Check for updates

What Makes Digital Support Effective? How Therapeutic Skills Affect Clinical Well-Being

WENJIE YANG* and ANNA FANG*, Carnegie Mellon University, USA
RAJ SANJAY SHAH, Georgia Institute of Technology, USA
YASH MATHUR, Carnegie Mellon University, USA
DIYI YANG, Stanford University, USA
HAIYI ZHU, Carnegie Mellon University, USA
ROBERT KRAUT, Carnegie Mellon University, USA

Online mental health support communities, in which volunteer counselors provide accessible mental and emotional health support, have grown in recent years. Despite millions of people using these platforms, the clinical effectiveness of these communities on mental health symptoms remains unknown. Although volunteers receive some training on the therapeutic skills proven effective in face-to-face environments, such as active listening and motivational interviewing, it is unclear how the usage of these skills in an *online context* affects people's mental health. In our work, we collaborate with one of the largest online peer support platforms and use both natural language processing and machine learning techniques to examine how one-on-one support chats on the platform affect clients' depression and anxiety symptoms. We measure how characteristics of support-providers, such as their experience on the platform and use of therapeutic skills (e.g. affirmation, showing empathy), affect support-seekers' mental health changes. Based on a propensity-score matching analysis to approximate a random-assignment experiment, results shows that online peer support chats improve both depression and anxiety symptoms with a statistically significant but relatively small effect size. Additionally, support providers' techniques such as emphasizing the autonomy of the client lead to better mental health outcomes. However, we also found that the use of some behaviors, such as persuading and providing information, are associated with worsening of mental health symptoms. Our work provides key understanding for mental health care in the online setting and designing training systems for online support providers.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: online communities, mental health, peer support, social computing

ACM Reference Format:
Wenjie Yang*, Anna Fang*, Raj Sanjay Shah, Yash Mathur, Diyi Yang, Haiyi Zhu, and Robert Kraut. 2024. What Makes Digital Support Effective? How Therapeutic Skills Affect Clinical Well-Being. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 190 (April 2024), 29 pages. https://doi.org/10.1145/3641029

1 INTRODUCTION

Mental health problems continue to rise globally and under-treatment of serious mental health problems remains a major problem, with more than one in ten people living with a mental health

Authors' addresses: Wenjie Yang*; Anna Fang*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Raj Sanjay Shah, Georgia Institute of Technology, USA; Yash Mathur, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Diyi Yang, Stanford University, USA; Haiyi Zhu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Robert Kraut, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

ACM 2573-0142/2024/4-ART190
https://doi.org/10.1145/3641029

disorder [20]. Although there is significant evidence supporting the effectiveness of professional treatments, such as therapy, a substantial proportion of people with mental health problems fail to receive any treatment because they lack access to services or have needs unmet by health services [47, 64]. As a result, peer-to-peer support through online mental health communities (OMHCs) has emerged as an accessible tool for achieving mental and emotional support. OMHCs include sites like 7 Cups and TalkLife, which usually provide free 24/7 peer-to-peer support from volunteers; online support communities are thus able to provide support at scale for a wide variety of mental and emotional problems [25, 30].

Despite the many benefits available through OMHCs, support providers on online platforms receive relatively little training [102] in contrast to the extensive training for mental health professionals and even volunteers for crisis intervention programs [10, 34, 66, 70]. OMHCs often require volunteer support providers to complete short training sessions along with optional specialized training; for example, 7Cups.com provides short training sessions based on skills established in face-to-face mental health support such as active listening, empathy, and motivational interviewing (MI) techniques. Another online support platform, TalkLife, provides small-group and module-led training for peer support techniques including asking questions, showing empathy, and reflection. However, there is little known about whether these therapeutic skills and other "common factors", which are effective for improving people's mental health when used by professionals in offline settings [43, 56, 68], are effective when used by peer supporters online. Resolving this uncertainty is especially important, as doing so can inform the design of online support providers' training.

To address these questions, we collaborated with a large online peer support platform in order to examine the longitudinal and clinical effectiveness of online peer-to-peer support. Using natural language processing and machine learning techniques, we analyze changes in support-seekers' depression and anxiety in a dataset spanning over two years containing over 8-million anonymous text-based support chats between volunteer support counselors and online support-seekers. We explore the effectiveness of support behaviors, like asking open-ended questions and providing reflections involved in motivational interviewing and expressing empathy, used by these OMHC volunteers and examine how these behaviors are associated with long-term improvement in support-seekers' depression and anxiety. Our work extends past research, which has largely assessed outcomes of OMHCs on various non-clinical outcomes (e.g. mood, retention [2, 84, 97]) in cross-sectional studies. Instead, we measure the effectiveness of support chats on changes in clinical outcomes in a longitudinal study. Assessing the presence and severity of mental illness symptoms are largely under-studied in the online context despite being the most important outcomes used in important primary care settings, from diagnosis and severity measurement [48, 82] to monitoring responses to mental health treatments [49, 54].

Specifically, our study addresses the following research questions:

RQ1. Does participating in support conversations with volunteers on online peer support platforms improve people's depression and anxiety?

RQ2. What volunteer characteristics and the therapeutic techniques they use in these conversations affect mental health outcomes?

Our study found that participating in online peer support chats with volunteer counselors improves symptoms of depression and anxiety with a statistically significant but relatively small effect size. In particular, having a single online support chat on average improved depression symptom scores by 1.6% and anxiety symptoms by 0.6%. Notably, these improvements were larger for people who were initially more distressed. We also found some established therapeutic techniques were associated with improvements in symptoms of depression and anxiety while others were associated with more symptoms. For example, the motivational interviewing (MI) technique of reflecting (capturing and returning something the client has said themselves) is effective for improving depression symptoms, while the MI technique of information-giving was actually associated with worse symptoms for both depression and anxiety. Overall, our findings provide insight into the efficacy of online support platforms and can lead to future work on training for online mental health support.

2 RELATED WORK

Below, we review prior work on online mental health communities, therapeutic skills that may lead to better mental health outcomes in both online and offline settings, and success metrics for evaluating OMHCs.

Table 1 situates our work among past literature, which is further reviewed in the sections below. We organize the most relevant prior work and differentiate our study though its use of clinical outcomes, large-scale longitudinal data, and evaluation of support providers techniques.

2.1 Online Mental Health Communities

OMHCs include support groups within general-purpose social networks, such as Facebook groups dedicated to specific health issues [9, 35, 37, 99], and platforms aimed entirely at health support, such as CrisisTextLine [1], 7Cups.com [2], and TalkLife [3] that connect support seekers with volunteer peer counselors for supportive chats. Peer support through online communities is able to fill gaps in accessing health services for many, helping those whose needs are unmet by traditional resources [32, 62, 78] or individuals who lack adequate peer networks in their daily lives to achieve needed support [44, 46, 89, 91]. Support through OMHCs has been found to have numerous benefits, such as yielding meaningful relationships and increasing trust in getting mental health treatment [28, 74, 81]. Additionally, OMHCs circumvent many key barriers that prevent help-seeking and allow immediate addressing of mental health needs [23]. Both general purpose social networks and ones focused exclusively on mental health are important in spreading health information, reducing harmful thoughts, empowering help-seeking for stigmatized populations, reducing suicidal ideation, and fundraising or spreading awareness for health issues [21, 22, 36, 41, 77]. There is evidence, though, to support that support provided specifically in online mental health platforms may deviate from conventional practices in traditional, offline therapy contexts; for example, there are less distinct boundaries between support-seeker and support-provider, and volunteer counselor behaviors are influenced by training that takes place through platform-provided resources [102].

2.2 Skills for Counseling and Therapy

Effective mental health support depends on numerous behaviors and strategies used [18]. In professional therapy, therapeutic alliance – the extent to which clients and therapists work collaboratively on common goals and connect emotionally – has been the most studied and demonstrably effective element in psychotherapy [67, 90]. Therapist behaviors, including exploration, showing empathy, reflection, accurate interpretation, and active listening, lead to better therapeutic alliance and mental health outcomes [1, 68, 69, 95]. Given that there exist core sets of mental health support skills ("micro-skills") that transcend specific therapies [68], our work explores how validated counseling techniques (e.g. expressing empathy, motivational interviewing techniques) affect online support-seekers' clinical outcomes. As noted by Hall and Horvath regarding mental health providers, "*micro-skills are the building blocks of effective therapeutic communication … [While they] do not encompass the entire therapeutic skill set, they are useful and accessible in the early stages of*

[1] crisistextline.org
[2] 7cups.com
[3] talklife.com

Table 1. Relevant prior work classified across dimensions including sample size, platform specifications, mental health outcomes, and analysis predictors.

	Large-scale sample	**Sample size**	**1-on-1 support chats**	**Non-professional support providers**	**Clinical mental health outcomes**	**Based on standardized therapy skills**
Our Work	✓	74K people (depression) 42K people (anxiety)	✓	✓	✓	✓
Althoff et al. [2]	✓	15K chats	✓			
Chikersal et al. [15]	✓	54K people		✓	✓	
Sharma et al. [87]	✓	34 mil. posts by 1.3 mil. people				
Doherty et al. [24]		18 people		✓	✓	
Saha et al. [83]	✓	300K posts by 39k people		✓		
Zhang et al. [104]	✓	1 mil. chats	✓	✓		
Wang et al. [97]	✓	1.7 mil. chats	✓	✓		
Pérez-Rosas et al. [73]		259 chats				✓
Chen and Xu [14]	✓	>60K posts by 46K people		✓		

counselor training... micro-skills are transtheoretical [and] are the ingredients of the therapeutic process" [100]. Some of these skills include reflecting on clients' feelings, asking open-ended questions, and providing empathy [39, 72]. Additionally, regardless of specific skills used, simply just having a therapeutic space to express oneself may be beneficial on its own [45, 71].

In online peer counseling, past work has often examined utterance-level skills of support providers [2, 15, 73, 103]. For example, high-quality interactions involve counselors discussing support-seekers' motivations and encouraging them, while lower-quality counseling involves expressing higher persuasion and uncertainty [73]. Peer counselors have also been found to have more success when expressing fewer negative emotions and displaying more positive sentiment [73, 83]. Our work also contributes to growing work trying to understand how mental health support may differ in the online context compared to the traditional offline context; for example, professional therapists may not be expected to discuss their own issues in traditional therapy sessions, but greater self-disclosure by online peers has been shown to help support-seekers feel more comfortable [29, 50, 88, 93, 98].

Below we briefly review frameworks used in our study: (1) empathy, which includes emotional expression, interpretation, and exploration, and (2) motivational interviewing (MI) – a set of validated client-centric counseling techniques used by therapists to help clients make lifestyle changes or work towards recovery [56, 58, 63]. We build on work by Sharma et al. that outlines and classifies three forms of empathy – emotional expression, interpretation, and exploration [87]. In addition, we study how online support providers use the behaviors that form the basis of motivational interviewing, as defined by the Motivational Interviewing Treatment Integrity (MITI) coding manual Version 4, which includes various effective, behavior-based techniques for talk therapy including affirmation, open questions, reflections, and informational support [58, 85]. Some techniques such as affirmation or reflective listening have been found to be consistent with typical therapeutic goals, while other therapist behaviors like confrontation are unlikely to lead to positive mental health changes in clients [3, 59].

2.3 Measuring Effectiveness of Online Support Communities

Over the past few decades, there has been ample research trying to understand the impact and effectiveness of support in online communities. There is a growing perspective over the past few years in HCI advocating for the inclusion of holistic models of well-being, such as the biopsychosocial model, which can take into account biological, psychological, and social factors that contribute to a person in order to identify better support methods and/or understand a person's experience better [60, 94]. While our work contributes a medical perspective of depression and anxiety, we also fully acknowledge that quantitative measurements of mental health symptoms is simply one of many ways for considering one's health, in addition to other holistic factors.

In terms of evaluating people's well-being from a quantitative perspective, previous work has used a variety of proxies. For example, past work has used client feedback immediately following a counseling session [29, 97, 104] and their continued engagement on the platform [87] as indications of high-quality peer support. Wang et al. found that some linguistic predictors are positively related to people giving higher evaluation scores to peer counselors but negatively related to their retention on the platform. Unfortunately, key limitations exist with using these outcome constructs as they have been shown to be unclear in value and meaning in the online mental health context [17, 97] and likely measure fundamentally different aspects of well-being [97]. Retention to the platform, for example, has been a notable focus of past research to measure the success of online health communities [14, 98, 101]; however, retention can be complex in interpretation in this context as Massimi et al. found that people may opt not to return to online mental health platforms due to their positive improvement offline [61]. Additionally, many of these past measures have been limited to the immediate effects of counseling sessions rather than observing longer-term improvements.

Other evaluation metrics have relied on human assessment, such as defining high-quality counseling sessions based on literature and coding these sessions with these predefined criteria [40, 73]. For instance, because reflective listening is a valuable technique in conventional offline therapy, researchers have measured the extent to which reflective listening is occurring online as a metric for the quality of online counseling sessions [63]. Despite there being relatively little prior work using clinical outcomes to measure improvement from digital mental health support [15, 24], the gold standard for clinical research in mental health is the use of self-report questionnaires measuring mental health symptoms (e.g. PHQ-9 for depression, GAD-7 for anxiety) or mental health evaluations by trained professionals. While PHQ-9 and GAD-7 are widely adopted and verified assessments in mental health research and practice, low data availability and response rate often make it impossible to study at scale. Our work extends the past literature as shown in Table 1, and analyzes clinical outcomes from chat-based volunteer support in a large-scale online peer support community using short versions of the PHQ and GAD questionnaires for depression and anxiety symptom assessment, respectively.

3 DATA

Our study's research site is one of the largest online peer support platforms, which we will refer to anonymously as the "online support platform". The online support platform is a psychological support service where individuals ("support-seekers") can engage in anonymous text-based chats with volunteer support providers on a variety of mental health problems. We will use the terms "support-seekers" and "support providers" for the remainder of this paper. Our dataset spans January 2020 to May 2022 and contains over 8 million chats, involving over 1.5 million support-seekers, and 288 thousand support providers. A typical (median) chat lasts for 23 minutes and, over the span of our two-year dataset, support-seekers chatted with 12 distinct providers on average. Our data contains all chat messages in a chat session, including both support-seeker and support-provider messages.

Privacy, Ethics, and Disclosure. This paper used anonymous behavioral log data obtained through a collaboration with an online support platform to conduct our analysis. All data was anonymized throughout the entire research process including analysis and no personally identifiable information was used in this study. No authors on this paper are affiliated with the support platform nor did this research receive any funding from the platform. Our study does not introduce interventions or alter people's experiences on the platform in any way given the potential dangers and ethical concerns in doing so, and we have chosen instead to only use people's prior log data to understand how they currently experience the platform. We also find it worth noting ethical concerns in a popular thread within online mental health research generally centering around predicting mental health status from online community behavior [13]. For example, prior work has often used vague definitions of mental health (such as generalized terms rather than medical definitions of "anxiety" or "depression") and lack theoretical or clinically rigorous grounding in measuring mental health disorders [13]. Although mental health can be measured through several different manners, our study hopes to address these rightful concerns through referencing only the clinical definitions of depression and anxiety (as opposed to colloquial or general reference) as well as using the established assessments for these conditions in medical and psychiatry contexts (i.e. PHQ and GAD assessments). Generally, we hope our study contributes to a growing perspective in the field for understanding people's support-seeking and support-providing behaviors, and based on people's self-reported data rather than predicting or labeling individual's statuses or diagnoses. We also take note from [12] about dehumanizing terminology used in research studies about mental health. As a result, in this paper we avoid using the terms "users", "subjects", "participants", or "accounts" when referring to the people who come to our study's online platform for mental health support; instead, we have chosen the term "support-seeker" (and, at times, "people with depression/anxiety" and "individuals"). Additionally, our team does not name nor consider support-seekers on our paper's research site as "patients" given our study does not include any professional mental health help and we characterize the relationship between support-seekers and support-providers on our study's platform as closer to a peer relationship.

3.1 Counselor Training

Support providers are required to complete a roughly one-hour, psychology-based training that is based on active listening and MI skills. Support-provider training includes learning skills such as asking guiding questions, reflecting concerns back to support-seekers, and showing empathy. After this initial training, support providers must pass an exam on the site to begin chatting with support-seekers. Support providers can also receive awards on their profiles from completing additional training modules such as specialized courses for specific conditions (e.g. ADHD, Depression) as well as advanced general skill courses (e.g. “Active Listening”, “Managing Emotions”).

3.2 Clinical Assessments

The online support platform encourages support-seekers to take free clinical mental health exams. Support-seekers can complete PHQ-9 and PHQ-2 to measure their depression and GAD-7 and GAD-2 to measure anxiety. These assessment instruments are routinely used in both the research and medical setting for diagnostic and severity measurement [4, 8, 55, 82, 96]. PHQ-2 includes the first two questions of PHQ-9, which is the 9-item questionnaire measuring depression presence and severity. Similarly, GAD-2 includes the first two questions of GAD-7, the 7-item questionnaire measuring anxiety severity. Support-seekers can take the PHQ and GAD exams every two weeks to measure any changes in their depression and anxiety symptoms. Our study's analysis presented later in this study will use support-seekers' repeated and longitudinal questionnaire results to measure their changes in depression and anxiety over time.

Although the 2-item versions for PHQ and GAD are not as accurate as the longer questionnaires, PHQ-2 and GAD-2 have been shown in prior work to maintain relatively high accuracy including both considerable sensitivity and specificity for measuring depression and anxiety, respectively [4, 53, 75]. Sensitivity and specificity for PHQ-2 are 86% and 78%, respectively [4]; GAD-2 has a sensitivity of 76% and specificity of 81% [75]. A complete list of items in PHQ-2 and GAD-2 can be found in the appendix. Because these scales ask about the frequency of symptoms of depression and anxiety in the past two weeks, a *higher* score on PHQ and GAD indicates *worse* mental health, while a lower score indicates better mental health. For both PHQ-2 and GAD-2, scores range from 0 (better mental health) to 6 points (worse mental health).

Our study's dataset includes 343,599 support-seekers' responses for PHQ-2 and 186,446 for GAD-2 between January 2020 to May 2022. Because we are interested in changes in mental health, we focus on the support-seekers who completed the questionnaires at least twice – 74,219 support-seekers for PHQ-2 and 42,296 support-seekers for GAD-2. These support-seekers show evidence of clinically significant symptoms of depression and anxiety in over 60% of of the PHQ and GAD they completed (see Table 2). Descriptive statistics for the dataset are included in Table 2 while the detailed distribution of the number of responses per support seeker is shown in Figure 1.

4 RQ1. DOES PARTICIPATING IN PEER-SUPPORT CHATS IMPROVE SUPPORT-SEEKERS' DEPRESSION AND ANXIETY?

We first present our methods and results to answer RQ1.

Because PHQ-2 and GAD-2 assessments were administered to all support-seekers independently of whether they had engaged in peer support chats with support providers, it is possible to differentiate between support-seekers who engaged in peer support chats versus those who did not. To answer RQ1, we compare changes in PHQ-2 and GAD-2 assessments between these two groups through propensity score matching (Section 4.1.1) and conduct analysis on the difference in their mental health assessments using regression analysis (Section 4.1.2).

4.1 Methods

Due to ethical concerns, we did not pursue a randomized controlled trial or other interventional study (e.g., randomly withholding treatments) to investigate our research questions. Instead, we used a pre-post observational study design to infer the treatment effects of engaging in one or more

Table 2. Descriptive statistics of PHQ-2 and GAD-2 questionnaires, as well as the chat session and messaging details in each observation period ("Obs. period") as described in Section 4.1 Methods. Note a "chat session" refers to back-and-forth messages exchanged without a break of more than 24 hours. The length of a chat session is roughly measured by the duration between the first and last messages exchanged.

	PHQ-2	**GAD-2**
# of users, taken at least 2x	74,219	42,296
Mean score (out of 6)	3.55	3.58
% of users with significant symptoms (i.e. score of ≥ 3)	63%	61%

	Obs. period, between PHQ assessments	**Obs. period, between GAD assessments**
Median # of chat sessions	2	3.20
Median # of messages, in total chat sessions	286	314
Median duration of total chat sessions, in days	12.33	18

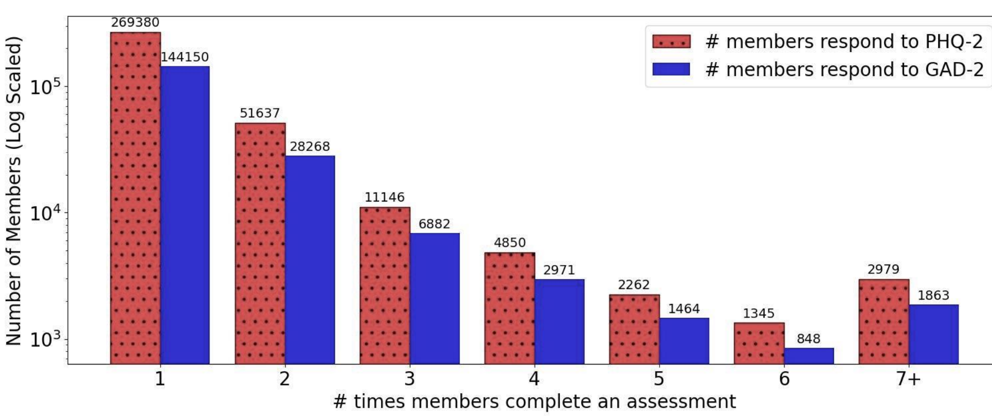

Fig. 1. Distribution of the number of times support-seekers completed the PHQ-2 and GAD-2 clinical assessments

chats on support-seekers' mental health. Specifically, we adopted an "observation-level" random-effects regression with propensity score matching to predict support-seekers' post-assessment mental health scores from their pre-assessment scores, depending on whether they chatted with a support-provider during the observation period and other independent variables. Detailed statistics of chat sessions during observation periods are included in Table 2.

In order to monitor mental health changes, we first selected an "observation period" as our unit of analysis. This period is defined by a support seeker completing two consecutive PHQ-2 or GAD-2 assessments. We refer to the first questionnaire in this sequence as the "pre-assessment", and the following one as the "post-assessment". These observation periods enable repeated observations for the same individuals at different time points; thus, this panel data allows us to use multilevel models like random effects to account for potential omitted-variable bias.

4.1.1 Mitigating Selection Bias via Propensity Score Matching. To account for individual differences between people, such as their prior experience on the site, their initial mental health status, or their age, that may bias whether support-seekers engage in support chats, we employed Propensity Score Matching (PSM), a quasi-experimental method that emulates the conditions of Randomized Controlled Trials (RCTs) prior to regression analysis. PSM is widely recognized for its robust performance in reducing the effects of confounding in observational studies [5] and has been successfully applied in OMHC scenarios similar to our own [14].

In our study, each observation is categorized into the treatment group if the support-seeker engaged in at least one chat during the observation period. Although another analysis method would include using the total number of chats during the observation period rather than splitting support-seekers into just two groups ("had a chat" versus "did not have a chat"), data limitations limit us from splitting "treated" support-seekers into many different (small) groups by their sum number of chats and matching based on this number. However, this method is worth exploring in future work. The aim of PSM is to match each observation in the treatment group with an observation in the control group (i.e. support-seekers who did not participate in chats) that is highly similar on confounding characteristics such as demographics and mental health status. To achieve this, PSM first estimates a propensity score for every observation – the probability of being in the treatment group conditional on certain covariates – and then pairs observations from the treatment and control groups who have similar propensity scores. This process results in treated and control pairs with similar distribution in terms of the covariates [5] but differing on whether they received “treatment” or not (see Section 2.2). Observations whose propensity scores differ by at most a pre-specified amount (the caliper width) are excluded.

Step 1. Propensity Score Estimation.

We adopted a logistic regression model to estimate propensity scores, which predict the probability a support-seeker engages in chats in a given observation period based on the covariates described below.

Although demographic factors of help-seekers such as age, gender, and ethnicity are frequently considered in past studies [26, 31, 33, 38, 52, 80], the OMHC which is our research site collects limited background information about support-seekers and support-providers.

- **Support-seekers’ age**: The mean age of support-seekers is 24.52 (±10.61), reflecting how people using the peer support platform are generally young. Age is the only demographic variable we can reliably measure because it is a required field seekers must provide when signing up for the online support platform.

Another key factor is help-seekers’ prior experience on the platform [31]. We operationalize this through three metrics:

- **Experience in chats**: the total number of chats support-seekers have participated in from the time of their registration on the platform until the start of the observation period.
- **Tenure**: the number of days since a support-seeker registered on the platform
- **Active days**: the total number of days a support-seeker has engaged in any activity on the platform.

Lastly, we consider:

- **Support-seekers’ initial mental health status**: measured by the support-seeker’s PHQ-2 or GAD-2 pre-assessment score.

Table 3. Covariate balance for support-seekers before and after 1:1 caliper matching without replacement based on logit propensity scores. Descriptive statistics and balance measures are presented for treatment and control groups, with conditions denoted as 'U' for unmatched and 'M' for matched.

		PHQ-2				GAD-2			
		Mean ± SD			%reduct	Mean ± SD			%reduct
Covariate		Treated	Control	%bias	|bias|	Treated	Control	%bias	|bias|
Age	U	23.39 (±9.31)	25.33 (±11.38)	-16.5	88.1	23.44 (±9.50)	25.31 (±11.41)	-16	90.6
	M	23.52 (±9.57)	23.42 (±9.79)	2		23.56 (±9.85)	23.51 (±10.05)	1.5	
Initial MHS	U	3.54 (±1.91)	3.54 (±1.87)	0	-2452.1	3.75 (±1.94)	3.68 (±1.92)	3.4	95.0
	M	3.58 (±1.87)	3.55 (±1.86)	1.2		3.77 (±1.90)	3.76 (±1.90)	-0.2	
Experience	U	13.30 (±51.06)	2.62 (±14.35)	67.7	96.5	13.87 (±50.39)	2.71 (±13.50)	72.6	97.1
	M	3.46 (±14.01)	4.41 (±19.12)	-2.3		3.45 (±12.41)	4.34 (±17.52)	-2.1	
Tenure	U	325.97 (±555.04)	317.92 (±579.10)	12.5	97.6	373.03 (±579.95)	364.49 (±609.93)	11.8	85.4
	M	285.93 (±535.96)	277.18 (±519.64)	0.3		320.32 (±552.83)	310.39 (±537.71)	1.7	
Active Days	U	19.88 (±52.55)	22.60 (±76.97)	18.4	96.8	21.16 (±51.26)	28.32 (±85.69)	12	93.1
	M	10.29 (±33.47)	13.34 (±45.41)	-0.6		11.26 (±32.69)	14.31 (±44.74)	0.8	

Step 2. Matching and Balance Estimation.

We employed a one-on-one matching strategy without replacement, based on the logit of the estimated propensity score. We paired each observation in the treatment group with a counterpart in the control group that had the closest logit within the defined caliper. The caliper was set at 0.2 of the standard deviation of the propensity score's logit, following Austin's recommendations [6] – specifically, 0.205 for PHQ outcomes and 0.226 for GAD outcomes.

Following matching, we identified 43,249 matched pairs for the PHQ analysis, and excluded 12,854 people in the treatment group who did not have an adequate match. The GAD analysis includes 26,224 matched pairs, excluding 9,747 people in the treatment group who did not have an adequate match. Table 5.3 details the balance evaluation after matching. Most of the variables show an 80-90% reduction in bias following matching, and all the covariates exhibit a bias of less than five, the threshold recommended by Austin [6]. Although there was no reduction in bias for pre-assessment mental health score, this is because the treated and control group did not differ in their pre-assessment mental health scores, as is shown by Figure 2 ("Pre-assessment PHQ-2/GAD-2").

4.1.2 Regression Analysis. After the matching process, we conducted a random-effects, multi-level regression with observations nested within support-seekers . Because approximately half of the sample had only a single observation, fixed effects regressions would not be appropriate. This regression predicts support-seekers post-assessment mental health (i.e., PHQ and GAD scores) from their pre-assessment mental health, dependent on whether they had at least one support chat during the observation period (treatment group) or not (control groups). Including the support-seeker's **pre-assessment mental health scores** (i.e., PHQ-2 or GAD-2) in the regression models effectively examines how the independent variables predict *changes* in mental health [16]. To test whether participating in a support-chat has stronger effects for those with worse mental health, we include the interaction between participating in a chat with pre-assessment mental health. We also include as a control variable the number of support-chats the seeker had before the observation.

4.2 Results

Table 4 shows the results for RQ1, the effects of chats on changes in mental health. As a reminder, because the PHQ-2 and GAD-2 scales ask about the frequency of symptoms, positive coefficients in the regression analyses indicate that a variable was associated with an increase in symptoms (ie.,

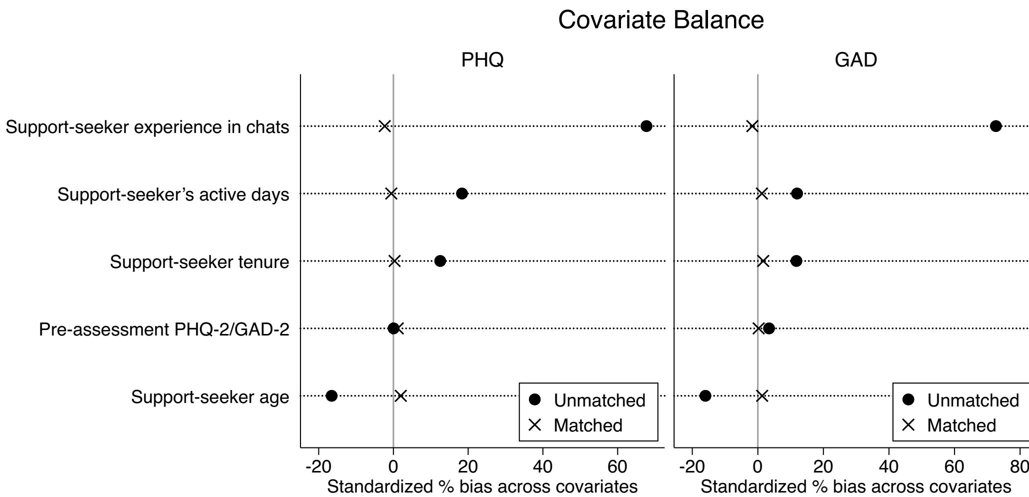

Fig. 2. Love plot showing covariate balance for PHQ (left) and GAD (right) before ("Unmatched") and after ("Matched") matching process.

Table 4. Random-effects regression predicting support-seekers' post-assessment PHQ-2 or GAD-2 scores, comparing matched seekers who had or did not have a support chat during the observation period, controlling for their pre-assessment mental health and the number of chats they had prior to the observation period and the pre-assessment mental health X number of prior chat interactions.

	PHQ-2		**GAD-2**	
	Coeff.	Robust SE	Coeff.	Robust SE
Support-seeker experience in chats (log)	-0.052***	0.006	-0.065***	0.008
Pre-assessment mental health score (MHS)	0.668***	0.004	0.683***	0.005
Treatment: Participated in $\geq$ 1 chats during observation period	-0.126***	0.01	-0.036**	0.013
Treatment × MHS	-0.155***	0.006	-0.145***	0.007
Constant	3.620***	0.007	3.739***	0.01
N_respondents	53915		32222	
N_obs	86498		52448	

$*p < 0.05; **p < 0.01; ***p < 0.001$

worse mental health). In terms of control variables, as expected, both depression (β =.67, p<0.001) and anxiety (β =.68, p<0.001) were very consistent over the several weeks between pre- and post-assessments. In addition, support-seekers who had more on-platform conversations prior to the current observation period had lower post-assessment depression (β =-.05, p<0.001) and anxiety (β =-.07, p<0.001) scores.

Most importantly, the results show that the treatment – participation in at least one support-chat during the observation period – was associated with small but highly reliable improvements in mental health. Participation in support-chats was associated with improvements in symptoms of both depression (β =-0.126, p < 0.001) anxiety, although to a lesser degree (β = -0.036, p < 0.01). In addition, the Treatment × pre-assessment mental health interactions show the benefits of participating in the support chats were greater for those who initially had worse mental health for both depression ($\beta = -.155, p < .001$) and anxiety ($\beta - .145, p < .001$). As illustrated in Figure 3, the dotted blue lines consistently fall beneath the solid red ones. This indicates that support-seekers who participate in chats exhibit fewer symptoms of depression (PHQ-2) and anxiety (GAD-2) (indicative of improved mental health) than those who do not participate. In addition, the $TreatmentXMHS$ interactions were significant and negative for both the PHQ-2 scale (β = -0.155, p < 0.001) and the GAD-2 scale (β = -0.155, p < 0.001), indicating that participating in a chat was associated with improved mental health more for those who initially had more depression and anxiety symptoms respectively. These interactions are illustrated in Figure 3, showing larger differences between those participating in chats versus not when the pre-assessment measure of mental symptoms was high (i.e. 1) rather than low (i.e. 0). That is, people with above-average (or major) depression or anxiety seemed to derive more benefit from engaging in conversations with volunteers compared to those with average or mild symptoms.

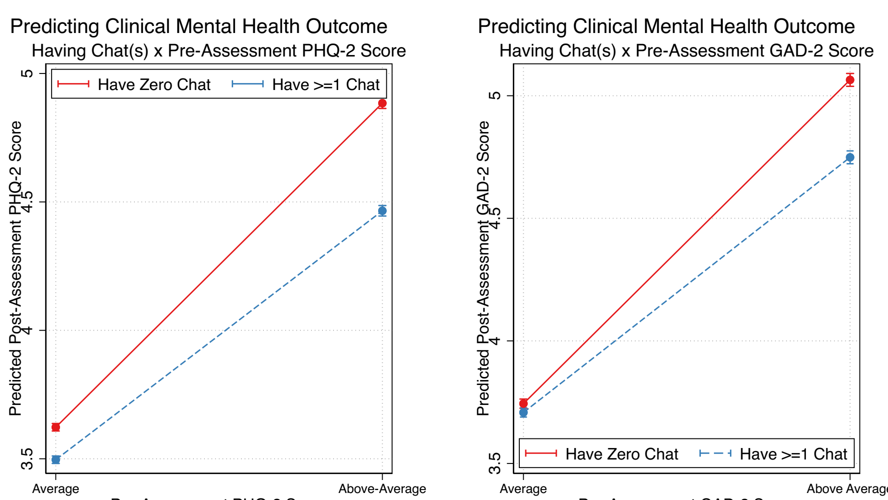

Fig. 3. Marginal plot that illustrates the interaction effect between treatment and pre-assessment score, for support-seekers who have average vs. above-average initial symptom severity. Note that a higher score indicates *worse* symptoms (i.e. greater symptom severity).

5 RQ2. WHAT CHARACTERISTICS OF THE SUPPORT CONVERSATIONS AFFECT CLINICAL MENTAL HEALTH OUTCOMES?

5.1 Methods

To answer RQ2, we also adopted a regression analysis to study the association between therapeutic techniques that occurred during support conversations and changes in support-seekers' mental health. This research involved three steps.

(1) Firstly, we collected consecutive PHQ-2 and GAD-2 assessments from support-seekers and arranged them in pairs, creating an observational period.
(2) For support-seekers who had a support chat during the observational period, we measured factors that the prior literature suggests should be associated with changes in support-seekers' mental health. These include characteristics of the support providers and support-seekers before the observation period started and characteristics of the support conversations that occurred during the observations period.
(3) We used random-effects regression models to predict how the providers', seekers', and chat characteristics predict changes in mental health assessments (i.e., post-assessment controlling for pre-assessment).

For each observation period in which a support-seeker has at least one conversation, we used seven independent variables as listed below in 5.1.1 and 5.1.2.

5.1.1 Support-provider characteristics. We measured five characteristics of the support providers with whom the support-seeker chats during the observation period:

- **Chats with repeat support providers**: capturing the percentage of chats the support-seeker had with a support-provider they previously interacted with before the time of pre-assessment. This tracks the relationship between support-seekers and support providers, a crucial factor in counseling processes as suggested by previous studies [1, 95].
- **Support-providers' age**: a continuous variable representing the average age of the support providers interacted with by the support-seeker.
- **Support-providers' experience in chats**: a variable measuring the average total number of chats these support providers have engaged in from their registration until the time of pre-assessment.
- **Support-providers' tenure**: time since the support provider registered on the online support platform.
- **Support-providers' training modules**: the average number of awards for support providers for completing additional training modules by the platform.

5.1.2 Support-seeker characteristics. We measured two characteristics of the support-seeker before the start of an observation period.

- **Support-seekers' pre-assessment mental health scores (MHS)** (i.e., PHQ-2 or GAD-2): By including pre-assessment score, this lagged dependent variable model effectively examines how the independent variables predict changes in mental health status [16].
- **Support-seekers' tenure**: As a measure of the support-seekers' experience with the platform, we included time since the support-seeker registered on the online support platform. Because this measure had a long tail, we used the log transformation.

5.2 Chat characteristics.

We concatenated all the communication the support-seeker had with one or more providers during an observation period and measured the following attributes of these conversations:

- **Total conversation turns**: The number of conversational turns during an observation period is a measure of the length of the conversation. We include it to gauge the "dosage effect" of peer-support chats on mental health. Just as the analysis for RQ1 showed that having a chat was more strongly associated with improvements in mental health for those who initially had worse mental health at the beginning of the observation period, we included interaction term between total turns and pre-assessment mental health to test whether longer conversations had stronger effects for those who initally had worse mental health.

- **Sentiment in support-seekers' language**: The positivity or negativity of the support-seekers language during the conversations may be an indicator of their mental health. In addition, previous research suggests the positivity or negativity of support-seekers' language may influence how support providers respond to them. For example, providers may offer more emotional support in response to seekers using emotionally negative language [98]. To measure sentiment, we adopted the off-the-shelf VADER classifier from Hutto et al.'s [42]. VADER is a rule-based model for sentiment analysis that performs exceptionally well for social media text, which makes it particularly relevant for our study's case. It outperforms human raters in social media datasets with an F1 score of 0.96. We applied the model to the support-seekers' language during the chats. The VADER classifier outputs an inclusive three-class label (negative, neural, and positive), and we included the negative and positive labels as predictors.
- **Support providers' therapeutic language**: We used natural-language processing techniques (NLP) to measure the support providers' language. We focused on Motivational Interviewing behaviors and empathy, two kinds of therapeutic techniques that have been widely shown to be effective in improving clients' mental health in traditional counseling contexts and are used for training volunteers in online support platforms.
  - **Motivational Interviewing behaviors**: We measured seven Motivational Interviewing behaviors from the MITI 4.2 manual [65], including asking questions, offering affirmations, and reflecting back the seekers' thoughts, using the classifiers developed by Shah et al [85]. The MITI code classifiers are transformer-based models (i.e., BERT) that were pre-trained on 120 million chat messages on the same platform as our study and fine-tuned based on manually annotated datasets. These manually annotated data consists of 14,797 utterances from 734 conversations annotated with 17 Motivational Interviewing behaviors adapted for the online nature of the platform of study by Shah et al [85]. Each model takes in a support-provider's utterance as input, as well as the immediately previous utterance, and outputs a binary label indicating whether the current utterance contains the corresponding MITI code. Descriptions of these measures, including definitions, examples, and classification accuracy are shown in Table 5.
  - **Empathy**: We measured three components of empathy (emotional reaction, exploration, and interpretation) based on classifiers developed by Sharma et al. [87]. Descriptions of these measures, including definitions, examples, and classification accuracy are shown in Table 6. The classifiers include two independently pre-trained RoBERTa-based encoders that accept one turn of clients' and support providers' messages separately and identify empathy shown in the support providers' responses. Following Sharma et al.'s settings, we fine-tuned the models on a public Reddit dataset [88] that contains many mental health communities. As reported in their paper, the models have F1 scores ranging from .626 to .745 in terms of identifying three kinds of empathy on the Reddit dataset.

5.2.1 Regression Analysis. In order to understand how support-provider characteristics, support-seeker characteristics, and characteristics of the conversations predict changes in mental health, we used a similar random-effect model specification as in RQ1 with additional predictor and control variables. However, because the RQ2 analysis contains many more predictor variables and all predictors are continuous rather than binary, we cannot use propensity score matching as a pre-processing step. Results are shown in Table 7. Because some of the measures of MI behaviors are highly correlated with the measures of empathy, resulting in multi-co-linearity, we conducted separate analyses using the MI interviewing behaviors and empathy as predictors. The correlation between MI codes and Empathy codes are shown in Figure 4.

Table 5. Description, examples, and classifier accuracy scores for our study's seven MI (MITI 4.2) codes [65]. Classifiers are from [85].

MI Technique	Description	Example	F1 Score
Giving Information	Support-provider gives general information or feedback in a neutral tone.	*"This website has a lot of materials with coping skills for anxiety if you would like to do more research"*	0.574
Question	Support-provider asks a question that maybe either open (leaving space for support-seekers' response) or closed (implies a short or restricted answer)	*"Hello, what would you like to talk about?"*	0.927
Reflection	Support-provider makes a reflective listening statement, capturing and returning something the support-seeker recently said.	*"It seems like you are having a hard time figuring out the way forward."*	0.692
Affirm	Support-provider accentuates something positive about the support-seeker, such as their strengths, intentions, or behavior	*"It's important to you to be a good parent, just like your folks were for you."*	0.624
Seeking Collaboration	Support-provider attempts to share power or acknowledge the client's expertise (e.g.support-provider seeks consensus with the client regarding tasks)	*"Would it be alright if we spend some time discussing the standards for consuming alcohol during pregnancy."*	0.500
Emphasizing Autonomy	Support-provider focuses on support-seekers'sense of control, freedom, and ability to decide their actions.	*"You're the one who knows yourself best here. What do you think ought to be on this treatment plan?"*	0.785
Persuade with Permission	Support-provider emphasizes collaboration or autonomy support while persuading. Per-mission is present when e.g. client asks for opinion or support-provider asks for permission.	*"I have some information about your risk of problem drinking and I wonder if I can share with you."*	0.770

5.3 Results

Table 7 shows that some characteristics of the support-providers were associated with statistically significant changes in seekers' mental health. We note that our findings generally show small R-squared values, though; we further discuss this in Section 6.1.1. We had expected that chatting with providers with more experience would be associated with improvements in mental health, but this expectations was only partially confirmed. In particular, talking with older support-providers was associated with reductions in depression ($\beta = -.036, p < .001$). This may be because the older volunteers had richer life experiences, which especially benefited the relatively young support-seekers

Table 6. Description, examples, and classifier accuracy scores for our study's three empathy codes [87]. Classifiers are from [87].

Empathy Technique	Description	Example	F1 Score
Emotional Reactions	Support-provider expresses warmth, compassion, and concern.	*"I'm so sorry that you experienced this"*	0.745
Interpretations	Support-provider communicates that they understand the feelings/experiences of the support-seeker, and may attempt to describe or relate.	*"Honestly I've been through the same thing. My parents got divorced when I was 18...I understand why you would be guarded."*	0.626
Exploration	Support-provider attempts to explore, probe, or describe the feelings/experiences not explicitly expressed by the support-seeker.	*"Do you feel like you have a short temper at the moment or that you're more snappy than usual"*	0.726

on this site [29]. However, the association was not significant in predicting anxiety. Surprisingly, support-providers' experience on the site, including past chatting experience and having repeated interactions with the support seeker during the observation period, was not associated with changes in either depression or anxiety ($p>0.05$). In fact, talking with support-providers who completed more training modules on the site was associated with increases in anxiety ($\beta = .037, p < .001$). One possibility is that poorer support-providers are the ones who seek out more training, while another is that the training on the site is counter productive. Given correlational nature of the evidence we cannot determine the causation behind this association. We discuss these findings more in Section 6 (Discussion).

We now turn to the associations of changes in mental health with support-seekers' characteristics. Support-seekers with a unit higher pre-assessment PHQ-2 score had higher post-assessment scores as well ($\beta = 0.446, p < 0.001$). Similarly, those with with a unit higher pre-assessment GAD-2 anxiety score reported higher anxiety during post-assessment (β=0.46, $p<0.001$). The sentiment of support-seekers in their messages predicted changes in their mental health. In particular, support-seekers who expressed more negative sentiment in conversations during an observation period reported worse depression ($\beta = 0.208, p < 0.001$) and anxiety ($\beta = 0.193, p < 0.001$) on their post-assessments. In addition, the associations with positive sentiment followed a similar but weaker pattern. Support-seekers who expressed more positive sentiment in conversations during an observation period reported slightly lower depression ($\beta = -0.024, p < 0.01$), but no change in their levels of anxiety ($\beta = -0.001, p > .05$) on their post-assessments.

We now review the associations of changes in mental health with characteristics of the support conversations. Our analysis for RQ1 demonstrated reliable improvements in depression and anxiety associated with participating in online peer support conversations, with the improvements larger for those who initially had worse mental health. The RQ2 analysis shows analogous results among those who participated in the support-chats. We consider the length of the support chats as a "dosage" of a support conversation; people who participated in longer chats with more turns have slightly improved mental health for depression ($\beta = -.023, p < .05$) but not anxiety ($\beta = .012, p > .05$), with larger effect size for those who initially showed worse symptoms. Having longer chats was more strongly associated with improvements in mental health for those who initially had worse

Table 7. Random effects regression predicting support-seekers' ("seeker") post-assessment PHQ-2 and GAD-2 mental health score, from characteristics of support providers ("provider"), seekers, and chats. *: p<0.05, **: p<0.01, ***: p<0.001

	PHQ-2				GAD-2			
	(1) MITI		(2) Empathy		(3) MITI		(4) Empathy	
Independent Variables	β	Robust SE	β	Robust SE	β	Robust SE	β	Robust SE
Constant	3.526***	0.009	3.526***	0.009	3.737***	0.011	3.737***	0.011
Provider's age	-0.036***	0.009	-0.039***	0.009	-0.019	0.011	-0.016	0.011
Provider's chats (log)	-0.001	0.009	0.003	0.009	-0.012	0.011	-0.011	0.011
Provider's training mods. (log)	0.017	0.009	0.011	0.009	0.037***	0.011	0.030**	0.011
% chats with same supporter	0.015	0.009	0.012	0.009	0	0.011	-0.002	0.011
Seeker's time on platform (log)	-0.005	0.009	-0.007	0.009	-0.01	0.011	-0.01	0.011
Pre-assessment MHS	0.446***	0.005	0.446***	0.005	0.460***	0.007	0.460***	0.007
Total turns (log)	-0.023*	0.009	-0.021*	0.01	0.012	0.012	0.023	0.012
Total turns x MHS	-0.071***	0.005	-0.071***	0.005	-0.056***	0.006	-0.056***	0.006
Seeker's negative sentiment	0.208***	0.009	0.208***	0.009	0.193***	0.012	0.193***	0.011
Seeker's positive sentiment	-0.024**	0.009	-0.028**	0.009	-0.011	0.011	-0.018	0.011
Provider gives information	0.019*	0.008			0.017	0.01		
Provider persuades w/ perm.	0.022**	0.009			0.033**	0.011		
Provider questions	0.011	0.009			0.005	0.012		
Provider reflection	-0.022*	0.009			-0.008	0.011		
Provider affirmation	-0.007	0.008			0.007	0.01		
Provider seek collaboration	0.016*	0.008			0.001	0.011		
Provider emphasize autonomy	-0.020*	0.008			0.005	0.01		
Provider emotional reaction			0.028***	0.009			0.048***	0.011
Provider exploration			0.014	0.009			0.002	0.011
Provider interpretation			0.012	0.009			0.008	0.011
N_respondents	25536		25536		15241		15241	
N_obs	41751		41751		25689		25689	
R-squared overall	0.316		0.316		0.344		0.344	

Fig. 4. Correlation heat map between all MI codes (Giving info, Persuading, Question, Reflection, Affirm, Seeking Collab, Emphasizing Auto) and Empathy codes (EmoReactions, Exploration, Interpretation).

depression ($\beta = -.071, p < .001$) and for those who initially had worse anxiety ($\beta = -.056, p < .001$). These interaction effects are illustrated in Figure 5.

In terms of the content of the support-conversations, our RQ2 results suggest that some MI and empathy behaviors predict improvements in mental health, but others are actually associated with declines in mental health. Of the seven MI behaviors, Reflections and Emphasizing Autonomy were significantly associated with improvements in depression symptoms. Specifically, when people received one standard deviation more Reflection and Emphasizing Autonomy during an observation period their PH-2 depression scores improved by 0.022 ($p < .05$) and 0.02 ($p < .05$) units respectively. In contrast, when they received a standard deviation more Information, Persuasion with Permission, and Seeking Collaboration their PH-2 depression outcomes got worse (p<0.05). However, we note that Seeking Collaboration and Giving Information both have only low to moderate accuracy in our MI classifiers (0.500 and 0.574, respectively), so both of these significant results may need to be interpreted with some caution. In terms of the components of empathy, only Emotional Reaction

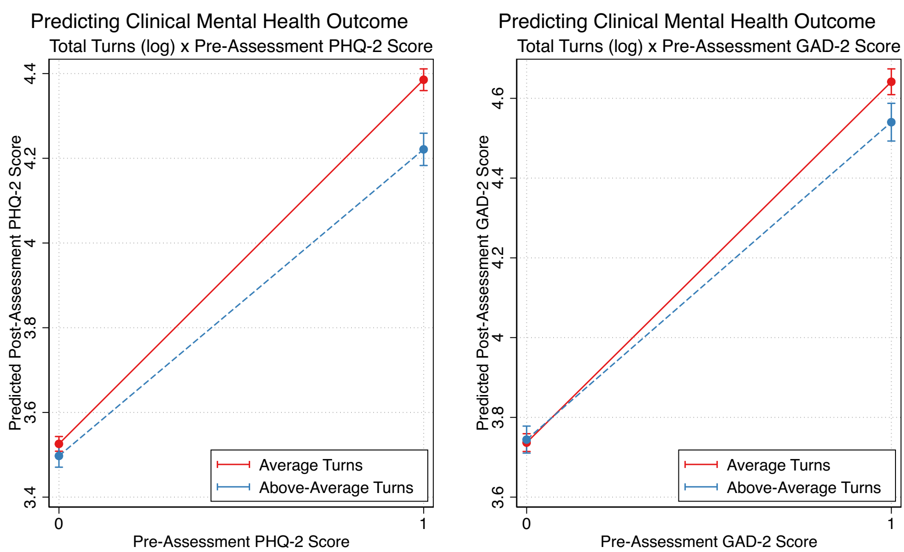

Fig. 5. Marginal plot that illustrates the interaction effect between total conversation turns and support-seekers' pre-assessment scores, comparing effects for those with average versus above-average initial symptom severity.

had a significant association with changes in mental health; Exploration and Interpretation did not. Inconsistent with much research on the common factors that seem responsible for success in conventional psychotherapy [68], receiving Emotional Reactions during the observation period was associated with a small, but statistically reliable increase in depression (0.028 ($p < .001$) points on the PHQ-2 score).

In terms of anxiety symptoms, none of the MITI codes was associated with statistically reliable improvements in anxiety. In contrast, like the results for depression, a standard deviation increase of Persuasion with Permission was associated with worsening of anxiety symptoms (β=0.033, p<0.001). Also like the depression analysis, support-seekers had worse anxiety when provided with more Emotional Reaction. One standard deviation increase in receiving Emotional Reaction during the observation period was associated with a small, but statistically reliable increase of 0.048 ($p < .001$) point increase in their GAD-2 score.

6 DISCUSSION

Our research strongly suggests that online peer support chats are beneficial for improving clinical depression and anxiety outcomes, albeit with a relatively small effect size. Additionally, we identified effective MI skills like Reflection and Emphasizing Autonomy that contribute to this clinical effectiveness. However, several MI techniques including Seeking Collaboration, Persuading (with Permission), Give Information, and the empathy component of Emotional Reaction were associated with worsening of mental health outcomes. Our work has significant implications for understanding how online support chats affect people's mental health symptoms as well as the specific methods contributing to the effectiveness of training provided to online volunteer support providers in many OMHCs.

6.1 Effectiveness of Online Peer Support Chats

Support-seekers who chat on the platform report less frequent symptoms of depression and anxiety compared to the matched seekers who did not engage, with the change highly reliable for both depression and anxiety ($p < .001$). Moreover, support-seekers who initially had more severe depression and anxiety showed greater improvement compared to those with milder initial symptoms. These results are based on a quasi-experimental analysis that matches people participating and not participating in chats on factors likely to be associated with mental health and willingness to engage in support chats. Furthermore, our analysis for RQ2 shows a parallel result, with the dosage of the chats (i.e., more turns) being associated with improved mental health, and especially so among support-seekers' who initially had worse depression and anxiety. Thus, findings for both RQ1 and RQ2 suggest that online support chats are more effective when the support-seeker has worse symptoms. Our study contributes a clinical perspective on existing literature [2, 24, 97] about the general effectiveness of online communities for mental health outcomes. Although the effect sizes are small, these results are consistent with the hypothesis that participating in peer-to-peer support-conversations improves people's mental health.

6.1.1 Low Effect Size of Support-Provider Chats. Although having support conversations seems to improve support-seekers' mental health, we acknowledge that these effects are not powerful. That is, while having at least one chat on the platform during an observation period was significantly associated with a reduction in the frequency of depression and anxiety symptoms, the improvements were small (a 1.6% improvement for depression scores and and a 0.6% improvement for anxiety). There are potentially several reasons for these findings. First, people may not be having enough volume of messaging on the platform to experience significant improvement. Results for RQ2 show a dosage effect, in which improvements in both depression and anxiety were greater with longer conversations. Having one or even several relatively brief conversations over the course of a several-week observation period may not be a strong enough intervention to have large effects on well-being. Another more problematic explanation, however, is that the volunteer counselors may not be trained well enough to provide effective support. In contrast to the months-long training that professional counselors receive, volunteers' roughly one-hour training may not provide the skills needed to be effective support-providers. Finally, on a more methodological note, the 2-item depression and anxiety measures used on the outcome measure in this research may not be sensitive or accurate enough to capture all the change that is occurring.

6.2 Characteristics of conversations

Although the results suggest that peer-to-peer support can improve mental health, our research did not conclusively identify the techniques that the volunteers counselors used that account for this success. Abundant research demonstrates what works in professional counseling. For example, meta-analyses show that counselors who express empathy are more effective than those who do not, with a moderate effect size [68] and that counselors who use appropriate Motivational Interviewing techniques (e.g. Reflection, Emphasizing Autonomy) while refraining from "MI-inconsistent" techniques (e.g. persuading) are more effective [56–58]. Consistent with this prior research, our research suggests that use of the MI-consistent techniques of Reflection and Emphasizing Autonomy was associated with improved depression outcomes. Additionally, the use of the MI-inconsistent techniques like Giving Information and Persuasion was associated with worse depression outcomes. Note, though, that our paper's classification of Persuasion includes the support-seeker giving permission rather than persuasion being unsolicited.

However, other results were inconsistent with the prior research on the effective skills used by professional counselors. For example, one component of empathy – Emotional Reaction – was associated with a significant increase in both the PHQ-2 and GAD-2 scores indicating worse mental health. In addition, use of the MI-consistent technique of Seeking Collaboration was associated with worse depression outcomes. One possible explanation for these findings is that the lightly trained volunteers on our research site don't execute these techniques well enough. Moreover, the machine learning models we used to measure MI behavior did not distinguish between "authentic" versus "superficial" execution of these behaviors. Moreover, it is likely that support-seekers elicit Emotional Reactions or Seeking Collaboration when they describe their mental health problems or express more negative sentiment. A major limitation of our analysis method is that we labeled support-providers' utterances independently of the context of the support chat and support-seekers' messages. However, in reality, it is likely that support-providers' chat behaviors largely depend on the support-seekers' messages that might have elicited them. Thus, it is possible that when support-seekers reveal more mental health problems in the discussion this causes the support-providers' empathy and MI-consistent behavior. Unfortunately, our analysis methods do not allow us to test for possible reverse causation.

In addition, no motivational interviewing techniques were significantly associated with improving anxiety outcome. Moreover, motivational interviewing techniques one might expect to improve mental well-being, like asking open-ended questions or providing affirmations, were not significantly associated with improving depression. One possible explanation for why these techniques don't have the positive effects suggested by the prior literature is that what the support-providers say in the conversation is less important than merely being present and acting as an audience. This explanation is consistent with research showing that personal disclosure and emotional expression in general is healthy for both emotional and physical well-being [45, 71]. As Pennebaker's research suggests [71], it may be that merely engaging in the conversations and having an opportunity to talk about ones problems and express ones emotions is all that is necessary. According to this view, role of the support-provider is to provide an audience to encourage fruitful self-expression rather than to act as an amateur therapist. Additionally, prior work has found that distraction from negative emotions leads to improved mood. This distraction may occur through just casually chatting through online social channels like social media platforms and OMHCs, where people also engage in casual chatter not about any mental health issue [19, 85].

6.3 Designing Training for Online Support Platforms

Our study generally contributes to the design of training for digital support platforms. In particular, our study identified how the therapeutic techniques that guide training in online support platforms, which are widely established in therapy and social work contexts, can significantly help or harm people. We found evidence that several established therapeutic techniques translate to effective clinical outcomes in the online context; for example, our findings suggest that focusing training on client-led techniques like emphasizing a support-seekers' autonomy and reflecting back their own stated feelings has effective outcomes, while behaviors that could be thought of as more support-provider-led (persuading, giving information) are harmful for people's mental health outcomes.

We found insignificant results for some behaviors that would presumably contribute to effective care (e.g. affirmation). As we have mentioned, the ways that our work classifies MI techniques and empathy are context-blind; we only label an utterance as containing a technique or not rather than whether the technique is being used *appropriately*. This is a problem prevalent in this research area generally, where online mental health interventions aim to increase the frequency at which people use support techniques with minimal consideration of *when* or *how* these techniques should be used. For example, Sharma et al's HAILEY system, which helps volunteer supporters write empathetic responses, doesn't focus on when to offer empathy [86]. This focus on quantity rather than appropriateness may contribute both to some support-provider behaviors showing insignificant results and low effect size even for statistically significant findings.

Additionally, some of our findings seem counterintuitive, like providers' emotionally empathetic reactions predicting *more* depression and anxiety or conversations with support-providers with more training modules predicting *more* anxiety. It may be that unmeasured confounds account these surprising results. For example, providers' may offer more empathetic responses when talking to someone describing worse depression or anxiety. Similarly, the support-providers who feel less confident or perform poorly in support chats may take these additional training modules. Further analysis is needed to better understand the conditions under which these skills harm support-seekers and how to change online training programs to account for this. We also note that, as reviewed previously in Section 2.2, skills used in the online context versus traditional in-person care may differ in effect. Given that our existing frameworks such as MI are based in the traditional context, many general concepts of care that show in our analysis to be insignificant (or even harmful) may remain relevant in online support but appear differently when presented in text-based conversations. Alternatively, MI techniques and empathy may have unique behaviors when combined with behaviors that are specific to the online context (i.e. chit-chatting, demographic self-disclosure [29, 85]). There may be context-specific online behaviors which are not covered by existing frameworks; for example, effective techniques like casual chit-chatting that can distract from negative emotions may be helpful for mental health outcomes [19, 85]. We suggest that future work may study techniques within the online context specifically rather than just attempting to emulate face-to-face contexts; existing research suggests that emotions are expressed differently through text-based conversations due to reasons such as lack of audio/visual cues and asynchronous communication, and trusting relationships are also different compared to in-person interactions [7, 27, 79, 92]

In general, our study is most relevant to the growing research interest in developing digital, self-guided, and interactive training to support online health communities [11, 51, 76]. Digital interactive training environment can help to train novice support providers in using the non-specific clinical micro-skills and communication styles that are common across evidence-based therapies, such as Motivational Interviewing and Cognitive Behavioral Therapy (CBT). This type of training could provide scalable training to hundreds of thousands of volunteer support providers on online mental health platforms. Although we focus on mental health volunteers, a similar approach could be used to supplement the more formal education that medical practitioners receive. Although further work is needed to understand the exact skills that should guide online training, our analysis suggests promising effects of online support chats for depression and anxiety. Our analysis also found that (at least a subset of) existing frameworks of motivational interviewing and empathy do translate to beneficial effects for online support-seekers. Our work contributes to studying behavioral interventions and therapy strategies built for the online context, such as iCBT [15] and augments the findings of [85, 97] in that outcome measures for measuring the effectiveness of these frameworks can be highly variable; some therapeutic behaviors may significantly affect proxies like support-seeker satisfaction but do not necessarily translate to clinical symptom improvement.

6.4 Limitations

Our work has the following limitations. First, we studied one peer support platform, and cannot guarantee our findings generalize to all other online platforms. Furthermore, our findings may not be applicable to the OMHCs that do not include peer support chats or volunteer counselors, such as those that connect people to licensed therapists for teletherapy (e.g. BetterHelp.com) or provide support primarily through forum-based help (e.g. mentalhealthforum.net). Second, our analysis may have suffered from some limitations in measuring variables. Our analysis can only detect changes in mental health between any two PHQ-2 and GAD-2 questionnaires. Although our analysis attempted to identify the results of engaging in one or more peer support chats during that time, people can engage in support chats with several different support providers, engage in other platform activities, and more; as a result, it is difficult to isolate and assess the effects of having any given chat. Third, although our study uses longitudinal data on assessment score changes and uses propensity-score matching, the analyses and still correlational and cannot conclusively identify causal directions. Lastly, we acknowledge the limited accuracy on some of our study's classifiers (see Table 5). Although the majority of our therapeutic technique classifiers have high accuracy, the lowest F1 accuracies for MITI classifiers are 50% for Seeking Collaboration and 57% for Giving Information. Inaccuracies in identifying MITI or empathy codes being used could have reduced our analysis' predictive accuracy. Although we saw statistically significant results for both classifiers, as discussed in our results, in general we have analyzed any results from these low accuracy MI code classifiers cautiously.

7 CONCLUSION

We have presented an analysis on the longitudinal and clinical effectiveness of online peer support platforms for depression and anxiety outcomes. Through a pre-post observational study design, we found that having online peer support chats with volunteer counselors improves depression and anxiety symptoms. Additionally, our analysis found that, although online peer support chats are effective, the specific therapeutic behaviors that underlie these platforms, such as support, affirmation, and empathy, do not explain this effectiveness. Our work has important implications for the future of designing online mental health communities and for digital training for online support providers.

ACKNOWLEDGMENTS

We first sincerely thank our colleagues for their wonderful feedback and support, including Milo Fang and colleagues at the Social AI Lab at Carnegie Mellon University. We also express sincere gratitude to our partner platform for their invaluable work in this area, support of mental health research, and collaboration with our research team. This work was supported by the Center for Machine Learning and Health at Carnegie Mellon University and the National Science Foundation (NSF) under Award No. 1952085.

REFERENCES

[1] Steven J Ackerman and Mark J Hilsenroth. 2003. A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clin. Psychol. Rev.* 23, 1 (Feb. 2003), 1–33.

[2] Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Trans Assoc Comput Linguist* 4 (2016), 463–476.

[3] Timothy R Apodaca and Richard Longabaugh. 2009. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. *Addiction* 104, 5 (2009), 705–715.

[4] Bruce Arroll, Felicity Goodyear-Smith, Susan Crengle, Jane Gunn, Ngaire Kerse, Tana Fishman, Karen Falloon, and Simon Hatcher. 2010. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann. Fam. Med.* 8, 4 (2010), 348–353.

[5] Peter C Austin. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* 46, 3 (2011), 399–424.

[6] Peter C Austin. 2011. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics* 10, 2 (2011), 150–161.

[7] Kurt D Baker and Mike Ray. 2011. Online counseling: The good, the bad, and the possibilities. *Couns. Psychol. Q.* 24, 4 (Dec. 2011), 341–346.

[8] Evelyn Behar, Ilyse Dobrow DiMarco, Eric B Hekler, Jan Mohlman, and Alison M Staples. 2009. Current theoretical models of generalized anxiety disorder (GAD): conceptual review and treatment implications. *J. Anxiety Disord.* 23, 8 (Dec. 2009), 1011–1023.

[9] Jacqueline L Bender, Maria-Carolina Jimenez-Marroquin, and Alejandro R Jadad. 2011. Seeking support on facebook: a content analysis of breast cancer groups. *J. Med. Internet Res.* 13, 1 (Feb. 2011), e16.
[10] James F Boswell and Louis G Castonguay. 2007. Psychotherapy training: Suggestions for core ingredients and future research. *Psychotherapy* 44, 4 (Dec. 2007), 378–383.
[11] Anja Busse, Wataru Kashino, Sanita Suhartono, Narendra Narotama, Dicky Pelupessy, Annafi Avicenna Fikri, and Cecilia A Essau. 2021. Acceptability and feasibility of using digital technology to train community practitioners to deliver a family-based intervention for adolescents with drug use disorders during the COVID-19 pandemic. *Addict Behav Rep* 14 (Dec. 2021), 100357.
[12] Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. 2019. Who is the" human" in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.
[13] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 43.
[14] Yixin Chen and Yang Xu. 2021. Exploring the Effect of Social Support and Empathy on User Engagement in Online Mental Health Communities. *Int. J. Environ. Res. Public Health* 18, 13 (June 2021).
[15] Prerna Chikersal, Danielle Belgrave, Gavin Doherty, Angel Enrique, Jorge E Palacios, Derek Richards, and Anja Thieme. 2020. Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16.
[16] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.
[17] Tiago Cunha, David Jurgens, Chenhao Tan, and Daniel Romero. 2019. Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 318–328.
[18] Joe Curran, Glenys D Parry, Gillian E Hardy, Jennifer Darling, Ann-Marie Mason, and Eleni Chambers. 2019. How Does Therapy Harm? A Model of Adverse Process Using Task Analysis in the Meta-Synthesis of Service Users' Experience. *Front. Psychol.* 10 (March 2019), 347.
[19] Anne Dalebroux, Thalia R Goldstein, and Ellen Winner. 2008. Short-term mood repair through art-making: Positive emotion is more effective than venting. *Motiv. Emot.* 32, 4 (Dec. 2008), 288–295.
[20] Saloni Dattani, Hannah Ritchie, and Max Roser. 2021. Mental Health. *Our World in Data* (Aug. 2021).
[21] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. *ICWSM* 8, 1 (May 2014), 71–80.
[22] Munmun De Choudhury and Emre Kıcıman. 2017. The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk. *Proc Int AAAI Conf Weblogs Soc Media* 2017 (May 2017), 32–41.
[23] Koen Demyttenaere, Ronny Bruffaerts, Jose Posada-Villa, Isabelle Gasquet, Viviane Kovess, Jean Pierre Lepine, Matthias C Angermeyer, Sebastian Bernert, Giovanni de Girolamo, Pierluigi Morosini, Gabriella Polidori, Takehiko Kikkawa, Norito Kawakami, Yutaka Ono, Tadashi Takeshima, Hidenori Uda, Elie G Karam, John A Fayyad, Aimee N Karam, Zeina N Mneimneh, Maria Elena Medina-Mora, Guilherme Borges, Carmen Lara, Ron de Graaf, Johan Ormel, Oye Gureje, Yucun Shen, Yueqin Huang, Mingyuan Zhang, Jordi Alonso, Josep Maria Haro, Gemma Vilagut, Evelyn J Bromet, Semyon Gluzman, Charles Webb, Ronald C Kessler, Kathleen R Merikangas, James C Anthony, Michael R Von Korff, Philip S Wang, Traolach S Brugha, Sergio Aguilar-Gaxiola, Sing Lee, Steven Heeringa, Beth-Ellen Pennell, Alan M Zaslavsky, T Bedirhan Ustun, Somnath Chatterji, and WHO World Mental Health Survey Consortium. 2004. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA* 291, 21 (June 2004), 2581–2590.
[24] Gavin Doherty, David Coyle, and John Sharry. 2012. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1421–1430.
[25] Mitchell Dowling and Debra Rickwood. 2013. Online Counseling and Therapy for Mental Health Problems: A Systematic Review of Individual Synchronous Interventions Using Chat. *J. Technol. Hum. Serv.* 31, 1 (Jan. 2013), 1–21.
[26] Nicki A Dowling, Simone N Rodda, Dan I Lubman, and Alun C Jackson. 2014. The impacts of problem gambling on concerned significant others accessing web-based counselling. *Addictive behaviors* 39, 8 (2014), 1253–1257.
[27] Russell K Elleven and Jeff Allen. 2004. Applying technology to online counseling: Suggestions for the beginning e-therapist. *Journal of Instructional Psychology* 31 (2004), 223–226.
[28] Hanmei Fan, Reeva Lederman, Stephen P Smith, and Shanton Chang. 2014. How Trust Is Formed in Online Health Communities: A Process Perspective. *Communications of the Association for Information Systems* 34, 1 (2014), 28.
[29] Anna Fang and Haiyi Zhu. 2022. Matching for Peer Support: Exploring Algorithmic Matching for Online Mental Health Communities. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–37.

- [30] Ruben Fukkink. 2011. Peer counseling in an online chat service: a content analysis of social support. *Cyberpsychol. Behav. Soc. Netw.* 14, 4 (April 2011), 247–251.
- [31] Ruben G Fukkink and Jo MA Hermanns. 2009. Children's experiences with chat support and telephone support. *Journal of Child Psychology and Psychiatry* 50, 6 (2009), 759–766.
- [32] Vasudha Gidugu, E Sally Rogers, Steven Harrington, Mihoko Maru, Gene Johnson, Julie Cohee, and Jennifer Hinkel. 2015. Individual Peer Support: A Qualitative Study of Mechanisms of Its Effectiveness. *Community Ment. Health J.* 51, 4 (May 2015), 445–452.
- [33] Margaret M Giorgio, Leslie M Kantor, Deborah S Levine, and Whitney Arons. 2013. Using chat and text technologies to answer sexual and reproductive health questions: Planned Parenthood pilot study. *Journal of Medical Internet Research* 15, 9 (2013), e203.
- [34] Madelyn S Gould, Anthony Pisani, Carlos Gallo, Ashkan Ertefaie, Donald Harrington, Caroline Kelberman, and Shannon Green. 2022. Crisis text-line interventions: Evaluation of texters' perceptions of effectiveness. *Suicide Life Threat. Behav.* 52, 3 (June 2022), 583–595.
- [35] Jeremy A Greene, Niteesh K Choudhry, Elaine Kilabuk, and William H Shrank. 2011. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *J. Gen. Intern. Med.* 26, 3 (March 2011), 287–292.
- [36] Xinning Gui, Yu Chen, Yubo Kou, Katie Pine, and Yunan Chen. 2017. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 1–19.
- [37] Keith N Hampton, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. Social networking sites and our lives'. Pew Internet Research Centre. June 16 2011.
- [38] Dilys Haner and Debra Pepler. 2016. "Live Chat" clients at kids help phone: Individual characteristics and problem topics. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 25, 3 (2016), 138.
- [39] Clara E Hill. 2020. Helping skills: Facilitating exploration, insight, and action, 5th ed. 5 (2020), 485.
- [40] Xiaolei Huang, Xin Li, Lei Zhang, Tianli Liu, David Chiu, and Tingshao Zhu. [n. d.]. Topic model for identifying suicidal ideation in Chinese microblog. https://aclanthology.org/Y15-1064.pdf. Accessed: 2023-5-15.
- [41] Jina Huh. 2015. Clinical Questions in Online Health Communities: The Case of "See your doctor" Threads. *CSCW Conf Comput Support Coop Work* 2015 (2015), 1488–1499.
- [42] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.
- [43] Bhautesh Dinesh Jani, David N Blane, and Stewart W Mercer. 2012. The role of empathy in therapy and the physician-patient relationship. *Forsch. Komplementmed.* 19, 5 (Oct. 2012), 252–257.
- [44] I Kawachi and L F Berkman. 2001. Social ties and mental health. *J. Urban Health* 78, 3 (Sept. 2001), 458–467.
- [45] Eileen Kennedy-Moore and Jeanne C Watson. 2001. How and When Does Emotional Expression Help? *Rev. Gen. Psychol.* 5, 3 (Sept. 2001), 187–212.
- [46] Ronald C Kessler, Patricia Berglund, Olga Demler, Robert Jin, Kathleen R Merikangas, and Ellen E Walters. 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* 62, 6 (June 2005), 593–602.
- [47] Heejung S Kim, David K Sherman, and Shelley E Taylor. 2008. Culture and social support. *Am. Psychol.* 63, 6 (Sept. 2008), 518–526.
- [48] K Kroenke, R L Spitzer, and J B Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 9 (Sept. 2001), 606–613.
- [49] Kroenke Kurt and Spitzer Robert L. 2002. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr. Ann.* 32, 9 (Sept. 2002), 509–515.
- [50] Taisa Kushner and Amit Sharma. 2020. Bursts of Activity: Temporal Patterns of Help-Seeking and Support in Online Mental Health Forums. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2906–2912.
- [51] Eunjung Lee, Toula Kourgiantakis, and Marion Bogo. 2020. Translating knowledge into practice: using simulation to enhance mental health competence through social work education. *Soc. Work Educ.* 39, 3 (April 2020), 329–349.
- [52] Nicole Levitz, Erica Wood, and Leslie Kantor. 2018. The influence of technology delivery mode on intervention outcomes: Analysis of a theory-based sexual health program. *Journal of Medical Internet Research* 20, 8 (2018), e10398.
- [53] Bernd Löwe, Kurt Kroenke, and Kerstin Gräfe. 2005. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J. Psychosom. Res.* 58, 2 (Feb. 2005), 163–171.
- [54] Bernd Löwe, Kurt Kroenke, Wolfgang Herzog, and Kerstin Gräfe. 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J. Affect. Disord.* 81, 1 (July 2004), 61–66.
- [55] Bernd Löwe, Irini Schenkel, Caroline Carney-Doebbeling, and Claus Göbel. 2006. Responsiveness of the PHQ-9 to Psychopharmacological Depression Treatment. *Psychosomatics* 47, 1 (2006), 62–67.

[56] Brad W Lundahl, Chelsea Kunz, Cynthia Brownell, Derrik Tollefson, and Brian L Burke. 2010. A Meta-Analysis of Motivational Interviewing: Twenty-Five Years of Empirical Studies. *Res. Soc. Work Pract.* 20, 2 (March 2010), 137–160.

[57] Molly Magill, Timothy R Apodaca, Brian Borsari, Jacques Gaume, Ariel Hoadley, Rebecca EF Gordon, J Scott Tonigan, and Theresa Moyers. 2018. A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. *Journal of consulting and clinical psychology* 86, 2 (2018), 140.

[58] Molly Magill, Jacques Gaume, Timothy R Apodaca, Justin Walthers, Nadine R Mastroleo, Brian Borsari, and Richard Longabaugh. 2014. The technical hypothesis of motivational interviewing: a meta-analysis of MI's key causal model. *J. Consult. Clin. Psychol.* 82, 6 (Dec. 2014), 973–983.

[59] Molly Magill, Jacques Gaume, Timothy R Apodaca, Justin Walthers, Nadine R Mastroleo, Brian Borsari, and Richard Longabaugh. 2014. The technical hypothesis of motivational interviewing: A meta-analysis of MI's key causal model. *Journal of consulting and clinical psychology* 82, 6 (2014), 973.

[60] Gabriela Marcu and Jina Huh-Yoo. 2023. Attachment-Informed Design: Digital Interventions That Build Self-Worth, Relationships, and Community in Support of Mental Health. (2023).

[61] Michael Massimi, Jackie L Bender, Holly O Witteman, and Osman H Ahmed. 2014. Life transitions and online health communities: reflecting on adoption, use, and disengagement. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1491–1501.

[62] Margaret McBeath, Drysdale Maureen T B., and Nicholas Bohn. 2017. Work-integrated learning and the importance of peer support and sense of belonging. *Education + Training* 60, 1 (Jan. 2017), 39–53.

[63] William R Miller and Stephen Rollnick. 2012. *Motivational Interviewing: Helping People Change.* Guilford Press.

[64] R Mojtabai, M Olfson, N A Sampson, R Jin, B Druss, P S Wang, K B Wells, H A Pincus, and R C Kessler. 2011. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. *Psychol. Med.* 41, 8 (Aug. 2011), 1751–1761.

[65] Theresa B Moyers, Lauren N Rowell, Jennifer K Manuel, Denise Ernst, and Jon M Houck. 2016. The Motivational Interviewing Treatment Integrity Code (MITI 4): Rationale, Preliminary Reliability and Validity. *J. Subst. Abuse Treat.* 65 (June 2016), 36–42.

[66] L Mufson, H Fitterling, and P Wickramaratne. 2006. National survey of psychotherapy training in psychiatry, psychology, and social work. *Arch. Gen. Psychiatry* (2006).

[67] John C Norcross and Michael J Lambert. 2018. Psychotherapy relationships that work III. *Psychotherapy* 55, 4 (Dec. 2018), 303–315.

[68] John C Norcross and Michael J Lambert. 2019. *Psychotherapy Relationships that Work: Volume 1: Evidence-Based Therapist Contributions.* Oxford University Press.

[69] John C Norcross and Bruce E Wampold. 2011. Evidence-based therapy relationships: research conclusions and clinical practices. *Psychotherapy* 48, 1 (March 2011), 98–102.

[70] Amber Paukert, Brian Stagner, and Kerry Hope. 2004. The Assessment of Active Listening Skills in Helpline Volunteers. *Stress, Trauma, and Crisis* 7, 1 (Jan. 2004), 61–76.

[71] James W Pennebaker. 2012. *Opening Up: The Healing Power of Expressing Emotions.* Guilford Press.

[72] Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, Lawrence An, Kathy J Goggin, and Delwyn Catley. 2017. Predicting Counselor Behaviors in Motivational Interviewing Encounters. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers.* Association for Computational Linguistics, Valencia, Spain, 1128–1137.

[73] Verónica Pérez-Rosas, Xinyi Wu, Kenneth Resnicow, and Rada Mihalcea. 2019. What Makes a Good Counselor? Learning to Distinguish between High-quality and Low-quality Counseling Conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, Florence, Italy, 926–935.

[74] Paul N Pfeiffer, Michele Heisler, John D Piette, Mary A M Rogers, and Marcia Valenstein. 2011. Efficacy of peer support interventions for depression: a meta-analysis. *Gen. Hosp. Psychiatry* 33, 1 (2011), 29–36.

[75] Faye Plummer, Laura Manea, Dominic Trepel, and Dean McMillan. 2016. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *Gen. Hosp. Psychiatry* 39 (March 2016), 24–31.

[76] Pawel Posadzki, Malgorzata M Bala, Bhone Myint Kyaw, Monika Semwal, Ushashree Divakar, Magdalena Koperny, Agnieszka Sliwka, and Josip Car. 2019. Offline Digital Education for Postregistration Health Professions: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J. Med. Internet Res.* 21, 4 (April 2019), e12968.

[77] Julie Prescott, Amy Leigh Rathbone, and Gill Brown. 2020. Online peer to peer support: Qualitative analysis of UK and US open mental health Facebook groups. *Digit Health* 6 (Dec. 2020), 2055207620979209.

[78] Anoushka Rassau and Lucius Arco. 2003. Effects Of Chat-Based On-Line Cognitive Behavior Therapy On Study Related Behavior And Anxiety. *Behav. Cogn. Psychother.* 31, 3 (July 2003), 377–381.

[79] Derek Richards and Noemi Viganó. 2013. Online counseling: a narrative and critical review of the literature. *J. Clin. Psychol.* 69, 9 (Sept. 2013), 994–1011.

[80] Simone Rodda and Dan I Lubman. 2014. Characteristics of gamblers using a national online counselling service for problem gambling. *Journal of Gambling Studies* 30 (2014), 277–289.

[81] E S Rogers, S Harrington, M Maru, G Johnson, and others. 2015. Individual peer support: A qualitative study of mechanisms of its effectiveness. *mental health journal* (2015).

[82] Miguel A Ruiz, Enric Zamorano, Javier García-Campayo, Antonio Pardo, Olga Freire, and Javier Rejas. 2011. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. *J. Affect. Disord.* 128, 3 (Feb. 2011), 277–286.

[83] Koustuv Saha and Amit Sharma. 2020. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. *ICWSM* 14 (May 2020), 590–601.

[84] Stephen M Schueller, Kathryn Noth Tomasino, and David C Mohr. 2017. Integrating human support into behavioral intervention technologies: The efficiency model of support. *Clinical Psychology: Science and Practice* 24, 1 (2017), 27.

[85] Raj Sanjay Shah, Faye Holt, Shirley Anugrah Hayati, Aastha Agarwal, Yi-Chia Wang, Robert E Kraut, and Diyi Yang. 2022. Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. *Proc. ACM Hum. Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–24.

[86] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023), 46–57.

[87] Ashish Sharma, Adam S Miner, David C Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. (Sept. 2020). arXiv:2009.08441 [cs.CL]

[88] Eva Sharma and Munmun De Choudhury. 2018. Mental Health Support and its Relationship to Linguistic Accommodation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 641)*. Association for Computing Machinery, New York, NY, USA, 1–13.

[89] Cecilia H Solano. 1986. People Without Friends: Loneliness and Its Alternatives. In *Friendship and Social Interaction*, Valerian J Derlega and Barbara A Winstead (Eds.). Springer New York, New York, NY, 227–246.

[90] Sarah Elizabeth Gellhaus Thomas, Ronald Jay Werner-Wilson, and Megan J Murphy. 2005. Influence of Therapist and Client Behaviors on Therapy Alliance. *Contemp. Fam. Ther.* 27, 1 (March 2005), 19–35.

[91] R Jay Turner, B Gail Frankel, and Deborah M Levin. 1983. Social support: Conceptualization, measurement, and implications for mental health. *Res. Community Ment. Health* 3 (1983), 67–111.

[92] Anne Vermeulen, Heidi Vandebosch, and Wannes Heirman. 2018. Shall I call, text, post it online or just tell it face-to-face? How and why Flemish adolescents choose to share their emotions on- or offline. *Journal of Children and Media* 12, 1 (Jan. 2018), 81–97.

[93] Tatiana A Vlahovic, Yi-Chia Wang, Robert E Kraut, and John M Levine. 2014. Support matching and satisfaction in an online breast cancer support community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1625–1634.

[94] Derick T Wade and Peter W Halligan. 2017. The biopsychosocial model of illness: a model whose time has come. , 995–1004 pages.

[95] Bruce E Wampold and Christoph Flückiger. 2023. The alliance in mental health care: conceptualization, evidence and clinical applications. *World Psychiatry* 22, 1 (Feb. 2023), 25–41.

[96] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (March 2018), 1–26.

[97] Tony Wang, Haard K Shah, Raj Sanjay Shah, Yi-Chia Wang, Robert E Kraut, and Diyi Yang. 2023. Metrics for Peer Counseling: Triangulating Success Outcomes for Online Therapy Platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23, Article 483)*. Association for Computing Machinery, New York, NY, USA, 1–17.

[98] Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 833–842.

[99] Barry Wellman and University of Toronto. Centre for Urban and Community Studies. 2001. *Does the Internet Increase, Decrease, Or Supplement Social Capital? : Social Networks, Participation, and Community Commitment.* Centre for Urban and Community Studies, University of Toronto.

[100] Raymond Witte and G Susan Mosley-Howard. 2014. *Mental Health Practice in Today's Schools: Issues and Interventions.* Springer Publishing Company.

[101] Diyi Yang, Robert Kraut, and John M Levine. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. *Proc SIGCHI Conf Hum Factor Comput Syst* 2017 (May 2017), 6363–6375.

[102] Zheng Yao, Haiyi Zhu, and Robert E Kraut. 2022. Learning to become a volunteer counselor: Lessons from a peer-to-peer mental health community. *Proc. ACM Hum. Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–24.

[103] Justine Zhang and Cristian Danescu-Niculescu-Mizil. 2020. Balancing Objectives in Counseling Conversations: Advancing Forwards or Looking Backwards. (May 2020). arXiv:2005.04245 [cs.CL]

[104] Justine Zhang, Robert Filbin, Christine Morrison, Jaclyn Weiser, and Cristian Danescu-Niculescu-Mizil. 2019. Finding Your Voice: The Linguistic Development of Mental Health Counselors. (June 2019). arXiv:1906.07194 [cs.CL]

A FULL PHQ-2 AND GAD-2 QUESTIONNAIRES

PHQ-2: Over the last 2 weeks, how often have you been bothered by the following problems?

(1) Little interest or pleasure in doing things
  - Not at all
  - Several days
  - More than half the days
  - Nearly every day

(2) Feeling down, depressed or hopeless
  - Not at all
  - Several days
  - More than half the days
  - Nearly every day

GAD-2: Over the last 2 weeks, how often have you been bothered by the following problems?

(1) Feeling nervous, anxious or on edge
  - Not at all
  - Several days
  - More than half the days
  - Nearly every day

(2) Not being able to stop or control worrying
  - Not at all
  - Several days
  - More than half the days
  - Nearly every day

Received July 2023; revised October 2023; accepted November 2023