Machine and Human Understanding of Empathy in Online Peer Support: A Cognitive Behavioral Approach

Sara Syed*
sara_syed@brown.edu
Brown University
USA

Zainab Iftikhar*
zainab_iftikhar@brown.edu
Brown University
USA

Amy Xiao
amy_xiao@brown.edu
Brown University
USA

Jeff Huang
jeff_huang@brown.edu
Brown University
USA

ABSTRACT

Online peer support provides space for individuals to connect with others and seek support. However, while empathy is critical for effective support, studies have found that highly empathetic support on these platforms can be rare. Using data from online peer support platforms, we conducted a mixed-methods analysis to study the factors that lead to support seekers' perceived empathy. We found that CBT techniques like active listening and reflective restatements, along with fostering a space for exploration, increase perceived empathy, whereas rigid adherence to structure, misalignment of concerns, and lack of emotional validation can contribute to low perceived empathy. In addition, despite the high levels of empathy reported by most support seekers (85%), computational models reported low averaged empathy (1.69/6). Lastly, we propose that empathy is not a quantifiable metric and that future algorithmic empathy measurements require human perspectives.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**; • **Computing methodologies** → **Discourse, dialogue and pragmatics**.

KEYWORDS

mental health, computational psychology, social computing, peer support, cognitive behavioral therapy, deep learning

---

*Both authors share first authorship.

---

1 INTRODUCTION

More than 50% of adults with mental illnesses in the United States do not receive mental health services [1] due to barriers like high treatment costs, stigma, and lack of trained professionals [3]. As a result, alternative cost-effective interventions—like internet-based therapy and peer support platforms—have become ubiquitous and accessible solutions to mental health care. Among these alternatives, Internet-based Cognitive Behavioral Therapy (iCBT) has gained prominence, with studies suggesting that guided iCBT can be as effective as face-to-face CBT [2]. However, iCBT faces some similar challenges as in-person treatment, including limited availability of trained professionals, paving way for peer support platforms to emerge as valuable resources. By removing the need of one-on-one time from a trained professional, peer support platforms have been found to confer therapeutic value [31], minimize wait times, reduce treatment costs [11], and mitigate stigma [32].

While there are various formats for online peer support groups, including face-to-face settings [50, 51], video-conferencing [33, 36], and voice-only [9, 21], there is an increasing adoption of text-based formats [27, 30] since certain attributes of texting can help reduce feelings of social anxiety and inhibition [23]. However, conveying empathy in text-based communication can be challenging due to the absence of tone, irony, body language and other in-person social cues [23, 44]. Previous research has found that online peer support groups can suffer from a lack of empathy [39], a necessary element of effective therapeutic relationships [16, 47] leading to a need in understanding how empathy is communicated and perceived in this format. To address this limitation, one potential solution, adopted by text-based peer support platforms [42, 43], includes training peer support providers in empathetic techniques, like restatements and open-ended questions. Since peer supporters often lack formal training in mental health interventions, providing support in a text-based platform can be challenging since support seekers cannot perceive empathy through the peer supporter's facial expressions, body language, and other visual cues. Hence, a training manual focusing on empathetic techniques could improve therapeutic outcomes as it provides peer support providers with a structured approach to navigate challenging situations by writing empathically [37].

In addition to training, as online text-based mental health peer support communities continue to grow, efforts are made to increase computational methods for assisting peer supporters, such as a

machine in the loop for suggesting [39] and evaluating [38] empathetic responses to support seeker's posts. This could be useful for increasing the efficacy of platforms, as previous work has highlighted that individuals may have difficulty self-assessing their own levels of empathy [7]. However, little work has been done to evaluate and understand if these computational approaches align with users' perceptions of empathy, as well as to understand factors of text-based therapeutic communication reported by support seekers and providers that can cause low and high perceived empathy. Understanding such elements may not only assist the users but also directly benefit the structure and training of these platforms.

In this paper, we explore the presence and expression of empathy in iCBT-based peer support conversations. Through a mixed-methods analysis of computational models, session dialogue and feedback from peer support sessions, we aim to further our understanding of current machine and human approaches to measuring and understanding factors that contribute to support seekers' perceived empathy. We ask the following research questions:

- What are the underlying factors that contribute to support seekers' perceived empathy in CBT-based peer support sessions?
- How is the support seeker experience reflected in a state-of-the-art approach to automatically measuring empathy, and are the above underlying factors captured?

We found that while 85% of support seekers reported high empathy across 100 sessions, computational models reported a low averaged empathy (1.69/6), implying that even when peers adopt empathetic techniques, computational measures may still indicate low empathy scores. Our findings highlight a discrepancy between the human experience and computational interpretation of empathy, suggesting a potential gap in how deep learning models are capturing the complexity of human empathy. While computational approaches may have high accuracy at labeling data according to the exact parameters that they were trained on, they have difficulty interpreting aspects of empathy, such as human connection and context, that don't fall inside their narrow definition. Consequently, this research aims to uncover these contextual and subjective factors that influence human empathy and support seekers' experience.

2 RELATED WORK

2.1 Empathy in Therapeutic Relationships

Empathy in CBT refers to how well the therapist can go into the client's world and see and experience their life [5]. As CBT relies on the examination of thoughts, feelings, and behaviors and their relationship to a person's experiences, empathy can help support providers better understand both the emotional reaction and the meaning of the experiences of a client [46]. Previous work has highlighted the importance of empathy within a mental health support system showing that irrespective of the support method used or the qualifications of a therapist, empathy in a therapist-patient relationship is necessary for effective treatment outcomes [16, 47]. Practitioners expressing empathy has been found to have a beneficial causal effects across a wide variety of fields, including amongst CBT groups for depression and clinical groups for cancer patients. Benefits include improvements in recovery and increased patient satisfaction [24]. More specifically, it is the patient's perception
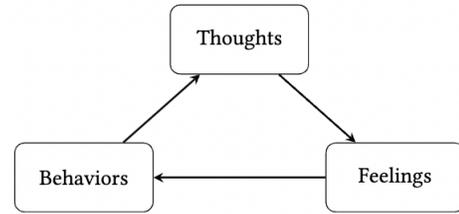

Figure 1: Cognitive Restructuring Framework, illustrating connection between thoughts, feelings, and behaviors

of empathy levels that are most strongly associated with successful outcomes [12], as there is a need for patients to feel that their therapist has empathy for them. However, previous research has shown that therapists may have difficulty evaluating their own levels of empathy in comparison to how they are perceived by their clients [7], highlighting a challenge that therapists may have with self-assessing their conveyed empathy and thus the efficacy of their treatments. As such, automated approaches to measuring patients' perceived levels of empathy may be able to provide therapists with real time feedback on the amount of empathy they are conveying.

2.2 The Role of Empathy in CBT: Enhancing Outcomes through Active Listening

Empathy plays a pivotal role in CBT, serving as a foundation upon which therapeutic alliances and treatment success are built [8]. Several studies have shown that CBT improves individuals' empathy. For instance, Song et al. reported that empathy levels in chronic pain patients increased following CBT which in turn improved interpersonal relationships [41]. Additionally, Gentry et al. discussed the significance of empathy training for efficient leadership [13], while Salem et al. explored the role of empathetic skills' training and their potential to mitigate cyberbullying [37].

Central to CBT is the principle that thoughts, feelings, and behaviors are interconnected, meaning that altering one can lead to changes in the others. This interconnectedness, as illustrated by Figure 1, is particularly evident when considering the practice of cognitive restructuring, which targets unhelpful thought patterns – often termed 'cognitive distortions' – to alleviate psychological distress. Empathy is crucial for the success of cognitive restructuring, and it is effectively practiced through the Active-Empathetic Listening Scale (AELS) [6], which consists of three components: 'sensing'— the careful observation of a client's verbal and non-verbal cues, 'processing'— the thoughtful interpretation of these cues through actions such as note-taking and summarizing, and 'responding'— the clear communication back to the client, reinforcing the connection between thoughts, feelings, and behaviors. This empathetic approach not only deepens therapist-client understanding but also aids in accurately addressing cognitive distortions and guiding clients toward more helpful thought patterns.

2.3 Digital Empathy & Computational Measurements

Recent research on digital empathy—empathy conveyed through digital settings—in mental health and peer support has primarily

focused on video conferencing [14, 18, 20, 28], neglecting to explore how empathy is communicated in text-based interactions. Due to the absence of vocal and physical cues in text-based communication, this omission is particularly significant as it has been estimated that around 90% of face-to-face communication is conveyed non-verbally [10, 17]. Prior studies have also centered on untrained peer support in discussion-board style forums [19, 34, 39], overlooking empathy's expression in one-on-one text-based dialogues. Such one-on-one settings offer private, context-sensitive exchanges crucial for deeply personal communication without fear of group judgment. The scalability of peer support models, compared to the therapist-dependent iCBT, notably enhances CBT accessibility. Training laypeople as peer supporters circumvent the scarcity of trained therapists, addressing capacity challenges inherent in iCBT platforms [29].

Few methods for measuring empathy exclusively from text have been proposed [15, 25, 49]; however, most that do, do not provide publicly available datasets or models, making it difficult to study their capabilities in applied settings. EPITOME [39] is a deep learning model for measuring situational empathy in peer support from the text. To the best of our knowledge, it is the only automated approach to measuring situational text-based empathy that considers messages from both the person receiving and the individual providing support. By considering text from both users, the model better encapsulates the CBT-based definition of empathy of if the supporter "can go into a client's world" by including the context of the information the person seeking help has shared. While other research has employed components of the EPITOME model to measure empathy [26, 45, 52], there has been a lack of research regarding understanding how digital empathy scores relate to the experience of a person seeking support.

3 DATA

3.0.1 Cheeseburger Session. To understand empathy in online peer support conversations, we use data from a CBT-based peer support platform, Cheeseburger Therapy [42]. The platform operates on a 'pay-if-it-helps' model designed to offer users an accessible form of mental health support without the up-front costs often associated with therapeutic services. Users are encouraged to contribute a payment of $25, but only if they feel that the service has been beneficial to them. The fees are directed to the support providers. The website is managed and maintained by a team that includes, a licensed family therapist, designers, software engineers, and individuals trained in CBT.

We chose this dataset because of the nature of the CBT-based interactions and the sessions' emphasis on empathy. The platform allows anyone with an internet-accessible device who is seeking support to sign up for an approximately one-hour session where they communicate with a trained peer through text. Cheeseburger Therapy utilizes CBT-based techniques to provide individuals seeking help (referred to as *thinkers* on the website) to openly express their distress and engage in discussions with trained peers (referred to as *helpers*). During a peer support session, helpers employ therapeutic and empathetic techniques like *active listening*, *open-ended questions*, *reflective restatements*, *cognitive restructuring*, and *thought records* to guide the session. They are taught to inquire thinkers

about something that is troubling them in life and then work with them to identify their cognitive distortion, related feelings, and behaviors. Helpers then assist the support seekers in completing the process of cognitive restructuring by creating a new thought in place of the original unhelpful thought. Anyone can sign up to become a helper, but it requires completing a training process that constitutes completing the CBT manual and going through practice sessions with another support provider in training, usually requiring a minimum of 20–30 hours.

3.0.2 Dataset description. The dataset consists of 116 CBT-based peer support conversations that took place from mid-November 2021 through May 2022. Sixteen sessions were "client sessions" in which the helper was an individual who had completed the CBT-based training program, and the support seeker could be any user. In those sessions, the individuals consented to the full transcription of their conversation to be released publicly at the end of the session and made available on the platform's website.

The remaining 100 sessions in the dataset were *buddy sessions*, sessions in which two peers in training conduct a session together in which one acted as the thinker and the other as the helper. Thinkers are encouraged to think of troubling situations that are personally affecting them, so the session is an authentic session and not a role-playing scenario. Helpers use the CBT-based methodology as they would if they were conducting a session with a real user. However, if they need to communicate with one another for any reason outside of the normal session context, for example, for assistance or communication regarding scheduling, they could do so through the *back-channel*. To communicate via the back channel, helpers were taught to send their message texts either within square brackets [] or parentheses (). As back-channel conversation falls outside of the normal session structure, all buddy session data was parsed, and the back-channels were removed. Approximately 12.6% of messages sent had at least some text sent via the back-channel. Analysis of back-channel conversations and their impact on thinkers' perceived empathy is not included in the scope of this paper. From each session, collected data included the text messages exchanged between the helpers and thinker, the notes that were taken during the session, and coded anonymous participant IDs. After completion of buddy sessions, thinkers and helpers were encouraged to complete a form (Figure 2), providing feedback on the session, how successful they found it, and advice for their buddy. Thinkers were asked to answer radio box style questions on if they felt certain outcomes of the session were met, as well as answer a sliding scale question related to the session's success as a whole. They were also provided a feedback box to leave comments or suggestions for the helper. Helpers answered sliding scale questions on how they felt throughout the session and the strength of their skills. Helpers were also provided a feedback box to leave any additional notes. Of the 100 buddy sessions, 72 contained post-session feedback provided by the thinker. On average, approximately 123 messages were exchanged per session, with about 61 messages being sent from the helper and 62 from the thinker.

3.1 Privacy, Ethics, and Disclosure

Data used in this study was obtained from the Cheeseburger Therapy platform [42], with proper licensing and consent. This research

Figure 2: Thinker's (left) and Helper's (right) feedback forms

involved analyzing data that had been previously collected by the Cheeseburger Therapy platform and did not contain any personally identifiable information (PII). When users registered for a session, they were informed that their sessions may be shared with academic collaborators for research purposes. Cheeseburger Therapy states its purpose as a research collaboration and that data will be used to understand how to make quality therapy more accessible by training everyday people to provide support. The authors did not have any direct contact with human subjects during this study, only accessing the data as secondary analysis by requesting it from Cheeseburger Therapy. As a retrospective examination with only de-identified data, the research does not provide any treatment recommendations or make any diagnostic claims. This work was approved by the authors' university's Institutional Review Board.

4 METHODS & ANALYSIS

We applied a two-step mixed methods approach to analyze iCBT-based peer support sessions. We first conducted a qualitative content analysis on post-session feedback to understand what factors contributed to thinkers' perception of empathy. Next, we applied a deep learning model to quantify empathy levels in peer support sessions and analyzed the results to understand the components of empathy that the algorithmic model encapsulated.

4.1 Step 1: Support Seekers' Perspective on Factors Contributing to Empathy – Qualitative Content Analysis

We conducted a thematic content analysis to qualitatively analyze post-session feedback and identify common themes and patterns related to sessions that had low perceived empathy (sessions in which the thinker identified that they did not "feel heard and understood"). Analysis was conducted on feedback, which encompassed reflections on their experience, assessments of the helper's strengths, and suggestions for potential areas of improvement. We followed an inductive open-coding approach to identify parts of a session that lead the thinker to self-identify as having felt or not felt "heard and understood".

4.1.1 Coding Procedure. To identify distinct themes and patterns related to perceived empathy, we first divided all sessions into two datasets: 1) sessions where thinkers self-identified as feeling "heard & understood" – high perceived empathy (n = 85) and 2) sessions where users did not feel "heard and understood" – low perceived empathy (n = 15). Each author then separately read through all feedback comments, pulling quotes for each newly introduced idea to create an exhaustive list of all points that were mentioned in the feedback. Authors then met to identify codes to create one list that could accurately represent all the diverse themes and patterns within the pulled quotes. Coding results were then discussed in a second round, where we removed overlapping codes and combined related codes into larger encompassing categories. The coding process was completed when the authors determined that all ideas originally identified in the feedback could be categorized into at least one of the codes. To improve the objectivity of the coding schema, authors agreed on strict definitions for each code. Two researchers each then separately applied the coding scheme to the data by re-reading through all the original thinker's feedback and marking each text with any codes that apply. Feedback often encompassed multiple ideas, so a feedback text could be marked to belong to multiple code groups. In case of rating conflicts, a third author independently rated the sessions. There was a strong agreement between the three coders, with an inter-rater reliability (IRR) of 0.81 using Cohen's Kappa. The coded data was then analyzed to discern themes and patterns associated with thinkers' perceived empathy within iCBT-mediated text-based peer support platforms.

4.1.2 Code Scheme. In this section, we introduce the code scheme that was derived from the dataset. Our scheme is composed of 11 distinct codes, divided into empathic and non-empathic categories. Figure 3 illustrates the frequency of which codes were found in the feedback.

Code (No Empathy):

(N1) *Lack of emotional validation*: Expressed feeling ignored or lacking emotional support from the helper.

(N2) *Felt rushed*: Expressed feeling rushed or hurried through the session.

(N3) *Did not align on main concern*: Expressed that their main trouble was missed or ignored.

(N4) *Pressure to conform to CBT techniques*: Expressed feeling constrained by the structure or tools used within the session.

(N5) *Too many questions*: Expressed that the helper asked too many questions.

(N6) *Redundant statements*: Expressed that the helper kept repeating information already acknowledged or discussed.

Code (Yes Empathy):

(Y1) *Validated emotions*: Expressed that their feelings surrounding their concerns were understood and affirmed.

(Y2) *Did not feel rushed*: Expressed having space to fully discuss thoughts and concerns.

(Y3) *Externalized feelings*: Expressed that the helper facilitated the verbalization and processing of the thinker's feelings.

(Y4) *Gained a new thought*: Expressed that they found a new thought because of the session.

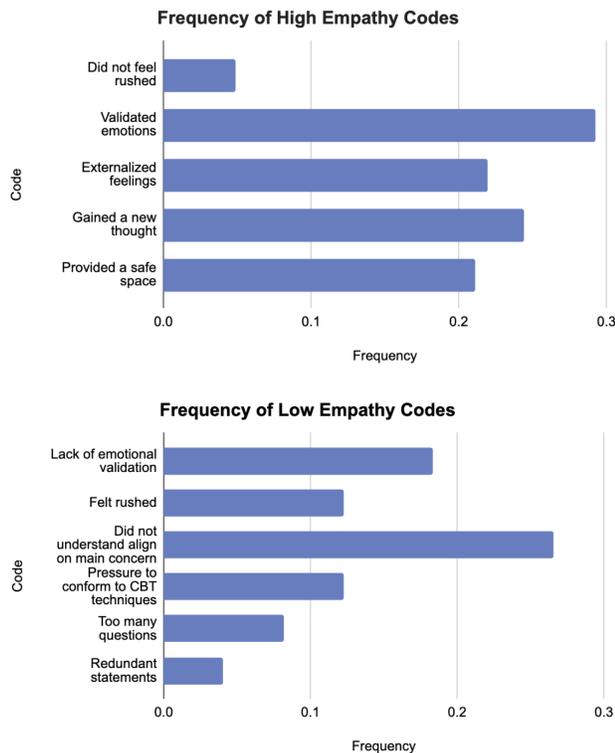(Y5) *Provided a safe space*: Expressed that the helper provided a

Figure 3: Frequency of high empathy and low empathy codes across participant feedback.

safe environment to freely explore their feelings and share their thoughts.

4.2 Step 2: Computational Approach to Measuring Empathy

Next, we applied an existing state-of-the-art deep learning model to quantify empathy computationally. The analysis of these measurements seeks to understand how automated approaches of empathy correlate to and encompass the components of iCBT sessions that were identified in the thematic content analysis that lead a thinker to feel empathy.

4.2.1 Calculating Empathy Scores. To calculate the levels of empathy conveyed by the helper in a session, we applied EPITOME, a deep learning model [39]. The model was trained using EPITOME's published dataset of Reddit[1] posts and replies that were taken from threads of 55 mental health-focused sub-Reddit groups [40]. EPITOME takes as input the text from a person seeking advice and the text from a person giving advice, then calculates the empathy level in a provider's response to a seeker's initial text. The generated empathy score measures 3 components of empathy: *Emotional Reactions*, *Interpretations*, and *Explorations*. Each of these three categories are scored as either (0) no communication, (1) weak communication, or (2) strong communication, relaying the graded extent to

which helpers conveyed the communication method in their reply. A 0 represented that the support provider did not employ the empathetic technique, a 1 that they weakly employed the technique and a 2 that they strongly employed the technique by relating it back to a specific component of the support seeker's original message. The EPITOME paper outlined a specific feature set to distinguish between a 0, 1, and 2 in each of the three sections. For example, a 1 on the exploration scale would indicate that the support provider generically inquired about more information, whereas a 2 on the exploration scale would be earned if the support seeker explicitly outlined the specific experiences and feelings in which they want to learn more about. A minimum empathy score would constitute receiving a 0 in all three sections, whereas a maximum empathy response would constitute a 2 in all sections (for a total of 6). Examples are shown in Table 1.

The authors of EPITOME have never stated that an empathy score of 6 is the aim for all empathetic responses. In fact, in their proof of concept for increasing empathy in peers' messages, they only achieved total empathy scores of 3 out of 6. Instead of empathy scores being interpreted linearly, these scores should be examined relative to one another (a 2 is better than a 1, but not necessarily two times better). Questions remain regarding the level of empathy required over continued interaction for thinkers to feel an overall sense of empathy from their helper.

To transform Cheeseburger session data into inputs compatible with the EPITOME model, each session conversation was converted into pairs of the thinker's messages and the helper's replies. Any subsequent messages sent by the thinker were concatenated into a single thinker message, and all of the helper's subsequent reply messages were concatenated together. For example, referencing Figure 4, messages (3) and (4) would be concatenated together as the thinker's text and message (5) would be the helper's text, and an empathy score would be computed given this information. Message (6) as the thinker's text, and Messages (7) and (8) concatenated together as the helper's response would be another set of inputs. Each session had, on average, 33 pairs of thinker messages and helper messages for which an empathy score was calculated. Texts were not altered when creating these data points and concatenating messages, as the EPITOME model was also trained using a dataset of uncleaned text. As participants for both the Cheeseburger dataset and the Reddit training dataset communicated exclusively through text, typos in their messages, or stylistic decisions[2] may have an impact on the relayed empathy. For example, if a helper consistently responds with multiple typos, this may lead the thinker to believe the helper was rushed, and thus feel less of a sense of patience, understanding, and empathy from the helper. By not cleaning texts, we are training and testing our empathy scores with the true texts that helpers and thinkers interacted with.

To quantitatively understand the amount of empathy conveyed by helpers throughout a session, we calculated empathy on a session level by averaging these scored pairs of helper and thinker replies, as previous research has found that averaging the empathy scores of individual speaking turns across a session, correlates with session-wide empathy levels [49]. Thus, we computed one averaged emotional reactions score, one averaged interpretations score, one

---

[1]https://www.reddit.com

[2]For example, "I am happy!!!" versus "I am happy!".

	no (0)	weak (1)	strong (2)
Emotional Reactions	"I see."	"Oh no, I hope everything will be fine."	"I am so sorry for your loss."
Interpretations	"Too bad."	"I can understand your situation."	"I can only imagine how anxious you feel."
Explorations	"I understand."	"Why is that?"	"Why do you think this made you feel so hurt?"

Table 1: Paraphrased example responses and EPITOME scores based on our private dataset

Figure 4: Example of the start of a conversation between a helper and a thinker, where they discuss the thinker's troubles.

averaged explorations score, and one averaged total score (sum of emotional reactions, interpretations, and explorations subscales) per session.

4.2.2 Error Analysis. To determine the accuracy of the automatically-produced empathy scores, we selected a random sample of 100 data points. For each data point in the sample, two authors hand-rated them on the emotional reactions, interpretations, and explorations sub-scales from 0–2. The authors used the rubric outlined in the published EPITOME paper to rate these data points. To determine inter-rater reliability, as well as the accuracy of the model, we calculated the percentage of agreement, Cohen's kappa, and linearly weighted Cohen's kappa for Rater 1 and the EPITOME Model, Rater 2 and the EPITOME model, and Rater 1 and Rater 2. Results are reported in Table 2.

While the performance of the model on the Cheeseburger Therapy dataset is high, as a randomly produced score would have an expected accuracy of 33%, these results are lower than the accuracy reported by EPITOME using their published Reddit dataset. Performance was particularly weaker on the Interpretations scale.

In line with the findings of the model's original publication, we observed that the model often over-scored for exploration responses that contained questions but were not aligned with the intent of a specific exploration of the thinker's feelings or situation. For example, the response "helpful?" received an exploration score of two despite its lack of specificity. Additionally, strong interpretation reactions from the helper were often mislabeled and instead given a 0 on the interpretation scale. In these cases, an additional point was often rewarded to the emotional reaction sub-scale when, in fact, the helper was expressing an understanding of the thinker's situation or feeling and not an emotional reaction. However, on occasion, short replies from the helper, such as "haha" or "absolutely," were incorrectly judged and given an interpretation score of 2. Interpretation and Exploration sub-categories produced binary results, with the model only ever rewarding 0 or 2 and never grading any of the helper's responses with a 1. Factors like the thinker's message length did not appear to play a major role in the quality of the outputted score.

5 RESULTS

5.1 Strong Emotional Validation is a Key Factor in High Perceived Empathy

The content analysis of participant feedback from peer support sessions revealed emotional validation as a major determinant in whether a thinker felt heard. 38% (23/60) of participant feedback specifically referenced their helper's validation as a defining feature of their experience (Codes N1, Y1). In particular, participants who experienced both high and low empathy expressed that attempts by helpers to relate to shared experiences helped facilitate feelings of emotional connection and empathy, making note that *"my helper helped me to feel understood by sharing some personal experience with the trouble I was going through"* (P15) and *"my buddy was a great listener and shared similar experiences, so they were able to validate my feelings"* (P8). On the other hand, some participants noted that general affirmations could be effective as well, explaining *"just a few 'I hear you', 'I can imagine', 'I see you really care', 'this must be difficult' type phrases would go a long way"* (P33).

Interestingly, several participants also delineated emotional validation and understanding within their feedback. For example, one participant commented that their helper was *"great at verifying their understanding,"* through restatements but that they ultimately *"missed the element of empathy and care"* (P33). While another participant explained that though they *"didn't entirely connect with some of the [cognitive distortions] we went through,"* their *"helper helped me to feel understood by sharing some personal experience with the trouble I was going through"* (P15). These dichotomous statements reflect the importance of emotional validation in shaping a thinker's perception of empathy in a peer support session,

Table 2: Accuracy and Inter-rater reliability of the EPITOME model and two raters. Accuracy is defined as the percentage of agreement. Cohen's kappa (k) and the linearly weighted Cohen's kappa are also reported.

	Emotional Reactions			Interpretations			Explorations		
	Accuracy	k	Weighted k	Accuracy	k	Weighted k	Accuracy	k	Weighted k
Rater 1 & EPITOME	85%	0.73	0.76	59%	0.27	0.33	74%	0.54	0.66
Rater 2 & EPITOME	68%	0.47	0.54	52%	0.18	0.21	66%	0.42	0.50
Rater 1 & Rater 2	80%	0.67	0.72	79%	0.67	0.73	83%	0.73	0.79

Table 3: In all subscales, the accuracy of EPITOME on the Cheeseburger Therapy dataset was high, but lower than that which was originally published in the EPITOME paper. The reported Cheeseburger Therapy accuracy was computed by taking the mean of the two raters' percentage of agreement with the model.

	Published EPITOME Accuracy	Cheeseburger Therapy Accuracy
Emotional Reactions	79.43%	76.50%
Interpretations	84.04%	55.50%
Explorations	92.61%	70.00%

showing that helpers can fail to understand a participant's experience fully and still make them feel heard, and on the flip side, can fully understand a participant's experience yet make them feel ignored.

5.2 Over-reliance on CBT can Foster a Disconnect between Support Seekers and Providers

Content analysis of session feedback also highlighted thinkers' inconsistent experiences with helpers' use of CBT tools and technique, with 13% (8/60) of feedback responses calling out feeling pressure to conform to the CBT model at the expense of their authentic expression (Code N4). One participant stated that their experience *"seemed to call for something of a different approach than the usual method"* (P28) and others emphasized that their connection with the support provider was *"just off"* (P16).

These findings reflect the importance of a more client-centered approach, in which CBT techniques are adapted to fit the needs and communication style of the thinkers. Specifically, participant comments emphasized the need for slowing down to establish a therapeutic connection early on in the session before focusing on CBT methods. Thinkers expressed that a perceived lack of connection hindered rapport and understanding, with comments such as *"I think simply slowing down a bit at the beginning, to be sure the thinker feels heard and a connection/rapport with you is key"* (P28). Multiple participants did note that they *"felt understood towards the end of the session"* and even walked away with some *"helpful insights"* (P34), but reported an overall feeling that they were not heard, implying that the session failed to provide proper support when advice was not predicated on a therapeutic connection. This may contextualize the most common concern within the low-empathy group, with 7 participants identifying a helper's failure to align with their main trouble (Code N3).

5.3 Overuse of Open-Ended Questions Negatively Correlates with Empathy

The thematic content analysis of post-session feedback also shed light on thinkers' expressed concerns about the excessive use of questions (Code N5) at times in the session when they needed more space to process their thoughts, with 6.7% (4/60) of feedback responses including explicit references to the overuse of questions. Thinkers noted frustration with sessions going in circles and lacking direction. As one participant explained, *"It would have felt nice to have felt a little more of a sense of spaciousness to explore what was coming up for me, but it felt like a bit of a pressure at times to sort of get to the point, and as such, I never really got a very deep understanding of what was coming up for me"* (P24). Overuse of open-ended questions, also seemed to add to the confusion of the thinkers themselves, with one participant stating, *"I got a little confused going about in circles in this session"..., "My buddy could benefit from asking one question at a time in order to focus the session in a way that everyone can follow along"* (P21). Another stated, *"A lot of questions can sometimes get the thinker in more head space, and less connected to the feelings"* (P45). These sentiments may provide an explanation for the observed correlation between more questions and low perceived empathy. This negative relationship was also witnessed via analysis of the computational empathy measurement.

5.3.1 Exploration scores had a weakly negative correlation with the thinker's overall rating of the session. We compared the EPITOME subscores to post-session feedback in which the thinker rated the session overall from "suffering" (-1) to "enlightenment" (1). The averaged emotional reactions score ($\beta = -0.155$, $p = .486$, $R^2 = 0.01$) and the averaged interpretations score ($\beta = 0.01878$, $p = 0.9306$, $R^2 = 0.00$) did not have a high correlation with the thinker's feedback score. However, the averaged exploration score ($\beta = -0.565$, $p < 0.001$, $R^2 = 0.25$) signified a weak model of a *negative* correlation between explorations and the thinker's experience (Figure 5). This was a surprising result, as it was expected that the empathy score and all of its sub-scores would be positively correlated with the thinker's feedback on their experience. However, the sentiments identified in the content analysis may provide an explanation for
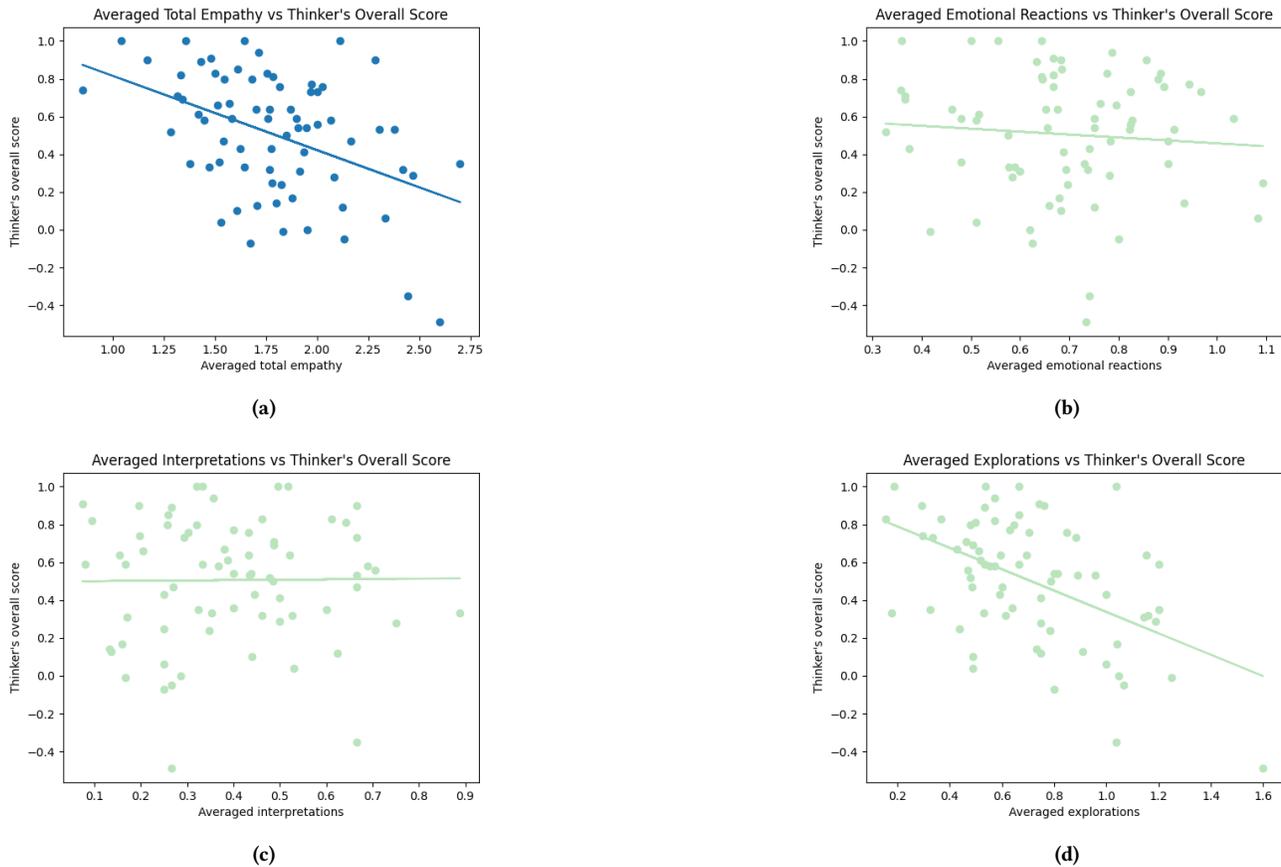
Figure 5: Averaged empathy scores vs thinker's overall score

the observed negative correlation between EPITOME explorations scores and the thinker's experience.

5.3.2 High explorations scores were negatively correlated with thinkers self-reporting that they "felt heard and understood". To further understand this negative relationship, we analyzed the conjecture (i) overuse of explorations may lead the thinker to feel they were not understood by the helper, comparing the averaged explorations score to the radiobox style feedback where the thinker indicated if they felt "heard and understood" (true or false). The results $(t(70) = 3.22, r_{pb} = -0.36, p = 0.0019)$, shown in Figure 6, indicated that the mean of the averaged explorations scores for sessions in which the thinker indicated that they did not feel "heard and understood" (M= 0.95, SD=0.29) was significantly higher than the mean of averaged explorations for sessions in which the thinker indicated that they did feel "heard and understood" (M= 0.66, SD=0.27), confirming a negative correlation between increased averaged explorations and the thinker feeling understood. There were, however, some notable outliers in this data. For example, there was a session in which the thinker indicated they did feel "heard and understood," and the averaged explorations score was 1.25, the second highest explorations score in this dataset.

## 5.4 Computational Models Reflect Training Effects on Empathy Scores

Computational analysis of the Cheeseburger Therapy dataset indicated an impact of training and CBT techniques on empathy within digital peer support. The sessions yielded an average empathy score between 0.85 and 2.7 out of a possible 6. While these scores are on the lower end, they are notably higher than the average empathy score of 1.09 published in the original EPITOME study using a dataset of untrained peers from Reddit. The distribution of results shown in Table 4 illustrates the increase in both the frequency and intensity of empathetic responses—emotional reactions and explorations. These findings reflect the efficacy of therapeutic training and signify that despite inherent limitations, computational models can partially capture nuance introduced by training in empathy assessment.

5.5 Discrepancy Between Support Seekers' Feedback and Computational Empathy Scores

85% of thinkers reported that they did feel "heard and understood". However, the averaged total empathy score per session calculated

Table 4: Distribution of results. We found a low percentage of strong scores but still a higher percentage than those of untrained Reddit peers for the Emotional Reactions and Explorations subcategories.

	Cheeseburger Therapy			Reddit (testing data)		
	no	weak	strong	no	weak	strong
Emotional Reactions	42.4%	49.1%	8.5%	66.0%	29.2%	4.8%
Interpretations	80.4%	0.0%	19.6%	53.4%	3.7%	42.9%
Explorations	68.1%	0.0%	31.9%	84.4%	3.4%	12.2%
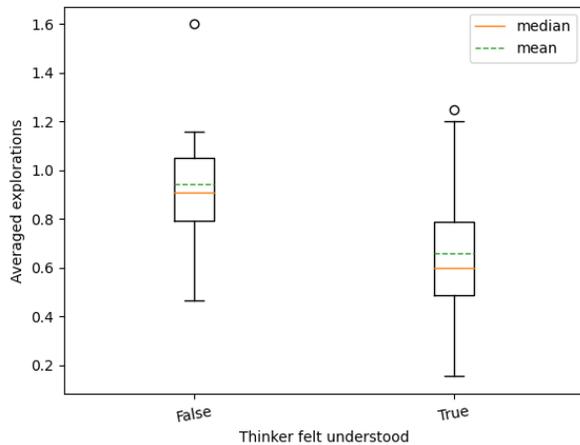
Figure 6: Distribution of scores for thinker's feedback "I felt heard and understood". The median and mean of averaged exploration scores is lower for sessions in which the thinker felt "heard and understood" than for which they did not feel "heard and understood".

by the EPITOME model was low. The averaged interpretation score across sessions was the lowest, at approximately 0.39 versus 0.66 and 0.64 for averaged emotional reactions and averaged explorations scores, respectively.

The helpers' responses scored low on the empathy scales, with average total empathy scoring an average of 1.69 out of 6. While the results suggest low empathy across sessions, this total empathy score is significantly higher than the one reported in the EPITOME original study, where they observed an average total empathy score of 1.09 out of 6. The distribution of results shown in Table 4 suggest that helpers scored a higher percentage of weak and strong emotional reactions scores and a higher percentage of strong explorations scores than that of the peers in EPITOME's Reddit dataset, who did not go through an evidence-based therapeutic training process. The difference in EPITOME scores and thinkers' self-reporting that they felt "heard and understood" could indicate that attributes that lead to support seekers' perceived empathy are not being encapsulated by the EPITOME model or add to the interpretation of the empathy scores to confirm that a 6 out of 6 is not a *necessary* score for a support seeker to feel empathy from a helper.

5.6 Explainable Biases in Empathy Rating: Interpreting Beyond the EPITOME Criteria

While conducting the Error Analysis (Section 4.2.2), the two authors noted that there were multiple cases in which the data point did not meet the qualifications of the EPITOME rubric to earn a weak or strong empathy score, but in which the authors felt that they still conveyed the sentiment that the subscale was meant to account for. The two authors examined these individual cases to identify potential biases in the EPITOME rubric. Examples where the automatically-calculated scores disagreed with the raters' opinion from this labeled sample are outlined below:

Emotional Reactions:
- In some instances, syntactical decisions conveyed additional emotional reactions. For example, *"riiiiiiight"* in the context of one of the peer supporter's replies indicated an intense feeling of relating. The rubric on which the EPITOME model was trained does not outline distinctions between syntactical decisions, despite its ability to affect a reader's interpretation.

Interpretations:
- The helper admitting that they don't currently understand what the thinker has shared, showing clear desire to understand the thinker.
- The helper sharing that they are still reading through the messages. For example, *"give me a moment to read through what you have shared"*.
- The helper conveying that they are taking notes throughout the session (that were shared with the thinker).
All of these examples convey an intent to better understand the thinker, which would not be awarded a point on the interpretation scale.

Explorations:
- The helper conveying that the thinker should feel comfortable expressing if any information in the shared notes is missing or incorrect.
- The helper expressing an intended blanket goal of encouraging sharing without asking direct question. For example, *"we can make sense of this"* or *"I welcome you to be honest with me here if there is something you want to open up about"*.
While not asking direct questions related to the thinker's feeling or situation, these examples show the intent to continue exploring and encourage a safe space.

Authors also rated the data points based on the extent to which they felt the empathetic technique was conveyed based on their CBT training, even if it did not explicitly meet a qualification on the EPITOME rubric. While the error analysis 4.2.2 helped us to

validate the EPITOME model as reported by the accuracy, the "human ratings" (Table 5) helped us identify the discrepancies and bias these models could have in rating helpers' posts. The inter-rater reliability is also reported.

The lower accuracy between the raters and the EPITOME model suggests that there are components of Emotional Reactions, Interpretations, and Explorations that are not encompassed by the EPITOME definition. However, as no clear rubric was defined between raters, the IRR between Rater 1 and Rater 2 was lower. Given that these examples were only generated from a small sample of the dataset, we anticipate that there are many other scenarios in which a helper could have conveyed one of these emphatic techniques but not met the EPITOME criteria.

6 DISCUSSION

Through a mixed-methods analysis, we investigated machine and human understanding of presence and expression of empathy in online CBT-based therapeutic conversations. We now summarize and present insights into our posed research questions on measuring and analyzing perceived empathy in text-based communication.

6.1 The Impact of Empathetic Techniques in Online Peer Support Platforms

Our results indicated that peer support platforms that adhere their sessions to some of the empathetic techniques of evidence-based psycho-therapeutic treatment, like CBT, can exhibit higher levels of empathy. Utilizing the same deep learning model for analyzing conversations, Cheeseburger Therapy helpers quantitatively conveyed higher levels of empathy than previously reported untrained peers [39] implying that following the training manual tends to increase helpers' conveyed empathy. This is particularly significant as the averaged empathy scores from our work was from sustained hour-long conversations and thus required consistent use of empathetic responses, whereas the untrained peers' data was from one direct interaction [38]. These findings suggest that (i) through initial CBT-based training, helpers learn higher empathy or (ii) trained helpers may have more incentive to be empathetic in their responses than the average internet user who is responding in mental health-related forums.

Prior research continuously highlights the therapeutic value behind CBT methods [22, 48] and the effectiveness of training in empathetic communication [29]. While CBT techniques enabled peer support providers to employ emphatic techniques, rigid adherence to the method without prioritizing genuine connection and validating emotions fostered a disconnect between support providers and seekers at the expense of their authentic expression. It is important that providers also remember to maintain flexibility in their session to avoid thinkers from feeling a lack of empathy due to a formulaic approach. Low empathy can be perceived even when support providers utilize empathetic techniques like active listening and restatements, implying that empathy is contextual. It is not solely determined by specific techniques but also depends on the broader context and the emotional connection between individuals involved in the interaction.

Our findings, highlighted by both the qualitative and quantitative analysis suggest that when leveraging too many explorations,

helpers risk causing the support seekers to feel a lack of understanding regarding their situation. While it is important for helpers to use questions to explore their understanding of the thinker's situation and perspective, limiting the number of questions may be necessary to ensure the support seeker feels the helper understands them. While we had originally hypothesized that all facets of the algorithmic empathy score would be positively correlated with the thinker's experience, this finding regarding the negative correlation between increased use of questions and the thinker experience is in line with previous research [35]. It is further backed by the results from Sharma et al. [39] in which approximately 28% of replies that received an exploration score of 0 were liked by the person seeking help. Whereas only approximately 15% of posts that were scored as 1 or 2 on the exploration sub-scale were liked by the person seeking help, indicating that high exploration scores were less often associated with liked posts.

6.2 Empathy and Beyond: The Multifaceted Aspects of Peer Connection

We found that averaged levels of total quantified empathy was low (1.69 out of 6). The highest averaged quantified empathy of any session was 2.7, suggesting that aiming for EPITOME scores of a 6 may be an unrealistic and unnecessary goal for helpers since, despite the low EPITOME scores, 85% of thinkers reported that they felt "heard and understood".

We argue that conveying the highest levels of empathy (6 for EPITOME) is not ideal since it requires consistent use of lengthy responses that may feel unnatural to the back-and-forth flow of the session in order to convey all three sub-scale components strongly. Additionally, there is not a particular need for empathy in all replies from the helper, as not all data inputs were ones in which the support seeker was specifically sharing or seeking advice. For example, in addition to model biases (Section 5.6), many data points included moments where peers were making small talk, sharing similar experiences, communicating Wifi issues, or talking about a different topic to ease in. While these utterances do not linearly associate with the three sub-components of quantified empathy, they still contribute to other components of a successful CBT session, for instance, therapist alliance, social presence, and deeper connection.

Prior literature has explored ways to increase social presence in text-based communication, such as through real-time text [23], implying that the design of peer support spaces holds value in helping individuals connect, share and communicate more. Specifically, simple design changes encompassing as little as typing indicators can help individuals feel validated and listened to [23] calling for their applicability in deep personal conversations and therapeutic communication. This implies that perceived empathy extends beyond the mere act of rewriting, as it encompasses various facets, including the design of text-based platforms. Increasing empathy involves not only the process of *task rewriting* [38] but also the consideration of how we design these text-based mental health platforms to facilitate nonverbal communication.

6.3 Moving beyond Quantified Empathy

The accuracy scores reported through the error analysis and EPITOME versus human ratings, combined with qualitative analysis,

Table 5: Accuracy of subscales based on the extent to which authors felt the empathetic technique was conveyed, even if they did not explicitly meet elements of the EPITOME rubric.

	Emotional Reactions		Interpretations		Explorations	
	Accuracy	Cohen's Kappa	Accuracy	Cohen's Kappa	Accuracy	Cohen's Kappa
Rater 1 & EPITOME	45%	.20	35%	.10	48%	.25
Rater 2 & EPITOME	52%	.18	35%	.07	58%	.34
Rater 1 & Rater 2	55%	.39	74%	.56	74%	.43

indicate that empathy is not a quantifiable metric, especially when measured over an entire session. Empathy is more than sentence re-writing, which computational approaches tend to prioritize, calling for a more comprehensive approach that measures empathy in a multifaceted manner rather than simplifying it into a one-dimensional quantitative score.

In addition, the state-of-the-art practice of assigning a single score to a support provider's response may introduce bias, especially for those who aren't native English speakers. This method also shifts their attention from other aspects that can increase empathy, such as establishing a connection early on in the session, creating safe spaces for exploration, and providing emotional validation, which can often be done by sharing similar experiences, as reported in the qualitative content analysis. While there may be benefits to being able to quantify empathy, such as relaying real time feedback to support providers, or suggesting edited responses as a means to encourage increased relayed empathy, such an evaluation method places an unnecessary strain on support providers, despite research showing that we can enhance empathy using design strategies [23] and content like CBT training [29].

7 FUTURE WORK & LIMITATIONS

While this work has begun to uncover some of the nuances related to understanding online text-based empathy, we call for future research to continue to investigate what measurements are necessary for effective support. Questions remain regarding whether a baseline level of empathy needs to be achieved in order for the support seeker to ultimately feel understood and if consistent deployment of empathetic responses is required throughout the entire session.

Through a quantitative and qualitative error analysis (Section 4.2.2), it was found that EPITOME applied to the Cheeseburger Therapy data set provided particularly high accuracy results specifically for the emotional reactions and explorations sub-scales. However, the binary results that were predicted from the interpretations and explorations scores may have been a limiting factor in the results. The paper addressed communication that occurred during the CBT structure. Future research should seek to understand how communication outside of the CBT method, such as back-channel conversations, impacts thinkers' perceived levels of empathy.

Future work also needs to develop more human-centered metrics for measuring empathy and establish a framework for selecting these metrics. Given the lack of a specified rubric in section 2.5, we encourage future work to propose a new rubric that takes into consideration some of the components of empathy outlined in the qualitative content analysis and examples in 5.6. Future studies

should investigate the effects of retraining machine learning models with respect to thinker's self-reported empathy scores.

We also acknowledge that cultural factors may affect different users' deployment and perception of empathy [4]. Given our desire to protect the anonymity of users, we did not collect any PII regarding participants, so this work also did not consider the demographic breakdown of support providers or support seekers within the analysis. Future work should investigate the perception of text-based empathy with respect to different user groups in order to provide a fuller understanding of digital empathy.

8 CONCLUSION

In this paper, we evaluated human and machine perceptions of empathy within iCBT-based peer support conversations. By conducting a mixed-methods approach that included analyzing computational models, session dialogue, and feedback from sessions, we found that CBT techniques like active listening and reflective restatements, in addition to relaying shared experiences and creating space for exploration contribute to support seeker's perceived empathy in text-based peer support settings. However, rigid adherence to the method can have opposite results. Our findings revealed that while a majority of support seekers (85%) reported experiencing high empathy during the sessions, computational models, on average, rated empathy lower (1.69 out of 6). This mismatch highlights the complexity of human empathy, proposing that empathy is not a quantifiable metric. Our study has broader implications for both mental health and AI-mediated peer support. By revealing an inconsistency between the human experience and machine interpretation of empathy, our work invites more refinement in the deep learning models used to scale empathy in iCBT. These insights also hold potential to guide the training and structure of online peer support programs, leading to more effective text-based support.

ACKNOWLEDGMENTS

REFERENCES

[1] Substance Abuse et al. 2021. Key substance use and mental health indicators in the United States: results from the 2020 National Survey on Drug Use and Health. (2021).
[2] Gerhard Andersson and Per Carlbring. 2017. Internet-assisted cognitive behavioral therapy. *Psychiatric Clinics* 40, 4 (2017), 689–700.

[3] C Holly A Andrilla, Davis G Patterson, Lisa A Garberson, Cynthia Coulthard, and Eric H Larson. 2018. Geographic variation in the supply of selected behavioral health providers. *American journal of preventive medicine* 54, 6 (2018), S199–S207.

[4] David Atkins. 2014. *The Role of Culture in Empathy: the Consequences and Explanations of Cultural Differences in Empathy at the Affective and Cognitive Levels.* Ph. D. Dissertation. University of Kent.

[5] Aaron T Beck. 1979. *Cognitive therapy of depression.* Guilford press.

[6] Graham D Bodie. 2011. The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. *Communication Quarterly* 59, 3 (2011), 277–295.

[7] David D Burns and Arthur Auerbach. 1996. Therapeutic empathy in cognitive-behavioral therapy: Does it really make a difference. *Frontiers of cognitive therapy* (1996), 135–164.

[8] Robyn L Croft, Courtney T Byrd, and Ellen M Kelly. 2022. The influence of active listening on parents' perceptions of clinical empathy in a stuttering assessment: A preliminary study. *Journal of communication disorders* 100 (2022), 106274.

[9] Vernon R Curran and Jon G Church. 1998. Not alone: Peer support through audio teleconferencing for rural women with breast cancer. *CMAJ* 159, 4 (1998), 379–381.

[10] Michael H. Eaves and Dale Leathers. 2017. *Successful Nonverbal Communication: Principles and Applications.* Routledge.

[11] David Daniel Ebert, Anna-Carlotta Zarski, Helen Christensen, Yvonne Stikkelbroek, Pim Cuijpers, Matthias Berking, and Heleen Riper. 2015. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PloS one* 10, 3 (2015), e0119895.

[12] Robert Elliott, Arthur C Bohart, Jeanne C Watson, and Leslie S Greenberg. 2011. Empathy. *Psychotherapy* 48, 1 (2011), 43.

[13] William Gentry, Todd Weber, and Sadri Golnaz. 2016. *Empathy in the workplace: A tool for effective leadership.*

[14] Abhishek Ghosh, Tathagata Mahintamani, Kshitiz Sharma, Geetesh Kumar Singh, Renjith R Pillai, BN Subodh, Debasish Basu, and Surendra K Mattoo. 2022. The therapeutic relationships, empathy, and satisfaction in teleconsultation for substance use disorders: Better or worse than in-person consultation? *Indian Journal of Psychiatry* 64, 5 (2022), 457.

[15] James Gibson, Nikolaos Malandrakis, Francisco Romero, David C. Atkins, and Shrikanth S. Narayanan. 2015. Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *Proc. Interspeech 2015.* ISCA, 1947–1951.

[16] Simon B Goldberg, Tony Rousmaniere, Scott D Miller, Jason Whipple, Stevan Lars Nielsen, William T Hoyt, and Bruce E Wampold. 2016. Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of counseling psychology* 63, 1 (2016), 1.

[17] Daniel Goleman. 1996. Emotional intelligence. Why it can matter more than IQ. *Learning* 24, 6 (1996), 49–50.

[18] Frédéric Grondin, Anna M Lomanowska, Geneviève Belleville, Maya A Yampolsky, Marie-Pier B Tremblay, and Philip L Jackson. 2023. Therapist and client perceptions of empathy in simulated teletherapy sessions. *Behaviour & Information Technology* (2023), 1–16.

[19] Sarah Hargreaves, Peter A Bath, Suzanne Duffin, and Julie Ellis. 2018. Sharing and empathy in digital spaces: qualitative study of online health forums for breast cancer and motor neuron disease (amyotrophic lateral sclerosis). *Journal of medical Internet research* 20, 6 (2018), e222.

[20] Erik Hedman-Lagerlöf, Per Carlbring, Frank Svärdman, Heleen Riper, Pim Cuijpers, and Gerhard Andersson. 2023. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *World Psychiatry* 22, 2 (2023), 305–314.

[21] Michele Heisler, Lakshmi Halasyamani, Kenneth Resnicow, Marie Neaton, Jan Shanahan, Stephanie Brown, and John D Piette. 2007. "I am not alone": The feasibility and acceptability of interactive voice response-facilitated telephone peer support among older adults with heart failure. *Congestive heart failure* 13, 3 (2007), 149–157.

[22] Stefan G Hofmann, Jade Q Wu, and Hannah Boettcher. 2014. Effect of cognitive-behavioral therapy for anxiety disorders on quality of life: a meta-analysis. *Journal of consulting and clinical psychology* 82, 3 (2014), 375.

[23] Zainab Iftikhar, Yumeng Ma, and Jeff Huang. 2023. "Together but not together": Evaluating Typing Indicators for Interaction-Rich Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 1–12.

[24] Bhautesh Dinesh Jani, David N Blane, and Stewart W Mercer. 2012. The role of empathy in therapy and the physician-patient relationship. *Forschende Komplementärmedizin/Research in Complementary Medicine* 19, 5 (2012), 252–257.

[25] Hamed Khanpour, Cornelia Caragea, and Prakhar Biyani. 2017. Identifying Empathetic Messages in Online Health Communities. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers).* Asian Federation of Natural Language Processing, Taipei, Taiwan, 246–251.

[26] Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. 2021. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes. (2021). arXiv preprint arXiv:2109.08828.

[27] Reeva Lederman, Greg Wadley, John Gleeson, Sarah Bendall, and Mario Álvarez-Jiménez. 2014. Moderated online social therapy: Designing and evaluating technology for mental health. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 1 (2014), 1–26.

[28] Sophie Leuchtenberg, Daniel Gromer, and Ivo Käthner. 2023. Videoconferencing versus face-to-face psychotherapy: Insights from patients and psychotherapists about comparability of therapeutic alliance, empathy and treatment characteristics. *Counselling and Psychotherapy Research* 23, 2 (2023), 389–403.

[29] Steven J Linton, Ida K Flink, Emma Nilsson, and Sara Edlund. 2017. Can training in empathetic validation improve medical students' communication with patients suffering pain? A test of concept. *Pain reports* 2, 3 (2017), e600.

[30] Robert R Morris, Stephen M Schueller, and Rosalind W Picard. 2015. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *Journal of medical Internet research* 17, 3 (2015), e72.

[31] Robert Robert Randall Morris. 2015. *Crowdsourcing mental health and emotional well-being.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[32] John A Naslund, Kelly A Aschbrenner, Lisa A Marsch, and Stephen J Bartels. 2016. The future of mental health care: peer-to-peer support and social media. *Epidemiology and psychiatric sciences* 25, 2 (2016), 113–122.

[33] Claire E Peck, Michelle H Lim, Melanie Purkiss, Fiona Foley, Liza Hopkins, and Neil Thomas. 2020. Development of a lived experience-based digital resource for a digitally-assisted peer support program for young people experiencing psychosis. *Frontiers in Psychiatry* 11 (2020), 635.

[34] Ulrike Pfeil and Panayiotis Zaphiris. 2007. Patterns of empathy in online communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 919–928.

[35] Erik Rautalinko. 2013. Reflective listening and open-ended questions in counselling: Preferences moderated by social skills and cognitive ability. *Counselling and Psychotherapy Research* 13, 1 (2013), 24–31.

[36] Sabirat Rubya and Svetlana Yarosh. 2017. Video-mediated peer support in an online community for recovery from substance use disorders. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.* ACM, New York, NY, USA, 1454–1469.

[37] Ashraf Atta MS Salem, Amthal H Al-Huwailah, Mahfouz Abdelsattar, Nadiah AH Al-Hamdan, Esraa Derar, Sheikhah Alazmi, Mosaad Abu Al-Diyar, and Mark D Griffiths. 2023. Empathic skills training as a means of reducing cyberbullying among adolescents: an empirical evaluation. *International journal of environmental research and public health* 20, 3 (2023), 1846.

[38] Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2021. Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach. In *Proceedings of the Web Conference 2021 (WWW '21).* ACM, 194–205.

[39] 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. (2020). arXiv preprint arXiv:2009.08441.

[40] 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.* ACM, New York, NY, USA, 1–13.

[41] Man-Kyu Song, Soo-Hee Choi, Do-Hyeong Lee, Kyung-Jun Lee, Won Joon Lee, and Do-Hyung Kang. 2018. Effects of cognitive-behavioral therapy on empathy in patients with chronic pain. *Psychiatry Investigation* 15, 3 (2018), 285.

[42] Cheeseburger Therapy Team. 2018. Cheeseburger Therapy. https://cheeseburgertherapy.org/ Accessed: 2023-22-12.

[43] Koko Team. 2020. Koko. https://www.kokocares.org/ Accessed: 2023-22-12.

[44] Christopher Terry and Jeff Cain. 2016. The Emerging Issue of Digital Empathy. *American Journal of Pharmaceutical Education* 80, 4 (2016), 58.

[45] İlkay Unay-Gailhard, Kati Lawson, and Mark A Brennan. 2023. An examination of digital empathy: When farmers speak for the climate through TikTok. *Journal of Rural Studies* 102 (2023), 103075.

[46] Jana Vyskocilova, Jan Prasko, and Milos Slepecky. 2011. Empathy in cognitive behavioral therapy and supervision. *Activitas Nervosa Superior Rediviva* 53, 2 (2011), 72–83.

[47] Fred D Wright, Aaron T Beck, and Cory F Newman. 1993. Cognitive Therapy of Substance. *Behavioral treatments for drug abuse and dependence* 137 (1993), 123.

[48] Kevin Wright. 1999. The communication of social support within an on-line community for older adults: A qualitative analysis of the SeniorNet community. *Communication Quarterly* 47, 4 (1999), S33.

[49] Bo Xiao, Dogan Can, Panayiotis G Georgiou, David Atkins, and Shrikanth S Narayanan. 2012. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference.* IEEE, 1–4.

[50] Svetlana Yarosh. 2013. Shifting dynamics or breaking sacred traditions?: the role of technology in twelve-step fellowships. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3413–3422.

[51] Sarah E. Zemore and Maria E. Pagano. 2008. *Kickbacks from Helping Others: Health and Recovery.* Springer New York, New York, NY, 141–166.

[52] Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. Comae: A multi-factor hierarchical framework for empathetic response generation. (2021). arXiv preprint arXiv:2105.08316.