

November 2019

Consensual Qualitative Research: Replicability of Results and Social Reliability of Process

Nicholas Morrison
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2



Part of the [Psychological Phenomena and Processes Commons](#)

Recommended Citation

Morrison, Nicholas, "Consensual Qualitative Research: Replicability of Results and Social Reliability of Process" (2019). *Doctoral Dissertations*. 1799.

<https://doi.org/10.7275/14792839> https://scholarworks.umass.edu/dissertations_2/1799

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Consensual Qualitative Research: Replicability of Results and Social Reliability of
Process

A Dissertation Presented

by

NICHOLAS R. MORRISON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2019

Clinical Psychology

Consensual Qualitative Research: Replicability of Results and Social Reliability of
Process

A Dissertation Presented

by

NICHOLAS R. MORRISON

Approved as to style and content by:

Michael J. Constantino, Chair

Harold D. Grotevant, Member

Brian Lickel, Member

Sarah A. Fefer, Member

Caren M. Rotello, Department Chair
Department of Psychological and Brain Sciences

DEDICATION

To Charles H. Bear

“He’s the best bear, and that IS true.”

ACKNOWLEDGEMENTS

It would have been impossible for me to achieve what I have thus far without the considerable support I have received throughout my life. I would first like to thank my parents, stepparents, and grandparents for all they have done, and my three brothers for their love and support. I would not be the person I am today without my family. I would also like to thank George N. Parks, whose Drum Major Academy and Minuteman Marching Band introduced me to UMass, and whose leadership training I rely on each and every day.

Many teachers and supervisors shaped and facilitated my experiences during my undergraduate and graduate years. Although there are too many to list here, I would especially like to thank Dr. Richard Halgin for his mentorship through my undergraduate career and toward my post-baccalaureate experiences that would help prepare me for graduate school, and Dr. Tammy Rahhal for her years of mentorship around teaching and advising.

I would also like to thank my graduate school peers for their constant support, including my lab mates throughout the years: Rebecca Ametrano, Samantha Bernecker, Alice Coyne, Brien Goodwin, and Heather Muir. The teamwork you always exhibited helped us all to succeed (and laugh) together. Additionally, I thank my graduate school cohort: Shayl Griffith, Marykate Oakley, Gennarina Santorelli, and (adopted coho-mate) Rachel Herman. Your encouragement got me through some of the roughest patches of my doctoral training. I am especially grateful for Genna's support, without which I could not have done it; I will cherish our memories together always in my heart.

I also thank Psi Chi, the Department of Psychological and Brain Sciences, and the Graduate School of the University of Massachusetts Amherst for funding my dissertation. Moreover, I would like to thank Dr. James Boswell for his willingness to share his data for this project. This dissertation would not have been possible without these supports.

Additionally, I am grateful to the many individuals whose tremendous efforts allowed my dissertation to come to fruition: the study judges and data auditors (whose names I cannot share here to protect their study anonymity, but to whom I will always be grateful!), and research assistants (Lindsey Chiarelli, Shannon Gair, Lauren Haliczzer, and Kristin Heideman). Also, thank you to my dissertation committee members, Drs. Sarah Fefer, Harold Grotevant, and Brian Lickel, for their willingness to serve and for facilitating my efforts related to this project.

Lastly, it is with the most heartfelt and sincere gratitude that I thank my dissertation chair and advisor, Dr. Michael Constantino. You have truly been a wonderful teacher, advocate, mentor, and friend over this last decade. You were always there to guide me through the periods of doubt and uncertainty. Thank you for pushing me to get to where I am now, and more importantly for believing I could do it; your advice, mentorship, and unwavering support mean more to me than you know.

ABSTRACT

CONSENSUAL QUALITATIVE RESEARCH: REPLICABILITY OF RESULTS AND SOCIAL RELIABILITY OF PROCESS

SEPTEMBER 2019

NICHOLAS R. MORRISON, B.A., UNIVERSITY OF MASSACHUSETTS AMHERST
M.S., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Michael J. Constantino

To solidify further their scientific footing, qualitative approaches would ideally demonstrate that they yield replicable information about a phenomenon under study. Although consensual qualitative research (CQR; Hill, 2012) proposes a rigorous, multistep method to enhance interjudge reliability and instill confidence in the results, it remains unclear if multiple uniformly trained teams analyzing the same stimulus set would arrive at similar analytic output (i.e., *replicability*—a high form of trustworthiness). Moreover, it is unclear if replicability (or lack thereof) might be influenced by the process through which CQR judges arrive at their output (i.e., *social reliability*). Addressing these gaps, this exploratory study employed mixed methods to evaluate replicability and social reliability between 2 teams that each consisted of 4 randomly assigned judges. These judges were uniformly trained in CQR before the teams separately analyzed 12 transcripts of semi-structured interviews assessing mental health care consumers' perspectives on using provider performance information to inform their treatment decisions. Replicability was examined quantitatively and qualitatively by comparing the output elements established by the CQR teams (i.e., domains, categories, core ideas, and core idea exemplars). Social reliability was examined quantitatively and qualitatively by comparing the teams on objective group process and self-reported group

climate. Replicability results were fairly nuanced. Whereas the teams tended to perceive similar *content* that comprised domains, categories, and core ideas, they notably differed in their level of abstraction. The teams also remarkably differed in how representative they saw the information discussed among the interview participants. Moreover, the team that demonstrated more vs. less abstraction also generated more representative findings, spent more time analyzing transcripts, equitably divided time spent discussing their perspectives, evidenced fewer auditor disagreements, and reported more positive group climate than the other team. Results preliminarily inform the practical utility of existing CQR findings, and future methods for optimizing CQR process and the replicability of its output.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	v
ABSTRACT.....	vii
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xv
CHAPTER	
1. INTRODUCTION.....	1
2. METHOD.....	7
2.1 Participants.....	7
2.1.1 CQR Trainer.....	7
2.1.2 CQR Judges.....	7
2.1.3 CQR Data Auditors.....	8
2.1.4 Output Replicability Judge.....	8
2.1.5 Social Reliability Raters.....	8
2.2 CQR Stimulus Set.....	9
2.3 Measures.....	10
2.3.1 Replicability.....	10
2.3.2 Social Reliability.....	10
2.3.2.1 Objective Group Process.....	11
2.3.2.2 Self-Reported Group Climate.....	11
2.4 Procedure.....	12
2.4.1 Training.....	13
2.4.2 Domain Generation.....	14

2.4.3 Open Coding.....	15
2.4.4 Categorization and Cross Analysis.....	15
3. RESULTS.....	17
3.1 Replicability.....	17
3.1.1 Domains.....	17
3.1.1.1 Team 1 Domain 1 vs. Team 2 Domains 1 and 6.....	17
3.1.1.2 Team 1 Domain 2 vs. Team 2 Domains 2 and 5.....	18
3.1.1.3 Team 1 Domain 3 vs. Team 2 Domains 3, 4, 5, and 6.....	18
3.1.1.4 Team 1 Domain 4 vs. Team 2 Domain 7.....	19
3.1.2 Categories.....	19
3.1.3 Core Ideas.....	21
3.2 Social Reliability.....	22
3.2.1 Quantitative Analyses.....	22
3.2.2 Qualitative Analyses.....	25
3.2.2.1 Team 1’s Focus Group.....	26
3.2.2.1.1 Feelings About the Project Changed Over Time.....	26
3.2.2.1.2 Particular Concerns Related to Time and Efficiency of the Project.....	27
3.2.2.1.3 Group Dynamic Was Positive and Comfortable Overall.....	28
3.2.2.1.4 Factors Conducive to Project Completion and That Would Be Helpful in the Future.....	28

3.2.2.1.5 All Steps of the Process Presented Unique Difficulties.....	29
3.2.2.2 Team 2’s Focus Group.....	30
3.2.2.2.1 Members of Group Had Unique Strengths But Worked Well Together.....	30
3.2.2.2.2 Interest in Topic and Positive Work Environment Made the Project More Enjoyable.....	31
3.2.2.2.3 Suggestions for Improving Project in the Future.....	32
3.2.2.2.4 Factors Conducive to Project Completion and That Would be Helpful in the Future.....	33
3.2.2.2.5 Unknown and Freedom Were Frustrating but Led to Better Understanding and Performance.....	34
3.2.2.2.6 Each Step Brought Challenges, but Cross-Analysis Was the Most Difficult.....	35
3.2.2.3 Focus Group Comparison.....	35
3.2.2.3.1 Similarities.....	36
3.2.2.3.2 Differences.....	37
4. DISCUSSION.....	39
APPENDICES	
A. CONSUMER TELEPHONE INTERVIEW ITEMS.....	86
B. CQR PROCESS MEASURE.....	89

C. FOCUS GROUP SEMI-STRUCTURED INTERVIEW INSTRUCTIONS AND ITEMS.....	90
D. CQR DISSERTATION ANONYMOUS FEEDBACK FORM.....	92
E. STUDY FLOW CHART.....	93
F. CQR DISSERTATION PARENT SURVEY IMMERSION QUESTIONNAIRE & EXPECTATIONS AND BIASES.....	95
BIBLIOGRAPHY.....	102

LIST OF TABLES

Table	Page
1. Participant Demographic Characteristics and Presenting Problem Domains.....	50
2. Total Number of Domains and Categories Generated by Teams.....	51
3. Team 1 Domain 1: Experience with Mental Health Care Services.....	52
4. Team 1 Domain 2: Experience Selecting Mental Health Services.....	53
5. Team 1 Domain 3: Attitudes About the Selection Process.....	54
6. Team 1 Domain 4: Attitudes on Preferred Providers.....	55
7. Team 2 Domain 1: Past and Current Experiences with Mental Health Services.....	57
8. Team 2 Domain 2: History with Mental Health Care Provider Selection Process.....	58
9. Team 2 Domain 3: Important Factors Selecting a Mental Health Provider.....	59
10. Team 2 Domain 4: Availability and Validity of Provider’s Background Information.....	60
11. Team 2 Domain 5: Possible Improvements to Mental Health Care Selection Process.....	61
12. Team 2 Domain 6: Experience with Surveys/Questionnaires.....	63
13. Team 2 Domain 7: Opinions about Well-Matched Provider.....	64
14. Team 1 Exemplars for Domain 1: Experience with Mental Health Care Services.....	65
15. Team 1 Exemplars for Domain 2: Experience Selecting Mental Health Services.....	67
16. Team 1 Exemplars for Domain 3: Attitudes About the Selection Process.....	69
17. Team 1 Exemplars for Domain 4: Attitudes on Preferred Providers.....	71
18. Team 2 Exemplars for Domain 1: Past and Current Experiences with Mental Health Services.....	73
19. Team 2 Exemplars for Domain 2: History with Mental Health Care Provider Selection Process.....	74

20. Team 2 Exemplars for Domain 3: Important Factors Selecting a Mental Health Provider.....	76
21. Team 2 Exemplars for Domain 4: Availability and Validity of Provider’s Background Information.....	77
22. Team 2 Exemplars for Domain 5: Possible Improvements to Mental Health Care Selection Process.....	78
23. Team 2 Exemplars for Domain 6: Experience with Surveys/Questionnaires.....	80
24. Team 2 Exemplars for Domain 7: Opinions About Well-Matched Provider.....	81
25. CQR Process Measure Scores across Time.....	82

LIST OF FIGURES

Figure	Page
1. Between-team differences in within-team climate change across all process measure administrations.....	83
2. Team 1's focus group themes and sub-themes.....	84
3. Team 2's focus group themes and sub-themes.....	85

CHAPTER 1

INTRODUCTION

Qualitative approaches have been increasingly employed in psychological research, with considerable emphasis in counseling and psychotherapy (Hays, Wood, Dahl, & Kirk-Jenkins, 2016; Levitt, 2015; McLeod, 2000). However, some within the psychological community continue to doubt the scientific legitimacy of these methods, viewing their yield as anecdotal description versus valid and reliable inference (Williams & Morrow, 2009). To establish further their scientific footing, qualitative approaches, no matter their guises or epistemological underpinnings, would ideally demonstrate that they yield replicable information about a phenomenon under study (Moret, Reuzel, Van Der Wilt, & Grin, 2007; Morrow, 2005; Sousa, 2014; Thomas & Magilvy, 2011).

Replicability in qualitative research, just as is the case in quantitative traditions, is concerned with the integrity and trustworthiness of the data, the generalizable meanings generated from the data, and how those meanings are communicated. Higher replicability would suggest that information (e.g., central themes) derived from text-based data is more versus less consistent between different analysts. With greater replicability, analytic output would be viewed as more dependable and generalizable, as opposed to anecdotal or unique to a given observer or observer team. To date, approaches for enhancing the replicability of, and thus trust in, qualitative analysis include methods like data triangulation to reduce investigator bias, data auditing to “check” primary raters, and rater reflexivity to balance participant meaning and researcher interpretation (Kisely & Kendall, 2011). These methods, though, are inherently more subjective than the statistical parameters to establish replicability in quantitative analysis (e.g., a numerical index of

interjudge reliability). Thus, qualitative methods and their yield remain open to skepticism.

Perhaps a central challenge to establishing replicability, and stability of scientific footing, in qualitative methods is the number of traditions from which these methods arise, including post-positivist (which assumes the existence of an objective reality that is only imperfectly apprehendable), constructivist-interpretive (which assumes the existence of multiple apprehendable, and equally valid, realities), and critical-ideological (which privileges the researcher's values and intention to disrupt and challenge the status quo; Guba & Lincoln, 2005; Ponterotto, 2005; Schwandt, 1994). These diverse philosophical approaches to scientific inquiry pull for varied epistemological, ontological, and methodological considerations from both qualitative researchers and general consumers of qualitative findings (Staller, 2013). Moreover, the differences between these traditions are not always clear, and they continue to evolve and diversify in their goals and procedures (Gergen, 2014). Hence, there is no unified or definitive voice in how best to establish replicability of qualitative results, nor is replicability always a goal.

Despite these paradigmatic differences, some researchers have made trans-epistemological efforts to advance the replicability and integrity of qualitative analysis (Onwuegbuzie & Leech, 2007; Whittemore, Chase, & Mandle, 2001). For example, the Society for Qualitative Inquiry in Psychology, a section of Division 5 (Quantitative and Qualitative Methods) of the American Psychological Association (APA), appointed a *Task Force on Resources for the Publication of Qualitative Research*. This initiative culminated in recommendations for designing, presenting, and reviewing qualitative research in psychology (see Levitt, Motulsky, Wertz, Morrow, & Ponterotto, 2017),

though, importantly, the Task Force acknowledged the diversity and complexity of qualitative methods. They stated,

We do not propose to replace methods themselves, close down discussion of differences among research designs, nor hinder their development by setting in place a new set of fixed procedural rules. Rather, we propose foundational principles that can complement discussions of specific research methods, promote dialogue, and support the continued evolution of qualitative methods. (p. 7)

Among these principles were (a) fidelity to the subject matter under investigation, and (b) utility in achieving research goals. Thus, although embracing diversity of approach, the general community of qualitative inquiry has seemed to acknowledge that it is important and possible to establish the replicability and integrity of analytic output.

In one contemporary approach, consensual qualitative research (CQR; Hill, 2012) integrates the role of multiple perspectives, relying on a team of judges to prevent individual biases and data auditors to minimize groupthink. Originally developed and characterized as a rigorous and feasible alternative to other qualitative methods (Hill et al., 2005), and one that possesses the qualitative analogue to quantitative reliability in data analysis, CQR has developed burgeoning support across the social sciences, in part, because of its team-based approach and step-by-step training manual (Hill, 2012).

Although rigorous, multistep, and team-based methods like CQR were developed with scientific soundness in mind, they are not without limitation, including with regard to replicability of method and rater perspective. Arguably, one of the most compelling markers of replicability (and, hence, greater trustworthiness of qualitative results) would be high similarity of the analytic output between two different teams trained uniformly and responding to the same stimulus material. Such demonstration would go a long way toward convincing the scientific community that such output was not simply (even if

richly) a highly contextualized and idiosyncratic “collection of anecdotes” (Williams & Morrow, 2009). Whereas team-based methods like CQR incorporate checks and balances for *within*-team reliability, they do not inherently involve procedures for assessing *between*-group replicability.

Given the increasing use of CQR across the social sciences, and the virtual absence of empirical investigation into the between-group replicability of this method, the field has called for research of this kind. As Williams and Morrow (2009) aptly noted,

We would be interested to see the outcome of two teams using one analytic strategy (such as CQR) to analyze the same set of data. We would also be interested to see the outcome of two teams using different analytic strategies (e.g., CQR and grounded theory) analyzing the same data. In either case, how would the composition of the team or the use of a different analytic strategy alter the final analysis of the data? (p. 581)

Additionally, the CQR manual itself states, “finding innovative ways to demonstrate trustworthiness could advance the credibility of the CQR method in the scientific community” (Williams & Hill, 2012, p. 182). These calls are consistent with the spirit of methodological integrity at a broader qualitative level established by the aforementioned APA Division 5 Task Force (Levitt et al., 2017).

I am aware of only one study to date that has examined the replicability of output between two independent research teams of experienced judges using the CQR paradigm (Ladany et al., 2012). Although the results suggested overlap in the central themes that emerged from the teams’ analysis of interview-based data, considerable between-group divergences were revealed in their more detailed output. Given that the two teams had been uniformly trained in CQR, and had worked together on many other studies, the fact that they still produced different outputs from the same stimulus set suggests that the issue of replicability merits additional consideration. Moreover, the issue of replicability

may be particularly salient when CQR is conducted with judges new to qualitative inquiry.

Also, to the extent that different CQR teams produce convergent and/or divergent output, it seems important to examine the process through which different teams' raters arrive at their analytic output. In fact, the multiple perspectives characteristic of CQR provides its own challenges to within-team reliability. Such interjudge reliability (i.e., the degree to which judges similarly view the phenomenon under study) is often challenged by the interplay of team dynamics, especially if degrees of power distinguish team members (e.g., an undergraduate student and postdoctoral fellow, or a graduate student and a professor), or if a general group climate is experienced unfavorably (Sanders & Cuneo, 2010). Additionally, factors like interpretive and philosophical differences, stylistic variance, and emotionality of team members can all influence the analytic process in diverse ways. As Sanders and Cuneo stated, "reliability, we contend, is not a strictly statistical or logical entity but is a socio-emotional entity and process; hence, our term, 'social reliability'" (p. 339). Thus, the concept of social reliability presents its own unique challenges to team-based approaches, including CQR.

Considering the state of the literature, it seems important to establish systematically the between-team replicability of CQR output when investigating the same material, and to explore the process through which the teams arrive at their output (as a *possible* determinant of output convergence or divergence). To this end, in the present exploratory study, I used multiple methods to evaluate the similarities and differences that emerged between two uniformly-trained CQR coding teams of neophyte judges in terms of their (a) analytic yield based on the same stimulus set (i.e., replicability), and (b)

coding process (i.e., social reliability). Replicability was examined quantitatively and qualitatively on the CQR output elements of domains, categories, core ideas, and core idea exemplars. Social reliability was examined quantitatively and qualitatively by comparing the teams on objective group process and self-reported group climate. Squaring with CQR's extensive use in the field of psychotherapy research, the interview data that were analyzed derived from a parent study that examined mental health care (MHC) consumers' values and preferences regarding the use of MHC provider performance data (grounded in patient-completed routine outcomes measures; see Boswell et al., 2018).

Given the rich, descriptive nature of CQR, like all qualitative methods, it is near impossible to establish definitively a "black-and-white" conclusion about between-group replicability. In the same way that qualitative approaches deal in shades of gray, replicability, and its corresponding trustworthiness, exist on a spectrum. Additionally, depending on the paradigmatic lens through which one interprets this study's findings (e.g., post-positivist, constructivist-interpretive, critical-ideological), the results could be understood in myriad ways. Indeed, Hill (2012) acknowledges *both* the constructivist and post-positivist elements that constitute CQR. Consequently, I did not hypothesize about whether the present study would yield results considered to be trustworthy, but rather focused on presenting both rich descriptive and preliminary quantitative data to allow readers to draw their own conclusions in accordance with their own epistemological frameworks. Consistent with the efforts of Levitt and colleagues (2017), I hope that this carefully conducted exploratory study informs the practical utility of existing CQR findings, as well as future methods for optimizing CQR process and output integrity.

CHAPTER 2

METHOD

2.1 Participants

2.1.1 CQR Trainer

As an investigator with extensive experience administering and training others to administer CQR, I trained together two independent rater teams and data auditors on the CQR protocol (Hill, 2012). After conducting the training, I was not directly involved in the analysis of the stimulus materials. Thus, all participants experienced the same training and training climate, with no interference from me on the coding process or outcome, thereby allowing CQR output elements and the social reliability indices to be unconfounded by my presence/influence. I also trained four raters to conduct thematic analyses on the social reliability data stemming from the focus groups.

2.1.2 CQR Judges

The two rater teams each consisted of four undergraduate research assistants who were working for credit (in the Psychotherapy Research Laboratory at the University of Massachusetts Amherst) and who consented to take part in the study. The judges, none of whom had prior experience with CQR, were randomly assigned to team. Judges ranged in age from 20 to 24 ($M = 21.25$, $SD = 1.39$). Half of the judges on each team had utilized mental health services in the past. Team 1's team consisted of three women and one man. Two judges identified as White, one identified as East Asian, and one identified as other race/ethnicity. Team 2's team also consisted of three women and one man. Two judges

identified as White, one identified as East Asian, and one identified as African American.

2.1.3 CQR Data Auditors

Each coding team was randomly assigned one data auditor. To prevent the merging of ideas or confounding of output between teams as a result of only one study auditor, each team was assigned its own independent auditor. Each auditor had earned a bachelor's degree and was working in the Psychotherapy Research Laboratory; one was a 26-year-old first-year graduate student who identified as a White woman and the other a 23-year-old project coordinator who also identified as a White woman. Neither auditor had prior experience with CQR.

2.1.4 Output Replicability Judge

I, a 30-year-old who identifies as a White man, served as the sole judge of output replicability in that I decided how best to display the findings of the two coding teams; however, these data emerged directly from the analyses of the two teams and were presented as reported by each team. The only degree of subjectivity that emerged was how I chose to present areas of convergence/divergence between team output. Given that I trained all study participants and reviewed the output generated by both teams, I was deeply immersed in the data and keenly aware of all results. This rich immersion is consistent with the well-established constructivist framework of “researcher-as-instrument” (Ellis & Berger, 2003; Pezalla, Pettigrew, & Miller-Day, 2012).

2.1.5 Social Reliability Raters

Two graduate students and two undergraduate students with no familiarity with the current project, and with no close relationships to the CQR judges, thematically analyzed the transcripts of the focus groups that assessed the CQR judges' experiences of their own team's process and climate. The social reliability raters were all White women who ranged in age from 20 to 29 ($M = 24.00$, $SD = 4.69$).

2.2 CQR Stimulus Set

The text-based data qualitatively analyzed in this study derived from the aforementioned parent study examining MHC consumers' values and preferences regarding the use of MHC provider performance data (Boswell et al., 2018). Participants were community MHC patients presenting for treatment at one of 12 community mental health centers (CMHCs) in eastern Massachusetts or central New York. The CMHCs employed diverse providers delivering a range of behavioral health services, including pharmacotherapy, individual and group psychotherapy, couples and family therapy, and case management.

To be included in the parent study, patients had to be (a) seeking or receiving MHC at a participating outpatient clinic, and (b) responsible for the MHC decisions for oneself, a family member, or a significant other who was unable to participate on his or her own behalf (e.g., due to severe cognitive impairment). There were no exclusion criteria for the type of service being sought (or received), presenting problem or diagnosis, or demographic profile. Of the subsample of participants ($n = 36$) who engaged in a semi-structured interview (see Appendix A for the Consumer Telephone Interview protocol), I randomly selected 12 interview transcripts specifically from

participants making MHC decisions for *themselves* for the present analysis (a number commensurate with the suggested sample size for CQR; Hill, 2012). Table 1 provides a summary of the demographic characteristics and presenting problems for the 12 interviewees whose transcripts were analyzed. All interviews were previously transcribed and de-identified by research personnel not involved in the present study. As per CQR protocol, judges generated the typical CQR output from the 12 transcripts; that is, *domains* (a list of the meaningful and unique topic areas examined in the interview), *core ideas* (summaries of the data that capture the essence of the participant's statement in fewer words), *categories* (clusters of similar core ideas identified by common elements or themes across participants), and *exemplars* (representative core ideas to be used in a report to bring a content area to life). These steps are described in more detail below.

2.3 Measures

2.3.1 Replicability

I assessed replicability between rater groups with both quantitative and qualitative methods. The objective, numerical output indices included frequency counts of domains, categories, and core ideas generated by each team, for which I then qualitatively compared their contents and descriptively presented their similarities and dissimilarities. Given the large number of core ideas generated by both teams, only the core idea exemplars were qualitatively compared.

2.3.2 Social Reliability

I assessed social reliability on multiple levels across all 12 transcripts (i.e., the process of moving through the CQR steps to yield the domains, categories, and core ideas). These levels included objective group process and self-reported group climate, as per the indices/measures described below.

2.3.2.1 Objective Group Process

The objective group process indices for each rater team were: the average length of time (in minutes) that each team member spent analyzing study transcripts, the discrepancy in total time (in minutes) spent talking (across all transcripts) between the most and least talkative group member, and the number of times that the auditor disagreed with the team and thereby prompted consensus discussion and resolution.

2.3.2.2 Self-Reported Group Climate

To assess self-reported group climate, the CQR judges completed the study-specific CQR process measure, which included eight items adapted from the group psychotherapy (Burlingame, McClendon, & Alonso, 2011) and psychosocial working environment (Pejtersen, Kristensen, Borg, & Bjorner, 2010) literatures (see Appendix B). Specifically, the measure assessed, from low to high, each judge's experience of adaptive group climate, both in terms of individual contributions to the group experience and the group-as-a-whole. The items were rated on a 7-point scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), with one item reverse scored. The theoretical range of the total score was 8 to 56, with a higher score representing more positive perceived climate. Judges completed the measure specifically within the context of tasks involving

consensus coding (i.e., generation of content from the coding team's perspective of the data). The sample's internal consistency for this measure was adequate (average $\alpha = .70$).

After completing the CQR analysis, judges on each team also participated in a 1-hour focus group designed to elucidate greater insight into their experience of the CQR process (see Appendix C). The focus groups, rather than individual interviews, capitalized on the joint perspectives of the participants to attend to group climate and experience (Krueger & Casey, 2009; Liamputtong, 2011). Similar focus groups have a precedent in the psychotherapy literature (e.g., Chronister, Chou, Kwan, Lawton, & Silver, 2015; Constantino, Morrison, MacEwan, & Boswell, 2013). Additionally, to ensure judges felt comfortable sharing all aspects of their experience, including information they may not have felt comfortable sharing in a focus group setting, they were asked to complete an anonymous feedback form after the focus group via Qualtrics, a secure, web-based platform (see Appendix D).

2.4 Procedure

Consistent with CQR methodology, the timeline for the present study accounted for a training phase (3 months), coding phase (6 months), and an analysis of coding process (6 months). A detailed description of each phase is presented below, and a study flow chart is presented in Appendix E, which delineates the time points at which the CQR process measure was administered. All aspects of the study that relied on consensus coding were both video- and audio-recorded to analyze objective group process. The University's institutional review board approved and provided oversight of this study, and the participants were treated in accordance with the APA's Ethics Code (American

Psychological Association, 2010).

2.4.1 Training

In the training phase, judges and auditors completed a series of tasks to ensure that they were intensely familiar with the CQR paradigm. They first attended a project orientation to discuss the timeline of the project and the study protocol. As a single group, the judges and auditors then read the CQR manual and two empirical papers that relied on CQR for data analysis (i.e., Knox, Burkard, Johnson, Suzuki, & Ponterotto, 2003; Vivino, Thompson, Hill, & Ladany, 2009). The judges and auditors then watched a series of three training videos created by the PI. These videos not only reviewed the key elements of CQR described in full below, but also reviewed and standardized the study-specific procedures (e.g., tracking progress of transcript coding, storing and accessing study-specific materials, formatting and preparing documents for data auditors). All videos reviewed the steps needed for independent generation of content (i.e., one judge's perspective of the data), consensus generation of content (i.e., the coding team's perspective of the data), and auditing of the content (i.e., the role of the data auditor in relation to the team of judges). In the first video, the judges and auditors learned how to generate domains. In the second video, judges and auditors learned the steps necessary to develop core ideas. In the third video, judges and auditors learned how to generate categories for the data and analyze data across transcripts.

After completing the video trainings, the judges reviewed the interview questions asked of the parent study participants and recorded their expectations and biases per CQR protocol; moreover, to further immerse judges in the perspectives of the parent study

participants, judges were asked to complete via Qualtrics the same demographic form completed by these participants (see Appendix F). At this point, the judges and auditors were randomly assigned to one of two coding teams. Subsequent to this random assignment, the two teams functioned entirely independently of one another for the remainder of the project.

Next, in order to develop a uniform coding style, judges independently completed open coding of two pilot transcripts (neither of which were one of the 12 transcripts selected for the present study) before coming to consensus as a team. Each team's auditor had the opportunity to audit the pilot materials, and all team members were able to discuss the process and ask questions before proceeding to domain generation and open coding of the actual study transcripts. If any questions arose throughout the coding process, judges were instructed to first consult with members of their coding team, then consult with their team's data auditor, and finally to contact the PI via email (if questions remained unanswered). In the latter case, I would not answer questions about CQR-specific queries (judges and auditors were encouraged to refer to the CQR manual and the other aforementioned training materials), but would address logistical questions related to the study by including all members of both teams in my email responses. Thus, all study judges would receive the same information from me regardless of which team member asked the question.

2.4.2 Domain Generation

After completing the training phase, the teams began generating initial domains for the study data. Each judge (nested within team) independently reviewed the 12 study

transcripts to immerse themselves in the data, generated domains, and established memos as they read. After generating domains independently, the teams came to consensus on the initial set of domains. Although domains could potentially shift throughout analysis, the initial domains constituted the core structure and initial lens through which the judges analyzed the data. After completing the consensus meeting for domains, judges completed the CQR process measure. Each team's auditor reviewed the domains and made recommendations for adjustments as needed, and the teams accepted and rejected the recommendations accordingly.

2.4.3 Open Coding

During the open coding phase, the judges independently established core ideas for the data, moving through one transcript at a time. Next, the team came together to achieve consensus on core ideas, placing each into one or more of the domains they created. Per CQR guidelines, study judges were free to decide the length of each core idea and the degree of detail captured by each core idea. Team meetings took place weekly to foster consensus and to avoid rater drift. After completing every third consensus meeting, judges completed the CQR process measure. Each team's auditor reviewed the core ideas and made recommendations for adjustments as needed, and the teams accepted and rejected the recommendations accordingly.

2.4.4 Categorization and Cross Analysis

After completing open coding, judges established categories for each domain. Each category consisted of the consensus core ideas established by the coding team, and

teams achieved consensus on finalized categories. Each team's auditor reviewed the categories and made recommendations for adjustments as needed, and the teams accepted and rejected the recommendations accordingly.

Lastly, during cross analysis, judges achieved consensus on the placement of core ideas into categories within their respective domains across all study transcripts, and provided exemplar core ideas for each category to provide the reader with additional context. Per CQR protocol, judges placed each core idea into as many categories and/or subcategories as they deemed appropriate for that core idea. After finishing the cross-analysis meeting, judges completed their final CQR process measure. Each team's auditor reviewed the cross analysis and made recommendations for adjustments as needed, and the teams accepted and rejected the recommendations accordingly. This last phase concluded the CQR analytic process. Judges then participated in a focus group interview and subsequently completed the anonymous feedback form.

CHAPTER 3

RESULTS

3.1 Replicability

In this section are the frequency counts of domains, categories, and core ideas, as well as my qualitative comparisons between the groups. As this qualitative analysis was the primary assessment of output replicability, I used narrative description to present points of convergence and divergence.

3.1.1 Domains

The sets of domains generated by each team are presented in Table 2. Team 1 generated 4 domains and Team 2 generated 7 domains, and both sets largely mapped onto the *major* questions inherent to the Consumer Telephone Interview Items (see Appendix A). However, as reflected in the count, there was a difference between the groups with regard to abstraction level. Team 1 generally developed broader, overarching domains. Judges on this team attended to the *experiences* and *attitudes* of study participants. Conversely, Team 2 developed domains that focused on more specific elements of participant experiences. Judges on this team attended to the *past* and *future language* of study participants.

3.1.1.1 Team 1 Domain 1 vs. Team 2 Domains 1 and 6

Team 1's Domain 1 (Experience with mental health care services) was most similar to Team 2's Domain 1 (Past and current experiences with mental health services). In both cases, the language of these domains was similar to the first item of the interview.

Related to engagement in MHC, Team 2 also generated their Domain 6 (Experience with surveys/questionnaires), whereas Team 1 classified such experience as a category within their Domain 1 (as reviewed in the category results below).

3.1.1.2 Team 1 Domain 2 vs. Team 2 Domains 2 and 5

Team 1's Domain 2 (Experience selecting mental health services) was most similar to Team 2's Domain 2 (History with mental health care provider selection process), and to a lesser extent Domain 5 (Possible improvements to mental health care selection process). Again, Team 1 maintained a broader abstraction at the domain level with a focus on participants' general experience, whereas Team 2 had finer distinctions between domains by focusing on *both* past and possible future experiences.

3.1.1.3 Team 1 Domain 3 vs. Team 2 Domains 3, 4, 5, and 6

These domains were the most divergent between the teams, though the general pattern held. That is, whereas both teams captured similar participant experiences, they differed as to whether they adopted a more global or nuanced focus in content. Team 1's Domain 3 (Attitudes about the selection process) centered on participants' global attitudes toward the various aspects of the MHC selection process. As before, Team 1 chose to make more nuanced distinctions at the level of categories and subcategories (e.g., providers, services, insurance), as discussed below. Conversely, Team 2 generated domains that addressed more specific aspects of MHC selection, including information related to the MHC provider, as identified in Domains 3 (Important factors selecting a mental health provider) and 4 (Availability and validity of provider's background

information), and logistical improvements across systems, as indicated in Domain 5 (Possible improvements to mental health care selection process).

3.1.1.4 Team 1 Domain 4 vs. Team 2 Domain 7

Team 1's Domain 4 (Attitudes on preferred providers) matched most closely to Team 2's Domain 7 (Opinions about well-matched provider). Both teams generated very similar language pertaining to participant attitudes about and/or opinions on the selection of preferred or well-matched providers.

3.1.2 Categories

Overall, the lowest output replicability between the teams occurred in the generation of categories, subcategories, and sub-subcategories whose representativeness were classified as follows: (a) "general" (emerged for 11 or 12 participants), (b) "typical" (emerged for 7-10), and (c) "variant" (emerged for 2-6). Per CQR protocol, categories that were deemed "rare" (i.e., endorsed by only 1 participant) were omitted from the displayed results. As indicated in the Table 2 frequencies, Team 2 generated twice as many categories as Team 1 (40 and 20, respectively) and more than twice as many subcategories (152 and 63, respectively). Part of this difference was accounted for by Team 1's decision to include sub-subcategories in their analytic output (38 total), which Team 2 did not (0 total). However, even when accounting for Team 1's sub-subcategories, Team 2 still generated considerably more subcategories. See Tables 3-6 for Team 1's categories, subcategories, and sub-subcategories (organized from most to least representative) for each of the four domains that this team generated; see Tables 7-

13 for Team 2's categories and subcategories (organized from most to least representative) for each of the seven domains that this team generated.

For both teams, many of the subcategories and sub-subcategories simply represented forms of "nominal" data. For example, Team 1's Domain 2, Category 4 (Patient's say in the mental health care selection process) included Subcategory 4.1 (Total say), Subcategory 4.2 (No say), and Subcategory 4.3 (Some say). As another example, Team 2's Domain 4, Category 3 (Preferred way to access PPI) included Subcategory 3.1 (Would prefer information online) and Subcategory 3.2 (Would prefer information via handout). The categories generated by both teams in this manner were especially comparable. For example, Team 1's Domain 2, Category 1 (Patient's knowledge about a provider prior to the first appointment) and Subcategories 1.1 (Some prior knowledge) and 1.2 (No prior knowledge) map closely onto Team 2's Domain 2, Category 5 (Knowledge of provider before first appointment) and Subcategories 5.1 (Some knowledge of provider before first appointment), 5.2 (Minimal knowledge of provider before first appointment), and 5.3 (No knowledge before first appointment).

However, category and subcategory representativeness differed between teams. In the aforementioned example, Team 1 suggested that all 12 participants (general representation) discussed Domain 2, Category 1 in some capacity; Team 2 did not place any core ideas into their Domain 2, Category 5, which renders it unclear how many participants addressed this theme in some capacity. However, the reader can see that Team 1 noted that 8 participants (typical representation) discussed having some prior knowledge about a provider prior to the first appointment, whereas Team 2 noted that only 3 participants (variant representation) indicated having this level of prior knowledge.

Even if Team 2's Subcategories 5.1 (Some knowledge of provider before first appointment) and 5.2 (Minimal knowledge of provider before first appointment) were consolidated into one subcategory, the results for Team 2 would still yield variant representation. This example highlights a comparative theme throughout the results; even in instances where considerable overlap existed between established categories and subcategories, the representativeness of the participants who discussed them was often notably different between teams.

As another way of contrasting the representativeness of their findings, Team 1 reported a range of general, typical, and variant findings across all of their domains. Conversely, Team 2 did not report any general findings across any of their domains, and only a limited number of typical findings. Team 2's findings were overwhelmingly variant. Thus, whereas Team 1's findings suggest that a number of shared experiences emerged between participants, Team 2's findings suggest there was little commonality between these very same participant responses.

3.1.3 Core Ideas

In total, Team 1 generated 291 core ideas and Team 2 generated 314 core ideas. Although the frequencies were rather close, the execution was disparate between the teams. Specially, during the cross analysis, Team 1 generally placed their core ideas into a category and, when applicable, an appropriate subcategory or subcategories within that category. Conversely, Team 2 often chose only to place a core idea into an appropriate subcategory of a category, and would only place that core idea at the category level when the core idea did not fit into a corresponding subcategory. For example, Team 1 placed

the core idea “P has worked with LCSW, LMHC, unlicensed mental health counselors, psychologists, and psychiatrists for medication management” in Subcategories 1.1 (Doctorate-level providers [psychiatrists, psychologists, PCP]) and 1.2 (Masters-level and other providers [LMHC, LCSW, other counselors]) and Category 1 (Mental healthcare providers worked with) of their Domain 1, whereas Team 2 placed their similar core idea “P has previously received therapy from doctors, social workers, doctoral psychologists and licensed mental health care workers; P received medication and counseling from a psychiatrist” in *only* Subcategory 2.1 (Has worked with multiple types of providers) and not also Category 2 (Types of providers worked with) of their Domain 1 (see Tables 3 and 7, respectively).

Given the large number of core ideas, only exemplars selected by each team for each category are presented in Tables 14-24. Consistent with CQR protocol, there were some instances in which the teams selected core ideas as exemplars for multiple categories. The core idea exemplars presented in Tables 14-24 have been displayed to mirror the categories/subcategories presented in Tables 3-13. (As noted, Team 2 did not always place core ideas at the category level.) As revealed in Tables 14-24, the content of the core ideas (and, thus, the exemplars) were notably similar between the two teams.

3.2 Social Reliability

3.2.1 Quantitative Analyses

The context for social reliability indices was the consensus coding of core ideas, as the teams approached the categorization, cross analysis, and auditing aspects of their

work differently. For example, Team 1 generated their categories together as a team, whereas Team 2 generated them independently before coming together to accept from their individual lists which categories they felt best fit the data. Additionally, when it came time to audit throughout the entirety of the process, Team 1 reviewed written feedback sent by their auditor via email, whereas Team 2 would begin each coding meeting by teleconferencing with their auditor to discuss the audits. Thus, to provide the most standardized comparison of time spent discussing uniform aspects of the CQR process, the objective social reliability indices were specific to the time study judges spent coming to consensus on core ideas across the 12 study transcripts, not including time judges spent accepting/rejecting auditor feedback.

Regarding time spent analyzing study transcripts, Team 1's average judge spent a total 114.11 minutes ($SD = 22.18$) analyzing all study transcripts, as compared to 88.84 minutes ($SD = 58.58$) for Team 2's average judge. Overall, this difference represented a moderate effect ($g = 0.50$). When descriptively examining each transcript individually, Team 1 spent more time discussing transcripts than Team 2 on 10 of 12.

Regarding the discrepancy in minutes spent talking when analyzing a given transcript, Team 1's most talkative group member spent a total (across all transcripts) of 145.88 minutes talking, or 32% of that team's total time spent consensus coding core ideas, whereas the least talkative group member spent a total of 96.75 minutes talking, or 21% of that team's total time spent consensus coding. For Team 2, the most talkative group member spent a total of 173.37 minutes talking, or 49% of that team's total time spent consensus coding, whereas the least talkative group member spent a total of 39.42 minutes talking, or 11% of that team's total time spent consensus coding.

Regarding the total number of auditor disagreements requiring consensus discussion and resolution, Team 1's auditor disagreed an average of 5.50 times per transcript ($SD = 3.92$) and Team 2's auditor disagreed an average of 7.00 times per transcript ($SD = 4.88$). This between-group difference represented a moderate effect ($g = -0.33$). Auditors addressed multiple concerns throughout the coding process, including "tie breaks" (e.g., resolving discrepancies when two judges advocated for a core idea written one way and two judges advocated for the same core idea written another way), clarification for study judges (e.g., helping study judges to write a core idea when they were unsure about the best way to capture a participant's thoughts), and general disagreements (e.g., asking study judges to revise content on which they had already come to consensus). Auditor feedback often addressed a variety of these aforementioned concerns. For example, when addressing a tie break for a core idea, an auditor may have instead advocated for rewriting it into several core ideas. Thus, auditors had free reign to address the judges' output as they saw fit. When examining each transcript individually, Team 2 had more disagreements than Team 1 for six transcripts, Team 1 had more disagreements than Team 2 for four transcripts, and the Teams had the same number of disagreement for two transcripts.

Regarding self-reported group climate as assessed by the CQR process measure, scatter plots revealed no outliers (any score $> 2 SDs$ from the mean) that would have had an undue influence in small samples. Descriptive statistics are presented separately by team in Table 25. The average group climate score across all time points was 53.57 ($SD = 2.59$) for Team 1 and 51.39 ($SD = 2.85$) for Team 2. This between-group difference represented a moderate-to-large effect (average $g = 0.70$).

Additionally, I visually examined between-team differences in within-team climate change across all process measure administrations. As depicted in Figure 1, although the teams started with similar average ratings of group climate, Team 1 generally reported an increasingly positive average climate over time, whereas Team 2 generally reported a decreasing average climate over time. Descriptively, within teams, there were no time points for which all members agreed on the same direction of change from the previous time point. However, Team 1 members generally reported greater within-team agreement on directional shifts in climate than Team 2. For example, the largest within-team discrepancy in perceived change from the previous time point was 3 points for Team 1 (i.e., one member perceived a 3-point increase in climate and another perceived no change from the previous time point) and 6 points for Team 2 (i.e., one team member perceived a 5-point decrease in climate and another perceived a 1-point increase from the previous time point). Finally, Team 1 perceived smaller timepoint-by-timepoint changes in their climate (M absolute change from one time to the next = 0.58) compared to Team 2 (M absolute change from one time to the next = 1.25), suggesting that Team 2's climate may have been more volatile.

3.2.2 Qualitative Analyses

For the qualitative assessment of social reliability, the independent raters transcribed the focus group responses and applied thematic analysis to the text-based data, an inductive method for gaining a rich understanding of participants' perceptions of target phenomena (Braun & Clarke, 2006). I trained the independent raters on the method, which included attending a project orientation, recording their biases and

expectations before seeing the data, reviewing Braun and Clarke’s protocol, and completing pilot coding of a de-identified focus group from a different project and coming to consensus on the themes. After training, the raters completed the current study’s thematic analyses independently by (a) initially reading though the focus groups’ text with an accompanying video recording, (b) re-reviewing the text while “memoing” potential themes, and (c) reviewing the text a third time to establish codes on which to build their themes. To reduce bias, half the raters analyzed Team 1’s focus group first, whereas the other half analyzed Team 2’s focus group first. After all raters had reviewed both transcripts independently, the raters then came to consensus on final themes.

3.2.2.1 Team 1’s Focus Group

The analysis of Team 1’s focus group data yielded 5 distinct themes, which in some instances were comprised of sub-themes (see Figure 2). Presented below are the themes and their core components, along with exemplar quotes from focus group participants.

3.2.2.1.1 Feelings About the Project Changed Over Time

The judges’ feelings about the project changed over time and revolved around multiple points, including: (1) learned how to make core ideas more concise (“I think the issue I had with the core ideas in the beginning was I didn’t really know what we were doing with them, so I didn’t how I should structure and parse them in the beginning”); (2) different levels of confidence in codes (“I guess I found myself a couple times when people had confidence, I would find myself kind of just going along with it”); (3) unsure

of what questions the study was trying to answer and whether they were on the right track (“I don’t know if my ideas are going to be even remotely on track, I was like so nervous”); (4) steps were not straightforward (“I feel like none of [the project aspects] were straightforward”); (5) consensus became easier over time (“After the first couple transcripts, I kind of got into a rhythm and it went faster”); (6) tried to piece together core ideas at first and shifted method over time (“I think as the process continued, I structured my individual core ideas kind of in the same format as the consensus that we were coming to”); and (7) initially stuck on the wording of core ideas (“I thought it was hard at first juggling everyone else’s version of the core idea, but I got better at looking for the differences and not getting caught up in the wording and stuff”).

3.2.2.1.2 Particular Concerns Related to Time and Efficiency of the Project

Judges commented on logistical concerns related to the project, particularly in the context of time: (1) coding process was generally time-consuming (“I never was sitting there thinking I didn’t want to be doing this, it was just like this is a long time frame to be doing this in one sitting”); (2) mentally checked out at times, especially at the end (“Towards the end of some consensus meetings I was mentally already checked out”); (3) logistically complicated with lots of documents to manage (“The last bit took a lot, it was time consuming I think a little bit because of the excel spreadsheets, just flipping through the pages”); and (4) project turn-around time was rushed/tedious (“I would just say the turn-around time made it, I wouldn’t use the word burnt out, but it was kind of hard to find the motivation to go through this whole transcript for the next meeting”).

3.2.2.1.3 Group Dynamic Was Positive and Comfortable Overall

Judges discussed a number of instances in which they felt positive about their dynamic and supported by their fellow judges, touching on a variety of elements: (1) comfortable with group members and got to know each other over time (“I personally had a good time, like I think we all got to know each other so well”); (2) comfortable speaking up and providing feedback or expressing a different opinion (“I felt so comfortable with [fellow judges] that I honestly never even was hesitant to speak up, like it just felt normal”); (3) enjoyed the work, especially compared to other group projects (“I feel like so many other group projects I work on there’s always like that one person that doesn’t do anything...or is very firm on their way, and I really don’t feel like that happened that much”); (4) learning experience (“Doing it on our own and coming in and comparing and seeing what others had taken from a core idea sparked interest, but then it was also like a learning experience”); (5) good communication across the group (“Communication was way easier than working with other people that I worked with in past years...so that kind of helped make everything faster and move forward”); (6) valued others’ input and perspectives (“I don’t think anyone thought they were right all of the time...we were very flexible about considering other perspectives and valuing each other’s input”); and (7) overcame challenges (“Everyone here wanted the result and outcome to be the best it could be I think, and so I just feel like that works better”).

3.2.2.1.4 Factors Conducive to Project Completion and That Would Be Helpful in the Future

Judges identified a number of factors they found helpful (or limited) for project

completion and which they believed would be helpful for CQR projects in the future: (1) research articles (“I think the research articles might have been more helpful than reading [the CQR manual] because they sum up the result, like how we’re going to use all the information and present it”); (2) video training and outlines (“There wasn’t too much referencing the manual, most of the sense I got of what we should be doing came from the training videos”); (3) prior project experience (“Last year I helped with [thematic analysis], and it was the same concept so I saw the big picture, so it was easier for me to kind of make core ideas and then do categorization”); (4) limits to the CQR manual (“I don’t think the manual really explained well what quality results would look like, like what I would be shooting for to begin with”); and (5) Google hangout/docs (“For example, say that someone was typing and there was a typo, one of us would have to say there’s a typo, whereas if it was Google docs, we could’ve easily just one of us while she’s typing gone up and fixed it”).

3.2.2.1.5 All Steps of the Process Presented Unique Difficulties

Judges acknowledged that each aspect of the project presented its own challenges: (1) transcripts were sometimes vague and it was hard to extrapolate and figure out important points (“Some of the responses were very vague and we had to be really careful not to extrapolate, and then some of them were way too detailed that we had to sort through all of it and figure out what’s important”); (2) domains were challenging/involved and overwhelming (“Domains too were challenging because we just weren’t sure what was going to fit in them”); (3) making core ideas was difficult, and was the longest step (“Going back and editing and re-organizing all the core ideas would’ve

been a long process, and I don't think it's feasible to do that in the project timeline"); (4) felt comfortable with core ideas ("I personally enjoyed making core ideas because everyone has different views; when we read the same sentence we have all different core ideas sometimes, so it was really fun to watch"); (5) categorization was more straightforward due to the structure ("I think categorization, like making our own category was easier and more straightforward along with core ideas"); (6) cross-analysis was the hardest part ("The last two meetings we were kind of time crunched and we thought some categories weren't exactly what we want them to be"); and (7) group-think throughout could lead to wrong conclusion ("I felt like a couple times where we had come to a consensus about doing things one way and we all agreed, and then the next meeting without realizing it we would come to a consensus about doing it differently").

3.2.2.2 Team 2's Focus Group

The analysis of Team 2's focus group data yielded 6 distinct themes, which in some instances were comprised of sub-themes (see Figure 3). Presented below are the themes and their core components, along with exemplar quotes from the focus group.

3.2.2.2.1 Members of Group Had Unique Strengths But Worked Well Together

Judges discussed the interplay between the unique strengths they brought to the process as individuals and how they worked as a group: (1) there was a team effort even when there were differences in opinions ("It was like they care enough to say their opinion and why they feel that way, that's a really good environment to be in that they can speak up"); (2) personality type likely influenced choice of roles ("It's interesting to

see who chooses to get the computer, and whose personality decides to set everything up, and just be there, and who chooses to be at the computer and who prefers not to be”); (3) one judge spoke more frequently than others (“I felt like I was pretty outspoken and there were meetings where I wondered if I should maybe be a little quieter...and I did feel a little of a difference in those meetings”); (4) whichever judge had access to the computer often took the lead (“I think if I had the computer I already had set where I was going to put something, so I subconsciously put it there and then asked what people thought about it”); (5) group members were not afraid to speak their minds if they disagreed or agreed (“When you have a strong stance on a particular thing, obviously that would compel you to speak up, so I feel like when I do have something that’s not what everyone is agreeing on, I will speak up about that”); (6) judges felt more motivated when others spoke up and showed they care (“I saw that they’re putting in a lot of work and whenever anyone disagrees with me, it made me think that they care enough to speak up and have this conversation, it was actually motivating for me if someone disagrees”); (7) often one person would catch things another missed via checks and balances (“When it comes to domains and categorizations, we did independent work, but we all missed something, so it’s good to have a team to add something when you missed something”); and (8) groupthink may have played a role during times of feeling tired/drained (“I just think groupthink affected us most when we were a little bit drained, when we’ve been at it so long we just left it there, so I feel like that’s where groupthink may have affected us”).

3.2.2.2.2 Interest in Topic and Positive Work Environment Made the Project More Enjoyable

Judges commented on how the degree to which they felt invested in the topic and the working environment affected their perception of the project: (1) interest in topic made the judges more likely to express their opinions (“If I thought this is an idea that I don’t feel happens enough, I really want to make sure it’s included because I think it should be included in the work, so people can know that this is what patients want”); (2) judges got a better sense of the real research process (“I liked the way you read the booklet, then a video, then you actually do the task, and that’s what I think of this project because you really had to self-teach the qualitative method, so for me that’s what I think of the process of learning”); (3) judges felt more involved than they had in other projects (“I feel like in this project I’m really involved in this process because I work in some other labs and all we did was just run the participants, and we weren’t actually involved in the research process”); (4) seeing effort and quality of work was motivating (“He did a really good job with his core ideas, like excellent, thorough, and pretty concise, and I think seeing how much effort he put in like motivated me also”); (5) group members liked each other (“We all like each other, so I feel like that’s really great and helps a lot, we’re not like strangers thinking let’s just get this done and be out of here”); (6) laughing fostered a warmer environment (“I think we would laugh a lot when we were stuck, and I think it was just fun”); and (7) judges felt supported by the principal investigator (“[The PI is] very detailed especially in emails telling us what to do...that really, really helped us a lot, and the training videos were great as well”).

3.2.2.2.3 Suggestions for Improving Project in the Future

Judges offered various suggested improvements for future coding projects: (1)

experience with the process in past project was helpful, especially with speaking up (“I wondered if maybe I spoke up a little bit more because I already did this similar process, especially the domain thing, so I wonder if that affected anything”); (2) template for core ideas and domains would be helpful (“I feel like if we can have a template for the whole process, like one is for core ideas and then one for domains”); (3) would like to receive feedback on practice coding (“I feel like pilot coding, we did it but then we wouldn’t know from [the principal investigator] if how we’re doing it is right”); (4) reviewing each step before moving forward would have been helpful (“Looking back to core ideas or domains before getting to another new phase is beneficial because we might change our minds when we finish one core idea to one domain”); and (5) more general rotation of roles among group members would be preferable (“I think to what was said, rotating it, having that be a mandatory thing...but then again it’s interesting to see who chooses to get the computer”).

3.2.2.2.4 Factors Conducive to Project Completion and That Would be Helpful in the Future

Similar to the previous theme, judges identified factors that were helpful to them throughout the coding process: (1) doing pilot coding was helpful (“I think to practice coding with a transcript was very good so I could get a sense of the research we were doing”); (2) the parent study’s interviewer summaries in the transcripts were helpful to decipher participants’ main take-away points (“The interviewer comes back and tries to basically summarize this is what the participant means, to summarize just to make sure [the interviewer] got what was said”); and (3) auditor helpful to project overall (“When

we were stuck, we didn't have to bulldoze over each other, we'd put a memo, get an outsider's opinion, see what she thinks, and then decide, so I think that helped the group dynamic a lot just knowing that if we were disagreeing no one would have to win that").

3.2.2.2.5 Unknown and Freedom Were Frustrating but Led to Better Understanding and Performance

The judges described a sense of unknown at the outset of the project, which felt overwhelming and frustrating in some instances, but ultimately led to a better understanding of the project and investment in their performance: (1) there was a sense of unknown at start about what was coming next ("Just the uncertainty of what we're doing and sometimes what's coming next, like okay we're done with coding, what are we doing next?"); (2) transitions from one step to the next were difficult and abstract ("I feel like everything's fine, just each transition from core ideas to domains and domains to categorizations, the transition part is really hard, because although we watched videos and read [the manual] it's just still abstract for me"); (3) process became less confusing over time ("Once you jump into it you realize that it's like actually less confusing because you don't have restrictions and you are doing it right"); (4) it was hard to decipher meanings from the transcripts ("Sometimes a participant would contradict himself, which was really a pain"); (5) it was helpful to figure things out on own ("With this project there's a lot of leeway, so that uncertainty is what's difficult for me, but we did figure it out and went on with it, and that also gave us power as a team to say this is what we accept, and this is how we're moving forward"); and (6) the broadness and conceptualization of the project was generally confusing ("I think the project getting

broader and broader, like going from core ideas to domains to categories, it got confusing like how broad we should be going with that, so I struggled with that the most I think”).

3.2.2.2.6 Each Step Brought Challenges, but Cross-Analysis Was the Most Difficult

Although the judges acknowledged difficulty throughout the coding process, they felt that cross analysis was the most difficult of the project phases: (1) manual was difficult to read (“Reading the [CQR manual] was a bit tough because I thought this is new, what am I reading, I’m not really getting what I’m reading”); (2) cross-analysis was confusing (“For me, I feel like there was a lot going on in the cross analysis, so I feel like there were little things in there that had to play with like why I was confused but also I just didn’t understand, it was just so big and I thought I was just a little bit lost”); (3) some parts felt disorganized (“I think we realized that we didn’t do the best job organizing things because we ended up having some subcategories with nothing in them, so I think as a team that was where we kind of didn’t do so hot”); (4) many parts were more time consuming than expected (“I’d say [cross analysis] took longer than I thought, I thought it was going to be really quick but it felt pretty long when we were doing it”); (5) there were too many sub-categories (“I wonder what the other team had, but we had so many subcategories that probably didn’t need to be there”); and (6) felt burnt out toward end of cross-analysis (“I think I started to feel burnout towards the end of the process and then there was also one point I thought ‘are these getting longer or am I just getting tired of it,’ but they were getting longer so that got like a little frustrating”).

3.2.2.3 Focus Group Comparison

The qualitative raters were instructed to pay particular attention to pronounced similarities and/or differences that emerged between the two focus groups. This qualitative “compare and contrast” approach has a precedent in the psychological literature (e.g., Lysaker et al., 2015; Morrison et al., 2017).

3.2.2.3.1 Similarities

Although more nuanced similarities may have existed between the teams, the themes presented here were those discussed by multiple judges. Both teams discussed elements related to confusion, unfamiliarity, and challenge at the outset of the project that eventually gave way to clarity. Team 1 discussed feeling uncertain and overwhelmed initially, but eventually felt that project goals became clearer and more manageable over time, whereas Team 2 discussed how the unknown of the project was frustrating, but ultimately led to better overall understanding and performance. Despite the clarity that reportedly emerged for both teams over time, each team acknowledged challenges that presented themselves across multiple stages of the coding process. Team 1 explicitly described all steps of the process as having presented unique difficulties, and Team 2 stated each step of the process brought its own confusion and challenges. Moreover, each team described notable difficulty with cross analysis, although Team 2 was more adamant about their difficulty during this phase. The teams’ difficulty with cross analysis, which was the last phase of coding, may have been exacerbated by the project timeline, as judges needed to complete their coding before graduating in order to ensure they were able to engage in all aspects of the project. Both teams discussed the experience of feeling rushed toward the end of the project, which also may have contributed to their

experiences of burnout and instances in which they felt most susceptible to groupthink.

Both teams also commented on the positive nature of their coding meetings and experiences together. Team 1 emphasized this theme more than Team 2, but both teams discussed factors that contributed to this experience, including commitment to the project and engagement with one another. Additionally, both teams shared only one identical theme, noting that certain factors were conducive to the completion of the project, and certain factors would be helpful to implement in the future (likely as a result of the questions asked of them during the focus group). Both teams noted that the video trainings were helpful in their understanding of the coding process. Regarding the assigned readings, both teams stated the empirical articles were helpful, but acknowledged limitations to the utility of the CQR manual, as they found themselves confused at points. Relatedly, each team emphasized both the importance of more extensive training throughout the project and that a model, template, or example of other projects would have been useful.

3.2.2.3.2 Differences

Generally, fewer differences in themes emerged between teams. The most notable may have been Team 2's acknowledgement that one of its judges dominated discussion across all aspects of the coding process. Although members of Team 2 described their experience of a positive work environment, they also elaborated on individual differences between team members, and vocal and task-related dominance within their team. Relatedly, Team 1 emphasized the equal distribution and rotation of project responsibilities, which was recognized as a subtheme of what led to a positive and

comfortable dynamic, whereas Team 2 reported that an adjustment of their project roles could be helpful for improving projects in the future. Although both teams acknowledged various challenges that presented themselves throughout the process, Team 2 spent considerably more time during their focus group discussing areas of confusion and uncertainty than Team 1.

CHAPTER 4

DISCUSSION

The purpose of this exploratory study was to evaluate the similarities and differences that emerged between two uniformly-trained CQR coding teams of neophyte judges in terms of their (a) analytic yield based on the same stimulus set (i.e., replicability), and (b) coding process (i.e., social reliability). Regarding replicability, the two teams generated the most similar output at the level of core ideas, both in terms of frequency counts and content. Teams also yielded comparably similar domains, and appeared to cover the broad themes addressed in the study transcripts. The teams were most discrepant in their generated categories. Whereas some categories overlapped thematically, the teams diverged in the level of abstraction and the number of categories/subcategories/sub-subcategories that they generated. Relatedly, the teams were inconsistent in their reports of categorical representativeness of the sample; Team 1 saw the categories as more general and typical of participant experiences, whereas Team 2 saw the categories as more variant of participant experiences.

Regarding social reliability, Team 1, compared to Team 2, generally spent more time analyzing transcripts, had its team members more evenly distribute their time spent discussing a transcript, and exhibited fewer disagreements with its auditor. Additionally, Team 1 reported a better average group climate than Team 2 to a moderate-to-large degree. Although the teams started with similar ratings of group climate, Team 1 generally reported an improving climate over time, whereas Team 2 generally reported a worsening, though also more volatile, climate over time. However, during the focus groups, both teams addressed the positive and collaborative environment in which they

worked. Additional process similarities between the teams were experiences of confusion throughout the project and interest in additional training, whereas process differences between the teams were most pronounced in the degree of dominance vs. egalitarianism experienced by the judges.

The present findings underscore the epistemological frameworks within which CQR is situated; namely, the post-positivist and constructivist paradigms. For example, regarding the construction of core ideas, the CQR manual states, “The purpose of creating core ideas is to capture the content of the interview data in a succinct manner, staying grounded in the data and not interpreting participant intentions” and, “One of the advantages of having so many people review the core ideas is that if important content is missed by one person, another will catch it” (Hill, 2012, p. 113). Additionally, Hill has noted, “CQR also has a flavor of post-positivism because the emphasis on consensus among team members and auditors implies that team members are working to co-construct a ‘truth’” (p. 26). Thus, the core idea consensus process is most consistent with a post-positivist framework, recognizing that there is in fact a truth the interviewee is speaking, but for which the study judges will only imperfectly be able to apprehend with a core idea. It is not surprising, then, that the process of judges staying as “grounded” in participants’ language as possible yielded such similar output between the core ideas of the two teams, both in terms of frequency and of content. Post-positivism is also evident, in part, at the level of domain generation, which often relies heavily on the questions posed to study participants. Hill (2012) has noted, “The post-positivist component of the epistemology of CQR is evident in the use of a standard semi-structured interview protocol, with flexibility to query for further information where needed...” (p. 27). Thus,

this more structured framework may also contribute to the similarities between the teams' domain output.

However, as the CQR process began to require extrapolation of participant language and embrace more elements of the constructivist framework, divergences in output between the teams became more prominent. Arguably, the most important phase of CQR is the cross analysis of data into categories established by study judges, which pulls strongly from constructivism. Hill captured this best when stating, "From an epistemological standpoint, CQR is primarily constructivist in its recognition of the mutual influence between researcher and participant" (p. 26). It may be the case that the categories the CQR teams created and the cross analysis of each team were not replicable, in part, because of the idiographic lenses through which teams viewed and constructed categorical extrapolations. In fact, the greatest degree of similarity between teams at the category level occurred when dichotomous choices were presented to participants by the parent study interviewers (e.g., whether higher performing providers should or should not be reimbursed at higher rates). As the two teams analyzed participant responses to more open-ended questions, the teams' category yield was more divergent. These findings may have been even more pronounced if a more free-ranging or unstructured interview had been employed.

Such divergent categorization can be looked at differently depending on one's epistemological lens. For those leaning toward the positivist, quantitative tradition, such divergence may be disappointing in that it might fail to empirically *support* theory. However, for those leaning toward the interpretivist, qualitative tradition, the differences might be welcomed contributions that help *shape* theory (Carminati, 2018), which

provides a unique contribution to a field in its own right (Polit & Beck, 2010). Of course, CQR integrates both positions, which requires a dialectical balance. In the end, the epistemological leanings and lenses through which one views the differences in output between the two teams (at the level of categories) will shape the reader's interpretation of this finding.

As noted, there were also discrepancies in how representative the teams viewed the information discussed by the interviewees. These prominent differences appeared to be driven by Team 2's overwhelmingly variant findings, which poses a notable problem for CQR-based results. As Hill (2012) has stated, "It would be difficult to defend the trustworthiness of a study that had only variant categories in a CQR study because those categories are not reflective of even the sample used in that particular study" (p. 179). Given this questionable trustworthiness for Team 2, it seems important to speculate on *how* the team arrived at predominantly variant results.

As one possibility, although randomly assigned, this team may have consisted of judges who were less likely to, or who had difficulty with, summarizing material at higher levels of abstraction. Put differently, and as exemplified in the higher frequencies of domains and categories for Team 2 vs. Team 1, Team 2 judges may have tended to see the data as so individualized that they rarely developed an overarching connecting thread (hence, variant categorization). Alternatively, it is possible that Team 2 relative to Team 1, perhaps linked to more inherent difficulties with abstraction, was more affected by time constraints. Both teams were logistically bound to an academic year to complete their training and analyses, which is on the shorter end of the spectrum for data analyses as recommended by the CQR manual. Perhaps if Team 2 had more time to complete the

cross-analysis process, they may have re-established categories that yielded more representative, and theoretically saturated, findings. As yet another possibility, procedural differences between teams in the coding process (e.g., incorporation of auditor feedback, generation of different levels of categories), discussed further below, may also have accounted for differences in representativeness. Although determining the precise reasons behind overly variant categorization requires additional research, it would seem reasonable to suggest now that if a CQR team produces only variant categories in a given study, the researchers may need to challenge the team to find the connecting threads across some of the categories (greater abstraction), and give team members more time to do so.

Notably, the present study's findings concerning replicability somewhat parallel those of Ladany and colleagues (2012), the only prior study to compare CQR output between two teams of judges. In Ladany et al.'s study, the teams generated overlapping domains, yet diverged considerably in their categorization and cross-analyses. The authors did not report on the teams' core ideas, so direct comparisons on this output dimension cannot be made with the present study. However, for the comparisons that can be made, there was convergence between Ladany et al.'s teams of seasoned CQR judges and the present study's teams of neophyte judges. Thus, the aforementioned post-positivist and constructivist elements at the various stages of CQR appear to be consistent across the two replicability studies that have now been completed, regardless of the phenomena under study or the degree of experience of study judges. It will, of course, be important for future research to continue examining replicability between CQR teams in the service of improving our understanding of trustworthiness of the method's output and

the processes that may contribute to it.

With regard to such processes, the present study was the first to examine social reliability between two CQR coding teams. The output of the teams is obviously subjective, making it difficult to state whether one team “more accurately” completed the analyses or did a “better job,” especially if one subscribes to the constructivist model. However, it could be argued that Team 1’s output was more consistent with the CQR manual’s call for representativeness of findings and clarity of results. This finding is in some ways unsurprising, given the greater amount of time that Team 1’s judges spent analyzing each transcript, the more egalitarian process with which they approached coding, the fewer disagreements they exhibited with their auditor, and the more positive perception of their group climate. Team 1’s approach to analyses appears to be more consistent with the aims of CQR and adherent to the recommendations and spirit of the manual. Although no definitive conclusions can be made about causality, these variables may have contributed to Team 1’s arguably “superior” (and perhaps more meaningful) output.

Elaborating on these social reliability differences, Team 1’s processes squared more with CQR’s emphasis on careful immersion in the data. Specifically, Team 1 spent more time coming to consensus on core ideas than Team 2 on most transcripts, which suggests they may have more carefully and closely completed this analysis compared to Team 2. Relatedly, Team 1 judges more evenly distributed their time spent discussing their opinions than Team 2 judges, which, at least in spirit, is more consistent with the aims of the CQR process. Interestingly, both teams appeared to have been attuned to these processes. For example, during their focus group, Team 1 judges discussed their

quality communications, their valuing of others' input and perspectives, and their equal distribution and rotation of responsibilities across the project. Conversely, during their focus group, Team 2 acknowledged that more general rotations of roles among group members would be preferable in the future.

Remarkably, though, no Team 2 judges explicitly reported negative group climate in their focus group, despite the fact that they generally endorsed decreasing and volatile climate ratings on the CQR process measure. One study judge noted in the anonymous feedback form that they were bothered by the tardiness of fellow group members, but it is unclear which team this judge represented. However, the qualitative raters noted that Team 2 acknowledged during their focus group that vocal and task-related dominance was a concern throughout the CQR process. Again, this was born out by the objective group process indices, and this subtle perception may have been enough to shift perceived group climate. Dominance in this team may have kept more emotional disagreements from emerging throughout the coding process, which in some instances can actually promote group solidarity and shift coding toward greater inter-coder agreement (Sanders & Cuneo, 2010). Moreover, conflict can allow team members to fully air differences in viewpoints, resulting in higher quality team decisions (Ayoko, Callan, & Härtel, 2008). Critically, the dominance present in Team 2 may have contributed to more frequent *auditor* disagreements and poorer output. Although Team 2 judges explicitly stated that they worked well together, more cohesive groups have been found to render poorer quality decisions when additional antecedent conditions of groupthink, including directive leadership, are present (Mullen, Anthony, Salas, & Driskell, 1994). Hill (2012) has acknowledged that personality variables have not been

formally assessed in CQR research teams, but noted that researchers of team effectiveness have consistently found that emotional stability, extraversion, openness to new experience, agreeableness, conscientiousness, and, perhaps most notably in this instance, lack of dominance, are important attributes to consider when composing a team (Driskell, Goodwin, Salas, & O'Shea, 2006; Driskell & Salas, 1992). Hence, future research should address the role of personality in CQR team composition and output.

Additionally, future research might examine the degree to which momentary social transactions affects CQR output. For example, external raters might consider employing Benjamin's (1974) *Structural Analysis of Social Behavior* to more thoroughly analyze helpful and detrimental exchanges between study judges. Additionally, it may be helpful for future CQR teams to monitor their process in real time throughout the project. In the same way that routine outcome monitoring can be helpful in the psychotherapy setting (Boswell, Kraus, Miller, & Lambert, 2015), this approach may help shift maladaptive processes in the CQR research setting. For example, team members might use a quantitative measure like the one developed for the present study, to track process, and then qualitatively explore issues in the group, which again has a precedent in the psychotherapy literature (Hill, Chui, & Baumann, 2013). Of course, the impact of such monitoring on CQR process and outcome requires empirical examination.

Despite the teams' differences in process, similarities were also noted. Namely, both teams discussed commitment to task during their focus groups, and commented on how this facilitated dedication to the project and increased motivation, participation, and engagement. This finding is consistent with theoretical and empirical literatures that suggest that team goal commitment affects team performance, quality of group

experience, and viability (e.g., Aubé & Rousseau, 2005; Renn, 2003). Furthermore, this commitment to task may have mitigated some of Team 2's frustration with team dominance, especially given their report that their interest in the topic made the project more enjoyable.

Both teams also discussed the experience of confusion and uncertainty throughout the project. This may partly stem from the unique nature of these teams, which varied from more typical CQR teams. For example, study judges had no prior experience with CQR and were unfamiliar with the topic they were investigating. However, judges noted that they often felt more comfortable as they progressed throughout the study, and their lack of experience in some ways made them ideal candidates for learning and implementing this paradigm. Future CQR project leaders might mitigate some of these concerns by generating well-developed templates and exemplars to help provide additional context to study judges who are new to this paradigm, and to revisit training and piloting at each major phase of the CQR process.

Several limitations characterized this study. First, the standardized CQR training delivered via pre-recorded video instruction, though necessary for study design, may have been less effective for judges than an in-person training or workshop in which judges would be free to ask questions as they arose in the moment. Although judges were free to email questions after viewing the training videos, it is possible that a more hands-on training in which questions could be addressed organically may have facilitated greater familiarity with the CQR method.

Second, study judges never received feedback from the principal investigator/study trainer to determine if they were following CQR guidelines with

fidelity. Judges were left to consult other members of their team, the team auditor, the training materials, and/or the CQR manual whenever questions arose. As this was a novel experience for all judges, it is possible that they may have coded differently if they received ongoing feedback (as is allowable, and perhaps even preferred, in typical CQR administrations). Given that both teams reported some degree of confusion during their focus group, particularly regarding the cross-analysis phase, feedback from an experienced trainer in conjunction with the CQR manual may have mitigated rater drift.

Lastly, although the use of only two teams allowed for a rich comparison and contrast of output and process, the results of this study may have been different with the introduction of an additional CQR coding team. Given the differences that can emerge when relying on such small sample sizes, it is possible that the results, especially the qualitative findings, may have emerged differently with a third group. Findings may also have varied with the incorporation of a group using a different form of qualitative analysis. Consistent with the call of Williams and Morrow (2009), future research might compare the output of multiple CQR teams with a grounded theory analysis, thematic analysis, and/or other qualitative approaches.

Despite these limitations, this study highlights the consistency of the CQR method by paralleling Ladany and colleagues' (2012) process regarding replicability. Further, it extends the existing research by having examined social reliability indices as possible determinants of similarities and differences in CQR output. Although the epistemological lens through which one conducts qualitative research will inform the interpretations of this study's findings, it is clear that both similarities and differences in output are possible between distinct coding teams. Moreover, it may be possible to affect team output by

directly targeting social processes within a CQR team. It will be important for future research to examine specifically how shifting group climate (e.g., mitigating dominance, encouraging diversity in project roles) may lead to differences in output.

Table 1

Participant Demographic Characteristics and Presenting Problem Domains (N = 12)

Demographic		Presenting Problem ^a	% yes (n)
Age <i>M</i> (<i>SD</i>)	39.75 (9.75)	Anxiety	83.33 (10)
		Depression	83.33 (10)
Gender % (<i>n</i>)		Bipolar	41.67 (5)
Female	75.00 (9)	Substance Use	41.67 (5)
Male	25.00 (3)	Trauma	41.67 (5)
		Psychotic	8.30 (1)
Ethnicity % (<i>n</i>)		Other	16.67 (2)
White	58.33 (7)		
Hispanic	16.67 (2)		
African-American	16.67 (2)		
Native American	8.30 (1)		
Marital status % (<i>n</i>)			
Single	66.67 (8)		
Married/domestic partnership	16.67 (2)		
Separated/divorced	16.67 (2)		
Education % (<i>n</i>)			
Completed high school/GED	50.00 (6)		
Some college	16.67 (2)		
Associates degree	16.67 (2)		
Four-year degree	16.67 (2)		
Household income % (<i>n</i>)			
Less than \$25K	91.67 (11)		
\$25K - \$50K	8.30 (1)		
CMS % yes (<i>n</i>)	75.00 (9)		
Previous treatment % yes (<i>n</i>)	100.00 (12)		

Note. GED = general education development; CMS = Center for Medicare/Medicaid (participants).

^a Presenting problem domains do not sum to 12 because participants could endorse multiple domains.

Table 2

Total Number of Domains and Categories Generated by Teams

Team	Domain	No. of Categories	No. of Subcategories	No. of Sub-Subcategories
Team 1				
	1. Experience with mental health care services	4	13	12
	2. Experience selecting mental health services	4	13	11
	3. Attitudes about the selection process	6	20	5
	4. Attitudes on preferred providers	6	17	10
	Total	20	63	38
Team 2				
	1. Past and current experiences with mental health services	4	11	0
	2. History with mental health care provider selection process	6	24	0
	3. Important factors selecting a mental health provider	6	32	0
	4. Availability and validity of provider's background information	6	21	0
	5. Possible improvements to mental health care selection process	8	34	0
	6. Experience with surveys/questionnaires	4	16	0
	7. Opinions about well-matched provider	6	14	0
	Total	40	152	0

Note. The total number of categories, subcategories, and sub-subcategories in this table is greater than the number of categories, subcategories, and sub-subcategories in subsequent tables, as some of these groups were omitted from the final presentation of results per CQR protocol when only 1 parent study participant was represented (i.e., a “rare” representation).

Table 3

Team 1 Domain 1: Experience with Mental Health Care Services

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Mental health care providers worked with		17	11 (G)
	1.1 Doctorate-level providers (psychiatrists, psychologists, PCP)	12	11 (G)
	A. Other doctorate-level providers (psychiatrists, PCP)	10	9 (T)
	B. Doctorate-level psychologists (PhD and PsyD)	6	6 (V)
	1.2 Masters-level & other providers (LMHC, LCSW, other counselors)	10	9 (T)
	1.3 Nurses/nurse practitioners, physicians assistants	5	3 (V)
2. Questionnaires & Surveys		23	11 (G)
	2.1 Completion of provider satisfaction surveys	11	11 (G)
	A. Completed sometimes/rarely or unsure	7	7 (T)
	B. Completed never	4	4 (V)
	2.2 Completion of personal difficulty questionnaires	11	11 (G)
	A. Completed never	4	4 (V)
	B. Completed sometimes/rarely or unsure	4	4 (V)
	C. Completed frequently	3	3 (V)
2.3 Patient's understanding of how survey and questionnaire data is used	2	2 (V)	
3. Types of mental health treatment received		13	11 (G)
	3.1 Individual counseling	10	10 (T)
	3.2 Medication management	8	8 (T)
	3.3 Hospitalizations	3	3 (V)
4. Experiences with mental health services		17	9 (T)
	4.1 Negative experiences	12	8 (T)
	A. Clinic issues	3	3 (V)
	B. Insurance	3	3 (V)
	C. Provider affect/match to patient	3	3 (V)
	D. Turnover	3	3 (V)
	4.2 Positive experiences	3	3 (V)
4.3 Non-valenced experiences	2	2 (V)	

Note. G = General; T = Typical; V = Variant; PCP = Primary Care Physician; LMHC = Licensed Mental Health Counselor; LCSW = Licensed Clinical Social Worker.

Table 4

Team 1 Domain 2: Experience Selecting Mental Health Services

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Patient's knowledge about a provider prior to the first appointment		13	12 (G)
	1.1 Some prior knowledge	9	8 (T)
	1.2 No prior knowledge	5	5 (V)
2. Difficulty in finding mental health services		25	11 (G)
	2.1 Level of difficulty	11	9 (T)
	A. Easy to find services	4	4 (V)
	B. Mixed experiences finding services	3	3 (V)
	C. Difficult to find services	3	3 (V)
	2.2 Factors causing difficulty	15	9 (T)
	A. Insurance	5	4 (V)
	B. Availability	5	4 (V)
	C. Administrative	3	3 (V)
	D. Accessibility	4	3 (V)
E. Lack of information about the providers	4	3 (V)	
2.3 Factors causing ease of selection	5	3 (V)	
A. Insurance	2	2 (V)	
B. Administrative	4	2 (V)	
3. Methods of finding mental health care providers		17	11 (G)
	3.1 Clinics/hospitals/agencies	8	7 (T)
	3.2 Personal connections	8	7 (T)
	3.3 Providers/professionals	6	6 (V)
	3.4 Online search	5	4 (V)
4. Patient's say in the mental health care selection process		11	11 (G)
	4.1 Total say	7	7 (T)
	4.2 No say	4	4 (V)
	4.3 Some say	2	2 (V)

Note. G = General; T = Typical; V = Variant.

Table 5

Team 1 Domain 3: Attitudes About the Selection Process

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Opinion on access to PPI		21	12 (G)
	1.1 Patient would like access	18	10 (T)
	A. Access through website/online system	10	8 (T)
	B. Access through a handout	4	4 (V)
	1.2 Patient would not like access	3	3 (V)
2. Improvements to provider selection process		25	12 (G)
	2.1 More provider info/patient input/unified info system	16	10 (T)
	2.2 Other	7	6 (V)
	2.3 Better matching/referral system	5	5 (V)
	2.4 Insurance-related issues/information	2	2 (V)
	2.5 More funding/more awareness for mental health care system (i.e., advertisements)	3	2 (V)
3. Potential downsides to using PPI		18	12 (G)
	3.1 Patient bias	12	8 (T)
	3.2 No downsides	4	4 (V)
	3.3 Confidentiality	2	2 (V)
4. How survey/questionnaire data should be used		13	10 (T)
	4.1 Treatment/service improvement and research	10	8 (T)
	A. Tailoring treatment	4	4 (V)
	B. Making a diagnosis	3	3 (V)
	4.2 Improvement of selection/matching process	3	3 (V)
5. How survey/questionnaire data is being used		9	8 (T)
	5.1 Treatment/service improvement and research	7	7 (T)
	5.2 Unsure	2	2 (V)
6. Benefits of accessing provider PPI		10	8 (T)
	6.1 Help make provider comparisons for selection	6	6 (V)
	6.2 Receiving additional info on providers	4	4 (V)

Note. G = General; T = Typical; V = Variant; PPI = Provider Performance Information.

Table 6

Team 1 Domain 4: Attitudes on Preferred Providers

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Selecting a well-matched provider		26	12 (G)
	1.1 List vs. one provider	14	12 (G)
	A. Patient would prefer a list of providers in patient's area	13	11 (G)
	1.2 Patient choosing vs. provider choosing for patient	14	12 (G)
	A. Patient would prefer to choose themselves	9	9 (T)
	B. Patient would prefer a provider choose for patient	5	4 (V)
2. Important factors when selecting a provider		24	12 (G)
	2.1 Provider personal qualities (i.e., empathetic, respectful)	13	9 (T)
	A. This is the most important factor for patient	6	6 (V)
	2.2 Provider practice qualities (i.e., knowledgeable, experienced, therapeutic style/method/focus)	16	9 (T)
	A. This is the most important factor for patient	7	6 (V)
	2.3 Logistics	8	5 (V)
	A. This is the most important factor for patient	6	4 (V)
3. Opinion on paying higher performing providers more		11	11 (G)
	3.1 No	9	9 (T)
	A. System-level bias	6	6 (V)
	B. Concerned about affecting client payment/rates	3	3 (V)
	3.2 Yes	4	4 (V)
4. Patient's willingness to wait to see matched/ preferred			

providers		11	11 (G)
	4.1 Yes	5	5 (V)
	4.2 Depends	5	5 (V)
5. Patient's willingness to pay more to see a matched/preferred provider		11	11 (G)
	5.1 No	5	5 (V)
	5.2 Depends	3	3 (V)
	5.3 Yes	3	3 (V)
6. Perceived importance of seeing well-matched provider		10	9 (T)
	6.1 Less important than factors in Category 1	4	4 (V)
	6.2 More important than factors in Category 1	4	4 (V)

Note. G = General; T = Typical; V = Variant.

Table 7

Team 2 Domain 1: Past and Current Experiences with Mental Health Services

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Kinds of services provided		7	7 (T)
	1.1 Currently uses medication	4	3 (V)
	1.2 Currently receives therapy	2	2 (V)
2. Types of providers worked with		3	2 (V)
	2.1 Has worked with multiple types of providers	12	9 (T)
	2.2 Has worked with only one type of provider (psychologist, social worker, etc.)	6	5 (V)
3. Satisfaction with provider		-	-
	3.1 Change in insurance coverage	3	2 (V)
	3.2 Satisfied with provider	2	2 (V)
4. Instances of hospitalization		-	-
	4.1 Multiple instances	2	2 (V)

Note. T = Typical; V = Variant.

Table 8

Team 2 Domain 2: History with Mental Health Care Provider Selection Process

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Amount of control during selection process		2	2 (V)
	1.1 Had complete control during selection process	4	4 (V)
	1.2 No say during selection process	6	3 (V)
2. Method of referral to provider		-	-
	2.1 Referred through friend/family member	7	7 (T)
	2.2 Referred by another provider	9	5 (V)
	2.3 Found provider through own research	4	3 (V)
	2.4 Referred by hospital	3	3 (V)
3. Difficulty of provider selection process		-	-
	3.1 Found selection process difficult	5	5 (V)
	3.2 Found selection process easy	5	4 (V)
4. Difficulties during selection process		-	-
	4.1 Insurance coverage issues	7	4 (V)
	4.2 High turnover rate among providers	4	4 (V)
	4.3 Long wait time to see a provider	5	3 (V)
	4.4 Poor transportation to reach provider	2	2 (V)
5. Knowledge of provider before first appointment		-	-
	5.1 Some knowledge of provider before first appointment	4	3 (V)
	5.2 Minimal knowledge of provider before first appointment	3	3 (V)
	5.3 No knowledge before first appointment	3	3 (V)
6. Pros of provider selection process		-	-
	6.1 Helpfulness of referrals	4	3 (V)
	6.2 Insurance referrals	2	2 (V)

Note. T = Typical; V = Variant.

Table 9

Team 2 Domain 3: Important Factors Selecting a Mental Health Provider

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Factors relating to provider's personality		2	2 (V)
	1.1 Provider sensitivity	4	3 (V)
	1.2 Provider open-mindedness	2	2 (V)
	1.3 Provider gender	2	2 (V)
2. Factors relating to provider's ability		-	-
	2.1 Provider's experience	5	3 (V)
	2.2 Provider treatment approach	3	3 (V)
	2.3 Provider specialization	2	2 (V)
	2.4 Provider's education	2	2 (V)
3. Logistical provider selection factors		-	-
	3.1 Importance of provider proximity to patient	3	3 (V)
4. Rank of provider selection factors		-	-
	4.1 All factors as equally important	3	3 (V)
5. Preference of certain factors over matched provider		-	-
	5.1 Prefers other factors over being match with provider	2	2 (V)
	5.2 Prefers being matched with provider over other factors	2	2 (V)
6. Most important factor when selecting provider		-	-
	6.1 Provider's personality	2	2 (V)

Note. V = Variant.

Table 10

Team 2 Domain 4: Availability and Validity of Provider's Background Information

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Opinion on having PPI		-	-
	1.1 Would prefer to have PPI	10	9 (T)
2. Downfall to have PPI		-	-
	2.1 Reviews from patients are subjective	8	6 (V)
	2.2 No downfall to have this information	4	4 (V)
	2.3 Treatment also is dependent on patient cooperation	2	2 (V)
3. Preferred way to access PPI		-	-
	3.1 Would prefer information online	6	6 (V)
	3.2 Would prefer information via handout	3	3 (V)
4. Opinion on higher reimbursement for higher performing providers		-	-
	4.1 Does not think higher performing providers should be paid more	4	4 (V)
	4.2 Does think higher performing providers should be paid more assuming insurance covers costs	3	3 (V)
	4.3 Unsure if higher performing providers should be paid more	2	2 (V)
5. Pros to have PPI		-	-
	5.1 Can help patient to make a more informed decision	3	3 (V)

Note. T = Typical; V = Variant; PPI = Provider Performance Information.

Table 11

Team 2 Domain 5: Possible Improvements to Mental Health Care Selection Process

Category	Subcategory	No. of Core Ideas	No. of Cases
1. Process of locating a provider		-	-
	1.1 Prefers list of multiple provider recommendations	11	8 (T)
	1.2 Prefers list of multiple provider recommendations and one single provider recommendation	2	2 (V)
2. How to improve the selection process		-	-
	2.1 Say in the selection process	4	3 (V)
	2.2 Access to provider track record	3	3 (V)
	2.2 Streamlined intake process	2	2 (V)
3. Method of choosing mental health provider		-	-
	3.1 Prefers to make own personal provider choice	3	3 (V)
	3.2 Prefers doctor or other professionals to make choice	2	2 (V)
	3.3 Is okay with making own provider selection or having a professional make selection	2	2 (V)
4. Suggestions to providers		-	-
	4.1 Use surveys to determine patients needs	3	3 (V)
5. Things patient would like to know/want before making a selection		-	-
	5.1 Website with specialization of provider	3	2 (V)
	5.2 The specialization of provider	2	2 (V)
6. Use of provider performance surveys		-	-
	6.1 Used to better understand patients and improve services	2	2 (V)

	6.2 Used to better match patients and providers together	2	2 (V)
7. Improvements to availability of provider information		-	-
	7.1 Wants provider information available online	2	2 (V)

Note. T = Typical; V = Variant.

Table 12

Team 2 Domain 6: Experience with Surveys/Questionnaires

Category	Subcategory	No. of Core Ideas	No. of Cases
1. How often patient fills out surveys		3	3 (V)
	1.1 Often	3	3 (V)
2. Previous experience with surveys		2	2 (V)
	2.1 Has completed personal wellbeing surveys	8	8 (T)
	2.2 Has completed provider satisfaction surveys	5	5 (V)
	2.3 Has not completed provider satisfaction surveys	3	3 (V)
	2.4 Has not completed personal wellbeing surveys	2	2 (V)
3. How patient thinks surveys are used		-	-
	3.1 Helps providers evaluate themselves	3	3 (V)
	3.2 Unsure of how surveys are used	3	3 (V)
	3.3 Allows provider to inform treatment decisions	2	2 (V)
4. Suggested use of surveys		-	-
	4.1 Highest quality of services	2	2 (V)

Note. T = Typical; V = Variant.

Table 13

Team 2 Domain 7: Opinions about Well-Matched Provider

Category	Subcategory	No. of Core Ideas	No. of Cases
1. How would patient prefer to find a matched provider?		3	3 (V)
	1.1 Would be comfortable selecting a provider from a list of matched providers	4	4 (V)
2. The importance of having a well matched provider		2	2 (V)
	2.1 Not as important as other factors	2	2 (V)
3. Willingness to travel far for matched provider		2	2 (V)
4. Willingness to wait to see matched provider		-	-
	4.1 Willing to wait	6	5 (V)
	4.2 Willing to wait if mentally stable at the time	3	3 (V)
5. Willingness to pay more for matched provider		-	-
	5.1 Is not willing to pay more	4	3 (V)
	5.2 Is willing to pay more	3	3 (V)
	5.3 If the copay is reasonable then client would pay more for a matched provider	2	2 (V)

Note. V = Variant.

Table 14

Team 1 Exemplars for Domain 1: Experience with Mental Health Care Services

Category	Subcategory
<p>1. P has worked with LCSW, LMHC, unlicensed mental health counselors, psychologists, and psychiatrists for medication management.</p>	<p>1.1 P is currently receiving outpatient therapy from a PsyD and medication management from a psychiatrist. A. P had received counseling from PhDs, licensed mental health care workers, and social workers. P has received medication management from a psychiatrist and P's PCP. B. P is currently receiving outpatient therapy from a PsyD and medication management from a psychiatrist. 1.2 P has worked with LCSW, LMHC, unlicensed mental health counselors, psychologists, and psychiatrists for medication management. 1.3 P is currently receiving counseling and sees a nurse practitioner for medication.</p>
<p>2. P believes that P has completed provider satisfaction surveys.</p>	<p>2.1 P has been asked to complete provider satisfaction surveys—and would be willing to complete them—but has not done so yet. A. P has been asked to complete a questionnaire about provider satisfaction once in a while with different insurance companies. B. P has never been asked to complete a survey about provider satisfaction. 2.2 P has completed questionnaires about difficulties such as symptoms, functioning, or quality of life. A. P has never been asked to complete a questionnaire asking about difficulties such as symptoms, functioning, and quality of life, but would be interested in doing so. B. P has been asked to complete questionnaires about difficulties such as symptoms, functioning, and quality of life. C. P has almost always been asked to complete a questionnaire about symptoms, functioning, and quality of life. 2.3 P's therapist has told P that the questionnaire information is used by the therapist to improve P's treatment. P also thinks that the information is reported to the Department of Mental Health, and thinks that it is used to monitor trends in mental health care and to gather information on demographics.</p>

3. P currently sees a talk therapist once a week and an RN with a specific concentration in psychiatric medicine once a month for medication management.

3.1 P had received counseling from PhDs, licensed mental health care workers, and social workers. P has received medication management from a psychiatrist and P's PCP.

3.2 P has received counseling and medication management both currently and in the past.

3.3 P has been hospitalized seven times, both pediatric and adult.

4. P stopped therapy in the past due to disappointment about turn over rates. When P's psychiatrist recommended P start therapy again, P voiced concerns about building trust with a provider only for them to leave.

4.1 P stopped therapy in the past due to disappointment about turn over rates. When P's psychiatrist recommended P start therapy again, P voiced concerns about building trust with a provider only for them to leave.

A. About a year to a year-and-a-half ago, P went to another clinic for pain management, but found the clinic very unprofessional.

B. P is not currently receiving mental health services because P's insurance recently changed and is trying to find a provider that accepts it.

C. When selecting a provider, empathy is the most important factor for P because P has been discouraged by providers who do not display empathy.

D. P stopped therapy in the past due to disappointment about turn over rates. When P's psychiatrist recommended P start therapy again, P voiced concerns about building trust with a provider only for them to leave.

4.2 P feels that P's current counselor is much better than P's previous counselor, despite having decades less experience.

4.3 P has had to switch providers involuntarily because P moved or changes in insurance due to income. P has also voluntarily switched providers.

Note. P = Patient; PCP = Primary Care Physician; LMHC = Licensed Mental Health Counselor; LCSW = Licensed Clinical Social Worker; RN = Registered Nurse.

Table 15

Team 1 Exemplars for Domain 2: Experience Selecting Mental Health Services

Category	Subcategory
<p>1. P did not know anything about the providers prior to the first appointment, but would have liked to.</p>	<p>1.1 Before the first appointment, P usually knows where the provider went to school, how long the provider has been in practice, and if the provider is nearby.</p> <p>1.2 P did not know anything about the providers before the first appointment.</p>
<p>2. P has had difficulty finding a mental health care provider, especially when there is a lack of information about them. This is the case with P's current mental health care agency.</p>	<p>2.1 P found it easy to find a provider because of the social worker's referral, but P found it difficult to find a center that accepted P's insurance.</p> <p>A. It has been easy for P to find a provider.</p> <p>B. P found it easy to find a provider because of the social worker's referral, but P found it difficult to find a center that accepted P's insurance.</p> <p>C. P has found it difficult to find a provider because there is no way to tell a provider's style from reading about them. P feels that finding a provider has been trial-and-error.</p> <p>2.2 P has had difficulty finding a mental health care provider, especially when there is a lack of information about them. This is the case with P's current mental health care agency.</p> <p>A. P found it easy to find a provider because of the social worker's referral, but P found it difficult to find a center that accepted P's insurance.</p> <p>B. P feels like P would take almost any therapist after getting past the logistical concerns of transportation and finding a provider that is taking new patients.</p> <p>C. After having their insurance change because of reaching Medicare age, P encountered many insurance and administrative complications, including having to pay out of pocket for psychiatrists, having difficulty finding availability, and being abandoned by a satellite clinic that neither P nor P's primary care doctor could reach after they had moved without informing P. P feels that they almost had a mental health crisis during this because of almost running out of medication before finding another provider. P reached a point where they needed a provider and the age and gender of the provider did not matter.</p> <p>D. P has had difficulty finding a mental health care provider primarily due to outpatient clinics being inaccessible by public transit, P's primary form of transportation.</p>

E. P has found it difficult to find a provider because there is no way to tell a provider's style from reading about them. P feels that finding a provider has been trial-and-error.

2.3 P feels that P had a say in the selection process in that the intake worker understood P and matched P with a therapist P is extremely happy with. P has never felt trapped with a therapist or as though P did not have a choice.

A. P feels that their insurance and an additional Medicare supplement plan provided P with choice and flexibility when choosing a provider.

B. P feels that P had a say in the selection process in that the intake worker understood P and matched P with a therapist P is extremely happy with. P has never felt trapped with a therapist or as though P did not have a choice.

3. P relied on P's psychiatrist to find a therapist within the clinic because P had no data to make an informed decision.

3.1 P was assigned a provider when P was in the hospital. In the past, P's psychologist recommended a psychiatrist.

3.2 P has used online searches and referrals from friends to find providers. P has also looked up providers in the phonebook. When P was a minor, P's parents found the providers.

3.3 P relied on P's psychiatrist to find a therapist within the clinic because P had no data to make an informed decision.

3.4 P gives providers a try after looking up bios/information online and talking to people. P acknowledges it is hard to tell ahead of time and has had to try a few before settling with one counselor or medication management option.

4. P feels that P had a say in the selection process in that the intake worker understood P and matched P with a therapist P is extremely happy with. P has never felt trapped with a therapist or as though P did not have a choice.

4.1 P feels that P had a say in the selection process in that the intake worker understood P and matched P with a therapist P is extremely happy with. P has never felt trapped with a therapist or as though P did not have a choice.

4.2 P feels that P did not have a say in the selection process.

4.3 P feels that P had a say in the selection of providers most of the time.

Note. P = Patient.

Table 16

Team 1 Exemplars for Domain 3: Attitudes About the Selection Process

Category	Subcategory
<p>1. P thinks it would be helpful to view provider performance information and reviews on a website that can reliably verify actual patient reviews.</p>	<p>1.1 P would like to access provider performance information because P would like to know that provider has been successful in treating other people.</p> <p>A. P would prefer to access provider performance information online because P is on the computer a lot.</p> <p>B. P would want to receive provider performance information from a flyer or handout rather than online or by referral.</p> <p>1.2 P would not necessarily want access to provider performance information because that information is difficult to generalize because there are such strong individual factors.</p>
	<p>2.1 P feels that it would be helpful to have a list of providers available at a particular clinic, including information on their current and past specialties and the issues with which they have had experience. Particularly, P would like to know if the providers are sensitive to GLBT issues.</p> <p>2.2 P thinks that a larger provider pool in general would help improve the selection process. Additionally, a more streamlined intake process (potentially online) would benefit patients who rely on transit to get to the clinic. Ideally, P would like better transportation to the providers' locations.</p> <p>2.3 P feels that the provider selection process could be improved by matching patients with providers of similar personalities to ensure that the patient is comfortable and get along with the provider. P also feels that matching patients and providers based on similar values and beliefs would be helpful because P feels that a large part of a patient is ignored when providers do not consider the patient's religion. P also would prefer if providers are less forceful when prescribing medications to patients who do not wish to take medications for personal reasons as well as health concerns.</p> <p>2.4 P does not think that provider should be able to choose what insurances they accept, because it makes it harder for people to find the services they need.</p> <p>2.5 Based on personal experience, P feels that more advertised awareness for mental health services would be useful in reducing stigma and helping at-risk populations find treatment providers.</p>
<p>2. P feels that the provider selection process could be improved by access to more information about the provider, maybe an informal conversation with a potential therapist, and more patient input.</p>	

3. P thinks that a potential downfall of using provider track records is that information can be biased by reviewers who had a bad experience with the provider or by reviewers who have a personal connection to the provider.

4. P feels that the personal wellbeing questionnaires should be used to tailor treatment, gain insight into how P was feeling, or make a diagnosis.

5. P thinks that questionnaires are used by facilities to track how patients are doing and to ensure that therapists are providing sufficient care.

6. P would like access to provider performance information because P can determine if the provider has a good track record in treating patients with similar needs to P.

3.1 P thinks that a potential downfall of using provider track records is that information can be biased by reviewers who had a bad experience with the provider or by reviewers who have a personal connection to the provider.

3.2 P cannot think of any negatives to using provider track records to choose providers.

3.3 P feels that confidentiality of clients and providers is a potential concern of using provider track record information.

4.1 P thinks that survey data should be used to determine if providers are fulfilling patients' needs; this is personally helpful for P in that providers can better their care or change treatment strategies. P also feels that survey data can inform a broad spectrum of research.

A. P feels that the personal wellbeing questionnaires should be used to tailor treatment, gain insight into how P was feeling, or make a diagnosis.

B. P feels that the personal wellbeing questionnaires should be used to tailor treatment, gain insight into how P was feeling, or make a diagnosis.

4.2 P thinks that survey data about client satisfaction should be used to ensure a good match between a provider and a client.

5.1 P feels that the information from questionnaires are helpful to clinics and the Department of Mental Health so that areas of need within the population can be identified and additional trainings can be given to providers on these issues.

5.2 P does not know how provider satisfaction surveys are used, and does not have an opinion on how they should be used.

6.1 P would want access to provider performance data and thinks it would be useful for making comparisons, or for helping providers keep their services on par with the services of other providers.

6.2 P thinks that provider performance information could be useful to see what a provider specializes in, and if the provider is sensitive to the needs of certain communities (GLBT). P thinks it is necessary to supplement raw data with patient feedback, as outcomes are not the whole picture.

Note. P = Patient; GLBT = Gay/Lesbian/Bisexual/Transgender.

Table 17

Team 1 Exemplars for Domain 4: Attitudes on Preferred Providers

Category	Subcategory
<p>1. P would prefer to have direct access to a ranked list of well-matched providers rather than just one provider, because P would like more options.</p>	<p>1.1 P would prefer a list of providers in P’s area to choose from over one particular provider who seemed to be the most well-matched. A. P would prefer a list of providers in the area over one specific recommendation so that P can make a decision. 1.2 Given a list of matched providers, P would rather make a provider selection based on that list than have an intake worker make the decision. A. If P were given a list of well-matched providers, P would feel more comfortable making the selection themselves rather than their primary care doctor or another provider making their selection for them because P is not very comfortable with their primary care provider. B. If P were to access provider performance information, P would like to receive the information through another provider, as P would be able to know the referring provider’s opinion.</p>
<p>2. P thinks that empathy, compassion, and knowledge of P’s issues are important factors when selecting a provider.</p>	<p>2.1 P thinks that empathy, compassion, and knowledge of P’s issues are important factors when selecting a provider. A. For P, the most important factor in the provider matching process is that the provider is a good listener. The location is a less important factor for P. 2.2 P thinks that empathy, compassion, and knowledge of P’s issues are important factors when selecting a provider. A. P wants the sharpest and most knowledgeable provider in the field, even if their personality is problematic to some clients, because P wants a provider that is able to figure out how to deal with P’s issues. 2.3 P feels the most important factor is the logistical issue of accessing a therapist. The second most important concern is that the therapist is respectful, experienced working with P’s demographic, and are capable of treating P’s mental health issues. A. P feels the most important factor is the logistical issue of accessing a therapist. The second most important concern is that the therapist is respectful, experienced working with P’s demographic, and are capable of treating P’s mental health issues.</p>

3. P does not think that higher performing providers should be reimbursed at a higher rate because the data is very subjective. It would be hard to pay based on merit and experience, especially in mental health care.

4. P would be willing to wait to see a preferred or matched provider.

5. P would be willing to pay more to see a preferred or matched provider, as long as it was under \$20 per co-pay (in comparison to a \$0 co-pay currently).

6. P feels that a matched provider would intrinsically include P's important factors, if they were truly matched with P. P would be willing to wait for a matched provider if it was a long term treatment plan.

3.1 P does not think that higher performing providers should be reimbursed at a higher rate because the data is very subjective. It would be hard to pay based on merit and experience, especially in mental health care.

A. P does not think that higher performing providers should be reimbursed at a higher rate because P does not have much money and because P thinks that paying higher performing providers more could lead to biases against certain types of providers, such as biblical counselors. P would possibly be okay with higher performing providers being reimbursed more if the higher payment was coming from insurance companies.

B. P would be fine with the insurance company reimbursing higher performing providers at a higher rate, but does not think that providers should be paid more if it would cost P more.

3.2 P would be fine with the insurance company reimbursing higher performing providers at a higher rate, but does not think that providers should be paid more if it would cost P more.

4.1 P would be willing to wait to see a preferred or matched provider.

4.2 P's willingness to wait to see a preferred provider would depend on how time sensitive P's current issues are.

5.1 P would probably not be willing to pay more to see a preferred provider.

5.2 P would be willing to pay more to see a preferred or matched provider, as long as it was under \$20 per co-pay (in comparison to a \$0 co-pay currently).

5.3 P would be willing to pay more to see a matched provider.

6.1 For P, it would probably be less important to see a matched provider than seeing a provider with good interpersonal skills or clinical flexibility. P would not completely believe track record data.

6.2 P is willing to travel a little bit further to see a matched provider if P is going to get better treatment.

Note. P = Patient.

Table 18

Team 2 Exemplars for Domain 1: Past and Current Experiences with Mental Health Services

Category	Subcategory
1. P has received and currently receives counseling and medication management.	<p>1.1 P currently sees a psychiatrist and receives mood stabilizing treatment once every five weeks.</p> <p>1.2 P currently receives individual behavioral therapy once every 2 weeks.</p>
2. P has worked with therapists in the past but is not sure whether they were psychologists or social workers.	<p>2.1 P has previously received therapy from doctors, social workers, doctoral psychologists and licensed mental health care workers. P received medication and counseling from a psychiatrist.</p> <p>2.2 P is in therapy with one provider for pain management relief.</p>
3. N/A	<p>3.1 P is not receiving any mental health services because P's insurance recently changed. P is looking for a new provider who will take P's current insurance.</p> <p>3.2 P had a social worker who P really liked that sent P to a facility that was easy to make appointments and easy to make payments.</p>
4. P has received mental health services since P was 8. P also has been hospitalized several times and has attended individual and group therapy sessions, mostly for women.	4.1 P has been hospitalized seven times in pediatric and adult units.

Note. P = Patient.

Table 19

Team 2 Exemplars for Domain 2: History with Mental Health Care Provider Selection Process

Category	Subcategory
<p>1. P felt as though they didn't have much control over their first provider but had more of a say in selection of their second provider.</p>	<p>1.1 P feels that P has had a say in the provider selection process. 1.2 P believes P did not have a say in the mental health care provider selection process.</p>
<p>2. N/A</p>	<p>2.1 P found out about the agency P receives treatment at because of a family member who received treatment at same clinic. 2.2 A psychologist recommended a psychiatrist for P once. 2.3 P will look at providers' bios and websites and give them a try if they accept P's insurance. P will also look at insurance company's list. 2.4 P found provider by referral at the hospital and P's cousin who is in the mental health field.</p>
<p>3. P has had mixed experiences finding mental healthcare providers; sometimes it works, sometimes it doesn't.</p>	<p>3.1 P found it difficult to find a mental health provider. 3.2 P has found it easy to find a mental healthcare provider.</p>
<p>4. P believes one of the most difficult barriers to finding a provider is logistics. Once P gets past the logistic concerns, P will just take anybody.</p>	<p>4.1 P has found the provider selection process difficult; P went without therapy for 4 years because it was hard to find somebody due to insurance issues. 4.2 P has had to go through the provider selection process several times because P's first two providers left for a better job. P feels like they had a say in the recent selection of providers. However, P feels that P had no say in selection process as a minor. 4.3 In P's opinion, it is typical for a patient to wait about four to six weeks to see a provider in P's area. 4.4 P thinks it's difficult to find providers because the transportation to providers is not convenient for P. Therapists in</p>

that area usually have a high load of patients so usually it takes time to see a therapist.

5. P has never actually seen a list of well matched providers.

5.1 P usually has information on provider's education, experience and location before P's first appointment.

5.2 P has previously known nothing about providers before the first appointment besides a few details from the intake process (what the provider specializes in and who they typically treat).

5.3 P didn't know anything about providers before first appointment.

6. N/A

6.1 P felt it was helpful to have someone to help P find a provider.

6.2 P thinks your insurance can find someone for you if need be.

Note. P = Patient.

Table 20

Team 2 Exemplars for Domain 3: Important Factors Selecting a Mental Health Provider

Category	Subcategory
1. P thinks providers' education, work experiences and personality are important factors to select providers.	1.1 After looking at logistical issues (proximity and convenience), P would prefer a provider that is respectful of P's demographic, has worked well with people of P's demographic, and has experience working with issues related to P (anxiety and mood disorders).
	1.2 It is important for P to have a female provider as well as someone who is interactive and open minded when choosing treatment providers.
	1.3 It is important for P to have a female provider as well as someone who is interactive and open minded when choosing treatment providers.
2. N/A	2.1 P also thinks a provider's graduate program or internship is important for future success because of the various skills, information, and experiences learned during program.
	2.2 P thinks knowing the provider's personality and their main emphasis on treating clients, including the provider's view on medication, would be helpful in making a provider selection.
	2.3 P considers the area of expertise as an important factor in choosing a provider.
	2.4 P thinks providers' education, work experiences and personality are important factors to select providers.
3. N/A	3.1 P thinks location is also an important factor for selecting a provider.
4. N/A	4.1 P considers professionalism, experience, transparency, and comfort level with provider as all equally important.
5. N/A	5.1 P believes comfortableness, knowledge within field, and openness and being non-judgmental are more important factors than choosing a provider that is considered matched or preferred.
	5.2 P thinks a matched provider is more important than the gender of provider, therapeutic approach and their religious views.
6. N/A	6.1 P ranks a provider who is friendly as the most important factor in selecting a treatment provider and P also considers education, knowledge, and not dismissing P as also important factors.

Note. P = Patient.

Table 21

Team 2 Exemplars for Domain 4: Availability and Validity of Provider’s Background Information

Category	Subcategory
1. N/A	1.1 P thinks it would be helpful to have provider performance information available (i.e. success rates).
2. P feels that it may be difficult to have all information accessible for all providers because different information regarding insurance policy is different for different providers.	2.1 P believes provider track record can be biased by more negative experiences as well as the legitimacy of people writing the reviews. 2.2 P can think of no potential downfall from using provider performance information. 2.3 P would like to receive provider performance information, but acknowledges that it may not be accurate. A patient’s mental health is not just dependent on the provider, but also on the patient’s willingness to cooperate.
3. N/A	3.1 P thinks if the information is online will be convenient for a lot of people. 3.2 P would like provider performance information given in a flyer or hand out.
4. P thinks all providers should be performing on a higher level and should not be reimbursed more. If higher performing providers could be identified, P thinks they should be paid more.	4.1 P does not think that a high performing provider should be paid more. 4.2 P thinks higher performing providers could get paid more for their work as long as it is within the range that insurance companies would be willing to pay. 4.3 P is not sure if higher performing providers should be reimbursed more because P is unsure if the criteria providers are judged upon is objective or not. P thinks creating these divisions will continue to decrease the amount of mental health care providers in the system.
5. N/A	5.1 P would like to receive provider performance information because it helps P make a more informed decision choosing a provider who has success with P’s specific needs.

Note. P = Patient.

Table 22

Team 2 Exemplars for Domain 5: Possible Improvements to Mental Health Care Selection Process

Category	Subcategory
1. N/A	<p>1.1 P would want a list of providers to choose from rather than just one provider.</p> <p>1.2 P thinks it will be helpful to have a list of well matched providers as well as a recommendation for one person.</p>
<p>2. P thinks provider selection process can be improved in the following ways: providing more information on list of well matched providers, properly funding the mental health system, and making mental health care services readily available by removing barriers to finding a provider.</p>	<p>2.1 To improve the provider selection process, P thinks it's important for patients to have more say in their healthcare. This would make them content with the services they receive and ultimately increase the ratings of providers since they will be properly matched.</p> <p>2.2 P thinks having provider success rates will be helpful in improving the provider selection process.</p> <p>2.3 P believes a larger provider pool, a more streamlined intake process and better transportation to the location would improve the mental health care selection process.</p>
3. N/A	<p>3.1 P would rather make the provider selection than have a primary care doctor make the selection because P knows P's own issues and needs best.</p> <p>3.2 P would trust P's doctor to make a referral for a provider.</p> <p>3.3 P would be comfortable with either P or another professional making a provider selection based on a list of well-matched providers, but would prefer that P make the provider selection.</p>
<p>4. P also thinks it would be great if counselors could collaborate with psychiatrists.</p>	<p>4.1 P thinks providers should use surveys to determine whether a certain provider is fulfilling people's needs or not and researchers should use the information to improve providers care and find new ways of treatment.</p>
5. N/A	<p>5.1 P would want a website with specialization of providers and</p>

success rates.

5.2 P would have liked to know more information regarding a provider's area of expertise before making a selection.

6. P thinks clinic can use the data to make adjustment like hiring more providers or hiring some providers with specialties that are needed by the community.

6.1 P thinks the questionnaire asking about difficulties should be used to understand patients more and to better the psychiatrist and the therapists.

6.2 P thinks provider performance questionnaires should be used to ensure patients and providers are good matches.

7. P would like to have more information about services that providers have and any helpful information about mental illness. P would prefer to find a place that has a therapist and a psychiatrist working together. P would want access to more self-care information and tips on how to manage stress.

7.1 P would like a website that is easy to access, with provider background and style.

Note. P = Patient.

Table 23

Team 2 Exemplars for Domain 6: Experience with Surveys/Questionnaires

Category	Subcategory
1. P has completed a few surveys related to provider satisfaction feedback.	1.1 P has completed many surveys related to symptoms and quality of life.
2. P does not think that P has filled out questionnaires regarding service satisfaction.	<p>2.1 P has filled out a survey relating to symptoms, functioning, and quality of life.</p> <p>2.2 P has completed a questionnaire about satisfaction with provider.</p> <p>2.3 P has never completed provider satisfaction questionnaires.</p> <p>2.4 P didn't complete a questionnaire asking about difficulties, symptoms, functioning, and quality of life.</p>
3. N/A	<p>3.1 P thinks provider satisfaction questionnaires are used in groups to compare and adjust treatment services to improve and correct them. P thinks they should compare questionnaire results with other agencies to see similarities/differences.</p> <p>3.2 P is unsure of how symptoms and functioning questionnaire information is used but hopes it is used to guide P's therapist to finding the best way to help P with P's diagnosis.</p> <p>3.3 P feels that the questionnaires P filled out regarding personal well being provide insight into how P is feeling and allows provider to make more informed treatment decisions.</p>
4. N/A	4.1 P has completed provider satisfaction questionnaires and calls them quality assurance questionnaires. P thinks these questionnaires are used so providers make sure they are giving the best services to patients.

Note. P = Patient.

Table 24

Team 2 Exemplars for Domain 7: Opinions About Well-Matched Provider

Category	Subcategory
1. P would rather find a provider himself/herself rather than have a primary care doctor find one for P.	1.1 P would feel comfortable making a selection from a list of matched providers.
2. P would rank seeing a matched provider equal to the other factors P listed such as experience and accessibility.	2.1 P would rank seeing a well-matched provider as less important than finding a provider who is knowledgeable, invested, and friendly.
3. P thinks provider being a good listener is more important than the location, P is willing to go further to see a matched provider.	
4. P would not be willing to wait to see a matched provider if they were currently unavailable due to the severity of P's mental illness.	4.1 P is willing to wait longer to see a preferred provider. 4.2 P would be willing to wait for provider if P feels mentally stable but will not if P was just starting out.
5. N/A	5.1 P is not willing to pay more to see a matched provider. 5.2 P is willing to pay more for a matched provider. 5.3 If the copay was reasonable, P is willing to pay more for a matched provider.

Note. P = Patient.

Table 25

CQR Process Measure Scores across Time

	<u>Team 1</u>		<u>Team 2</u>		<u>Mean Difference</u>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Hedges <i>g</i> *
Time 1	52.50	3.32	52.75	1.89	-0.08
Time 2	53.75	2.63	52.00	3.46	0.50
Time 3	53.75	2.63	50.75	3.59	0.83
Time 4	53.75	2.87	51.25	2.63	0.79
Time 5	54.00	2.16	50.25	3.69	1.08
Time 6	53.75	2.22	51.00	3.16	0.87
Time 7	53.50	2.65	51.75	3.30	0.51
Mean Climate	53.57	2.59	51.39	2.85	0.70

Note.

*Positive values indicate that Team 1 reported more positive group climate than Team 2, whereas negative values indicate that Team 2 reported more positive group climate than Team 1.

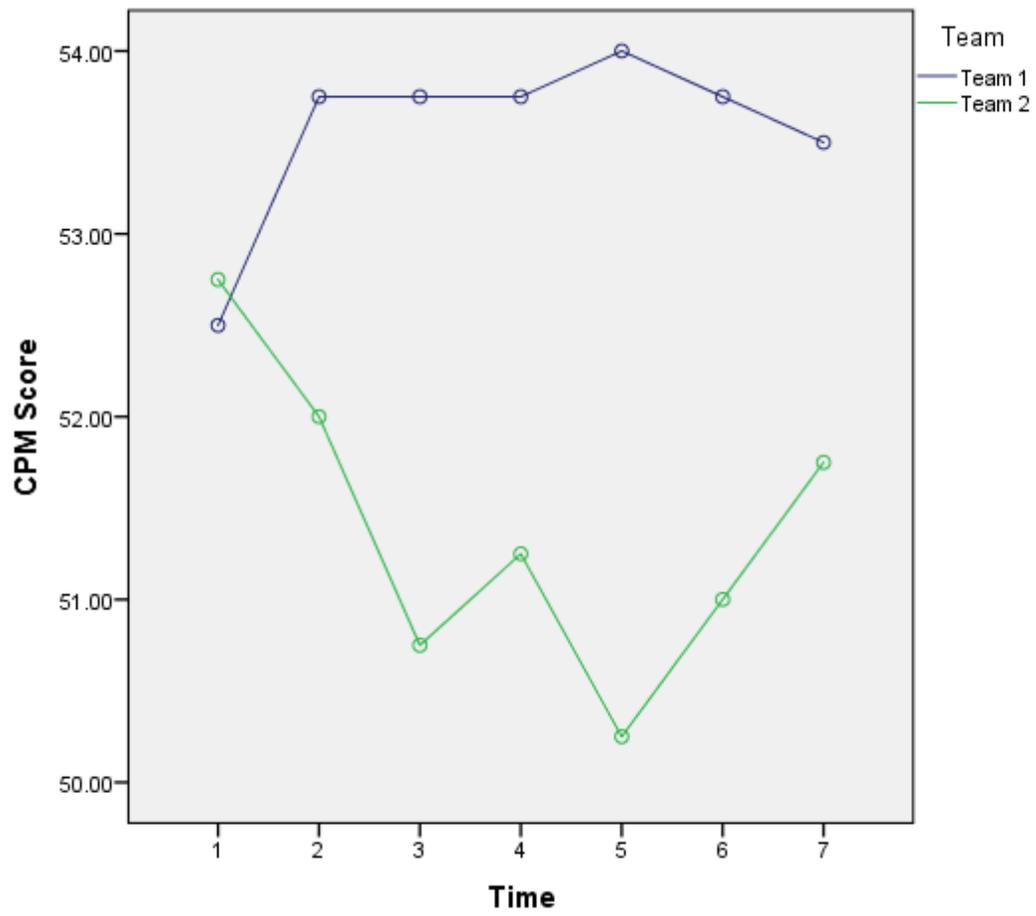


Figure 1. Between-team differences in within-team climate change across all process measure administrations.



Figure 2. Team 1's focus group themes and sub-themes.

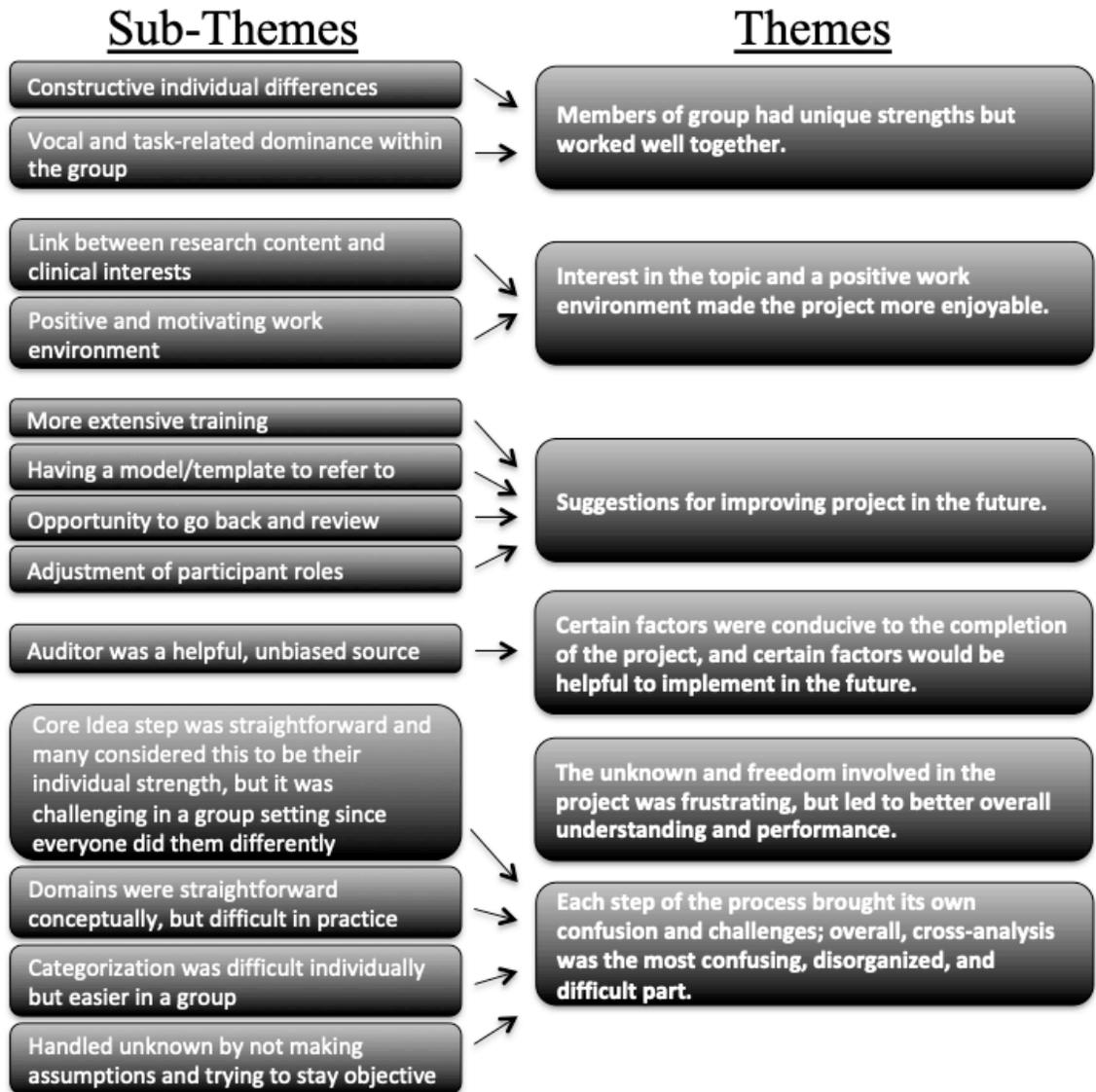


Figure 3. Team 2's focus group themes and sub-themes.

APPENDIX A

CONSUMER TELEPHONE INTERVIEW ITEMS

Orientation: Health care is devoting more attention and resources to things such as “quality” or “outcome” assessment. For example, hospitals collect a lot of information regarding how long a typical patient stays on a unit, how many people are re-admitted after a procedure due to complications, and how more chronic diseases are managed. Hospitals receive quality ratings, and health care providers can be similarly tracked; for example, certain surgeons have better cardiac surgery outcomes than others. A variety of information can be used to compare and contrast treatment providers, and this is also the case for mental health care when changes in symptoms, functioning, employment, quality of life, frequency of hospitalizations, etc. are tracked. The field is capable of identifying “higher” and “lower” performing providers based on the average improvement (or lack thereof) of their clients. With these capabilities in mind, we would like to get your perspectives on the use of provider performance information for mental health care decision-making.

This interview will be audio recorded, and anything that you disclose to me will remain confidential to the research team and used for research purposes only. Transcripts of the interviews will remove all personally identifying information. Please be assured that only members of our research team will have access to these transcripts.

1. What mental health services are you (or the identified client) currently receiving?
Have received in the past?
2. What types of mental health care providers have you (or the identified client) worked with?
3. Can you describe, in as vivid detail as possible, what your experience has been when trying to identify a mental health care provider with whom to work (or with whom the identified can work)?
 - How have you found providers in the past?
 - Has it been easy or difficult to find someone?
 - Did you know anything about your provider(s) before the first appointment?
 - Did you feel as though you had a say in selection?
4. What do you think would have been helpful/what would you have wanted to help you make a more informed choice regarding your treatment provider?

5. In the context of receiving services, have you ever been asked to complete a questionnaire asking you about your (or the identified client's) difficulties (e.g., symptoms, functioning, quality of life) and/or satisfaction with your provider or services?
 - a. If so, how do you think that information is used?
 - b. How should it be used?
6. Do you, or would you, want access to provider performance information (i.e., provider success rates in treating their previous clients)? Why or why not?
 - a. If you do want access, how would you prefer to receive it? E.g., Online? Through another provider who is referring you? As a handout?
7. What factors are important to you when selecting a treatment provider (or when considering a provider for the identified client)?
 - a. Are these factors more, less, or equally important? Can you rank them in terms of importance?
8. Given its capability of identifying "higher" and "lower" performing providers based on their track record, the field also could potentially match clients to providers who seem particularly well suited for them. Would you be willing to wait to see a "preferred" or matched provider if he or she was currently unavailable?
 - a. Would you pay more?
 - b. Are there other things in your previously stated list of important factors in making provider decisions that you would rank lower than seeing a "preferred" or matched provider were it possible to do so?
9. If you were given a list of "well-matched" providers, would you feel comfortable making a provider selection based on that list?
 - a. Alternatively, would you feel more comfortable if other professionals use such a list to make a specific recommendation or recommendations? That is, would you want your primary care doctor or outpatient care setting to make a specific referral recommendation based on this information rather than have direct access to it yourself?
 - b. Would you prefer a list of "preferred" providers in your area, or the identification of a specific provider?
10. What are the potential costs or pitfalls to using provider track record information for provider selection?

11. Should higher performing providers be reimbursed at a higher rate?
12. Please discuss, in as much detail as you can, the steps that you think should be taken to improve the mental health care provider selection process. What would be most helpful to you? We are very interested in your voice on this important matter.

APPENDIX B
CQR PROCESS MEASURE

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
1. I feel able to freely speak my mind about the data and content under discussion during meetings.	1	2	3	4	5	6	7
2. I voice my concerns when I disagree with fellow group members during meetings.	1	2	3	4	5	6	7
3. I feel comfortable and positive about meeting with my team members to come to consensus.	1	2	3	4	5	6	7
4. The opinions and perspectives that I bring to team meetings is an important part of the CQR process.	1	2	3	4	5	6	7
5. My opinions and perspectives are valued by other members of the coding team.	1	2	3	4	5	6	7
6. I adequately prepare for the team meetings by completing my independent coding appropriately.	1	2	3	4	5	6	7
7. I experience feelings of burnout from the amount of time I have dedicated to this project.	1	2	3	4	5	6	7
8. Overall, our team has done a good job thus far completing the tasks inherent in the CQR process.	1	2	3	4	5	6	7

Note. Item 7 is reverse-scored.

APPENDIX C

FOCUS GROUP SEMI-STRUCTURED INTERVIEW INSTRUCTIONS AND ITEMS

Think of today's focus group interview as a brainstorming session as a team; you can agree and/or disagree with your peers, add to their perspectives or not, and choose to comment or not for each topic. This focus group will allow me to get a better understanding of your *experience* of the data, the project, and of each other. We have 2 hours set aside for today's focus group, so don't worry about saying anything you think is "off topic;" anything you feel is relevant is fair game. Thus, even though there are some structured questions, you have a lot of latitude to touch on the things YOU feel are important. This interview will be audio and video recorded, and anything that you disclose to me will remain confidential to the research team and used for research purposes only. Transcripts of the interviews will remove all personally identifying information. Please be assured that only members of our research team will have access to these transcripts. You may cease participation in this focus group at any time.

For the interview, I have several prompting questions, but I will also give you ample time and space to discuss your responses *in full and vivid detail*. Most importantly, this is a safe environment, and there are genuinely no "right" or "wrong" answers about anything.

Are there any questions or concerns about anything before we get started?

Overall/General Questions

1) Looking back at your time during the CQR study, briefly discuss your overall experience of participating in this project.

-What immediately comes to mind when you think about your time on this project?

2) What aspects of the project did you find to be most straightforward? What did you find to be most confusing?

-What could be done in the future to make those aspects more clear or understandable?

Feelings of Competency

3) What were the aspects of the project you feel you handled best as an individual? As a team?

4) What do you think you struggled with or found most difficult as an individual? As a team?

Experiences During Consensus

5) Discuss what role, if any, groupthink may or may not have played in your consensus meetings.

-How did your auditor help you to address these concerns?

6) What aspects of consensus meetings did you find to be most frustrating or distressing?

-What were the most or least helpful ways of navigating these issues?

7) What aspects of consensus meetings did you find to be most satisfying or rewarding?

-How, if at all, did your perception of these aspects affect your work as a team?

8) What do you think compelled you to speak up or remain silent when you had an idea during team meetings?

-How might you shift this to help foster greater balance for team dynamics in the future?

Wrap Up

9) All things considered, what do you feel could be done differently for coding teams in the future to foster trustworthy findings?

10) Is there any aspect of the project you feel we did not touch on today?

APPENDIX D

CQR DISSERTATION ANONYMOUS FEEDBACK FORM

This space below is designed to let you provide any ENTIRELY ANONYMOUS feedback about any aspect of the CQR dissertation project. In addition to raising any points you may not have felt comfortable discussing during the focus group interview, you are encouraged to share anything else you may feel it would be helpful for me to know about the CQR process or your experience of it. Your response may be long or short.

Potential examples (which are not exhaustive) might include commenting about your experience of not feeling comfortable speaking up in meetings or feeling that I did not do a good job facilitating a certain aspect(s) of the project. This is entirely so I can learn about how to improve this process for future CQR teams.

To be crystal clear, all 8 members of the project are receiving this same survey, so unless you indicate any identifying information about yourself or your team in your response I will have no way to identify you. In fact, I encourage you to leave your responses entirely anonymous. I will also NOT follow up your response unless you indicate explicitly that you would like for me to do so. You may also simply submit “no feedback” if you don’t have any thoughts, which is entirely fine, too. This is just a final way to ensure you objectively have an anonymous means through which to submit study feedback.

APPENDIX E
STUDY FLOW CHART

Study Phase	CQR Process Measure
Training	
Project orientation	
Read CQR materials	
Watch 3 training videos	
Review interview questions	
Record expectations & biases	
Pilot coding #1	
Independent, consensus, audit, finalize	
Pilot coding #2	
Independent, consensus, audit, finalize	
Domain Generation	
Immersion in 12 transcripts	
Establish memos	
Independent, consensus, audit	✓
Open Coding	
Transcript #1	
Independent, consensus, audit, finalize	✓
Transcript #2	
Independent, consensus, audit, finalize	
Transcript #3	
Independent, consensus, audit, finalize	
Transcript #4	
Independent, consensus, audit, finalize	✓
Transcript #5	
Independent, consensus, audit, finalize	
Transcript #6	
Independent, consensus, audit, finalize	
Transcript #7	
Independent, consensus, audit, finalize	✓
Transcript #8	

Independent, consensus, audit, finalize

Transcript #9

Independent, consensus, audit, finalize

Transcript #10

Independent, consensus, audit, finalize

✓

Transcript #11

Independent, consensus, audit, finalize

Transcript #12

Independent, consensus, audit, finalize

✓

Categorization

Independent, consensus, audit, finalize

Cross Analysis

Independent, consensus, audit, finalize

✓

Focus Groups

Conduct interviews

Administer anonymous feedback form

Transcribe interviews

Quantitative and Qualitative Analyses

Statistical analyses

Thematic analysis

Study Write-Up

APPENDIX F

CQR DISSERTATION PARENT SURVEY IMMERSION QUESTIONNAIRE & EXPECTATIONS AND BIASES

Please read this statement carefully and in its entirety:

We ask that you please set aside approximately 45-60 minutes to complete this survey in its entirety, as we would like you to reflect carefully, especially on the open-ended questions.

To help better understand your expectations and biases about this project, from both *OUR* and *YOUR* perspectives as researchers, we ask that you please complete the same demographic questionnaire asked of the original study participants. It is our hope that these questions will help you think about your own mental health experiences, or lack thereof, as you prepare to study participant responses. Keeping your own expectations and biases in mind during qualitative analysis is an important step in the process. Although we hope in some ways to check this bias, we also recognize as qualitative researchers that the data we see and interpret passes through our own lenses and experiences. Thus, it will be important to keep your expectations and biases in mind throughout data analysis.

It is important to note that you may leave any question blank that you do not feel comfortable answering. That being said, your responses will remain entirely de-identified, and will not be linked to you in any way. That is, all answers will be scrambled when they are submitted, and can only be attributed to your team. For example, there will be no way for the principal investigator to tie together race, gender, and mental health experience variables, as they will all be scrambled upon submission. The principal investigator will only see the data aggregated by team, and the data will only be analyzed in aggregate form. Your confidentiality is paramount to the principal investigator and will be carefully safeguarded.

1. Age in years: _____

2. Gender:

_____ female

_____ male

_____ transgender

_____ Other gender

3. Race/ethnicity:

_____ Hispanic/Latino(a)

_____ East Asian

_____ South Asian

- African American/Black
 - Native American
 - European American/White
 - Hawaiian/Pacific Islander
 - Biracial/multiracial
 - Other race/ethnicity
4. Marital status:
- separated or divorced
 - married/in a domestic partnership
 - single
 - widowed
5. Household annual income:
- Less than \$25,000
 - \$25,001 - \$50,000
 - \$50,001 - \$75,000
 - \$75,001 - \$100,000
 - \$101,001+
6. Religion – if no affiliation, please write "none":
-
7. Highest education level:
- Did not complete high school or GED
 - Completed high school or GED
 - Some college
 - Completed an Associate's degree
 - Completed a four year college degree
 - Some graduate or medical school
 - Completed a master's degree
 - Completed a doctoral or MD degree

As a reminder, it is important to note that you may leave any question blank that you do not feel comfortable answering. That being said, your responses will remain entirely de-identified, and will not be linked to you in any way. That is, all answers will be scrambled when they are submitted, and can only be attributed to your team. For example, there will be no way for the principal investigator to tie together race, gender, and mental health experience variables, as they will all be scrambled upon submission. The principal investigator will only see the data aggregated by team, and the data will only be analyzed in aggregate form. Your confidentiality is paramount to the principal investigator and will be carefully safeguarded.

Treatment History:

8. Mental health service(s) currently receiving or seeking. Please check all that apply:

Individual psychotherapy

Group psychotherapy

Medication

Marital therapy

Family therapy

Case management

In home therapy

Family support

School based services

Community support

Other, please

describe: _____

9. Who referred you or played a key role in your referral to your current treatment setting?
Check all that apply:

Primary/family care doctor

Other mental health care provider

Insurance company

Friend or family member

Self

Other, please describe: _____

10. Which of the following problems best capture your reason for seeking services at this time? Check all that apply:

Depression

Anxiety

Trauma

Eating disorder

Marital or family problems

Behavioral problems

Development disability

Learning disability

Attentional problems

Brain injury

Psychotic symptoms

Bipolar disorder

_____ Substance use
_____ Other, please
describe: _____

11. Do you receive Medicare or Medicaid (for Massachusetts residents, includes MassHealth)?

_____ Yes

_____ No

12. Have you received mental health services in the past?

_____ Yes

_____ No

13. How many different mental health therapists or counselors have you seen in your lifetime? _____

14. How many different providers have prescribed you medication for your mental health in your lifetime? _____

15. Were there any times in your life when you wanted a mental health provider and could not find one?

_____ Yes

_____ No

16. Has it been hard to find a mental health provider who you were confident could help you?

_____ Yes

_____ No

17. Has a health care provider or agency ever recommended a specific mental health care provider to you?

_____ Yes

_____ No (if No, please skip to question 18)

17a. If you responded "Yes" to question 17, did the person or agency that gave you the recommendation explain what the recommendation was based on?

_____ Yes

_____ No

18. Has any professional ever discussed with you the pros and cons of choosing one mental health provider vs. another?

_____ Yes

_____ No

19. Have you ever used a consumer satisfaction rating website, such as Angie's List or Healthgrades, to find a mental health care provider?

Yes

No (if No, please skip to question 20)

19a. If you responded “Yes” to question 19, did you find the website helpful in finding a provider?

Yes

No

20. Do you believe that all mental health care providers are capable of helping you?

Yes

No

21. Imagine that you could see a list of mental health providers’ track records in helping people with issues like your own (that is, a list of the percentage of people who they have helped versus the percentage of people who they have not helped). Would you trust these data and how they were collected?

Yes

No

22. Imagine that you could see a list of mental health providers’ track records in helping people with issues like your own. Would you use this list to help you select your provider?

Yes

No

23. Imagine that a health care professional like your primary care doctor is giving you a referral for a mental health provider. Would you feel more confident about your options if you knew that this person had reviewed providers’ track records in helping people like you?

Yes

No

24. Imagine that your insurance company is giving you a referral for a mental health provider. Would you feel more confident about your options if you knew that your company had reviewed providers’ track records in helping people like you?

Yes

No

25. Would you pay more out of pocket to see a mental health care provider who is listed as highly effective in treating the problems that you have?

Yes

No

26. Should mental health care consumers have access to information on the track records of providers in the local area?

Yes

No

27. Would it be important for you to be assigned or referred to a mental health care provider based on their track record in helping people with issues like your own?

_____ Yes

_____ No

28. Would it be more important than usual for you to be assigned or referred to a mental health care provider based on their track record in helping people with issues like your own **IF** you previously have not benefited from mental health treatment for that problem?

_____ Yes

_____ No

29. Do you think access to information on the track records of mental health care providers would increase the likelihood of someone being helped by treatment?

_____ Yes

_____ No

30. Do you think that matching a consumer with a provider who has a track record of helping people with similar issues would increase the likelihood of that consumer being helped by treatment?

_____ Yes

_____ No

First, please consider your expectations REGARDING THE STUDY CONTENT. Expectations can be defined as “beliefs that researchers have formed based on reading the literature and thinking about and developing the research questions.” Please review the “Provider Performance Interview Questions,” located in the Box folder under “Other.” What, broadly or specifically, do you expect to find REGARDING THE STUDY CONTENT? There is no right or wrong answer; rather, it is only important to critically reflect on what you might find. Please write out your response in a word document before copying and pasting into this box.

Next, please consider your expectations REGARDING THE STUDY PROCESS. That is, what are your expectations regarding training? What are they for your individual performance? Do you have any thoughts about what the group process might look like as you come to consensus? There is no right or wrong answer; rather, it is only important to critically reflect on what you might expect. Please write out your response in a word document before copying and pasting into this box.

Now, please consider your biases REGARDING THE STUDY CONTENT. Biases in CQR can be defined as “personal issues that make it difficult for researchers to respond objectively to the data.” We all have biases, and our biases can be based anything, including the literature we've read, the people we've talked to, and our own lived experiences. There is no right or wrong answer; rather, it is only important to critically reflect on what your biases might be. Please write out your response in a word document before copying and pasting into this box.

Now, please consider your biases REGARDING THE STUDY PROCESS. Biases in CQR can be defined as “personal issues that make it difficult for researchers to respond objectively to the data.” We all have biases, and our biases can be based anything, including the literature we’ve read, the people we've talked to, and our own lived experiences. There is no right or wrong answer; rather, it is only important to critically reflect on what your biases might be. Please write out your response in a word document before copying and pasting into this box.

BIBLIOGRAPHY

- American Psychological Association. (2010). *American Psychological Association ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/code/>
- Aubé, C., & Rousseau, V. (2005). Team goal commitment and team effectiveness: The role of task interdependence and supportive behaviors. *Group Dynamics: Theory, Research, and Practice, 9*, 189-204. doi:10.1037/1089-2699.9.3.189
- Ayoko, O. B., Callan, V. J., & Härtel, C. E. J. (2008). The influence of team emotional intelligence climate on conflict and team members' reactions to conflict. *Small Group Research, 39*, 121-149. doi:10.1177/1046496407304921
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review, 81*, 392-425. doi:10.1037/h0037024
- Boswell, J. F., Constantino, M. J., Oswald, J. M., Bugatti, M., Goodwin, B. J., & Yucel, R. (2018). Mental health care consumers' relative valuing of clinician performance information. *Journal of Consulting and Clinical Psychology, 86*, 301-308. doi:10.1037/ccp0000264
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25*, 6-19. doi:10.1080/10503307.2013.817696
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101. doi:10.1191/1478088706qp063oa
- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy, 48*, 34-42. doi:10.1037/a0022063
- Carminati, L. (2018). Generalizability in qualitative research: A tale of two traditions. *Qualitative Health Research, 28*, 2094-2101. doi:10.1177/1049732318788379
- Chronister, J., Chou, C., Kwan, K. K., Lawton, M., & Silver, K. (2015). The meaning of social support for persons with serious mental illness. *Rehabilitation Psychology, 60*, 232-245. doi:10.1037/rep0000038
- Constantino, M. J., Morrison, N. R., MacEwan, G., & Boswell, J. F. (2013). Therapeutic alliance researchers' perspectives on alliance-centered training practices. *Journal of Psychotherapy Integration, 23*, 284-289. doi:10.1037/a0032357
- Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics, 10*, 249-271. doi:10.1037/1089-2699.10.4.249

- Driskell, J. E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors*, *34*, 277-288. doi:10.1177/001872089203400303
- Ellis, C., & Berger, L. (2003). Their story/my story/our story: Including the researcher's experience in interview research. In J. A. Holstein & J. F. Gubrium (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 467-493). Thousand Oaks, CA: Sage.
- Gergen, K. J. (2014). Pursuing excellence in qualitative inquiry. *Qualitative Psychology*, *1*, 49-60. doi:10.1037/qup0000002
- Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 191-215). Thousand Oaks, CA: Sage.
- Hays, D. G., Wood, C., Dahl, H., & Kirk-Jenkins, A. (2016). Methodological rigor in *Journal of Counseling & Development* qualitative research articles: A 15 year review. *Journal of Counseling & Development*, *94*, 172-183. doi:10.1002/jcad.12074
- Hill, C. E. (Ed.). (2012). *Consensual qualitative research: A practical resource for investigating social science phenomena*. Washington, DC: American Psychological Association.
- Hill, C. E., Chui, H., & Baumann, E. (2013). Revisiting and reenvisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. *Psychotherapy*, *50*, 68-76. doi:10.1037/a0030571
- Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of Counseling Psychology*, *52*, 196-205. doi:10.1037/0022-0167.52.2.196
- Kisely, S., & Kendall, E. (2011). Critically appraising qualitative research: A guide for clinicians more familiar with quantitative techniques. *Australasian Psychiatry*, *19*, 364-367. doi:10.3109/10398562.2011.562508
- Knox, S., Burkard, A. W., Johnson, A. J., Suzuki, L. A., & Ponterotto, J. G. (2003). African American and European American therapists' experiences of addressing race in cross-racial psychotherapy dyads. *Journal of Counseling Psychology*, *50*, 466-481. doi:10.1037/0022-0167.50.4.466
- Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). Los Angeles: Sage.

- Ladany, N., Inma, A. G., Hill, C. E., Know, S., Crook-Lyon, R. E., Thompson, B. J., ... Walker, J. A. (2012). Corrective relational experiences in supervision. In L. G. Castonguay & C. E. Hill (Eds.), *Transformation in psychotherapy: Corrective experiences across cognitive behavioral, humanistic, and psychodynamic approaches* (pp. 335-352). Washington, DC: American Psychological Association.
- Levitt, H. M. (2015). Qualitative psychotherapy research: The journey so far and future directions. *Psychotherapy, 52*, 31-37. doi:10.1037/a0037076
- Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology, 4*, 2-22. doi:10.1037/qup0000082
- Liamputtong, P. (2011). *Focus group methodology: Principles and practice*. London: Sage.
- Lysaker, P. H., Kukla, M., Belanger, E., White, D. A., Buck, K. D., Luther, L., ... Leonhardt, B. (2015). Individual psychotherapy and changes in self-experience in schizophrenia: A qualitative comparison of patients in metacognitively focused and supportive psychotherapy. *Psychiatry, 78*, 305-316. doi:10.1080/00332747.2015.1063916
- McLeod, J. (2000). The contribution of qualitative research to evidence-based counselling and psychotherapy. In N. Rowland & S. Goss (Eds.), *Evidence-based counselling and psychological therapies: Research and applications* (pp. 112-126). New York, NY: Routledge.
- Moret, M., Reuzel, R., Van Der Wilt, G. J., & Grin, J. (2007). Validity and reliability of qualitative data analysis: Interobserver agreement in reconstructing interpretative frames. *Field Methods, 19*, 24-39. doi:10.1177/1525822X06295630
- Morrison, N. R., Constantino, M. J., Westra, H. A., Kertes, A., Goodwin, B. J., & Antony, M. M. (2017). Using interpersonal process recall to compare patients' accounts of resistance in two psychotherapies for generalized anxiety disorder. *Journal of Clinical Psychology, 73*, 1523-1533. doi:10.1002/jclp.22527
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology, 52*, 250-260. doi:10.1037/0022-0167.52.2.250
- Mullen, B., Anthony, T., Salas, E., & Driskell, J. E. (1994). Group cohesiveness and quality of decision making: An integration of tests of the groupthink hypothesis. *Small Group Research, 25*, 189-204. doi:10.1177/1046496494252003

- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity*, *41*, 233-249. doi:10.1007/s11135-006-9000-3
- Pejtersen, J. H., Kristensen, T. S., Borg, V., & Bjorner, J. B. (2010). The second version of the Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health*, *38*, 8-24. doi:10.1177/1403494809349858
- Pezalla, A. E., Pettigrew, J., & Miller-Day, M. (2012). Researching the researcher-as-instrument: An exercise in interviewer self-reflexivity. *Qualitative Research*, *12*, 165-185. doi:10.1177/1468794111422107
- Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, *47*, 1451-1458. doi:10.1016/j.ijnurstu.2010.06.004
- Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on paradigms and philosophy of science. *Journal of Counseling Psychology*, *52*, 126-136. doi:10.1037/0022-0167.52.2.126
- Renn, R. W. (2003). Moderation by goal commitment of the feedback-performance relationship: Theoretical explanation and preliminary study. *Human Resource Management Review*, *13*, 561-580. doi:10.1016/j.hrmr.2003.11.003
- Sanders, C. B., & Cuneo, C. J. (2010). Social reliability in qualitative team research. *Sociology*, *44*, 325-343. doi:10.1177/0038038509357194
- Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118-137). Thousand Oaks, CA: Sage.
- Sousa, D. (2014). Validation in qualitative research: General aspects and specificities of the descriptive phenomenological method. *Qualitative Research in Psychology*, *11*, 211-227. doi:10.1080/14780887.2013.853855
- Staller, K. M. (2013). Epistemological boot camp: The politics of science and what every qualitative researcher needs to know to survive in the academy. *Qualitative Social Work: Research and Practice*, *12*, 395-413. doi:10.1177/1473325012450483
- Thomas, E., & Magilvy, J. K. (2011). Qualitative rigor or research validity in qualitative research. *Journal for Specialists in Pediatric Nursing*, *16*, 151-155. doi:10.1111/j.1744-6155.2011.00283.x
- Vivino, B. L., Thompson, B. J., Hill, C. E., & Ladany, N. (2009). Compassion in psychotherapy: The perspective of therapists nominated as compassionate. *Psychotherapy Research*, *19*, 157-171. doi:10.1080/10503300802430681

- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research, 11*, 522-537. doi:10.1177/104973201129119299
- Williams, E. N., & Hill, C. E. (2012). Establishing trustworthiness in consensual qualitative research studies. In C. E. Hill (Ed.), *Consensual qualitative research: A practical resource for investigating social science phenomena* (pp. 175-185). Washington, DC: American Psychological Association.
- Williams, E. N., & Morrow, S. L. (2009). Achieving trustworthiness in qualitative research: A pan-paradigmatic perspective. *Psychotherapy Research, 19*, 576-582. doi:10.1080/10503300802702113