The Meaning Behind Responses And Changes In The IAPT Outcome Measures: Seeking Perspectives Of IAPT Service Users And Clinicians

Huen Tung Pandora Lee

D.Clin.Psy. thesis (Volume 1), 2020

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: [redacted]

Name: Huen Tung Pandora Lee

Date: 08/07/2020

Overview

The conceptual introduction of this thesis reviewed the existing literature and studies of responses to questionnaire measures of mental health and wellbeing. In the conceptual introduction I found that a range of factors can impact on the response process but also highlighted the lack of a unified theory and approach in this area. The empirical paper details a three-phase qualitative study on service users' and clinicians' views of the meaning behind responses and changes in questionnaires used in the IAPT setting. I identified a range of personal, interpersonal, contextual and questionnaire factors that influenced questionnaire responses. The study supported the acceptability of the Global Rating of Improvement measure and the utility in using this alongside the symptom measures to capture a holistic sense of wellbeing. Clinicians and service users might hold different understanding of the questionnaire scores, and the study emphasised the importance of taking a collaborative approach to interpret responses and changes of questionnaire scores for them to be clinically meaningful and beneficial. I subsequently proposed a conceptual framework of questionnaire response in a clinical setting to illustrate the findings and guide future research in this area. The critical appraisal includes further thoughts and personal reflections on this study and the wider research topic.

Impact Statement

In this thesis I set out to understand people's response to questionnaires used in the Improving Access to Psychological Therapies (IAPT) setting. Both the methods and findings have the potential to impact on future research and clinical practice by adding to and amending the way that questionnaires are discussed with IAPT service users.

The review in this thesis helps to identify the gap of existing knowledge in the study of questionnaire response. It suggests that the current evidence is limited by the lack of unified theory and approach, and urges further study in this area. The design of the current study serves as a model for future similar studies to build on. In particular, it demonstrates the benefit of having a three-stage sequential qualitative design involving both service users and clinicians in providing an in-depth and multidimensional exploration of the topic of interest.

The study puts emphasis on the clinical implications of the findings. Based on the findings, the study offers suggestions on ways to better support service users who might experience difficulties responding to IAPT questionnaires. It also helps to identify issues that are particularly important to clinical practice and risk management (e.g. the underscoring on risk question). A comprehensive summary of the current study alongside recommendations for clinical practice will be shared with the participating services through internal communications. Findings will be shared with the wider IAPT service network through knowledge-exchange platforms such as journal papers, seminars and conferences. The involvement of clinicians in the study also ensures the relevance of the research to clinical practice.

Given the centrality of questionnaire scores on service decisions, findings from this study call for a more cautious approach to interpreting changes in the questionnaire scores due to issues with response difficulties. It suggests the utility of the newly introduced Global Rating of Improvement question, and the possible advantage of using both types of measures to inform service decisions. It demonstrates the need to consider alternative or revised approaches such as computerized adaptive testing to alleviate response burden and to enhance the validity of the measures.

Many service user participants expressed genuine interest in the research topic, and the involvement of their views in this research helped to bring their voice to the IAPT services and wider settings. Findings will be shared with participants who have indicated an interest in receiving an email about the outcome from this study. To see that their views and concerns are being disseminated in the study might make them feel acknowledged and heard. Ultimately, the goal of the study is to use the knowledge generated from the service users to benefit their wellbeing, as well as future service users, by mobilizing changes in research, clinical and service practice.

Table Of Contents

List Of Tables And Figures

Acknowledgements

The completion of this undertaking could not have been possible without the support, encouragement and assistance of so many people whose names may not all be enumerated. Their contributions are sincerely appreciated and gratefully acknowledged. In particular, I would like to express my indebtedness to the following:

My research supervisors, Dr Joshua E J Buckman and Dr Rob Saunders, for their patience and guidance throughout the project.

My course tutor and the Research Director, Dr John King, for his kindness and endless support not only for my thesis but also during my training.

To all the participants who have entrusted me with their valuable insights on the topic.

To Ms Samantha Brown for her contributions to the data analysis process.

To all the staff from the participating services and the UCL DClinPsy Programme and for their administrative and practical support.

Above all, to all my family, friends and my DClinPsy cohort who have been and will always be my greatest source of strength in life.

Part 1: Conceptual Introduction

Understanding Responses To Questionnaires Measuring Mental Health And Wellbeing

Abstract

The growth of interest in making use of mental health and wellbeing measures in healthcare settings such as the Improving Access to Psychological Therapies (IAPT) services urges an exploration of existing knowledge in this area. In this conceptual introduction I performed an article search to identify relevant theories and studies. Based on the search results, it offers an overview of theories related to the understanding of cognitive processes and other factors underlying questionnaire response and the change in response. Evidences related to the dimensions outlined in these models are reviewed under the following subheadings: (a) cognitive processes, (b) respondent characteristics, (c) motivation, (d) task variables, (e) social context, (f) emotions, and (g) assessment of change. Although qualitative findings suggested these aspects could have an impact on questionnaire response, most studies fell short in providing direct support for their hypothesis due to methodological issues and sample size constraints. The lack of unified models and approaches also makes it hard to integrate research findings. Given the limited studies in this area, and that factors influencing response to mental health and wellbeing measures in healthcare setting are likely to be different from those identified in previous studies, this conceptual introduction highlights the need to conduct relevant studies that directly investigate questionnaire response in a naturalistic setting (e.g. in IAPT).

Introduction

Routine data collection and monitoring is a main feature in the Improving Access to Psychological Therapies (IAPT) services. A range of self-reported measures of mental health is completed by IAPT service users on a regular basis. Data generated from these measures serve as benchmarks for service performance, and have implications on service management and commissioning (NCCMH, 2018). However, despite the centrality of the outcome measures in the IAPT services, it is less clear what might underline individual service users' responses to their questionnaires and what factors might influence the response process.

Existing literature and research in questionnaire development have suggested that the ways that respondents make sense of questionnaires are influenced by a range of factors, including the properties of the instrument, their own personal characteristics, their immediate environments and socio-cultural contexts (Johnson, Pennell, Stoop, & Dorer, 2018; Kane, 2013; Schwarz, 2007; Tourangeau, Rips, & Rasinski, 2000). However, most of the previous studies on questionnaire response were conducted during the initial development or adaptation of a measure. Limited research has been done on validated measures that are being applied to clinical settings or being filled out routinely on a sessional basis. It remains unclear if factors that influence questionnaire response in the initial phase of development will continue to exert their effects in later implementation, and if new issues might arise when the questionnaires are used for clinical or service purposes. Insufficient understanding of factors that might influence questionnaire response or the response process might lead researchers, clinicians, and policy makers to interpret questionnaire scores at "face value", draw inaccurate or invalid conclusions, and make decisions that might be irrelevant or even detrimental to the respondent's health (Barkham, Hardy, & Mellor-Clark, 2010; Wolpert, 2013, 2014).

This research project set out to investigate people's response to routine questionnaires used in the IAPT settings, where questionnaire scores play a central role in the delivery and evaluation of the clinical work conducted in the services. Understanding people's response process to questionnaires in this context might help to assess response validity, identify sources of response bias, inform a better practice of questionnaire use and assist joint decision making between clinicians and service users. To lay the groundwork for the research study, the conceptual introduction will review the existing knowledge in the field around people's questionnaire response. Based on the results from the literature search, it will offer an overview of existing theories in the area, followed by an outline of factors that have been identified from the models and theories as influential to questionnaire response. An evaluation of the research evidence in this area is used to identify knowledge gaps, which the empirical paper in this thesis attempts to begin to fill. This conceptual introduction ends with further suggestions for future research.

Background

Factors that influence responses to questionnaires have been extensively investigated in the field of cognitive aspects of survey methodology (CASM). CASM researchers draw on cognitive psychology, statistics and survey design to construct cognitive models of people's response processes. Findings from cognitive interviews, empirical studies, and experimental surveys are used to resolve flaws in questionnaire design, reduce bias, and enhance response accuracy (Schwarz, 2007). While response accuracy in behavioural and factual questions can be established by comparing the answers to some objective measures of the same concepts, such an approach is often not applicable to subjective questions. Since people's thoughts, emotions and attitudes are often context-dependent and unstable (Tourangeau et al., 2000), a range of personal, cognitive and situational influences might obscure people's responses to subjective measures (Krumpal, 2013; Tourangeau & Yan, 2007).

As such, the study of subjective responses calls for a different approach, with the purpose not to rectify the response processes but to understand them as part of the subjective experience. Contemporary CASM research adopts an interpretivist framework and focusses on the phenomenological aspects of questionnaire response. Respondents' unique ways of responding are seen as meaningful in their own right, and the study of factors that influence these responses informs inferences that can be drawn from the data (Chepp & Gray, 2014; Johnson et al., 2018; Kane, 2013; Schwarz, 1995, 2007; Tanur, 1992). Moreover, mental health and wellbeing measures that are widely used in healthcare settings often assess constructs that have high personal significance to the respondents. Thus, the way respondents attribute meaning to questionnaire items can also reveal aspects of their subjective world (Mallinson, 2002).

The interest in people's subjective experience is particularly relevant to two recent developments in healthcare. Namely, the move towards a person-centred approach in healthcare, and the increased use of outcome measures in clinical decision making (NHS England, 2020). Under these frameworks, healthcare services are motivated to actively engage their users in ensuring the development, implementation and interpretation of outcome measures are aligned with service users' personal aims, values, and priorities in life (The Better Care Fund, 2015). Qualitative understandings of the responses and the response process can add richness and meaning to the quantitative data, and situate the service users at the core of clinical practice and policymaking (Miller & Daly, 2013). The next section describes the article search performed on existing literature and studies on responses to questionnaires measuring mental health and wellbeing.

Article Search

Key Terms.

The searches of the existing literature and studies were conducted on the PsycINFO, MEDLINE, and Embase databases on 12th January 2020. The search strategy was refined after initial scoping searches in November 2019; the final searches included the terms surrounding questionnaires and measures (*"questionnaire" "outcome measure", or "survey"*), types of measures or outcomes (*"change", "improve", "recover", "psychology", or "mood"*), response and meaning (*"response", "response style", "meaning", "interpretation", "understanding", "view", or* *“perception”*), known factors and research methods (*“cognitive”, “cognitive factor”, or “cognitive interviewing”*), and user groups (*“respondent”, “service user”, “client”, “clinician”, or “therapist”*). A total of 543 individual articles was returned from three databases.

Inclusion/ Exclusion Criteria

Each article was evaluated in terms of its relevance to the topic of interest. Specifically, articles that help to enrich the understanding of people’s responses to questionnaires measuring mental health and wellbeing were included. Whilst studies that merely provide a descriptive account of the psychometric tests without further exploration of the factors underlying the quantitative findings were excluded. After removal of duplicates and irrelevant articles, this resulted in 59 articles from the searches for inclusion in this extended introduction.

Search Results.

The 59 articles consisted of a mix of theoretical pieces and research papers. The types of research studies included (a) reports on the pretesting phase of questionnaire development, (b) adaptations of existing measures based on language or culture, (c) qualitative validation of established measures, and (d) experimental studies on questionnaire design.

Study Methodology. While some of the questionnaire development studies were purely quantitative, the majority of the studies employed a qualitative or a mixed method approach. The most common research methodology involved the use of statistical modelling to establish psychometric properties of the measures (principal components analyses and item response theory based analyses) and to identify underlying latent constructs (such as factor analyses), in combination with qualitative techniques such as cognitive interviewing and focus groups to understand the quantitative findings. Cognitive interviewing is a technique that aims to elicit people’s cognitive processes underlying their questionnaire response. It can be carried out following a think aloud protocol, where respondents are asked to describe their thinking process as they fill out the measure, or using interview prompts, where respondents are asked to report retrospectively on the response process (Beatty & Willis, 2007).

Type Of Measures. The returned studies included measures on Quality of Life (QOL), Patient-Rated Outcome Measures (PROMs), symptomology (depression, anxiety, and psychotic symptoms), behaviours (risky and sexual behaviours, drug use, and anger management), health (complementary medicine use, and healthcare experience), and other psychological constructs (therapeutic relationship, trust, mindfulness, resilience, and experience of torture).

Following the identification of relevant literatures, the next section introduces theories described in the returned articles.

Overview Of Theories

The cognitive models and theories of questionnaire response provide a conceptual framework to understand what factors might impact on the response process and how these factors might interact to influence questionnaire response. As informed by the search results and their relevance to the understanding of responses to mental health and wellbeing measures, this section will describe (a) the cognitive processes of questionnaire response, (b) the satisficing model, (c) the belief sampling model, (d) the cognitive social-psychological model of question-answering process, (e) the risk and loss hypothesis, (f) mood as information, and (g) response shift.

Cognitive Processes Of Questionnaire Response

Four distinct but partially overlapping components, that is (a) Comprehension, (b) Retrieval, (c) Judgment, and (d) Response, were outlined in the cognitive processes of questionnaire response. It is proposed that, to respond to a survey question, the respondent has to interpret from the wording and other cues of what is being asked of them, retrieve information that they think might be relevant, make judgments based on the available cognitions, and map their internal answer to the response options provided by the question. However, respondents might employ these processes in different order or omit some stages on some occasions. Factors that influence any of the four components would have an impact on the response process and outcome (Fienberg, Loftus, & Tanur, 1985; Tanur, 1992; Tourangeau et al., 2000).

The Satisficing Model

The thorough execution of all stages of cognitive processes outlined above would allow a respondent to arrive at a high-quality and accurate answer, but at a cost of significant energy and diligence from the respondent (Krosnick, Alwin, & Cannell, 1987). Alternatively, people might respond in a less than ideal manner by satisficing, where one or more of the cognitive processes are compromised or completely omitted, and the respondent might formulate their answer based on their beliefs, feelings or other heuristic cues (Krosnick, 2018; Krosnick, Narayan, & Smith, 1996). The satisficing model (Krosnick et al., 1987) suggested that the likelihood of satisficing would depend on three factors: task difficulty; respondent's ability; and motivation. Thus, the respondent mightl be more likely to satisfice if the task difficulty is high, their ability is compromised, or their motivation is low.

Belief Sampling Model

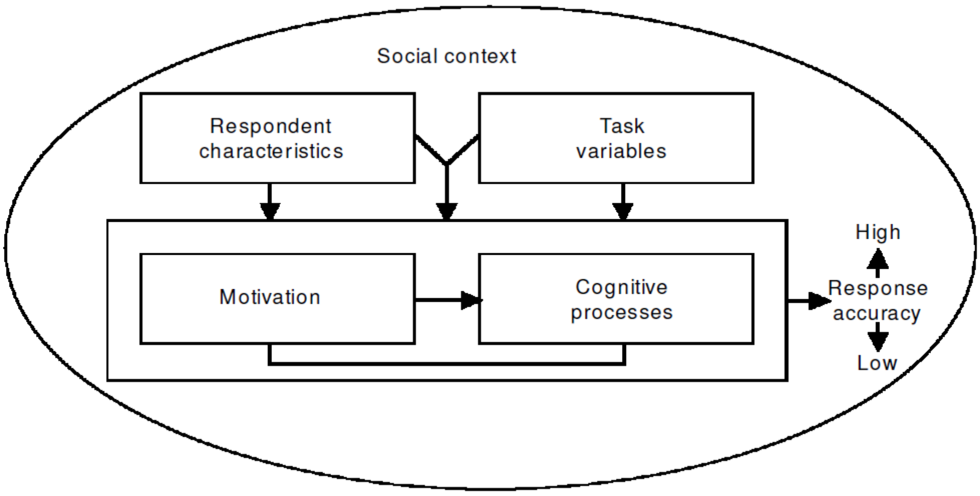
The belief sampling model (Tourangeau & Rasinski, 1988) was originally constructed to conceptualise context effect on attitude questions, and has been applied to understand response to self-reported well-being measures (Diener, 1994). The model postulated that, respondents construct their answers based on a sample of considerations accessible at the time of responding. Considerations refer to beliefs, feelings, experience, memories and knowledge that are relevant to the question. Some of the considerations are more chronologically stable; while others might be more temporarily constructed. Contextual factors would impact on the accessibility and the evaluative process of the retrieved cognitions, thus contributing to response instability.

The Cognitive Social-Psychological Model Of Question-Answering Process

Drawing on existing theories on questionnaire response (including the ones outlined above) and studies on measures of alcohol and drug use, the cognitive social-psychological model of question-answering process (Babor, Brown, & Del Boca, 1990; Del Boca & Noll, 2000) was developed to understand factors underlying questionnaires that measures addictive behaviours. Perhaps due to the sensitive nature of the topic, this model puts emphasis on the social context, in which four inter-related components of the response process are embedded (Figure 1; Babor et al., 1990; Del Boca & Darkes, 2003; Del Boca & Noll, 2000). This model proposed that response accuracy is a result of interaction between (a) respondent characteristics, (b) task variables, (c) motivation and (d) cognitive processes.

Figure 1

Schematic Illustration Of The Cognitive Social-Psychological Model Of Question-Answering Process (Del Boca & Noll, 2000)



The Risk And Loss Hypothesis

The acceptability of a question depends on its intrusiveness within a given cultural context, respondent's perceived threat of disclosure, and social desirability (Tourangeau & Yan, 2007). The risk and loss hypothesis suggested that respondents' motivation to report truthfully would be contingent on their evaluation of the potential consequences of disclosure (the loss) and the likelihood of it happening (the risk) (Rasinski, Baldwin, Willis, & Jobe, 1994), as well as thier assessment of the potential gains and losses associated with accurate reporting (Willis, Sirken, & Nathan, 1994). In addition, social desirability, as defined both a personal characteristic of approval seeking and impression management (Crowne & Marlowe, 1964), and the social acceptability of the question topic or the response (Krumpal, 2013), were suggested to play a central role in determining truthful responding on sensitive questions.

Mood As Information

Cognitive models such as the belief sampling model (Tourangeau & Rasinski, 1988) considered mood and emotions as contextual factors that could induce response bias (Rothman & Schwarz, 1998; Schwarz, Bless, et al., 1991). Taking a different perspective, the mood as information proposition suggested that mood and emotions could be useful in the judgment process (Clore & Wyer Jr, 2001; Schwarz & Clore, 2003). Specifically, it suggested that affect could serve to direct our attention to relevant areas of judgment (Wyer & Carlston, 1979) and allow us to adopt the most appropriate processing model based on the state of our environment (Bless & Schwarz, 1999; Mackie, Asuncion, & Rosselli, 1992).

Response Shift

The use of routine measures to monitor change rests on the assumption that people's responses and their interpretations of the constructs remain constant over time. However, research on health and quality of life (QOL) measures (Andrykowski & Hunt, 1993; Tilton, Bach, & Wang, 1993) suggested that, changes in questionnaire response over time could be a result of the interaction of four components, (a) the change in health status, (b) the predisposed individual characteristics, (c) the behavioural, cognitive and affective processes involved in accommodating the changes in health status, and (d) response shift (Sajobi, Brahmbatt, Lix, Zumbo, & Sawatzky, 2018). Response shift, as defined by a change in the way of responding, can occur when the processing of personal experience (such as life events, education, treatment) leads to a change in internal standards (recalibration), values (reprioritization) and the idiosyncratic definition of the construct (reconceptualization) (Sprangers & Schwartz, 1999).

Interim Summary

The cognitive processes of questionnaire response (Tourangeau et al., 2000) outlined the cognitive components that are involved in response to measures. The statisficing model (Krosnick et al., 1987), the belief-sampling model (Tourangeau & Rasinski, 1988), the risk and loss hypothesis (Rasinski et al., 1994) and the mood as information proposition (Schwarz & Clore, 2003) offer suggestions of how the various cognitive components might operate under the influences of various internal and external factors. The cognitive social-psychological model of question-answering process (Babor et al., 1990) can be seen as an attempt in taking a broader perspective to encompass factors that were mentioned in the other models. Although this model specifically concerned measures of alcohol and drug use, it could have the potential to be generalised to other mental health and wellbeing measures that might be considered as sensitive or intrusive. An additional dimension of questionnaire response is the measure of changes over time and the issue with response shift (Sprangers & Schwartz, 1999). The next section will draw on results from the literature search to review evidence around factors outlined in these models and theories.

Review Of Articles

Based on the theories outlined above and the returned articles from the literature search, this section will be structured under the following subheadings: (a) cognitive processes, (b) respondent characteristics, (c) motivation, (d) task variables and (e) social context (f) the role of emotions, and (g) assessment of change.

Cognitive Processes

The cognitive processes of questionnaire response (Tourangeau et al., 2000) was developed based on the literatures and evidence from psycholinguistic theories (Grice, 1975), memory studies (Bartlett, 1932; Brewer, 1996), decision making (Tversky & Kahneman, 1974) and judgment (Parducci, 1965). This is the most commonly referred conceptualisation amongst the returned articles from our literature search. However, the majority of the studies did not specifically investigate the nature of cognitive processes underlying questionnaire response. Instead, these cognitive processes were mostly studied as part of the overall aim to address response difficulties in newly developed or existing measures through a qualitative or mixed method design.

Issues related to comprehension (Brodey, Addington, et al., 2018; Brodey, Purcell, et al., 2018), retrieval (Wiklund, Holmstrom, Stoker, Wyrwich, & Devine, 2013), judgment (Godderis, Adair, & Brager, 2009; Paterson, 2004) and response mapping (Sentell & Ratcliff-Baird, 2003) were identified in these studies, and were linked to response errors in the measure. For instance, using cognitive interviewing techniques, a validation study of a newly developed self-report screening instrument showed that, religious individuals might interpret items asking about religious hallucination as referring to their religious belief, thus causing a false high base-rate on those items for this group of respondents (Brodey, Addington, et al., 2018; Brodey, Purcell, et al., 2018). Cognitive interviews also helped to identify that issues with long recall period (Wiklund et al., 2013) and uncertainty of standard of comparison (Paterson, 2004) can lead to recall bias or variation in the way of responding amongst participants.

The use of mixed method approaches allows the use of quantitative data to validate some of these qualitative findings through convergent results. For example, a validation study of a state mindfulness measure showed that, items that were identified to yield comprehension difficulties from cognitive interviews also had lower factor loadings on the questionnaire construct (Ullrich-French, Cox, Cole, Rhoades Cooper, & Gotch, 2017). A mixed method study using interview and statistical modelling also suggested that, the application of item analysis using Rasch modelling could help to identify issues (i.e. patient's trust in the case) that might not be apparent from interview data due to the complexity of the questionnaire construct (David, Hitchcock, Ragan, Brooks, & Starkey, 2018) .

Studies described above were predominantly pre-tests of newly developed measures, with the aim to identify and rectify response errors. While validation studies conducted with well-established measures of clinical symptomology can have important implication for existing clinical and research practice. In a validation study of the revised Anxiety Sensitive Index (ASI; Reiss, Peterson, Gursky, & McNally, 1986), researchers found from their cognitive interviews that, most respondents answered questions that asked about their beliefs and cognitions around anxiety based on their actual experience, rather than their cognitive appraisal of what might hypothetically happen as intended by the measure (Brown, Hawkes, & Tata, 2009). Based on their findings, the researchers suggested that responses to the ASI were more likely to reflect another aspect of trait anxiety rather than the cognitive appraisal of anxiety. Thus, they urged a reconsideration of theories of anxiety that were developed based on studies using this measure. Response difficulties have also been identified with measures of depression (Godderis et al., 2009; Sentell & Ratcliff-Baird, 2003). A cognitive interviewing study with patients in primary care settings showed that, people might interpret the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001; Malpass et al., 2016), a measure of depressive symptoms, in a diverse manner. Respondents from their study also had difficulties mapping their internal answers to the response options.

In line with the cognitive processes of questionnaire response (Tourangeau et al., 2000), results from the articles showed that factors affecting any of the comprehension, memory, judgment and response processes would have an impact on questionnaire response. It suggested the utility of integrating item-level quantitative analysis to identify problematic question and qualitative methodology to understand the cause of the issues. However, the studies cited above that adopted this specific type of mixed method design (David et al., 2018; Ullrich-French et al., 2017) had a sample size of twelve and fifteen. Although this would be a reasonable size for cognitive interviews, it might limit the conclusion that could be drawn from quantitative data in these studies. Validation studies of well-established measures showed that issues with cognitive processes continue to exert influences after the initial pre-testing period and highlighted the importance of conducting validation studies in subsequent implementation of a measure. Inattention to the response process might lead to a misinterpretation of questionnaire responses, as demonstrated in the case of the ASI measure (Brown et al., 2009).

Respondent Characteristics

The returned articles showed that personal characteristics such as age, literacy level, personality and psychological states could have an influence on the cognitive processes. These factors were studied by correlating personal dimensions with response difficulties experienced by respondents in relation to a specific measure (e.g. Sentell & Ratcliff-Baird, 2003; Ullrich-French et al., 2017) or by investigating response difficulties encountered by specific groups of intended respondents (e.g. Iezzoni et al., 2017; Magwood, Jenkins, & Zapka, 2009).

Studies showed that respondents with a lower education or cognitive level, or those who are unfamiliar with the psychological terminology, might struggle to understand the concepts described in the questionnaire (Iezzoni et al., 2017; Magwood et al., 2009; Sentell & Ratcliff-Baird, 2003). A validation study of the Beck Depression Inventory (Beck, Steer, & Brown, 1996) demonstrated a significant correlation between the literacy level and the amount of comprehension difficulties experienced by adult learners, senior respondents and oncology patients. The effect held even after other factors such as age and education level were controlled (Sentell & Ratcliff-Baird, 2003). In this study, comprehension of the questionnaire was assessed by the percentage of items that a participant was able to correctly paraphrase as judged by a set of standardised criteria. In another validation study of a diabetes health-related quality-of-life instrument amongst fifteen older African Americans, the type and amount of response difficulties were identified by performing content analysis on cognitive interview data. By examining characteristics of the five respondents who experienced the most response difficulties, the researchers concluded thatm response difficulties were associated with being older, having a lower education level and a longer history of ill-health (Magwood et al., 2009). The influence of age was examined in a validation study of a State Mindfulness Scale for Physical Activity. In this study, they divided the respondents into groups below and above ten years old, and based on the observed difference between the two groups, they suggested that those who were ten years old or older achieved a higher comprehension rate of the questionnaire items (Ullrich-French et al., 2017). However, apart from the first study (Sentell & Ratcliff-Baird, 2003), the other two studies (Magwood et al., 2009; Ullrich-French et al., 2017) did not perform any quantitative analysis on the observed differences to support their suggestion, thus undermining the strength of this evidence.

Personal characteristics such as self-esteem can cause bias in the judgment process of subjective measures. A study asked recently unemployed adults to fill out standardised measures on (a) social support, (b) job-seeking and (c) depression at baseline and four months later, as well as questions on their perceived changed on these aspects (e.g. “how much decrease or increase has there been in how upset you feel about things?”) at the four-month time point (Foster & Caplan, 1994). Data on self-esteem, judgment stability and negative mood were also collected as potential sources of response bias. Using multiple regression on the data, researchers concluded that people with higher self-esteem were more likely to under-estimate negative changes and over-estimate positive changes, a pattern indicative of an optimistic response bias. Negative mood and judgment instability also tended to amplify the response bias. Apart from self-esteem, other aspects such as how much the respondents enjoy effortful cognitive tasks can impact on the degree to which they engage in all cognitive processes involved in the questionnaire response (Haugtvedt & Petty, 1992).

In addition, questionnaire studies with people with addiction problems (Brown, Kranzler, & Del Boca, 1992; Del Boca & Darkes, 2003) and homeless persons with mental disorders (Adair et al., 2012) suggested thatm psychological disturbance such as acute mental health conditions or substance use can have an impact on people's memory and ability to process information, thus leading to response difficulties. However, these studies were mostly based on qualitative findings. Without the use of control group and quantitative evidence, it is hard to establish any association between response difficulties and the proposed personal factors.

Studies on respondent characteristics were often conducted as part of a validation study of the measures with the intended respondents, thus the study design was mostly driven by the presumption researchers had about what types of response difficulties their respondents might experience (e.g. investigation of comprehension difficulties amongst older or younger respondents). Two studies have used multiple regression to demonstrate the link between response difficulties or response patterns with respondents' literacy level and self-esteem (Foster & Caplan, 1994; Sentell & Ratcliff-Baird, 2003), while others did not perform any statistical comparison to support their conclusions (Magwood et al., 2009; Ullrich-French et al., 2017). Overall, the existing findings were inconclusive due to methodological constraints.

Motivation

The satisficing model outlined motivation as one of the factors that impacted on the response strategies people employ in answering a questionnaire, alongside task difficulties and the person's abilities (Krosnick, 1991). While early evidence for this proposition mostly came from experimental survey studies involving socio-political attitude questions or cognitive task, one of the studies was conducted using a health survey (Rothman & Schwarz, 1998). In this study, participants were asked to give three or eight examples of health behaviours that might increase or decrease risk of heart disease, followed by rating of their own risk to the condition. It was found that participants who reported having a family history of heart disease were motivated to engage in the cognitive process, and therefore less susceptible to give a response based on heuristic such as the ease of recall (Rothman & Schwarz, 1998).

The motivation to give a truthful response might be determined by factors such as socio-cultural context, risk and loss of disclosure, social desirability and personal traits (Crowne & Marlowe, 1964; Del Boca & Noll, 2000; Johnson et al., 2018; Krumpal, 2013; Rasinski et al., 1994; Tourangeau & Yan, 2007). This could have particular relevance to routine outcome measures in clinical settings due to their sensitive or intrusive nature. A review of research in this area summarised the following as factors affecting truthful reporting on sensitive topics: the mode of administration; the data collection setting; the presence of others; and the question wordings (Tourangeau & Yan, 2007). The reviewers highlighted that it was not the mere presence of others, but the perceived risk and loss of the surveyor and others in the environment knowing the answer, that determined the likelihood of a truthful response. In line with this, another study also showed that, while the presence of an interviewer might hamper truthful responding, if the privacy of the setting was ensured, participants were more likely to self-disclose when the measures were administered by an interviewer (Rasinski, Willis, Baldwin, Yeh, & Lee, 1999). Research on people's considerations when responding to sensitive questions suggested that, the current confidentiality assurances practice may fail to address respondent's concerns. In particular, respondents were found to have concerns about revealing private information to the surveyor, to people that might be handling the data, or to those present in the immediate environment (e.g. other patients in the clinic) (Rasinski et al., 1999). These considerations would have important implication for data collection practice of routine measures in clinical settings.

Issues with acceptability of the questions or measures were often explored as part of the validation study through cognitive interviewing or other qualitative methods. For instance, in an interview study regarding the use of mental health measures with homeless people, the respondents commented that it was not only the content of the question, but also the way that the questions were being asked that would affect how forthcoming they could be in their responses (Adair et al., 2012). In a validation study of a measure of delinquency and risky behaviours amongst Indian youth, participants suggested that the role of question order (i.e. placing questions about strength and positive aspects before questions about delinquency and risky behaviours) could facilitate disclosure (Pavkov, Priest, & Fox, 2012). These studies demonstrated the value in engaging respondents' views in addressing issues with truthful responding through questionnaire design and delivery.

Other studies using cognitive interviewing techniques have revealed issues with truthful responding in clinical settings. In a qualitative validation of the Edinburgh Postnatal Depression questionnaire, questions such as "the thought of harming myself has occurred to me" could easily evoke feelings of shame and guilt amongst patients in the maternity ward. Respondents were found to express hesitation in answering the question due to concerns about the potential implication it might have on their care, and the consequences on their rights to care for their babies (Godderis et al., 2009). A similar finding has also been established in a primary care setting, where patients expressed hesitation about the suicidal risk question in the self-reported depression measure (Malpass et al., 2016).

As demonstrated, motivation can be related to various factors ranging from personal ability, task difficulties to question order (Haugtvedt & Petty, 1992; Krosnick, 1991; Narayan, 1995; Wicker, Park, McCann, & Hamman, 1995). In particular, the motivation to respond truthfully is highlighted for questions or measures that might be considered as intrusive or sensitive, such as around mental health and addiction (Del Boca & Noll, 2000; Tourangeau & Yan, 2007). While the majority of these studies were not conducted in a healthcare setting, some preliminary evidence was obtained in the studies with hospital patients (Godderis et al., 2009) and with primary care patients (Malpass et al., 2016). The clinical relevance highlights the importance of further research in this area.

Task Variables

Task variables can refer to questionnaire features, the mode of administration and task difficulties. Numerous CASM research has shown that changes in question wordings, response options, question order, and presentation format could have a direct impact on people's responses to the questionnaire (e.g. Huybers, 2016; Parducci, 1965; Schwarz, 1990; Schwarz & Hippler, 1987; Schwarz, Strack, Hippler, & Bishop, 1991; Spörrle, Gerber-Braun, & Försterling, 2007; Tourangeau, Couper, & Conrad, 2004). These variables were usually studied by giving participants different versions of the survey or by asking them to fill out the survey under different experimental conditions. The impact of the manipulation was established by the comparison of response generated from different experimental conditions.

Questionnaire features might also interact with personal characteristics to influence the response process. For example, positivity bias might influence how numbers on a scale are perceived and lead to the overuse of the positive end of the scale (Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991; Sears, 1983). This effect is particularly prominent on performance scales, which might also be explained by people's general assumption of positive relations with others and their wish to present themselves in a positive light (Zajonc, 1968). This could have implication for self-reported measures or questionnaires used in healthcare settings, as respondents' judgment of their own mental health and wellbeing might be impacted by their overall self-perception and by their existing relationship with their clinicians.

Another aspect to the processing of a questionnaire item is the cognitive effort it involves (c.f. task difficulties in the Satisficing Model; Krosnick, 1991; Krosnick et al., 1987; Krosnick et al., 1996). Processing can vary depending on the level of similarity between a given item and those adjacent to it in the questionnaire (Hamby, Ickes, & Babcock, 2016). For example, having a statement followed by one that is written in an opposite direction (e.g. "I feel that I have a number of good qualities" followed by "All in all, I am inclined to feel that I am a failure") was found to induce a higher misreporting rate in the second item (Ickes et al., 2019), presumably due to the higher cognitive effort involved in interpreting the statement. Similarly, text features that increase cognitive burden (such as the use of uncommon words and complex syntax) were found to increase the response time and increase the tendency to satisfice (Lenzner, Kaczmirek, & Lenzner, 2010). The cognitive ability might also be impacted by the item position in the questionnaire. Specifically, respondents might be more likely to misreport on the last items of a length or cognitively intensive questionnaire compared to the first few items (Krosnick et al., 1987; Narayan, 1995). Although increasing the number of response categories might help respondents to distinguish nuances within their answers, it might not be helpful if respondents are overloaded by the number of response options and lose their ability to meaningfully distinguish between response categories (Miller, 1956). In support of this, a study on an anger measure has shown that, reducing the number of items and response categories could help to improve data quality while retaining the psychometric properties of the measure (Hawthorne, Mouthaan, Forbes, & Novaco, 2006).

In addition to the studies outlined above, many of the aforementioned studies (e.g. Brodey, Addington, et al., 2018; Brodey, Purcell, et al., 2018; Magwood et al., 2009; Sentell & Ratcliff-Baird, 2003; Wiklund et al., 2013) showed that issues people have with cognitive processes of questionnaire responses were related to questionnaire features such as questionnaire wordings and response categories.

Social Context

According to the cognitive social-psychological model of question-answering process (Figure 1; Babor et al., 1990), all of the above factors are embedded in a social context. The psycholinguistic view on questionnaire response suggested that, people’s response behaviours can be understood as an act of communication (Schwarz, 1995). Thus, respondents make use of questionnaire features and other contextual cues to make implicit assumptions respondents hold about the surveyors' intent, which consequentially impact on their questionnaire response (Schober, 1999; Schwarz, 1995; Tanur, 1992).

Studies that look into the relationship between the role of surveyors or survey organisation and response outcome might offer insight into the social context of response process. In an experimental study, participants were told that the surveyor was either a social scientist or a personality psychologist, and were issued a questionnaire that asked for their opinions about the reasons for mass-murder (Norenzayan & Schwarz, 1999). It was found that respondents' gave more social or personality variables according to the role of the surveyor. Based on this finding, researchers concluded that respondents' answers were given based on the perceived interest of the surveyors or the research intent. However, it could also be that the role of surveyors served as a prime and made relevant constructs more available to the respondents. The impact of setting or organisation affiliated with the survey is coined "sponsorship effect". Studies that examined sponsorship effect found little support in terms of the overall differences in response patterns (Allen & Roberts, 2016; Crabtree, Kern, & Pietryka, 2018; Etter, Perneger, & Rougemont, 1996). However, there is some evidence that the effect can be observed on an item level (Gibson & Mayhall, 2019) and in the willingness to engage with the task (Boulianne, Klofstad, & Basson, 2011). In one of these studies, ex-patients from a hospital were issued with the Herth Hope Index (Herth, 1992) to assess their mental wellbeing after discharge. In one condition, participants were told that the chaplaincy was involved in the study, and in the control condition they were told that it was conducted by the hospital. Although no mean difference was found in the questionnaire scores in both conditions, results from principal component analysis showed that the factor structure of the chaplaincy condition deviated more than the control condition from what might be expected from the measure. This has led the researchers to conclude the presence of a sponsorship effect. In addition, they also noticed that a greater variation in the response to the question ("I have a faith that gives me comfort") in the chaplaincy condition than in the control. Based on their hypothesis, they suggested that some respondents in the chaplaincy condition might have interpreted the word "faith" as referring to religious belief based on the role of the survey sponsor, thus leading to a greater variation in response (Gibson & Mayhall, 2019). However, with the lack of supports from qualitative investigation in this study, their hypothesis about what might have underlied these observation could remain tentative.

The effect of socio-cultural factors is the main focus of the cultural adaptation of questionnaires. Culture can be seen as an enduring system of meaning, a set of implicit rules that governs people's understanding, decision making and personal values (Heise, 2010) and a framework for people to organise their personal experience (Lee, Liu, & Hu, 2017; Nisbett, Peng, Choi, & Norenzayan, 2001). Differences in the conceptual framework might raise the question about comparability of questionnaire findings across cultures (Johnson et al., 2018). As seen in the validation studies of translated measures of health, certain words and concepts might not have an equivalent translation in another language or culture (Gomutbutra et al., 2012; Scorza et al., 2013). The perception of mental health and the acceptability of a measure can vary hugely across culture (Kaiser et al., 2019). Even within the same society, meanings of certain constructs might also change over time. To address this issue, researchers of the questionnaire on complementary medicine use adopted an iterative design with regular reviews to ensure changes in the cultural perception of relevant practices were being integrated into the measure (Stussman, Bethell, Gray, & Nahin, 2013). In this particular study, the questionnaire was used to elicit information rather than to quantify and measure changes. It is unclear how this practice can be translated to routine outcome measures, and how to address issues with comparability if the same measure has to be changed over time.

Overall, validation studies of culturally adapted measures demonstrated the impact social context could have on the way respondents understand and respond to questionnaire measures. Studies on sponsorship effect suggested the possible influence of the survey context on the questionnaire response. However, with the limited evidence, the findings remain inconclusive. It would be of interest to further investigate whether the service context, relationship with clinicians or presence of others in the clinics might influence response to measures used in clinical setting.

Emotions

The role of emotion in questionnaire response is featured in the belief-sampling model (Tourangeau & Rasinski, 1988) and mood as information (Clore & Wyer Jr, 2001; Schwarz & Clore, 2003). These theories draw on studies on memory and decision making to support their propositions.

Based on the observation that most mental health questionnaires rely on people's self-report of frequencies of behavioural activities and emotional experience, one of the studies was set out to investigate the recall accuracy and strategies for both types of experience (Brown, Williams, Barker, & Galambos, 2007). Their study first asked participants to record the frequency of behaviours and emotions in a web-based diary, and subsequently asked them to provide an estimate of each types of experience at a later time point and described the strategies they used to derive these estimates. Their results showed that while estimates of behaviours and emotions were equally accurate, respondents were more likely to recall and count individual events to generate the behavioural frequencies, but would use what their self-knowledge about their average emotional state to infer emotional estimates.

The role of emotions in questionnaire response has also been explored in a conceptual discussion paper on issues that might arise in designing a survey to assess the experience of torture. Researchers carefully considered issues such as the tendency for traumatic memories to be repressed or not encoded, the prevalence of cognitive deficits amongst torture survivors, and denials of torture experiences due to lack of trust (Willis & Gonzalez, 1998).

The previous section on motivation has explored issues with truthful responding to sensitive questions, but what has yet to be explored is the range of emotions that can be evoked in the face of the perceived threat in the question or the measure. Qualitative studies have suggested that the emotional reaction usually precedes the cognitive evaluation of the issue, thus further highlighting the intricacy of issues around personal sensitivity, and the importance of creating a safe and private environment in order to elicit truthful responses (Andreenkova & Javeline, 2018). The feeling of embarrassment can lead people to disguise their true answers in order to avoid offending the interviewer by a socially unacceptable response (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996)

The finding that different cognitive strategies were employed in retrieving behavioural and emotional information could have implication in questionnaire design. Specifically, it raised the question of whether frequency is the most appropriate way to measure emotional experience. Questions or measures that tap into sensitive topics or traumatic experience could have the potential to evoke strong emotions, which calls for attention when considering the clinical use of questionnaires and the impact it could have on the respondents.

Assessment of Change

Changes can be assessed by comparing questionnaire scores between time points (e.g. pre-post treatment) or by asking respondents about their subjective view on progress (e.g. the global rating of improvement question). Response shift is a concept that changes observed in questionnaire response could be due to the change in internal standards, values and conceptualization of the construct (Sprangers & Schwartz, 1999).

Early studies on response shift used a particular type of research method called then-test study. In conventional pre-post study design, participants are usually asked to fill out measures at baseline and at a second time point (e.g. pre- and post-treatment) based on their concurrent states. In then-test studies, participants also fill out an additional measure at the second time point that asks them to retrospectively report how they were at baseline. The difference between the concurrent and retrospective baseline scores is taken as support for response shift (Sajobi et al., 2018). To illustrate, in a then-test study on quality of life measures among HIV patients undergoing intensive treatment, respondents were asked to fill out retrospective and concurrent measures of their subjective state of wellbeing at baseline and post-treatment, as well as an assessment of change question that asked them to rate on a seven-point scale ranging from feeling “much worse” to “much better” (Nieuwkerk, Tollenaar, Oort, & Sprangers, 2007). Researchers then correlated physiological markers of recovery (such as CD4-cell count, plasma viral-load, body mass index and haemoglobin concentration) with the following: (a) the difference between concurrent pre-post treatment scores, (b) the difference between retrospective baseline and concurrent post-treatment scores and (c) the assessment of change response. A significantly greater association was found between physiological markers of recovery and the difference in the retrospective baseline and concurrent post-treatment scores. Although this could be taken as support for the response shift, as the researchers pointed out, their research methodology could not rule out other reasons such as recall bias, social desirability bias and cognitive dissonance reduction (Nieuwkerk et al., 2007). The difficulty to eliminate alternative explanations has been the main criticism of using this type of then-test study design to investigate response shift (Sajobi et al., 2017).

The advancement of statistical procedures has opened up a new realm of investigation through statistical modelling. Structural Equation Modelling and Item Response Theory have enabled researchers to identify changes in response patterns on a statistical level (Schwartz, Finkelstein, & Rapkin, 2017). For instance, by applying Confirmatory Factor Analysis to pre- and post-treatment scores from the Beck Depression Inventory (Beck et al., 1996), it was found that people who have undergone psychotherapy demonstrated an overestimation of depressive symptomatology, smaller measurement errors, and higher correlations between constructs of depression (Fokkema, Smits, Kelderman, & Cuijpers, 2013). The researchers related these statistical findings to the psychoeducational elements of the Cognitive Behaviour Therapy treatment the participants received, and suggested that the results indicated higher self-awareness of their symptoms and more within-group agreement about the understanding of their conditions following therapy. Such results could offer another way to analyse and interpret the data (e.g. to infer changes in psychological understanding) beyond symptom severity.

While then-test studies and statistical modelling can merely provide indirect evidence for response shift, questionnaires such as the use of Quality-of-Life Appraisal Profile (Rapkin & Schwartz, 2004) have been developed to capture individual's appraisal processes underlying responses to QOL measures. It was suggested that the understanding of response process to the questionnaire measure can shed light on individual's psychological adaptation to significant events in life (Fokkema et al., 2013; Sajobi et al., 2018). This view is particular relevant in the understanding of questionnaire response of attendees of psychological treatments.

Response shift can also be applied to understand the minimal clinically important difference (MCID), which is defined as the smallest difference in outcome measure scores that constitute a clinically meaningful outcome for the service users (McGlothlin & Lewis, 2014). This can be established by anchoring changes in a clinical outcome measure (e.g. symptom measure of depression) with a subjective measure that asks about people's perception of change over time (e.g. the global rating of improvement) (Button et al., 2015). Responsiveness to clinically important change is subject to contextual influences such as the characteristics of the target group, the treatment type, the timing of data collection, the construct being investigated (Beaton, Boers, & Wells, 2002) and the direction of change (i.e. positive changes tend to be valued more than decline of the same degree) (Cella, Hahn, & Dineen, 2002). Without the understanding of the appraisal process behind the scores on the global rating of change (GRC), the interpretation of change might be inconclusive. For example, using the concept of response shift, the disproportional perception of positive and negative change in GRC can be understood as an indication of the recalibration process (Schwartz, Finkelstein, et al., 2017).

One of the studies used logistic modelling to map out the response patterns of respondents that showed discrepancy between GRC and a range of patient-rated measures of health and wellbeing (e.g. RAND-12 Health Status Inventory; Johnson & Maddigan, 2004) and the association with their cognitive appraisal processes assessed by the appraisal profile survey (Schwartz, Powell, &

Rapkin, 2017). Their results showed that GRC were more associated with people's appraisal patterns than with changes in their health measures. They also identified from the appraisal profile that, those who reported feeling worse on GRC than on the health measure tended to focus on illness experience and define their QOL in terms of their illness. They were also more likely to compare themselves to the doctor's prediction and gave greater emphasis to big changes in their life. Discrepancy between GRC and symptom measures of depression have also been explored in primary care setting using semi-structured interviews (Robinson et al., 2017). It was found that, recent life events, perception of the condition, motivation, and recall difficulties were some of reasons for the mismatch.

The ability of making use of questionnaires to capture changes in people's mental health and wellbeing is crucial to the routine use of outcome measures in clinical setting. This section reviewed the concept of response shift and other statistical and research development in the assessment of change using questionnaire. Aspects such as reconceptualization and psychological adaptation to life challenges will have significant impact on how changes in measures of mental health and wellbeing can be understood. In particular, studies that investigated people's appraisal process through profiling measure or qualitative interviews suggested that people's response to a measure can indeed change over time due to different perceptions of their conditions, standard of comparisons and construct. These studies also offered valuable insight into how people make sense of their mental health and wellbeing, aspects that are clinically meaningful and important for individual's recovery.

Discussion

This conceptual introduction set out to review existing theories and literature in the field in order to understand people's response to questionnaires measuring mental health and wellbeing, especially those that are used routinely in clinical settings (e.g. IAPT services). However, despite the attempts made to select studies and theories that are most relevant to the topic of interest, the majority of research in this field and consequently the majority of examples included in this extended introduction are not of measures of mental health and wellbeing used in healthcare settings. Therefore the conclusions drawn from the studies are mostly tentative in nature and are used to highlight directions for future research.

Overall, the existing literature and studies reviewed in this extended introduction provide rich resources in informing the thinking about responses to measures of mental health and wellbeing used routinely in healthcare settings. A range of cognitive, motivational, psychological, questionnaire, and social-cultural factors are all likely to play a role in questionnaire responses. Each of these factors will have their specific relevance depending on the setting, the measures, and the target group of respondents. Mental health and wellbeing measures in healthcare settings have generally focussed on assessing people's emotional experiences and gathering their views on their own progress, both of which could be sensitive in nature. People's emotional experiences are encoded and processed differently from behavioural information, and they have the potential to reveal important aspects of the respondents' view about themselves. While current mood might affect the judgment process, it can also serve an informative function. The assessment of change is more than a matter of recalling relevant information, it reflects the way respondents make sense of their personal experience. In healthcare, concepts of response shift might relate to people's belief about the treatment process and their own recovery. Finally, the issue of sensitive questions is one that needs careful consideration. Different from questionnaire developers, clinicians have a duty of care to their patients when it comes to issues of risk (e.g. suicidality). Inherently, the use of psychological measures in healthcare settings that assess personally sensitive issues might carry a risk that respondent's data can be misinterpreted or misused to make judgments and clinical decisions about the patient, heightening the importance of accurate responding.

Research and theories in CASM offer their unique perspectives on questionnaire response by drawing on psycholinguistic theories and cognitive psychology. A selection of theories have been reviewed in this extended introduction. Although most of the theories were not developed specifically for measures of mental health and wellbeing, models such as cognitive processes of questionnaire response (Tourangeau et al., 2000) have been applied to understand questionnaire response on symptom measures of depression and anxiety. The cognitive social-psychological model of question-answering process (Babor et al., 1990; Del Boca & Noll, 2000) was developed to understand factors underlying response to measures of addictive behaviours, and could potentially be generalisable to other mental health and wellbeing measures. However, the article search did not identify any studies that directly investigated the components and the structure of these models. Most of the review articles can only provide indirect support for the cognitive processes. The satisficing model (Krosnick, 1991), belief sampling model (Diener, 1994; Tourangeau & Rasinski, 1988), risk and loss hypothesis (Rasinski et al., 1994) and mood as information (Schwarz & Clore, 2003) offer suggestions of the mechanisms of how different factors such as questionnaire features, motivation, cognitive ability, emotions and beliefs might interact to derive the questionnaire response. These models provide a direction for hypotheses to be developed and tested in experimental or correlation studies in future.

The reviewed articles demonstrated the utility of using a mixed method approach to understand questionnaire response. For example, by using item-level analysis to identify issues with specific questions on a measure and using qualitative interviews to understand the response process. Although qualitative methods such as cognitive interviewing can offer valuable insights into the understanding of questionnaire response, they are time consuming and costly. In addition, there is no unified approach to conduct these interviews, and to analyse and report data generated from this approach, making it difficult to integrate findings from the field. However, the development of qualitative surveys such as the quality-of-life appraisal profile (Rapkin, Ryniker, Morgenstern, & Schwartz, 2011; Rapkin & Schwartz, 2004) can provide a convenient way to capture the aspects of people’s thinking process behind their questionnaire response in a large sample size (e.g. Schwartz, Powell, et al., 2017).

Limitations

The major limitation of the existing studies is that most of them were not specifically set out to understand questionnaire response, this subsequently limited the conclusions that can be drawn from the literature reviewed in this conceptual introduction. For most of the studies, the focus was on validating a questionnaire measure, therefore there is a huge variation in the way factors and processes underlying questionnaire response were being studied and reported in the reviewed literature. Some of the findings are highly specific to a particular measure or a context of application, thus limiting the generalisability to other settings. The majority of the studies were only conducted with a targeted respondent group without a control group or any randomised allocation process, making it hard to establish whether response difficulties are associated with certain characteristics of the respondent group, and whether these differences will induce a systematic response bias. Such aspects would have to be investigated using quantitative methods such as correlation studies or statistical modelling. In addition, the sample size in these studies varied hugely depending on the research methodology, ranging from twelve in a mixed method study using cognitive interviewing and Rasch modelling (David et al., 2018) to 858 in a secondary data analysis of longitudinal data (Schwartz, Finkelstein, et al., 2017). While the sample size of twelve might be sufficient for cognitive interviews, it limited the inference that can be drawn from the quantitative data in the mixed method studies.

It is also noteworthy that the major limitation to this conceptual introduction was that it was not a full systematic review. Although 59 studies were identified, it might be the case that important studies were missed or that some relevant findings from the included studies were missed. For instance, with a less structured approach to the article search, other possible search terms such as (e.g. “measure”, “inventory”, “rating scale”, “perspective”, “patient”) were not used. This might have resulted in further relevant papers being overlooked in this study. In addition, systematic reviews are typically conducted by teams of reviewers co-reviewing in order to avoid such occurrences. However, such methods were not appropriate for this extended introduction.

Future Directions

The theories and ideas drawn from CASM and cognitive studies of questionnaire response present ample possibilities in gaining insight into the response to psychological measures in healthcare settings. Given the limitations, future studies would benefit from using research methodology such as cognitive interviewing to decipher factors underlying people's questionnaire response in a naturalistic setting such as Improving Access to Psychological Therapies (IAPT) services, where outcome measures are routinely used. In addition, research and studies included in this essay have only focussed on the response process, it has not explored how the response is being interpreted by the respondents themselves, or by the clinicians, researchers and policy makers. It will also be interesting to find out if different stakeholders might have the same or different understanding of what might underlie questionnaire responses. The answer to this research question will have important implications in the translation of measure outcomes to clinical or policy decisions that would be meaningful to the respondent.

References

Adair, C. E., Holland, A. C., Patterson, M. L., Mason, K. S., Goering, P. N., Hwang, S. W., & Home, A. (2012). Cognitive interviewing methods for questionnaire pre-testing in homeless persons with mental disorders. *Journal of Urban Health, 89*(1), 36-52.

Allen, P., & Roberts, L. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. *First Monday, 21*(2).

Andreenkova, A., & Javeline, D. (2018). Sensitive questions in comparative surveys. *Advances in comparative survey methods: Multicultural, multinational and multiregional contexts (3MC) NJ: John Wiley & Sons, Inc*.

Andrykowski, M. A., & Hunt, J. W. (1993). Positive psychosocial adjustment in potential bone marrow transplant recipients: Cancer as a psychosocial transition. *Psycho-Oncology, 2*(4), 261-276.

Babor, T. F., Brown, J., & Del Boca, F. K. (1990). Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Behavioral Assessment*.

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*: John Wiley & Sons.

Bartlett, F. (1932). C.(1932). Remembering. *Cambridge: CambridgeUniversityPress*.

Beaton, D. E., Boers, M., & Wells, G. A. (2002). Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current opinion in rheumatology, 14*(2), 109-114.

Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly, 71*(2), 287-311. doi:10.1093/poq/nfm006

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78*(2), 490-498.

Bless, H., & Schwarz, N. (1999). Sufficient and necessary conditions in dual-mode models: The case of mood and information processing.

Boulianne, S., Klofstad, C. A., & Basson, D. (2011). Sponsor prominence and responses patterns to an online survey. *International Journal of Public Opinion Research, 23*(1), 79-87.

Brewer, W. F. (1996). What is recollective memory. *Remembering our past: Studies in autobiographical memory*, 19-66.

Brodey, B. B., Addington, J., First, M. B., Perkins, D. O., Woods, S. W., Walker, E. F., . . . Putz, J. (2018). The Early Psychosis Screener (EPS): item development and qualitative validation. *Schizophrenia research, 197*, 504-508.

Brodey, B. B., Purcell, S. E., Rhea, K., Maier, P., First, M., Zweede, L., . . . Brodey, I. S. (2018). Rapid and accurate behavioral health diagnostic screening: initial validation study of a web-based, self-report tool (the SAGE-SR). *Journal of medical Internet research, 20*(3), e108.

Brown, G. P., Hawkes, N. C., & Tata, P. (2009). Construct validity and vulnerability to anxiety: A cognitive interviewing study of the revised Anxiety Sensitivity Index. *Journal of anxiety disorders, 23*(7), 942-949.

Brown, J., Kranzler, H. R., & Del Boca, F. K. (1992). Self-reports by alcohol and drug abuse inpatients: factors affecting reliability and validity. *British Journal of Addiction, 87*(7), 1013-1024.

Brown, N. R., Williams, R. L., Barker, E. T., & Galambos, N. L. (2007). Estimating frequencies of emotions and actions: a web-based diary study. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*(2), 259-276.

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., . . . Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychol Med, 45*(15), 3269-3279. doi:10.1017/S0033291715001270

Cella, D., Hahn, E. A., & Dineen, K. (2002). Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Quality of life research, 11*(3), 207-221.

Chepp, V., & Gray, C. (2014). Foundations and new directions. *Cognitive interviewing methodology*, 7-14.

Clore, G., & Wyer Jr, R. (2001). Affective feelings as feedback: Some cognitive consequences. Theories of mood and cognition: A user's handbook. LL Martain and GC Clore. In: Mahwah, NJ, Lawrence Erlbaum Associates.

Crabtree, C., Kern, H. L., & Pietryka, M. T. (2018). Sponsorship effects in online surveys.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*: New York: Wiley.

David, S. L., Hitchcock, J. H., Ragan, B., Brooks, G., & Starkey, C. (2018). Mixing interviews and Rasch modeling: Demonstrating a procedure used to develop an instrument that measures trust. *Journal of Mixed Methods Research, 12*(1), 75-94.

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction, 98*, 1-12.

Del Boca, F. K., & Noll, J. A. (2000). Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction, 95*(11s3), 347-360.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology, 70*(5), 979.

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research, 31*(2), 103-157.

Etter, J.-F., Perneger, T. V., & Rougemont, A. (1996). Does sponsorship matter in patient satisfaction surveys? A randomized trial. *Medical care*, 327-335.

Fienberg, S. E., Loftus, E. F., & Tanur, J. M. (1985). Cognitive aspects of health survey methodology: an overview. *The Milbank Memorial Fund Quarterly. Health and Society*, 547-564.

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment, 25*(2), 520.

Foster, D. A., & Caplan, R. D. (1994). Cognitive influences on perceived change in social support, motivation, and symptoms of depression. *Applied Cognitive Psychology, 8*(2), 123-139.

Gibson, C. B., & Mayhall, T. B. (2019). Comprehension Context and Sponsor Effects in a Hospital Mental Health Study. *Sociological Methods & Research, 48*(3), 677-697.

Godderis, R., Adair, C. E., & Brager, N. (2009). Applying new techniques to an old ally: A qualitative validation study of the Edinburgh Postnatal Depression Scale. *Women and Birth, 22*(1), 17-23.

Gomutbutra, P., Aramrat, A., Sattapansri, W., Chutima, S., Tooprakai, D., Sakarinkul, P., & Sangkhasilapin, Y. (2012). Reliability and validity of a Thai version of assessment of chronic illness care (ACIC). *Journal of the Medical Association of Thailand, 95*(8), 1105.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58): Brill.

Hamby, T., Ickes, W., & Babcock, M. (2016). Evidence for context switching in the effects of average item length and item-length variability on internal consistency. *Journal of personality assessment, 98*(5), 491-502.

Haugtvedt, C. P., & Petty, R. E. (1992). Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of personality and social psychology, 63*(2), 308.

Hawthorne, G., Mouthaan, J., Forbes, D., & Novaco, R. W. (2006). Response categories and anger measurement: do fewer categories result in poorer measurement?: development of the DAR5. *Soc Psychiatry Psychiatr Epidemiol, 41*(2), 164-172. doi:10.1007/s00127-005-0986-y

Heise, D. R. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*: John Wiley & Sons.

Herth, K. (1992). Abbreviated instrument to measure hope: development and psychometric evaluation. *Journal of advanced nursing, 17*(10), 1251-1259.

Huybers, T. (2016). Exploring the use of best-worst scaling to elicit course experience questionnaire responses. *Assessment & Evaluation in Higher Education, 42*(8), 1306-1318. doi:10.1080/02602938.2016.1270256

Ickes, W., Babcock, M., Hamby, T., Park, A., Robinson, R., & Taylor, W. (2019). Side Streets and U-Turns: Effects of Context Switching, Direction Switching, and Factor Switching on Interitem Correlations and Misresponse Rates. *J Pers Assess, 101*(3), 326-339. doi:10.1080/00223891.2018.1450262

Iezzoni, L. I., Matulewicz, H., Marsella, S. A., Warsett, K. S., Heaphy, D., & Donelan, K. (2017). Collaborative design of a health care experience survey for persons with disability. *Disability and health journal, 10*(2), 231-239.

Johnson, J. A., & Maddigan, S. L. (2004). Performance of the RAND-12 and SF-12 summary scores in type 2 diabetes. *Quality of life research, 13*(2), 449-456.

Johnson, T. P., Pennell, B.-E., Stoop, I. A., & Dorer, B. (2018). *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*: John Wiley & Sons.

Kaiser, B. N., Ticao, C., Anoje, C., Minto, J., Boglosa, J., & Kohrt, B. (2019). Adapting culturally appropriate mental health screening tools for use among conflict-affected and other vulnerable adolescents in Nigeria. *Global Mental Health, 6*.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Krosnick, J. A. (2018). Questionnaire design. In *The Palgrave handbook of survey research* (pp. 439-455): Springer.

Krosnick, J. A., Alwin, D. F., & Cannell, C. (1987). Satisficing: A strategy for dealing with the demands of survey questions.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation, 1996*(70), 29-44.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47*(4), 2025-2047.

Lee, S., Liu, M., & Hu, M. (2017). Relationship between Future Time Orientation and Item Nonresponse on Subjective Probability Questions: A Cross-Cultural Analysis. *J Cross Cult Psychol, 48*(5), 698-717. doi:10.1177/0022022117698572

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology, 24*(7), 1003-1020.

Mackie, D. M., Asuncion, A. G., & Rosselli, F. (1992). The impact of positive affect on persuasion processes.

Magwood, G. S., Jenkins, C., & Zapka, J. (2009). Validation of diabetes health-related quality-of-life instruments using cognitive interviewing with older African Americans. *Journal of nursing measurement, 17*(3), 195-220.

Mallinson, S. (2002). Listening to respondents:: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Social Science & Medicine, 54*(1), 11-21.

Malpass, A., Dowrick, C., Gilbody, S., Robinson, J., Wiles, N., Duffy, L., & Lewis, G. (2016). Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *Br J Gen Pract, 66*(643), e78-84. doi:10.3399/bjgp16X683473

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: defining what really matters to patients. *JAMA, 312*(13), 1342-1343.

Miller, E., & Daly, E. (2013). *Understanding and measuring outcomes*. Retrieved from Institute for Research and Innovation in Social Services:

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

Narayan, S. S. (1995). *Satisficing in attitude surveys: The impact of cognitive skills, motivation, and task difficulty on response effects.* The Ohio State University,

NCCMH. (2018). The Improving Access to Psychological Therapies Manual. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual.pdf

NHS England. (2020). Personalised care and support planning. Retrieved from https://www.england.nhs.uk/ourwork/patient-participation/patient-centred/planning/

Nieuwkerk, P. T., Tollenaar, M. S., Oort, F. J., & Sprangers, M. A. (2007). Are retrospective measures of change in quality of life more valid than prospective measures? *Medical care, 45*(3), 199-205.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological review, 108*(2), 291.

Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology, 29*(8), 1011-1020.

Parducci, A. (1965). Category judgment: a range-frequency model. *Psychological review, 72*(6), 407.

Paterson, C. (2004). Seeking the patient's perspective: a qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Quality of life research, 13*(5), 871-881.

Pavkov, T. W., Priest, J., & Fox, K. (2012). Use of cognitive question testing methodologies in participant action research: implementing the Three-Step Test-Interview Method in Indian county. *Cultur Divers Ethnic Minor Psychol, 18*(1), 82-90. doi:10.1037/a0026844

Rapkin, B. D., Ryniker, L., Morgenstern, B., & Schwartz, C. (2011). User manual for the quality of life appraisal profile. *Concord, MA: DeltaQuest Foundation Inc*.

Rapkin, B. D., & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and quality of life outcomes, 2*(1), 14.

Rasinski, K. A., Baldwin, A. K., Willis, G. B., & Jobe, J. B. (1994). *Risk and loss perceptions associated with survey reporting of sensitive behaviors.* Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 13*(5), 465-484.

Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour research and therapy, 24*(1), 1-8.

Robinson, J., Khan, N., Fusco, L., Malpass, A., Lewis, G., & Dowrick, C. (2017). Why are there discrepancies between depressed patients' Global Rating of Change and scores on the

Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open, 7*(4), e014519. doi:10.1136/bmjopen-2016-014519

Rothman, A. J., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin, 24*(10), 1053-1064.

Sajobi, T. T., Brahmbatt, R., Lix, L. M., Zumbo, B. D., & Sawatzky, R. (2018). Scoping review of response shift methods: current reporting practices and recommendations. *Qual Life Res, 27*(5), 1133-1146. doi:10.1007/s11136-017-1751-x

Schober, M. F. (1999). Making sense of questions: An interactional approach. *Cognition and survey research*, 77-93.

Schwartz, C. E., Finkelstein, J. A., & Rapkin, B. D. (2017). Appraisal assessment in patient-reported outcome research: methods for uncovering the personal context and meaning of quality of life. *Qual Life Res, 26*(3), 545-554. doi:10.1007/s11136-016-1476-2

Schwartz, C. E., Powell, V. E., & Rapkin, B. D. (2017). When global rating of change contradicts observed change: examining appraisal processes underlying paradoxical responses over time. *Quality of life research, 26*(4), 847-857.

Schwarz, N. (1990). What respondents learn from scales: The informative functions of response alternatives. *International Journal of Public Opinion Research, 2*(3), 274-285.

Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review/Revue Internationale de Statistique*, 153-168.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21*(2), 277-287. doi:10.1002/acp.1340

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: another look at the availability heuristic. *Journal of personality and social psychology, 61*(2), 195.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological inquiry, 14*(3-4), 296-303.

Schwarz, N., & Hippler, H.-J. (1987). What response scales may tell your respondents: Informative functions of response alternatives. In *Social information processing and survey methodology* (pp. 163-178): Springer.

Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*(4), 570-582.

Schwarz, N., Strack, F., Hippler, H. J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*(3), 193-212.

Scorza, P., Stevenson, A., Canino, G., Mushashi, C., Kanyanganzi, F., Munyanah, M., & Betancourt, T. (2013). Validation of the “World Health Organization disability assessment schedule for children, WHODAS-child” in Rwanda. *PloS one, 8*(3).

Sears, D. O. (1983). The person-positivity bias. *Journal of personality and social psychology, 44*(2), 233.

Sentell, T. L., & Ratcliff-Baird, B. (2003). Literacy and comprehension of Beck Depression Inventory response alternatives. *Community mental health journal, 39*(4), 323-331.

Spörrle, M., Gerber-Braun, B., & Försterling, F. (2007). The Influence of Response Lines on Response Behavior in the Context of Open-Question Formats. *Swiss Journal of Psychology, 66*(2), 103-107. doi:10.1024/1421-0185.66.2.103

Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine, 48*(11), 1507-1515.

Stussman, B. J., Bethell, C. D., Gray, C., & Nahin, R. L. (2013). Development of the adult and child complementary medicine questionnaires fielded on the National Health Interview Survey. *BMC complementary and alternative medicine, 13*(1), 328.

Tanur, J. M. (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*: Russell Sage Foundation.

The Better Care Fund. (2015). How to understand and measure impact. In (Vol. Issue 4): NHS England.

Tilton, M., Bach, J., & Wang, T. (1993). Life satisfaction measures: comparison of ventilator supported spinal cord injured individuals with those free of ventilatory support. *J Am Paraplegia Soc, 16*, 253.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68*(3), 368-393.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological bulletin, 103*(3), 299.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin, 133*(5), 859.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Ullrich-French, S., Cox, A., Cole, A., Rhoades Cooper, B., & Gotch, C. (2017). Initial validity evidence for the state mindfulness scale for physical activity with youth. *Measurement in Physical Education and Exercise Science, 21*(4), 177-189.

Wicker, F. W., Park, Y., McCann, E., & Hamman, D. (1995). Task involvement and satisficing with motivational scales. *Psychological reports, 77*(1), 211-223.

Wiklund, I., Holmstrom, S., Stoker, M., Wyrwich, K. W., & Devine, M. (2013). Are treatment benefits in neuropathic pain reflected in the self assessment of treatment questionnaire? *Health and quality of life outcomes, 11*(1), 8.

Willis, G. B., & Gonzalez, A. (1998). Methodological issues in the use of survey questionnaires to assess the health effects of torture. *The Journal of nervous and mental disease, 186*(5), 283-289.

Willis, G. B., Sirken, M., & Nathan, G. (1994). The cognitive aspects of responses to sensitive survey questions (Working Paper Series, No. 9). *Hyattsville, MD: National Cemer for Health Statistics, Cognitive Methods Staff*.

Wolpert, M. (2013). Do patient reported outcome measures do more harm than good? *BMJ, 346*, f2669. doi:10.1136/bmj.f2669

Wolpert, M. (2014). Uses and abuses of patient reported outcome measures (PROMs): potential iatrogenic impact of PROMs implementation and how it can be mitigated. *Adm Policy Ment Health, 41*(2), 141-145. doi:10.1007/s10488-013-0509-1

Wyer, R. S., & Carlston, D. E. (1979). *Social cognition, inference, and attribution*: Psychology Press.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology, 9*(2p2), 1.

Part 2: Empirical Paper

The Meaning Behind Responses And Changes In The IAPT Outcome Measures: Seeking Perspectives Of IAPT Service Users And Clinicians

Abstract

Aims. Questionnaires are routinely used in IAPT services to monitor changes in symptoms and functioning, informing everything from clinical decisions to commissioning. However, service users and clinicians have noted that questionnaires might not always represent patient's subjective sense of wellbeing, and little is known about the factors underlying their response process and the meaning of their responses. This study aimed to investigate (a) factors underlying responses to IAPT questionnaires, (b) the relationship between changes in questionnaire scores and a subjective sense of wellbeing, (c) the acceptability of the recently introduced Global Rating of Improvement (GRI) question, and (d) service users and clinicians' perspectives on pre-post treatment symptom changes.

Method. This study adopted a three-phased qualitative approach. In phase one, 121 service users from two participating IAPT services completed an online survey, from which 19 participated in individual cognitive interviews in phase two. Ten of their treating clinicians then completed a survey in phase three. Reflexive thematic analysis and a multiple case study methodology were used to extract themes and compare views of service users and clinicians.

Results. A range of personal, interpersonal, contextual and questionnaire factors underlying responses to the IAPT questionnaires were identified. The GRI question was acceptable and was able to capture aspects of wellbeing beyond symptom change.

Conclusions. The simultaneous use of both standard IAPT questionnaires and the GRI routinely might improve the understanding of service users' degrees of post-therapy improvement. The importance of involving service users in interpretation of their questionnaire responses is emphasised by the findings of this study.

Keywords: Questionnaire Response, Measures, Subjective Sense of Wellbeing, Cognitive Interviewing, Global Rating of Improvement, Service User And Clinician Perspectives, IAPT

Background

Context Of The IAPT Programme

The Improving Access to Psychological Therapies (IAPT) programme aims to provide evidence-based interventions to people with depression and anxiety within the National Health Service (NHS) in England (NCCMH, 2018). To justify the use of public funding, the IAPT programme is mandated to collect routine outcome data. The aggregated data are used to judge the performance of services based on the government's mental health policy (HMG/DH, 2011) and the NHS outcome frameworks (DoH, 2010). The "No Health Without Mental Health" policy document (HMG/DH, 2011) defines recovery as "a way of living a satisfying, hopeful and contributing life, even with limitations caused by the illness" (Anthony, 1993, p. 527), and acknowledges that the concept cannot be fully captured with any existing validated outcome measures. Thus the document re-defines recovery in terms of employment and quality of life, and specifies that the mental health minimum dataset (MDS) collected in the IAPT services can serve as a proxy indicator to assess attainment of this objective. In IAPT services, the term 'recovery' is defined as a change in the scores on symptom measures of depression and anxiety between the start and the last attended treatment session. Metrics such as recovery rate[1], reliable change and reliable recovery are used to establish clinical improvements. IAPT services are set a target of at least 50% of service users moving to recovery and 65% reliably improving on at least one symptom measure pre-post therapy (NCCMH, 2018).

The Role Of Outcome Measures In IAPT

Well-validated measures of depression (Patient Health Questionnaire; PHQ-9; Kroenke et al., 2001) and anxiety (General Anxiety Disorder Scale; GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006) are used in IAPT services as part of the minimum dataset (MDS). Disorder-specific measures

---

[1] All service users that start treatment "at caseness" (i.e. above the clinical cut-off for either or both of the depression and anxiety measures) and attend at least two treatment sessions are counted as having "moved to recovery" if they score below caseness on both symptom measures at their last appointment. Recovery rate is the percentage of the number of referrals moved to recovery divided by those that completed a course of treatment and were "at case" at the start of treatment.

such as the Social Phobia Inventory (SPIN) for social phobia (Connor et al., 2000) and the Impact of Event Scale-Revised for Post-Traumatic Stress Disorder (Weiss, 2007) are used to measure symptom change for those specific conditions. Using changes in symptom measures as primary outcomes of recovery, however, might fail to capture the diverse definition of recovery in the field and important aspects that are valued by service users (Lewis & Killaspy, 2014; McPherson, Evans, & Richardson, 2009). Recently, a more generic measure of respondent's own sense of pre-post treatment improvement, i.e. the Global Rating of Improvement (GRI) measure, has been used to determine minimal clinically important differences on the PHQ-9 and GAD-7 (Button et al., 2015), and has recently been used routinely in two IAPT services. The study described here draws on the experiences of service users and IAPT clinicians in using the MDS and the GRI measure.

Literature On Questionnaire Response

Research from cognitive psychology and questionnaire development have shown that a range of factors can influence responses to questionnaires[2]. Factors such as the mode of administration, question order, response options and type of questions being asked all play a role in people's questionnaire response (Babor et al., 1990; Bless & Schwarz, 1999; Brown et al., 2007; Schwarz, 1999; Stone, Broderick, Schwartz, & Schwarz, 2008; Tourangeau et al., 2000). Social desirability, relationship with the clinicians, and the setting affiliated with the measure can hinder truthful responding (Krumpal, 2013; Rasinski et al., 1994; Tourangeau & Yan, 2007). People might rely on their mood and other contextual cues to inform their responses when their cognitive ability or motivation is compromised (Krosnick et al., 1987; Krosnick et al., 1996). Their way of responding to the same questionnaire can also change over time due to changes in psychological constructs and internal standards (Fokkema et al., 2013; Rapkin & Schwartz, 2004; Robertson et al., 2009; Sajobi et al., 2018; Schwartz, Finkelstein, et al., 2017; Sprangers & Schwartz, 1999). All of the above factors could have significant impacts on responses to MDS questionnaires used in IAPT.

---

[2] These are discussed in detail in the Conceptual Introduction to this Thesis.

Research On Responses To Symptom Measures And GRI

There has been very limited research on questionnaire response in IAPT settings, but relevant studies on symptom measures and GRI question in primary care and other clinical settings might offer some insights into the topic.

A recent qualitative study using cognitive interviewing techniques showed that, patients in primary care settings have issues with comprehension and response mapping with the measure, and expressed hesitations in responding truthfully to the PHQ-9 question about suicidal ideation (Malpass et al., 2016). Beyond response difficulties, investigation into the way service users interpreted their questionnaire scores and recovery process revealed that service users might define “good outcomes” in terms of gaining a sense of empowerment, continual development, and personal balance in face of ongoing struggle (De Smet et al., 2020). Important clinical progress can take place even when there was a lack of change in pre-post treatment questionnaire scores (De Smet, Meganck, Van Nieuwenhove, Truijens, & Desmet, 2019). These qualitative studies yielded different understandings of responses to symptom measures from those generated from a quantitative and statistical perspective.

Various studies have explored the complex relationship between GRI and symptom measures among patients in general practice. While quantitative studies have identified that people with more severe symptom scores are less likely to report improvement on GRI (Hobbs et al., 2020), qualitative investigations have offered possible reasons behind the mismatch between GRI and symptom measure scores. Themes such as the perception of GRI being a more accurate depiction of their mental states, impact of life events on one but not both types of measures, purposefully underscoring on the PHQ-9 for self-motivation, and retrieval difficulty were common amongst primary care patients with depression (Robinson et al., 2017).

Beyond service users’ views of outcome measures, clinicians also have their unique perspective to contribute. Studies on clinical use of routine outcome measures found that both service users and clinicians agreed on the benefits of using questionnaires to steer conversations and the direction in therapy (Unsworth, Cowie, & Green, 2012; Wolpert, Curtis-Tyler, & Edbrooke-Childs, 2016). Other studies have revealed divergent views on questionnaire use between clinicians and service users (Dowrick et al., 2009). For instance, compared with their patients, clinicians might have more concerns about the limits of standardised symptom measures in capturing the complexity of service context, clinical conditions, interpersonal dynamics and personal circumstances (Norman, Dean, Hansford, & Ford, 2014; Truijens et al., 2019; Wolpert et al., 2016).

Rationale Of The Study

In IAPT, service users receive psychological therapies for depression, anxiety and other common mental health problems, and the MDS is made up of a broader range of symptom measures and data are routinely collected on a sessional basis rather than pre-post treatment only as in other studies. Thus, the question remains as to whether previous findings are generalizable to the IAPT setting given the difference in treatment model, psychological conditions and data collection. In extension to this, the impact and acceptability of using the GRI on a sessional basis is unknown. The assessment of service users' intentions behind their GRI rating may be pertinent to the joint treatment decision making between service users and IAPT clinicians. Clinicians might also provide an additional perspective on the MDS and GRI measures, given their knowledge about recovery targets and the centrality of questionnaire response at all levels of the IAPT service.

Objectives

1. To explore factors underlying responses to the MDS questionnaires and the GRI question
2. To explore the relationship between changes in scores on these measures
3. To assess the acceptability of the GRI question to clinicians and service users
4. To compare the way changes in the outcome measures are perceived by service users and their clinicians

Method

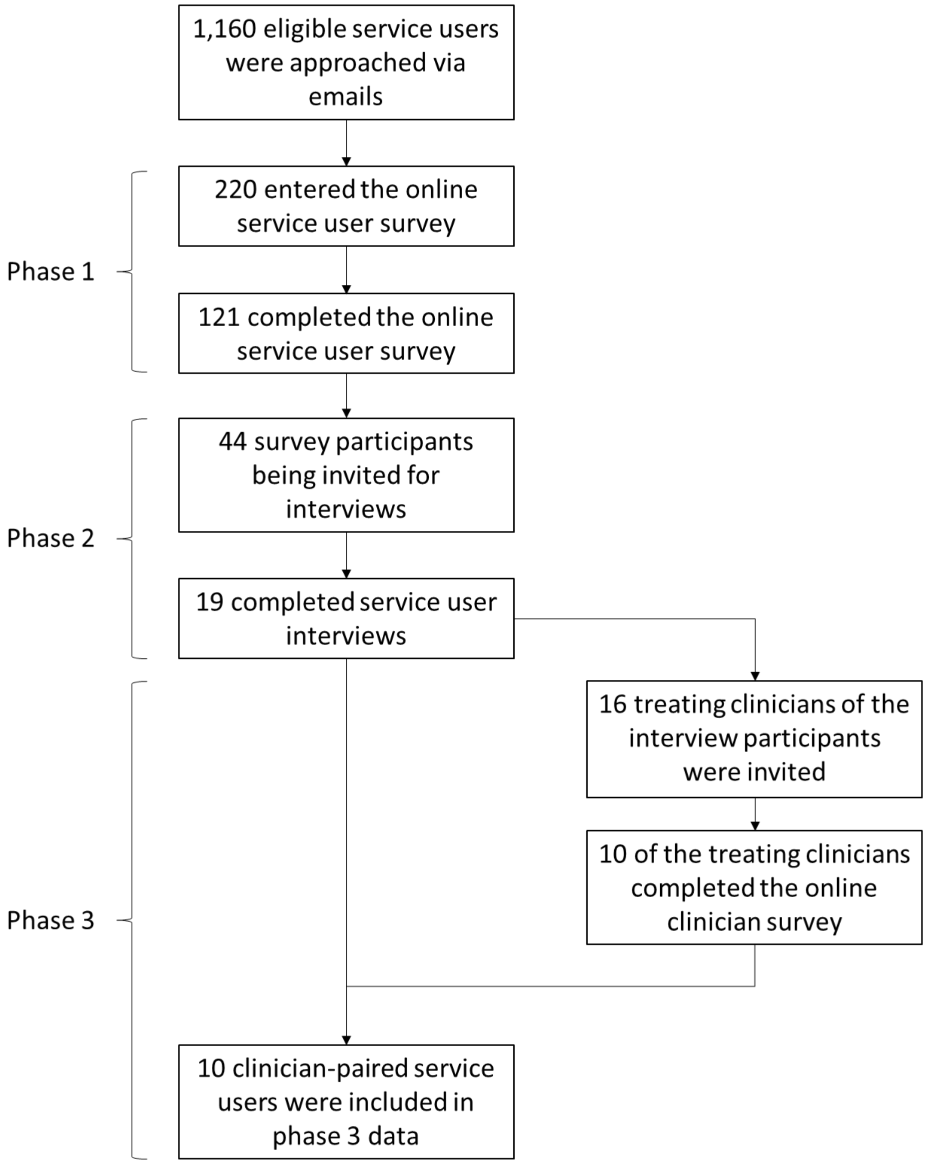
This is a three-phase sequential qualitative study which consisted of a service user survey, individual service user interviews and a clinician survey. Findings from earlier stages were used to inform subsequent exploration (Appendix A and B). Ethics approval for this study has been obtained from UCL Research Ethics Committee and North West – Preston NHS Research Ethics Committee (See Appendix C for the ethics approval letter).

Participants

Participants were service users and staff of two London IAPT services. Data were collected between 24th January 2020 and 8th May 2020. Eligible participants had to have given consent to be contacted for research purpose, have completed two treatment sessions, were discharged between 1st April 2019 (the date when the GRI question was introduced to the service) and 24th January 2020, to be willing and able to provide written informed consent, to read and write in English, and to have access to an electronic device to complete the online survey. Clinicians would only be contacted with the consent from their service users and had to be contracted by the IAPT services at the time of contact (Figure 1).

Figure 1

Participant Flow Diagram



Phase 1: Service User Surveys

The screening of eligible service users was done on IAPTus, the electronic health record system used in the participating IAPT services. Invitation emails with a link to the REDCap online survey and a unique participant code were sent securely to all eligible service users (N = 1,160). Participants had to read through the Participant Information Sheet and agree to the terms listed on the consent form before they could proceed to the survey (Appendix D). A question about their interest in participating in the phase two interview was also included. A total of 121 complete survey responses were received, with a response rate of 10.4% and a completion rate of 55.0%.

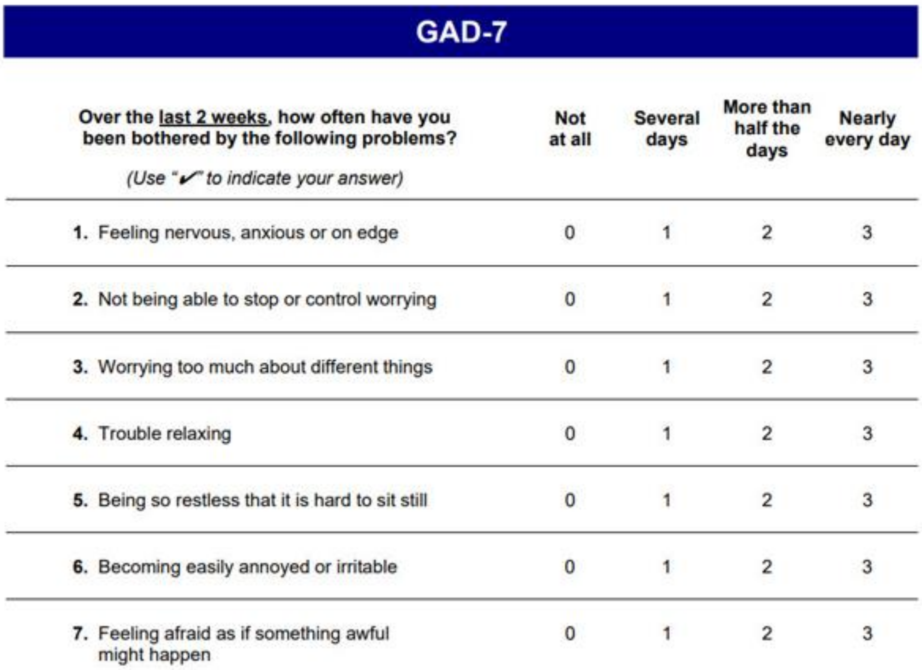
The survey (Appendix D) asked about their response to the MDS measures and GRI question. To optimise the study design and to increase the specificity of the findings, this study has chosen to focus on the two MDS measures that all respondents would have filled out at least once by default in their initial assessment, i.e. PHQ-9 and GAD-7. Based on the existing literature, the survey included questions on the following dimensions: (a) comprehension, (b) memory, (c) judgment, (d) effort, (e) response options and (f) perception (Dillman & Redline, 2004; Fayers, Langston, Robertson, & Group, 2007; Robertson et al., 2009; Tourangeau et al., 2000). Participants were asked to select which factor(s) played a role in their response and to rank each dimension in terms of its influence on the questionnaire response. The survey included free-text boxes to prompt elaboration on their answers and to capture additional factors. It also asked about the acceptability of the GRI question and respondents' perceptions of both the MDS measure and GRI question. Images of the measures were included as prompts (Figure 2 and 3). It took around 15 minutes to complete the survey. All survey participants were entered to a prize draw of a £50 amazon voucher, unless they had chosen to opt out from it. There was no reimbursement for service users completing the survey.

Figure 2

Image of the MDS measures

PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use "✓" to indicate your answer)*	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

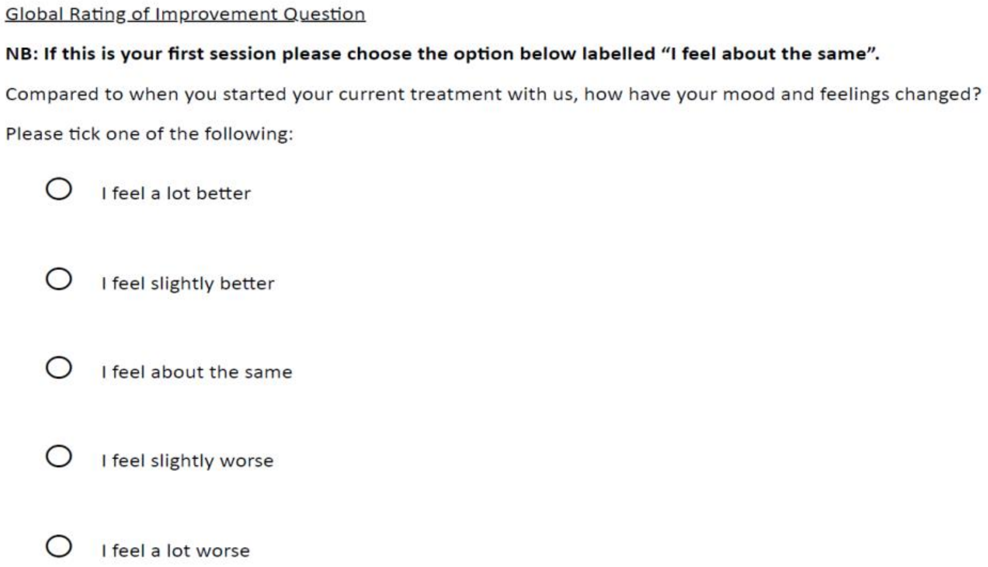


GAD-7

Over the last 2 weeks, how often have you been bothered by the following problems? *(Use "✓" to indicate your answer)*	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3

Figure 3

Image of the GRI question



Global Rating of Improvement Question

NB: If this is your first session please choose the option below labelled "I feel about the same".

Compared to when you started your current treatment with us, how have your mood and feelings changed?

Please tick one of the following:

○ I feel a lot better

○ I feel slightly better

○ I feel about the same

○ I feel slightly worse

○ I feel a lot worse

Phase 2: Service User Interviews

A follow-up invitation email was sent to participants who had completed the survey and had indicated an interest in taking part in an individual interview. Interviews were arranged based on the service users' preference, room availability and the researcher's schedule. This resulted in 19 interviews being completed and audio-recorded. All interviews were conducted by the researcher (PL). The interview sample size was comparable to recommendations from the existing literature (Beatty & Willis, 2007; Willis, 2004). Face-to-face interviews were held in individual rooms in one of the two IAPT services, two interviews were conducted over telephone in a private residential space[3].

The interviews were structured using cognitive interviewing techniques with an interpretive framework (Chepp & Gray, 2014). Retrospective probes (Appendix E) were scripted based on past studies (Malpass et al., 2016; Tourangeau et al., 2000), and other factors arose from the phase one survey. Interviewees were encouraged to reflect on their experience of filling out the questionnaire

---

[3] This arrangement was due to social distancing rules as part of the governmental response to the COVID-19 pandemic.

and other factors beyond those already mentioned in their surveys. The interviews took on average 30 minutes to complete and interview participants were given £10 to compensate for their time.

Phase 3: Clinician Surveys

All interview participants granted us consent to contact their treating clinicians to fill out an online survey on REDCap (Appendix F). In the invitation email, clinicians were provided with the IAPTus number for them to know which service user the survey would refer to. Clinicians had to read through the Participant Information Sheet and agree to the terms on the consent form before proceeding to the survey questions. The survey asked about their perceptions of pre-post treatment clinical changes of their service users, and their understanding of what may have explained the changes in the measures beyond symptom change. The survey took around 10 minutes to complete. Two clinicians have left the service at the time of contact, one of which had two participants on their caseload. This resulted in sixteen treating clinicians being invited, and ten completed the survey.

Data Analysis

Survey responses were downloaded from REDCap, with the demographic and clinical information extracted from IAPTus. Data were anonymised, processed and stored in UCL Data Safe Haven. A professional company contracted by the institution and abided by the data protection policy transcribed all the audio-recordings of the interviews. All personal identifiable information was removed from all data before analysis. Free-text survey responses and interview transcripts were manually analysed in NVivo 12. Quantitative data were processed in SPSS 22. Credibility of the study was enhanced by involvement of IAPT staff members during data analysis, the triangulation of different data sources and a reflexive journal keeping (Berger, 2015; Guest, MacQueen, & Namey, 2011; Patton, 1999).

Analysis of the free text responses in the service user surveys and the interview transcripts was performed under a phenomenological framework. It followed the recursive and iterative six-stage process of reflexive thematic analysis (Braun & Clarke, 2006, 2014; Clarke & Braun, 2018). In addition, an IAPT staff member was involved after the initial round of data analysis to analyse half of the survey data and two of the interview transcripts. The researcher and the staff member then review the themes and codes generated by each other. Insight gained from the review was used to inform subsequent rounds of data analysis conducted by the researcher, from which the report was produced.

A multiple case study (Yin, 1981a, 1981b) was performed to triangulate findings from the clinician surveys, clinical record from IAPTus and the clinician-paired service users' surveys and interviews. Data from the clinician-paired participants were re-analysed in the context of their clinicians' responses. The methodological flexibility enabled the researcher to shift the focus according to the research intent, and helped to capture the complexity and depth of the phenomena (Sandelowski, 2011; Simons, Lathlean, & Squire, 2008). The comparative analysis allowed areas of disagreement among different data sources and similar patterns across different pairs to be identified.

Results

This section presents collated findings from service users survey (SS) and interview (SI) under the sections "Service User's Perspectives" and "Clinicians' And Service Users' Perspectives". All participants were assigned a pseudonym for the purpose of this manuscript only.

Service Users' Perspectives

Most of the survey respondents were female (n = 86, 71.1%), white (n = 92, 76.1%) and in employment (n = 89, 73.5%). The mean age was 34.95 (SD = 13.12). The majority of them self-referred (n = 97, 80.2%) to the service for depressive and anxiety disorders (n = 100, 82.6%), with their MDS scores above clinical thresholds, i.e. at caseness (n = 117, 93.4%) when they entered the service. The mean number of sessions attended were 7.55 (SD = 4.69), with reliable improvement at 74.4% and recovery rate at 57.0%. Compared with the latest available IAPT report (NHS Digital, 2019a), these participants received a higher number of sessions, and were slightly more likely to meet criteria for reliable improvement and recovery[4]. As a subgroup of the survey respondents, the demographic and clinical features of the interview participants reflected the diversity of the overall sample. Chi-square test revealed that interview participants tended to have more favourable clinical outcomes of reliable improvement at 94.7%, X^2 (2, N = 121) = 6.55, p = .038, and reliable recovery at 73.7%, X^2 (1, N = 121) = 3.91, p = .048. They also received a higher mean number of clinical sessions (M = 9.95, *SD* = 4.76) compared to the those who only participated in the survey (M = 7.11, *SD* = 4.56), $t(119) = -2.48$, p = .015. (Appendix G).

Perceived Factors Underlying Responses to MDS Measures

In the survey, we asked people whether their responses to the MDS measures were affected by: (a) their understanding, (b) memory, (c) judgment, (d) effort, (e) the response options available and the (f) presentation of the measures. Eighty-eight percent of the survey participants selected at least one of these factors as being influential in their response process. Of which, 'response options' (n = 65) and 'judgment' (n = 60) were the most frequently selected factors. 'Effort' was the least frequently selected factor (n = 18) (Figure 4). Further, comprehension, memory, judgment and response were more likely to be ranked as the most influential factors, while presentation and efforts being ranked as the least influential (Figure 5, Appendix H).

Analysis from free text response showed that respondents' description on how individual cognitive processes and other factors influenced their response to the MDS measures were highly overlapping (Appendix I). For instance, the lack of specificity of questionnaire wordings could simultaneously impact on comprehension, memory, judgment, and response. The interrelated nature of these processes makes it less helpful to categorise the themes extracted from the free-text responses into individual cognitive components. Instead, it was noticed that respondents tended to describe particular aspects related to the measures, context, their personal characteristics and/ or other people either make the response process easier or more difficult for them. Therefore, the

---

[4] the national average was 6.9 sessions, with reliable improvement at 67.4% and recovery rate at 52.1%

responses were reconsidered as a whole and reorganised into two sets of intersecting higher order themes based on the type of factors (personal, interpersonal, contextual and questionnaire) and its effect on the response process (facilitative vs. influencing). A factor was defined as facilitative if the respondent described it as helping to make the response more reflective of their internal state and wellbeing or the response process easier. Conversely, an influencing factor would be one that posed threat to the accuracy of response or made the response process more difficult for the respondent. The occurrence of each theme in the service user survey and the number of responses from unique participants in interviews were included to reflect the relative importance and to direct the attention of this study (Table 8 and 9; Appendix J).

Figure 4

Bar Chart Of Frequency Of Cognitive Processes Being Selected As Influential In The Response Process

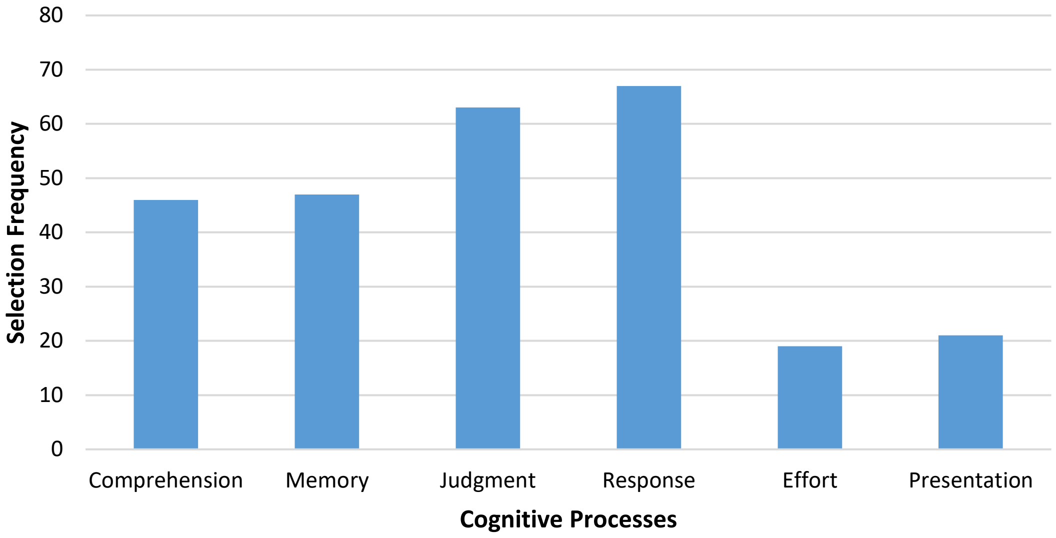
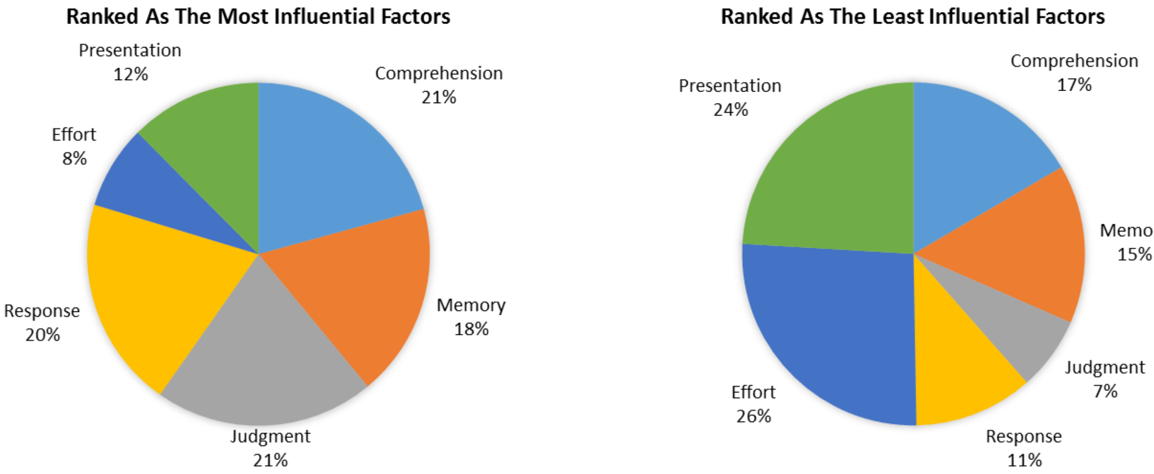


Figure 5

Pie Chart Of Factors Ranked As The Most And Least Influential Factors In The Response Process



Notes. Most Influential Factors are defined by factors with a rank score of 5 or 6, while least influential factors are defined by factors with a rank score of 1 or 2.

Table 9

Facilitative Factors Underlying Responses To MDS Measures

Facilitative Factors	Service Users Survey (SS)	Service Users Interview (SI)
Personal Factors *Occurrence in SS: 20* *Unique Responses in SI: 9*	– Easy to understand and answer (13) – Familiarity with questionnaire facilitates understanding (2) – Use of response options and previous response as reference (2) – Easier to remember personally relevant information (1) – Easier to decide when symptoms are stable (1) – Feeling of personal experience being normalised (1)	– ***Attempt to keep consistency in judgment process (2)*** – ***Improved accuracy due to inability to remember previous response (1)*** – ***Familiarity with questionnaire (5)*** – Positive feelings being evoked by questionnaire - feeling validated and normalised (1)
Interpersonal Factors *Occurrence in SS: 3* *Unique Responses in SI: 13*	– Discussion with therapist facilitates comprehension and makes it meaningful (3)	– ***Relationship with clinicians facilitate truthful responding (4)*** – ***Discussion with clinicians to make questionnaire more personally and clinical relevant (5)*** – ***To gain clinician's view on the scores and own progress (2)*** – Discussion with clinician facilitates comprehension - Clarify questionnaire use and way of responding (5) – ***To steer direction of therapy (3)***
Contextual Factors *Occurrence in SS: 4* *Unique Responses in SI: 1*	– Reminder to complete (2) – Easily accessible on device decreases effort required (2)	– ***Scores being automatically sent to clinicians (1)***

Facilitative Factors	Service Users Survey (SS)	Service Users Interview (SI)
Questionnaire Factors *Occurrence in SS: 40* *Unique Responses in SI: 5*	– Accurate description of experience and align with own feelings (12) – Questionnaire as a tool for self-reflection and progress tracking (7) – Response range and multiple-choice format facilitates decisions and makes it less effortful (4) – Brevity of response and questionnaire length (4) – Clear presentation (4) – The repeated use of same measures promotes understanding and perception, and helps to track progress (3)	– Specificity, clarity, appropriate length and discrete response categories facilitates responding (4) – Repeated use of same measures - serves as an anchor to track progress (1) – ***Timeframe and specificity of symptom measures help to mitigate impact of current mood state (2)*** – Matching between recall period and session frequency (1)

Notes. Numbers in brackets in SS column indicate the number of unique responses in each category (each participant might contribute more than one unique responses in each category)

Numbers in brackets in SI column indicate responses from unique participants

Accented items in SI column denote new themes from SI

Table 10

Influencing Factors Underlying Responses To MDS Measures

Influencing Factors	**Service Users Survey (SS)**	**Service Users Interview (SI)**
Personal Factors *Occurrence in SS: 94* *Unique Responses in SI: 17*	– Cognitive processes impacted by fluctuation of conditions and mental states (12) – Negative emotions (anxiety about over-reporting, reinforce negative self-view, amplify negative feelings, unease, overthinking) evoked by questionnaires (12) – Struggle to quantify feelings (10) – Struggle to understand response options and language (9) – Attention span, memory and casual responding impacted by energy level and effort required (8) – Concerns about response accuracy (8) – Unsure about the frame of reference impact on decision (7) – Understanding, memory and judgment limited by self-awareness (7) – Limited personal relevance impact on memory, judgment, effort and response (7) – Struggle to recall and use of availability heuristic (6) – Lack of motivation (3)	– Motivation impacted by perceived personal relevance, uncertainty about the purpose and intention of the questions, feeling of being obliged to fill in (9) – Struggle to recall emotional experience, and specificity of the episodes, issues with timeframe (9) – Negative feelings and self-view (anger, closed down, embarrassment, overwhelmed, uselessness) being evoked by questionnaire (8) – Cognitive processes impacted by fluctuation of conditions and mental states (8) – Idiosyncrasy of interpretation (4) – ***Unfamiliarity with the type of question (4)*** – Limited by level of insight (3) – Interaction between different cognitive processes (2) – Understanding of language (2) – ***Recalibration - change of perception of self or the past (2)*** – Unsure about frame of reference (2)
Interpersonal Factors *Occurrence in SS: 17* *Unique Responses in SI: 11*	– Consideration of how the response might be perceived or used in the service context impact on response and judgment (8)	– ***Over- or under-reporting due to perceived clinical implication of scores (9)*** – ***Lack of discussion about the clinical relevance of scores (7)***
Influencing Factors	**Service Users Survey (SS)**	**Service Users Interview (SI)**
Interpersonal Factors (contd.)	– Perceive the measures are completed for the service or clinicians (3) – Inability to check meanings and progress with others (3) – Infer severity of symptoms from being in the service (2)	– ***Unable to explain to the clinician the discrepancy between feelings and scores (1)*** – ***Spending too much time focusing on the questionnaire (1)***

Influencing Factors	Service Users Survey (SS)	Service Users Interview (SI)
	– Sensitivity of information impacted on honesty (1)	
Contextual Factors *Occurrence in SS: 75* *Unique Responses in SI: 14*	– Recency Effect and Current Mood impact on memory, judgment and response (51) – Time pressure, environment and time of completion impact on attention and effort (24)	– Recency Effect and Current Mood impact on memory, judgment and response (11) – Time pressure and stress increases likelihood of responding careless or more negatively (5) – Immediate environment impact on mood and authenticity (3) – ***Anxiety about discharge (4)*** – ***Timing of question being asked - the GRI questions being asked too early on in therapy (3)*** – *Technical issues with internet connection (1)*
Questionnaire Factors *Occurrence in SS: 115* *Unique Responses in SI: 5*	– Issues with response options impact on memory, decision and response (70) – Lack of specificity of questionnaire wordings impact on understanding, retrieval, judgment and response (21) – Issues with recall periods (mismatch with frequency of sessions, recall period too long) impacted on memory and response (14) – Does not differentiate causes of behaviours and feelings (6)	– Specificity, clarity, appropriate length and discrete response categories facilitates responding (4) – Repeated use of same measures - serves as an anchor to track progress (1) – ***Timeframe and specificity of symptom measures help to mitigate impact of current mood state (2)*** – Matching between recall period and session frequency (1)
Questionnaire Factors (contd.)	– Technical issues affect understanding, effort and perception (5) – Perceived as unfriendly and not straightforward (3) – Severity of symptoms does not reflect the level of distress (2) – Length of questionnaires impacted on effort (2)	

Notes. Numbers in brackets in SS column indicate the number of unique responses in each category (each participant might contribute more than one unique responses in each category)

Numbers in brackets in SI column indicate responses from unique participants

Accented items in SI column denote new themes from SI

Overall, there were more themes categorised as influencing than facilitative. Personal factors captured aspects related to individual cognitive processes, psychological functions and emotional states. Interpersonal factors referred to relational conditions and concerns. Contextual factors described external influences such as the immediate environment and time. Questionnaire factors were dimensions related to questionnaire designs.

Personal Factors.

Many participants reported struggles with cognitive processes as impacted by the stability of their symptoms, energy levels, and levels of self-awareness. Perceived lack of relevance and response difficulties can lead to a decrease in motivation and increase in causal responding. As depicted, a respondent who “didn't find the [MDS] questions helpful or relevant” would “mainly guessed how I felt.” (P11, SS) and another respondent was unsure if they were “evaluating myself against what I think a normal person feels, or… against myself”, and therefore “tended to just rush through.. and rely on instinct in giving my score” (P114, SS). Familiarity, personal relevance, use of heuristics and symptoms stability were seen as facilitative factors for their response process. Having their symptoms precisely described in the questionnaire statements could facilitate the response process, help to normalise feelings captured in the questionnaire.

The Role Of Emotions.

Negative emotions were commonly reported in both surveys and interviews. The questionnaire could evoke “worry about exaggerating” (P99, SS), “feelings of unease” (P84, SS) and negative self-views. It could feel “arbitrary” by “just assigning numbers to these deeply personal feeling(s)” (P57, SS). Anger or frustration was often associated with not seeing the questionnaires as relevant but feeling obliged to complete them in return for receiving treatment. Respondents might infer their subjective states from the level of response difficulties they were experiencing. For example, it could feel “distressing” when they “had a week or 2 which were difficult to remember”. Their recall difficulties might make them “believe I was wasting everyone's time which was a difficult feeling during a hard time" (P50, SS). Conversely, emotions could also lead to response difficulties, as described, "I just felt more worried about it and had to read more carefully and think about it. That's just because I was already stressed, but there was nothing in the setting out that made it difficult." (P100, SI). Negative feelings could be heightened by being reminded of their negative experience in the questionnaire items, and might have biased their self-evaluation: "discretely each event isn't that much of a big deal, but because you're forced to think about it in one place…it's going to come together and amplify how you feel." (P80, SI)

Interpersonal Factors.

Interpersonal factors were more prominent in interview findings than in the survey. Responses could be influenced by "how I believe the person who is reading my questionnaire will interpret my symptoms" (P30, SS). This might lead to the tendency of "biasing your own responses by wanting to show an improvement" (P105, SS) or underscoring as "not wanting to create alarm" (P78, SS). When MDS scores were not discussed or considered for clinical decision it could decrease the perceived utility of the questionnaire. As demonstrated in this situation that "it's almost the end of the sessions and I was still not ready for the end, so then when I heard that it was going to be the end… I didn't feel like there was much point of these questionnaires" (P63, SI). Some described positive use of the measures in facilitating conversation in sessions and suggested that a positive relationship with the clinician could motivate truthful responding.

Over- Or Under-Reporting On Measures.

Our findings suggested that over- or underreporting on the MDS measures was often a result of interpersonal concerns. Over-reporting was associated with the perception that access to the service was contingent on the level of symptom severity indicated by the MDS questionnaire scores. This could lead to hesitation to admit improvement. As described, " I don't want to be too happy just so they [the service] don't cut me off when I've still got a handful of things that I actually could do with a chat about" (P91, SI). There could be tension between over-reporting to reflect the need for further support and under-reporting due to concerns about service constraints. As illustrated*,*

> “am I wasting their time? Am I taking this time away from someone who’s having it worse? When in reality in that two-week period I might not have had a major panic attack but that doesn’t mean everything’s fine long-term” (P50, SI).

Under-reporting could be due to positivity bias, impression management and worries about potential clinical consequences. A positive relationship with therapist might in this case hamper truthful responding. As described, “I feel bad a lot of the time. Saying that I feel worse, because I felt like it was not just me working on it, but also the therapist… So then I would try and say that it was better… or the same” (P63, SI). Underscoring on the risk question on PHQ-9 (i.e. “Thoughts that you would be better off dead, or of hurting yourself”) was mainly driven by the intention to avoid follow-up actions being undertaken. This was particularly the case for some participants that had previous relevant experience, and considered past actions based on their score on the suicidal item as being disproportional to their subjective sense of suicidality. As demonstrated in the scenario below,

> “That’s not very representative of me being or not being suicidal… because it’s obviously quite hard to put emotions and the mental state into numbers, but I remember that one specific time when it went very high and they needed to contact my GP… I think I [was] just scared to say that again…in case it then creates some sort of domino effect of things going into motion.” (P91, SI)

Perceived Clinical Relevance And Use.

Our finding showed that respondents made sense of the purpose of the questionnaires and the use of scores through their interactions with the clinicians and service context.

Six of the nineteen interview participants reported that they were unsure what the questionnaires were used for and felt that it was irrelevant to their clinical experience. This could lead to decreased motivation and increased response difficulties. One participant who felt "this form's a waste of time" had a conversation with their clinician, but were only repeatedly told that, "I know, but you have to do it" without being offered any explanation (P40, SI).

Despite not having a clear understanding of the intended purpose of these measures, in general participants were actively thinking about how their questionnaire scores might be interpreted and used (Table 12). In most cases participants suggested it could have multiple purposes and the perceived use could be interpreted in both a positive or negative manner. For example, for those perceiving the questionnaires as intended for the service's use only, a participant described having "a huge amount of respect for the NHS… so if you ask me I'll try and help" (P87, SI), whilst another mentioned how "It didn't feel like there was any interest in helping me or understanding my issues, it was just a way to show you had good numbers" (P29, SS).

In comparison, participants' views were unequivocally positive if they considered the questionnaires could be used for themselves. In a sense, all the perceived personal uses were related to seeing the questionnaires as an extension of the therapy, either as a tool to enhance their therapeutic progress or to facilitate clinical discussion. As depicted in the scenario below,

> "I think having done the questionnaire meant that during the week when I was displaying those symptoms, I was more conscious of it… also in my sessions, my therapist would always have a look at it… It was already useful for me as a reflective exercise, but it felt even more useful that it was actually discussed in the sessions." (P100, SI)

Generally, participants expressed a more favourable view if they considered the perceived use as being in line with their treatment goal, or at least not in conflict with their personal interest. This could have a direct effect on how they decided to respond. Using two opposite responses to the risk question on PHQ-9 as an example,

> "I used to say that to [my clinician] I'm not suicidal. So if I put nearly every day I think I would be better off dead … I wouldn't expect people to come along and cart me off somewhere… so… I'm completely honest." (P40, SI)

In contrast, another participant shared the major obstacle for them to provide a truthful response to the risk question was that:

> "you have to trust your therapist, that they're not going to go back and… contact your employers, and that was one of the things I was worried about… if I was honest, that it would then go back and question my ability to work." (P121, SI).

While the perceived use might account for the difference observed in these two scenarios, arguably it might be more to do with how the scores were being understood and whether the clinical decision was founded on sufficient communication around the interpretation and use of questionnaire scores. Participants might also vary their response strategies based on the perceived use of the questionnaire and the score, casting a direct impact on their answers (Appendix K).

Table 12

Perceived Use Of Questionnaire Scores

User Group	Perceived use
For the service	To justify access to service To assess severity of symptoms or give diagnosis To decide treatment lengths To evaluate service quality and efficacy of practice
For the clinician	To evaluate clinician's performance To monitor risk To gain understanding of service user's conditions To plan for the session To monitor service user's treatment progress
For the service user	To monitor own progress To reinforce subjective sense of improvement To increase self-awareness To gain clinician's views on scores To use it to open up conversations about own feelings To motivate engagement in therapy

Contextual Factors.

Contextual factors such as current mood, immediate environment, time of the day and time pressure could have a huge impact on people's questionnaire response. People's response to the MDS measure might be "depended more on the mood I was at the moment I was answering them, rather than remembering the whole week" (P60, SS). Contextual factors seemed to mostly affect respondents' attention and effort. For instance, it "[d]epends on how focus can you be in the situation you're answering the questions. Not the same being at home, taking your time than for example being in group sessions therapy" (P9, SS). Facilitative contextual factors included the automatic reminders and the accessibility on device.

Questionnaire Factors.

Issues related to questionnaire design made up the largest category in both facilitative and influencing factors. Most issues reported related to difficulties in mapping their internal response to the response options available in the measures. Specifically, the number of days options failed to capture the duration and intensity of symptom experiences and could not differentiate between a symptom occurring "a few minutes one day, or many hours on another" (P15, SS). Respondents also reported that their "[a]nswers often really fell in-between choices" (P119, SS). There was a qualitative difference perceived between the response categories, such that, the scale might be regarded as "non-linear". It could be "hard to decide between several and more than half the days" due to the perception of a "much less difference between those two options than the other 'steps'" (P71, SS).

Problems with the wording of questionnaire items included issues with double- and triple-barrelled statements (e.g. 'poor appetite or overeating'), which could led to response confusion, as "[b]y ticking a number however, it would seem as if I agree with the whole question" (P115, SS). Some felt unsure about the meaning of words, as it could be "different for different people." (P27, SS). The lack of considerations of the context and origins of behaviours and feelings also caused response difficulties. As a respondent suggested: “My episodes of anxiety are associated with trigger events. If there are no events I might have no symptoms but this does not mean I am better” (P33, SS). Moreover, the presence of certain feelings or behaviours could be unrelated to their conditions, for instance, “[a]s a young mother to a baby I would obviously be tired and lacking energy” (P77, SS).

There were issues with the two-week recall period being too long and suggestions that this could result in guessing at answers. As one respondent put it, “at some point over two weeks I had almost always felt these things in some way and it’s hard to remember to what degree” (P87, SS). A mismatch between the two-week recall period and weekly sessions also caused confusion for some.

On the other hand, some saw the item descriptions “outline the exact symptoms and feelings that someone can experience” (P90, SS), and the response options being “distinct” enough for them to “choose the answer which felt most representative” (P100, SS). The act of filling out the questionnaire also helped them to “reflect on how I was getting on” (P116, SS) and to “notice my emotional changes” (P102, SS). In relation to this, the repeated use of the same measures appeared to aid progress tracking and response. Succinctness of the response options, multiple-choice format, and questionnaire length all helped to make the response process easier.

Acceptability Of GRI

Overall, GRI demonstrated a high acceptability amongst the participants, with 80.3% seeing the value of being asked the GRI question, 57.9% regarded it as being extremely or somewhat easy to respond to, and 37.6% felt it was extremely or very useful (Figure 7). The majority (76.1%) have had a discussion with their clinicians about their GRI scores. Eighty-six survey responses referred to the positive values and uses of the GRI question, while eleven mentioned some limitations (Appendix L).

GRI was being valued for its ability to promote insight and self-awareness, to monitor progress, and for the service to evaluate treatment outcomes and clinicians' performance. It has also been used to facilitate therapy, serving as an "easy way to start the conversation" (P44, SS) or "letting people know how you feel" (P117, SS). The positive focus of the questionnaire might help to "boost their overall view of themselves" (P90, SS) and "reinforce the positive progress" (P89, SS). One participant recalled that "towards the end of my treatment, being able to select 'I feel a lot better' gave me a pride of how far I'd come and what I'd done" (P11, SS). Conversely, some respondents concerned that the GRI might make them feel "obliged to answer in a certain way" (P72, SS), and create the "pressure to 'improve'". Especially, "a number of consecutive 'negative answers could lead the person to think they're not doing well at all" and resulting in them "feeling worse than they actually are" (P96, SS). In addition, GRI is no exception to factors that influence other measures, such as memory functioning, specificity of questions and response options, and current mood (Appendix M).

Underlying The Subjective Sense Of Improvement

Understanding what came to mind when people responded to the GRI question could also highlight aspects that were important to individuals' sense of wellbeing (Appendix N). Interview participants described taking into account aspects beyond symptom change when judging their sense of improvement, such as daily functioning, life-style, medication, work and life experience.

Taking a broader perspective of their wellbeing and thinking about their “overall feeling of the world” could help to “transcend” their symptoms (P87, SI). For some, taking the active step in accessing help was a milestone in their recovery process. They also judged their improvement based on the attainment of therapy goal and the ability to continue implementing coping skills they learned from the session in face of ongoing life challenges. Some described the sense of improvement as a general feeling. A positive therapeutic relationship was also mentioned as directly contributing to their sense of improvement or indirectly through motivating them to engage in the therapeutic process.

Difference Between the GRI And MDS Measures

Forty-six percent of the survey respondents reported a difference in the way they responded to GRI and MDS measures. A slightly higher portion of respondents favoured GRI over MDS in capturing their feelings about the progress they made in therapy (36.8% and 31.7% respectively) (Figure 8, Appendix L). Overall, participants commented that GRI was “easier to answer” and mapped “accurately to my own experience” (P71, SS), while MDS has the benefit of being “more specific and so I would definitely [put] more thought into them” (P72, SS). Participants took into account a broader range of considerations in responding to GRI than MDS. It was suggested that two types of measures could be complementary of each other, with the MDS “highlighting specific topics” while the GRI “works like a summary” (P75, SS). They were seen as “representing maybe two different aspects of how my brain thinks about my mental health*”* (P100, SI). Being able to define their own sense of improvement could have clinical significance in the context of fluctuations of symptoms, as “I could still suffer from periods of acute low mood whilst still feeling like I was getting better overall" (P58, SS). It is also important to note the differential emotional impact brought by filling out these questionnaires, with GRI focusing on a more positive outlook and MDS having the potential to trigger negative emotions and thoughts in the respondents (Table 17).

Figure 7

Pie Charts Of The Perceived Ease Of Responding And Usefulness Of GRI

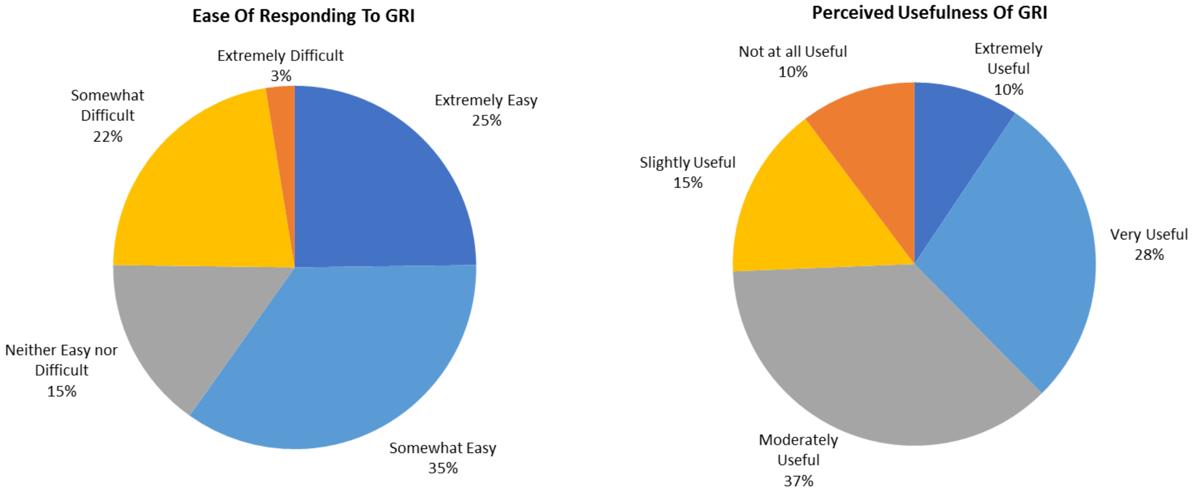


Figure 8

Pie Charts Of The Perceived Difference In The Way They Responded To GRI And MDS Measures And Their Ability To Reflect The Respondents' Feelings About Their Therapy Progress

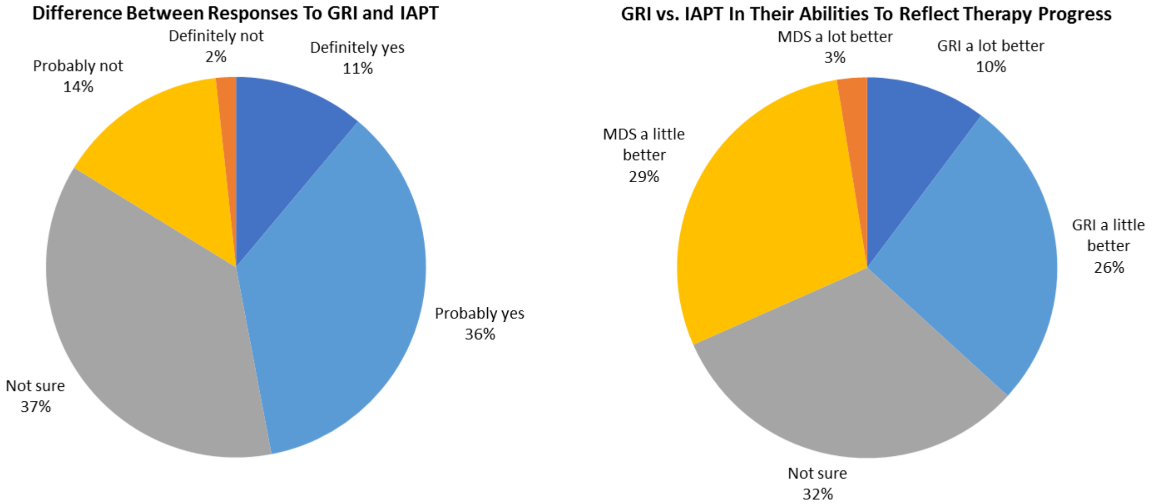


Table 17

The Perceived Difference In The Way Of Responding To GRI And MDS Measures (Data Analysis From Service User Survey)

	GRI	**MDS**
Effort	Less motivated to think through (4) Easier to answer (4) Struggle to condense progress into a brief response (1)	Motivated to think through (2) More effortful (2)
Considerations	Daily functioning (1) Current Mood (3) Overall clinical change (3) Pressure to improve (1) More honest (1)	Average of past experience (1) Weekly changes and occurrence of episodes (3)
Generic vs. Specific	Cannot capture the nuance of experience (3) Harder to answer and prone to overthink (2)	Capture improvement in specific and broader areas related to conditions (6) Easier to answer due to specificity (1)
Perception of the questions	As a measure of self-perception (1) Personal relevance (2) Assessment of clinicians' performance (1) Focus on holistic wellbeing, emotions, and subjective experience (5)	Direction questions to prompt self-reflection of feelings (1) Lack of personal relevance (1) Objective and quantitative measures (3) Focused on symptoms rather than subjective sense of improvement (2)
Positive vs. Negative-focused	Promote optimistic self-view and self-kindness (2)	Evoke negative self-views and emotions (3)

Notes. Numbers in brackets indicate the number of unique responses in each category (each participant might contribute more than one unique responses in each category)

Perceived Clinicians' View Of Change

Fourteen of the interview participants shared their views on how their clinicians might understand the scores in their measures. Of which, ten of them expected a shared understanding based on their previous discussion or positive therapeutic relationship. Four participants supposed that their clinicians might be more able to "draw patterns", while service users themselves are "less easy to be objective" (P14, SI). To the contrary, a participant who did not think the symptom measures were relevant to their clinical experience described the discussion about the scores as being “administrative”, and that the clinician did not seem to understand how the questionnaire was merely “a vague measure… quantifying what is quite an emotional thing” (P95, SI).

Clinicians’ And Service Users’ Perspectives

The clinician-paired participants were all at caseness at referral, and all achieved reliable improvement by the end of treatment (Appendix O and P). All but one participants had at least six face-to-face appointment with their clinicians (P15 had two digital therapy sessions). Clinicians were assigned a matching code with their paired service users, thus C1 referred to the clinician of the participant P1. Table 24 summarised findings from a comparative analysis of information drawn from clinician survey (CS), service user survey (SS) and interview (SI). Areas with conflicting views between clinicians and service users were highlighted. Cases were ordered based on the degree of overlap within the clinician-service user pair.

Perceived Factors Underlying Questionnaire Responses Beyond Symptom Changes

Changes in life circumstances were perceived by the clinician as the main factor underlying responses to MDS measures beyond symptom change, while none of their service users mentioned this as their considerations. Instead, the major influencing factors reported by their service users were response difficulties and immediate context. Only one clinician (C14) realised that their service user might have had response difficulty in capturing the intensity of their symptoms with the options available in the measures. Interpersonal concerns were mentioned by only two of the clinicians but by half of the service users.

The majority of the clinicians expected life circumstances to be an important consideration for their service users’ response to the GRI question, and three also mentioned the attainment of therapy goals. A larger overlap between service users and clinicians’ views was observed in GRI than in MDS, with seven of the cases at least in partial agreement. Half of the service users indeed considered changes in their life circumstances, and six described therapy-related factor such as changes in coping strategies and perception of life in assessing their improvement. None of the clinicians was aware of the influence of current mood, response difficulties, contextual and interpersonal factors.

Eight clinicians were unsure which measures would better reflect their service users' subjective state, whereas nine of their service users had a definite preference. Service users who preferred the MDS found the specificity of the measure useful in prompting self-awareness of their symptoms, and those that favoured GRI felt it was more personal and could guide them to reflect on their overall wellbeing beyond their conditions. Even those who indicated a preference for the MDS still found it valuable to be asked the GRI question.

Match Between Changes In The MDS Scores And GRI Ratings

All but one participating clinician was positive about how well the changes in the MDS scores mapped onto the service users' sense of wellbeing. In a broad sense, most participants agreed that the changes in MDS scores matched their subjective sense of improvement, but emphasised the importance of understanding the changes in both the MDS and GRI measures in "the context of your appointment to get a proper picture" (P14, SI). Those that expressed greater ambivalence reported previous experience of having had their scores being misinterpreted and led to clinical conclusions or actions that they thought were unwarranted.

In the open responses, seven of the clinicians elaborated on how they judged the match between changes in the MDS scores and their perceived level of distress in their service users. Four mentioned taking into considerations service users' subjective report, of which, two also based on achievements of therapy goals. Both clinicians and service users might infer the perceived or subjective sense of wellbeing from changes in the MDS scores. For instance, service users described that "if I noticed the [MDS] scores seemed to be [improved], then I felt more confident and willing to say yes, things are better." (P78, SI). In addition, clinicians might apply statistical concepts such as reliable improvement and recovery rate as their judgment criteria of the perceived level of distress in their service users. One clinician (C35) speculated that changes in the MDS measure could be a better proxy of improvement for their service user as it might be less masked by “ongoing struggle with life stressors”. Interestingly, their service user (P35) also had doubts about their subjective sense of improvement, but attributed this to their judgment at the time being clouded by the anxiety of leaving therapy.

Discussion Of Scores And Shared Understanding

Overall, most clinicians and service users reported having had a discussion about changes in the measures, but there was a marked difference in how the scores were discussed. Service users that had a higher level of shared understanding with their clinicians described a more collaborative process in co-constructing meaning behind their questionnaire response. Such conversations might include exploring the context in which the symptoms happened, or checking the validity of their response. Whereas those who felt that a discussion was absent or insufficient described the scores as being interpreted by clinicians or external parties (e.g. GP) in a prescriptive manner without consulting their own views.

In general, clinicians and service users expected a high degree of shared understanding from each other. Even when both of the pair expressed an ambivalent view (e.g. C50 – P50), they were in good agreement on where the discrepancy might lie. For instance, the clinician (C50) who shared an ambivalent view with their service user (P50) observed the “conservative” attitude of their service user towards the measures and reflected on the insufficient discussion about the utility and validity of the MDS measures. These concerns mapped precisely onto the service user’s (P50) view in hoping that they could have had more discussion about the response difficulties they had with the MDS measures. Similarly, another service user (P35) who reported often being overwhelmed by response difficulties felt that a conversation around how the use of questionnaire scores would help to guide their response process. Both occasions suggested that a discussion about the validity and the intended use of the measures would have mitigated some of the negative impacts. Notwithstanding, discussion about questionnaires does not have to be extensive. Service users who had a greater shared understanding with their clinicians shared the usefulness of having a detailed conversation at the beginning to determine the validity of the measures in relation to their symptom presentations, and to define a way of responding and understanding that is most relevant to the therapy. Subsequent discussion should be shorter with the goal to “get us on the same page quicker and more efficiently” (P91, SI).

In the case where both expected a mismatch (C78 – P78), the difference was viewed positively by the service user as they believed the clinician would be more able to utilise their expert knowledge in drawing out the clinical pattern and offered a different perspective on their conditions. Likewise, their clinician (C78) expected a mismatch based on their observation that changes in the service users’ (P78) MDS scores preceded their subjective report of improvement. It seemed that the different criteria they used to determine improvement could serve a therapeutic purpose.

In addition to topics highlighted in clinician surveys, other factors such as treatment intensity, session length, memory (as inferred by discharge date) and match between presenting problem and MDS measures were also considered (Table 25) to exclude or include alternative explanations of the findings. Amongst which, perceived therapeutic use and attribution of change to therapy seemed to have relevance in the degree of agreement. For a more comprehensive exploration see Appendix Q.

Table 24

Comparative Analysis Of Clinicians' And Service Users' Views

	Participant Code	**Match between Score Change and Perceived/ Subjective Sense of Wellbeing**		**GRI vs MDS**		**Other factors influence Response to MDS**		**Other factors on general sense of wellbeing**		**Discuss scores with each other**		**Shared understanding of score change**	
	Data Source	***CS***	***SI***	***CS***	***SS, SI***	***CS***	***SS, SI***	***CS***	***SI***	***CS***	***SS, SI***	***CS***	***SI***
Agreement	**P89**	O, SR, T	O	NS	MDS	IP		T	T	O	O	O	O
	P90	O	O, SC	NS	NS	L, M		T	T, M	O	O	O	O
	P14	O	O	NS	MDS	R	R, C	L	R, T, L, M	O	O	O	O
	P100	O	O, SC	NS	MDS	L	C	L	L, T, M, C	O	O	O	O
	P78	O, RCI	O, SC	NS	GRI	L	R, IP	L	L	O	O	X	X
	P50	Δ, T, SR	Δ	GRI	GRI	IP	R, IP, C	L	R, T	O	X	Δ	Δ
	P35	O, C	O, C, SC	NS	MDS	L	R, C	L	R, M, C	X	O	O	O
	P91	O, RCI	X	GRI	GRI	L	R, C, IP	L	T, IP	O	O	O	O
	P15	O, SR	O	NS	MDS	L	R, M, IP	L	L, M	X	X	O	N/A
Mismatch	**P80**	O, RCI, SR	Δ	NS	GRI	M	R, IP	T, L	L, M	O	X	O	O

Notes. Theme and response Coding (O = yes, X = No, Δ = ambivalent views, NS = Not Sure), Underlying factors (R = Response Difficulties, L = Life circumstances, T = Therapy, M = Mood, C = Context, IP = Interpersonal Factors).

Table 25

Other Factors

	Participant Code	**Treatment Type**	**Rank by session length (1 = longest)**	**Rank by discharge date (1 = most recent)**	**Problem Descriptor**	**Perceived Irrelevant Items**	**Perceived primary use**	**Additional notes**
	Data sources	***D***	***D***	***D***	***D***	***SI***	***SI, SS***	***SI***
Agreement	**P89**	S2 CBT	6	5	AD	X	T, S	Felt well supported by the service and clinicians, can respond honestly
	P90	S2 CBT	6	9	DD	X	T, S	Felt well supported by the service and clinicians, can respond honestly
	P14	S3 CBT	1	3	DD	O	T, S, TD	Can respond honestly
	P100	S2 CBT	6	7	AD	X	T, S, SC	Feedback facilitates authentic response
	P78	S2 CBT	6	8	DD	O	T, S, RK	Negative emotions evoked by measures
	P50	S3 CBT	2	2	Other	X	SC, Cl, Δ	Negative emotions evoked by measures
	P35	S3 CBT	3	10	AD/DD	X	S, Δ	Unsure of questionnaire use leads to anxiety
	P91	S3 CBT	5	4	DD	X	S, SC, TD, RK	Hesitate to admit improvement
	P15	Online S2 CBT	7	1	DD	X	Cl	Improvement not due to therapy; edit response due to interpersonal concerns
Mis-match	**P80**	S2 CBT	6	6	DD	X	S, D	Improvement not due to therapy; validity of measures depends on purpose

Notes. Problem Descriptor (AD = Anxiety Disorders; DD = Depressive Disorders); Theme Codes (X = No; O = Yes) Perceived primary use (S = self; T = for therapy; TD = treatment decision; SC = access to service; Cl = for clinician only; RK = to monitor risk; Δ = unsure; D = Diagnosis)

Recommendations From Service Users

In addition to responses to the pre-set questions in the service user survey, many respondents also shared some of their ideas about changes of the measures in the free-text response, as listed below in Table 26.

Table 26

Recommendations From Service Users

Change of Response Categories in MDS	Use Never, Occasionally (i.e. Once or Twice), Usually or All the time
	Use Frequency and Duration instead of Days
	To include the option Everyday
	To break down into week 1 and week 2
Add timeframe to GRI	To specify recent weeks/ last month - to promote acceptance of temporary relapses
Include Open Response	For respondents to add emotions
	To give further description
Use of memory aid	To keep a log of information and experience
Change of Presentation	Calendar view or visual scales
Change of Frequency of Data Collection	To ask questions on a daily basis

Discussion

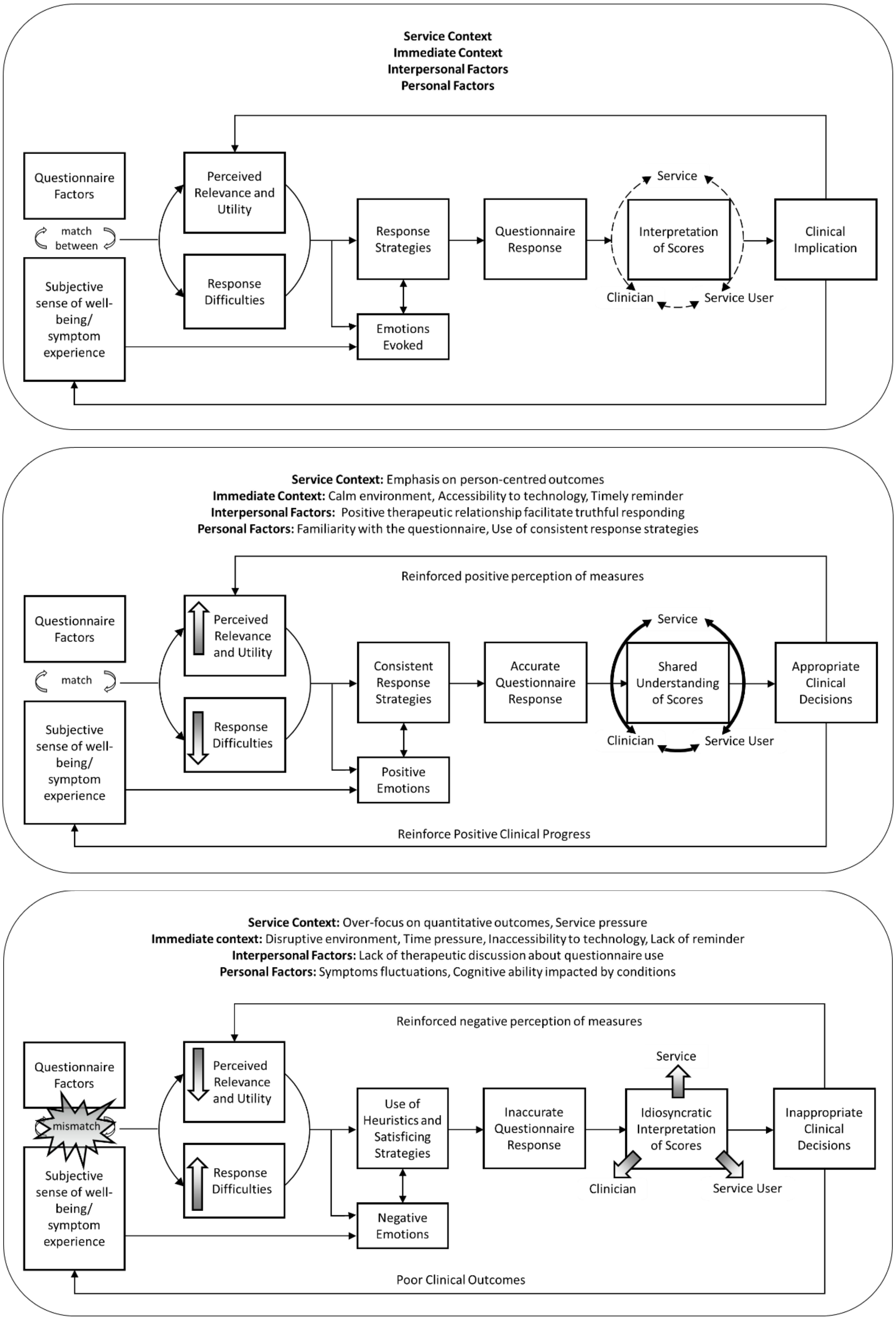
Factors Underlying Questionnaire Response

This study revealed a range of personal, interpersonal, contextual, and questionnaire factors underlying responses to questionnaires used in IAPT services. The results replicated findings from a previous study in general practice, and identified issues with truthful responding to sensitive questions and interpreting double- and triple-barrelled statements in the PHQ-9 (Malpass et al., 2016). Due to the level of response difficulties experienced by the respondents, many expressed uncertainty about the validity of their questionnaire responses, especially when their own views and wider context were not considered in the interpretation of their scores. Our findings here suggested that the choice of response strategies participants adopted in face of the cognitive and emotional challenges of responding to the measures was dictated by the perceived purpose of the questionnaires and the type of response difficulties experienced, and could have a direct impact on responses. The ability to adhere to the same response strategy is crucial for the validity of changes in the MDS scores and the GRI ratings. In addition to uncertainty about questionnaire use, emotion and current mood also impacted on response stability. The role of emotions in the response process was highlighted in our findings. Apart from being the focus of symptom measures, emotions could be evoked by or perpetuate response difficulties. Having a detailed conversation in the beginning of therapy to agree on questionnaire use and address uncertainty might help to mitigate response difficulties and reduce the cognitive and emotional burden on the respondents.

Figure 9 serves as a conceptual illustration of factors and processes identified in this study. A discussion about how this diagram related to existing theories can be found in Appendix R. This framework highlights issues that can arise in different stages of questionnaire response, which can serve as a roadmap to guide clinical use and discussion of the measures. For instance, the initial scene setting conversation can focus on the match between different types of measures and service users' perception of their conditions and wellbeing. After the initial questionnaire use, clinicians may want to explore any response difficulties arise and support the respondents to derive a set of response strategies to facilitate the response process and produce the most informative answers. Clinicians may also encourage their service users to adhere to the same response strategies to improve the validity of changes in their questionnaire scores. The model highlights that emotions are not only as part of the clinical presentation but can also be evoked by the questionnaire and the response process, and would worth clinical considerations. Service context, immediate context, interpersonal and personal factors are included in the background to illustrate their influence on multiple stages of the process. The framework further emphasises the interpretation of questionnaire response or changes in the measures as a collaborative process between service user, clinician and the service. This can then lead to joint decision making and a clinical outcome that is person-centred, fruitful and empowering. Using two opposite hypothetical scenarios, it also illustrates how different factors might interact to give rise to differential outcomes of questionnaire response in a clinical setting.

Figure 9

Conceptual Framework Of Questionnaire Response In A Clinical Setting (Top); A Positive Hypothetical Scenario (Middle); A Negative Hypothetical Scenario (Bottom)



Service Users' And Clinicians' Views On GRI and MDS

The introduction of the GRI question is a positive step towards a more person-centred approach to data collection in clinical settings, and enables meaningful investigation into the relationship between MDS measures and subjective sense of improvement. Both service users and clinicians perceived the GRI question has the potential to tap into important aspects of wellbeing that might not have been previously captured by the MDS measures, such as life circumstances and therapy progress. Findings supported the utility of using both types of measures to provide a holistic picture of their wellbeing.

Comparisons between clinicians and service users' perspectives showed that clinicians might be less aware of the role of response difficulties and over-assumed the influence of life circumstances on their service users' response to the MDS measures. Clinicians might also rely on information only accessible to them (e.g. reliable improvement and recovery criteria) to make judgement about their service users' wellbeing. Comparatively, there was a greater consensus between service users and clinicians on what the GRI scores might be referring to. The comparative analysis also further emphasised on the co-construction of meaning between service users and clinicians for the questionnaire scores to be clinically useful and valuable.

Clinical Implications

Many respondents recalled that the response process was most stressful for them when they filled out the questionnaires for the first time due to their mental states, unfamiliarity with the questionnaire and the service settings. For many, being asked the GRI question can help to validate their progress in a person-centred and positive-focused manner. However, if the question is being asked too frequently, or if fluctuation has not been explored as a feature of the recovery process, service users might feel the pressure to give a positive answer. A positive relationship with the clinician can also hamper truthful responding if the service user perceived the purpose of the question was to evaluate the clinician's performance. The finding that people might underscore on the risk question on PHQ-9 has major implication for risk management. It raised the concern of whether a single question is sufficient to assess such a sensitive and complicated matter, especially given the response difficulties people might experience.

It is suggested that a discussion of the purpose and use of the IAPT MDS questionnaires and GRI prior to them being filled out the first time would facilitate response process and response accuracy. In addition to direct support, IAPT services might consider the use of explanatory text to clarify the purpose of data collection and free text boxes for respondents to raise any issues. Following a thorough conversation about questionnaire use at the start of the clinical engagement, subsequent discussion about changes in the questionnaire scores can be more incisive by building on this shared understanding.

Current mood and recency effects played a significant role in both the MDS response difficulties and GRI rating. It might be worth checking with the service users whether their response to the MDS measure reflect their condition in the past two weeks or only the past few days. For those who are seeing their service users weekly might wish to explore alternative clinical tools that can capture the weekly changes in addition to the MDS. For GRI, clinicians might consider guiding the service user to use a consistent reference point (e.g. a particular time before they started therapy) as an anchor to their response.

To address issues with anxiety around discharge, the service might consider introducing a follow-up period (e.g. one-month) to re-issue the measures. This could allow the service users to make more accurate judgments of their own ability to manage their conditions without the treatment and can consequentially lead to better informed clinical decision making. It might also help intervene earlier to prevent relapse or recurrences of the conditions.

Our findings highlighted concerns respondents have about their questionnaire response being misinterpreted and lead to misinformed clinical decisions. Using a stand-alone score to infer symptom severity and recovery status required a comparison between an individual’s response to a normative data, which might not be valid given the multiple factors underlying individual's questionnaire response. Therefore, while concepts such as caseness, reliable improvement and recovery rate are useful in interpreting data for service evaluation, their clinical utility should be treated with caution. Beyond statistics and numbers, it is always important to be reminded that questionnaire serves as an interface for communication between service users, clinicians and the service. Thus, rather than only seeing response editing and instability as a threat to validity of the questionnaire scores, it should be understood as a manifestation of the respondents' concerns and distress.

To support the implementation of clinical recommendations, the issue would also need to gain wider acknowledgement in the service guidelines. Currently, there are no specific guidelines in the IAPT manual (NCCMH, 2018) on how questionnaires should be used for therapeutic benefits. Most guidance regarding questionnaire use is for improving data quality, apart from a brief mention that a discussion about individual items on disorder-specific measures can help to distinguish if a decrease in symptom levels may be due to avoidance or a genuine reflection of improvement. Future versions of the IAPT manual might expand on the current recommendation by suggesting a detailed discussion on the purpose and utility of the symptom measures in the initial IAPT sessions.

Strengths and Limitations

The current study might serve an important role in bringing to light subjective experiences and meanings into the IAPT setting, where outcomes are more traditionally defined in quantitative terms. Different from most exiting studies, the three-phase approach involving both service users and clinicians allowed this study to capture the complexity of issues around questionnaire response from multiple perspectives. The use of cognitive interviewing within an interpretative framework was crucial to provide a safe and open space for the in-depth exploration of sensitive topics such as response editing and interpersonal concerns. The sample characteristics suggested a possible selection bias (Patel, Doku, & Tennakoon, 2003), in that service users who have benefited from the service might have been more willing to participate in the study than those that did not benefit from treatment. This sample characteristics might generate an overall more positive view on questionnaires. In relation to this, the characteristics and research method of the current study limited exploration of whether response pattern might vary as treatment progresses. The reliance on retrospective recall in this study might make the findings more prone to recall bias. Thus, participants might have remembered the use of questionnaires more positively as their conditions have subsequently improved. Further studies might make use of statistical modelling to investigate the response pattern on both measures over time and its association with symptom severity. This could also add another dimension for data triangulation to determine the extent to which the findings were a result from response bias.

Using online surveys has the benefits of reaching out to a wider sample and collect views that might be more difficult to express in person, but such methods are also susceptible to response effects. Here the researcher attempted to mitigate response difficulties raised by respondents in emails. However, the completion rate of the online survey suggested that around half of the data were excluded due to non-completion. Further studies would benefit from having service user participation and consultation in the research design from the outset. This might have changed what questions are being asked, how they are being understood and the way the study is being conducted.

Having the interview in phase two helped to validate findings from phase one and elicited additional information that were not captured in the survey format. It reiterated the benefits of using multiple qualitative methods in a single study. At the same time, there might be issues with comparability between data drawn from the service user interviews and clinician surveys due to the different modes and depth of investigation. Future studies can consider conducting interviews with both clinicians and service users for better comparability. Having the service user survey as the initial phase of study might be more appealing for people that find questionnaire helpful to participate. It could be that an alternative research method, such as having a focus group or interview as the initial phase, might be more able to capture additional struggles people might have with the questionnaire format. Time and locations were mentioned by at least two respondents as the reason why they were not able to attend the interview. There have been changes to guidelines around remote working since this study started, options such as video calls or non-office hours slots could have made the interviews more accessible for people who have to work full-time or do not live close to the interview sites. Overall, these limitations mean that the current findings might be less reflective of the views of people who might hold a less favourable view of the service or questionnaire, or those who found the study method hard to engage, and making these results less generalisable to the overall IAPT service population.

Having the interviews held in one of the IAPT services might give an impression that the researcher was part of the clinical team and make it harder for participants to express negative views about the service or clinicians. This could generate response bias due to demand characteristics. To mitigate this, the researcher explained her independent role and the anonymity of their data in the beginning of the interview. Similarly, clinicians might also have concerns about the research purpose or see the survey as a test of their clinical ability or knowledge. This could have made them more hesitant to participate in the study if they thought the service users have not significantly improved. Future studies might consider allowing respondents to offer their views in a completely anonymised manner, such as through anonymised online feedback form, in order to decrease the influence of the interviewer and the setting.

Future Directions

The advanced development in statistic models such as item response theory and the use of a computerised adaptive testing framework have the potential to decrease response difficulties and increase the perceived relevance of the measures (e.g. Gibbons et al., 2013). This approach can also serve as an effective way to identify suicidal risk (Delgado-Gomez, Baca-Garcia, Aguado, Courtet, &

Lopez-Castroman, 2016). Clinical training might consider including modules on the communicative and interpersonal aspects of questionnaire use, and the potential response difficulties people might encounter in deriving their response.

Conclusion

The current study offered an in-depth qualitative exploration of responses to IAPT MDS questionnaires and the GRI used in two IAPT services. It revealed a range of personal, interpersonal, contextual and questionnaire factors underlying people's response to the MDS measures. It supported the acceptability of GRI question amongst service users and clinicians and suggested the benefits of using both the MDS measures and GRI rating to capture a holistic picture of service users' sense of wellbeing. The importance of communication between service users, clinicians and the service in making sense of questionnaire response was emphasised throughout different stages of the study.

References

Anthony, W. A. (1993). Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. *Psychosocial rehabilitation journal, 16*(4), 11.

Babor, T. F., Brown, J., & Del Boca, F. K. (1990). Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Behavioral Assessment*.

Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly, 71*(2), 287-311. doi:10.1093/poq/nfm006

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative research, 15*(2), 219-234.

Bless, H., & Schwarz, N. (1999). Sufficient and necessary conditions in dual-mode models: The case of mood and information processing.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Braun, V., & Clarke, V. (2014). What can “thematic analysis” offer health and wellbeing researchers? *International journal of qualitative studies on health and well-being, 9*.

Brown, N. R., Williams, R. L., Barker, E. T., & Galambos, N. L. (2007). Estimating frequencies of emotions and actions: a web-based diary study. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*(2), 259-276.

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., . . . Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychol Med, 45*(15), 3269-3279. doi:10.1017/S0033291715001270

Chepp, V., & Gray, C. (2014). Foundations and new directions. *Cognitive interviewing methodology*, 7-14.

Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research, 18*(2), 107-110.

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *The British Journal of Psychiatry, 176*(4), 379-386.

De Smet, M. M., Meganck, R., De Geest, R., Norman, U. A., Truijens, F., & Desmet, M. (2020). What "good outcome" means to patients: Understanding recovery and improvement in psychotherapy for major depression from a mixed-methods perspective. *J Couns Psychol, 67*(1), 25-39. doi:10.1037/cou0000362

De Smet, M. M., Meganck, R., Van Nieuwenhove, K., Truijens, F. L., & Desmet, M. (2019). No Change? A Grounded Theory Analysis of Depressed Patients' Perspectives on Non-improvement in Psychotherapy. *Front Psychol, 10*, 588. doi:10.3389/fpsyg.2019.00588

Delgado-Gomez, D., Baca-Garcia, E., Aguado, D., Courtet, P., & Lopez-Castroman, J. (2016). Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. *Journal of affective disorders, 206*, 204-209.

Dillman, D. A., & Redline, C. D. (2004). Testing Paper Self-Administered Questionnaires: Cognitive Interview and Field Test Comparisons. *Methods for testing and evaluating survey questionnaires*, 299-317.

DoH. (2010). The NHS Outcomes Framework 2011/12. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213789/dh_123138.pdf

Dowrick, C., Leydon, G. M., McBride, A., Howe, A., Burgess, H., Clarke, P., . . . Kendrick, T. (2009). Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ, 338*, b663. doi:10.1136/bmj.b663

Fayers, P. M., Langston, A. L., Robertson, C., & Group, P. T. (2007). Implicit self-comparisons against others could bias quality of life assessments. *Journal of clinical epidemiology, 60*(10), 1034-1039.

Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment, 25*(2), 520.

Gibbons, R. D., Hooker, G., Finkelman, M. D., Weiss, D. J., Pilkonis, P. A., Frank, E., . . . Kupfer, D. J. (2013). The CAD-MDD: A computerized adaptive diagnostic screening tool for depression. *The Journal of clinical psychiatry, 74*(7), 669.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*: Sage Publications.

HMG/DH. (2011). No Health Without Mental Health: A Cross-Government Mental Health Outcomes Strategy for People of All Ages.

Hobbs, C., Lewis, G., Dowrick, C., Kounali, D., Peters, T. J., & Lewis, G. (2020). Comparison between self-administered depression questionnaires and patients' own views of changes in their mood: a prospective cohort study in primary care. *Psychol Med*, 1-8. doi:10.1017/S0033291719003878

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Krosnick, J. A., Alwin, D. F., & Cannell, C. (1987). Satisficing: A strategy for dealing with the demands of survey questions.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation, 1996*(70), 29-44.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity, 47*(4), 2025-2047.

Lewis, G., & Killaspy, H. (2014). Getting the measure of outcomes in clinical practice. *Advances in Psychiatric Treatment, 20*(3), 165-171.

Malpass, A., Dowrick, C., Gilbody, S., Robinson, J., Wiles, N., Duffy, L., & Lewis, G. (2016). Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *Br J Gen Pract, 66*(643), e78-84. doi:10.3399/bjgp16X683473

McPherson, S., Evans, C., & Richardson, P. (2009). The NICE Depression Guidelines and the recovery model: Is there an evidence base for IAPT? *Journal of Mental Health, 18*(5), 405-414. doi:10.3109/09638230902968258

NCCMH. (2018). The Improving Access to Psychological Therapies Manual. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual.pdf

NHS Digital. (2019). Psychological Therapies: Annual report on the use of IAPT services (England 2018-19). Retrieved from https://files.digital.nhs.uk/1C/538E29/psych-ther-2018-19-ann-rep.pdf. https://files.digital.nhs.uk/1C/538E29/psych-ther-2018-19-ann-rep.pdf

Norman, S., Dean, S., Hansford, L., & Ford, T. (2014). Clinical practitioner's attitudes towards the use of Routine Outcome Monitoring within Child and Adolescent Mental Health Services: a qualitative study of two Child and Adolescent Mental Health Services. *Clin Child Psychol Psychiatry, 19*(4), 576-595. doi:10.1177/1359104513492348

Patel, M. X., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment, 9*(3), 229-238.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health services research, 34*(5 Pt 2), 1189.

Rapkin, B. D., & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and quality of life outcomes, 2*(1), 14.

Rasinski, K. A., Baldwin, A. K., Willis, G. B., & Jobe, J. B. (1994). *Risk and loss perceptions associated with survey reporting of sensitive behaviors.* Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Robertson, C., Langston, A. L., Stapley, S., McColl, E., Campbell, M. K., Fraser, W. D., . . . Group, P. T. (2009). Meaning behind measurement: self-comparisons affect responses to health-related quality of life questionnaires. *Qual Life Res, 18*(2), 221-230. doi:10.1007/s11136-008-9435-1

Robinson, J., Khan, N., Fusco, L., Malpass, A., Lewis, G., & Dowrick, C. (2017). Why are there discrepancies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open, 7*(4), e014519. doi:10.1136/bmjopen-2016-014519

Sajobi, T. T., Brahmbatt, R., Lix, L. M., Zumbo, B. D., & Sawatzky, R. (2018). Scoping review of response shift methods: current reporting practices and recommendations. *Qual Life Res, 27*(5), 1133-1146. doi:10.1007/s11136-017-1751-x

Sandelowski, M. (2011). When a cigar is not just a cigar: Alternative takes on data and data analysis. *Research in nursing & health, 34*(4), 342-352.

Schwartz, C. E., Finkelstein, J. A., & Rapkin, B. D. (2017). Appraisal assessment in patient-reported outcome research: methods for uncovering the personal context and meaning of quality of life. *Qual Life Res, 26*(3), 545-554. doi:10.1007/s11136-016-1476-2

Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American psychologist, 54*(2), 93.

Simons, L., Lathlean, J., & Squire, C. (2008). Shifting the focus: sequential methods of analysis with qualitative data. *Qualitative Health Research, 18*(1), 120-132.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine, 166*(10), 1092-1097.

Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine, 48*(11), 1507-1515.

Stone, A. A., Broderick, J. E., Schwartz, J. E., & Schwarz, N. (2008). Context effects in survey ratings of health, symptoms, and satisfaction. *Medical care, 46*(7), 662.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin, 133*(5), 859.

Truijens, F. L., Desmet, M., De Coster, E., Uyttenhove, H., Deeren, B., & Meganck, R. (2019). When quantitative measures become a qualitative storybook: a phenomenological case analysis of validity and performativity of questionnaire administration in psychotherapy research. *Qualitative Research in Psychology*, 1-44.

Unsworth, G., Cowie, H., & Green, A. (2012). Therapists' and clients' perceptions of routine outcome measurement in the NHS: A qualitative study. *Counselling and Psychotherapy Research, 12*(1), 71-80. doi:10.1080/14733145.2011.565125

Weiss, D. S. (2007). The impact of event scale: revised. In *Cross-cultural assessment of psychological trauma and PTSD* (pp. 219-238): Springer.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*: Sage publications.

Wolpert, M., Curtis-Tyler, K., & Edbrooke-Childs, J. (2016). A Qualitative Exploration of Patient and Clinician Views on Patient Reported Outcome Measures in Child Mental Health and Diabetes Services. *Adm Policy Ment Health, 43*(3), 309-315. doi:10.1007/s10488-014-0586-9

Yin, R. K. (1981a). The case study as a serious research strategy. *Knowledge, 3*(1), 97-114.

Yin, R. K. (1981b). The case study crisis: Some answers. *Administrative science quarterly, 26*(1), 58-65.

Part 3: Critical Appraisal

This thesis has focused on the meaning of questionnaire responses, and the empirical study itself also carries meaning in the wider context. In this critical appraisal I would like to focus on topics that I consider as pertinent to my learning from this study and the research process, these are: (a) the choice of research method, (b) a qualitative study in a quantitative oriented setting, (c) Computerised Adaptive Testing and (c) ethical concerns of data usage. However, before I shared my critical reflections on these areas, I would like to contextualize my views by offering a reflexive statement.

Reflexive Statement

During my clinical psychology training, I had a year-long placement in my first year in one of the sites I then conducted my thesis research study in. Returning to my previous placement setting afresh, as a researcher and a final year trainee, was a unique experience. My familiarity with the setting simultaneously gave me the advantage of understanding some of the issues first-hand but also made me more prone to personal biases. Keeping a reflexive diary (Braun & Clarke, 2006, 2014; Guest et al., 2011) has allowed me to maintain a reflexive stance in navigating my role as a researcher in the context of a clinical training, and integrating knowledge from the two domains. For instance, I reflected on how my existing relationship with some of the clinicians might lead to unwarranted biases in my data analysis. As a result, to eliminate the influence of the clinicians' identity, I chose not to include clinicians' names in my resulting data. I also wondered whether the concern about anonymity might have made some of the clinicians more reluctant to participate. In hindsight, this could have been better managed by having another researcher rather than myself contact clinicians that I knew from the time of my placement. Despite these challenges, to have a research project in the setting where I started my clinical training provided a rare opportunity for closure for me, thus making this study carry more personal significance than it would have otherwise done.

The Choice Of Research Method

Two major principles, methodological pluralism and pragmatism, have formed the basis of the methodological decision for this study. Methodological pluralism acknowledges the value in various research approaches, and pragmatism posits that different approaches can help to address different research questions. Based on these views, the choice of a qualitative approach is guided by the research interest in meaning and people's subjective experience around their questionnaire responses (Pistrang & Barker, 2012). There is, however, more nuance in the decision amongst different qualitative methods. Considerations about practicality, previous studies and other qualitative approaches all come into play.

Practical aspects such as manpower, time and room availability posed major limitations on the scope and scale of the study. To maintain the consistency of the interview procedure, I was the only interviewer throughout the study. The arrangement of interviews was a negotiation among my training schedule, room availability at the research site, and the participants' availability. This condition made it harder for some participants (e.g. those who are working 9to-5, or those that do not live in close proximity to the research sites) to attend the interviews. During the Covid-19 pandemic, two interviews were held over the phone. On reflection, telephone interviews could have been offered as an option to make the study more accessible for a wider range of participants. Yet, in practice it was slightly more difficult without the visual guidance of which questionnaires the participants were referring to during the interviews. With the increased use and familiarity with digital teleconferencing technology in the population during the pandemic, perhaps it might be easier for future interviews to be carried out via video-calls. The decision to have the clinicians to fill out a survey instead of doing an interview was also due to practical concerns. It would have been unlikely, given the length of the study period, that I would have been able to conduct interviews with each of the pairs of service users and clinicians, and fully transcribe and analyze the data in the necessary depth. Further, it is unlikely that the requisite number of clinicians would have been afforded the time to partake in full qualitative interviews, particularly given additional time constraints as a result of the pandemic.

Cognitive interviewing is a common approach in research into questionnaire responses and has been used in previous studies in similar settings (e.g.Malpass et al., 2016; Robinson et al., 2017). There are various approaches to conducting cognitive interviews, such as the think aloud procedure, and with the use of retrospective probes. The choice of using retrospective recall over a think aloud protocol is made when the interest is in the respondents' own perspective of the overall experience of questionnaire response, not just their struggles with specific wordings or the available response options. In addition, for the context of this study where participants were service users of a routine care setting, the think aloud procedure would not have afforded the ability to capture the contextual influence during the time when the participants were still using the IAPT service. By using an interpretive framework (Chepp & Gray, 2014), it brings in a non-judgmental stance to the cognitive interview, which fits in better with the person-centred focus on individual experience and is more in line with the nature of the self-reported measures where "response accuracy" is defined in subjective terms. The combined use of service user surveys and interviews helped to validate the results from each phase and build a stronger argument with convergent findings.

Within the field of qualitative research, there are multiple approaches that address the issue of meaning, with the most apparent comparison drawn between the current approach and that of an interpretative phenomenological analysis (IPA; Smith & Shinebourne, 2012). IPA consists of a clear outline of the theoretical orientation and research protocol, and aims to provide a more in-depth understanding of a smaller sample. Although IPA would have served the purpose of providing a detailed investigation of the research question, it requires a limited sample size and a very specific focus. As a preliminary exploration, this study would aimed to capture a broader spectrum of experience amongst IAPT service users so it was necessary to both have a greater breadth of focus and to capture the opinions and experiences of more than the handful of participants that would have been involved if the study were run in an IPA framework. Additionally, the study was not only interested in the meaning itself, but also the way meaning is derived as part of the response process. Reflexive thematic analysis (Braun & Clarke, 2006) has its advantage in being able to be flexibly applied to data generated from surveys and cognitive interviews (Guest et al., 2011). This means that the approach is able to allow the research to draw on quantitative aspects apparent from the surveys or interviews, such as the frequency of occurrence of each theme, to illustrate its relative importance in the study sample.

A Qualitative Study In A Quantitative-Oriented Setting

The quantitative focus of the IAPT setting is evident in a number of ways, including the selection of Key Performance Indices (such as waiting time, percentage of people being seen, recovery rate, etc.), the emphasis on randomized controlled trial evidence as the “ideal” basis for evidence based practice in IAPT (NCCMH, 2018), and the predominantly quantitative nature of relevant studies informing the effectiveness of both the interventions used in IAPT and of the stepped care model upon which IAPT services are structured (e.g. Clark, 2011; Clark et al., 2018; Gyani, Shafran, Layard, & Clark, 2013; Radhakrishnan et al., 2013). Questionnaire measure outcomes are interpreted in statistical terms using concepts such as ‘caseness’ and ‘reliable change index’ to infer symptom severity and recovery status. Our qualitative findings revealed a complex and intricate dynamic behind the quantitative phenomena and highlighted the role of collaborative clinical discussion to make the data meaningful and relevant. It also captured service users’ voice, a fundamental aspect of person-centred care that is arguably yet to be sufficiently covered in publications using IAPT data and national reports on the performance of IAPT services. For the findings to be fully utilised to benefit the service and its stakeholders, they will have to be presented and received well in a quantitative-oriented setting. Several features in this study might help to make it more appealing to a quantitative mind-set. For example, having a comparatively larger scale study as opposed to small scale in-depth qualitative investigation might give more weight to the findings in the minds of the stakeholders. While the inclusion of the occurrence of themes might not be a common practice in reflexive thematic analysis (Braun & Clarke, 2006, 2014; Clarke & Braun, 2018), the use of number could have its advantage in conveying the prevalence of each theme (Guest et al., 2011). The use of tables and diagrams is also an attempt to summarise the findings in a clear and concise manner that can read well in a quantitative-oriented setting.

Another route of dissemination is through the integration of recommendations into clinical training. Drawing on my own training experience, the potential benefit of routine outcome measurement is widely acknowledged from a researcher's, clinician's or service's perspective. Less commonly explored are the emotions or thoughts that might be evoked for service users by the questionnaires. There is also insufficient emphasis on the need to validate the service users' response. It is important to bring clinicians' awareness to the idea that a score can have multiple meanings on an individual and group level. On a group level interpretation of data, individual differences or idiosyncratic ways of responding are considered as random errors. Concepts such as caseness and recovery status are generated from a comparison between individual and normative data. However, in clinical settings, those "random errors" or individual differences are what make each service user unique as a person. The only way to make sense of it in a clinical context is through a collaborative discussion with the service user about the score in relation to their therapy goal(s). Although the importance of shared decision making has been highlighted in the clinical training materials (e.g. Richards & Whyte, 2009), explicit focus on how to manage scenarios where the questionnaire outcomes and subjective reports dispute would be a beneficial addition.

Computerised Adaptive Testing

In addition to human factors, questionnaire features such as the relevance of specific items and questionnaire length also contribute to the response burden. The proliferation of statistical advancement in questionnaire development gives a promising future of tailoring measures based on individual needs. The application of item response theory might allow individual questionnaires to be constructed from a calibrated item bank. Algorithms and models such as computerised adaptive testing and decision trees can assist the item selection process by suggesting subsequent items based on the respondents’ previous answers and on pre-set criteria (Unick, Shumway, & Hargreaves, 2008). Existing research has applied these models to depression, anxiety and mental health measures, and demonstrated that it is entirely possible to retain measurement precision with a significantly smaller number of questionnaire items (Fliege et al., 2005; Gibbons et al., 2013; Gibbons et al., 2008). As the item banks collate items across different measures of a specific domain, they can also offer the flexibility to elicit more detailed responses by presenting the respondent with more follow-up questions when needed. Our current finding showed that the single item on suicidal risk on PHQ-9 might not be able to accurately reflect suicidal risk, and could lead to underscoring due to concerns about clinical consequences. By drawing across different measures on suicidality, it is possible to include discriminative items that are less direct (e.g. “I often feel “empty” inside), and might help to more accurately identify people who are at risk. Research has indeed showed that it is possible to precisely identify suicidal behaviours with a four-item measure using a decision tree model (Delgado-Gomez et al., 2016). In addition to assessment and symptom monitoring, our research also showed the benefit of using measure that highlight positive aspects of recovery. Computerised adaptive testing has also been applied to quality of life measures to capture aspects that are pertinent to an individual’s recovery process alongside condition-specific items (Paap, Kroeze, Terwee, van der Palen, & Veldkamp, 2017). With the advancement of statistical procedures, it is hoped that the future of routine outcome measures can give the flexibility to clinicians and service users to personalise a set of items that would be most relevant and responsive to the clinical presentation and therapy goals, thus alleviating response burden, improving response validity and allowing comparison across different measures and clinical presentations.

Ethical Concerns About Data Usage

Service users are the information providers for the routine outcome measures. The interest in service users’ perspectives on questionnaire response and the emphasis on informing them about data usage is not only of clinical importance, but also of ethical necessity. Our study period coincided with the introduction of the General Data Protection Regulation (GDPR). As a result, there were changes to the research protocol (such as using a different online survey platform that allowed data to be exported direct from the secured cloud drive on the survey site to UCL Data Safe Haven) and additional statements were included in the Patient Information Sheet regarding data usage and their rights. Although this led to a delay in the study process, I personally thought it was much a needed procedure to protect personal data and the rights of information givers.

GDPR also has impacted on the data collection practices of IAPT services. Additional guidelines and statements have been written about personal data rights (NHS Digital, 2019b, 2020b). GDPR outlines that people should have the right to be informed, the right to get access to their data, the right to rectification and erasure, the right to restrict processing, the right to data portability, the right to object, and the rights in relation to automated decision making and profiling (Information Commissioner's Office, 2020a, 2020b). The extent to which information givers are granted these rights are based on the legal basis of data collection. The NHS guidelines on data usage specifies that “the IAPT data is to be used for secondary care rather than direct patient care, patients using NHS-funded IAPT services must be made aware that their confidential data will be used for this beneficial, additional purpose i.e. to improve care.” (NHS Digital, 2020b). The guidelines focus almost exclusively on data usage beyond clinical settings, and outline the rights of service users to opt out from having their data used for secondary purposes. Yet, from the service user’s perspective, it could be more relevant for them to be made aware of how their own care will be directly or indirectly impacted by actions or decisions made based on the information they provide. The current guideline specifically denies service users’ right (a) to erase or remove their data, (b) to move, copy or transfer it, (c) to object to it being processed or used, and (d) to know if a decision was made by a computer rather than a person (NHS Digital, 2020a). Based on the GDPR, the service users should have the right to be informed how their information is being used and the limitations to their rights. However, from our findings, it does seem that such conversation rarely happens, or at least rarely happens to a sufficient level, before data are collected. In addition, it is unclear whether this guidance is set up specifically for secondary use of data, or if it also applies to clinical practice. To date, there is no equivalent guidance on the primary use of data in IAPT settings. In line with the growing awareness in people's right to their personal data, perhaps it would be something the service might look to develop in the near future.

Conclusion

In this section I have shared some of the reflections and thoughts I have had about the methods and findings of my empirical study, specifically these were related to the choice of the methods utilised, the difficulty in presenting qualitative findings in a quantitative-oriented service setting, the development of computerised adaptive testing and the socio-political context around the use of routine outcome measurement data. It is hoped that the future of IAPT and data-informed health care practices can become more aligned with the ethos of person-centred care, foregrounding the importance of service users' opinions and concerns about the data they provide to NHS services. I would also like to envision that the leading force in healthcare research (e.g. NICE guidelines) would take a methodological pluralism and pragmatism perspective in acknowledging the utility of and evidence generated from both qualitative and quantitative approaches.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Braun, V., & Clarke, V. (2014). What can "thematic analysis" offer health and wellbeing researchers? *International journal of qualitative studies on health and well-being, 9*.

Chepp, V., & Gray, C. (2014). Foundations and new directions. *Cognitive interviewing methodology*, 7-14.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *Int Rev Psychiatry, 23*(4), 318-327. doi:10.3109/09540261.2011.606803

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet, 391*(10121), 679-686.

Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research, 18*(2), 107-110.

Delgado-Gomez, D., Baca-Garcia, E., Aguado, D., Courtet, P., & Lopez-Castroman, J. (2016). Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. *Journal of affective disorders, 206*, 204-209.

Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of life research, 14*(10), 2277.

Gibbons, R. D., Hooker, G., Finkelman, M. D., Weiss, D. J., Pilkonis, P. A., Frank, E., . . . Kupfer, D. J. (2013). The CAD-MDD: A computerized adaptive diagnostic screening tool for depression. *The Journal of clinical psychiatry, 74*(7), 669.

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., . . . Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*(4), 361-368.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*: Sage Publications.

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behav Res Ther, 51*(9), 597-606. doi:10.1016/j.brat.2013.06.004

Information Commissioner's Office. (2020a). Guide to the General Data Protection Regulation (GDPR). Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

Information Commissioner's Office. (2020b). Your data matters. Retrieved from https://ico.org.uk/your-data-matters/

Malpass, A., Dowrick, C., Gilbody, S., Robinson, J., Wiles, N., Duffy, L., & Lewis, G. (2016). Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *Br J Gen Pract, 66*(643), e78-84. doi:10.3399/bjgp16X683473

NCCMH. (2018). The Improving Access to Psychological Therapies Manual. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2018/06/the-iapt-manual.pdf

NHS Digital. (2019, 21 May 2019). Transparency notice: how we use your personal data. Retrieved from https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register

NHS Digital. (2020a). GDPR Register - Improving Access to Psychological Therapies (IAPT). Retrieved from https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register/improving-access-to-psychological-therapies-iapt

NHS Digital. (2020b, 23 April 2020). IAPT data set: fair processing guidance. Retrieved from https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/improving-access-to-psychological-therapies-data-set/iapt-data-set-fair-processing-guidance#suggested-wording-for-fair-processing-information

Paap, M. C., Kroeze, K. A., Terwee, C. B., van der Palen, J., & Veldkamp, B. P. (2017). Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Quality of life research, 26*(11), 2909-2918.

Pistrang, N., & Barker, C. (2012). Varieties of qualitative research: A pragmatic approach to selecting methods.

Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., & Lafortune, L. (2013). Cost of improving Access to Psychological Therapies (IAPT) programme: an analysis of cost of session, treatment and recovery in selected Primary Care Trusts in the East of England region. *Behav Res Ther, 51*(1), 37-45. doi:10.1016/j.brat.2012.10.001

Richards, D., & Whyte, M. (2009). *Reach Out: National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions*: Rethink Mental Illness.

Robinson, J., Khan, N., Fusco, L., Malpass, A., Lewis, G., & Dowrick, C. (2017). Why are there discrepancies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open, 7*(4), e014519. doi:10.1136/bmjopen-2016-014519

Smith, J. A., & Shinebourne, P. (2012). *Interpretative phenomenological analysis*: American Psychological Association.

Unick, G. J., Shumway, M., & Hargreaves, W. (2008). Commentary: Are We Ready for Computerized Adaptive Testing? *Psychiatric Services, 59*(4), 369-369.

Appendix A. Data Flow

Screening eligible service users

- Searching through IAPTus database for eligible service users who meet the inclusion criteria and have given consent to be contacted for research purpose.
  - Carried out by PI or other members of the research team on-site in the IAPT services on secured computer provided by the trust
  - The list of IAPTus number of eligible service users will be generated on the trust computer and stored on an encrypted and secured electronic USB device provided by UCL
  - The list of IAPTus number of eligible service users will be uploaded to UCL data safe haven and deleted from the USB device

Invitation and distribution of survey

- Based on the information on IAPTus, eligible service users will be contacted via their preferred means of contact (Email or phone)
- If contacted by phone, eligible service users will be asked for a contact email address for the online REDCap survey link to be sent
- A list of emails of eligible service users will be generated and stored on the secured USB device and UCL data safe haven
- Invitation emails are sent securely via REDCap with a personalized link to the survey generated automatically by uploading the list of distribution emails onto REDCap
- Email addresses of eligible service users will have to be entered on the REDCap site.
- Their email addresses will be removed if they did not respond to the survey after 3 email invitations were sent or if they decline to participate.

Survey Data Collection and Storage

- Participants will fill out the survey following the personalized link included in the invitation email
- Participants will be presented with the Participant Information Sheet followed by the consent form
- They will have to agree to the terms outlined on the consent form in order to proceed to the survey questions
- Survey responses will be stored and uploaded directly onto UCL Safe Haven the UCL data safe haven

Invitation of Interview Participants

- Survey participants who have indicated willingness to be contacted for an individual follow-up interview will be invited via email previously used in sending their online survey
- If there is no further response received after 3 email contacts or if the participants change their mind and are no longer willing to take part in the interview, no further contact will be made
- Their participation or not in the interview will have no impact on the data collected in their survey unless they indicate a complete withdrawal from the study
- The researcher will provide their contact number but the main means of contact will be by email. No clinical information will be included in the email apart from practical arrangement for the interviews.

Interview Data Collection

- Interviews will be conducted in individual room in UCL building at Torrington Place or in one of the IAPT services involved in the research study during official working hours (9am – 5pm).
- Interviews will be recorded on an encrypted voice recorder provided by UCL
- The audio recording will be stored under the participant number and transferred onto a secured and encrypted USB device on a secured trust or UCL computer
- The audio recording will be deleted from the device once being uploaded to UCL Data Safe Haven

Interview Data Processing and Storage

- The audio recording will be used to transcribe the interview, all personal identifiable information will be removed from the verbatim. The audio recording will be deleted from the UCL Data Safe Haven once being transcribed.
- Transcription will be done by PI or other members of the research team ant the verbaatim will be stored on the secured and encrypted USB device as well as uploaded to UCL Data Safe Haven using the unique participant number

Invitation of Clinician Survey

- With the consent from Interview participants, the treating clinicians of the participants will be invited by email to fill out an online survey on REDCap.
- The email to clinicians will be sent between NHS accounts and therefore will be secured. The IAPTus number of the interview participants under their previous care will be included in the email for the clinicians to give response in relation to their clinical understanding of the particular participant.
- In the invitation email, the clinicians will be provided an unique code which they have to enter in the survey for the researchers to map their response to the response from the service user participants

Clinician Survey Data Collection

- Participants will fill out the survey following the link included in the invitation email
- Participants will be presented with the Participant Information Sheet followed by the consent form
- They will have to agree to the terms outlined on the consent form in order to proceed to the survey questions
- They will have to enter to unique code provided in the invitation email for the researchers to map their response to the response from the service user participants
- Their responses to the survey will be stored directly on UCL Safe Haven and are mapped onto their email address in this database

Clinician Survey Data Storage

- The clinician survey response will be stored under the unique participant number previously assigned to the service user participants. This will be stored separately from the service user response.
- The processed survey response and outcome measures will be saved on the secured and encrypted USB device provided by UCL, which will be deleted after being uploaded to the UCL data safe haven

Data Analysis and Write Up

- No identifiable information will be contained in the report and it will be stored on password protected personal electronic device and also secured UCL computer.

Dissemination of Result

- The results from this study will be communicated to service users of the participating IAPT services via newsletter to those that have stated that they would like such information and via a presentation to the service user advisory groups.
- Results will be disseminated to clinical staff of IAPT via service-wide and Trust-wide newsletters and in a presentation at one of the IAPT Research Seminars held termly at UCL.
- Results will be included in a doctoral thesis which will be published online via the UCL repository and a paper copy will be held at the UCL Science Library.
- Manuscripts detailing the findings of the study will be submitted for publication in peer-reviewed journals and will be made Open Access wherever possible, using funds from the Wellcome Trust Fellowship Grant of the primary supervisor of this project at UCL. All investigators involved in the study will be invited to be authors/co-authors of any such journal publications.

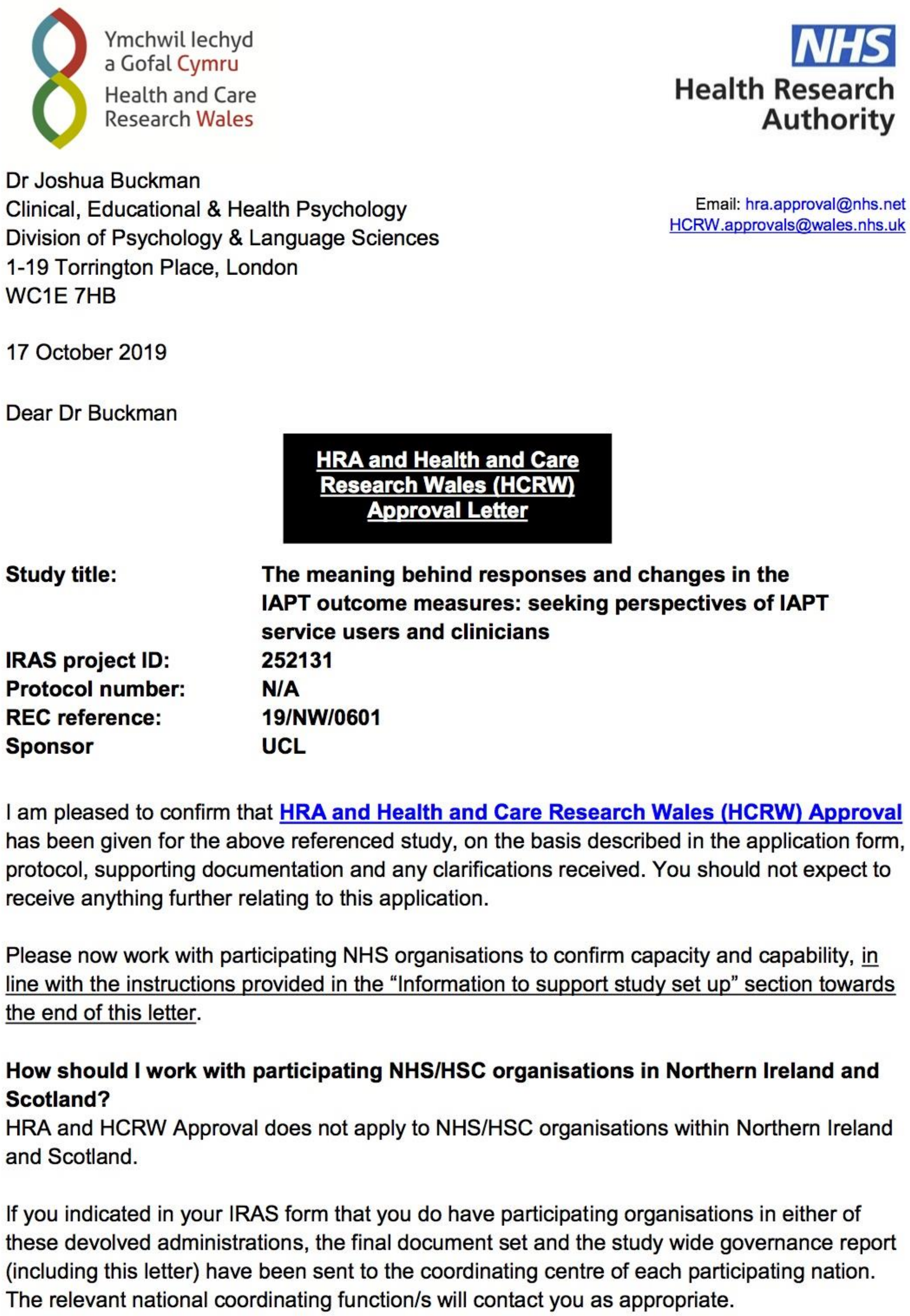
Archiving

- UCL and each participating site recognise that there is an obligation to archive study-related documents at the end of the study (as such end is defined within this protocol).
- The PI or other members from the research team confirm that they will archive the study master file at UCL for the period stipulated in the protocol and in line with all relevant legal and statutory requirements.
- All study documents will be archived for 5 years and in line with all relevant legal and statutory requirements.
- No research document will be stored in individual participating site beyond the study period.

Appendix B. Implementation schedule

Activity	Time in Weeks and Dates	Participant Numbers
Recruit participants for survey, send survey link on REDCap to be completed by participants	24 January - 28 February 2020	n = 121
Begin initial analysis and start recruitment for interview participants	7 February - 6 March 2020	n = 121
Process Survey Response via REDCap and stored on UCL Safe Haven, and begin initial analyses	7 February - 6 March 2020	n = 121
Invite participants to interviews and conduct interviews	14 Februaury - 19 March 2020	n = 121
Contact treating clinicians by email to fill out survey on REDCap	25 February - 8 May 2020	n = 131
Transfer audio recordings of interviews to DSH. Transcribe interview data.	19 March - 6 April 2020	n = 131
Analyse all data, review codes and themes with IAPT clinician	10 March - 17 April 2020	N/A
Report Writing and Dissemination activites	17 April - 21 September 2020	N/A

Appendix C. Ethics Approval Letter



Ymchwil Iechyd
a Gofal Cymru
Health and Care
Research Wales

NHS
Health Research
Authority

Dr Joshua Buckman
Clinical, Educational & Health Psychology
Division of Psychology & Language Sciences
1-19 Torrington Place, London
WC1E 7HB

Email: hra.approval@nhs.net
HCRW.approvals@wales.nhs.uk

17 October 2019

Dear Dr Buckman

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title: **The meaning behind responses and changes in the IAPT outcome measures: seeking perspectives of IAPT service users and clinicians**
IRAS project ID: **252131**
Protocol number: **N/A**
REC reference: **19/NW/0601**
Sponsor **UCL**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?

HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **252131**. Please quote this on all correspondence.

Yours sincerely,

Michael Pate
Approvals specialist

Email: hra.approval@nhs.net

Copy to: *Miss Delasi Apraku – Sponsor contact.*

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Contract/Study Agreement template [Email Template]		22 July 2019
Contract/Study Agreement template [Email Template]	2	14 October 2019
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)		22 July 2019
Interview schedules or topic guides for participants	1	22 July 2019
IRAS Application Form [IRAS_Form_28082019]		28 August 2019
Letter from funder		
Letter from funder		03 May 2019
Letter from sponsor [Confirmation of UCL sponsorship]		14 August 2019
Letters of invitation to participant [Email Invitation for Service Users]	2	14 October 2019
Letters of invitation to participant [Email Invitation for Clinicians]	2	14 October 2019
Non-validated questionnaire [Survey questions_service users]	2	14 October 2019
Non-validated questionnaire [Survey questions_clinicians]	2	14 October 2019
Non-validated questionnaire [Online Survey Logic Flow - Service User]	1	22 July 2019
Non-validated questionnaire [Online Survey Logic Flow - Service User]	2	14 October 2019
Non-validated questionnaire [Online Survey Logic Flow - Clinicians]	1	22 July 2019
Non-validated questionnaire [Online Survey Logic Flow - Clinicians]	2	14 October 2019
Organisation Information Document		
Other [Expenses Claim Form]		
Other [Participant Receipt Form]		
Other [Delegation Log]		
Participant consent form [Consent Form_Service User Survey]	2	14 October 2019
Participant consent form [Consent Form Clinician Survey]	2	14 October 2019
Participant consent form [Consent Form_Service User Interview]	2	14 October 2019
Participant information sheet (PIS) [PIS - Service User Survey]	1	22 July 2019
Participant information sheet (PIS) [PIS_Service Users]	2	14 October 2019
Participant information sheet (PIS) [PIS_Clinician]	2	14 October 2019
Referee's report or other scientific critique report [Peer review report]		19 November 2018
Referee's report or other scientific critique report [Peer review report]		30 September 2018
Research protocol or project proposal [Research Protocol]	2	14 October 2019
Schedule of Events or SoECAT [Assessed]		02 September 2019
Summary CV for Chief Investigator (CI) [Joshua Buckman's CV]		
Summary CV for student [CV - Pandora Lee]		
Summary CV for supervisor (student research) [CV - Supervisor RS 1]		
Summary CV for supervisor (student research) [CV - Supervisor RS 2]		
Validated questionnaire [Examples of IAPT Measures and GRI]	1.0	21 June 2019

IRAS project ID	252131

Information to support study set up

The below provides all parties with information to support the arranging and confirming of capacity and capability with participating NHS organisations in England and Wales. This is intended to be an accurate reflection of the study at the time of issue of this letter.

Types of participating NHS organisation	Expectations related to confirmation of capacity and capability	Agreement to be used	Funding arrangements	Oversight expectations	HR Good Practice Resource Pack expectations
One site type.	**Formal confirmation.** Research activities should not commence at participating NHS organisations in England or Wales prior to their formal confirmation of capacity and capability to deliver the study.	**Organisation information document acts as the agreement** An organisation information document has been submitted and the sponsor is not requesting and does not expect any other site agreement to be used.	No funds are being provided to the site by the sponsor.	Local PI.	Where arrangements are not already in place, research staff not employed by the NHS host organisation undertaking any of the research activities listed in the research application would be expected to obtain a Letter of Access based on standard DBS checks and occupational health clearance.

Other information to aid study set-up and delivery

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales in study set-up.
This study has not been put forward for adoption to the NIHR Portfolio.

Appendix D. Online Service User Survey

Understanding Meaning of Responses to IAPT Measures (Service User Survey)

Please complete the survey below.

Thank you!

IRAS ID: 252131
REC Reference: 19/NW/0601

Participant Information Sheet: Service Users

Title of Project: The meaning behind responses and changes in the IAPT outcome measures: seeking perspectives of IAPT service users and clinicians (Student Study)

Participant Information Sheet

You are being invited to take part in a research study. Before you decide whether to take part, we would like you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to discuss with others if you wish. If there is anything that is not clear or if you would like more information, please ask us.

Summary

A range of questionnaires which measure common symptoms of low mood, anxiety, and any difficulties in doing day-to-day tasks are routinely used in the Improving Access to Psychological Therapies (IAPT) services. Changes in the scores on these measures can have different implications on a clinical, service, and policy level, but based on feedback from our Service User advisory groups we are interested in better understanding what the questions and scores on those measures actually mean for you. More specifically, we hope to find out more about (1) the things influence the way you respond on the questionnaires, (2) how you and your clinician make sense of changes in your scores on those questionnaires, and (3) whether or not the scores reflect your own sense of wellbeing.

To do this, we would like to invite you to fill in a survey that asks you about your thinking processes when you fill in the questionnaires routinely used in IAPT. If you choose to participate in the survey, you can let us know if you are happy to be contacted for the second part of the study. The second phase involves a one-off individual interview with you to explore the issues outlined above in greater depth. In addition, with your consent, your clinicians will be invited to fill out an online survey to offer their perspectives on the changes in your routine IAPT questionnaire measure scores. The results from this study will help clinicians and researchers to better understand what might be being communicated in the responses to routine questionnaire measures, and whether the answers given on those questionnaires and the corresponding scores from those answers match up with how clinicians think about the experiences of their clients, and how our IAPT and the service users think about their overall sense of wellbeing and any improvements in wellbeing throughout therapy. This might then inform how we do things in our local IAPT services in future, or potentially how things are done across a number of IAPT services nationwide.

Why have I been invited?

We are inviting everyone who has given consent to be contacted for research purposes and is receiving treatment or completed their treatment between the end of April and end of October 2019 in the IAPT services in **[Trust location obscured for confidentiality reason]**take part in this study.

Do I have to take part?

Your participation in this study is completely voluntary. There will be no impact on your routine care in this IAPT service regardless of whether you choose to participate or not in this study. If you do decide to participate, you can take part in only the survey or in both the survey and the interview. You are free to withdraw your participation at any time and we will remove all your information from our research data upon your request. You do not need to give any reasons for your withdrawal. If you do not take part, you will still receive your routine care in the IAPT service. You can indicate that you have read through and understand the information on this Patient information sheet by ticking the corresponding boxes in the consent page of the survey. We will contact you via your preferred means of contact (email/ phone) as indicated on our record for up to three times, and if we do not hear from you after three attempted contacts we will assume that you are not interested in taking part of the study and will not make further contacts.

What will happen if I choose to take part?

If you agree to take part, you can follow the link to the consent form. Once you have indicated that you have understood and agreed with the information presented on this sheet, you can proceed to the online survey questions.

As part of your treatment in the [**Trust's name**] IAPT service, you would have completed a range of self-reported questionnaires regarding the symptoms you were experiencing at different time points. You might have filled out these measures in a pen and paper version or in an electronic version that was sent to you via email before each therapy session. The survey used in this study asks about the way you thought about those questionnaires and how you decided on the answers you would give to questions on those questionnaires. By completing the survey, you can help clinicians and researchers to better understand the meaning of the answers that our service users give to the questions within the routine questionnaire measures that we ask everyone to complete at every session in IAPT. The survey should take around 15 minutes to complete.

The second part of the study involves a one-to-one interview separately only with those of you who have indicated interest to participate. After which, we will seek consent from you to contact your clinician for them to fill out an online survey. This will only take place once you have completed your treatment at the service. You can agree to complete the survey without having to commit to take part in the interview, or to give consent for us to invite your clinician for the survey. You can decide whether to participate in the study at a later time point when you finish your treatment at the IAPT service. There will be an option in the consent form for you to indicate if you will be happy to be contacted for the interview with you and for us to contact your clinician. The interview will either be arranged in one of the clinical rooms in this IAPT service or in the UCL buildings in Torrington Place, depending on what is more convenient for you. We aim to interview approximately 20 people, the nature of this type of research means that we may find that fewer than 20 is sufficient to give us the rich information we are hoping to gain as part of the study. If that should happen, or if many more than 20 indicate a desire to take part in the interview, whilst we will try and accommodate all those that would like to take part, if this is not possible we will contact you via your preferred means and may not be able to conduct an interview with you.

The interviewer will ask you questions about how you make sense of the changes in the questionnaires that you have filled out as part of your treatment in the IAPT services, and if the changes in the scores map onto how you feel about your psychological wellbeing. The interview will last for 25 minutes. Afterwards, with your consent, we will invite your clinician to fill out an online survey regarding changes in your measures and how they understand it. We will have to inform your clinicians that you have participated in the study and have agreed for us to contact them in order to invite them to fill out the online survey based on their understanding of changes in your measures. However, your clinician will not receive any information about your responses to our survey and interview (if applicable) apart from your status of participation. You will not receive information about your clinician's response to the online survey. A unique ID will be used in the study to map your clinicians' response to yours to protect your identity.

What are the possible benefits of taking part?

By taking part in this study, you will contribute to a better knowledge of what is being communicated by the questionnaire responses and how changes in the questionnaires should be understood by clinicians and researchers. This will have the potential to improve the way we evaluate psychological wellbeing in the services, and have an impact on future service management and policy making. As a survey participant, you will automatically enter into a prize draw of a £50 amazon voucher, and as an interview participant, you will be offered £10 for your time.

What are the possible disadvantages or risks of taking part?

As you will be asked to complete survey and interview questions regarding your previous thinking processes and your sense of wellbeing, you may find it slightly uncomfortable to discuss certain details. You can always withdraw from the study at any point by clicking the exit button on the survey webpage. For interview participants, your interviewer will address any concerns you have before the interview takes place, and you will always have an option not to answer any of the questions if you find them too distressing. You can communicate you wish to terminate the interview at any point of the process. There is no consequence for withdrawing your participation in the study.

If you feel upset or concerned about filling in the survey involved in this study, please share your concerns with your IAPT clinician who will be able to help. If you experience any distress during the interview, please notify the interviewer and we will arrange for you to speak to clinicians from the IAPT service or from our team. We will also ask you to contact your GP or your clinician if we become concerned about your welfare during the interview. We will provide a list of local support and crisis services to all of our participants.

Who has reviewed the study?

All research in the NHS is reviewed by an independent group of people called a Research Ethics Committee. They are there to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable ethical opinion by the North West - Preston NHS Research Ethics Committee.

What will happen to my information?

We will follow ethical and legal practice under the Data Protection Act and the General Data Protection Regulation 2016/679 (2018), so that all information about you will be kept secure and handled in confidence. All information collected about you during the study will be kept strictly confidential and stored in secure premises at UCL and on password protected files. Your name and contact details will be stored separately from the data collected with a unique code assigned to your information so that you cannot be recognised from it. Personal information such as your name, age, email and telephone number will be kept for 6 - 12 months. The non-identifiable data collected as part of the study will be kept for 5 years, after which it will be disposed of securely. This is in order to meet the data storage obligations and policies at UCL, this allows for appropriate auditing of studies and of any events or complaints that require investigation after the completion of the study.

The information collected during the study will be looked at by authorised persons from University College London who are organizing and conducting this research study. They may also be looked at by authorised auditors to check that the study is being carried out correctly.

What you say during the research survey and interview is confidential. However, if you say anything to us that we think puts you or anyone else in danger, we may consider it necessary to report this to the appropriate authority.

Limits to confidentiality

Confidentiality will be respected except under the circumstances described above where there is reason to believe that maintaining confidentiality could result in serious harm to someone. Under such circumstances we would inform you of any decisions that might limit your confidentiality. To protect your confidentiality, please do not enter any information that may make yourself identifiable in the free text boxes on the online survey.

What will happen to the results of the research project?

We aim to publish the results of this study as soon as possible after the completion of the study. The results will be published in the form of scientific journal articles, conference presentations, a doctoral thesis, and we will also be fed-back to service user and staff consultation groups. If you wish to obtain a copy of the published results, please feel free to let the researcher know and we will send you a summary of the findings either by email or by post, based on your preference. We will make sure that none of our participants will be identified in any report or publication. We hope that data collected during the course of the project can be used to support additional or subsequent research and to inform policy making.

What happens when the research study stops?

Throughout the study and afterwards, your IAPT clinician or therapist will continue to treat you as they feel as best for your and with your agreement.

Who is organising and funding the research?

This research is funded by Research Department of Clinical, Educational and Health Psychology, University College London (UCL) and Wellcome Trust. It is organised by UCL in collaboration with **[Name of local service obscured for confidentiality reasons]**.

Local Data Protection Privacy Notice

UCL is the sponsor for this study based in the United Kingdom. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. UCL will keep identifiable information about you for 5 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting UCL Data Controller contact data-protection@ucl.ac.uk.

A Data Protection Impact Assessment has been carried out to identify and mitigate any risks to the privacy of the data you provide us as part of the study. The completed form has been registered with the Data Protection Office as part of the research registration forms.

[**IAPT service name obscured**] will keep your name, patient identification number and contact details confidential, and will not pass this information to UCL.[**IAPT service name obscured**] will use this information as needed, to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Certain individuals from UCL and regulatory organisations may look at your electronic patient record held at [**IAPT service name obscured**]l and research records to check the accuracy of the research study. UCL will only receive information without any identifying information. The people who analyse the information will not be able to identify you and will not be able to find out your name, patient identification number or contact details.

[**IAPT service name obscured**] will keep identifiable information about you from this study for 6-12 months after the study has finished.

UCL will collect information about you for this research study from [**IAPT service name obscured**] This information will include your IAPTus number, your contact details and your responses to the IAPT measures (MDS and GRI question), which is regarded as a special category of information. We will use this information to contact you for research purpose and to gain more understanding of your responses to the online survey.

If you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation and in other organisations. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research. This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance.

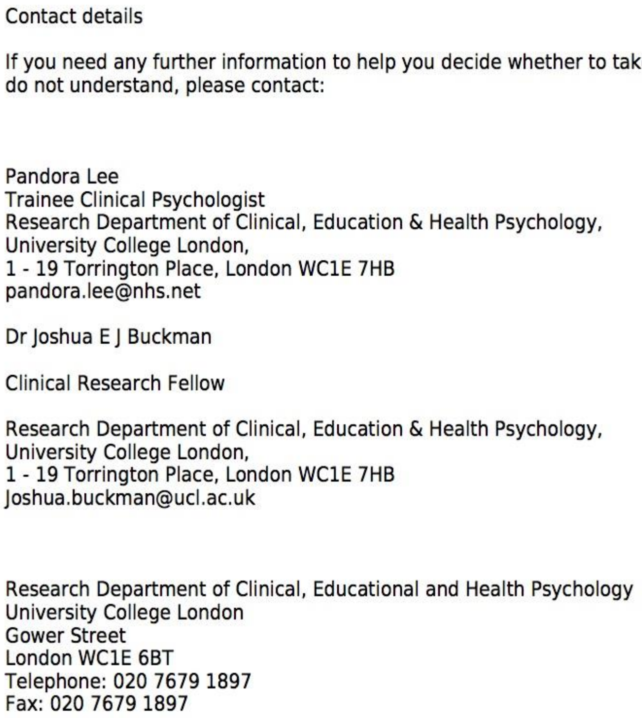
If I have any concerns

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated by members of staff due to your participation in the research, NHS or UCL complaints mechanisms are available to you. Please ask your IAPT clinician if you would like more information on this. In the unlikely event that you are harmed by taking part in this study, compensation may be available. If you suspect that the harm is the result of negligence on the part of the Sponsor (University College London) or your IAPT service, then you may be able to claim compensation. After discussing this with your IAPT clinician, please make the claim in writing to Dr. Joshua Buckman (Joshua.buckman@candi.nhs.uk) who is the Chief Investigator for the research and is based at the University College London (using the address below). The Chief Investigator will then pass the claim to the Sponsor's insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this.

In addition, you may contact PALs:

Details of local services
obscured for confidentiality reasons

Details of local services obscured for confidentiality reasons



Contact details

If you need any further information to help you decide whether to take part in the study, or if there is anything you do not understand, please contact:

Pandora Lee
Trainee Clinical Psychologist
Research Department of Clinical, Education & Health Psychology,
University College London,
1 - 19 Torrington Place, London WC1E 7HB
pandora.lee@nhs.net

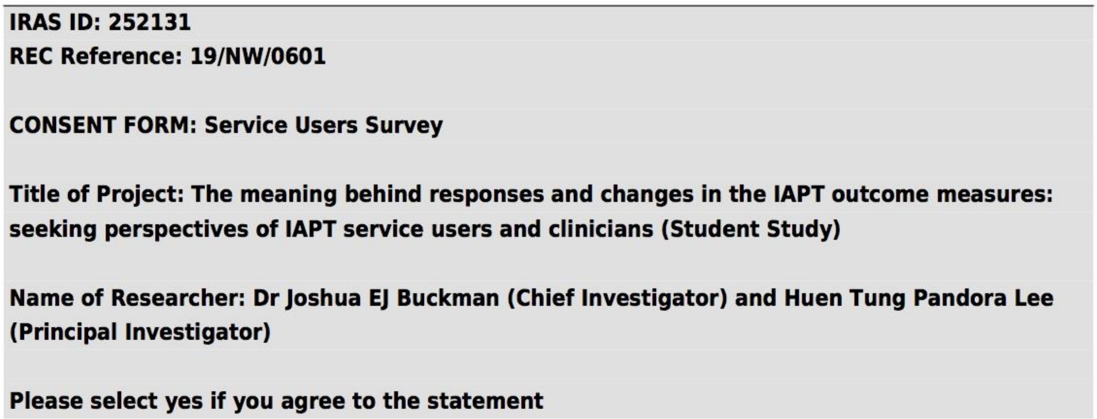
Dr Joshua E J Buckman

Clinical Research Fellow

Research Department of Clinical, Education & Health Psychology,
University College London,
1 - 19 Torrington Place, London WC1E 7HB
Joshua.buckman@ucl.ac.uk

Research Department of Clinical, Educational and Health Psychology
University College London
Gower Street
London WC1E 6BT
Telephone: 020 7679 1897
Fax: 020 7679 1897

Details of local services obscured for confidentiality reasons



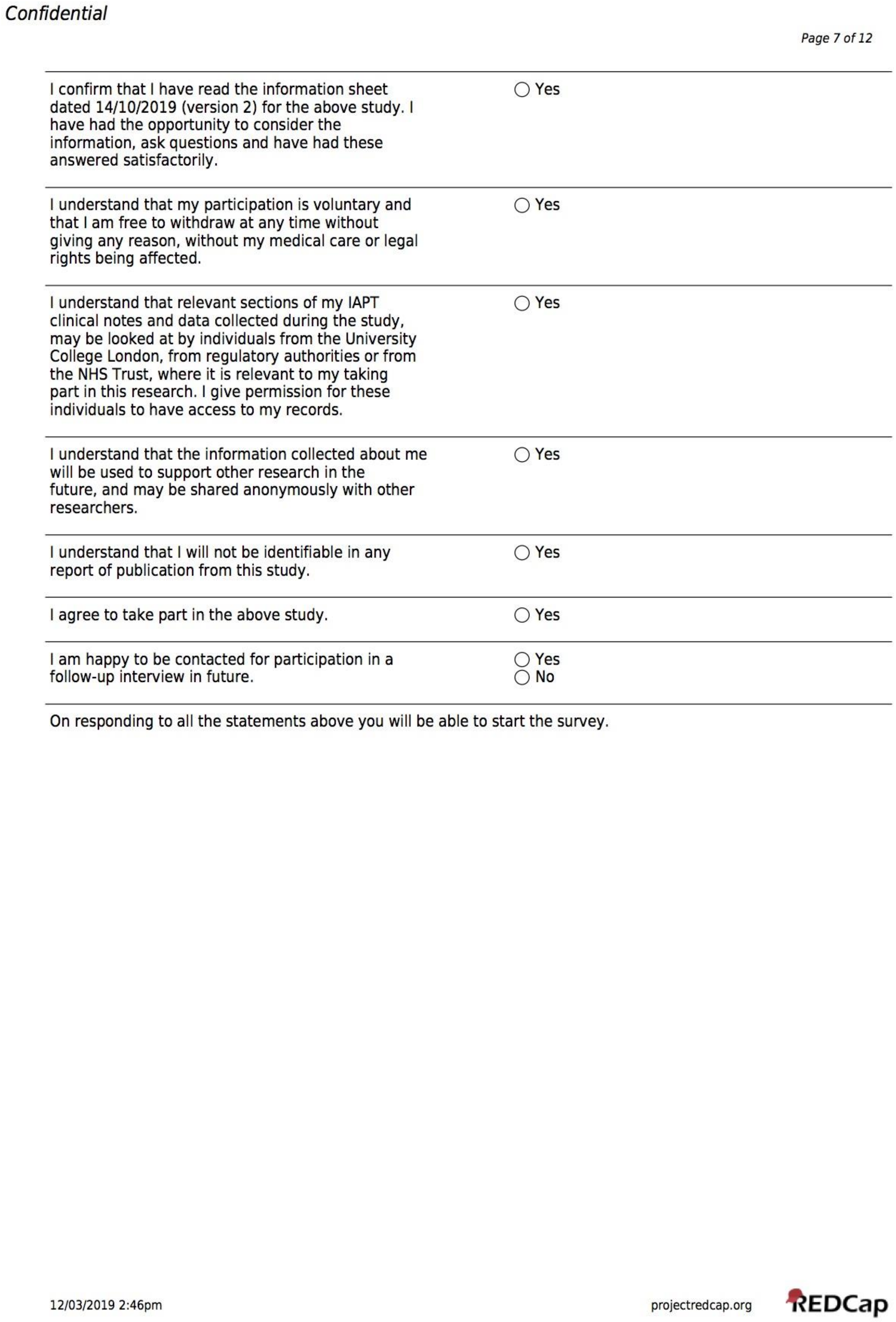
IRAS ID: 252131
REC Reference: 19/NW/0601

CONSENT FORM: Service Users Survey

Title of Project: The meaning behind responses and changes in the IAPT outcome measures: seeking perspectives of IAPT service users and clinicians (Student Study)

Name of Researcher: Dr Joshua EJ Buckman (Chief Investigator) and Huen Tung Pandora Lee (Principal Investigator)

Please select yes if you agree to the statement



I confirm that I have read the information sheet dated 14/10/2019 (version 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	○ Yes
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.	○ Yes
I understand that relevant sections of my IAPT clinical notes and data collected during the study, may be looked at by individuals from the University College London, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.	○ Yes
I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers.	○ Yes
I understand that I will not be identifiable in any report of publication from this study.	○ Yes
I agree to take part in the above study.	○ Yes
I am happy to be contacted for participation in a follow-up interview in future.	○ Yes ○ No

On responding to all the statements above you will be able to start the survey.

The questions below ask about your responses to the questionnaires used in the IAPT services that you will have been asked to complete prior to every session. You should have filled out at least two sets of some of them throughout your contact with the services and probably many more than that.

Click here to display a sample of the IAPT questionnaires that we are asking about

[Attachment: "PHQ9+GAD7.jpg"]

Do you think any of the factors outlined below might have influenced your responses to the IAPT questionnaires? (you may choose as many options as appropriate)

☐ How well you understood the questions
☐ How easy it was to remember the information you were being asked about
☐ How easy it was to decide which of the available response options to choose
☐ How much effort it took to complete the questionnaires
☐ Whether the available response options matched how you would most like to have answered the questions
☐ How well the questionnaires were presented, or how user-friendly they were
☐ None of above

Please give more details on how your understanding of the questions influenced your responses to the questionnaires. __________________________________

Please give more details on how your ability to remember the information being asked about in each question influenced your responses to the questionnaires. __________________________________

Please give more details on how the ease or difficulty in deciding responses to each question influenced your responses to the questionnaire. __________________________________

Please give more details on how your effort required to complete the questionnaires influenced your responses to the questionnaires? __________________________________

Please give more details on how well the available response options matched with the way you would like to have answered the questions influenced your responses to the questionnaires. __________________________________

Please give more details on how your perception of how the questionnaires were presented influenced how you responded to the questionnaires. __________________________________

Can you rank the influence each of the factors had on the way you responded to the IAPT questionnaires, with 6 being the most influential and 1 the least influential.

	1 (least influential)	2	3	4	5	6 (most influential)
How well you understood the questions	○	○	○	○	○	○
How easy it was to remember the information you needed to answer the questions	○	○	○	○	○	○
How easy it was to decide which of the available response options to choose	○	○	○	○	○	○
How much effort it took to complete the questionnaires	○	○	○	○	○	○
Whether the available response options matched how you would most like to have answered the questions	○	○	○	○	○	○
How well the questionnaires were presented, or how user-friendly it was	○	○	○	○	○	○

What other factors might have played a role in the way you responded to the IAPT questionnaires? ______________________________

Thinking now about the question you were asked each session on whether or not you felt better compared to when you started treatment (The Global Rating of Improvement Question) and answer the following:

Click here to see a sample of the Global Rating of Improvement question

[Attachment: "GRI.png"]

How easy did you find it to answer the Global Rating of Improvement each session?	○ Extremely easy ○ Somewhat easy ○ Neither easy nor difficult ○ Somewhat difficult ○ Extremely difficult ○ N/A - I was not asked this question
How useful did you find it to be asked the Global Rating of Improvement question each session?	○ Extremely useful ○ Very useful ○ Moderately useful ○ Slightly useful ○ Not at all useful
Did you discuss your response to the Global Rating of Improvement question with your clinician/therapist?	○ Yes ○ No
Do you think it is valuable to be asked a question on how much better/worse you feel (The Global Rating of Improvement question) each session?	○ Yes ○ No
If yes, why?	______________________________

The questions below concern the comparison between the IAPT questionnaires and the Global Rating of Improvement question

Please click here to see a sample of the questionnaires/ question we are asking about

[Attachment: "compare.jpg"]

Do you think there was a difference in the way you responded to the IAPT questionnaires and the Global Rating of Improvement question?	○ Definitely yes ○ Probably yes ○ Not sure ○ Probably not ○ Definitely not
If yes, in what way was it different?	______________________________
Did the Global Rating of Improvement better reflect your feelings about the progress you made during therapy, than the IAPT questionnaires?	○ Definitely yes - The Global Rating of Improvement was a lot better than the IAPT questionnaires ○ Probably yes - The Global Rating of Improvement was a little better than the IAPT questionnaires ○ Not sure - The Global Rating of Improvement was somewhat similar to the IAPT questionnaires ○ Probably not - The IAPT questionnaires were a little better than the Global Rating of Improvement question ○ Definitely not - The IAPT questionnaires were a lot better than the Global Rating of Improvement question

End of Survey

Participants of this survey will be entered into a prize draw of a £50 amazon voucher. The winner will be announced on 28th February 2020.	○ Yes, I agree to be entered into the prize draw ○ No, I would like to opt out from the prize draw
Would you like us to send you an email regarding the result of this study?	○ Yes ○ No
You have indicated interest in taking part in a follow-up individual interview. For arrangement purpose, can you please select the days, times and locations that would be most convenient for you to attend the interview? The researcher will contact you in a subsequent email with further details of the interview and a proposed arrangement based on your indicated preference. *Please kindly note that interviews are only held during office hours on weekdays either in UCL (1 - 19 Torrington Place, WC1E 7HB) or one of the participating IAPT services.	☐ Monday morning (9am - 12pm) ☐ Monday afternoon (1pm - 4pm) ☐ Tuesday morning (9am - 12pm) ☐ Tuesday afternoon (1pm - 4pm) ☐ Wednesday morning (9am - 12pm) ☐ Wednesday afternoon (1pm - 4pm) ☐ Thursday morning (9am - 12pm) ☐ Thursday afternoon (1pm - 4pm) ☐ Friday morning (9am - 12pm) ☐ Friday afternoon (1pm - 4pm)
Your preferred locations (Please choose as many as appropriate):	☐ ☐ ☐ ☐ **Details of local services obscured for confidentiality reasons**

Appendix E. Service User Interview Schedule

The interview will be divided into two parts, the first part is interested in the way you respond to your questionnaires and the second part aims to find out how you make sense of changes in your questionnaire scores.

A. Cognitive Interviewing

I would like you try to recall as much as you can about how things were for you when you filled out the questionnaires.

1) Comprehension of the question
    a. What was it like to make reading/hearing the questions for the first time?
    b. how easy did you find to understand the questions being asked?
    c. Were there any difficulties with filling the questionnaires in?
    d. **How did your understanding of the questions or any difficulties with knowing what the questions were asking impact on the way you responded to the questions?**
2) Retrieval from memory of relevant information
    a. What was it like trying to remember the relevant information to answer the questionnaire questions?
    b. how easy did you find recalling the relevant information in order to answer the questions?
    c. how confident were you in the accuracy of your memories?
    d. **How did your memory of the relevant information or your confidence in the memories impact on the way you responded to the questions?**
3) Decision processes
    a. What was it like deciding how to respond to the questionnaire questions?
    b. how easy did you find it to make the decision of how to respond?
    c. Did you have any difficulties with the decision-making process?

    d. **How did the decision-making process and the ease or difficulty you had in making a decision of how best to respond impact on the way you responded to the questionnaire questions?**

4) Response processes
    a. What was it like trying to match your own idea of what you would say in response to the questionnaire questions with the response options provided on the questionnaires?
    b. how easy did you find the response process?
    c. Were there any difficulties with it?
    d. **How did the process of responding to the questions in the way required by the questionnaires impact on your response?**
5) Perception
    a. did you fill out the questionnaires online on a computer, on your phone, or in pen and paper, or more than one of these? What was it like to read the questionnaires on a screen/ on paper?
    b. how easy did you find reading it on screen/ on paper?
    c. Were there any difficulties with that?
    d. **How did it impact on the way you responded to the questions?**
6) Other factors
    a. When you were filling in your questionnaires, was there other information (not asked about) that influenced how you responded to the questions?
    b. Were there other things on your mind while you were filling out the questionnaires?
    c. **How did having other things on your mind or using information not asked about in the questionnaire to influence your responses impact on the accuracy of your questionnaire responses?**

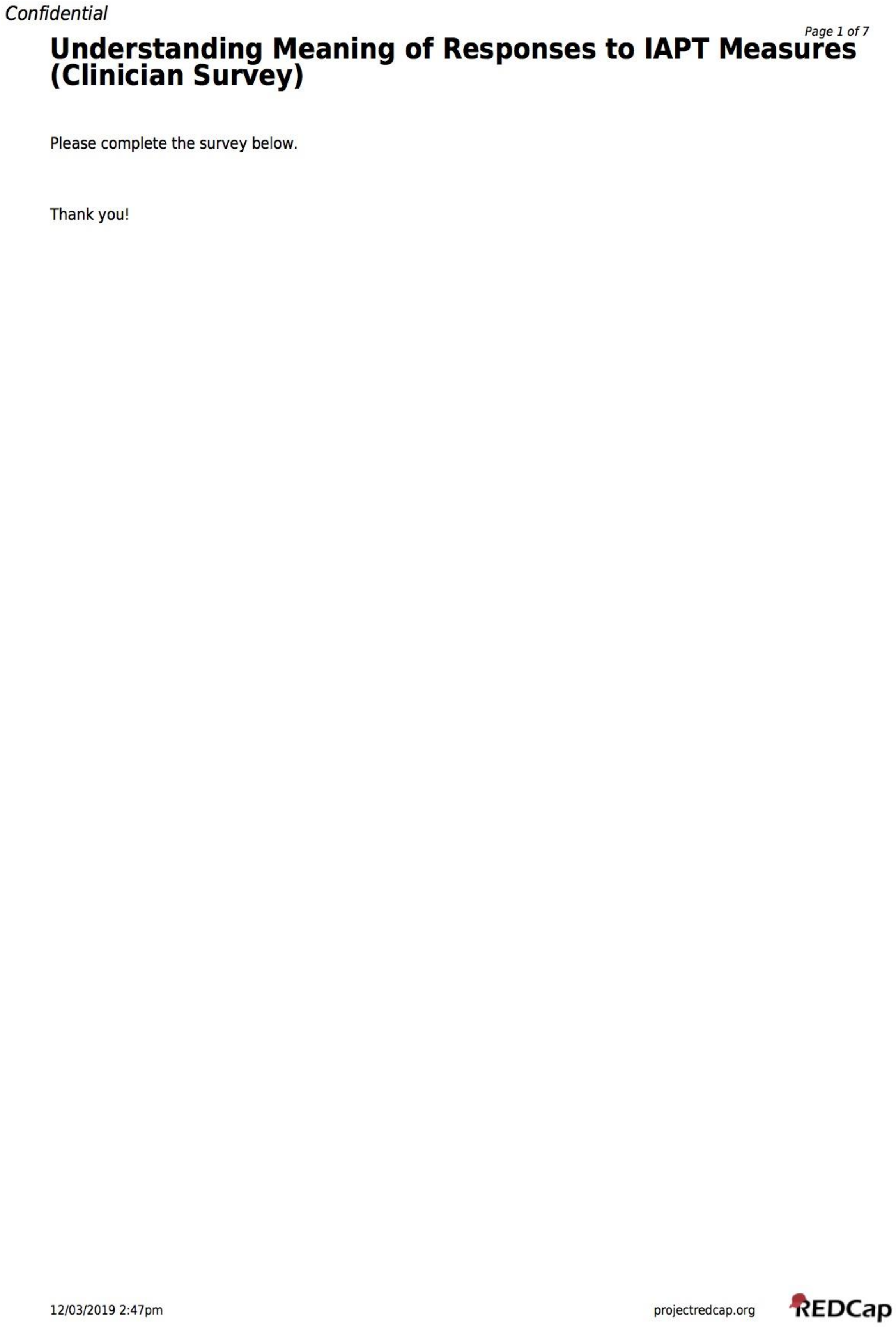
B. Making sense of changes

We are interested in how people make sense of the changes in their questionnaire measure scores during and at the end of therapy.

1. Now that you have completed your treatment in the IAPT service, how do you feel compare to when you started?
2. Do you think there were any changes in the overall scores on the symptom related questionnaire you filled out from the beginning to the end of your treatment? Do you think the scores have gone up (symptoms got worse), gone down (symptoms improved), or not changed? If you don't know do say so.
3. Do you think there were any changes in the individual symptoms (questions asked about in the questionnaires) from the beginning to the end of your treatment? If so, what do you think changed and how?
4. Do you think the changes in your questionnaire scores matched your general sense of wellbeing at the time of your last treatment session? And did this match up with your sense of if you got better, worse or experienced little change from the beginning to the end of your treatment? In what ways did these things match or differ?
5. Besides the level of symptoms you experienced, what other factors do you think contributed to your general sense of wellbeing at the time of your last treatment session?

Do you think that your clinician would understand the changes in your questionnaire scores the same way as you? If not, in what way might their understanding differ from yours?

Appendix F. Online Clinician Survey



Understanding Meaning of Responses to IAPT Measures (Clinician Survey)

Please complete the survey below.

Thank you!

IRAS ID: 252131

REC Reference: 19/NW/0601

Participant Information Sheet: Clinicians

Title of Project: The meaning behind responses and changes in the IAPT outcome measures: seeking perspectives of IAPT service users and clinicians (Student Study)

Participant Information Sheet

You are being invited to take part in a research study. Before you decide whether to take part, we would like you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to discuss with others if you wish. If there is anything that is not clear or if you would like more information, please ask us.

Summary

A range of questionnaires which measure common symptoms of low mood, anxiety, and any difficulties in doing day-to-day tasks are routinely used in the Improving Access to Psychological Therapies (IAPT) services. Changes in the scores on these measures can have different implications on a clinical, service, and policy level, but based on feedback from our Service User advisory groups we are interested in better understanding what the questions and scores on those measures actually mean. More specifically, we hope to find out more about (1) the things influence the way your clients respond on the questionnaires, (2) how you and your clients each make sense of changes in scores on those questionnaires, and (3) whether or not the scores reflect your client's own sense of wellbeing.

To do this, we have invited the service users from your IAPT service to fill in a survey that asks information on the thinking processes behind their responses to the routine outcome measures used in IAPT services. Some of your service users have also taken part in one-to-one interviews to explore in greater depth the factors underlying their questionnaire responses and the way they make sense of the changes in their outcome measure scores throughout their treatment in your IAPT service. We are inviting you as their clinicians to fill out a 10-minute online survey to offer your perspectives on the changes in their outcome measure scores. The result from this study will help clinicians and researchers to better understand what might be communicated in the routine outcome measures, how this corresponds to the scores on those measures, and whether or not there are any similarities or differences in the way that the clinicians and IAPT the service users each consider the service user's sense of improvement throughout therapy. This might also inform future policy development, service management and clinical practice.

Why have I been invited?

You are being invited as one or more of the service users that you have seen in this IAPT services as part of the **[IAPT service name obscured]** have taken part in the study and have given us consent to contact you to offer your view on the changes in their outcome measures.

Do I have to take part?

Your participation in this study is completely voluntary. There will be no impact on your role in this IAPT service regardless of whether you choose to participate or not in this study. If you do decide to participate, you are free to withdraw your participation at any time and we will remove all your information from our research data upon your request. You do not need to give any reasons for your withdrawal. We will contact you via your **[IAPT service name obscured]** internal email for up to three times, and if we do not hear from you after three attempted contacts we will assume that you are not interested in taking part of the study and will not make further contacts.

What will happen if I choose to take part?

If you agree to take part, you can follow the link to the consent form. Once you have indicated that you have understood and agreed with the information presented on this sheet, you can proceed to the online survey questions. Your service user will not receive any information about your responses to our survey or your status of participation. You will not receive information about your service response to the online survey and interview (if applicable). A unique ID will be used in the study to map your response to your service user to protect your identity.

What are the possible benefits of taking part?

By taking part in this study, you will contribute to a better knowledge of what is being communicated by the questionnaire responses and how changes in the questionnaires should be understood by clinicians and researchers. This will have the potential to improve the way we evaluate psychological wellbeing in the services, and have an impact on future service management and policy making.

What are the possible disadvantages or risks of taking part?

Taking part in this study might involve taking a small amount of time out from your working day. You are welcome to discuss with your manager or supervisor if this can be regarded as something that is part of your work as a clinician in this IAPT service. Your participation in this study should not have any impact on your role in this service. If you have any concerns regarding this, please feel free to contact the researcher to discuss.

Who has reviewed the study?

All research in the NHS is reviewed by an independent group of people called a Research Ethics Committee. They are there to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable ethical opinion by the North West - Preston NHS Research Ethics Committee.

What will happen to my information?

We will follow ethical and legal practice under the Data Protection Act and the General Data Protection Regulation 2016/679 (2018), so that all information about you will be kept secure and handled in confidence. All information collected about you during the study will be kept strictly confidential and stored in secure premises at UCL and on password protected files. The data collected as part of the study will be kept for 5 years, after which it will be disposed of securely. You will be contacted with your **[service name obscured]** email address as a staff member of the team.

The information collected during the study will be looked at by authorised persons from the University College London who are organising the research. They may also be looked at by authorised people to check that the study is being carried out correctly.

Limits to confidentiality

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case we would inform you of any decisions that might limit your confidentiality. To protect your confidentiality, please do not enter any information that may make yourself identifiable in the free text boxes in your response.

What will happen to the results of the research project?

We aim to publish the results as soon as possible after the completion of the study. The results of the study will be published in the form of scientific journal articles, conference presentations, a doctoral thesis, and we will also feedback the results to service user consultation groups and IAPT clinician's forums. If you wish to obtain a copy of the published results, please feel free to let the researcher know and we will send you a summary of the findings either by email or by post, based on your preference. We will make sure that none of our participants will be identified in any report or publication. We hope that data collected during the course of the project can be used to support additional or subsequent research and to inform policy making.

Who is organising and funding the research?

This research is funded by Research Department of Clinical, Educational and Health Psychology, University College London (UCL) and Wellcome Trust. It is organised by UCL in collaboration with **[Local Service name obscured for confidentiality reasons]**

Local Data Protection Privacy Notice

UCL is the sponsor for this study based in the United Kingdom. This means that we are responsible for looking after your information and using it properly. UCL will keep identifiable information about you for 5 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting UCL Data Controller contact data-protection@ucl.ac.uk.

UCL will collect information about you for this research study from **[IAPT service name obscured]**. This information will include your view about your understanding of the changes in your service users' outcome measures.

A Data Protection Impact Assessment has been carried out to identify and mitigate any risks to the privacy of the data you provide us as part of the study. The completed form has been registered with the Data Protection Office as part of the research registration forms.

If I have any concerns

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated by other members of staff due to your participation in the research, NHS or UCL complaints mechanisms are available to you. Please ask the researcher if you would like more information on this. In the unlikely event that you are harmed by taking part in this study, compensation may be available. If you suspect that the harm is the result of negligence on the part of the Sponsor (University College London) or your IAPT service, then you may be able to claim compensation. After discussing this with your IAPT clinician, please make the claim in writing to Dr. Joshua Buckman (Joshua.buckman@candi.nhs.uk) who is the Chief Investigator for the research and is based at the University College London (using the address below). The Chief Investigator will then pass the claim to the Sponsor's insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this.

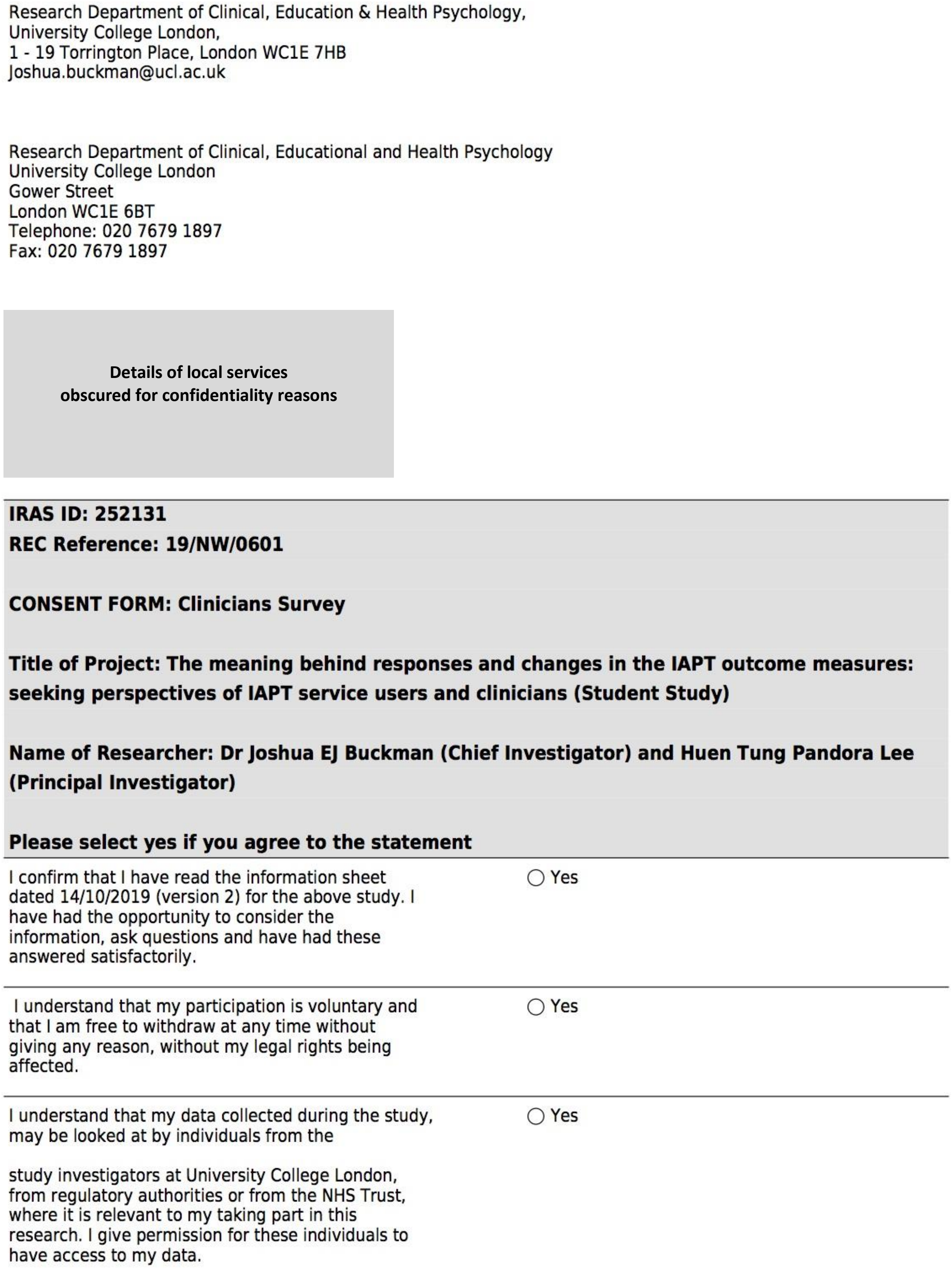
Contact details

If you need any further information to help you decide whether to take part in the study, or if there is anything you do not understand, please contact:

Pandora Lee
Trainee Clinical Psychologist
Research Department of Clinical, Education & Health Psychology,
University College London,
1 - 19 Torrington Place, London WC1E 7HB
pandora.lee@nhs.net

Dr Joshua E J Buckman

Clinical Research Fellow



Research Department of Clinical, Education & Health Psychology,
University College London,
1 - 19 Torrington Place, London WC1E 7HB
Joshua.buckman@ucl.ac.uk

Research Department of Clinical, Educational and Health Psychology
University College London
Gower Street
London WC1E 6BT
Telephone: 020 7679 1897
Fax: 020 7679 1897

Details of local services
obscured for confidentiality reasons

IRAS ID: 252131
REC Reference: 19/NW/0601

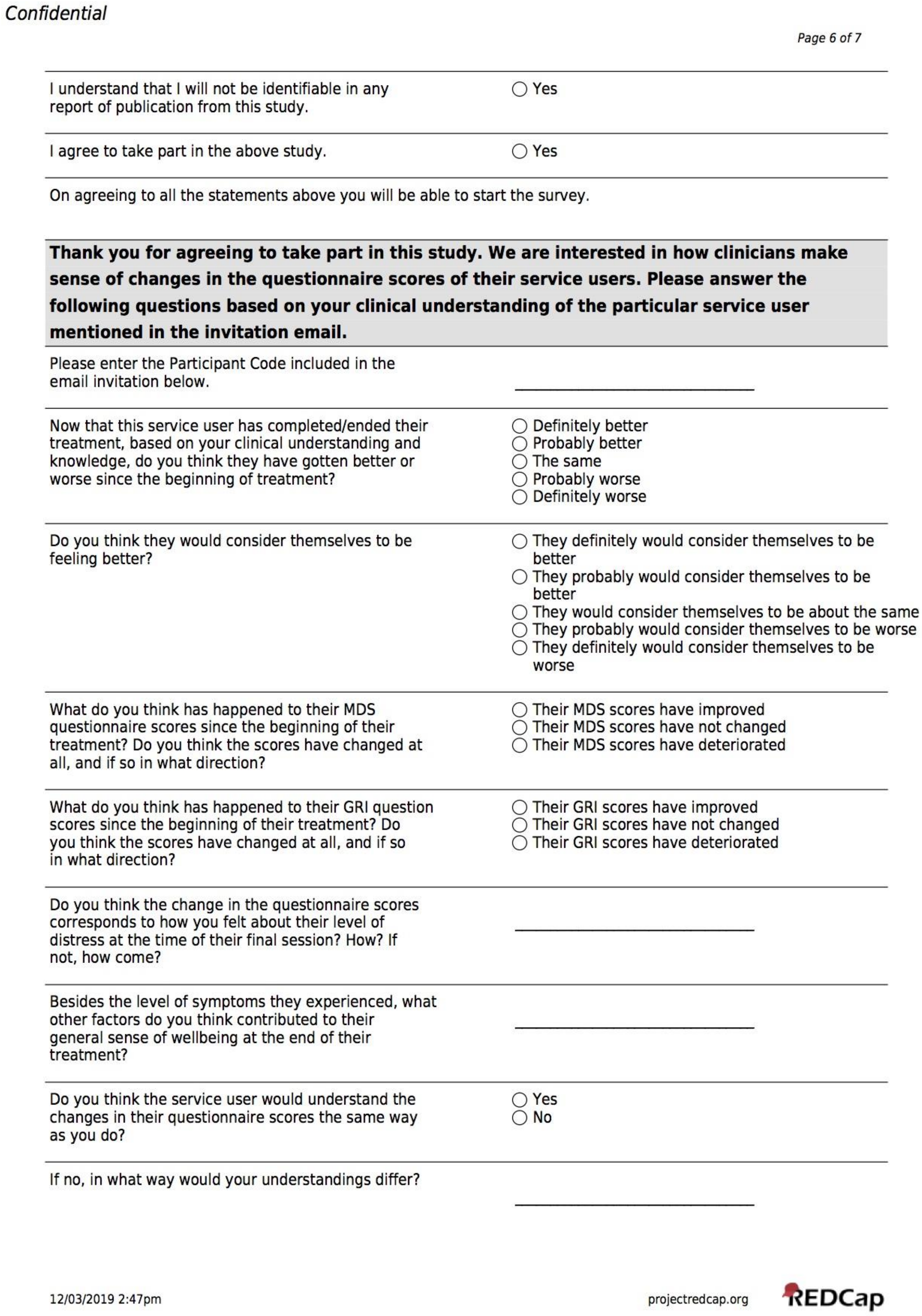
CONSENT FORM: Clinicians Survey

Title of Project: The meaning behind responses and changes in the IAPT outcome measures: seeking perspectives of IAPT service users and clinicians (Student Study)

Name of Researcher: Dr Joshua EJ Buckman (Chief Investigator) and Huen Tung Pandora Lee (Principal Investigator)

Please select yes if you agree to the statement

I confirm that I have read the information sheet dated 14/10/2019 (version 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	○ Yes
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my legal rights being affected.	○ Yes
I understand that my data collected during the study, may be looked at by individuals from the study investigators at University College London, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data.	○ Yes



Confidential

Page 6 of 7

I understand that I will not be identifiable in any report of publication from this study.	○ Yes
I agree to take part in the above study.	○ Yes

On agreeing to all the statements above you will be able to start the survey.

Thank you for agreeing to take part in this study. We are interested in how clinicians make sense of changes in the questionnaire scores of their service users. Please answer the following questions based on your clinical understanding of the particular service user mentioned in the invitation email.

Please enter the Participant Code included in the email invitation below.	__________________________
Now that this service user has completed/ended their treatment, based on your clinical understanding and knowledge, do you think they have gotten better or worse since the beginning of treatment?	○ Definitely better ○ Probably better ○ The same ○ Probably worse ○ Definitely worse
Do you think they would consider themselves to be feeling better?	○ They definitely would consider themselves to be better ○ They probably would consider themselves to be better ○ They would consider themselves to be about the same ○ They probably would consider themselves to be worse ○ They definitely would consider themselves to be worse
What do you think has happened to their MDS questionnaire scores since the beginning of their treatment? Do you think the scores have changed at all, and if so in what direction?	○ Their MDS scores have improved ○ Their MDS scores have not changed ○ Their MDS scores have deteriorated
What do you think has happened to their GRI question scores since the beginning of their treatment? Do you think the scores have changed at all, and if so in what direction?	○ Their GRI scores have improved ○ Their GRI scores have not changed ○ Their GRI scores have deteriorated
Do you think the change in the questionnaire scores corresponds to how you felt about their level of distress at the time of their final session? How? If not, how come?	__________________________
Besides the level of symptoms they experienced, what other factors do you think contributed to their general sense of wellbeing at the end of their treatment?	__________________________
Do you think the service user would understand the changes in their questionnaire scores the same way as you do?	○ Yes ○ No
If no, in what way would your understandings differ?	__________________________

12/03/2019 2:47pm projectredcap.org REDCap

Have you discussed the changes in their MDS and GRI scores with the service user?	○ Yes ○ No
Apart from the symptoms asked about in each questionnaire what other factors might the service user have been considering when completing the MDS questionnaires?	______________________________
Might any of these other things better explain the change and direction of change in symptom measure scores pre-post treatment vs the level of symptomology experienced by them at the time?	○ Yes - these other factors might better explain the change in the MDS scores ○ Not sure - these other factors might be as important as the symptom change in accounting for changes in the MDS scores ○ No - these other factors do not better explain the changes in the MDS scores
Do you think the Global Rating of Improvement question better reflected the progress this service user made during therapy than the IAPT MDS questionnaires?	○ Definitely yes - The Global Rating of Improvement question was a lot better than the IAPT questionnaires ○ Probably yes - The Global Rating of Improvement question was a little better than the IAPT questionnaires ○ Not sure - The Global Rating of Improvement question was somewhat similar to the IAPT questionnaires ○ Probably not - The IAPT questionnaires were a little better than the Global Rating of Improvement question ○ Definitely not - The IAPT questionnaires were a lot better than the Global Rating of Improvement question

Appendix G. Demographic And Clinical Information (Phase One And Two)

Table 1

Frequency Table Of Demographic Information

		Survey Respondents	Interview Participants
		N = 121 (%)	**N = 19 (%)**
Gender	*Male*	35 (28.9)	7 (36.8)
	Female	86 (71.1)	12 (63.2)
Age Group	*18 - 24*	20 (16.5)	6 (31.6)
	25 - 44	75 (62.0)	9 (47.4)
	45 - 64	21 (17.4)	3 (15.8)
	65 or above	5 (4.1)	1 (5.3)
Ethnicity	*White British*	55 (45.5)	11 (57.9)
	Any other white	37 (30.6)	5 (10.5)
	Black, Asian and other ethnic minority	17 (14.0)	3 (15.8)
	Other mixed	12 (9.9)	-
Employment Status	*Employed full-time*	62 (51.2)	10 (52.6)
	Employed part-time	17 (14.0)	3 (15.8)
	Self-employed	10 (8.3)	2 (10.5)
	Full-time student	15 (12.4)	3 (15.8)
	Full-time carer	2 (1.7)	-
	Unemployed	11 (9.1)	1 (5.3)
	Retired	4 (3.3)	-

Table 2

Frequency Table Of Clinical Information

		Survey Respondents	Interview Participants
		N = 121 (%)	**N = 19 (%)**
Referral Source	*Self*	97 (80.2)	18 (94.7)
	Non-self	24 (19.8)	1 (5.3)
Problem Descriptors[a]	*Depressive disorders*	57 (47.1)	11 (57.9)
	Anxiety disorders	43 (35.5)	6 (31.6)
	Other	14 (11.5)	2 (10.6)
	N/A	7 (5.8)	-
Risk	*No risk*	33 (27.3)	6 (31.6)
	Low risk	78 (64.5)	13 (68.4)
	Moderate risk	3 (2.5)	-
	Risk not assessed	7 (5.8)	-
Medication	*Not prescribed*	74 (61.2)	11 (57.9)
	Prescribed but not taking	5 (4.1)	-
	Prescribed and taking	40 (33.1)	8 (42.1)
	Not stated	2 (1.7)	-
Treatment Type	*Assessment only*	1 (0.8)	-
	Step 2 interventions[b]	74 (61.2)	9 (47.4)
	Step 3 interventions[c]	38 (31.4)	9 (47.4)
	Mixed Step 2 and 3	8 (6.6)	1 (5.3)
Initial Caseness	*Not at caseness*	8 (6.6)	-
	Either at caseness	29 (24.0)	4 (21.1)
	Both at caseness	84 (69.4)	15 (78.9)
Reliable improvement	*No reliable improvement*	27 (22.3)	-
	Reliable improvement	90 (74.4)	18 (94.7)
	Reliable deterioration	4 (3.3)	1 (5.3)
Moved to Recovery		69 (57.0)	14 (73.7)
Reliable Recovery		64 (52.9)	14 (73.7)
Last GRI	*I feel a lot better*	46 (38.0)	11 (57.9)
	I feel slightly better	43 (35.5)	5 (26.3)
	I feel about the same	19 (15.7)	3 (15.8)
	I feel slightly worse	5 (4.1)	-
	I feel a lot worse	-	-
	Was not asked	8 (6.6)	-

Notes.

[a] Depressive disorders include Depressive Episode and Recurrent Depressive Disorder, Anxiety disorders include General Anxiety Disorder, Panic Disorder, Social Phobia, Obsessive Compulsive Disorder, Hypochondriacal Disorder; Other include Mixed anxiety and depressive disorder, Somatoform Disorder, Primary Insomnia and Adjustment Disorder.

[b] Step 2 interventions included Guided Self-Help, Workshops, Online Cognitive Behavioural Therapy (CBT) program, Employment Support.

[c] Step 3 interventions included High Intensity CBT, counselling, Interpersonal Psychotherapy, Dynamic Interpersonal Therapy and mindfulness group

Table 3

Descriptive Statistics Of Clinical Information

	Survey Participants (N = 121)			**Interview Participants (N = 19)**		
	Minimum	***Maximum***	***Mean (SD)***	***Minimum***	***Maximum***	***Mean (SD)***
Days in Service	32	607	189.3 *(116.4)*	63	511	229.4 *(137.8)*
Number of Sessions attended	1	27	7.6 *(4.7)*	2	18	10 *(4.8)*
Sets of Measures completed	2	32	9.4 *(5.6)*	3	23	12.5 *(6.3)*
Clinical Time (mins)	45	1740	464.6 *(353.7)*	90	1255	617.9 *(374.7)*
Initial PHQ9	2	25	13.2 *(4.8)*	8	22	14.3 *(3.7)*
Initial GAD7	3	21	12.3 *(4.3)*	5	20	12.7 *(3.8)*
Last PHQ9	0	27	7.2 *(5.2)*	0	18	6.3 *(4.7)*
Last GAD7	0	21	6.7 *(4.5)*	0	16	5.3 *(4.2)*

Appendix H. Descriptive Statistics Of Survey Responses (Cognitive Processes)

Table 4

Frequency Table Of The Cognitive Processes Being Selected By Respondents As Influential Factors In Their Response To MDS Measures

		Survey Respondents	Interview Participants
		N = 121 (%)	**N = 19 (%)**
Cognitive Processes	*Comprehension*	43 (35.5%)	11 (57.9%)
	Memory	46 (38.0%)	10 (52.6%)
	Judgment	60 (49.6%)	13 (68.4%)
	Response	65 (54.7%)	12 (63.2%)
	Effort	18 (14.9%)	3 (15.8%)
	Presentation	20 (16.5%)	6 (31.6%)
	None of Above	14 (11.6%)	1 (5.3%)

Table 5

Mean Rank Score of Each Cognitive Process Based On The Perceived Influence Of The Correspondent Processes On The Response To MDS Measures

		Survey Respondents	Interview Participants
		N = 121 Mean (SD)	**N = 19 Mean (SD)**
Cognitive Processes	*Comprehension*	3.83 (1.68)	4.26 (1.28)
	Memory	3.84 (1.58)	4.00 (1.76)
	Judgment	4.12 (1.36)	4.42 (1.31)
	Response	3.95 (1.49)	4.05 (1.68)
	Effort	2.95 (1.50)	2.79 (1.58)
	Presentation	3.17 (1.61)	3.53 (1.50)

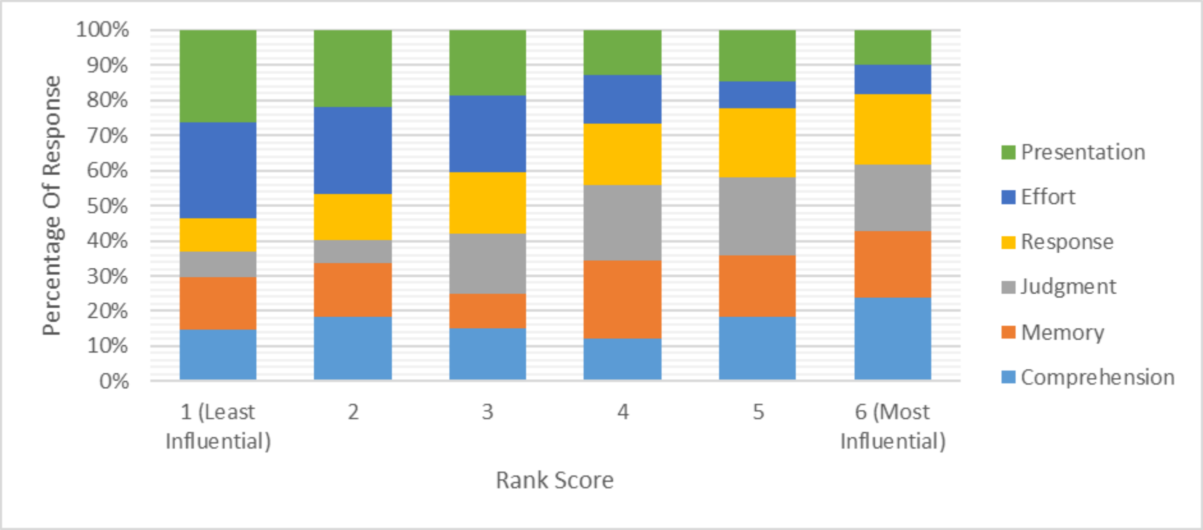
Table 6

Frequency And Percentage Of Response In Each Rank Score Category Based On The Perceived Influence Of The Correspondent Processes On The Response To MDS Measures

Cognitive Factors	**Rank Score N = 121 (%)**					
	1 (Least Influential)	**2**	**3**	**4**	**5**	**6 (Most Influential)**
Comprehension	14 (11.6)	17 (14.0)	21 (17.4)	19 (15.7)	24 (19.8)	26 (21.5)
Memory	14 (11.6)	14 (11.6)	14 (11.6)	35 (28.9)	23 (19.0)	21 (17.4)
Judgment	7 (5.8)	6 (5.0)	24 (19.8)	34 (28.1)	29 (24.0)	21 (17.4)
Response	9 (7.4)	12 (9.9)	25 (20.7)	27 (22.3)	26 (21.5)	22 (18.2)
Effort	26 (21.5)	23 (19.0)	31 (25.6)	22 (18.2)	10 (8.3)	9 (7.4)
Presentation	25 (20.7)	20 (16.5)	26 (15.7)	20 (16.5)	19 (15.7)	11 (9.1)

Figure 6

Stacked Bar Chart Of Percentage Of Response In Each Rank Score Category Based On The Perceived Influence Of The Correspondent Processes On The Response To MDS Measures



Appendix I. Survey Responses As Categorized By Cognitive Processes

Table 7

Facilitative factors as categorized by their effect on individual cognitive processes

Cognitive Processes	**Facilitative Factors**			
	Personal Factors	***Interpersonal Factors***	***Contextual Factors***	***Questionnaire Factors***
Comprehension	Familiarity with questionnaire (2) Easy to understand (11)	Discussion with therapist (1)		
Memory	Personal relevance (1) Recall previous response as reference (1)			Specificity of questions and response options (1)
Judgment	Stability of symptoms (1) Using response options as reference (1) Accurate description of experience (1)			Broad Response Range (2) Clarity of timeframe and distinctiveness of response options (1) Multiple choice format (1)
Response				Specificity of response options (2) Response options aligned with own feelings (3)
Effort	Easy to answer (1)		Easily accessible on device (1) Reminder to complete (1)	Multiple choice format (1)

Cognitive Processes	**Facilitative Factors**			
	Personal Factors	***Interpersonal Factors***	***Contextual Factors***	***Questionnaire Factors***
Presentation			Online survey format (1)	Accurate description of symptoms (1) Repeated use of measures (1) Brevity of response and questionnaire (3) Clarity of response options (1) Consistency of response options and scale (1) Multiple choice format (1) Specificity of question (4) Clear presentation (4)
Other		Discussion with therapist (1)	Easily accessible on device (1) Reminder to complete (1)	Brevity of questionnaires (1) Repeated use of measures as reference (1) Simplicity of language use (1)

Notes. Numbers in brackets indicate occurrence of each theme in terms of the number of responses in each category

Table 8

Influencing Factors As Categorized By Their Effect On Individual Cognitive Processes

Cognitive Processes	**Influencing Factors**			
	Personal Factors	***Interpersonal Factors***	***Contextual Factors***	***Questionnaire Factors***
Comprehension	Idiosyncrasy of meaning (1) Level of insight (2) Understanding of language (1) Understanding of response options (2)	Inability to check meanings with others (2) Inference from being in the service (2) Inference from others response (1)		Lack of specificity (7) Technical issues with display (1)
Memory	Impacted by conditions (fluctuation of symptoms (5); influenced by symptoms (1)) (6) Level of insight (1) Limited by effort required (1) Limited personal relevance (1) Struggle to remember previous response (2) Inference from struggle to recall (Availability heuristics) (1)		Lack of distinctive events in the recall period (1) Recency Effect and Current Mood (12)	Including time before 2 weeks (1) Mismatch between frequency of sessions and recall period (1) Recall period being too long (10) Lack of specificity of recall cues (2) In-between response (1)

Cognitive Processes	Influencing Factors			
	Personal Factors	***Interpersonal Factors***	***Contextual Factors***	***Questionnaire Factors***
Judgment	Fluctuation of symptoms (5) Level of insight (4) Limited personal relevance (3) Struggle to quantify feelings or experience (5) Struggle to recall (5) Unsure about frame of reference (5)	Consideration of how the response might be perceived (2)	Recency Effect and Current Mood (4)	In-between response (10) Lack of specificity of response options (16) Nonlinearity between response options (4) Lack of specificity of questions (2)
Response	Idiosyncrasy of meaning (1) Limited personal relevance (1) Struggle to quantify feelings or experience (4) Struggle to understand response options (2)	Consideration of how the response might be perceived (1) Inference of surveyor's intent (2)	Time Pressure (1) Recency Effect and Current Mood (1)	Did not capture fluctuation or duration of symptoms (6) Does not differentiate causes of behaviours (6) In-between response (29) Lack of specificity (2) Mismatch between frequency of sessions and recall period (1) Nonlinearity between response options (1) Severity of symptoms does not reflect the level of distress (2) Unable to capture feelings (7)
Effort	Attention span (1) Limited personal relevance (1) Low energy level (2) Guessing (1)		Immediate environment (1) Time of the day (1)	Length of questionnaires (2) Technical issues with display (2)

Cognitive Processes	Influencing Factors			
	Personal Factors	***Interpersonal Factors***	***Contextual Factors***	***Questionnaire Factors***
Presentation	Having dyslexia (1) Struggle to quantify feelings or experience (1) Impacted by effort required (1)	Inference of surveyor's intent (1)		Lack of specificity of response options (1) Technical issues with display (1)
Other	Anxiety about over reporting (5) Energy level (1) Limited personal relevance (1) Motivational factors (3) Unsure about frame of reference (1)	Consideration of how the response might be perceived (4) Sensitivity of information (1)	Immediate environment (7) Time of the day (4) Time pressure (10) Recency Effect and Current mood (33)	Lack of specificity of wordings (1) Issues with response options (3) Technical issues with display (1) Time span between questionnaires (1)

Notes. Numbers in brackets indicate occurrence of each theme in terms of the number of responses in each category

Appendix J. Example Quotes For The Seven Most Common Themes

Table 11

Example Quotes To Illustrate The Seven Most Common Themes Of Influencing Factors Underlying Questionnaire Response

Themes	**Example quotes**
Issues with response options impact on memory, decision and response (70)	*"It seemed very dogmatic very set. Oftentimes I was more in the middle than everyday or not at all or half the time. Sometimes it was just in some mornings sometimes when I couldn't sleep etc"* (P120, SS)
Recency Effect and Current mood impact on memory, judgment and response (51)	*"I had a fairly selective memory with my low mood and anxiety. If I did the survey whilst feeling particularly low, my answers would score lower. If I was feeling good, I'd score a better mood."* (P11, SS)
Time pressure, environment and time of completion impact on attention and effort (24)	*"Depends on how focus can you be in the situation you're answering the questions. Not the same being at home, taking your time than for example being in group sessions therapy when your mind is not always where you want it to be"* (P9, SS)
Lack of specificity of questionnaire wordings impact on understanding, retrieval, judgment and response (21)	*"I had difficulty particularly with the question about worrying too much. I wasn't always sure whether this meant worrying about similar things, or worrying about completely separate things."* (P38, SS)
Issues with recall periods (mismatch with frequency of sessions, recall period too long) impacted on memory and response (14)	*"A key stumbling block for me was always the 2 week period... at some point over two weeks I had almost always felt these things in some way - and it's hard to remember to what degree."* (P87, SS)
Fluctuation of conditions and mental states (12)	*"I had to think because my feelings changed depending on my situation. I only felt claustrophobic in confined spaces and the panic attacks would come on suddenly."* (P115, SS)
Negative emotions (anxiety about over-reporting, reinforce negative self-view, amplify negative feelings, unease, overthinking) evoked by questionnaires (12)	*"Honestly I really hated this part of an already impersonal therapy. I can't afford talking therapy and CBT has always made me feel much worse. It just felt like 'hey, you aren't rich so instead of real treatment fill out this online survey.' Because of that I probably skipped through it quite quickly. I don't see what is so hard about just having a therapist ask what is bothering me."* (P57, SS)

Appendix K. Use Of Response Strategies

Given the complex dynamic between external influences and the response process, it would be of interest to explore what strategies they have utilised to manage these conflicting demands and generate their response. Table 13 gives an overview of response strategies, categorised based on the purposes they served and issues that they were used to address. Information in the table is drawn from both the survey response and interviews for a more comprehensive overview.

Five purposes were outlined, which include (a) to achieve an accurate reflection of the conditions, (b) to decrease cognitive load, (c) to adhere to the questionnaire instruction, (d) to make sense of the questionnaire, and (e) to maximise its clinical utility. These categories were derived for descriptive and exploratory purposes, and were set up to best map onto what has been described by the participants, rather than the researcher's speculation about the underlying mechanisms. It was highly likely that these purposes were not mutually exclusive and were interrelated. For instance, to achieve an accurate reflection of the conditions might serve clinical use, or the outcome from making sense of the questionnaire might direct which purposes would be prioritise in subsequent stages of responding.

It is worth noting that, while being faced with a similar issue, the respondent might adopt different response strategies based on their perceived purpose. Thus, in the case of issues with the recall period, a participant had chosen to adhere to the questionnaire instruction of reporting a 14-day period even though they felt scores would not reflect changes induced by their weekly sessions. Whereas another participant has decided to treat the recall period as an estimate to take into account events that fell outside of the recall period but were clinically significant. The perception of whether certain aspect was an issue would also be dependent on the purpose. Thus, whilst using previous responses as an anchor could help to decrease cognitive load, it could be counterproductive if the respondent aimed to derive a response based on the re-numeration of individual events in the recall period.

Another important aspect to consider was whether the respondents could maintain the same strategies, as this would have implication on the validity of changes in their MDS and GRI scores. Despite numerous reports of response difficulties, most respondents described an attempt in adhering to the same response strategies throughout, such that “I’ve interpreted the question one way to start with… I try to answer the same interpretation of it every time, so that at least… my responses will match up with my responses” (P81). As a result, even though some expressed concerns about how well the measures reflect their feelings, most were optimistic about its validity in mapping changes in their own conditions. At the same time, their ability to implement a consistent approach might be susceptible to contextual influences and their emotional state at the time.

Table 13

Response Strategies (Findings From Both Service Users Survey And Interviews)

Purpose	Issues	Strategies
To achieve an accurate reflection of the condition	In-between response	Circle two answers when the preferred response lies between two options
	Influence from previous response	Avoid remembering the previous response to reduce the influence
	Absoluteness of the "Not at all" category	Choose to report more negatively if it has not been a wholly positive experience
	Impact of current mood	Active self-reminder
		Focus on the specificity of the symptom measures
		Use of timeframe as a reminder
		Fill out the measure just before the session to get a most updated representation*
		Dedicate the same time and day of the week to fill out the measures*
		Try to average out
	Different responses to MDS and GRI	Crosspolination of the two measures – use the other measure as anchor
To decrease cognitive load	Too much effort required	Answer based on general impression rather than specific examples
		Guessing*
	Struggle with memory	Based memory on specific that are accessible
		Make use of calendar as memory aid
	Struggle with judgement and response	Recall previous response as anchor*
		Use response options as reference*
	Feeling worked up by the response process	Take a break and remind oneself the ability to explain their answers if there is ambiguity
	Similarity of questions	Tendency to give similar answers
To adhere to the questionnaire instruction	Mismatch between 2-week recall period and weekly sessions	Adhere to the 2-week recall period
	Conflict between subjective and objective accuracy	Based response solely on the objective frequency of behaviours or feelings rather subjective sense of severity*
To make sense of the questionnaire	Ambiguity of intended purpose of the measures	Draw on own academic or professional background to assume the intended purpose and meaning of the measures
	Ambiguity in meanings of wording	Interpret suicidal questions as describing a feeling rather than actual behaviours
		Clarify with clinician and agree on a shared meaning
		Substitute ambiguous word with one that has clearer and discrete meaning

Purpose	Issues	Strategies
To make sense of the questionnaire (contd.)	Ambiguity in the specificity of timeframe	Interpret the 2-week timeframe as an estimate rather than concrete number of days based on assumption of the intended meaning of the measures
	Ambiguity in the description of response categories	Mentally convert the response categories into number of days, e.g. more than half of the days would be equivalent to 8 or more days
	Uncertainty about the frame of reference	Use responses to other questions as anchor
To maximise its clinical utility	Mismatch between recall period and personally significant events	Interpret the 2-week timeframe as an estimate to take into account significant events that do not fit into the recall period
	Conflict between subjective and objective accuracy	Value subjective accuracy over objective accuracy
		Give an approximation on measures and elaborate in session
	Perceived consequences of response on risk item	Downplay response to avoid triggering reactions from services
	Justify access to service	Include benign events in the reporting
		Downplay improvement to continue treatment
	Impression management	Avoid reporting deterioration to avoid hurting clinician's feeling
	Inability to capture cause and context of behaviours and feelings	Give an approximation on measures and elaborate in session
	Uncertainty about the correct way of responding	Adhere to the same decision making and comprehension throughout to more accurately capture changes
	Response categories unable to capture intensity of symptoms	Treat categories as accumulative and increase frequency for more intense symptoms
	Uncertainty about the frame of reference	With reference to own past experience of mental health to determine severity of symptoms

Notes. * responses from service users surveys instead of interviews

Appendix L. Descriptive Statistics Of Survey Responses (Acceptability Of GRI)

Table 14

Frequency Table Of Survey Responses On The Acceptability Of GRI

		Survey Respondents	**Interview Participants**
		N = 117 (%)	**N = 19 (%)**
Ease of Responding to GRI	*Extremely Easy*	29 (24.8%)	3 (15.8%)
	Somewhat Easy	41 (35.0%)	10 (52.6%)
	Neither Easy nor Difficult	18 (15.4%)	2 (10.5%)
	Somewhat Difficult	26 (22.2%)	4 (21.1%)
	Extremely Difficult	3 (2.6%)	-
Usefulness of GRI	*Extremely Useful*	11 (9.4%)	3 (15.8%)
	Very Useful	33 (28.2%)	3 (15.8%)
	Moderately Useful	43 (36.8%)	6 (31.6%)
	Slightly Useful	18 (15.4%)	7 (36.8%)
	Not at all Useful	12 (10.3%)	-
Valuable to ask GRI	*Yes*	94 (80.3%)	15 (78.9%)
	No	23 (19.7%)	4 (21.1%)
Discuss with Therapist	*Yes*	89 (76.1%)	15 (78.9%)
	No	28 (23.9%)	4 (21.1%)
Difference between GRI and MDS	*Definitely Yes*	13 (11.1%)	3 (15.8%)
	Probably Yes	42 (35.9%)	9 (47.4%)
	Not Sure	43 (36.8%)	3 (15.8%)
	Probably Not	17 (14.5%)	3 (15.8%)
	Definitely Not	2 (1.7%)	1 (5.3%)
GRI vs. MDS	*GRI a lot better*	12 (10.3%)	2 (10.5%)
	GRI a little better	31 (26.5%)	6 (31.6%)
	Not sure	37 (31.6%)	4 (21.1%)
	MDS a little better	34 (29.1%)	7 (36.9%)
	MDS a lot better	3 (2.6%)	-

Appendix M. Example Quotes For Values Of GRI

Table 15

Example Quotes To Illustrate The Perceived Values And Limitations Of GRI

	Themes	**Example quotes**
Values of GRI	*Promote insight (35)*	*"It's good as self reflection, to consider if progress is being made, and if this reflects on my engagement (if insist my worrytime homework, for example)."* (P100, SS)
	Facilitate therapy or communication (14)	*"I think personally for me it was really important to be asked if I feel better or worse because sometimes I felt too ashamed or embarrassed to ask for things to be changed, or to ask to talk about particular things, so by being asked how I feel, I would indirectly give my therapist an insight into how I was doing and then he was able to ask me the right questions to get me speaking about what I wanted to speak about."* (P94, SS)
	Positive focused (10)	*"It gives you a baseline of where you started and where you finished your therapy process. I think for someone who is deeply depressed and had tad low self-esteem this does boost their overall view of themselves if there is even a slightest improvement. I know I felt better each time."* (P90, SS)
	Ability to capture subjective experience (3)	*"Sometimes this can be a more accurate reflection compared with other questions where the answers might not really fit how the person is feeling"* (P119, SS)
	For service's use (3)	*"To improve the service for others and make the therapy more evidence-based"* (P56, SS)
	For clinician's use (3)	*"it gives the therapist more information to work with, maybe each session would be too much but every other session would be a nice check-in point"* (P50, SS)
Limitations of GRI	*Pressure to improve (4)*	*"Not sure if that should be asked in every session, as this could be felt as a bit of pressure to 'improve'. Also, I believe a number of consecutive 'negative' answers could lead the person to think they're not doing well at all, and they may be suggested that they're feeling worse than they actually are."* (P96, SS)
	Impacted by response difficulties (4)	*"I do think the options provided may be unable to capture some responses due to the gap between the 'slightly' and 'a lot' options."* (P39, SS)
	Affected by Current Mood (2)	*"it is difficult to give an accurate answer. If my interview is on a particularly good or bad day, I believe that this would have a huge impact on my answers."* (P30, SS)
	Prefer a conversation about process (1)	*"I would've found a conversation about the whole therapy process beneficial."* (P38, SS)

Appendix N. Example Quotes For Considerations Underlying The Subjective Sense Of Improvement

Table 16

Considerations underlying the subjective sense of improvement

Themes	**Example quote**
Daily functioning and life experience (11)	*"I think the bigger picture is like your overall feeling of the world, and I think that transcends symptoms. Something negative may have happened and an incident may have happened, or you might have felt a particular way at a particular time, but the overall perception of your life experience can really not match that at all... So, I think the bigger picture is – unhelpfully, it's probably not quantifiable in the same way as the symptoms potentially."* (P87, SI)
Perceived increase in coping resources (10)	*"I've had therapy in the past that has been very talk base, which is good, but has not been the practical side of things that had this time, I had a workbook, and score sheets, which gives you something to reflect on. So, the sheets for example were a good indicator, just a reminder of what was going on and progress...And you know, coping strategies, which we all need. So, nothing's externally has changed, but I've got more structure within myself to sort of know what I need to do to you know, push back, or cope or whatever."* (P89, SI)
Goal of therapy (6)	*"definitely the goals thing helped to then look back… because it's quite hard to remember the state that I was in. You remember that you felt bad, but you don't remember exactly the thoughts you had…. It was kind of like a work in progress goals thing, but especially the short-term ones they were very vivid and specific."* (P91, SI)
General impression and emotional experience (5)	*"I guess because before I was, I was just in quite a bad place… that's why I went looking for help, whereas at the end I felt a bit more confident and I felt a bit more. So like even though I couldn't tell you exactly how I felt right at the beginning, I know that I felt better, yeah, I felt better."* (P95, SI)
Positive relationship with clinician (2)	*"actually, I liked her so much... I thought she was so lovely and helpful and … she seemed to know … I felt she related to what I was saying and I remember things she said. And I do work with those, which I hadn't done before when I did before. I mean, it helped me before but not like it did this time. Because I am using the strategies she suggested... I feel a lot better, I feel slightly better. I probably feel a lot better. Yeah."* (P40, SI)

Appendix O. Response Summary From Clinician Survey

Table 18

Response Summary From Clinician Survey

		Frequency **N = 10**
Clinician perceived improvement	*Definitely Better*	5
	Probably Better	5
	The same	-
	Probably Worse	-
	Definitely Worse	-
Service Users perceived improvement	*Definitely would consider themselves to be better*	4
	Probably would consider themselves to be better	6
	Would consider themselves to be the same	-
	Probably would consider themselves to be worse	-
	Definitely would consider themselves to be worse	-
MDS change	*Improved*	9
	No Change	1
	Deteriorate	-
GRI change	*Improved*	8
	No change	2
	Deteriorate	-
Same understanding	*Yes*	8
	No	2
Discuss with Service User	*Yes*	8
	No	2
Importance of other factors	*Other factors more important*	1
	Not sure - the same	8
	Symptom change more important	1
GRI better than MDS	*Definitely Yes*	1
	Probably Yes	1
	Not sure	8
	Probably not	-
	Definitely not	-

Appendix P. Background Information Of Clinician-Paired Service Users

Table 19

Frequency Table Of Demographic Information

		Clinician-Paired Participants
		N = 10
Gender	*Male*	5
	Female	5
Age Group	*18 - 24*	4
	25 - 44	5
	45 - 64	1
	65 or above	-
Ethnicity	*White British*	6
	Any other white	3
	Black, Asian and other ethnic minority	1
	Other mixed	-
Employment Status	*Employed full-time*	6
	Employed part-time	1
	Self-employed	-
	Full-time student	3
	Full-time carer	-
	Unemployed	-
	Retired	-

Table 20

Frequency Table Of Clinical Information

		Clinician-Paired Participants
		N = 10
Referral Source	*Self*	9
	Non-self	1
Problem Descriptors*	*Depressive disorders*	6
	Anxiety disorders	2
	Other	2
	N/A	-
Risk	*No risk*	3
	Low risk	7
	Moderate risk	-
	Risk not assessed	-
Medication	*Not prescribed*	4
	Prescribed but not taking	-
	Prescribed and taking	6
	Not stated	-
Treatment Type	*Assessment only*	-
	Step 2 interventions	6
	Step 3 interventions	4
	Mixed Step 2 and 3	-
Initial Caseness	*Not at caseness*	-
	Either at caseness	1
	Both at caseness	9
Reliable improvement	*No reliable improvement*	-
	Reliable improvement	10
	Reliable deterioration	-
Moved to Recovery		7
Reliable Recovery		7
Last GRI	*I feel a lot better*	6
	I feel slightly better	3
	I feel about the same	1
	I feel slightly worse	-
	I feel a lot worse	-
	Was not asked	-

Table 21

Descriptive Statistics Of Clinical Information

	Clinician-Paired Participants		
	N = 10		
	Minimum	***Maximum***	***Mean (SD)***
Days in Service	63	511	209.9 *(158.9)*
Number of Sessions attended	2	18	8.40 *(4.6)*
Sets of Measures completed	3	21	10.7 *(6.2)*
Clinical Time (mins)	90	1145	511 *(376.3)*
Initial PHQ9	9	22	15.2 *(3.58)*
Initial GAD7	8	19	12 *(3.02)*
Last PHQ9	0	10	5.9 *(3.64)*
Last GAD7	0	10	4.9 *(3.28)*

Table 22

Frequency Table Of Survey Responses On The Influences Of Cognitive Processes

		Selection Frequency	**Mean Rank Scores**
		N = 10	**Mean *(SD)***
Cognitive Processes	*Comprehension*	5	4.20 *(1.55)*
	Memory	5	4.50 *(1.78)*
	Judgment	8	5.00 *(0.82)*
	Response	6	4.30 *(1.25)*
	Effort	1	2.60 (1.96)
	Presentation	3	3.70 (1.57)
	None of Above	-	-

Table 23

Frequency Table Of Survey Responses on the Acceptability of GRI

		Clinician-Paired Participants
		N = 10
Easy of Responding to GRI	*Extremely Easy*	2
	Somewhat Easy	5
	Neither Easy nor Difficult	1
	Somewhat Difficult	2
	Extremely Difficult	-
Usefulness of GRI	*Extremely Useful*	2
	Very Useful	1
	Moderately Useful	2
	Slightly Useful	5
	Not at all Useful	-
Valuable to ask GRI	*Yes*	8
	No	2
Discuss with Therapist	*Yes*	7
	No	3
Difference between GRI and MDS	*Definitely Yes*	1
	Probably Yes	7
	Not Sure	-
	Probably Not	1
	Definitely Not	1
GRI vs. MDS	*GRI a lot better*	-
	GRI a little better	4
	Not sure	1
	MDS a little better	5
	MDS a lot better	-

Appendix Q. Other Factors (Clinician Survey)

Additional factors were explored in this section to consider alternative explanations of the observed patterns. While treatment model and intensity, session length and discharge date might impact on the type of therapy discussion, familiarity between the pair or their memory, none of these factors seemed to play a significant role in accounting for the pattern observed. The match between their problem descriptor and MDS measures were also considered, but those who reported a mismatch in the interview were in fact where a match would be expected based on the diagnosis and the type of symptom measure used (Table 25).

In addition, the perceived use of questionnaires and additional significant factors were drawn from the interview data. In line with findings from previous section on discussion about score change, those that had experience of it being explored in a therapeutic context had a more positive view about its utility in facilitate their clinical progress. In contrast, those that felt it was used to make treatment decision or determine access to service without a therapeutic discussion tended to perceive it more negatively and have greater mismatch with their clinicians' views on score change. Uncertainty about score use and lack of feedback also led to greater response difficulties, negative emotional impacts and response editing. Moreover, those that had the most incongruent views with their clinicians felt that their therapy did not contribute to their improvements.

Appendix R. Schematic Representation In Relation To Existing Models And Theories

Our findings corresponded to the psycho-linguistic perspective of questionnaire response, in that measures are interpreted with a communicative intent (Schwarz, 1995, 1999). It is likely that the communicative and interpersonal aspects are more salient due to the clinical implication of the measure outcomes and the existing relationship service users have with the clinicians and the service. The cognitive social-psychological model of questionnaire response (Del Boca & Noll, 2000) served as a good starting point to collate understandings gained from the existing study, but fell short in addressing the clinical impacts of the questionnaire scores. The satisficing model outlines motivation, personal ability and task difficulty as the three components to determine the likelihood of the use of satisficing strategies (Krosnick et al., 1987; Krosnick et al., 1996; Narayan, 1995). Task difficulty could depend on the match between subjective sense of wellbeing or change and questionnaire factors. Personal ability is included in the personal factor, and motivation is suggested to be a result of perceived relevance of the measures. In addition, perceived use of the measures can play an important role in the respondents' evaluation about the risk of disclosure on sensitive questions (Tourangeau & Yan, 2007). Previous literature (Clore & Wyer Jr, 2001; Schwarz & Clore, 2003) has acknowledged the role of emotions in cognitive processes, but it has not been incorporated to a comprehensive model of questionnaire response. It is proposed that, emotions would have an even more prominent role in a mental health setting and in response to measures that specifically assess emotional constructs. Emotions has therefore been included a separate factor. In addition, the model depicts the interpersonal component in the interpretation of scores as an interaction and communication among service user, clinician and the service. It is proposed that the interpretation of scores will lead to clinical decision or outcome, which will not only feed into the perceived relevance and use of the measures, but also contribute to subjective sense of wellbeing through clinical actions.