Investigating the Fidelity of Digital Peer Support: A Preliminary Approach Using Natural Language Processing

By

Arya Kadakia

Senior Honors Thesis
May 27, 2022
Department of Psychological and Brain Sciences
Dartmouth College

Advisor Reader: Karen Fortuna, Ph.D.

Second Reader: Luke Chang, Ph.D.

Acknowledgements

This thesis would not be possible without the guidance of my advisor Professor Karen Fortuna. Her work on the PeerTECH platform provided me with the essential data for this project. I was also greatly assisted by Mr. Adam Whitney in developing the fidelity measure as well as by Professor Luke Chang in navigating the honors thesis process. Additionally, I would like to thank Professor Sarah Masud Preum and graduate student Mr. Yunrui Lu for helping me run the Python analyses. Finally, I would like to give my gratitude the Kaminsky Family Fund Award for supporting this research.

Table of Contents

Abstract

Adults with serious mental illnesses are disproportionately affected by chronic health conditions that are linked to inadequately managed medical and psychiatric illnesses and are associated with poor lifestyle behaviors. Emerging intervention models emphasize the value of peer specialists (certified individuals who offer emotional, social, and practical assistance to those with similar lived experiences) in promoting better illness management and meaningful community rehabilitation. Over the last few years, there has been an increasing uptick in the use of digital services and online platforms for the dissemination of various peer services. However, current literature doesn't consider the factors that may limit their fidelity, or the ability to understand and competently implement interventions as expected. Moreover, there is little research on the use of natural language processing to recognize the fundamental components of digital peer support, which are critical to evaluate an intervention's fidelity. Hence, this research will attempt to understand the individual components of digital peer support to develop a corpus and use natural language processing to classify high fidelity evidence-based techniques used by peer support specialists in novel datasets. The research hypothesizes that a binary classifier can be developed with an accuracy of 70% through the analysis of digital peer support data.

Introduction

Adults with SMI, including individuals with schizophrenia/delusional disorder, bipolar disorder, and recurrent major depression, represent 4% of the U.S. population (CDC, 2021). However, this demographic is disproportionately affected by medical comorbidity, earlier onset of chronic health conditions, and have a 10–25-year reduced life expectancy compared to the general population (Schneider et al., 2019). Such high rates of morbidity and early mortality have largely been associated with poor self-management of physical and psychiatric illnesses, which necessitates interventions that help teach medical, emotional and role management to patients (Whiteman et al., 2016).

With an already overwhelmed clinical workforce, there is an increased need for task shifting services away clinicians. Task shifting is an approach to improving mental health care delivery by shifting key processes from highly trained providers to other individuals with less training (Hoeft et al., 2018). This allows providers to work at their peak capacity of practice while non-specialist workers or community perform basic tasks like intake assessment, monitoring progress, navigating the healthcare system, or teaching supplementary self-management resources and techniques that are also essential for recovery. This consequently frees up specialists to oversee a larger caseload and deal directly with more complex cases (Kanzler et al., 2021).

Peer Support

Task-shifting staff and mental health first-aid providers can also be peers of the individuals they serve, based on age, location, environment, developmental stage, or occupation. Certified peer support specialists have the potential to address both provider and patient-based barriers to the use of self-management programs among people with SMI as they comprise one of the fastest growing mental health workforces and have shown empirical support for their ability to promote engagement in self-management apps. These individuals are people diagnosed with a mental illness who are hired, trained, and certified to provide Medicaid reimbursable peer support services (Fortuna et al., 2022). Peer support has been defined as a combination of emotional and social support along with expertise, companionship, and sense of belonging that is mutually offered by persons with a lived experience of a mental health condition, trauma, or extreme states of distress to others sharing a similar lived experience to bring about a self-determined personal change (Solomon, 2004).

In the past 15 years, peer support/peer-supported services have radically expanded across the world. (White, 2020; Chinman et al., 2014; Fortuna et al., 2020). The services include inpatient, outpatient, and community-based support services for individuals with mental health challenges or substance abuse offered by individuals who themselves identify as experiencing similar challenges and are maintaining well or in recovery (Mead, Hilton, & Curtis, 2001; Solomon, 2004). Over 30,000 peer support specialists in the United States offer publicly reimbursable mental health services throughout 43 different states (Cronise et al., 2016; Daniels et al., 2017). As peer support services proliferate, there has been growing research on their effectiveness on service users (White, 2020; Chinman et al., 2014; Fortuna et al., 2020). Several reviews have even demonstrated that peer specialists added to a clinical team as a supplementary service or to deliver a specific recovery curriculum have shown outcomes such as decreased hospitalization, increased client activation, greater treatment engagement, more satisfaction with life situation and finances, better quality of life, as well as less depression and fewer anxiety symptoms (Chinman et al., 2014; Davidson, Chinman, Sells, & Rowe, 2006).

Peer Support Models

To become a certified peer support specialist, one needs to complete a course through a certified program. However, there is no single pioneering school for peer support. Training can be done through various programs with slightly varying principles and can be utilized in different contexts for different populations. Key tenets revolve around helping service users in understanding and defining their goals for recovery, helping them monitor their progress and learn new skills, supporting them in their treatment, modeling effective coping strategies, aiding them in self-advocacy, and developing and supervising recovery plans (*CPS Job Description*, n.d.). Recovery planning is an essential aspect of peer support that involves aspects of community engagement and self-management beyond mental health, and many specialists use the Wellness Recovery Action Plan (WRAP), an evidence-based wellness tool, to aid in the planning and implementation of recovery *(What Is WRAP?,* n.d.).

Some national peer support certifications (in-person and virtual) are outlined here.

1) <u>Mental Health America</u> developed the National Certified Peer Specialist (NCPS) Certification program to train peer supporters. It a voluntary, examination-based certification that can be completed along with state certifications to enhance the quality of peer support and create of network of certified individuals (*National Certified Peer Specialist (NCPS) Certification*, n.d.).
2) <u>Copland Center for Wellness and Recovery</u> is an approved Certified Peer Support Specialist (CPS) training by the Pennsylvania Office of Mental Health, Substance Abuse Services (OMHSAS) and the Pennsylvania Certification Board (PCB). It is 10-day hands-on training that teaches individuals various practices of peer support and recovery planning

(along with WRAP) through experiential learning methods (*Certified Peer Specialist Training (CPS)*, n.d.).

3) Intentional Peer Support is a training developed by Shery Mead to teach peers to develop greater awareness of personal and relational patterns and communicate more purposefully and in a trauma-informed way. It teaches individuals to redefine mental health through connection, improving their world view, and engaging in mutual relationships (*What Is IPS?*, 2013).
4) Georgia Peer Support Institute (GPSI) is a three-day immersion in peer support designed by the Georgia Mental Health Consumer Network. It aims to teach principles of recovery from behavioral health issues, characteristics of peer-run and recovery-oriented services, and skills to improve recovery (*GA Peer Support Inst.*, n.d.).
5) Appalachian Consulting Group trains and consults in behavioral health to promote a national workforce of peer specialists. They developed the Peer Specialist Core Recovery Curriculum Training, a 4-day session that teaches individuals basic recovery techniques as well as their own developed tools. For instance, the Catch it! Check it! Change it! Process helps combat negative self-talk while the five-step 'PICBA' process is used for problem solving (*Peer Specialist Core Recovery Curriculum Training | ACG*, n.d.).

PeerTECH Platform

New innovations aligned with the new definition of peer support have led to blended models of peer support that are enhanced with evidence-based techniques to support overall health and recovery (Chinman et al., 2014; Fortuna et al., 2020). The feasibility, acceptability, and preliminary effectiveness of peer-to-peer networks, peer-delivered interventions supported with technology, and use of asynchronous and synchronous technologies has also been demonstrated

(Fortuna et al., 2020). In fact, promising evidence from two pre/post pilot studies and one randomized pilot study suggest that a three-month peer-delivered self-management technology system (PeerTECH) can facilitate evidence-based medical and psychiatric self-management skills training among people with SMI with one or more chronic health conditions (Fortuna et al., 2018). In an effort to promote community-engaged research, the system was developed and produced in conjunction with peer support specialists.

PeerTECH (Figure 1) is a multicomponent intervention that consists of two features. First is a peer support specialist facing smartphone app that includes a scripted three-month curriculum that uses video and text to support peers in delivering self-management skill development. The curriculum includes prompts to offer lived experience of medical and psychotic challenges as well as scripted evidence-based trainings on coping skills, psychoeducation, medical management, social skills, self-advocacy, relapse prevention planning, and healthy lifestyle behaviors. Second is a patient facing app that offers self-management support through a personalized daily self-management checklist, an on-demand library of self-management resources such as peer-led recovery narratives, and text and video platforms to communicate with their assigned peer support specialist (Fortuna et al., 2018).

Figure 1: The PeerTECH dashboard

Fidelity

While the effectiveness of peer support has been assessed in various outcome studies, there has been little research into the quality of peer services. Fidelity is a measure of whether an intervention is being delivered as intended (Moncher & Prinz, 1991). Fidelity standards help to create a portrait of the ideal structures and processes of a model and provide a mechanism for monitoring adherence to program principles over time (MacNeil & Mead, 2003). However, no validated data has been produced regarding peer support fidelity, an intervention that especially requires supervision due to the involvement of a disabled workforce with chronic health conditions. Some meta-analyses on randomized control trials have found little impact of peer support (Fuhr et al., 2014; Lloyd-Evans et al., 2014) and there is a growing need to measure the degree to which peer specialist services are delivered with fidelity. Moreover, evidence suggests that the certain relational qualities of peer support, compared to clinical relationships, can be eroded in regulated healthcare environments, thus increasing the need to assess the quality of services being delivered (Giliard et al., 2021). Peer support specialists have also reported stigmatization, loyalty conflicts, lack of a clear job description and feelings of insecurity and disinterest among other staff members that can provide barriers to administering fidelity-adherent interventions (Wall et al., 2021).

Yet despite this need, it is quite difficult to create a concrete fidelity criterion due to the lack of clarity on what constitutes as peer support as well as the multiple perspectives, needs, and values of individuals that engage in peer support. Chinman et al. (2016) note a lack of evidence offering insight into whether the absence of effect demonstrated in several recent trials of peer support interventions is attributable to ineffective peer support or to the intervention not having been delivered as intended. Failure to appropriately measure peer specialist service fidelity in these studies may be in large part because no instrument exists to measure it. Measuring the fidelity of peer support can also possibly improve the role of the peer specialist. Bond et al. (2000) has described how fidelity tools can increase the clarity of treatment models and can help identify the critical components that have been associated with outcomes. Moreover, the consistent inclusion of a fidelity measure in future outcome studies could help better characterize the relationship between fidelity and outcomes, improving the conclusions of peer support research.

Fidelity can be measured in multiple ways. It can be assessed unobtrusively (using notes and logs), through direct or indirect (audio/video recordings) observation, by interviews, or even by self-report (Chinman et al., 2016). Yet, besides defining fidelity merely in terms of how long peers meet or the extent to which specialists use tools, measurement should assess the principles that characterize peer-to-peer relationships - which have yet to be concretely defined. Initial steps for measuring fidelity have been under development over the last few years. An ethnography survey found seven certain key standards of fidelity - promoting critical learning, providing community, having flexibility, using instructive meetings, maintaining mutual responsibility, keeping safety, and setting clear limits (MacNeil & Mead, 2003). Another study developed a fidelity index that assessed peer support in four principle-based domains; building trusting relationships based on shared lived experience; reciprocity and mutuality; leadership, choice, and control; building strengths and making connections to community (Gillard et al., 2021). Lastly, Chinman et al. (2016) conducted a comprehensive review of peer support fidelity through an extensive literature review, an expert panel, and cognitive interviews with peer support specialists. They use a job delineation framework to find overlap with the role of the peer support and identified key activities (*reducing isolation, focusing on strengths, being a role model and sharing their recovery story, and assisting with illness managemen*t) and processes (*promoting empathy, empowerment, hope, trusting relationships).* Besides these two sets, they also identify skill building, documentation and resource sharing, and professional development as aspects of peer support. In addition, they identify certain implementation factors that can affect fidelity including how they are integrated into the treatment team, the amount of collaboration with co-workers, and the quality of leadership support received by supervisors.

NLP Measures

Natural language processing (NLP) is a subfield of artificial intelligence that is concerned with how a computer recognizes and analyzes unstructured language in a dataset that is not premeditated or consciously planned. Currently, NLP models used to scale-up fidelity have focused primarily on clinical interventions. One notable instance is Lyssn.io, a behavioral health analytics company that developed an artificial intelligence assessment platform for recording and managing session files. Using automatic speech recognition and machine learning, their tool automatically summarizes the content of cognitive-behavioral therapy sessions, estimates the intervention’s competency and the clinician's level of empathy, and offers assessments to mental health professionals or behavioral health organizations to improve the quality of services they provide (*Predicting CBT Fidelity like a Human - Lyssn*, 2021). Furthermore, NLP has been implemented in other domains including medicine and the social sciences. Natural language is often used to extract medical information including diagnoses, medications, and the clinical experience. One medication information extraction approach for primary care visit conversations showed promising results, extracting about 27% more medication mentions from our evaluation set while eliminating many false positives in comparison to existing baseline systems (Ganoe et al., 2021). In addition, natural language processing has been used to analyze the language of mental health and self-reported diagnoses on social media platforms such as Twitter (Coppersmith et al., 2015).

Further application includes the identification of empathy in text (Sharma et al., 2020), development of a medical corpus to assist in clinical note generation (Shafran et al., 2020), and the measurement of counseling conversation outcomes based on linguistic aspects of conversations (Althoff, Clark & Leskovec, 2016). However, such tools developed for clinicians would not be helpful in the case of peer support since there is significantly different textual content. Unlike manualized CBT techniques and outcomes, peer support's focus is primarily on individual lived experiences, goal setting, and a person-centered, humanistic approach towards techniques and outcomes. In addition, current information extraction tools to support clinicians are developed using a lot of annotated data, feedback from domain experts, and medical datasets and medical knowledge bases that efficiently capture and represent domain knowledge. Such resources are not available for peer support or serious mental illnesses.

Current Research

There is a currently a gap in knowledge regarding what is generalized fidelity-adherence for peer support. No current certified program includes rigorous training of lay interventionists or real-time fidelity monitoring to ensure interventions are being delivered with fidelity and offering continuing education training. There is substantial evidence demonstrating that interventions can greatly benefit from greater fidelity adherence and the proposed research can help overcome the challenges associated with this. Furthermore, while many peer support services are using secure smartphone-based apps to deliver services, none are currently using natural language-informed content detection and flagging systems, auto-generated fidelity suggestions based on evidence-based practices, and data-informed peer support reminders and prompts to help organize intervention delivery. The proposed research will act as a first step to this by assessing the fidelity adherence of peer support using qualitative and quantitative methods. Using indirect observation, it will first examine the key components of digital peer support as offered through PeerTECH synchronous sessions to begin the development of a peer support corpus and then identify the overlap between the text with the certified manuals and fidelity tools discussed above. Furthermore, this research hopes to understand whether topic modeling of the data yields distinguishable and appropriate topics that can also be independently associated with the themes. It will also attempt to identify the fidelity of a text in a novel dataset using natural language processing through a binary classification model (high or low fidelity).

Primarily, this research hypothesizes:

- Null Hypothesis (H_0): The NLP model will not be able to distinguish between high and low fidelity texts.
- Alternative Hypothesis (H_1): The NLP model will achieve a level of accuracy defined as 70% when classifying between high and low fidelity texts.

Methods

Sample Data

The current sample includes data from the PeerTECH platform and from social media peer support groups. 27 audio-to-text transcripts of anonymized conversations between certified peer support specialists and service users were extracted from the PeerTECH platform. Each conversation was transcribed verbatim, showing a back-and-forth dialogue between a peer support specialist and a service user. From social media, a total of 104 posts (along with 416 comments) were scraped from 6 public Facebook support groups and a total of 1,444 comments were scraped from 6 Reddit subreddits (Table 1).

Table 1: Post and/or comment count for Facebook groups and Subreddits

Facebook Group	Post Count	Comment Count	Subreddit	Comment Count
Certified Peer Specialist (UNITED)	6	37	r/MentalHealth	179
Peer Specialist Forum	5	8	r/Depression	109
Texas Peer Support Specialists	6	12	r/Anxiety	246
Maryland Peer Recovery Specialist Connections	42	65	r/DepressionHelp	298
COVID Long-Haulers	15	120	r/SuicideWatch	156
Anxiety Depression Support Group	30	174	r/Recovery	456

Extraction and Coding

The 27 PeerTECH transcripts were converted to .txt files and then transferred to the extensible Human Oracle Suite of Tools (eHOST), a public domain software available on GitHub (Leng, 2015). eHOST enables researchers to annotate texts; thereby, marking the span of the text string that represents the information of interest. Facebook data (both the posts and comments) was extracted manually and copied onto a Google Sheets document. Reddit data was scraped using an external Python script (Guardati, 2021) also available publicly on GitHub that used the Python Reddit API Wrapper (a Python module that provides a legal access to Reddit's API) and built datasets for subreddit posts and comments. While the subreddit data was scraped confidentially, all identifying information was discarded from the Facebook support groups as well.

Data Dictionary Classification

A data dictionary provides the groundwork for preprocessing NLP data and helps explore the individual components of the text as well as their potential relationships. It comprises of classes (or groups) that are associated with multiple items (can be entities like single words or even phrases). The 27 PeerTECH transcripts were utilized for this process using eHOST. Since no natural language dictionary exists for peer support, the production of the dictionary was an iterative process, which meant that a grounded theory approach provided the optimum way to classify and identify the themes that constituted a digital peer support session. The framework outlined in Chun Tie et al. (2019) was roughly followed for this process. A preliminary reading was conducted to identify specific features and the transcripts were divided into three sets of five and one set of seven. For the initial coding stage, texts from the first set were broken into excerpts and words or phrases that were similar in content or shared sentiment were annotated. Each tag was given a code depending on the information it provided. In the intermediate coding stage, the next set was examined, and entities were given codes from the prior set depending on their similarity while repeated tags that didn't fit in any specified codes were given new ones. Since data saturation is debatable with such limited data, the codes were reviewed and compared at this point and those with conceptual reoccurrences were placed under a single category. In the advanced coding stage, a similar process was conducted with the third set. Categories were further grouped together based on whether they had a relationship due to an underlying theme, shared properties, or common function that they provided in the text. These larger grouped phrases were also given preliminary labels and acted as the precursor for the core themes. The last set was then reviewed and annotated using the labels (code, category, and theme) generated in the previous sets while anomalous tags were discarded. Finally, all the transcripts were reviewed together, cleaned of unique tags, and relevant tags missing categories and/or themes were classified with appropriate ones. The categories and themes were given definitions and renamed as attributes and classes respectively (based on eHOST convention).

<u>Example</u> (process outlined in Table 2)

Service User: I don't wanna be sick. I notice that when I'm sick, I have like no energy, don't wanna do anything, just sleep, so that's not very positive.

Peer Support Specialist: So, it's harder to be active. How do you feel physically if you're stressed or in a bad mood?

Table 2: Grounded theory process example

Codes	**Categories**	**Theme**
Don't wanna be sick	Illness Recovery	Goals
Be active	Fitness	

Sleep	Symptoms	Illness
No energy		
It’s harder	Validation	Peer Support
Bad mood	Lived experience	
Stressed		
Feel physically		(Removed: low occurrence)
Very Positive	(Removed: low occurrence)	

Fidelity Classification

Operationalizing fidelity required a novel approach as there is no universal certified fidelity measure designed using natural language processing. Moreover, the practice of peer support is not homogenous with various schools establishing slightly different measures based on their target goals. Due to the limitations of this project, its fidelity will only be explored as a binary classification problem wherein phrases or words from the conversations between peer support specialists and patients will be classified into two classes: high fidelity or low fidelity.

For this research, principles of four certified peer support training institutions (Mental Health America, Copland Center, Intentional Peer Support, and Appalachian Consulting Group) and two peer-reviewed articles on fidelity measures (Chinman et al., 2016 and MacNeil & Mead, 2003) were reviewed. Key parameters were outlined and divided into four categories through which fidelity could be examined: the process of peer support, the attitude of the peer support specialist, the content of the session with the service user, and the specific techniques employed to facilitate conversation.

A certified peer support specialist was also asked to independently validate and modify the operationalization of peer support features. Besides developing markers of fidelity adherence, the peer support specialist also highlighted certain indicators of low fidelity that could be present in the natural language data. 21 out of the 27 PeerTECH transcripts and all the comments from the Facebook groups and the subreddits were utilized for marking fidelity. The texts were transferred to Google Sheets and annotated using blue highlights for high fidelity and red for low. 1265 entities were marked for high fidelity and 516 for low fidelity. The peer support specialist was then given a random sample of texts (around 20% of the full dataset) and was tasked to find any discrepancies in the annotations, which were then corrected upon review.

Topic Modeling and Classifier

The overarching goal of developing a natural language model was to identify fidelity-adherence in the interaction between a peer support specialist and a service user. To do so, the fidelity annotations was first explored qualitatively using two topic modeling techniques. These tools help discover abstract "topics" that occur in a collection of unstructured texts. Not only can this group by content, but also by certain hidden semantic structures, potentially facilitating the development of identifiers for high and low fidelity. First, the data file had to be processed for analysis using the Natural Language Toolkit (NLTK) program. This required removing punctuation, removing stop words (common words), tokenization (diving strings into smaller units), stemming (removing affixes), and lemmatization (reducing words to their base form). Next, two topic model analyses were conducted:

1) Latent Dirichlet allocation (LDA): a generative statistical model that generates unobserved groups (topics) based on similarity between words in texts. LDA is useful since it not only

generates topics but also classifies and ranks words based on its relevance to the topic. The model was run on low and high-fidelity tags separately to identify relevant words.

2) BERTopic: a type of topic modeling technique that can create dense clusters of words that can be interpreted into topics as well as retains key words in the topic description itself. First, the texts are vectorized into its dense vector representation (assigning numerical representations to semantic meaning) and the dimensionality is reduced (input variables are reduced). Finally, similar text segments are clustered together and are given topical markers. BERTopic was run on all annotations together to independently identify clusters.

For the final part, a publicly available pre-trained Bidirectional Encoder Representations from Transformers (BERT) classifier from Hugging Face (an online Artificial Intelligence community) was used (*Distilbert-Base-Uncased-Finetuned-Sst-2-English*, n.d.). Google-developed BERT is a transformer-based language model, meaning that it differentially weights the significance of different items/words in an input sentence. It is quite useful for binary classifications, and therefore is appropriate to distinguish between high and low fidelity. DistilBERT was used for this analysis, a smaller version of BERT that was pretrained using the same text corpus and performs masked language modeling (hides 15% of the words in the input then run the entire sentence through the model to predict the masked words). Since this model is a version of the BERT base model, fine-tuning using the labeled fidelity phrases and their original texts was necessary. The texts were divided into three sets: the training set for fitting the parameters of the model (n = 1138), the validation set for finding the optimal values for the hyperparameters (n=285), and the testing set to evaluate the performance of the model (n=356).

Results

Data Dictionary Classification

The final data dictionary included 8 classes: Medication, Illness, Illness Management, Psychoeducation, Goals, Peer Support, Therapeutic Techniques, and Determinants. Definitions for each class is provided in Table 3. The attributes for each class along with their definitions are present in Appendix A.

Table 3: The 8 classes from the dictionary and their respective definitions

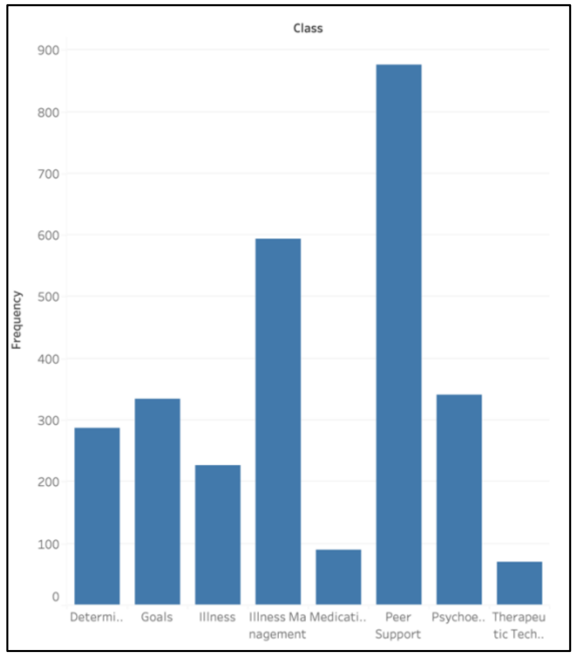
Class	**Definition**
Medication	Information about actual medication taken by peer support specialist or service user.
Illness	Details about peer support specialist’s or service user's current or past illness(es).
Illness Management	Additional activities/tasks that are used/advised by peer support specialist or service user to manage their illness. They are present or future-oriented and include specific behaviors.
Psychoeducation	General information from peer support specialists about mental/physical health and illnesses. NOT actual habits of the peer or service user.
Goals	Goals established by the peer or service user. They are future-oriented and abstract (not concrete steps).
Peer Support	Techniques used by the peer support specialist to help the service users to better process their thoughts, feelings, and actions.
Therapeutic Techniques	Techniques used by therapists that can benefit in the reduction of symptoms.
Determinants	Specific underlying *past* factors that magnify or ameliorate a physical or mental illness. Can be actual experience of the service users or talked about by the peer.

Figure 2 below demonstrates the reported frequencies of tags in each class. A table with specific values can be found in Appendix B.

The content of the classes was further explored through the generation of word clouds depicting the frequency of any word or phrase with greater than 2 instances in the text (Figure 3).

For each class, the top three most common tags (and their corresponding attributes) are presented below –

Figure 2: Frequency of tags across classes



1. Determinants: family (protective factor: family), my friends (protective factor: friends), stress (stressors)
2. Goals: healthier (illness recovery goals), happier (lifestyle goals), weight loss (fitness goals)
3. Illness: depression (diagnosis), label (illness attitude: negative), diabetic (diagnosis)
4. Illness Management: medication (relapse prevention: medication use), exercise (adaptive behavior: physical activity), tai chi (adaptive behavior: wellness activity)
5. Medication: weight gain (side effects), sedative (intended effects), insulin (name)
6. Peer Support: that's okay (validation), it's hard (lived experience), I think (reflection)
7. Psychoeducation: mind and body are connected (physical, social, and mental health connection), depression (mental illness), mental health (mental health)
8. Therapeutic Technique: positive self-talk (cognitive), distraction technique (behavioral), own choice (choice)

Figure 3: Words clouds showing highest frequency words or phrases in each class. From a-h: illness management, goals, psychoeducation, therapeutic techniques, medication, peer support, determinants, and illness.

Fidelity Classification

Table 4 outlines the indicators of high fidelity as evident in the process of peer support, the attitude of the peer support specialist, the content of the session with the service user, and the specific techniques employed to facilitate conversation (including how the sentence is structured).

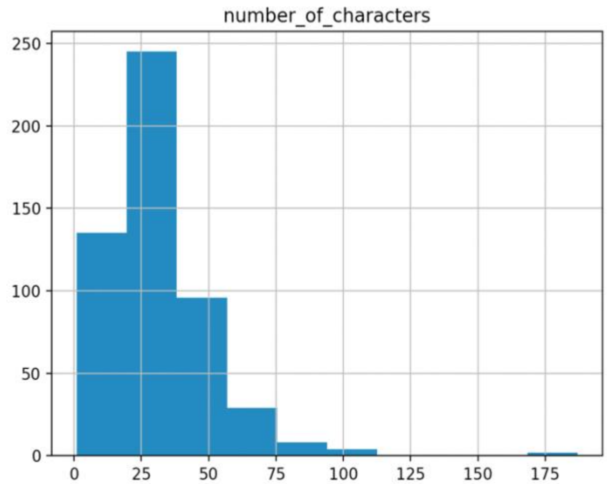
Table 3: The indicators of high fidelity separated into four categories

Process	Attitude	Content	Techniques
Voluntary and comfortable	Hopeful/Empowering	Encourage self-help/self-advocacy/self-determination	Reflective listening and open-ended questions
Mutual and reciprocal relationships	Open-minded, flexible	Help link clients to community resources	Restatement of dissatisfaction through goals
Equally shared power	Person-driven, seeking to understand cultural, family, and individual worldview	Share and reflect on each other's personal knowledge and lived experience of recovery	Statements with 'you' and 'we'
Strengths-focused	Creating comfort, honesty, and trust	Promotes critical learning and the renaming of experiences	Phrasing as moving towards what we want rather than moving away from our problems.
Transparent	Empathetic and relatable	Help reduce isolation by providing sense of community and engaging socially	Validation and minimal interruptions
Built on trust and rapport	Respectful	May act as liaison or proxy for the individual if desired	Use questions to help a peer get in touch with the life they want
	Honest and direct, genuine concern	Motivates change desired by the individual	teach coping skills to combat negative self-talk
	Instructive	Helps individuals to examine personal goals and define them in achievable ways	identify beliefs and values a peer holds that works against recovery
	Prioritize safety	Helps to navigate the system and manage illness	Promotes trauma-informed care (asking 'what happened,' not 'what's wrong')
	Make a commitment to change	Create environments that promote recovery	Use questions to help a peer identify and move through their fears
		Increase access to services and help peer prepare for a doctor's visit	Role modeling

Moreover, for low fidelity, the following specific markers were developed -

- Power and coercion
- Sharing of unsolicited advice or ambiguous/false information
- I-statements (depends on context)
- Use of extensive clinical jargon
- Disregard for sociocultural factors
- Encouraging involuntary treatment
- Asking questions for the sake of assessment rather than curiosity

A t-test between the length of characters in low and high-fidelity phrases yielded to be non-significant ($t = 0.183$, $p > .05$) with both categories averaging around 31 characters (Figure X)



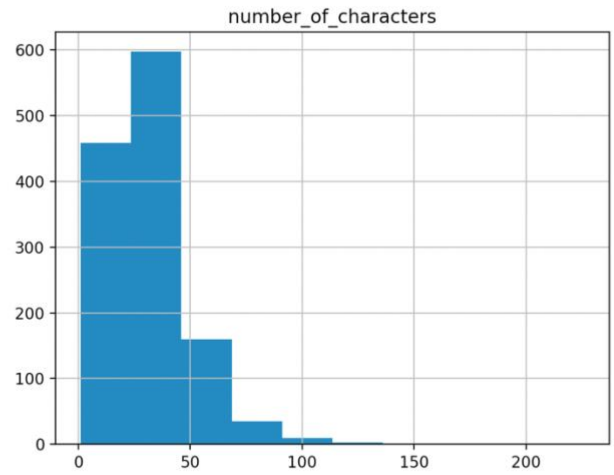


Figure 4: Histograms showing the distribution of the number of characters in low fidelity (left) and high fidelity (right).

Topic Modeling

LDA: The LDA generated two topics for both high and low fidelity each, however, they lacked any strong associations. This is evident in the coherence scores generated for each category (the degree of semantic similarity between high scoring words in a topic, with a score closer to zero representing stronger coherence). For two topics, low fidelity phrases had a score of -15.5 while high fidelity phrases had a score of -14.1. In addition, the LDA also indicated the top words present in low and high fidelity texts. Table 4 demonstrates the top twelve most frequently used words in both categories (italicized words were common in both categories).

High Fidelity		**Low Fidelity**	
Word	**Topic Weight**	**Word**	**Topic Weight**
Help	0.024	Don't	0.030
Like	0.022	*Get*	0.018
Tried	0.017	Need	0.016
Feel	0.017	*Go*	0.013
Get	0.010	*Help*	0.013
Go	0.010	*Feel*	0.012
Hard	0.009	Want	0.012
Good	0.008	Don't	0.012
Think	0.008	Sure	0.011
Take	0.008	Life	0.011
Time	0.007	Anxieties	0.011
Your	0.007	*Your*	0.010

Table 4: Topic weights for the top twelve most frequent words in high and low fidelity

Since this is nonparametric ordinal data, a Mann-Whitney *U* test was conducted to test the difference between the topic weights of high and low fidelity. However, the results were non-significant ($U = 43$, $p > .05$), suggesting that neither fidelity category had substantially greater topic weights than the other and both had similar distributions of weights for its words.

BERTopic: The BERTopic model generated 40 topics out of our given texts (Appendix C includes key words and topic count). Out of them, 29 topics appeared relevant due to some underlying similarity in content or meaning while the remaining 11 appeared to be clustered randomly. The BERTopic also generated a hierarchical cluster map (Figure 5) that placed topics together based on the cosine similarity matrix between topics (determines how similar two entities are irrespective of their size).

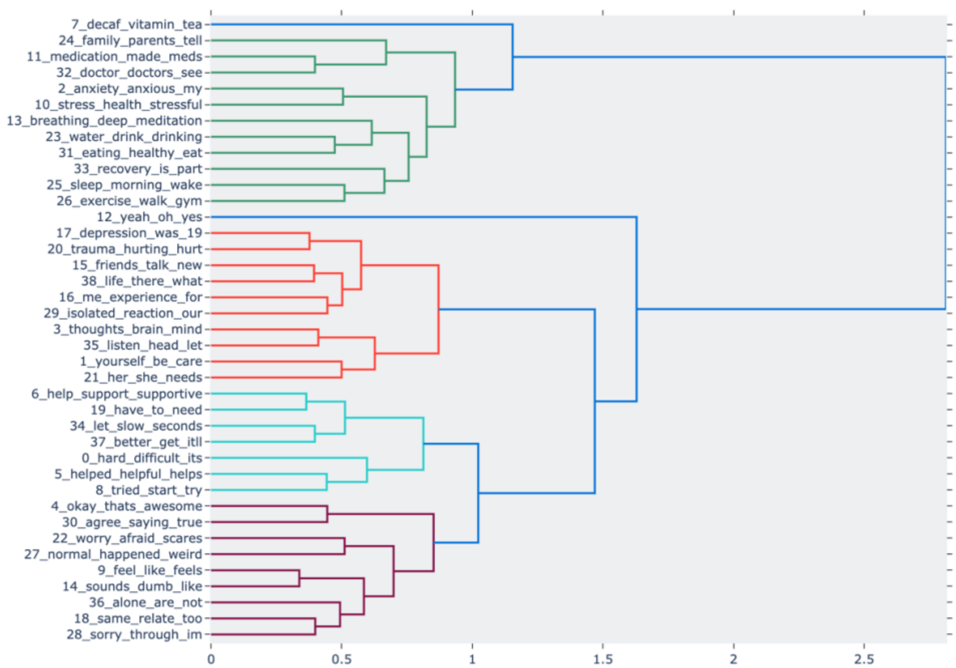


Figure 5: Hierarchical cluster map of the 40 topics generated by BERTopic.

Classifier

Using the DistilBERT model, the classifier predicted the label of phrases (either high or low fidelity) with an initial accuracy of 76%. The model also generated a loss value of 0.66 under these set parameters.

Discussion

Unpacking Qualitative Results

As desired, these results give insight not only into the common themes between the natural language data and existing fidelity standards but also how the generated topics from the NLP analysis relate to these themes. First, the data dictionary helped establish a framework to understand the individual components of digital peer support (with respect to PeerTECH). The entities in the dictionary's word clouds provide evidence for the kind of vocabulary that is present in digital peer support natural language and helps establish a starting point for the development of the corpus in this field.

Potential Grounded Theory Storyline: A certain trend emerged in the transcripts that showcased the relationship between classes. The texts demonstrated that there were two primary roles that either the peer support specialist or service user adopted. One individual would bring up their Illness and current Medication (if any), discuss their past Determinants, talk about their Goals loosely, and discuss any of their maladaptive Illness Management strategies. The other individual would then offer more adaptive Illness Management techniques and frame the Goals into more concrete steps while constantly employing Peer Support skills and occasionally some Therapeutic Techniques when necessary to facilitate dialogue. While both individuals could temporarily possess either of the roles, the certified peer support specialist more readily employed the latter skills in the conversation.

Furthermore, an overlap between the constructed fidelity measure and certain classes derived from the dictionary, namely Peer Support, Goals, Illness Management, and Psychoeducation (plus their respective attributes), is evident. For instance, the Peer Support attributes 'reflective listening', 'validation', 'reflection', and 'role modeling' along with multiple attributes from the Goals class such as ‘illness recovery’ and ‘lifestyle goals’ overlap with points in the Technique category. Similarly, Peer Support attributes of ‘lived experience’ and ‘social engagement’, multiple Illness Management attributes (specifically, the adaptive and relapse prevention behaviors), and Psychoeducation attributes like ‘mental illness’ and ‘medicine’ overlap with the Content category. Since the format of the data is text, the measurement of fidelity is limited primarily to these two categories. Attitude is slightly harder to ascertain without certain speech factors such as tone or pitch, and it’s also difficult to delineate the subjective process with small, independent text samples and a largely content-based dictionary.

Moreover, the topic modeling techniques developed topics from the unsegregated fidelity data that can also be associated with the prior results. Table 5 shows the key words of topics produced by BERTopic and their possible relationships with classes and attributes. Some are placed in multiple classes since their application depends heavily on the context of the sentence. They have been marked with an asterisk (*).

Table 5: Topics (and attributes) that can be classified in dictionary classes

Class	Topics (and possible attribute)
Peer Support	- hard, difficult, easy (lived experience) - okay, that’s awesome, good (validation) - helped, helpful, helps (role modeling) - tried, try, could (role modeling) - feels like, feels, could (reflective listening) - yeah, oh, yes, definitely (validation) - experience, experiences, personal (lived experience) - same, relate, feel (increase relatability) - worry, afraid, scares (lived experience) - normal, happens, happened (validation) - agree, saying, true (validation) - depression, depressed (lived experience) * - anxiety, anxious (diagnosis/symptoms) * - alone (lived experience)
Illness	- anxiety, anxious (diagnosis/symptoms) * - depression, depressed (diagnosis) *
Psychoeducation	- thoughts, brain, mind (mental health) - stress, health, physical, mental (stress effects/physical, social, and mental health connection)
Illness Management	- vitamin, decaf, tea, coffee (lifestyle) - medication, meds (medication use/advice)

	- breathing, deep meditation (wellness activity) - friends, chat, talk (community engagement) * - depression, therapist (therapy)* - water, drink, drinking (lifestyle) - sleep, morning, wake (lifestyle) * - exercise, walk, gym (physical activity) * - eating, healthy, eat, health (diet) * - doctor, doctors (doctor check-up) - support, supportive, others (community engagement) - family, parents, tell (community engagement) *
Goals	- friends, chat, talk (social goals) * - sleep, morning, wake (lifestyle goals) * - exercise, walk, gym (fitness goals) * - eating, healthy, eat, health (dietary goals) * - recovery, abuse (illness recovery goals)
Determinants	- Hurting, trauma, pain (trauma) - family, parents, tell (family)*

Interestingly, the classes (and many of the attributes) associated with these key topics - specifically Peer Support, Illness Management, and Goals - are also the classes that overlap with components of the fidelity measure. It is possible that there may be some potential between-class relationships that are worth exploring. For instance, the relationship between the attributes of Peer Support used and the current Illness Management strategies employed would be quite insightful. Additionally, one topic (key words - *have to, need you*) seems to be a marker for low fidelity because the phrases suggest an imperative sentence that can be coercive. Limited topics for low fidelity could possibly be due to the fewer samples for this category in the dataset.

Furthermore, analyzing the BERTopic hierarchical cluster map provides more information on the similarity of texts. Barring a few exceptions, clustered topics seem to largely fall under dictionary classes and it's worth noting that these topic clusters were formed independent of any identifying tags such as class or fidelity level. In the map, topics 7 (*decaf*, *vitamin*, *tea*), 24 (*family, parents*), 11 (*medication, meds),* 32 (*doctor, doctors*), 13 (*breathing, deep meditation*), 23 (*water, drink, drinking),* 31 (*eating, healthy, eat*), 25 (*sleep, morning, wake*), and 26 (*exercise, walk, gym*) are clustered together based on similarity (along with some outliers) and these topics all map onto the Illness Management class. The topics that overlap with the Goals class – 23, 31, 25, and 26 - are also closer together than the ones unique to Illness Management. In addition, a large Peer Support cluster is present in the map and it's interesting that the analysis successfully clusters within the class based on whether the topics are associated with content-related attributes or technique-related attributes. The middle cluster is related to the former, with topics like 17 (*depression*) and 16 (*experience, experiences*) that deal with *what* the peer support specialist talks about placed together along with topic 3 (thoughts, brain, mind), another content-related topic that's under the Psychoeducation class. Meanwhile, the large bottom cluster relates to *how* the peer support specialist phrases his words and relates to the Techniques category in the fidelity measure. Topics in this sub-cluster include 0 (*hard, difficult),* 5 (*helped, helpful*), 8 (*tried, try*), 4 (*okay, awesome),* 30 (*agree, true*), 22 (*worry, afraid*), 27 (*normal, happens*), 9 (*feel like, feels),* 36 (*alone*), and 18 (*same, relate*) seem to relate to subjective responses by peer support specialists or service users rather than concrete concepts. While there are illness-related topics like 2 (*anxiety, anxious*) and 17 (*depression*), the model was unsuccessful in predicting their similarity. Similarly, certain outliers that one would expect to belong to a certain class were independent, including topic 6 (*support, supportive, others*) related to Illness Management and topic 10 (*stress, health*) related to Psychoeducation. Moreover, as expected, topic 19 (*have to, need you*) that was relevant to low fidelity is clustered with the noise topics 34 (*let, slow, seconds*) and 37 (*better, get it'll*) because it substantially lacked any similarity with the other topics. Since Peer Support and Illness Management were the largest categories that overlapped with the fidelity measure, it isn't surprising that most topics fall under these categories. Nevertheless, the fact that the topics are also clustered by certain attribute within the class clusters, suggests that these tags can be easy to recognize and can possibly be distinguished by future NLP analyses.

Potential Implications of Quantitative Results

There was no significant difference between the distribution of the number of characters in high vs. low fidelity texts. This possibly suggests that the technique (how the sentence is structured) is more relevant than the specific content in *distinguishing* between fidelity adherence, supported by the fact that the primary topic for low fidelity in BERTtopic was related to phrasing (*have to, need you*) and this was also the case for many high fidelity topics in the Peer Support class. Similarly, the LDA demonstrates the words with the highest topic weights in both high fidelity (*help*: 0.024, *like*: 0.022, *tried*: 0.017) and low fidelity (*don't*: 0.030, *get*: 0.018, *need*: 0.016) not only fall under this technique category but also represent words from topics identified in BERTopic for high and low fidelity. However, any conclusions on the nature of fidelity using this would be too preliminary without a more in-depth exploration, especially of low fidelity data (that is particularly limited in this research).

Going back to the LDA, while there is overlap between certain words in high and low fidelity topics, the weights demonstrate that some words that one would expect to fall under high fidelity are more prominent for that category (for instance, *help*: 0.024 > 0.013 and *feel*: 0.017 > 0.012). While one would assume 'your' to be a high fidelity entity since it's a you-statement, it has a greater weight in the low fidelity category (0.010 > 0.007). It is possible that additional sentence context is required to determine whether a word benefits the fidelity of a peer support specialist or not. Moreover, since a non-significant Mann-Whitney *U* result suggests that words in high and low fidelity categories had a similar distribution of topic weights, it is likely that these words individually can't be considered as the distinguishing factor between fidelity categories.

Lastly, the result of the DistilBERT classifier rejects the null hypothesis and supports the alternative hypothesis. While an accuracy of 76% seems good for an initial model and suggests that its predictions are reasonably accurate with the true data, its value needs to be taken with a grain of salt. A moderately high loss value of 0.66 suggests that the model is making some large errors. It is possible that this is not only due to the small sample size but also due to the limited sample for low fidelity available, resulting in quite flawed predictions for some of the test data.

Strengths and Limitations

This research has quite a few strengths. It offers a comprehensive overview of digital peer support, breaks it up into its foundational elements through the bottom-up construction of a data dictionary, evaluates the significance of these elements in contributing to the fidelity of digital peer support, and successfully builds a classifier to flag texts as high or low fidelity. The development of a digital peer support corpus is an extensive process, and this research helps initiate that process for future research from the ground up. Moreover, having the assistance of a certified peer support specialist to validate the fidelity elements as well as review the texts from a random sample helps improve the credibility and construct validity of the fidelity measure. On the same note, it was quite valuable to involve the peer support specialist in the formulation and validation of the constructed fidelity measure. By partnering with a relevant stakeholder in conceptualizing the problem as well as coding and reviewing the data, this research offers solutions to community-identified challenges and promotes the value of incorporating marginalized communities as partners in the research process. In addition, the use of various topic modeling techniques like LDA that explored topic weights of words and BERTopic that examined the texts independent of its fidelity label to generate clustered topics helped corroborate some of the qualitative findings from the natural language data. Ethically, there is always concern with the use of public social media data without the official consent of the user. Nevertheless, the names of Facebook users were not taken from groups during the data extraction process and the Reddit scraping tool helped eliminate almost all identifying information in the subreddit data (solely extracted the comments and posts without user data).

However, the research is also hindered by some considerable limitations. First, it uses quite a small sample that is generally insufficient for the training of the datasets of NLP. Also, there is limited data on the attributes of the classes in the data. The eHOST software was insufficient to calculate the frequencies of attributes in the transcripts and this could have been useful to compare, for instance, the difference between maladaptive and adaptive illness management techniques used. There may also be an issue of dependency, the idea that content within a session/for a particular service user would be more similar than across sessions/users, when constructing the dictionary. Thus, it is possible that some classes and attributes are better represented by some sections of the data and that BERTopic picked up on these potential dependencies when making topics instead of generating clusters from independent samples. Moreover, the sample isn't quite representative of digital peer support as a whole and can't be generalized to other populations or settings, resulting in low ecological validity. Regarding the fidelity measure, the operationalization of low fidelity was slightly poor since there is little literature focusing specifically on it. Instead, low fidelity was mainly considered to be any indication that didn't adhere to or contradicted the high fidelity markers (along with some specific features). Lastly, while 76% is a reasonably good accuracy score, the lack of sufficient training iterations plus a reasonably big loss value suggests that the score isn't a perfect indicator to evaluate the classifier. The presence of false positives is quite possible with the limitations of the dataset. The fact that there was only one primary annotator (the author) for marking the data dictionary and fidelity measures means that it is likely that some contradictions exist in the tags that is resulting in error for the classifier.

Improvements

As a pilot study into the use of natural language processing to improve peer support fidelity, there are a couple of essential next steps to improve the validity of the results. First, a larger and more diverse sample would be quite beneficial to improve the accuracy of the classifier. Typically, a dataset of around 10,000 samples would be a sufficient starting point for the development of a classifier.

Furthermore, have additional annotators would be greatly beneficial to compare tags and develop more specific definitions. Calculating an inter-annotator Kappa score for a certain subsample would be useful to understand the agreement between annotators regarding the tags. Only when Kappa is greater than a certain threshold (for instance, > .80) would one then move on to the next stage. This is useful because of the nature of our fidelity measure is primarily divided into the Content and Technique aspects. While the former is quite concrete, the latter can be interpreted quite subjectively and can possibly vary for some annotators, thus requiring a more thorough operationalization to generate a stable measure that can consistently isolate both aspects of digital peer support.

Additionally, to combat the occurrence of false positives, calculating F1 scores (by measuring precision and recall) and constructing a confusion matrix can help identify errors in the classifier. If required, a Receiver Operating Characteristic (ROC) curve could also be drawn that summarizes the trade-off between the true positive rate and false positive rate for a predictive model using different probability thresholds. Along with a Precision-Recall curve, it is a useful tool when predicting the probability of a binary outcome (high vs. low fidelity).

Future Scope and Application

The next step is to develop a transfer learning-based NLP model that would require minimal data annotation and limited domain knowledge while still capturing the required information with reasonable accuracy - even from novel datasets. For this, the inclusion of relevant high-resource labelled datasets from other mental health domains such as cognitive behavioral therapy and empathy evaluation would also greatly benefit the fine-tuning of the model to high or low fidelity indicators. These additional datasets aren't directly related to peer support but can act as a parallel corpus for their classification task. Moreover, the correlation between the outcomes of the NLP algorithm and evidence-based medical, psychiatric, and social health manuals (such as the Chronic Disease Self-Management manual) can be examined to facilitate the iterative optimization of the NLP tool. In addition, the tool would also gain considerably from outcome data and self-reports from peer support specialists and service users. This can not only help ascertain whether a subjective assessment of fidelity matched what was determined by the model but also help identify other indicators of high, and particularly, low fidelity peer support. After the development of a reasonably accurate binary classification model (with an average F1 score approaching 0.8), the NLP model can hopefully be tasked to assign scores indicating the degree of fidelity to transcripts of conversations. Currently, common dialogue classifiers focus primarily on classifying the whole dialogue segments according to some pre-defined measure (e.g., usefulness of conversation between a virtual agent and a human user). However, the notion of fidelity adherence can also further be explored in the varying degrees of textual granularity (for instance, from whole transcript to individual dialogue lines). This can help generate insight on fidelity adherence with respect to both partial and whole interaction between the peer support specialist and service user.

Additionally, after sufficient investigation into text-based fidelity, research can turn focus to the importance of speech in order to address the other aspects of fidelity including Attitude and Process. For instance, Templeton et al. (2022) explores how social connection can be assessed in conversation by measuring the speed with which people respond to each other. Their research indicated that faster responders evoked greater feelings of connection. Moreover, some individuals with SMIs tend to speak more slowly and use more pauses due to speech impairments and cognitive deficits (Cohen et al., 2014). It would be interesting to measure how response time, pauses, and speech rate play a role in fidelity as such metrics can provide valuable data to assess the relationship between peer support specialist and service users.

If the approach is found to be feasible and effective, the development of a scalable fidelity feedback loop is possible, allowing third-party digital peer support specialists to gain automated w through their desired platform. If the development of the tool progresses as planned, it will not only be able to automatically ‘flag’ high and low fidelity texts but also provide evidence-based alternatives to low fidelity entities as well as empathetic rewriting, i.e., computationally transforming low-empathy conversational posts to higher empathy. Finally, a controlled experiment would also provide great insight into whether the feedback loop is improving service user engagement and their recovery process. Studying the use of the tool in an experimental condition against a control group can help improve the internal validity while investigating the feasibility and effectiveness of the tool.

Conclusion

As a preliminary investigation, none of these results can be meaningfully viewed in isolation. However, taken together, they offer a comprehensive understanding of the components of digital peer support fidelity. In the future, the methodology and classifier can also be adapted for other kinds of telehealth interventions beyond digital peer support (for instance, virtual outpatient visits, urgent care, pharmacy, etc.). If this tool is successfully implemented, it can benefit task shifting endeavors, reduce clinician load, and improve the self-determination of peer specialists. Fidelity scores can also be used to generate reports on quality metrics and key features of sessions, and when can potentially be combined with other mHealth tools (e.g., behavioral sensing, ecological momentary assessments) to monitor service users' progress more effectively. Despite its limitations, the development of a fidelity classifier using natural language processing was an incredibly valuable first step in effectively monitoring, assessing, and improving the quality of peer support services administered using digital services.

References

About Mental Health. (2021, November 23). https://www.cdc.gov/mentalhealth/learn/index.htm

Althoff, T., Clark, C., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463-476.

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, *2*(2), 75–87.

Certified Peer Specialist Training (CPS). (n.d.). Copeland Center for Wellness and Recovery. Retrieved May 26, 2022, from https://copelandcenter.com/our-services/certified-peer-specialist-training-cps

Chinman, M., McCarthy, S., Mitchell-Miland, C., Daniels, K., Youk, A., & Edelen, M. (2016). Early stages of development of a peer specialist fidelity measure. *Psychiatric Rehabilitation Journal*, *39*(3), 256–265.

Chinman, M., George, P., Dougherty, R. H., Daniels, A. S., Ghose, S. S., Swift, A., & Delphin-Rittmon, M. E. (2014). Peer Support Services for Individuals With Serious Mental Illnesses: Assessing the Evidence. *Psychiatric Services*, *65*(4), 429–441.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 2050312118822927.

Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech Deficits in Serious mental Illness: A Cognitive Resource Issue? *Schizophrenia Research*, *160*(0), 173–179.

Coppersmith, G., Leary, R., Whyne, E., & Wood, T. (2015). Quantifying suicidal ideation via language usage on social media. *Joint Statistics Meetings Proceedings, Statistical Computing Section, 110.*

Coppersmith, G,, Dredze, M., Harman, C., & Kristy Hollingshead. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 1-10.

CPS Job Description. (n.d.). GMHCN. Retrieved May 25, 2022, from https://www.gmhcn.org/cps-job-description

Cronise, R., Teixeira, C., Rogers, E. S., & Harrington, S. (2016). The peer support workforce: Results of a national survey. *Psychiatric Rehabilitation Journal*, *39*(3), 211–221.

Daniels, A. S., Bergeson, S., & Myrick, K. J. (2017). Defining Peer Roles and Status Among Community Health Workers and Peer Support Specialists in Integrated Systems of Care. *Psychiatric Services*, *68*(12), 1296–1298.

Davidson, L., Chinman, M., Sells, D., & Rowe, M. (2006). Peer support among adults with serious mental illness: A report from the field. *Schizophrenia Bulletin*, *32*(3), 443–450.

Distilbert-base-uncased-finetuned-sst-2-english · Hugging Face. (n.d.). Retrieved May 27, 2022, from https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english

Fortuna, K. L., Marceau, S. R., Kadakia, A., Pratt, S. I., Varney, J., Walker, R., Myers, A. L., Thompson, S., Carter, K., Greene, K., & Pringle, W. (2022). Peer Support Specialists' Perspectives of a Standard Online Research Ethics Training: Qualitative Study. *JMIR Formative Research*, *6*(2), e29073.

Fortuna, K. L., Brusilovskiy, E., Snethen, G., Brooks, J. M., Townley, G., & Salzer, M. S. (2020). Loneliness and its association with physical health conditions and psychiatric hospitalizations in people with serious mental illness. *Social Work in Mental Health*, *18*(5), 571–585.

Fortuna, K. L., Naslund, J. A., LaCroix, J. M., Bianco, C. L., Brooks, J. M., Zisman-Ilani, Y., Muralidharan, A., & Deegan, P. (2020). Digital Peer Support Mental Health Interventions for People With a Lived Experience of a Serious Mental Illness: Systematic Review. *JMIR Mental Health*, *7*(4).

Fortuna, K. L., Myers, A. L., Walsh, D., Walker, R., Mois, G., & Brooks, J. M. (2020). Strategies to Increase Peer Support Specialists' Capacity to Use Digital Technology in the Era of COVID-19: Pre-Post Study. *JMIR Mental Health*, *7*(7).

Fortuna, K. L., DiMilia, P. R., Lohman, M. C., Bruce, M. L., Zubritsky, C. D., Halaby, M. R., Walker, R. M., Brooks, J. M., & Bartels, S. J. (2018). Feasibility, Acceptability, and Preliminary Effectiveness of a Peer-Delivered and Technology Supported Self-Management Intervention for Older Adults with Serious Mental Illness. *The Psychiatric Quarterly*, *89*(2), 293–305.

Fuhr D.C., Salisbury T.T., De Silva M.J., Atif N., van Ginneken N., Rahman A., Patel V. (2014). Effectiveness of peer-delivered interventions for severe mental illness and depression on clinical and psychosocial outcomes: a systematic review and meta-analysis. *The International Journal for Research in Social and Genetic Epidemiology and Mental Health Services,* 49(11), 1691–1702.

GA Peer Support Inst. (n.d.). GMHCN. Retrieved May 26, 2022, from

Ganoe, C. H., Wu, W., Barr, P. J., Haslett, W., Dannenberg, M. D., Bonasia, K. L., Finora, J. C., Schoonmaker, J. A., Onsando, W. M., Ryan, J., Elwyn, G., Bruce, M. L., Das, A. K., & Hassanpour, S. (2021). Natural language processing for automated annotation of medication mentions in primary care visit conversations. *JAMIA Open*, *4*(3), ooab071.

Gillard, S., Banach, N., Barlow, E., Byrne, J., Foster, R., Goldsmith, L., Marks, J., McWilliam, C., Morshead, R., Stepanian, K., Turner, R., Verey, A., & White, S. (2021). Developing and testing a principle-based fidelity index for peer support in mental health services. *Social Psychiatry and Psychiatric Epidemiology*, *56*(10), 1903–1911.

Guardati, S. (2021). *Subreddit-comments-dl* [Python]. https://github.com/pistocop/subreddit-comments-dl.

Hoeft, T. J., Fortney, J. C., Patel, V., & Unützer, J. (2018). Task Sharing Approaches to Improve Mental Health Care in Rural and Other Low Resource Settings: A Systematic Review. *The Journal of Rural Health : Official Journal of the American Rural Health Association and the National Rural Health Care Association*, *34*(1), 48–62.

Kanzler, K. E., Kilpela, L. S., Pugh, J., Garcini, L. M., Gaspard, C. S., Aikens, J., Reynero, E., Tsevat, J., Lopez, E. S., Johnson-Esparza, Y., Ramirez, A. G., & Finley, E. P. (2021). Methodology for task-shifting evidence-based psychological treatments to non-licenced/lay health workers: Protocol for a systematic review. *BMJ Open*, *11*(2), e044012.

Leng, C. J. (2015). *New to NLP:* [Java]. https://github.com/chrisleng/ehost.

Lloyd-Evans, B., Mayo-Wilson, E., Harrison, B., Istead, H., Brown, E., Pilling, S., Johnson, S., & Kendall, T. (2014). A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. *BMC Psychiatry*, *14*, 39.

MacNeil, C. & Mead, S. (2003). Discovering the Fidelity Standards of Peer Support in an Ethnographic Evaluation. *The Journal of Community Psychology (in review).*

Mead, S., Hilton, D., & Curtis, L. (2001). Peer support: A theoretical perspective. *Psychiatric Rehabilitation Journal*, *25*(2), 134–141.

Moncher FJ, Prinz RJ (1991) Treatment fidelity in outcome studies. *Clin Psychol Rev 11*, 247–266.

National Certified Peer Specialist (NCPS) Certification—Get Certified! (n.d.). Mental Health America. Retrieved May 25, 2022, from https://www.mhanational.org/national-certified-peer-specialist-ncps-certification-get-certified

Peer Specialist Core Recovery Curriculum Training | ACG. (n.d.). Retrieved May 26, 2022, from https://acgpeersupport.com/services/cps-training/

Predicting CBT fidelity like a human—Lyssn. (2021, March 11). https://www.lyssn.io/predicting-cbt-fidelity-like-a-human/

Schneider, F., Erhart, M., Hewer, W., AK Loeffler, L., & Jacobi, F. (2019). Mortality and Medical Comorbidity in the Severely Mentally Ill. *Deutsches Ärzteblatt International*, *116*(23–24), 405–411.

Shafran, I., Du, N., Tran, L., Perry, A., Keyes, L., Knichel, M., Domin, A., Huang, L., Chen, Y., Li, G., Wang, M., Shafey, L. E., Soltau, H., & Paul, J. S. (2020). *The Medical Scribe: Corpus Development and Model Performance Analyses* (arXiv:2003.11531). arXiv.

Sharma, A., Miner, A. S., Atkins, D. C., & Althoff, T. (2020). *A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support* (arXiv:2009.08441). arXiv.

Solomon, P. (2004). Peer Support/Peer Provided Services Underlying Processes, Benefits, and Critical Ingredients. *Psychiatric Rehabilitation Journal*, *27*(4), 392–401

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, *119*(4), e2116915119.

Wall, A., Lovheden, T., Landgren, K., & Stjernswärd, S. (2022). Experiences and Challenges in the Role as Peer Support Workers in a Swedish Mental Health Context—An Interview Study. *Issues in Mental Health Nursing*, *43*(4), 344–355.

What is IPS? | Intentional Peer Support. (2013, July 9). https://www.intentionalpeersupport.org/what-is-ips/

What Is WRAP? (n.d.). *Wellness Recovery Action Plan*. Retrieved May 26, 2022, from https://www.wellnessrecoveryactionplan.com/what-is-wrap/

White, S., Foster, R., Marks, J., Morshead, R., Goldsmith, L., Barlow, S., Sin, J., & Gillard, S. (2020). The effectiveness of one-to-one peer support in mental health services: A systematic review and meta-analysis. *BMC Psychiatry*, *20*(1), 534.

Whiteman, K. L., Naslund, J. A., DiNapoli, E. A., Bruce, M. L., & Bartels, S. J. (2016). Systematic Review of Integrated Medical and Psychiatric Self-Management Interventions for Adults with Serious Mental Illness. *Psychiatric Services (Washington, D.C.)*, *67*(11), 1213–1225.

Appendix

A: List of attributes for each class and their definitions.

Medication

Attribute	**Definition**
Name	Medication label
Dose	How much medication is taken
Frequency & Routine	How often the medication is taken and reminders that are used (eg., before dinner).
Side Effects	Additional unexpected effects from the medication.
Intended Effects	Expected effects from the medication.

Illness

Attribute	**Definition**
Diagnosis	Clinical diagnosis of the individual
Symptoms & Warning Signs	Symptom experiences and warning signs
Severity	Behaviors that signify severity. For example, hospitalization.
Current State	How the individual feels currently about illness.
Illness Attitude: Positive	Positive approach towards illness.
Illness Attitude: Negative	Negative approach towards their own illness.

Illness Management

Attribute	**Definition**
Maladaptive Behavior: Smoking	Smoking habits and instances
Maladaptive Behavior: Substances	Substances (alcohol, drug) habits and instances
Maladaptive Behavior: Lifestyle	General lifestyle practices that are detrimental to health. For example, poor sleep, not showering, etc.
Maladaptive Behavior: Diet	Poor, unbalanced diets
Maladaptive Behavior: Community Withdrawal	Not interacting or socializing within a community

Adaptive Behavior: Physical activity	Engaging in physical exercise or working out
Adaptive Behavior: Community engagement	Interacting with a community either socially or for work.
Adaptive Behavior: Wellness Activity	Healthy, beneficial activities that help cope with illness, also including recreation
Adaptive Behavior: Diet	Healthy, balanced diets
Adaptive Behavior: Doctor check-up	Meeting with a doctor to review illness
Adaptive Behavior: Lifestyle	General, positive lifestyle practices that are good for heath. For example, sufficient sleep, good hygiene, etc.
Relapse Prevention: Medication Use	Using the right medication to help with illness effects
Relapse Prevention: Medication Advice	Strategies to remember and take the right medications.
Relapse Prevention: Therapy	Engaging in therapy for help with mental illness.

Psychoeducation

Attribute	**Definition**
Physical, social, and mental health connection	The fact that physical, social, and mental health are connected as well as ways in which they are.
Stress effects	Facts about what stress stress does to the mind and body
Mental Illness	Facts about certain mental illnesses such as depression, schizophrenia, and anxiety.
Medical Illness	Facts about certain physical, medical illnesses such as COPD, obesity, etc.
Mental Health	Facts about mental health in general, not relating to any illness
Physical Health	Facts about physical health in general, not relating to any illness
Mental Illness: Causes	Facts about what biological or psychological factors cause mental illnesses.
Mental Illness: Symptoms	Facts about what symptoms/warning signs are caused by a mental illness.
Mental Illness: Experience	Facts about what individuals typically experience with a certain illness.
Medical Illness: Causes	Facts about what biological or psychological factors cause medical illnesses.
Medical Illness: Symptoms	Facts about what symptoms/warning signs are caused by a medical illness
Substance Use	Information about substance use, abuse, risk factors, and precautions.
Psychology	General terms and theories in psychology relevant to the context.
Medication	Facts about certain general medications, not relating to the service user.

Goals

Attribute	Definition
Recovery Goals	Goals for the illness that are typically the product of illness management
Lifestyle Goals	Goals to accomplish in life, including general changes in lifestyle
Fitness Goals	Goals related to improving their body's fitness
Dietary Goals	Goals relating to improving quality/quantity of food consumed
Social Goals	Goals related to increasing community engagement and meeting people
Financial Goals	Goals related to money
Academic Goals	Goals for academic purposes such as going to college

Peer Support

Attribute	Definition
Reflective Listening	Repeating and rephrasing what service user said to make them feel heard
Validation	Agreeing with what the peer might have felt using empathy
Role Modeling	Suggestions on how to do things (you should do 'x', etc.). They are small, motivating tips, not major illness management techniques/behaviors.
Social Engagement	Sharing general information and chatting about each other's lives
Lived Experience: Peer	Subjective experience of the peer with the illness (it was hard, challenging, etc.)
Lived Experience: Service User	Subjective experience of the service user with the illness (it was hard, challenging, etc.)
Homework Review	Discussing past homework and new homework suggestions
Increasing Relatability	Explaining concepts and jargon in a way that the service user can understand. Through the use of examples and metaphors
Reflection	Reflecting on a certain situation, the emotional reaction, and the consequences.

Therapeutic Techniques

Attribute	Definition
Cognitive	Cognitive techniques that help identify maladaptive thoughts and restructuring them

Behavioral	Reward and punishment techniques that can help shape behavior
Choice	Showing service user that they have a choice when it doesn't seem like it
Contingency Management	Modifying the environment (people, activities, etc.) to modify behaviors and responses.
Catharsis	Emotional relief

Determinants

Attribute	**Definition**
Triggers	Immediate causes of stress in the environment
Protective Factor: Family	Supportive and beneficial family members
Protective Factor: Community	Supportive and beneficial community members
Protective Factor: Housing	Stable housing with good resources
Protective Factor: Finances	Stable and sufficient income
Protective Factor: Education	Met required education necessary for goals
Protective Factor: Employment	Has desired and stable employment
Risk Factor: Family	Unsupportive family members or loss of family member
Risk Factor: Lack of Community	Lack of any community members to interact with
Risk Factor: Housing	Unstable housing situation or loss of good housing
Risk Factor: Employment	Lack of suitable or stable employment
Risk Factor: Family History	Family history of illnesses
Risk Factor: Trauma	Past traumatic incidences that still affect the individual
Risk Factor: Finances	Poor or insufficient finances
Risk Factor: Community Influence	Poor influences including people and environment conditions that instill bad habits
Risk Factor: Education	Lack of sufficient education needed for goals

B: Frequency of tags in each class.

Class	**Frequency**
Determinants	287
Goals	335
Illness	226
Illness Management	593
Medication	90
Peer Support	876
Psychoeducation	341
Therapeutic Technique	70

C: Key words and for topic count each topic generated from BERTopic.

Topic	**Key words**	**Freq**	**Topic**	**Key words**	**Freq**
-1	you, to, the, it	776	19	Have to, need you	26
0	hard, difficult, easy	70	20	Hurting, trauma, pain	25
1	yourself, be, care	53	21	Her, she, needs, sad	21
2	anxiety, anxious	48	22	Worry, afraid, scares	20
3	thoughts, brain, mind	44	23	water, drink, drinking	19
4	okay, that’s awesome, good	43	24	family, parents, tell	18
5	Helped, helpful, helps	36	25	Sleep, morning, wake	17
6	support, supportive, others	35	26	Exercise, walk, gym	16
7	Vitamin, decaf, tea, coffee	35	27	normal, happens, happened	16
8	tried, try, could	33	28	Sorry, though, I’m	16
9	Feels like, feels, could	32	29	Isolated, reaction, our	15
10	stress, health, physical, mental	32	30	agree, saying, true	15
11	Medication, meds	32	31	Eating, healthy, eat, health	14
13	yeah, oh, yes, definitely	28	32	Doctor, doctors	13
13	breathing, deep meditation	28	33	Recovery, abuse	13
14	Sounds like, dumb	28	34	Let, slow, seconds	12
15	friends, chat, talk	27	35	Listen, head, let	12
16	Experience, experiences, personal	27	36	Alone, are not	11
17	Depression, depressed	26	37	Better, get, it’ll	11
18	Same, relate, feel	26	38	Life there, rewarding	10