

Article

The development of an experience measure for an online digital mental health community

Charlotte Mindel¹, Lily Mainstone-Cotton¹, Santiago de Ossorno Garcia^{1*}, Aaron Sefi², Georgia Sugarman¹, Louisa Salhi^{1,3}, Holly Brick¹, Katherine Jackson¹ and Terry Hanley⁴

¹ Kooth Plc, London, United Kingdom; research@kooth.com

² Department of Psychology, University of Exeter, Devon, United Kingdom; aaron@kooth.com

³ School of Psychology, University of Kent, Canterbury, United Kingdom; lsali@kent.ac.uk

⁴ School of Environment, Education and Development, The University of Manchester, Manchester, United Kingdom; terry.hanley@manchester.ac.uk

* Correspondence: research@kooth.com ; sossorno@kooth.com.

Abstract: Online digital mental health communities can contribute to users' mental health positively and negatively. Yet the measurement of outcomes and impact relating to digital mental health communities is difficult to capture. In this paper we demonstrate the development of an online experience measure for a specific children and young people's community inside a digital mental health service. The development is informed by three phases: (i) item reduction through Estimate-Talk-Estimate modified Delphi methods, (ii) user testing with participatory action research and (iii) a pilot within the digital service community to explore its use. Experts in the field were consulted to help reduce the items in the pool and to check their theoretical coherence. User experience workshops helped to inform the usability and appearance, wording, and purpose of the measure. Finally, the pilot results highlight completion rates, difference in scores for age and community roles and a preference to 'relate to others'; as a mechanism of support. Outcomes frequently selected in the measure show the importance of certain aspects of the community, such as safety, connection, and non-judgment previously highlighted in the literature. Self-reported helpfulness scales like this one could be used as indicators of active engagement within the community and its content but further research is required to ascertain its acceptability and validity. Phased approaches involving stakeholders and participatory action research enhances the development of digitally enabled measurement tools.

Keywords: Digital Mental Health community 1; Peer Online Community 2; Experience measure 3; Forums 4; User testing 5; Quality of Care 6; Measure design and development 7; Peer support 8; Affinity maps 9.

1. Introduction

Online peer communities can provide a platform of social interaction for young people. Children and young people are considered digital natives¹, most have lived with relative ease of access to the internet since childhood. This influences children and young people's attitude towards the internet and how they seek support, with most considering the internet as the first option for seeking information, advice, or emotional support for mental health problems^{2,3}. When paired with the importance of social peer's attitudes, beliefs, and behaviour during adolescence⁴ online peer communities can offer an important form of support for young people seeking information or struggling with mental health. Those online communities can be formed through instant communication tools,

social media networks, and asynchronous forums for communication where people share content with the intention of being seen by peers.

The importance of online peer communities in supporting adolescent mental health is shown by a strong but complex relationship between online social networks, mental health and wellbeing^{5,6}. When online social networks are used to seek support, reports of depressed moods are correspondingly minimised or maximised depending whether users were passive or active in their online use^{7,8}. However, some studies have reported low-quality connections and 'comparison effects' of specific social media platforms and its links to depression^{9,10}. Liu and colleagues⁶ meta-analysis highlights how different digital communication tools and media uses affects well-being depending on the intimacy of the medium and the different types activity. Lately, the harms of some online social networks have been explored in more detail and have also been found to predict an increase in body dissatisfaction¹¹. Therefore, it is important to recognise that the nature and characteristics of the online community will influence whether the impact on mental health of users is positive or negative. For example, visiting pro-anorexia sites was negatively associated with perception of appearance, drive for thinness and perfectionism for their users¹², but online social support has shown to act as both protective and risk factor mediating how web content is internalised¹³. Conversely, online mental health communities can be seen as the analogue to face to face support groups, especially for a subset of the population aiming for advice or to express emotions online^{14,15}. The anonymity and social connectedness created in these spaces can help people to overcome stigma and make positive disclosures of experiences and problems¹⁶. Nevertheless, others have demonstrated how dependency on these communities can hinder recovery from stigma, especially in spaces without moderation or supervision¹⁷.

It is when online mental health communities have appropriate characteristics (e.g. moderation, anonymity) that they can help individuals, and maximize support when experiencing mental health difficulties to enhance coping and recovery^{18,19}. The potential negative impact of online mental health communities on young people can be mitigated using moderation of content, creating safety, preserving anonymity and other mechanisms to create a boundary environment which avoids judgement. Observations of unmoderated platforms identified signs of self-harm normalisation and increase of suicidal ideation^{20,21}. Comparatively, users of moderated mental health forums report a reduction in frequency and severity of self-harm behaviour after starting to use the forum²². Given the mixed impacts of online communities, it is important to examine the impacts of specific online communities designed to provide mental health support with the aim to understand positive engagement, reduce risk of harm, preserve safety, and enhance well-being of online mental health support experiences.

1.1. Measurement in digital mental health communities

Determining how to evaluate the impact and effectiveness of online community mental health support is a key focus for platforms providing online peer support services. The indirect and asynchronous nature of community forum support presents challenges when using standardised measures for its evaluation. Especially when the community is diverse, user-led, and not focused on a specific mental health concern or topic leading to a specific outcome or mental health difficulty such as anxiety, eating difficulties or depression.

When standardised measures have been used in online communities' research, there have been mixed results. One online peer support group for young people found users improved in anxiety scores, but did not show any changes in depression²³. Others found a non-significant reduction in depression of forum users or no differences in body dissatisfaction between forums users and control group^{24,25}.

A clearer picture on the benefits of online mental health communities is found when qualitative and mixed methods are used. Horgan and colleagues²⁴ used thematic analysis on forum posts, alongside standardised outcomes for depression and anxiety. Young

people using the forum frequently discussed the immediate benefits of sharing their feelings on the forum and described a sense of not being alone. Forum posts also mentioned the benefits of individuals comparing themselves to others, and consequently believing that their situation was less bad than previously thought. In regards to a self-harm community investigated, young people reported that they felt they learnt more about mental health from other users, compared to information sites, and they felt it easier to disclose information online, in part because they were less likely to be judged than in real life ²⁶. More recent investigations have shown how self-efficacy and access to further support can increase thanks to the use of this communities ²⁷. They can also provide a sense of belonging, tackling feelings of loneliness regarding mental health experiences ²⁸. Qualitative studies also reveal how social modelling allows encouragement between peers to use pro-social behaviour and receive support within and outside the community ²⁹.

Qualitative methods do, however, have limitations in measuring outcomes for online communities at scale. They are time intensive and cannot be used repeatedly to track user experience and satisfaction, nor be used as a method to routinely collect information about the community. However, qualitative research does provide an in-depth understanding why young people use online mental health communities, what outcomes are achieved. The results can be used as the functional theory to develop an experience measure for an online community.

Online peer mental health communities aiming to support users require understanding on how their resources and content composing the community help or hinder the user well-being. Measurements can be collected and routinely aggregated on the self-reported impact of that community experience in the platform to the user. Developing a self-reported measure for this endeavour should aid understanding on what help the content can provide, and how different user may benefit. Ultimately, a self-reported helpfulness measure will provide an indicator of active engagement, going beyond the forum analytics often reported to define engagement in digital contexts (e.g., Views, clicks, time, popularity). Measuring the helpfulness of community content will provide insights on how resources consumed contributed to a positive, negative, or neutral experience. The measure should also understand the mechanisms that lead to the experience, and what types of outcomes users are achieving in relationship with the community and their experience. A peer online community experience measure administered in a specific online community may help to recommend personalised content or identify resources that are contributing to the recovery and support of individuals using an online community.

The Kooth.com (referred to from here onwards as the service) online community is a user-led forum inside a multi-component digital mental health service where the content revolves around the changing needs and experiences of the young people in its platform. The content of the community consists of three core types of posts: (1) a co-created magazine with a combination of psycho-educational, creative, and informative content written by the service users and practitioners, (2) discussion forums authored by users, providing direct interactions between peers but still moderated by professionals, and (3) mini-activities a specific type of content created with the intention of helping users build life skills and promote planned action. All user submitted content is moderated before being published on the platform to safeguard, categorise, and age-restrict content where necessary. When designing a measure for a specific online community and its characteristics, a framework to measure quality-of-care is required, these frameworks will be specially useful when the programme theory and mechanisms of change for the online support community have been previously investigated, so both can get combined to the develop a specific and relevant measurement.

1.2. The framework to develop a Peer Online Community Experience Measure for an online digital mental health community

Donabedian's ³⁰ quality-of-care framework recommends measuring care through assessing structure, process, and outcomes. For an online community, the focus of care is

posed in the peer support aspect taking place in the forum. The measure should also consider the role of the community member within the forum: (1) the role of an active contributor of the community generating content, or (2) the role of a consumer of the community reading and engaging with the content posted community.

The design, and structure of the Peer Online Community Experience Measure (POCEM) was divided in three parts, each representing one of the domains of Donabedian's measurement framework.

- Part one: Assesses the quality of the community resource through structure, focusing on helpfulness of magazine articles and forum discussions using a 'emoji'-based Likert scale.
- Part two: Assesses why online community users found these specific structures (community resources) helpful in respect to the area of support.
- Part three: Explore what outcomes are achieved, specific to those resources considered helpful in reference to the area of support received.

Part one was developed through consultations with experts in field to develop a quantitative scale for the measure. Part's two and three were initially developed using the researched theory of change previously developed within the online digital service, which used qualitative analysis of the community content to ascertain positive outcomes achieved the online community ³¹.

1.2.1. Part one: Assessing quality through structure

According to Donabedian seminal paper³², structure is about the 'adequacy of facilities' typically referring to relatively static characteristics of the care (e.g. personnel, buildings, resources). Within an online digital community, the structure is harder to define; it can relate to the design of the forum, its moderation, but also to the content included in the community or perhaps the community itself. Constructs like helpfulness for online communities is one option to encompass structure from the framework, allowing service users to reflect their experience of the online facilities. Some Patient Reported Experience Measures (PREMs) and Net Promoter Scores (NPS) are global rating scales used to offer a reflective experience for the user or patient through a single question, providing a numeric response from users to benchmark quality ^{33,34}. The NPS is a straightforward way of measuring experience satisfaction through asking whether an individual would recommend the service to a friend; however, it's unclear how useful it is for patients ³⁵. Defining the structure as the experience of the community and its helpfulness with the question 'Did you find this part of Kooth helpful', prompting a 5-point Likert scale response from 'No' to 'Loads' was determined to mediate the complexities of assessing the quality of the structure of the support from an online forum community.

1.2.2. Part two: Assessing quality through process

The service's theory of change identified a matrix of four possible domains or mechanisms of support that a young person may experience from community engagement ³¹. These relate to the commonly used differentiation between informational and emotional support ^{36,37}, also previously explored in the service's forum data ³⁸ and provide a part of the measure assessing the 'processes' in the online community. These four domains of support, defined by either an interpersonal experience (in relation to others) or an intrapersonal experience (in relation to oneself), were previously found to be the main experiences of adolescents and their goals in online therapy ³⁹. Within part two, in which users are asked about why they found this part of the service helpful, or in the instance they didn't, what they were hoping to have found, they were given the option of selecting one of four options representing each four domains of support (1: Emotional interpersonal; 2: Emotional intra-personal; 3: Informational inter-personal; 4: Informational intra-personal).

1.2.3. Part three: Assessing quality through outcomes

The final part looked at 'outcomes', which seek to capture whether the goals of care were achieved^{40,41}. Goals achieved and combined outcomes were previously identified through qualitative analysis of the service's community forums and magazine articles. Outcomes were found through qualitative exploration of the question '*what factors influence perceived positive behaviour change for Children and Young People who participated in an online peer support intervention?*'. Six main desirable outcomes were identified for the online community: 1) Relatedness and Self-Expression 2) Hope and Help Seeking 3) Building a Safe Community 4) Digital Altruism 5) Hope and Help Receiving 6) Making Change. These represent the potential outcomes that the part three ask the user as an indicator of outcome achieved. This last part of the measure assessing outcomes will only ask those who found the structure of the community helpful in the scale in the first part of the measure.

1.3. The present study

The present study describes the (i) development, (ii) user testing, and (iii) pilot results of the implemented measure in a dynamic and multifaceted digital mental health service; This phased approach involved different key stakeholders and participants that influenced each phase of its development. To ensure POCEM is a meaningful measure for the people using it, participatory action research practices were conducted to guide the development of items and structure of the measure. An iterative design process incorporated evidence collected from practitioners, researchers, design experts, and young people using the platform. The ethos of Donabedian's^{30,32} framework was applied to the development and design of this digital community experience measure.

This measurement and its development provide an opportunity to collect data on the peer support community and its structure, processes, and outcomes within the service. The study describes the development of the POCEM divided by three phases including the implementation of the measure with a pilot testing phase administered in the online community resources from the platform.

2. Methods

A multi-phased design process was used for the development of the service's community measure, guided by participatory action research, involving iterative development, reflective decision making and real-world application of the findings^{42,43}. The development of the service's community measure involved a group of practitioners, researchers, user experience design experts and young people. In hindsight, three key phases of design supported the development of the measure. Emphasising the digital context and acknowledging the iterative nature in measure development research as an incremental process⁴⁴:

1. Item generation and reduction: A three-part measure developed with digital product experts and designers. The content of the measure and items were created by combining qualitative thematic indicators of outcomes and mechanisms. Delphi rounds were used to reduce items and explore the content for the measure.
2. User testing: To directly explore, using participatory workshops, the face validity of the measure with young people. The focus was to verify the appropriateness of language and how design of the measure was experienced on the platform as a prototype.
3. Pilot study: A 10-week pilot of the measure within the digital online community. Exploring completion rates, average scores, item frequency selection at correlations between items at each part of the measure.

2.1. Phase 1: Item generation and reduction

The process of item generation and reduction for the first phase of POCEM development was carried out using the Estimate-Talk-Estimate Delphi technique⁴⁵. The technique is used to achieve expert consensus through an iterative process of multiple discussion

sessions between a panel of experts. The Estimate-Talk-Estimate method differs from the standard Delphi technique by then allowing for verbal interaction between panel members. This interaction allows for clarification and justification of decisions made by individuals, which can be impaired if using a fully anonymous and asynchronous process ⁴⁶. The stage of independent evaluation aims to prevent the dominance of individual voices in the process, and group pressure for conformity. The process includes anonymity in the initial responses, multiple iterations of the process with controlled feedback at each stage, statistical analysis of the group responses, and expert input throughout the process ⁴⁷. The Delphi technique is frequently used in healthcare research and has previously been used to modify a social responsiveness scale ⁴⁸, while the Estimate-Talk-Estimate variation has been used in developing a framework for mental health apps ⁴⁹.

To develop a measure of peer community experiences that reflected both young people's views and expert opinions, a two-stage Estimate-Talk-Estimate Delphi process was used. Stage one involved panel members with mental health practice expertise to compose an initial pool of items based on previous theory known about the service³¹. Stage two of the Delphi process involved discussion between researchers to assess the items generated, identify links between the generated items and the constructs from the theory, and reduce initial pool of items using an inter-rater agreement approach.

2.1.1. Participants

Two panel groups were recruited, with of a total of six expert panel for the Estimate-Talks-Estimate workshops. Most experts belonged to the service and one to a university institution. Experts registered their interest in the project via email and specified whether they were interested in participating in (a) the service's Theory of Change thematic analysis, (b) the item generation stage of the online community measure, or (c) the item reduction stage of the online community measure. Experts could volunteer for multiple parts of the project. All panel members had extensive experience researching or providing support and moderation in the digital mental health platform.

Panel members were recruited to participate in two projects concurrently: the development of the service's Theory of Change³¹, and the generation of the desirable outcomes that inform the three parts of POCEM. One panel member was involved in both the item generation stage and the later item reduction stage. The continued involvement of one panel member was used to ensure continuity between the two stages (Table 1).

Table 1. Expert panel members for item generation reduction

Panel member	Organisation role	Institution	Project involvement
PM1	Mental health practitioner	Kooth	Thematic analysis; Item generation
PM2	Mental health community practitioner	Kooth ¹	Thematic analysis; Item generation
PM3	Lecturer in counselling psychology	University of Manchester	Thematic Analysis; Item generation; Item reduction
PM4	Lead researcher	Kooth	Item reduction
PM5	Chief research officer	Kooth	Item reduction
PM6	Lead experience designer	Kooth	Item reduction

PM: Panel Member

2.1.3. Procedure

Four rounds of panel workshops were carried out with the expert participants. Workshops were held face-to-face or by videoconference. Asynchronous communications through e-mail was used to prepare the experts, and anonymous questionnaires for voting were provided in each round. The rounds had different aims regarding the content relevance and structure, reduction of proposed items and changes in wording of items to improve quality. All rounds were documented through field notes supervised by the research lead (TH).

2.1.3.1. Round 1

The first stage of the process was a face-to-face meeting group with the expert panel, wherein a broad list of initial items was generated. The items were generated deductively from the desirable outcomes found in the Theory of Change of the service ³¹. The experts involved in this initial item generation were concurrently involved in the Theory of Change research, allowing for a deeper understanding on the transcript’s findings and theoretical foundation of the Kooth online community outcomes and mechanisms. The process of generating the initial pool of items utilized a thematic analysis approach, consistent with Braun and Clarke ⁵⁰ analysis in psychology research. The thematic analysis investigated the factors influencing positive behaviour change for children and young people accessing an online peer support intervention and described the desirable outcomes for the online community. The items were generated by each panel member independently and decided in group through a panel discussion process. Items were generated using the framework for at least each of the desirable outcomes and mechanisms for positive change in the online community identified in the thematic analysis (Figure 4).

2.1.3.2. Round 2

This round with panel members focused on mapping the initial pool of items to the domains of support. The domains of support were mechanisms formulated in earlier research, and they are intended to represent the high-level ‘wants’ and ‘needs’ from children and young people asking for mental health support within an online digital service. These domains of support are covered in part-two of the POCEM as the process (mechanism) that lead to that online community resource being helpful (Figure 1). After mapping domains for each item in the initial pool, panel members were asked to select the items more likely to be selected by two different types of online community members: (1) those contributing; or (2) those consuming (accessing by reading) the community resources. The workshop with experts focused on which items were more relevant to each type of community member and provide rationale on their decisions.

	Emotional Support	Informational Support
Intrapersonal	Relates to emotional needs in relation to one’s self, with expected support targeting the feelings and goals associated with personal growth, inner resources, and emotional awareness and regulation.	Informational support in relation to one’s self, such as the provision of knowledge, facts, advice or feedback on actions associated with self-improvement and personal development.
Interpersonal	Relates to an individual’s emotional needs in relation to others (e.g. family, friends, teachers, therapists etc.), such as the feelings and goals associated with forming relationships, interacting with others and opening up.	Informational support in relation to others, such as the provision of knowledge, facts, advice or feedback on actions associated with relationship processes, interactions or conflict.

Figure 1. The service’s (Kooth.com) high-level outcome of support, ‘wants’ and ‘needs’ from children and young people accessing a digital mental health support service ³¹.

2.1.3.3. Round 3

Following the second round, the panel members voted on items' relevance to identify those to be discarded. The rating was done independently and anonymously by all panel members. Panel members were given two weeks to make their evaluations. A workshop was carried out to discuss the relevance items findings, and the relative ratings of the different members. The discussion focused on whether the items were repetitive, reflected the support domains they were matched with, and were representative of the outcomes from the thematic tree.

2.1.3.4. Round 4

Following the last round of talks, panel members asynchronously and independently evaluated the wording of the items and suggested additional items relevant to the peer online community. Panel members were given three weeks to make their assessment and present their review. The workshop focused on editing the wording of the items, and further reducing the items down due to similarity, the structure of the measure, and the previous theory used to develop it. The discussion considered the independent comments made prior to workshop. The output of the last round composed the statements of the three-part measure taken to prototype generation and for user testing with children and young people as main stakeholders of the group.

2.1.4. Analysis

Descriptive statistics were used to describe participants' characteristics and frequency in votes and selection was recorded for each expert. The field notes outputs from each round of the item generation and reduction phase were discussed sequentially, influencing the materials taken to panel of experts in each round. In round three, when panel members independently and anonymously rated their preference of the items an intra-class correlation coefficient was calculated to understand agreement between panel members on their decisions.

2.2. Phase 2: User Testing

2.2.1. Participants

A voluntary purposive sample of 11 young people was recruited amongst four primary and secondary schools in Manchester, UK. The sample was used to conduct the user testing workshops. The 11 young people aged 12-17 (7 female, 4 male) expressed their interest for participating in the workshops and parental consent was provided. The study was advertised through teaching staff, participants had no previous experience as users within the digital service (Kooth.com), parental and individual consent was sought for each participant and an incentive of £10 was given to participants to attend the 60-minutes workshops. Two researchers, one with participatory research expertise and a user experience designer conducted and analysed the user testing sessions.

2.2.2. Instruments and materials

2.2.2.1. Kooth Prototype: Clickable high-fidelity

A high-fidelity prototype is a smartphone-based interactive representation of the product, the web-based service, with strong similarities to the final design in terms of details and functionality. The high-fidelity prototype was developed with the vector graphic editor Sketch software⁵¹ and included the POCEM inside the online peer support community allowing the users to click around and interact with the whole platform. In the context of measure usability, a high-fidelity prototype allows exploration of wording, structure, relevance, and comprehensiveness of the measurement and its functionality in interaction with the whole platform.

2.2.2.2. Peer Online Community Experience Measure (POCEM)

The POCEM is an online community measure (specific to the service, Kooth.com) that contains three main stages for completion. Its aim is to measure satisfaction and quality that an online community resource has in relationship with the individual. The measure automatically differentiates between contributors of the online community and readers.

The first stage contains single item question (*"Did you find this part of Kooth helpful?"*) scored with a 1-5 Likert scale (1: No; 2: Not really; 3: Don't Know; 4: A bit 5: Loads!, all scores aided with emojis) to assess the helpfulness of the online community resource. The Likert scale scores determines the helpfulness as a benchmark and branch the measurement into stage two. Depending on the scoring in Stage one a new single-item question will prompt (*"What were you hoping for?"* for 1-2 scores and *"Why did you find it helpful?"* For 4-5 scores) to select between four multiple choice (single response) question between four quadrants representing high-level outcomes (or domains) of support from the service. Respondents who selected score 3: Don't Know in the scale are not shown parts two and three of the measure. The last stage is displayed only to those users who completed stage one and two and scored between 4-5 in the stage one in regards to the helpfulness of that part of the digital service, a single item (multiple response) question (*"What type of things have you learned?"*) with specific outcomes that relate to the domain response on part two will be prompted, readers can select between 23 outcomes (across domains) and contributors can select 14 specific outcomes.

2.2.2.3. Lookback.io: screen recording & audio

Lookback.io⁵² is a user testing software for mobile UX user recording tool. It allows recording of screen interactions alongside voice audio recordings when conducting supervised sessions of remote user testing⁵³. This software allows secure storage and organization of your user testing sessions for qualitative analysis. This tool allowed the recording of both screen behavior and audio from participants attending the user testing sessions.

2.2.3. Procedure

The user testing was structured in one-to-one sessions delivered at each of the four schools. Participants were provided with a smartphone which had a loaded a high-fidelity prototype of the measure within the platform. Sessions were facilitated by a user experience expert researcher and observed by other researcher to safeguard the session and take notes. The sessions were voice recorded and screen recorded, for further transcript and analysis.

The facilitator encouraged young people to verbalise their thoughts and perceptions using the think-aloud protocol⁵⁴ as they navigated their way through the platform while following the facilitator instructions with the prototype. Instructions followed a protocol of specific tasks within the prototype measure that aim to identify any issues with the interface, allowing facilitators to observed participant specific behaviour in relationship with the task. Facilitation tasks included asking about expectations in relationship to the next event that the interface will show during the session, and whether there were any issues with the wording clarity and relevance for the measure.

2.2.4. Analysis

Affinity diagrams or KJ methods are adopted in user testing for prototype interaction⁵⁵. They are a good technique tool to synthesize and organize large amounts of qualitative data post-task, the user testing sessions were synthesized in affinity cards representing each participant observations and quotes. Such cards are later jointly analyzed to create a diagram. Affinity cards for issues more frequently raised, and for higher severity reported by participants tend to take more priority to address as changes in the prototype.

We followed the adapted four stage (creating notes, clustering notes, walking the wall, and documentation) process from Lucero⁵⁶. Researchers worked in Microsoft Excel, using rows to represent participants and columns for each affinity note. A total of 236 affinity cards were collected from field notes, screen recordings and audio recordings from each session. Rounds of clustering by researcher identified two main clusters in

reference to the measure, and to the prototype and task performed (58.48% Measure affinity notes and 41.52% Prototype and tasks affinity notes). Twelve clustering issues were collected across clusters, some directly related with the measure such as including an ‘other’ personalized option, and issues with the platform and prototype such as difficulties in navigation. The affinity diagram was then created, walking the wall exercises with other researchers and experts at the service (n=3) provided with synthesis and identification of priority changes in the measure and prototype interaction by looking at frequency, feedback notes and quotes presented in the affinity diagram. Documentation on the output from the affinity diagram discussed by experts is provided in Figure 2.



Figure 2. User Testing – Affinity diagram output from workshops with POCEM interactive prototype at Kooth.com.

2.3. Phase 3: Face validity pilot testing

In contrast to content validity which is more concerned with having the breadth and accuracy of items to measure a construct, face validity assesses the degree of respondents judging the instrument and its items are appropriate for the targeted assessment ⁵⁷. We used a 10-week pilot of the tool to collect qualitative and quantitative data from the users in the online community completing the measure inside the platform (11-25 year old service users), as well as collecting opinions about the measure from those users during the pilot from those who consented to the study.

2.3.1. Participants

The measure was iteratively released onto the service’s platform. Online service users who either contributed to a forum or submitted an article during the testing period were presented with the contributors’ measure after submission. Service users who read an article or forum were presented with the readers measure at the end of the post. Only users who expressed consent for their data to be used in research were included in the sample. Data was collected between the 13th November 2019 and the 22th January 2020.

2.3.2. Procedure

The clickable prototype of the POCEM was implemented as a feature for service improvement in the online community at the service’s platform, changes from the previous Phase 2 were included in the measure for pilot. For a period of 10 weeks the measure was tested within the platform and data was collected on users engaging with the community at the digital mental health service. Routinely collected monitoring information was used

alongside peer support data to investigate the measure performance. All users accessing the platform community were able to complete and see the measure during the 10-week period.

2.3.3. Analysis

Frequencies and descriptive analysis were carried out on completion rates for users who accessed the online community and those who completed the measure. Descriptive statistics and frequency of scores on the three steps of the measure were calculated to understand if items were being selected sufficiently. The aim of the pilot analysis was to see if the POCEM is a measure of community resources that can be used as a proxy for active engagement or helpfulness of a community resource and its content, and the outcomes more frequently selected by users of the measure.

3. Results

3.1. Phase 1: Item generation and reduction

The process of item generation and reduction was done iteratively over four Delphi rounds. A flow chart of the outcomes from each round of the item generation and reduction process are shown in Figure 3.

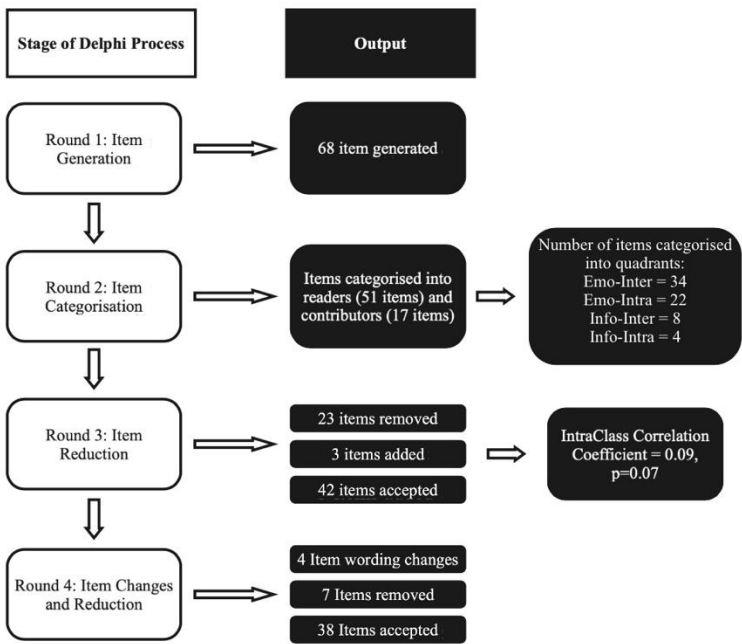


Figure 3. A flow chart of the results from each stage of the Delphi process for the generation and reduction of items.

3.1.1. Round 1

The items initially generated for the online community measure were produced based on the panel members understanding of previous literature on online peer support communities and service’s Theory of Change ³¹. The thematic analysis revealed desirable outcomes for positive behavior change in the online community: 1) Relatedness and Self-Expression 2) Hope and Help Seeking 3) Building a Safe Community 4) Digital Altruism

5) Hope and Help Receiving 6) Making Change. A diagram of the thematic tree (Figure 4) was used to generate the initial pool of 68 items based in these desirable outcomes and its mechanisms from the main themes and sub-themes from the tree (Supplementary Table A1).

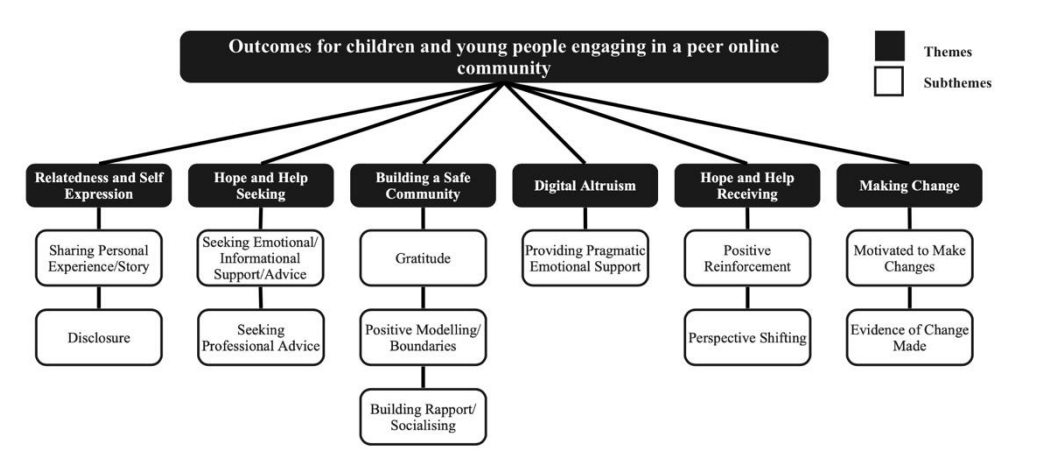


Figure 4. Online community thematic tree on factors of perceived positive behavior change.

3.1.2. Round 2

This round aimed to categorise the initial pool of items based on two criteria. First panel members categorized each item based on the type of community member that will find the item useful. Most of the items (75%; n=51), were classified as relevant to users reading or consuming resources in the community, whilst the remaining 17 (25%; n=25) were relevant for users contributing with content to the community.

The pool items were then categorized into domains of support and used to inform the second part of POCM. The emotional domains were more frequently used to categorize the items in the pool compared with informational domains (Table 2). Through a discussion process, participants agreed on a three-part structure to the measure, with respondents only shown items relevant to the selected domain.

Table 2. Frequency of items categorization into high-level domains (quadrants) of support after Round 1

Wording of quadrant	Domain of high-level support	Frequency of items
‘It helped me to relate to others’	Emotional-Interpersonal	34
‘I feel better about myself’	Emotional-Intrapersonal	22
‘I felt it was important to me’	Informational-Intrapersonal	8
‘I learned some skills I can try with others’	Informational-Interpersonal	4

3.1.3 Round 3

Each panel member voted on the items that they believed should be kept for the measure. An Intraclass correlation (ICC) estimate with 95% confidence intervals was calculated based on three raters, absolute agreement, with a 2-way mixed-effects model.

The inter-rater reliability between panel members was poor at this stage of the item reduction process (ICC =0.09, P=0.07).

In the following workshop the ratings were discussed amongst the panel. The outcome of the workshop was the removal of 23 items, and the addition of three further items. All of the items that were selected by at least two raters (12 items) were kept in the item list.

Two items were added to the domain ‘Informational-Intrapersonal’ for readers, bringing the total items to five in that domain. One last item was added to the domain ‘Informational-Interpersonal’ for readers as community members, increasing the total items in the domain to five too.

3.1.4. Round 4

Following independent and asynchronous evaluation of item wordings, nine suggestions were made regarding wording changes to the items. The six changes that were accepted in the workshop are shown in table 4, including the rationale for each change. In addition to the item changes, the fourth round also resulted in the removal of seven further items. The final 38-items and statements wording were tested in a prototype in the user testing phase.

Table 4: Changes made to the wording of items for the third part of the measure. Changes implemented during the Round 4.

Original Wording	Change in wording
I benefited from feedback from others	I benefited from comments from others
I feel optimistic for the future	I now feel more hopeful
I have implemented a suggestion from someone else	I have done something positive after a suggestion from someone else
I felt comfortable seeking support from my peers	I felt comfortable to share with others
I didn't feel judged by others	I feel good about not being judged
I felt connected	I felt connected to someone

3.2. Phase 2: User testing findings

The main findings collected from the affinity diagram identified issues and recommendations for the POCM. Many participants perceived the measure to be linked with the type of content consumed or accessed-at-the-time by the user inside of the community, being mainly forum posts and its subsequent comments. This is well illustrated by one of the participants quote when prompted in the session to explain what the measure is intending to do [Participant 3]: “How? in my experience was just reading the person and the comments under it”. Most young people reported that they will be more likely to complete the measure if the content of the forum post was helpful for them. The feedback

suggests here may be an agreement bias effect deterring users from providing negative feedback to a peer within the online community, or encouraging users to ignore the measure when the content is not perceived helpful; [Participant 9] said “If it is related to what I am doing, I would fill in the measure, to see similar articles” and [Participant 1] stated that connect with peers will be a key motivation to complete the measure: “...if I had trouble making friends I would say loads (of motivation)”. Most of participants found it normal that the measure will appear in each resource of the community platform, however some of the participants find difficult to find the measure in the platform without facilitators prompt within the sessions. This provided some evidence regarding the measure appearance and initiated changes that may affect measure completion.

In regard to the measure appearance, it was identified that one emoji under the scale had a mismatched emotion, [Participant 10] explained: “The ‘No’ just looks like they’re about to cry or something”, despite the majority appeal to the representation with emojis within the scale, for instance [Participant 1] stated his preference: “The emojis are more neutral not grumpy or red as might give wrong impression to others”. Findings also reveal difficulties from participants understanding who will see their responses. Four participants demonstrated doubts about the information being publicly available for the other peers to see in the community, for instance [Participant 8] showed: “I thought it would instill confidence in the author to write more”. User experience findings around physical appearance of the measure and its instructions led to changes for the live version of the measure taken to the pilot phase (Figure 5). Finally, user testing provided a scenario to review item wordings based on experiences of participants interacting with the prototype during the exercises. Some statements changes are presented based on the rationale given by participants extracted from the affinity diagram. Wording review steer two changes on the second part, and four wording changes in the last part of the measure (Table 5). Overall, user testing allowed identification of appearance issues, validated the focus of construct measurement (it measures the specific community resource), and allowed changes on wording by the intended population for the measure.

Phase 2: User Testing

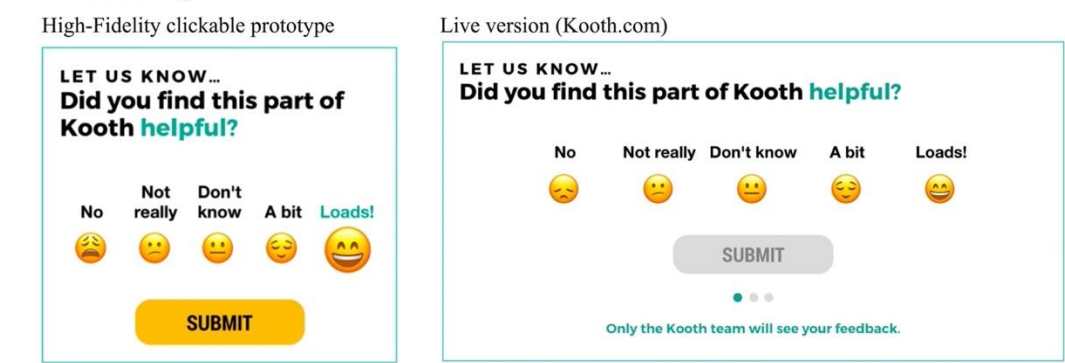


Figure 5. High-fidelity prototype (left) of the measure and its new appearance after user testing (right).

Table 5. User testing affinity diagram findings and changes in wording of the measure

Original state- ment	POCEM	Rationale [Participant]	Changed statement
It feels important to me	Part two	"It means what he said is important to you be- cause you relate to it" [N]	I learned something im- portant to me

I understand myself more	Part two	"Maybe you don't realise how you feel because maybe sometimes you can be sad but you don't know why then you read all this it can make sense to you"[N]	I feel better about myself
I feel safe in this community	Part Three	"Does it mean Kooth as a whole or the outside community?"[N]	I feel safe in the Kooth community
I feel good about now being judged	Part Three	"At first I was like what does it mean?"[N]; "Obviously when someone judges you its negative, but in the app you can feel more better about not being judged"[N]	It feels good not to be judged
I have done something positive after a suggestion of somebody else	Part Three	"I don't think you're gonna do this - I don't get how you're gonna do something positive? I don't think you're gonna come back and do the quiz"	I have learnt enough to make a positive change
I feel excited to support other people with my new found knowledge	Part Three	"It is kind of the same as I've done something positive (Item)..."	Item removed

3.3. Phase 3: pilot study results

The first part of the measure was completed 13,502 times, part two which was only shown to those who had selected responses other than 'don't know' was completed 6,365 times, and the third part, which excluded those who responded 'no' to the part two was completed 3,296 times. The measure was completed by 7,026 participants.

3.3.1. Completion rates

The measure was tested and released iteratively by parts on the platform. The pilot data collection started when the full measure and its three parts were released in the platform. The complete measure was tested between the 11th of December 2019 and the 20th of January 2020, with 2140 unique service users completing a total of 4897 administrations POCM. There was a total of 68439 views of community content on the site by service users who gave research consent during this time, and a total of 2425 contributions in the form of article or discussion posts. Completion rates were divided between readers and contributors to better understand overall completion of the instrument across community members (Table 6).

Table 6. The unique users, POCM completions, and proportion of completions.

Engagement Type	Participants	Number of Completions	Completion rates
Readers	2083	4685	6.85%
Contributors	57	212	8.74%
All	2140	4897	6.91%

3.3.2. Respondent Characteristics

The respondents ages range from 10 to 25, the range allowed in the community and the service. However, five respondents reported an age over 25 and were removed from the dataset, as these will be outliers of the service. The remaining respondents ages ranged from 10 to 25, with a mean of 13.47 (SD= 2.09). Most service users completing the POCEM were female, white, and aged between 10-14 years (Table 7).

Table 7: Demographic characteristics and frequencies of unique users completing one or more POCEMs

Demographic	Frequency
Gender	
Female	72.9%
Male	22.2%
Gender Fluid	3.3%
Agender	1.5%
Age	
10-14	72.6%
15-19	26.0%
20-25	1.4%
Ethnicity	
Asian	6%
Black	3.8%
Mixed	5.1%
White	81.9%
Other	3.2%

3.3.3. POCEM Part 1: Helpfulness of peer community content

The most frequently selected helpfulness score was 5: ‘Lots!’, indicating that the content helped the service user considerably. The frequency of the scores decreased as the helpfulness rating decreased, with the rating of 1: ‘Not really’ selected the least frequently (Figure 6).

Age of service users had a small significant impact on the perceived helpfulness of the community content, with a Kruskal-Wallis test showing a significant effect of age on helpfulness score ($H(2)=7.89, p=.02$). Post-hoc tests (using Dwass-Steel-Critchlow-Fligner pairwise comparisons) were carried out for the three groups and showed service users aged 10-14 gave a significantly ($p=.03$) higher helpfulness score ($M=3.8, SD=1.13$) compared to service users aged 15-19 ($M=3.7, SD=1.5$). There was no difference ($p=0.4$)

between service users aged 10-14 and aged 20-25 ($M=3.4$, $SD=1.43$), or between service users aged 15-19 ($M=3.68$, $SD=1.15$) and aged 20-15 ($p=.80$)

There were no significant differences between the helpfulness scores for service users of different genders ($H(3)=2.4$, $p=.40$), or between different ethnicities ($H(4)=8.4$, $p=.07$)

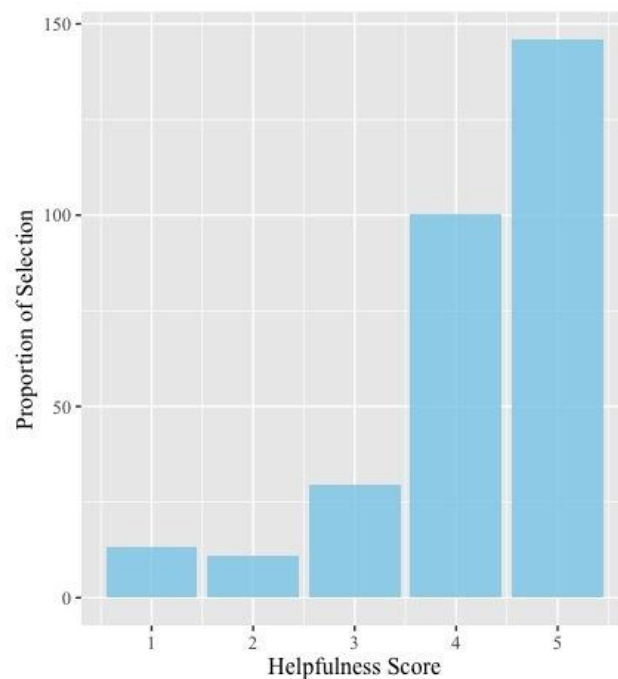


Figure 6: Frequency of selection for the helpfulness scores.

Service users who completed the POCeM after contributing to the community content selected the helpfulness score of 5: 'Lots!' substantially more frequently than service users who read the community content (Figure 8). T-test frequency comparisons reveal a significant difference in the mean helpfulness score ($t(247)=8.8$, $p<.001$) between readers and contributors. Contributors had an average helpfulness score of 4.52, whilst readers had an average helpfulness score of 4.00.

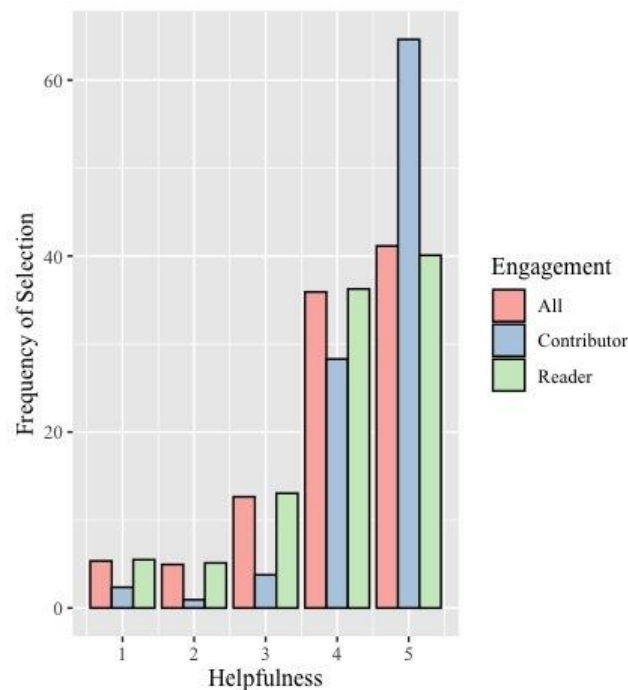


Figure 8. Frequency of selection across the five helpfulness scores, split by engagement

3.3.4. Part 2: Domain selection

Out of the 4897 completions of the measure, 14.2% of responses gave 3: 'Don't know' as the helpfulness score. For this response, the rest of the measure was not shown, and responses ($n=619$) were removed from the domain selection analysis. Out of the remaining 4278 responses with a score of 1, 2, 4, or 5, 10.05% of respondents stopped answering the measure after part 1 providing a helpfulness score, being those incomplete scores of the measure.

The most frequently selected high-level domain of support was 'Help me relate to others', with 55.1% of respondents selecting the option. When the responses were split by whether the helpfulness feedback was positive or negative (Figure 9). More than half (58.2%) of respondents who selected this domain of support gave a positive response selected compared to respondents giving a negative response (32.3%).

Giving 'No response' after the part 1 was most common for respondents giving a negative response in the initial rating, with 34.5% of respondents not selecting a domain (quadrant of support) after giving a negative helpfulness score compared to 6.8% of respondents who gave a positive helpfulness score.

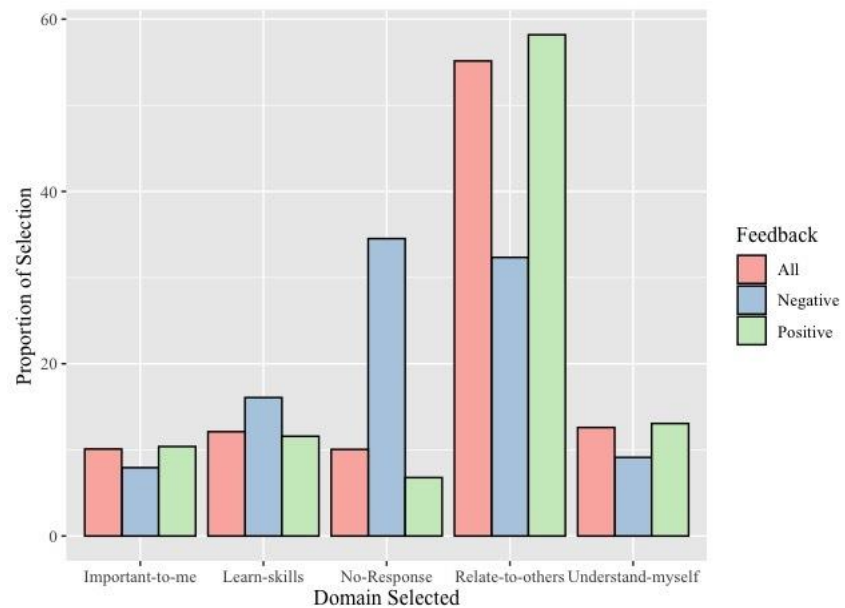


Figure 9. Frequency of selection for high-level support domains in part two of the measure, split by type of feedback

Across all the support domains, respondents who answer only part one on average gave a significantly lower helpfulness score. A Kruskal-Wallis test showed a significant effect of domain selection on helpfulness score ($H(4)=207.45, p<.001$). Post-hoc tests (using Dwass-Steel-Critchlow-Fligner pairwise comparisons) were carried out for the five groups and showed service users who did not give a response in part two gave a significantly lower helpfulness score ($m=3.28$), compared to those who selected the domains 'Important to me' ($M=4.29, p<.001$), 'Learn skills' ($M=4.09, p<.001$), 'Relate to others' ($M=4.31, p<.001$), and 'Understand myself' ($M=4.28, p<.001$) (Table 8). A bar plot of the mean helpfulness scores for the different domains (quadrants) reveal the differences between those who did not respond against the rest of the domains average scores (Figure 10).

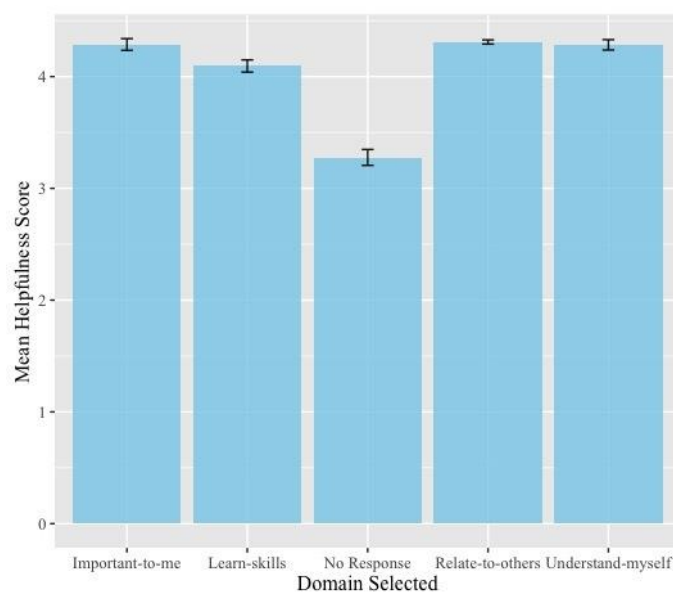


Figure 10: The mean helpfulness scores for each domain selected in part two of the measure. Standard error bars included.

Table 8. Dwass-Steel-Critchlow-Fligner pairwise comparisons for domain helpfulness scores

Item 1	Item 2	W	p
Relate-to-others	No Response	-19.784	< .001
Relate-to-others	Learn-skills	-1.855	.684
Relate-to-others	Understand-myself	1.991	.623
Relate-to-others	Important-to-me	2.278	.491
No Response	Learn-skills	12.733	< .001
No Response	Understand-myself	15.671	< .001
No Response	Important-to-me	15.011	< .001
Learn-skills	Understand-myself	2.683	.319
Learn-skills	Important-to-me	2.889	.246
Understand-myself	Important-to-me	0.397	.999

3.3.5. Part 3: Outcome selection

Based on the response frequencies for the items in the third part of the measure, one item was removed. The item 'I now feel able to ask for support outside of Kooth' was selected the least frequently out of the total items, with only 4.27% of respondents selecting the 'Emotional-Intrapersonal' domain choosing the item. Out of the frequency of selection for all possible items, this item it was selected only for 0.6% of responses (Table 9). When the frequency of the item selection is looked at for each domain, the option of 'No Response' was the most frequently selected for all but three of the domains, where "Relate to others" is the exception, which had the item 'Felt connection' as the most frequently selected item (Figure 11).

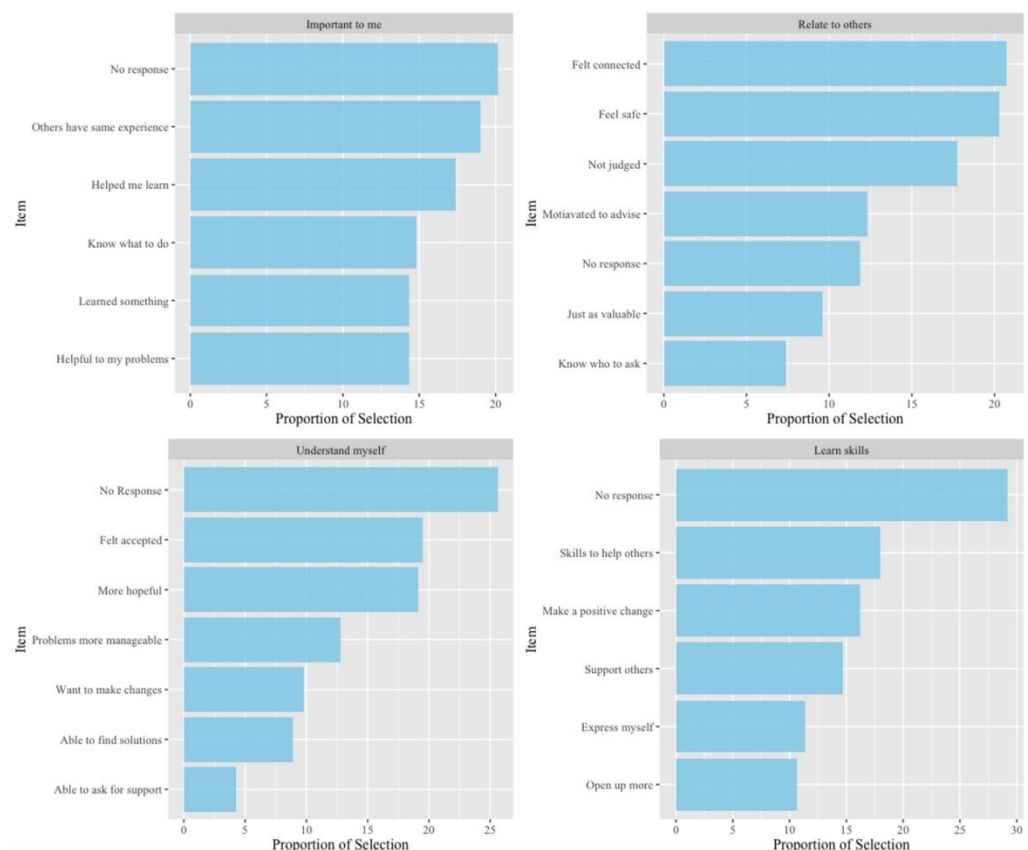


Figure 11: Selection of outcomes for each domain in Part 3: POCeM.

The item stage response with the lowest mean helpfulness score was 'No response', similarly to the helpfulness scores for the other domain options. The analysis was run after removing cases where respondents did not answer the domain section in part 2 ($n=430$). The Kruskal-Wallis test showed a significant effect of item selection on helpfulness score ($H(22)=407$, $p<.001$), and a post-hoc test (Dwass-Steel-Critchlow-Fligner) showed that there was a significant difference in the helpfulness score when respondents selected 'No Response' compared to each other item selection. There were no significant differences between the other items (Supplementary Table A2).

For the item's selection, 'No response' was also the most selected option overall, accounting for 25.38% of the item selections. The increase in the 'No response' selection for the item part of the measure, compared to part 2 of the measure where 10.05% of respondents stopped gave no response, will be in part due to all respondents having the option of not providing a response, compared to the other items that are only shown to a subsection of respondents.

Table 9. Mean helpfulness score, standard deviation, and frequency of selection for each item in Part 3: POCEM.

Item Selected	Domain	Frequency	% Of Total	Mean	SD
No Response	All	656	17.0 %	3.25	1.62
Able-to-ask-for-support	Relate to Others	23	0.6 %	4.57	0.507
Able-to-find-solutions	Relate to Others	48	1.2 %	4.67	0.476
Express-my-self	Relate to Others	59	1.5 %	4.47	0.751
Feel-safe	Relate to Others	479	12.4 %	4.46	0.67
Felt-accepted	Relate to Others	105	2.7 %	4.59	0.6
Felt-connected	Relate to Others	489	12.7 %	4.43	0.617
Helped-me-learn	Understand Myself	75	1.9 %	4.59	0.572
Helpful-to-my-problem	Understand Myself	62	1.6 %	4.66	0.571
Just-as-valuable	Learn Skills	227	5.9 %	4.48	0.64
Know-what-to-do	Learn Skills	64	1.7 %	4.56	0.588
Know-who-to-ask	Important to Me	174	4.5 %	4.5	0.727
Learned-something	Learn Skills	62	1.6 %	4.56	0.59
Make-a-positive-change	Important to Me	84	2.2 %	4.51	0.736
More-hopeful	Understand Myself	103	2.7 %	4.52	0.557
Motivated-to-advise	Important to Me	291	7.6 %	4.5	0.645

Not-judged	Important to Me	419	10.9 %	4.38	0.683
Open-up-more	Important to Me	55	1.4 %	4.64	0.677
Others-have-same-experiences	Learn Skills	82	2.1 %	4.55	0.57
Problems-more-manageable	Learn Skills	69	1.8 %	4.61	0.492
Skills-to-help-others	Understand Myself	93	2.4 %	4.46	0.76
Support-others	Understand Myself	76	2.0 %	4.42	0.753
Want-to-make-changes	Understand Myself	53	1.4 %	4.62	0.489

4. Discussion

In this paper, we outline and discuss a novel multi-phased method for designing an experience measure for young people within an online mental health peer community. We aim to provide a structure for design and testing, whilst reflecting on lessons learnt, to support the future research of other online healthcare platforms and designers. We highlight the value of using mixed methods in an iterative design process when transitioning through phases from initial measure development, refinement, and piloting the measure as well as limitations for each phase.

An experience measure can provide an evaluation of the quality of care received by service users, uncovering key insights into what is and it is not working in a service⁵⁸. By using Donabedian framework^{30,32} as the foundation for measure design, we aimed to assess how the structure, processes, and outcomes within a specific mental health online community are experienced and viewed by young people. An instrument to ascertain outcomes and mechanisms from the online community content reported by users may serve to improve and understand the quality and mechanism seek from engaging with an online community in the future.

The measure was designed with three main stakeholders in mind. First, experts to design the principles of the instrument. Second, young people as main users to provide feedback what they considered important to measure and their perceptions. Third, for practitioners to understand the mechanisms and positive outcomes within the community aspects of the platform. As such, it was important for experts, practitioners, and young people to be involved in the development and testing of the measure for each phase. In the development of the measure itself, service users were consulted through a user testing phase, with much of their feedback influencing the final design. Involving young people across the development and design was essential for ensuring an instrument was produced that accurately reflected their needs within a peer online community, and therefore improve its acceptability⁵⁹.

The use of adapted Estimate-Talk Delphi rounds with academic and clinical experts allowed for a narrowing and improved content of the measurement. The approach was non-standard, with each Delphi round composed of two parts; an initial, independent assessment of the items followed by a group workshop. Unlike systematic approaches to Delphi rounds⁶⁰ the independent and anonymous polling incorporated both qualitative and quantitative feedback. Whilst most Delphi rounds included independent voting, only round 2 required panel members to rate the items. In the other rounds, panel members

provided qualitative feedback. Although rounds may have benefited from a systematic assessment or relevance and clarity of instrument items, like Content Validity Indexes⁶¹, a more dynamic approach to the Delphi rounds provided rich qualitative feedback to influence the iterative design of the measure and enhance the design of the innovation⁶².

A consequence of placing service users, clinicians, and user experience experts at the centre of the design process was of an atypical structure to the measure. This uncommon structure may hinder its generalizability and may not adhere with assumptions required to further investigate the quality of a measurement and its validity in the literature. Feedback in the initial Delphi rounds suggested the three-part structure, with the helpfulness rating placed at the beginning, filtering the other parts of the instrument collecting information about mechanisms and outcomes of that community experience.

User testing is a fundamental step for measure development in online contexts. We recommend replicating the best fidelity prototype possible when conducting user testing research activities, despite low and high-fidelity prototypes have shown similar results when compared⁶³. We observed how young people benefited from structured activities and more realistic objects for the think aloud exercises, which in turn can help to influence changes in the appearance and quality of the workshop outputs and its findings. User testing is a time consuming and resource intense process, and the volume and complexity of data generated may contribute to longer periods of time needed for analysis. The KJ method can provide a good opportunity to analyze and synthesize these testing findings, requiring expertise and focus from researchers to present and facilitate the synthesis of the user testing activities. Affinity maps on the other hand can inform beyond the purpose of the research and provide ideas and improvements with general industrial value (e.g. user needs, product satisfaction). User testing will often use a volunteer or purposive sample, this means that added emphasis should be placed finding participants normally underrepresented. User testing methods like the KJ method may present challenges to integrate quantitative information or systematic agreement methods in the creation affinity maps, but this may help to improve researcher bias in the synthesis stage for this method and phase. User testing represents a new and additional phase for online measure development that provides with invaluable observations about the digital context that can influence the instrument design and its validity⁶⁴.

Pilot study results indicated that service users had a positive experience when accessing the online peer community content, with the helpfulness ratings being positively skewed. These initial results may indicate social desirability or acquiescence bias effects previously shown in digital contexts and scale creation^{65,66}. The potential influence of social desirability was highlighted in the user testing phase, where most of the young people reported that they were more likely to complete the measure if the specific resource was considered helpful, and 4 out of the 11 user testing participants indicated that they believed that after providing feedback to content, such as articles, the authors of the content would be notified of their feedback. To attempt to reduce this perception, qualitative feedback from the user testing phase revealed users' worry about their responses being seen by community members, changes in the instructions and text in the measure were applied in the pilot phase, reinforcing anonymity of responses (Figure 5). The limited disclosure of POCeM responses, along with the anonymity of the service, should help users to not anticipate a social consequence of their responses^{67,68}.

The pilot results showed the support domain 'Relate to others' to be the most frequently selected mechanism for service users providing positive feedback and was the second most selected domain for those with negative feedback too. The average helpfulness score for users selecting the domain was not significantly different to the other support domains, but service users were less likely to drop-out at the item selection stage when selecting this domain. Given that service users who dropped out of the measure early were more likely to give a negative helpfulness score, it is a reasonable assumption that users looking to relate to others through the community were more likely to be satisfied with their experience.

Previous research has illustrated key reasons for young people seeking out support in online communities are to feel less alone with their problems, find a space to talk with peers, and where they feel less likely to be judged⁶⁹. The most frequently selected outcomes in POCEM for positive experiences in the community were 'Felt connection', 'Feel safe', and 'Not judged'. Similar outcomes for a supportive online community have been found previously⁷⁰ and demonstrates how online communities may help users to feel less isolated and more supported around mental health⁷¹. The overall frequency of outcome selection at part three will, in part, be a consequence of the differences in frequency selection of domains at part two of the POCEM. When respondents selected an outcome in part three, rather than dropped out after selecting a domain in part two, there was no significant difference between the average helpful scores for each outcome. The positive average of helpfulness across the sample may accurately reflect a positive experience for service users in relationship to the outcomes selected in the instrument. On the other hand, it may also be a consequence of a ceiling effect in the measure⁷², but this indicates that all outcomes selected to be explored as part of the measure may be relevant experiences in an online community to account for.

There are several limitations to be considered for the development of the POCEM. This study offers insights into considerations that should be made in the early development of a measure for a digital context. By designing the instrument or measure with a specific service in mind, the ability to generalize the existing measure to other online peer communities is limited, the use of experts from the same context may provide a limited view during the Delphi rounds⁷³. It is unclear if these processes and outcomes can be applied to other supportive communities outside the digital service. Future research should look at the transferability of the measure to other communities.

Some of the lessons learnt in the development of the POCEM illustrate the benefits and challenges of designing and testing a measure in a digital environment. The low completion rates of service users starting the measure presents key challenge and threat to the acceptability of the measure by the community. Digital service users reading content in online communities are more likely to be 'lurkers', individuals who will read community content but not actively participate⁷⁴. Therefore, digital environments might be more prone to missing or misleading data around instrument completions and administration, different approaches to data imputation and how missing data is treated can have consequences in psychometric testing and measure performance⁷⁵. Researchers in digital contexts should be aware of these issues, report missing data, and think in advance what psychometric properties the measure should be tested upon.

A phased approach for instrument involving co-design participatory action research with multiple stakeholders can influence the structure and purpose of the measurement. This influence may be counterintuitive for the structure and administration of the instrument and may limit or breach assumptions to test measurement performance in psychological instrument research enhancing difficulties to understand the psychometric properties of the measure.

The next steps are to understand if the measure is a good proximal indicator of active engagement in the community to support mental health, so construct validity of the instrument can be established. The POCEM can help us understand consumption and use of mental health supportive online communities beyond web-based analytics (e.g., how long people read, or contribute, and how frequently they engage) within the service. Further exploration on acceptability and completion rates in relationship with time of engagement, time spent and other content properties is necessary to provide further evidence on what the instrument is intended to measure. Routinely collected information of this instrument in the digital service should help to understand the trends and commonalities deemed helpful in the community, as well as to explore differences across population characteristics so the measure can be evaluated beyond pilot data. Experience measures like POCEM can help services to understand their mechanisms and outcomes more frequently achieved by users of an online community. Online peer communities may use experience measures to understand what resources benefit or hinder the individual, active

engagement of the users with the content can be monitored and better understood beyond digital analytics, so a positive and safe community can be maintained to provide peer support, and other factors that contribute to the well-being of the individual engaging in a digital mental health community.

5. Conclusions

Developing an experience measure for an online community requires a multi-phased systematic process to inform its development and structure involving stakeholders. Different stakeholders can contribute to pieces of information leading to key decisions on the design and development of the instrument, a phased approach with multiple methodologies for appropriate time and stage of development is recommended. Delphi expert rounds and participatory approaches provide rich data that can influence the structure and construct validity of the measure. Further instruments and studies are required to understand psychosocial factors and causal explanations for supportive online communities' outcomes. Measurement of self-reported helpfulness of the community content can serve as an indicator for 'active engagement' and help to understand the main reasons users benefit from these communities and its content. The pilot findings collected information on outcomes achieved by users and is supported by previous literature on online supportive communities highlighting its importance to reduce isolation and enhancing support. Further work to develop acceptable and valid measures for online mental health communities across different contexts is required, so services and practitioners can make better use of these online communities and recommend users by informing about the main outcomes and mechanism a mental health online community should seek for positive behavioral change and recovery promotion in their users.

Supplementary Materials: Table A1 and Table A2 are provided in Appendix A.

Author Contributions: Measure and study Conceptualization, C.M. and A.S.; reviewed methodology, S.O.; analysis and validation, H.B, G.S., C.M.; validation, S.O.; formal analysis, L.M.; investigation, L.M., S.O.; resources, S.O., G.S., LM; data design and curation, L.M.; S.O.; writing—original draft preparation, C.M., L.M. S.O L.S.; writing—review and editing, S.O., L.M; visualization, L.M.; supervision, A.S.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded as part of service improvement and innovation by Kooth plc. There has been no external funding sought to carry out this work. As the funder of the project Kooth plc provided the resource to plan for, collect and analyse data, and write the report for this study.

Institutional Review Board Statement: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and its later amendments. Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements for service quality improvement and innovation. Written informed consent to participate in this study was provided by all the participants' legal guardian/next of kin and anonymity of subjects was always preserved.

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study. All online community anonymous users included in the study consented for their data to be used for research and publication purposes at the time of registering to access the service.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from Kooth plc and are available upon authors and organizational approval at research@kooth.com. Some qualitative field notes may not be available without the permission of Kooth plc and upon a reasonable request, so the privacy and consent of participants is preserved.

Acknowledgments: Tom Kayll for data extraction and database support, Katherine Jackson for her participation and recruitment contribution. Joseph Lloyd for activity safeguarding and note taking in the workshops. The product team at Kooth for implementing the measure. Dr. Terry Hanley for

his independent advice and support and Dr. Lynne Green and Lex Young their teams from the clinical content at Kooth.com

Conflicts of Interest: CM, LM, AS, SG, GS, HB and KJ are researchers employed and receive honorarium by Kooth plc. The funder remained independent it did not influence the design or outcome of the study. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Table A1: Initial pool of items and its mapping to high-level support domains, engagement type, and positive outcomes of online communities

Items	Engagement Type	Domain	Theme
I felt safe to share my problem	Contributor	Emo-inter	Sharing Personal Experience
I felt able to be myself	Reader	Emo-inter	Sharing Personal Experience
I didn't feel judged by others	Reader	Emo-inter	Sharing Personal Experience
I felt able to be honest with myself/others	Contributor	Emo-inter	Sharing Personal Experience
A sense of connection helped me share my personal story	Contributor	Emo-inter	Sharing Personal Experience
I felt able to share something personal with others	Contributor	Emo-inter	Disclosure
I am no longer afraid to share with others	Reader	Emo-inter	Disclosure
I felt comfortable seeking support from my peers	Reader	Emo-inter	Disclosure
I felt I could benefit from support from my peers	Reader	Emo-inter	Disclosure
I was able to say something for the first time today	Contributor	Emo-inter	Disclosure
Others helped me when I asked for help	Contributor	Emo-inter	Asking for Advice (Emotional and Informational)
I know who to ask for help	Reader	Emo-inter	Asking for Advice (Emotional and Informational)
I have influence over what this community is like	Reader	Emo-inter	Positive Modelling / Boundaries
Others have inspired or guided me	Reader	Emo-inter	Positive Modelling / Boundaries
I respect this community	Reader	Emo-inter	Positive Modelling / Boundaries
I feel safe in this community	Reader	Emo-inter	Positive Modelling / Boundaries
I felt connected	Reader	Emo-inter	Building Rapport / Socialising
I was able to connect with others	Reader	Emo-inter	Building Rapport / Socialising
I enjoyed engaging with others	Reader	Emo-inter	Building Rapport / Socialising
I was able to develop a meaningful relationship	Reader	Emo-inter	Building Rapport / Socialising
I have made new friendships	Reader	Emo-inter	Building Rapport / Socialising
I have felt warmth or affection for someone	Reader	Emo-inter	Building Rapport / Socialising
I can trust people in this community	Reader	Emo-inter	Building Rapport / Socialising
It was fun talking with others	Reader	Emo-inter	Building Rapport / Socialising
I now feel like I have helped someone	Contributor	Emo-inter	Digital Altruism

I now feel motivated to give advice to others	Reader	Emo-inter	Digital Altruism
I now feel motivated to learn more so I can help others	Reader	Emo-inter	Digital Altruism
I now feel able to achieve my goals	Reader	Emo-inter	Positive Reinforcement
I benefitted from feedback from others	Reader	Emo-inter	Positive Reinforcement
I now feel empowered by feedback from others	Reader	Emo-inter	Positive Reinforcement
I felt I was just as good as other people	Reader	Emo-inter	Perspective Shifting
It now feels like others believe in me	Reader	Emo-inter	Perspective Shifting
I am now willing to take suggestions from others	Reader	Emo-inter	Motivation for Change
Others have supported me to achieve my goal	Reader	Emo-inter	Evidence of Change Made
Sharing a personal story helped me understand my emotions	Contributor	Emo-intra	Sharing Personal Experience
I felt able to trust others	Contributor	Emo-intra	Disclosure
I felt able to ask for support	Contributor	Emo-intra	Asking for Advice (Emotional and Informational)
I now feel able to ask for support outside of Kooth	Reader	Emo-intra	Asking for Advice (Emotional and Informational)
I felt thankful for the community	Reader	Emo-intra	Gratitude
I now feel less alone	Reader	Emo-intra	Gratitude
Someone liked what I said and it made me feel good	Contributor	Emo-intra	Gratitude
I felt accepted	Reader	Emo-intra	Building Rapport / Socialising
I now feel more competent	Reader	Emo-intra	Positive Reinforcement
I now feel more hopeful	Reader	Emo-intra	Positive Reinforcement
I am now able to find solutions to my problems	Reader	Emo-intra	Positive Reinforcement
I can recognise the positive things I have done	Reader	Emo-intra	Perspective Shifting
I now believe things can improve	Reader	Emo-intra	Perspective Shifting
I feel optimistic for the future	Reader	Emo-intra	Perspective Shifting
My outlook is more positive	Reader	Emo-intra	Perspective Shifting
My problems now feel more manageable and in perspective	Reader	Emo-intra	Perspective Shifting
I have been inspired to make change	Reader	Emo-intra	Motivation for Change
I now want to make changes in my life	Reader	Emo-intra	Motivation for Change
I am now hopeful for change	Reader	Emo-intra	Motivation for Change
I now feel excited for the future	Reader	Emo-intra	Motivation for Change
I am now ready to make change	Reader	Emo-intra	Motivation for Change
I have achieved my goal	Reader	Emo-intra	Evidence of Change Made
I have learned how to express myself	Contributor	Info-inter	Sharing Personal Experience
I have developed skills to open up to others / myself	Reader	Info-inter	Disclosure
I was able to use my experience to help others	Contributor	Info-inter	Positive Modelling / Boundaries
I think I inspired or guided others	Contributor	Info-inter	Positive Modelling / Boundaries
I provided information to help others	Contributor	Info-inter	Digital Altruism
I have learned how to support others	Reader	Info-inter	Digital Altruism

I have the knowledge and skills to help others	Reader	Info-inter	Digital Altruism
I have implemented a suggestion from some-one else	Reader	Info-inter	Evidence of Change Made
I got information that helped me learn about myself	Reader	Info-intra	Asking for Advice (Emotional and Informational)
I learned something new today	Reader	Info-intra	Positive Modelling / Boundaries
Giving advice helped me feel positive about myself	Contributor	Info-intra	Digital Altruism
I now know what I need to do to feel better	Contributor	Info-intra	Positive Reinforcement

Table A2: The pairwise comparisons results for items.

Item 1	Item 2	W	p
able-to-ask-for-sup-port	able-to-find-solutions	1.16602	1.000
able-to-ask-for-sup-port	express-myself	-0.12622	1.000
able-to-ask-for-sup-port	feel-safe	-0.75609	1.000
able-to-ask-for-sup-port	felt-accepted	0.69006	1.000
able-to-ask-for-sup-port	felt-connected	-1.34521	1.000
able-to-ask-for-sup-port	helped-me-learn	0.50868	1.000
able-to-ask-for-sup-port	helpful-to-my-problem	1.46918	1.000
able-to-ask-for-sup-port	just-as-valuable	-0.55931	1.000
able-to-ask-for-sup-port	know-what-to-do	0.25428	1.000
able-to-ask-for-sup-port	know-who-to-ask	0.08608	1.000
able-to-ask-for-sup-port	learned-something	0.28603	1.000
able-to-ask-for-sup-port	make-a-positive-change	0.25600	1.000
able-to-ask-for-sup-port	more-hopeful	-0.31496	1.000
able-to-ask-for-sup-port	motivated-to-advise	-0.26611	1.000
able-to-ask-for-sup-port	not-judged	-1.60366	1.000
able-to-ask-for-sup-port	open-up-more	1.52822	1.000

Item 1	Item 2	W	p
able-to-ask-for-sup- port	others-have-same-experiences	0.03184	1.000
able-to-ask-for-sup- port	problems-more-manageable	0.51798	1.000
able-to-ask-for-sup- port	skills-to-help-others	-0.22505	1.000
able-to-ask-for-sup- port	support-others	-0.72098	1.000
able-to-ask-for-sup- port	want-to-make-changes	0.66100	1.000
able-to-ask-for-sup- port	No Response	-5.48308	0.020
able-to-find-solutions	express-myself	-1.58618	1.000
able-to-find-solutions	feel-safe	-2.88132	0.927
able-to-find-solutions	felt-accepted	-0.74670	1.000
able-to-find-solutions	felt-connected	-3.73231	0.544
able-to-find-solutions	helped-me-learn	-0.91193	1.000
able-to-find-solutions	helpful-to-my-problem	0.33739	1.000
able-to-find-solutions	just-as-valuable	-2.52674	0.982
able-to-find-solutions	know-what-to-do	-1.18678	1.000
able-to-find-solutions	know-who-to-ask	-1.57970	1.000
able-to-find-solutions	learned-something	-1.13685	1.000
able-to-find-solutions	make-a-positive-change	-1.20528	1.000
able-to-find-solutions	more-hopeful	-2.05960	0.999
able-to-find-solutions	motivated-to-advise	-2.16464	0.997
able-to-find-solutions	not-judged	-4.03466	0.378
able-to-find-solutions	open-up-more	0.46465	1.000
able-to-find-solutions	others-have-same-experiences	-1.54511	1.000
able-to-find-solutions	problems-more-manageable	-0.90084	1.000
able-to-find-solutions	skills-to-help-others	-1.83913	1.000
able-to-find-solutions	support-others	-2.41866	0.989
able-to-find-solutions	want-to-make-changes	-0.64918	1.000
able-to-find-solutions	No Response	-8.88934	< .001
express-myself	feel-safe	-0.89359	1.000
express-myself	felt-accepted	1.11123	1.000
express-myself	felt-connected	-1.75366	1.000
express-myself	helped-me-learn	0.83529	1.000
express-myself	helpful-to-my-problem	2.01148	0.999
express-myself	just-as-valuable	-0.58779	1.000

Item 1	Item 2	W	p
express-myself	know-what-to-do	0.49398	1.000
express-myself	know-who-to-ask	0.31794	1.000
express-myself	learned-something	0.53171	1.000
express-myself	make-a-positive-change	0.51077	1.000
express-myself	more-hopeful	-0.23233	1.000
express-myself	motivated-to-advise	-0.17097	1.000
express-myself	not-judged	-2.12753	0.998
express-myself	open-up-more	2.04741	0.999
express-myself	others-have-same-experiences	0.22214	1.000
express-myself	problems-more-manageable	0.84000	1.000
express-myself	skills-to-help-others	-0.12269	1.000
express-myself	support-others	-0.77151	1.000
express-myself	want-to-make-changes	0.98652	1.000
express-myself	No Response	-8.27002	< .001
feel-safe	felt-accepted	2.85871	0.932
feel-safe	felt-connected	-1.85142	1.000
feel-safe	helped-me-learn	2.20691	0.997
feel-safe	helpful-to-my-problem	3.61437	0.611
feel-safe	just-as-valuable	0.48816	1.000
feel-safe	know-what-to-do	1.63667	1.000
feel-safe	know-who-to-ask	1.94825	0.999
feel-safe	learned-something	1.66937	1.000
feel-safe	make-a-positive-change	1.78461	1.000
feel-safe	more-hopeful	0.83311	1.000
feel-safe	motivated-to-advise	1.37220	1.000
feel-safe	not-judged	-2.65356	0.968
feel-safe	open-up-more	3.52852	0.659
feel-safe	others-have-same-experiences	1.39607	1.000
feel-safe	problems-more-manageable	2.17005	0.997
feel-safe	skills-to-help-others	0.90182	1.000
feel-safe	support-others	-0.11159	1.000
feel-safe	want-to-make-changes	2.18974	0.997
feel-safe	No Response	-17.82899	< .001
felt-accepted	felt-connected	-4.01719	0.387
felt-accepted	helped-me-learn	-0.24557	1.000
felt-accepted	helpful-to-my-problem	1.19318	1.000
felt-accepted	just-as-valuable	-2.30651	0.994

Item 1	Item 2	W	p
felt-accepted	know-what-to-do	-0.59789	1.000
felt-accepted	know-who-to-ask	-1.06227	1.000
felt-accepted	learned-something	-0.54229	1.000
felt-accepted	make-a-positive-change	-0.62209	1.000
felt-accepted	more-hopeful	-1.64039	1.000
felt-accepted	motivated-to-advise	-1.83259	1.000
felt-accepted	not-judged	-4.43828	0.201
felt-accepted	open-up-more	1.28604	1.000
felt-accepted	others-have-same-experiences	-1.00872	1.000
felt-accepted	problems-more-manageable	-0.23217	1.000
felt-accepted	skills-to-help-others	-1.41032	1.000
felt-accepted	support-others	-2.12116	0.998
felt-accepted	want-to-make-changes	0.02168	1.000
felt-accepted	No Response	-11.86547	< .001
felt-connected	helped-me-learn	3.21637	0.814
felt-connected	helpful-to-my-problem	4.55426	0.163
felt-connected	just-as-valuable	1.99035	0.999
felt-connected	know-what-to-do	2.57180	0.978
felt-connected	know-who-to-ask	3.30176	0.775
felt-connected	learned-something	2.59158	0.975
felt-connected	make-a-positive-change	2.79915	0.945
felt-connected	more-hopeful	1.97314	0.999
felt-connected	motivated-to-advise	3.01166	0.890
felt-connected	not-judged	-0.92667	1.000
felt-connected	open-up-more	4.40141	0.215
felt-connected	others-have-same-experiences	2.43668	0.988
felt-connected	problems-more-manageable	3.15825	0.838
felt-connected	skills-to-help-others	1.94294	0.999
felt-connected	support-others	0.83304	1.000
felt-connected	want-to-make-changes	3.07149	0.871
felt-connected	No Response	-16.81621	< .001
helped-me-learn	helpful-to-my-problem	1.33819	1.000
helped-me-learn	just-as-valuable	-1.78138	1.000
helped-me-learn	know-what-to-do	-0.34066	1.000
helped-me-learn	know-who-to-ask	-0.69511	1.000
helped-me-learn	learned-something	-0.29129	1.000
helped-me-learn	make-a-positive-change	-0.34912	1.000

Item 1	Item 2	W	p
helped-me-learn	more-hopeful	-1.25917	1.000
helped-me-learn	motivated-to-advise	-1.33910	1.000
helped-me-learn	not-judged	-3.60596	0.616
helped-me-learn	open-up-more	1.41937	1.000
helped-me-learn	others-have-same-experiences	-0.70063	1.000
helped-me-learn	problems-more-manageable	0.01002	1.000
helped-me-learn	skills-to-help-others	-1.06817	1.000
helped-me-learn	support-others	-1.74735	1.000
helped-me-learn	want-to-make-changes	0.23104	1.000
helped-me-learn	No Response	-10.07349	< .001
helpful-to-my-problem	just-as-valuable	-3.17560	0.831
helpful-to-my-problem	know-what-to-do	-1.61051	1.000
helpful-to-my-problem	know-who-to-ask	-2.12085	0.998
helpful-to-my-problem	learned-something	-1.55401	1.000
helpful-to-my-problem	make-a-positive-change	-1.64990	1.000
helpful-to-my-problem	more-hopeful	-2.59450	0.975
helpful-to-my-problem	motivated-to-advise	-2.80196	0.944
helpful-to-my-problem	not-judged	-4.86597	0.087
helpful-to-my-problem	open-up-more	0.15018	1.000
helpful-to-my-problem	others-have-same-experiences	-2.01963	0.999
helpful-to-my-problem	problems-more-manageable	-1.32203	1.000
helpful-to-my-problem	skills-to-help-others	-2.33370	0.993
helpful-to-my-problem	support-others	-2.92344	0.916
helpful-to-my-problem	want-to-make-changes	-1.02833	1.000
helpful-to-my-problem	No Response	-10.19715	< .001
just-as-valuable	know-what-to-do	1.27342	1.000

Item 1	Item 2	W	p
just-as-valuable	know-who-to-ask	1.34460	1.000
just-as-valuable	learned-something	1.31081	1.000
just-as-valuable	make-a-positive-change	1.36788	1.000
just-as-valuable	more-hopeful	0.42951	1.000
just-as-valuable	motivated-to-advise	0.71488	1.000
just-as-valuable	not-judged	-2.64273	0.970
just-as-valuable	open-up-more	3.13018	0.849
just-as-valuable	others-have-same-experiences	0.99689	1.000
just-as-valuable	problems-more-manageable	1.76171	1.000
just-as-valuable	skills-to-help-others	0.52530	1.000
just-as-valuable	support-others	-0.39563	1.000
just-as-valuable	want-to-make-changes	1.83688	1.000
just-as-valuable	No Response	-14.12888	< .001
know-what-to-do	know-who-to-ask	-0.27376	1.000
know-what-to-do	learned-something	0.04441	1.000
know-what-to-do	make-a-positive-change	0.00319	1.000
know-what-to-do	more-hopeful	-0.83337	1.000
know-what-to-do	motivated-to-advise	-0.84381	1.000
know-what-to-do	not-judged	-2.95201	0.908
know-what-to-do	open-up-more	1.67402	1.000
know-what-to-do	others-have-same-experiences	-0.32167	1.000
know-what-to-do	problems-more-manageable	0.34836	1.000
know-what-to-do	skills-to-help-others	-0.67692	1.000
know-what-to-do	support-others	-1.34395	1.000
know-what-to-do	want-to-make-changes	0.53854	1.000
know-what-to-do	No Response	-9.11000	< .001
know-who-to-ask	learned-something	0.32209	1.000
know-who-to-ask	make-a-positive-change	0.29525	1.000
know-who-to-ask	more-hopeful	-0.70965	1.000
know-who-to-ask	motivated-to-advise	-0.76770	1.000
know-who-to-ask	not-judged	-3.82667	0.491
know-who-to-ask	open-up-more	2.14876	0.998
know-who-to-ask	others-have-same-experiences	-0.08747	1.000
know-who-to-ask	problems-more-manageable	0.69530	1.000
know-who-to-ask	skills-to-help-others	-0.53206	1.000
know-who-to-ask	support-others	-1.33435	1.000
know-who-to-ask	want-to-make-changes	0.87227	1.000

Item 1	Item 2	W	p
know-who-to-ask	No Response	-13.58795	< .001
learned-something	make-a-positive-change	-0.04250	1.000
learned-something	more-hopeful	-0.87398	1.000
learned-something	motivated-to-advise	-0.88773	1.000
learned-something	not-judged	-2.96517	0.904
learned-something	open-up-more	1.62042	1.000
learned-something	others-have-same-experiences	-0.36575	1.000
learned-something	problems-more-manageable	0.29928	1.000
learned-something	skills-to-help-others	-0.71619	1.000
learned-something	support-others	-1.37534	1.000
learned-something	want-to-make-changes	0.49146	1.000
learned-something	No Response	-9.01125	< .001
make-a-positive-change	more-hopeful	-0.87486	1.000
make-a-positive-change	motivated-to-advise	-0.91358	1.000
make-a-positive-change	not-judged	-3.20399	0.819
make-a-positive-change	open-up-more	1.70865	1.000
make-a-positive-change	others-have-same-experiences	-0.33727	1.000
make-a-positive-change	problems-more-manageable	0.35592	1.000
make-a-positive-change	skills-to-help-others	-0.71009	1.000
make-a-positive-change	support-others	-1.39635	1.000
make-a-positive-change	want-to-make-changes	0.54733	1.000
make-a-positive-change	No Response	-10.16984	< .001
more-hopeful	motivated-to-advise	0.11720	1.000
more-hopeful	not-judged	-2.46997	0.986
more-hopeful	open-up-more	2.59859	0.975
more-hopeful	others-have-same-experiences	0.53605	1.000
more-hopeful	problems-more-manageable	1.25693	1.000
more-hopeful	skills-to-help-others	0.11474	1.000
more-hopeful	support-others	-0.68037	1.000
more-hopeful	want-to-make-changes	1.38848	1.000

Item 1	Item 2	W	p
more-hopeful	No Response	-10.56530	< .001
motivated-to-advise	not-judged	-3.66487	0.582
motivated-to-advise	open-up-more	2.78140	0.948
motivated-to-advise	others-have-same-experiences	0.51426	1.000
motivated-to-advise	problems-more-manageable	1.32499	1.000
motivated-to-advise	skills-to-help-others	0.02790	1.000
motivated-to-advise	support-others	-0.88596	1.000
motivated-to-advise	want-to-make-changes	1.44418	1.000
motivated-to-advise	No Response	-16.01813	< .001
not-judged	open-up-more	4.68844	0.126
not-judged	others-have-same-experiences	2.86984	0.929
not-judged	problems-more-manageable	3.54168	0.652
not-judged	skills-to-help-others	2.39283	0.991
not-judged	support-others	1.28237	1.000
not-judged	want-to-make-changes	3.41045	0.722
not-judged	No Response	-15.33947	< .001
open-up-more	others-have-same-experiences	-2.06033	0.999
open-up-more	problems-more-manageable	-1.40497	1.000
open-up-more	skills-to-help-others	-2.34877	0.993
open-up-more	support-others	-2.90329	0.921
open-up-more	want-to-make-changes	-1.12335	1.000
open-up-more	No Response	-9.66516	< .001
others-have-same-experiences	problems-more-manageable	0.70516	1.000
others-have-same-experiences	skills-to-help-others	-0.39008	1.000
others-have-same-experiences	support-others	-1.11765	1.000
others-have-same-experiences	want-to-make-changes	0.87757	1.000
others-have-same-experiences	No Response	-9.93819	< .001
problems-more-manageable	skills-to-help-others	-1.06749	1.000
problems-more-manageable	support-others	-1.74347	1.000
problems-more-manageable	want-to-make-changes	0.22097	1.000
problems-more-manageable	No Response	-9.76210	< .001

Item 1	Item 2	W	p
skills-to-help-others	support-others	-0.73298	1.000
skills-to-help-others	want-to-make-changes	1.20252	1.000
skills-to-help-others	No Response	-10.02723	< .001
support-others	want-to-make-changes	1.82525	1.000
support-others	No Response	-8.59153	< .001
want-to-make-changes	No Response	-8.80683	< .001

References

1. Lusk B. Digital natives and social media behavior: an overview. *Prev Res.* 2010;17(S1):3-7. link.gale.com/apps/doc/A246449154/AONE?u=ucl_ttda&sid=googleScholar&xid=6eac2c71. Accessed 9 Dec. 2021. Accessed December 9, 2021.
2. Dooley BA, Fitzgerald A. *My World Survey : National Study of Youth Mental Health in Ireland*. Dublin, Ireland; 2012. <https://researchrepository.ucd.ie/handle/10197/4286>. Accessed December 9, 2021.
3. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: The place of the internet. *Soc Sci Med.* 2005;60(7):1467-1478. doi:10.1016/J.SOCSCIMED.2004.08.010
4. Coleman JS. The Structure of Society and the Nature of Social Research: <http://dx.doi.org/10.1177/107554708000100301>. 2016;1(3):333-350. doi:10.1177/107554708000100301
5. Gowen K, Deschaine M, Gruttadara D, Markey D. Young adults with mental health conditions and social networking websites: Seeking tools to build community. *Psychiatr Rehabil J.* 2012;35(3):245-250. doi:10.2975/35.3.2012.245.250
6. Liu D, Baumeister RF, Yang C, Hu B. Digital Communication Media Use and Psychological Well-Being: A Meta-Analysis. *J Comput Commun.* 2019;24(5):259-273. doi:10.1093/JCMC/ZMZ013
7. Frison E, Eggermont S. Toward an Integrated and Differential Approach to the Relationships Between Loneliness, Different Types of Facebook Use, and Adolescents' Depressed Mood: *Communic Res.* 2015;47(5):701-728. doi:10.1177/0093650215617506
8. O'Keeffe GS, Clarke-Pearson K, Mulligan DA, et al. The Impact of Social Media on Children, Adolescents, and Families. *Pediatrics.* 2011;127(4):800-804. doi:10.1542/PEDS.2011-0054
9. Pantic I, Damjanovic A, Todorovic J, et al. Association between online social networking and depression in high school students: Behavioural physiology viewpoint. *Psychiatr Danub.* 2012;24(1):90-93.
10. Chou HTG, Edge N. "They Are Happier and Having Better Lives than I Am": The Impact of Using Facebook on Perceptions of Others' Lives. *Cyberpsychology, Behav Soc Netw.* 2012;15(2):117-121. doi:10.1089/CYBER.2011.0324
11. de Vries DA, Peter J, de Graaf H, Nikken P. Adolescents' Social Network Site Use, Peer Appearance-Related Feedback, and Body Dissatisfaction: Testing a Mediation Model. *J Youth Adolesc.* 2016;45(1):211-224. doi:10.1007/S10964-015-0266-4/TABLES/2
12. Custers K, Van den Bulck J. Viewership of pro-anorexia websites in seventh, ninth and eleventh graders. *Eur Eat Disord Rev.* 2009;17(3):214-219. doi:10.1002/ERV.910
13. Kvardova N, Machackova H, Smahel D. The Direct and Indirect Effects of Online Social Support, Neuroticism, and Web Content Internalization on the Drive for Thinness among Women Visiting Health-Oriented Websites. *Int J Environ Res Public Heal* 2020, Vol 17, Page 2416. 2020;17(7):2416. doi:10.3390/IJERPH17072416

14. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. *Patient Educ Couns.* 2013;93(3):472-479. doi:10.1016/J.PEC.2013.07.009
15. Setoyama Y, Yamazaki Y, Nakayama K. Comparing support to breast cancer patients from online communities and face-to-face support groups. *Patient Educ Couns.* 2011;85(2):e95-e100. doi:10.1016/J.PEC.2010.11.008
16. Andalibi N, Ozturk P, Forte A. Sensitive Self--disclosures, Responses, and Social Support on Instagram: The Case of #Depression. *Proc 2017 ACM Conf Comput Support Coop Work Soc Comput.* doi:10.1145/2998181
17. Lawlor A, Kirakowski J. Online support groups for mental health: A space for challenging self-stigma or a means of social avoidance? *Comput Human Behav.* 2014;32:152-161. doi:10.1016/J.CHB.2013.11.015
18. Love B, Crook B, Thompson CM, et al. Exploring Psychosocial Support Online: A Content Analysis of Messages in an Adolescent and Young Adult Cancer Community. *Cyberpsychology, Behav Soc Netw.* 2012;15(10):555-559. doi:10.1089/CYBER.2012.0138
19. Saha K, Sharma A. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. *Proc Int AAAI Conf Web Soc Media.* 2020;14:590-601. <https://ojs.aaai.org/index.php/ICWSM/article/view/7326>. Accessed December 10, 2021.
20. Whitlock JL, Powers JL, Eckenrode J. The virtual cutting edge: the internet and adolescent self-injury. *Dev Psychol.* 2006;42(3):407-417. doi:10.1037/0012-1649.42.3.407
21. Horne J, Wiggins S. Doing being 'on the edge': managing the dilemma of being authentically suicidal in an online forum. *Sociol Health Illn.* 2009;31(2):170-184. doi:10.1111/J.1467-9566.2008.01130.X
22. Murray CD, Fox J. Do Internet self-harm discussion groups alleviate or exacerbate self-harming behaviour? *Aust e-Journal Adv Ment Heal* . 2006;5(3):225-233. doi:10.5172/JAMH.5.3.225
23. Ellis LA, Campbell AJ, Sethi S, O'Dea BM. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. *J Cyber Ther Rehabil.* 2011;4(4):461-467. doi:10.2/JQUERY.MIN.JS
24. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. *Arch Psychiatr Nurs.* 2013;27(2):84-89. doi:10.1016/J.APNU.2012.12.005
25. Low KG, Charanasomboon S, Lesser J, et al. Effectiveness of a Computer-Based Interactive Eating Disorders Prevention Program at Long-Term Follow-Up. *J Treat Prev.* 2006;14(1):17-30. doi:10.1080/10640260500403816
26. Jones R, Sharkey S, Ford T, et al. Online discussion forums for young people who self-harm: user views. *Psychiatrist.* 2011;35(10):364-368. doi:10.1192/PB.BP.110.033449
27. Prescott J, Rathbone AL, Brown G. Online peer to peer support: Qualitative analysis of UK and US open mental health Facebook groups. *Digit Heal.* 2020;6:2055207620979209. doi:10.1177/2055207620979209
28. Andy A. Studying How Individuals Who Express the Feeling of Loneliness in an Online Loneliness Forum Communicate in a Nonloneliness Forum: Observational Study. *JMIR Form Res.* 2021;5(7):e28738. doi:10.2196/28738
29. Ray S, Kim SS, Morris JG. The Central Role of Engagement in Online Communities. *Inf Syst Res.* 2014;25(3):528-546. doi:10.1287/ISRE.2014.0525
30. Donabedian A. The Quality of Care: How Can It Be Assessed? *JAMA.* 1988;260(12):1743-1748. doi:10.1001/JAMA.1988.03410120089033
31. Hanley T, Sefi A, Grauberg J, Prescott J, Etchebarne A. A Theory of Change for Web-Based Therapy and Support Services for Children and Young People: Collaborative Qualitative Exploration. *JMIR Pediatr Parent.* 2021;4(1). doi:10.2196/23193
32. Donabedian A. Evaluating the Quality of Medical Care. *Milbank Q.* 2005;83(4):691. doi:10.1111/J.1468-0009.2005.00397.X
33. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv Res.* 2019;54(5):1023-1035. doi:10.1111/1475-6773.13187

34. Bruyneel L, Tambuyzer E, Coeckelberghs E, et al. New Instrument to Measure Hospital Patient Experiences in Flanders. *Int J Environ Res Public Heal* 2017, Vol 14, Page 1319. 2017;14(11):1319. doi:10.3390/IJERPH14111319
35. Krol MW, Boer D de, Delnoij DM, Rademakers JDDJM. The Net Promoter Score – an asset to patient experience surveys? *Health Expect*. 2015;18(6):3099. doi:10.1111/HEX.12297
36. Wang YC, Kraut R, Levine JM. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. *Proc ACM Conf Comput Support Coop Work CSCW*. 2012:833-842. doi:10.1145/2145204.2145329
37. Deetjen U, Powell JA. Informational and emotional elements in online support groups: a Bayesian approach to large-scale content analysis. *J Am Med Informatics Assoc*. 2016;23(3):508-513. doi:10.1093/JAMIA/OCV190
38. Prescott J, Hanley T, Ujhelyi K. Peer Communication in Online Mental Health Forums for Young People: Directional and Nondirectional Support. *JMIR Ment Heal*. 2017;4(3). doi:10.2196/MENTAL.6921
39. Hanley T, Ersahin Z, Sefi A, Hebron J. Comparing Online and Face-to-Face Student Counselling: What Therapeutic Goals Are Identified and What Are the Implications for Educational Providers? *J Psychol Couns Sch*. 2017;27(1):37-54. doi:10.1017/JGC.2016.20
40. Ransom DC, La Guardia JG, Woody EZ, Boyd JL. Interpersonal interactions on online forums addressing eating concerns. *Int J Eat Disord*. 2010;43(2):161-170. doi:10.1002/EAT.20629
41. El Haj HI, Lambrini M, Rais N. Quality of care between Donabedian model and ISO9002V2008. *Int J Qual Res*. 2013;7(1):17-20.
42. Avison DE, Lau F, Myers MD, Nielsen PA. Action research. *Commun ACM*. 1999;42(1):94-97. doi:10.1145/291469.291479
43. MacDonald C. Understanding Participatory Action Research: A qualitative research methodology option. *Can J Action Res*. 2012;13(2):34-50. doi:10.33524/CJAR.V13I2.37
44. Egholm CL, Packness A, Stokholm J, et al. Questionnaire development for the Lolland-Falster Health Study, Denmark: An iterative and incremental process. *BMC Med Res Methodol*. 2020;20(1):1-7. doi:10.1186/S12874-020-00931-1/TABLES/1
45. Gustafson DH, Shukla RK, Delbecq A, Walster GW. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organ Behav Hum Perform*. 1973;9(2):280-291. doi:10.1016/0030-5073(73)90052-4
46. von der Gracht HA. Consensus measurement in Delphi studies. *Technol Forecast Soc Change*. 2012;79(8):1525-1536. doi:10.1016/J.TECHFORE.2012.04.013
47. Keeney S, Hasson F, McKenna HP. *The Delphi Technique in Nursing and Health Research*. 1st ed. (Keeney S, Hasson F, McKenna HP, eds.). New York City, United States: Wiley-Blackwell; 2011. doi:10.1002/9781444392029
48. Wright B, Phillips H, Le Couteur A, et al. Modifying and validating the social responsiveness scale edition 2 for use with deaf children and young people. *PLoS One*. 2020;15(12):e0243162. doi:10.1371/JOURNAL.PONE.0243162
49. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An Assessment Framework for e-Mental Health Apps in Canada: Results of a Modified Delphi Process. *JMIR mHealth uHealth*. 2018;6(7). doi:10.2196/10016
50. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. doi:10.1191/1478088706QP063OA
51. Sketch.com. <https://www.sketch.com/about-us/>. Accessed December 10, 2021.
52. Lookback.io. <https://www.lookback.com/>. Accessed December 10, 2021.
53. Larsen LB, Øvad T, Nielsen L, Larusdottir M. Remote User Testing - Experiences and Trends. In: Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, et al., eds. *INTERACT 2021: Human-Computer Interaction*. Vol 12936 LNCS. Springer, Cham; 2021:579-583. doi:10.1007/978-3-030-85607-6_81
54. Joe J, Chaudhuri S, Le T, Thompson H, Demiris G. The use of think-aloud and instant data analysis in evaluation research: Exemplar and lessons learned. *J Biomed Inform*. 2015;56:284-291. doi:10.1016/J.JBI.2015.06.001

55. Pernice K. Affinity Diagramming: Collaboratively Sort UX Findings & Design Ideas. *Nielsen Norman Gr.* February 2018. <https://www.nngroup.com/articles/affinity-diagram/>. Accessed December 13, 2021.
56. Lucero A. Using Affinity Diagrams to Evaluate Interactive Prototypes. In: *INTERACT 2015: Human-Computer Interaction*. Vol 9297. Springer, Cham; 2015:231-248. doi:10.1007/978-3-319-22668-2_19
57. Hardesty DM, Bearden WO. The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *J Bus Res.* 2004;57(2):98-107. doi:10.1016/S0148-2963(01)00295-8
58. Lavela SL, Gallan AS, Phd G, Phd AS. Issue 1 Inaugural Issue Article 5 2014 Part of the Health and Medical Administration Commons, Health Policy Commons, Health Services Administration Commons, and the Health Services Research Commons. *Patient Exp J.* 2014;1. doi:10.35680/2372-0247.1003
59. Proctor E, Silmere H, Raghavan R, et al. Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Adm Policy Ment Health.* 2011;38(2):65. doi:10.1007/S10488-010-0319-7
60. Fink A, Kosecoff J, Brook RH. *Consensus Methods: Characteristics and Guidelines for Use.*; 1984.
61. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Sastre-Fullana P, Pol-Castañeda S, De Pedro-Gómez JE. Content validity of a health science evidence-based practice questionnaire (HS-EBP) with a web-based modified Delphi approach. *Int J Qual Heal Care.* 2016;28(6):764-773. doi:10.1093/INTQHC/MZW106
62. Zartha Sossa JW, Halal W, Hernandez Zarta R. Delphi method: analysis of rounds, stakeholder and statistical indicators. *Foresight.* 2019;21(5):525-544. doi:10.1108/FS-11-2018-0095/FULL/PDF
63. Walker M, Takayama L, Landay JA. High-Fidelity or Low-Fidelity, Paper or Computer? Choosing Attributes when Testing Web Prototypes. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol 46. SAGE PublicationsSage CA: Los Angeles, CA; 2016:661-665. doi:10.1177/154193120204600513
64. de Ossorno Garcia S, Salhi L, Sefi A, Hanley T. The Session Wants and Need Outcome Measure: The Development of a Brief Outcome Measure for Single-Sessions of Web-Based Support. *Front Psychol.* 2021;12:4900. doi:10.3389/FPSYG.2021.748145/BIBTEX
65. Vesely S, Klöckner CA. Social Desirability in Environmental Psychology Research: Three Meta-Analyses. *Front Psychol.* 2020;11:1395. doi:10.3389/FPSYG.2020.01395/BIBTEX
66. Kuru O, Pasek J. Improving social media measurement in surveys: Avoiding acquiescence bias in Facebook research. *Comput Human Behav.* 2016;57:82-92. doi:10.1016/J.CHB.2015.12.008
67. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. *Behav Res Methods, Instruments, Comput.* 1999;31(3):433-438. doi:10.3758/bf03200723
68. Maryon F. King, Gordon C. Bruner. Social Desirability Bias: A Neglected Aspect of Validity Testing. *Psychol Mark.* 2000;17(2):79-103. doi:https://doi.org/10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0
69. Jones R, Sharkey S, Ford T, et al. Online discussion forums for young people who self-harm: User views. *Psychiatrist.* 2011;35(10):364-368. doi:10.1192/pb.bp.110.033449
70. Coulson NS, Smedley R, Bostock S, et al. The Pros and Cons of Getting Engaged in an Online Social Community Embedded Within Digital Cognitive Behavioral Therapy for Insomnia: Survey Among Users. *J Med Internet Res.* 2016;18(4). doi:10.2196/JMIR.5654
71. White M, Dorman SM. Receiving social support online: implications for health education. *Health Educ Res.* 2001;16(6):693-707. doi:10.1093/HER/16.6.693
72. Gullledge CM, Grace Smith D, Ziedas A, Muh SJ, Moutzouros V, Makhni EC. Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains among shoulder and knee patients undergoing nonoperative and operative treatment. *JBJS Open Access.* 2019;4(4). doi:10.2106/JBJS.OA.19.00015
73. Hsu C-C, Sandford BA. The Delphi Technique: Making Sense of Consensus. *Pract Assessment, Res Eval.* 2007;12:10. doi:10.7275/pdz9-th90

-
74. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: Improving community experiences for everyone. *Comput Human Behav.* 2004;20(2):201-223. doi:10.1016/j.chb.2003.10.015
 75. Xu X, Xia L, Zhang Q, Wu S, Wu M, Liu H. The ability of different imputation methods for missing values in mental measurement questionnaires. *BMC Med Res Methodol.* 2020;20(1). doi:10.1186/S12874-020-00932-0