Life"), a goal-setting tool, and a journaling tool. A mood-tracking feature was launched in February 2024. Additional UI/UX and content changes occurred during the evaluation period including the creation of additional chat modules.

Figure 1. Example chat within the Alongside platform.



Risk procedures

Alongside utilizes a proprietary machine learning safety monitoring system to screen all student free text inputs and identify any mention on self-harm, suicidal ideation, homicidal ideation, abuse or harassment. If potential risk is detected, the bot asks, "I am concerned you are at risk of serious harm, did I get that right?" The student either selects yes or no. If the student selects yes, they are asked, "Can you confirm what is happening?" Options include “thinking of hurting myself, “thinking of hurting someone else,” “worried about someone hurting me”, or “someone else needs help.” If students say no, one of Alongside’s clinical team members reviews the chat within 24 hours to ensure this was in fact a false trigger. If a student confirms a severe issue, the bot administers a self-reported severity measure such as the Columbia Suicide Severity Rating Scale (CSSRS), directs the student to call a 24/7 hotline such as 988, and walks the student through creating a safety plan based on their specific situation. Concurrently, an automated email and/or text message is generated to notify school staff of the concerning message and information gathered in the risk screening process. In addition, schools have access to a dashboard with the above information, allowing schools to easily track or refer back to important safety information. In addition to these standard procedures, if a student responded positively to “I have thought about harming myself” on the YP-CORE (as described below), the Alongside program automatically matched their user ID to their name/email and generated a real-time notification to the school reporting the student name, email, and response on the survey so the counseling team could follow up and provide support. School student wellness staff have access to a dashboard including de-identified data on app usage, common challenges or topics students are discussing, as well as

information on students with severe issues or who choose to reach out.

Recruitment

Five middle and high schools agreed to partner with Alongside to provide the program to their students and share de-identified data for research purposes. Table 1 provides general information about each school as reported by the National Center for Education Statistics [17].

Table 1. School characteristics.

School	Location	Type	Grades	Size	Urban/ Rural	% Free or Reduced Lunch	Diversity *
School 1	New Mexico	Public	9-12th	621	Remote Town	100%	76% BIPOC
School 2	Texas	Public	9-12th	1157	Distant Town	81%	90% BIPOC
School 3	Texas	Public	6th-8th	750	Suburban	86%	99% BIPOC
School 4	Texas	Public	6th-7th	561	Distant Town	74%	90% BIPOC
School 5	Texas	Public	6th - 8th	772	Large City	93%	94% BIPOC

*% BIPOC is the % enrollment by non-White race/ethnicity, i.e., the % enrollment of American Indian/Alaska Native, Asian, Black, Hispanic, Native Hawaiian/Pacific Islander, and Two or More Races

To implement Alongside, school counseling teams must complete a 45-minute orientation. Additionally, schools must determine parental consent/assent procedures prior to implementing Alongside. Each district has a protocol, largely determined by their superintendent, regarding whether parents must give active consent for their child to use Alongside, or whether parents may choose to opt out of providing consent for their child to use Alongside. All schools included in the current evaluation opted for a parental opt-out procedure for using Alongside and sharing de-identified data for research purposes. School 3 made the decision to exclude students under 13 years old from using the app. All other schools made the decision that this process was sufficient per their school/district/state regulations. Letters were sent by the schools to all parents via their typical communication method (e.g., email, Canvas) to provide information about Alongside, including a statement that de-identified data would be used for research purposes. Letters were sent in both English and Spanish. Parents were given the opportunity to opt their child out of accessing the Alongside program or opt out of de-identified data sharing at any time. No parents opted out of either.

Schools or districts who had active partnership agreements with Alongside but had not yet begun to implement the program with students by January 2024 were invited to participate in the research. All schools were recommended to roll out Alongside via a 20-minute teacher-led roll out activity during class. Schools are provided with posters and cardboard cutouts to promote awareness of Alongside. Additionally, school counseling and wellness staff receive a monthly newsletter with

topical suggestions. For example, in February schools received a newsletter with recommended journal prompts, videos, and chat starters focusing on healthy dating relationships and teen dating violence prevention.

Measures

Demographics

Upon app registration, users were asked to report their age, gender (female, male, prefer not to say), race (American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, or White) and whether they identified as someone who was part of the LGBTQ community (yes or no).

Young Person's CORE (YP-CORE)

The YP-CORE was our pre-registered primary outcome. This 10-item scale asks participants to indicate how often they experience symptoms of distress (e.g., “I’ve felt edgy or nervous”) on a 5-point Likert scale ranging from 0 (Not at all) to 4 (Most or all of the time). Scores are summed and range from 0 to 40, with higher scores indicating greater levels of distress [18]. Among the users who completed the baseline and one-month questionnaires, internal consistency was α = .80 at baseline and α = .80 at one-month. Among the users who completed the baseline and three-month questionnaires, internal consistency was α = .79 at baseline and α = .79 at three-month.

Patient Health Questionnaire-2 (PHQ-2)

This 2-item scale asks participants how often they’ve experienced symptoms of depression (e.g., “Little interest or pleasure in doing things”) over the last two weeks on a 4-point Likert scale ranging from 0 (Not at all) to 3 (Nearly every day). Scores are summed and range from 0 to 6, with higher scores indicating greater levels of depression [19]. Among the users who completed the baseline and one-month questionnaires, internal consistency was α = .69 at baseline and α = .72 at one-month. Among the users who completed the baseline and three-month questionnaires, internal consistency was α = .69 at baseline and α = .74 at three-month.

Generalized Anxiety Disorder Questionnaire-2 (GAD-2)

This 2-item scale asks participants how often they’ve experienced symptoms of anxiety (e.g., “Feeling nervous, anxious or on edge”) over the last two weeks on a 4-point Likert scale ranging from 0 (Not at all) to 3 (Nearly every day). Scores are summed and range from 0 to 6, with higher scores indicating greater levels of anxiety [20]. Among the users who completed the baseline and one-month questionnaires, internal consistency was α = .87 at baseline and α = .67 at one-month. Among the users who completed the baseline and three-month questionnaires, internal consistency was α = .81 at baseline and α = .81 at three-month.

Beck Hopelessness Scale-4 (BHS-4)

This 4-item scale asks participants to indicate their current level of hopelessness through rating their agreement with statements (e.g., “My future seems dark to me”) on a 4-point Likert scale from 0 (absolutely disagree) to 3 (absolutely agree). Scores are summed and range from 0 to 12, with higher scores indicating greater levels of hopelessness [21]. Among the users who completed the baseline and one-month questionnaires, internal consistency was α = .83 at baseline and α = .78 at one-month. Among the users who completed the baseline and three-month questionnaires, internal consistency was α = .84 at baseline and α = .86 at three-month.

UCLA Loneliness Scale-3 (ULS-3)

This 3-item scale asks participants to indicate how often statements reflecting loneliness (e.g. “I feel

left out") are descriptive of them on a 4-point Likert scale from 1 (Never) to 4 (Often). Scores are summed and range from 4-16, with higher scores indicating greater levels of loneliness [22]. Among the users who completed the baseline and one-month questionnaires, internal consistency was $\alpha = .77$ at baseline and $\alpha = .80$ at one-month. Among the users who completed the baseline and three-month questionnaires, internal consistency was $\alpha = .70$ at baseline and $\alpha = .82$ at three-month.

Expectancies

Participants' expectancies regarding mental health treatment were assessed using one item: "On a scale from 1 (not at all helpful) to 10 (extremely helpful), how helpful do you think individual therapy would be in reducing mental health problems, if you struggle/were to struggle with them?" [23].

App Usage Data

Alongside collected data over the three-month period on the following usage metrics per-user: number of app opens, number of chat sessions started, number of chat sessions completed, number of chat messages sent, number of goals created, number of journal entries, number of videos watched, number of mood log entries, and total number of activities completed.

Data Analysis

The RStudio Statistical Program was used to complete data analyses [24]. For each statistical test, a p-value of <.05 was considered statistically significant. Descriptive statistics for participant characteristics (e.g., sex, gender) as well as usage metrics (e.g., number of chats completed) were reported in terms of means, standard deviations, and percentages. Cronbach's Alpha was calculated to evaluate the internal consistency of each quantitative measure at each time point. We used listwise deletion to exclude users who did not have complete data available for outcomes of interest. We additionally conducted either paired t-tests or Wilcoxon signed-rank tests to determine whether demographic or clinical characteristics at baseline differed between participants who completed the one-month assessment and participants who completed the three-month assessment.

Hypothesis 1

To test hypothesis 1, we conducted either paired t-tests or Wilcoxon signed-rank tests. We first tested the assumption of normality (via the Shapiro-Wilk normality test). If the assumption was met, we conducted a paired t-test. If the assumption was not met, we conducted a Wilcoxon signed-rank test. We additionally calculated Cohen's d effect sizes (with 95% confidence intervals). We ran one set of tests comparing outcomes from baseline to the one-month timepoint, and another set of tests comparing outcomes from baseline to the three-month timepoint.

Hypotheses 2 & 3

To test hypotheses 2 and 3, we conducted multiple linear regressions. For hypothesis 2, the outcome variable was YP-CORE scores at the follow-up time-point, and predictor variables included baseline YP-CORE scores and demographic variables (age, race/ethnicity, gender, school, and whether or not they identified as a member of the LGBTQ community). We ran one model with YP-CORE scores at the one-month timepoint as the outcome variable and one model with YP-CORE scores at the three-month timepoint as the outcome variable. Race/ethnicity, sex, and LGBTQ status were dummy coded, with "White/Caucasian", "male," and "No" as the reference groups. School was dummy coded with School 1 as the reference group. For hypothesis 3, the outcome variable was YP-CORE

scores at the follow-up timepoint, and predictor variables included baseline YP-CORE scores and user engagement variables. Each user engagement variable was tested individually; therefore ten models were run (total number of activity sessions, total number of videos watched, total number of chat sessions, total number of journal entries, and total number of goals set; five models for one-month outcomes and five models for three-month outcomes).

Additional analyses

We additionally tested all hypotheses on a sub-sample of users who had clinically elevated distress at baseline (defined as YP-CORE of ≥ 14.1) [25]. We conducted tests to determine whether users who saw a clinically significant decrease in distress (defined as a decrease of ≥ 7.9) [25] differed in their usage data from users who did not see clinically significant decrease in distress. We conducted tests to determine whether LGBTQ users differed in their usage data from non-LGBTQ users.

Pre-registration

This analysis was pre-registered on Open Science Framework (OSF; https://osf.io/m8t6k). There were several deviations from the pre-registered analyses. We pre-registered that we would run longitudinal analyses including all three timepoints (baseline, one-month, and three-months). However, we did not have sufficient data to run these analyses. Instead, we analyzed outcomes from baseline to one-month, and baseline to three-months. Additionally, we pre-registered that we would analyze whether grade predicted changes in overall distress (hypothesis 2) and whether total number of minutes spent in the app predicted changes in overall distress (hypothesis 3). However, these data were unavailable. We did not pre-register but chose to run the following exploratory analyses: 1) testing hypotheses separately for users with elevated symptoms at baseline; 2) testing whether users who showed clinically significant decreases in distress differed in their app usage from users who did not show clinically significant decreases in distress; and 3) testing whether LGBTQ users differed from non-LGBTQ users in their baseline YP-CORE scores or app usage metrics. The decision to add this set of analyses was made in light of emerging questions regarding who the app is most effective for, and which components drove effectiveness. Particularly, the decision to test whether LGBTQ users differed in their app usage from non-LGBTQ users was made due to findings from hypothesis 2 as discussed in the results. Lastly, we pre-registered that we would exclude participants younger than 12 years old, however the sample size was smaller than we anticipated. To increase the available sample size and thereby the likelihood of detecting true effects, we elected to include participants who were at least 10 years old.

Results

Participant retention

Across the five schools, there were 3419 students who were eligible to use Alongside. There was a total of 3861 students at Schools 1-5 (see Table 1); however, School 3 excluded students 13 or below from this study, which removed 442 students from eligibility. A total of 2804 students signed into Alongside. Among these, 1356 consented to sharing de-identified data for research purposes. There were students who enrolled in the study on mobile, and the survey was not available on mobile; therefore, those students did not have the chance to participate. A total of 1114 students who consented to share de-identified data completed the baseline questionnaires upon registration; 27% of these students completed at least one chat outside of the onboarding session. A total of 199 students who consented to share de-identified data completed the one-month questionnaires; 66 (33%) of these students completed at least one chat outside of the onboarding session. A total of 300 students who consented to share de-identified data completed the three-month questionnaires; 116 (38%) of these students completed at least one chat outside of the onboarding session.

Participant Characteristics

A total of 66 users completed both the baseline and the one-month questionnaires. A total of 116 users completed both the baseline and the three-month questionnaires. A total of 26 users completed questionnaires at all three time points. Among users who completed both the baseline and the one-month questionnaires, 97.01% (n = 65) chose to complete the questionnaires in English. They ranged in age from 10-18 years old (M = 14.02, SD = 2.05). The majority were female (52.46%, n = 32). The most commonly self-identified race/ethnicity was Hispanic or Latino (66.10%, n = 39). Eleven users (18.33%) identified as part of the LGBTQ community. Among users who completed both the baseline and the three-month questionnaires, 91.45% (n = 107) chose to complete the questionnaires in English. They ranged in age from 11-18 years old (M = 13.83, SD = 2.17). The majority were female (52.25%, n = 58). The most commonly self-identified race/ethnicity was Hispanic or Latino (72.89%, n = 78). Fifteen users identified as part of the LGBTQ community (14.23%). Table 1 includes detailed information on users' demographic characteristics.

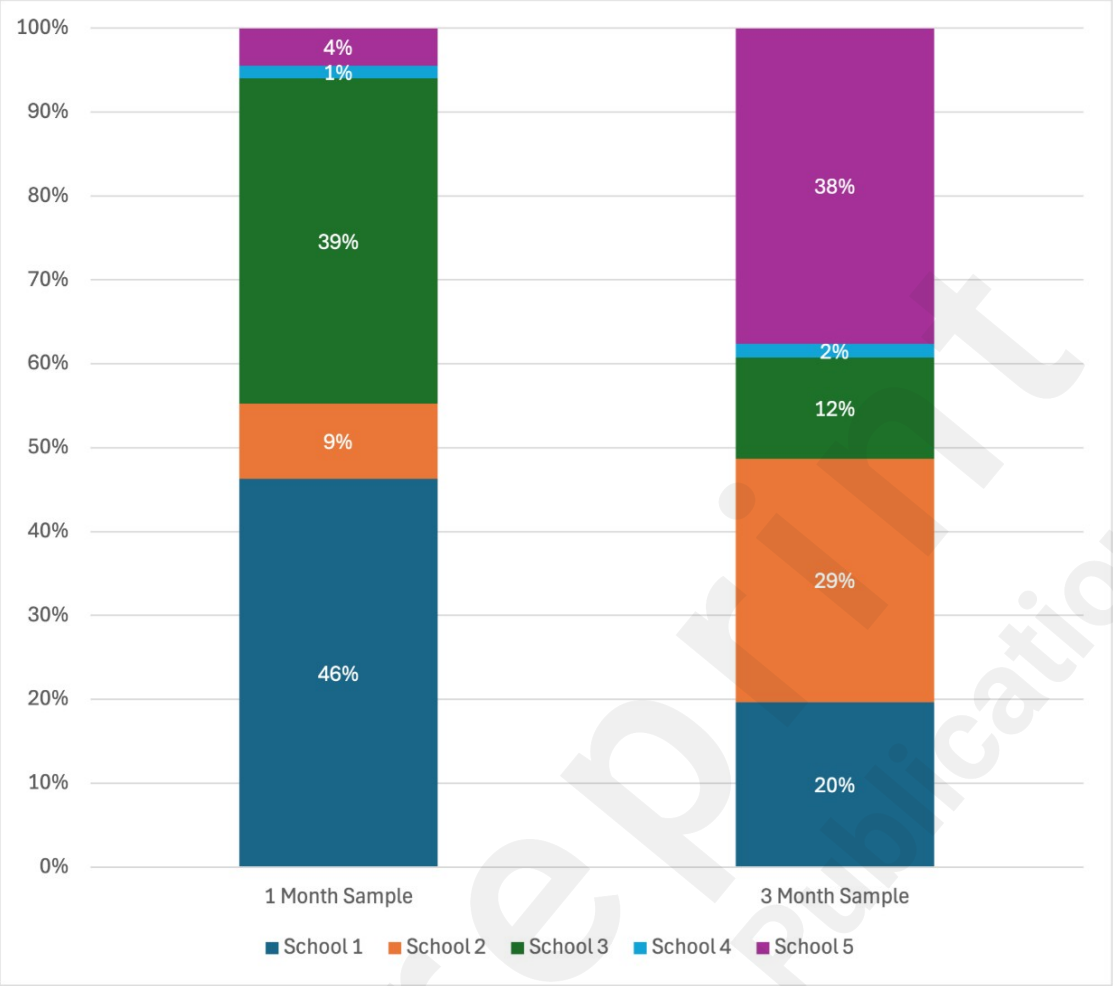
Results from Wilcoxon and Chi-square tests suggested that there were no significant differences in baseline demographic characteristics (age, gender, race/ethnicity, LGBTQ status) between the users who completed both baseline and 1-month questionnaires and users who completed both baseline and three-month questionnaires. Additionally, there was no evidence that the two groups differed in their baseline scores on the YP-CORE, PHQ2, GAD2, BHS4, or ULS3. There were two key differences between the two samples. 1) Users who completed baseline & one-month assessments had significantly lower expectancies scores at baseline (M = 5.62, SD = 2.92) compared to users who completed baseline and 3 month assessments (M = 6.71, SD = 2.58, W = 2572.5, p = 0.02); 2) There were significantly different proportions of students in each school who completed the 1-month vs. the 3-month assessments ($X2(4)$ = 50.64, p < 0.01). Figure 2 displays the different proportions of students from each school that are represented in the two samples.

Table 2. Users' Demographic Characteristics.

	Completed Baseline and Midline	**Completed Baseline and Endline**	**Completed Baseline, Midline, & Endline**
Language			
Spanish	3.0% (n=2)	8.5% (n=10)	3.8% (n=1)
English	97.0% (n=65)	91.5% (n=107)	96.2% (n=25)
Gender			
Female	52.5% (n=32)	52.3% (n=58)	43.5% (n=10)
Male	44.3% (n=27)	44.1% (n=49)	52.2% (n=12)
Prefer not to state	3.3% (n=2)	3.6% (n=4)	4.3% (n=1)
Race/Ethnicity			
American Indian or Alaska Native	3.4% (n=2)	0.9% (n=1)	4.3% (n=1)
Black or African American	8.5% (n=5)	15.0% (n=16)	4.3% (n=1)
White	20.3% (n=12)	6.5% (n=7)	21.7% (n=5)
Asian	1.7% (n=1)	3.7% (n=4)	4.3% (n=1)
Native Hawaiian or Other Pacific Islander	0% (n=0)	0.9% (n=1)	0% (n=0)
Hispanic or Latino	66.1% (n=39)	72.9% (n=78)	65.2% (n=15)
LGBTQ			

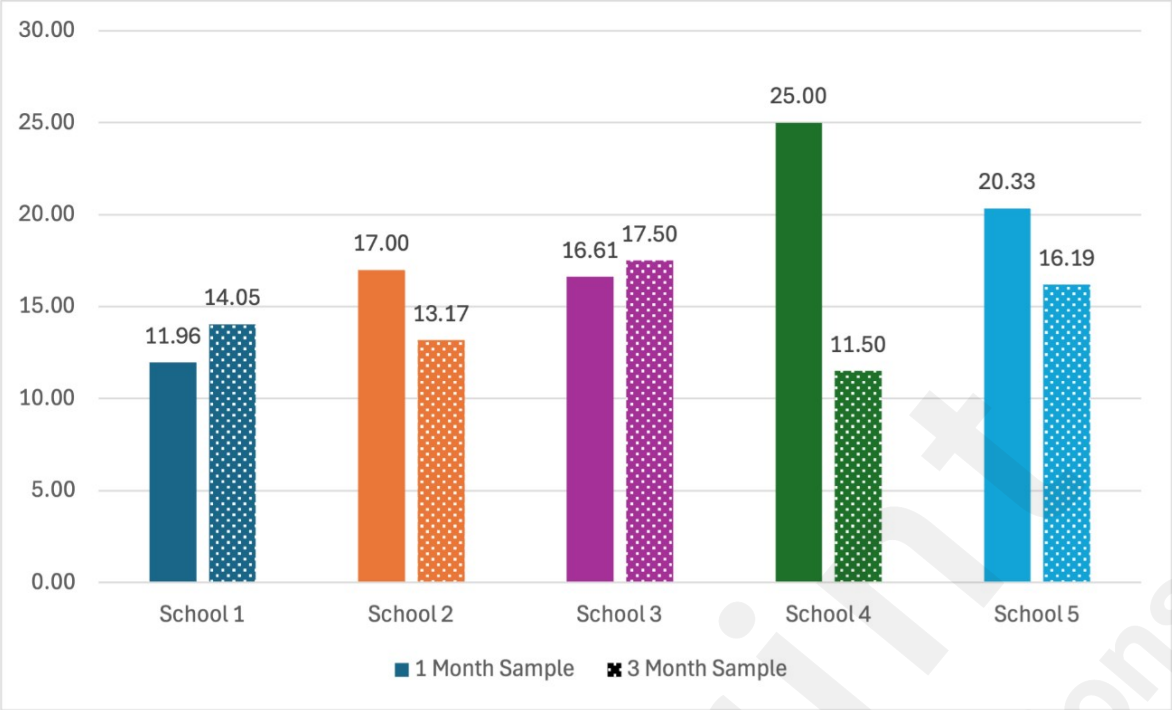
Yes	18.3% (n=11)	14.3% (n=15)	20.8% (n=5)

Figure 2. Proportion of sample from each school.



We additionally tested whether baseline YP-CORE scores differed between schools. Baseline YP-CORE scores were significantly different across schools among users completed both baseline and 1-month questionnaires, $F(1) = 8.41$, $p = .006$, but not among users who completed both baseline and 3-month questionnaires, $F(1) = 0.56$, $p = .46$. Figure 3 portrays baseline YP-CORE scores across schools for both samples.

Figure 3. Example YP-CORE scores at baseline per school.



Usage metrics

Table 3 includes data on users' app usage metrics. Results among the sample of users who completed both baseline and three-month assessments encompass the *entire* three-month period (not the period between one-month and three-months).

Table 3. Usage metrics.

Usage metric	Means and SDs for 1 month data	Means and SDs for 3-month data
App opens	2.85 (2.85)	4.19 (4.44)
Chat sessions	3.09 (3.15)	3.33 (3.72)
Goals created	0.87 (1.48)	1.14 (1.85)
Journals	0.42 (1.09)	0.29 (0.92)
Videos	0.21 (0.67)	0.40 (1.88)
Mood log	0.12 (0.56)	0.17 (0.65)
Time spent chatting (seconds)	988.5 (1211.66)	1103 (1392.89)
Total activities	4.70 (5.13)	5.33 (6.99)
Chat messages sent	44.45 (44.42)	48.19 (54.63)
Chats completed	1.43 (1.21)	1.49 (1.69)

Hypothesis 1: Changes in outcomes across three months

A paired t-test showed statistically significant within-person decreases in YP-CORE scores from baseline (N = 43, M = 14.33, SD = 7.43) to one-month (N = 43, M = 12.42, SD = 7.36; $t(42) = 2.21$, p = 0.03) with a small effect size (r = 0.34). However, a Wilcoxon signed rank test did not show statistically significant within-person decreases in YP-CORE scores from baseline (N = 85, M = 14.67, SD = 7.75) to three-month (N = 85, M = 14.05, SD = 8.06; W = 1821, p = 0.16).

Wilcoxon signed rank tests did not show statistically significant within-person changes in PHQ-2 scores from baseline (N = 50, M = 1.42, SD = 1.42) to one-month (N = 50, M = 1.2, SD = 1.55; W = 264, p = 0.3) or from baseline (N = 98, M = 1.5, SD = 1.61) to three-months (N = 98, M

=1.37, *SD* = 1.65; *W* = 772, *p* = 0.29).

Wilcoxon signed rank tests did not show statistically significant within-person changes in GAD-2 scores from baseline (*N* = 53, *M* = 1.39, *SD* = 1.63) to one-month (*N* = 53, *M* = 1.17, *SD* = 1.44; *W* = 220, *p* = 0.24) or from baseline (*N* = 102, *M* = 1.69, *SD* = 1.76) to three-months (*N* = 102, *M* =1.52, *SD* = 1.87; *W* = 1174, *p* = 0.16).

A Wilcoxon signed rank test did not show statistically significant within-person changes in BHS-4 scores from baseline (*N* = 51, *M* = 2.98, *SD* = 2.82) to one-month (*N* = 51, *M* = 3.0, *SD* = 2.79; *W* = 278, *p* = 0.75). However, a Wilcoxon signed rank test showed statistically significant within-person decreases in BHS-4 scores from baseline (*N* = 96, *M* = 3.62, *SD* = 3.28) to three-months (*N* = 96, *M* = 2.99, *SD*= 3.21; *W* = 1530, *p* = 0.006) with a moderate effect size (*r* = 0.32).

Wilcoxon signed rank tests did not show statistically significant within-person changes in ULS-3 scores from baseline (*N* = 49, *M* = 4.98, *SD* = 1.76) to one-month (*N* = 49, *M* = 4.82, *SD* = 1.73; *W* = 295, *p* = 0.56) or from baseline (*N* = 102, *M* = 5.19, *SD* = 1.82) to three-months (*N* = 102, *M* =4.88, *SD* = 1.89; *W* = 1400, *p* = 0.06).

Wilcoxon signed rank tests did not show statistically significant within-person changes in expectancies scores from baseline (*N* = 56, *M* = 5.46, *SD* = 2.94) to one-month (*N* = 56, *M* = 5.79, *SD* = 3.04; *W* = 450, *p* = 0.6) or from baseline (*N* = 103, *M* = 6.72, *SD* = 2.62) to three-months (*N* = 103, *M* = 6.15, *SD* = 2.99; *W* = 1778, *p* = 0.1).

Hypothesis 2: Demographic predictors of change in YP-CORE

Results from linear regressions showed there was no evidence that sociodemographic factors (age, race/ethnicity, LGBTQ identity, gender, school) significantly predicted changes in YP-CORE scores from baseline to one month. There was evidence that identifying as part of the LGBTQ community was a significant predictor of lower distress from baseline to three-months (*t* = -2.07, *p* = 0.04). There was no evidence that any other sociodemographic factors predicted changes in YP-CORE scores from baseline to three months.

Hypothesis 3: App usage predictors of change in YP-CORE

Results from linear regressions showed there was no evidence that total number of activity sessions, videos watched, chat sessions, journal entries created, or goals set predicted changes in YP-CORE scores at one-month or three-months.

Exploratory analyses

Elevated subsample

In a sub-sample of users who had clinically-elevated distress at baseline (defined as YP-CORE of ≥ 14.1) [25], a Wilcoxon signed rank test showed statistically significant within-person decreases in YP- CORE scores from baseline (*N* = 20, *M* = 20.6, *SD* = 5.17) to one-month (*N* = 20, *M* = 16.9, *SD* = 5.97; *W* = 128, *p* = 0.02) with a large effect size (*r* = .52). Additionally, a Wilcoxon signed rank showed statistically significant within-person decreases in YP-CORE scores from baseline (*N* = 42, *M* = 20.95, *SD* = 5.26) to three-months (*N* = 42, *M* = 17.71, *SD* = 7.51; *W* = 682, *p* = 0.004), with a moderate effect size (*r* = .45).

A Wilcoxon signed rank showed statistically significant within-person decreases in GAD-2 scores from baseline (*N* = 24, *M* = 2.38, *SD* = 1.84) to one-month (*N* = 24, *M* = 1.63, *SD* = 1.69; *W* = 122, *p* = 0.03), however there did not appear to be statistically significant within-person changes in GAD-2 scores from baseline (*N* = 44, *M* = 2.64, *SD* = 1.79) to three-months (*N* = 44, *M* = 2.11, *SD* = 2.03; *W* = 398, *p* = 0.08)

A paired t-test did not show statistically significant within-person decreases in BHS-4 scores from baseline (*N* = 20, *M* = 4.9, *SD* = 3.01) to one-month (*N* = 20, *M* = 4.05, *SD* = 2.96; *t*(19) = 1.29, *p* = 0.21), however a Wilcoxon signed rank test showed statistically significant within-person

decreases in BHS-4 scores from baseline (N = 40, M = 5.5, SD = 3.39) to three-months (N = 40, M = 4.3, SD = 3.45; W = 504, p = 0.007) with a moderate effect size (r = 0.32).

Results on other secondary outcomes (PHQ-2, ULS-3, expectancies) were not significant at one-month or three-months.

Results from linear regressions showed that, among users with clinically elevated distress at baseline, identifying as part of the LGBTQ community was a significant predictor of lower distress at 1-month (t = -3.04, p = 0.02) and at 3 months (t = -2.44, p = 0.02). There was no evidence that any other sociodemographic factors predicted changes in YP-CORE scores at one-month or three months.

There was no evidence that total number of activity sessions, videos watched, chat sessions, journal entries created, or goals set predicted changes in YP-CORE scores at one-month or three-months.

Differences in usage between users who did and did not see clinically significant decreases in distress

We categorized users based on whether they showed clinically significant changes in distress. According to Twigg et al (2015) the overall reliable change index value for the YP-CORE is 7.9 (including all age bands and genders). Among participants who completed both the baseline and 1-month YP-CORE, 13.95% (n = 6) saw a clinically-significant decrease in distress (scores decreased by 7.9 or more), 2.33% (n = 1) saw a clinically significant increase in distress (scores increased by 7.9 or more), and 83.72% (n = 36) did not see clinically significant changes (changes in scores were < 7.9 and > -7.9). Among participants who completed both the baseline and 3-month YP-CORE, 11.76% (n = 10) saw a clinically significant decrease in distress, 10.59% (n = 9) saw a clinically significant increase in distress, and 77.65% (n = 66) did not see clinically significant changes.

At one-month, users who saw clinically significant decreases in distress opened the app significantly more (N = 6, M = 5, SD = 3.22) compared to individuals who did not see clinically significant changes in distress (N = 36, M = 2.94, SD = 3.14, W = 163, p = 0.04). Users who saw clinically significant decreases in distress also saved their mood more often (N = 6, M = 0.83, SD = 1.6) compared to individuals who did not see clinically significant changes in distress (N = 36, M = 0.06, SD = 0.33, W = 141, p = 0.009), spent more seconds chatting (N = 6, M = 1463, SD = 1121) compared to individuals who did not see clinically significant changes in distress (N = 36, M = 1045, SD = 1417, W = 173, p = 0.02), and sent more chat messages (N = 6,M = 57.2, SD = 24.8) compared to individuals who did not see clinically significant changes in distress (N = 36, M = 47.8, SD = 54.5, W = 164, p = 0.045). No differences emerged for number of chat sessions, goals created, journal entries, videos watched, or chats completed.

At three-months, users who had a clinically significant decrease in distress saved their mood significantly more (N = 10, M = 0.7, SD = 1.34) compared to uses who did not have clinically significant changes (N = 66, M = 0.12, SD = 0.57; F(2, 82) = 3.56, p = 0.03). No differences emerged for number of app opens, chat sessions, goals created, journal entries, videos watched, time spent chatting, total activities, chat messages sent, or chats completed.

Differences between LGBTQ and non-LGBTQ users

There was no evidence that LGBTQ users differed in their baseline YP-CORE scores at in either the one-month (W =136.5, p = 0.13) or three-month samples (W = 354, p = 0.16). At one-month, LGBTQ users opened the app significantly more (N = 11, M = 5.27, SD =4.92) compared to non-LGBTQ users (N = 49, M = 2.49, SD = 2.05; W = 148, p = 0.02) and spent more time chatting in seconds (N = 11, M = 1689, SD =1554) compared to non-LGBTQ users (N = 49, M = 919, SD = 1154; W = 155, p = 0.03). At one-month, no differences emerged for number of chat sessions, goals created, journal entries, videos watched, mood log entries, chat messages sent, or chats completed. At

three-months, LGBTQ users used the journal tool significantly more (N = 15, M = 1.2, SD= 2.08) compared to non-LGBTQ users (N = 90, M = 0.18, SD = 0.51; W = 397, p < 0.01) At three-months, no differences emerged for number of app opens, chat sessions, goals created, videos watched, mood log entries, time spent chatting, chat messages sent, or chats completed.

Discussion

In light of increasing mental distress among youth and limited resources available in schools, innovative solutions are necessary. This paper reports findings from a pragmatic evaluation of Alongside, a digital mental health program designed to improve mental wellbeing among middle and high schoolers while remaining cost-effective and scalable.
We found evidence that overall distress (our primary outcome) decreased over one-month among all Alongside users who completed at least one chat outside of the initial onboarding session, although decreases were not sustained three months post-registration. In a subsample of users who had elevated distress at baseline, however, findings were both stronger and sustained over time. Findings on secondary outcomes related to depression, anxiety, loneliness, and treatment expectancies were largely null, although some positive effects were seen in hopelessness among the full sample and both anxiety and hopelessness among the elevated subsample. While positive results are promising, it is critical that we contextualize them alongside null findings and consider areas for future work.

There are multiple reasons why decreases in overall distress may have been seen among the full sample at one-month but not sustained at three-months. First, the Alongside program may be most effective in the short-term for in-the-moment distress, but not currently effective for implementing long-term change. This does not negate the meaningfulness of short-term improvements. In school settings where resources are spread thin, it may prove better to provide easily accessible tools that promote short-term improvements, as opposed to no support at all for those who have sub-clinical levels of distress (i.e., those not in need of immediate intervention), which is the common approach currently. Second, a decrease in usage of the Alongside program between one-month and three-months could explain why users did not see sustained benefits. When comparing app usage between the sample that completed the three-month questionnaires and the sample that completed the one-month questionnaires, we see marginal differences; for example, the average number of app opens at three months (across the entire three-month period) was 4.19 while the average number of app opens at one-month was 2.85, a difference of only 1.34. We discuss the unclear relationship between app usage and symptom change in more detail later in this discussion, but it is possible that a lack of benefit beyond one month could be attributed to low intervention uptake beyond one month. Lastly, there could be discrepancies between the sample of users who completed one-month questionnaires and users who completed three-month questionnaires. We found that the proportion of students from each school significantly differed between the two samples, and among the users who completed the one-month questionnaires, distress at baseline significantly differed between users from different schools. At the same time, there was no evidence that baseline distress or demographic characteristics significantly differed between users who completed one-month questionnaires and users who completed three-month questionnaires. Still, it is possible that the practical limitations and implementation challenges that Alongside faced in recruiting users to complete questionnaires, discussed in more detail below, contributed to discrepancies between findings at one-month and three-months.

Contrary to our hypotheses, we did not see effects on any secondary outcomes related to depression, anxiety, loneliness, or treatment expectancies among the full sample. We caution against interpreting null results in this initial evaluation as a sign to abandon the Alongside program, particularly in light of other potential benefits on short-term distress or benefits among the subsample of users with elevated distress; rather, null findings may demonstrate the need to continue improving and testing the Alongside program through iterative design processes if the program is intended to

effect these outcomes in a universal sample. Alternatively, the Alongside program may be better suited for users who were already experiencing distress. Our findings pointing to stronger effects among the clinically elevated subsample are unsurprising in light of previous research demonstrating larger effect sizes for school-based digital interventions among clinically elevated subsamples of students [26, 27]. It is possible that the Alongside program is better positioned to be a Tier 2 (selective) intervention in the Multi-Tiered Systems of Support (MTSS) model rather than a Tier 1 (universal) prevention tool, but more research is needed to consider the intentional implementation of Alongside at this level.

Our findings related to Hypothesis 2 suggest that LGBTQ users may see greater decreases in distress than non-LGBTQ users. Otherwise, there were no demographic predictors of changes in overall distress. LGBTQ users may have been more likely to benefit from using the Alongside program because LGBTQ youth often face unique barriers to accessing other mental health care [28]. We also found that LGBTQ users differed from non-LGBTQ users by using several app features more often than non-LGBTQ users. It is possible that the Alongside program is particularly well-suited for dissemination among LGBTQ students.

Our findings related to Hypothesis 3 suggest that there is not a relationship between users' degree of app usage and users' change in overall distress. It is possible that one use may be enough for someone to benefit from, while users who repeatedly come back to a program may do so because they are not getting what they need. Individual differences may better explain the complicated relationship between app usage and symptom change. At the same time, we found that users with clinically significant changes in distress used several app features more often than users who did not see clinically significant changes. It is possible that the relationship is not proportional (e.g., greater usage confers greater impact) but that there is a minimal level of usage that can confer some impact, and the degree of impact is individually variable; additional testing is necessary. Alternatively, the strength of the relationship between usage metrics and change in distress may exist but may be small enough that our limited sample size was not powered to detect it. In sum, we cannot discount that the degree of program usage influences the degree of clinical impact to some extent, but it does not appear to be a particularly meaningful contributor.

Multiple limitations impact our ability to confidently interpret the findings of the current evaluation. Firstly, this was not a randomized evaluation. Without a control group to compare to, we cannot confirm that any positive outcomes were due solely to use of the Alongside program. Simultaneously, we cannot determine whether there were preventative effects such that users who did not see decreases in distress may have otherwise seen increases in distress had they not had access to the program. Secondly, the primary and secondary outcomes were deficit-focused and did not include wellbeing measures. The dual-factor model of mental health [29] suggests that wellbeing and distress are distinct constructs; null findings related to distress do not preclude positive impacts on wellbeing, however, the current evaluation cannot elucidate the impact of the Alongside program on wellbeing. Lastly, Alongside faced a plethora of barriers to implementation that impacted the number of users who responded to questionnaires. One school had over 36% of students screen positive for thoughts of self-harm which contributed to hesitancy around pushing for follow-up assessments as the number of students identified overwhelmed the counseling staff. Additionally, the three-month endline overlapped with state testing in all Texas schools leading to difficulty promoting follow-up. Finally, one school asked students to return school devices before the 3-month endline and the survey was not available on personal devices. Only 26 users did questionnaires at all three timepoints, therefore we could not reliably analyze change over time including all timepoints. The timing of the study in the school year may also have contributed to low numbers of students who responded to questionnaires.

The current evaluation is strengthened by its pragmatic nature and pre-registered analysis. The research-to-practice gap is maintained due to a lack of research on real-world implementation, making findings from highly controlled research studies ungeneralizable to the constantly changing

and complicated settings in which the intervention is intended to be deployed. By testing an intervention while it is actively being implemented in a non-research setting, it is more likely that findings will be reflective of the ultimate impact an intervention can have. Additionally, this analysis was pre-registered and all deviations from the pre-registration are reported. Pre-registration addresses the “file drawer” problem within the scientific community, wherein only statistically significant results are shared, and null findings are not published. Pre-registration is particularly important when considering the lack of transparency that many commercial digital mental health tools maintain. Digital mental companies largely do not test their products through rigorous research at early stages of development. Transparent testing upfront is necessary for companies to then transparently improve their platforms.

This analysis highlights several opportunities for continued research. Randomized evaluations with large numbers of participants are necessary to reliably parse the effects of the Alongside program. Further evaluations may test future iterations of the Alongside program as it is updated to improve its features or add content that may directly influence secondary outcomes such as depression or anxiety. Additionally, hybrid effectiveness-implementation trials may evaluate the Alongside program as implemented via Tier 2 for students who are experiencing early signs of mental distress.

Conclusion

Providing digital mental health tools through schools may help youth to access needed health supports & resources. In this non-randomized pilot pragmatic evaluation of the Alongside digital mental health tool, we assessed the near-term impact of app usage on the self-reported mental health outcomes of middle and high school students. We found that Alongside can confer some short-term benefits, but that those benefits are not necessarily sustained in a universal sample. Future work may use a larger sample or randomized design to further understand when and for whom Alongside may be most effective.

Acknowledgements

This research was supported by a research grant from Alongside (PI: Schleider). Outside of the present work, Schleider and her research are supported by funding from the National Institutes of Health (DP5OD028123), the National Science Foundation (2141710), the Health Research and Services Administration (U3NHP45406-01-00), the Society of Clinical Child and Adolescent Psychology, Hopelab, the Child Mind Institute, Kooth, Understood, the Montana Department of Public Health and Human Services, the American Psychological Foundation, and the Florida Institute for Child Welfare. KAC is supported by funding from the National Institutes of Health (F31MH139185).

Conflicts of Interest

JLS has serves on the Scientific Advisory Board for Walden Wise and the Clinical Advisory Board for Koko (unpaid), has received consulting fees from United Health and Woebot, and receives book royalties from New Harbinger; Oxford University Press; and Little, Brown Book Group. She is co-founder and chief scientific advisor for Mindly. No Mindly products were used or are referenced in the present manuscript. KAC is a part-time employee at Lyra Health, Inc. EAF is an employee of Alongside. SF is an employee of Alongside.

References

1. Center for Disease Control and Prevention, *Youth Risk Behavior Survey Data Summary & Trends Report: 2009-2019* 2019, Center for Disease Control and Prevention. p. 61.
2. Fortuna, L.R., et al., *The Impact of COVID-19 on Anxiety Disorders in Youth: Coping with Stress, Worry, and Recovering from a Pandemic.* Child Adolesc Psychiatr Clin N Am, 2023. **32**(3): p. 531-542.
3. Gruber, J., et al., *Mental health and clinical psychological science in the time of COVID-19: Challenges, opportunities, and a call to action.* American Psychologist, 2021. **76**(3): p. 409.
4. Gulliver, A., K.M. Griffiths, and H. Christensen, *Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review.* BMC psychiatry, 2010. **10**: p. 1-9.
5. Colizzi, M., A. Lasalvia, and M. Ruggeri, *Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care?* International journal of mental health systems, 2020. **14**: p. 1-14.
6. Purtle, J., et al., *Population-based approaches to mental health: history, strategies, and evidence.* Annual Review of Public Health, 2020. **41**(1): p. 201-221.
7. Duong, M.T., et al., *Rates of mental health service utilization by children and adolescents in schools and other common service settings: A systematic review and meta-analysis.* Administration and Policy in Mental Health and Mental Health Services Research, 2021. **48**: p. 420-439.
8. Cohen, K.A., et al., *Brief school-based interventions targeting student mental health or well-being: a systematic review and meta-analysis.* Clinical child and family psychology review, 2024. **27**(3): p. 732-806.
9. Herrenkohl, T.I., *Cross-system collaboration and engagement of the public health model to promote the well-being of children and families.* Journal of the Society for Social Work and Research, 2019. **10**(3): p. 319-332.
10. Fazel, M., et al., *Mental health interventions in schools in high-income countries.* The Lancet Psychiatry, 2014. **1**(5): p. 377-387.
11. Heinrich, C.J., A. Colomer, and M. Hieronimus, *Minding the gap: Evidence, implementation and funding gaps in mental health services delivery for school-aged children.* Children and youth services review, 2023. **150**: p. 107023.
12. Panchal, N., C. Cox, and R. Rudowitz, *The landscape of school-based mental health services.* Kaiser Family Foundation, 2022. **6**.
13. Hoover, S., *Moving Upstream: Opportunities in School Mental Health.* Scanlan Center for School Mental Health practice briefs, 2023.
14. Curtis, B.L., et al., *Comparison of smartphone ownership, social media use, and willingness to use digital interventions between generation Z and millennials in the treatment of substance use: cross-sectional questionnaire study.* Journal of medical Internet research, 2019. **21**(4): p. e13050.
15. Pretorius, C., et al., *Young people seeking help online for mental health: cross-sectional survey study.* JMIR mental health, 2019. **6**(8): p. e13524.
16. Schleider, J.L., et al., *Future directions in single-session youth mental*

health interventions. Journal of Clinical Child & Adolescent Psychology, 2020. **49**(2): p. 264-278.
17. The Institute of Education Sciences - National Center for Education Statistics. *School Search*. Available from: https://nces.ed.gov/ccd/schoolsearch/.
18. Twigg, E., et al., *The Young Person's CORE: Development of a brief outcome measure for young people.* Counselling and Psychotherapy Research, 2009. **9**(3): p. 160-168.
19. Richardson, L.P., et al., *Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents.* Pediatrics, 2010. **125**(5): p. e1097-e1103.
20. Toussaint, A., et al., *Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7).* Journal of affective disorders, 2020. **265**: p. 395-401.
21. Perczel Forintos, D., et al., *Proposal for a short version of the Beck Hopelessness Scale based on a national representative survey in Hungary.* Community mental health journal, 2013. **49**: p. 822-830.
22. Hughes, M.E., et al., *A short scale for measuring loneliness in large surveys: Results from two population-based studies.* Research on aging, 2004. **26**(6): p. 655-672.
23. Schleider, J.L. and J.R. Weisz, *Parent expectancies and preferences for mental health treatment: the roles of emotion mind-sets and views of failure.* Journal of Clinical Child & Adolescent Psychology, 2018. **47**(sup1): p. S480-S496.
24. Allaire, J., *RStudio: integrated development environment for R.* Boston, MA, 2012. **770**(394): p. 165-171.
25. Twigg, E., et al., *Acceptability, reliability, referential distributions and sensitivity to change in the Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE) outcome measure: replication and refinement.* Child Adolesc Ment Health, 2016. **21**(2): p. 115-123.
26. Osborn, T.L., et al., *Single-session digital intervention for adolescent depression, anxiety, and well-being: Outcomes of a randomized controlled trial with Kenyan adolescents.* Journal of consulting and clinical psychology, 2020. **88**(7): p. 657.
27. Fitzpatrick, O.M., et al., *Project SOLVE: Randomized, school-based trial of a single-session digital problem-solving intervention for adolescent internalizing symptoms during the coronavirus era.* School Mental Health, 2023. **15**(3): p. 955-966.
28. Choi, K.R., L.E. Wisk, and B.T. Zima, *Availability of LGBTQ mental health services for US youth, 2014 to 2020.* JAMA pediatrics, 2023. **177**(8): p. 865-867.
29. Antaramian, S.P., et al., *A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning.* American Journal of Orthopsychiatry, 2010. **80**(4): p. 462.