



2015-06-01

Process Feedback in Group Psychotherapy: A Qualitative Inquiry into Leader Implementation of GQ/OQ Feedback

Sean Cameron Woodland
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Psychology Commons](#)

BYU ScholarsArchive Citation

Woodland, Sean Cameron, "Process Feedback in Group Psychotherapy: A Qualitative Inquiry into Leader Implementation of GQ/OQ Feedback" (2015). *All Theses and Dissertations*. 5606.
<https://scholarsarchive.byu.edu/etd/5606>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Process Feedback in Group Psychotherapy: A Qualitative Inquiry
into Leader Implementation of GQ/OQ Feedback

Sean Cameron Woodland

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Gary M. Burlingame, Chair
Bruce N. Carpenter
Robert L. Gleave
Mark E. Beecher
Joseph A. Olsen

Department of Psychology

Brigham Young University

June 2015

Copyright © 2015 Sean Cameron Woodland

All Rights Reserved

ABSTRACT

Process Feedback in Group Psychotherapy: A Qualitative Inquiry into Leader Implementation of GQ/OQ Feedback

Sean Cameron Woodland
Department of Psychology, BYU
Doctor of Philosophy

This dissertation explores what it means to “Act Upon” measure-based feedback in the group therapy context. Eleven group leaders at three college counseling centers were provided feedback completed by their group members using the Group Questionnaire (GQ) and the Outcome Questionnaire (OQ-45). Researchers selected two a priori ways in which the feedback could be “acted upon”: via GQ Use and GQ Value. Leaders reported their use and value of the feedback using two data sources—weekly leader slips and end-of-term debrief interview transcripts. Both sources of data were content analyzed across several phases of coding. The resultant categories are intended to provide a preliminary understanding of how leaders treat the feedback received. Dimensions were then added to consolidate the meaning of the categories into a temporal pattern. Finally, using the resultant data, a scheme for quantifying the “acted upon” construct is proposed in effort to develop a potential moderator or mediator variable for future quantitative analyses. Implications of the dissertation are then discussed.

Keywords: GQ, OQ, group psychotherapy, feedback, content analysis

ACKNOWLEDGMENTS

I would like to first acknowledge Gary Burlingame, my dissertation chair. He has been there to guide my professional steps since I was his undergraduate assistant in 2008. His skill with pushing me and my ambitions to greater heights was just what I needed over the last seven years. I am privileged to have had him as my mentor, and I will continue to be privileged to call him my friend. I would like to thank Bruce Carpenter, whose professional decorum always reminded me that I was being made into something special. I would also like to thank Mark Beecher and Robert Gleave, who helped me understand the importance of remembering what it means to study human beings. I would also like to thank Joseph Olsen, who was ever patient with me as I learned and re-learned the importance and application of statistical analyses.

I would also like to acknowledge the hard work of Kait Whitcomb, who engaged stride for stride with me in this research project. As a co-rater, she was invaluable in providing an extra perspective to the content of this work. She also was invaluable in her ability to manage the majority of the parent RCT, and in helping get signatures while I was away on internship.

Finally, I'd like to acknowledge my fiancée, Jennifer Davies. She has been there to encourage me through in what was in some ways the most difficult part of this project. She helped me when I felt burnt out, and has been there in times when I felt utterly alone in my final push toward the completion of this project. I will always be in debt for her patience and longsuffering on my behalf.

Table of Contents

Table of Contents	iv
List of Tables and Figures	v
Introduction.....	1
Literature Review.....	1
Effects of Using Outcomes-Based Feedback	3
Adding Process Measures to the Feedback Delivery.....	4
Outcome/Progress Feedback in Group Treatments	5
Feedback on Therapeutic Processes in Group Treatment	7
Overview of the Parent Randomized Clinical Trial.....	10
Rationale and Hypotheses.....	10
Participants.....	14
Procedures.....	16
Current Study	18
Method	19
Unitizing	20
Open Coding of Qualitative Content	21
Results.....	23
Discussion.....	59

List of Tables and Figures

Table 1	11
Table 2	13
Table 3	25
Table 4	26
Table 5	27
Table 6	29
Table 7	39
Table 8	43
Table 9	46
Table 10	52
Table 11	55
Table 12	56
Table 13	57
Figure 1	12
Figure 2	60

Process Feedback in Group Psychotherapy: A Qualitative Inquiry into Leader Implementation of GQ/OQ Feedback

It is well-known that the psychotherapy landscape continues to emphasize greater objectivity, measurability and accountability in clinical practice. Fittingly, group process and outcome measures have had an increasing influence on assisting therapists to assess the quality of their therapeutic work (Burlingame & Beecher, 2008; Lambert & Ogles, 2004). One method that is increasing in frequency is outcome and process measures that provide feedback to the clinicians regarding possible obstacles a client may be facing to achieve a good outcome. These measures provide measure-based alerts from domains such as depressive or anxious symptoms, the quality of the therapeutic relationship, social support, or motivation for change (Lambert et al., 2004). The effect of receiving feedback has been assessed using randomized controlled trials for individual therapy, and a necessary next step is to test the efficacy of similar measures in group treatment. The dissertation undertaken is an initial step in making this assertion, relying on data from a 30-month, multisite randomized clinical trial of group outcome and process feedback in college counseling centers (Burlingame & Beecher, 2012).

Literature Review: The Use of Outcomes-based Feedback in Individual Psychotherapy

The individual outcomes literature is full of examples using feedback to inform treatment progress and outcomes in what has been termed a *patient-focused* paradigm (Howard, Moras, Brill, Martinovich, & Lutz, 1996). Since the introduction of this paradigm there have been several quality assurance systems (e.g., Anker, Duncan, & Sparks, 2009; Barkham et al., 2001; Beutler, 2001; Kordy, Hannover, & Richard, 2001; Lambert, Hansen, & Finch, 2001) created that all strive for continuous monitoring of patient outcomes that leads to progress feedback to the clinician; often, this feedback is even delivered in real-time. The most empirically supported

feedback system is the OQ-Analyst online software program (Shimokawa, Lambert, & Smart, 2010). In fact, the OQ-Analyst is one of only two outcome feedback systems with sufficient empirical support to be formally recognized by the National Registry of Evidence-based Programs and Practices (NREPP; SAMHSA, 2012).

The OQ-Analyst provides three types of “alerts” for clinicians: (a) those related to change in a client’s score on an OQ measure (typically the OQ-45) across treatment sessions that is of statistical import (**change alerts**); (b) those related to how a client’s current score compares to normative values using empirically derived cut-off scores (**absolute alerts**); and, (c) those related to a client’s expected progress in therapy (**progress alerts**). **Change alerts** are triggered by the reliable change index (RCI; Jacobson & Truax, 1991). The RCI informs the clinician if change on an outcome or process measure is statistically significant or simply attributable to measurement error. **Absolute alerts** are produced when a client’s outcome score moves from one normative population to another. Typically, two normative populations (clinical and normal) are used to produce absolute alerts and these two populations are separated by a cutoff score that reflects the 50th percentile value that evenly separates the two normative samples. Other cut scores have been used to define extreme scores, e.g., 90th and 95th percentiles. **Progress alerts** are produced when a client’s outcome score changes (i.e., deteriorates or improves) beyond what can be expected from normative change expectations that are sometimes referred to as optimal recovery curves (Howard et al., 1996). These curves provide a clinician with information on whether a client is “on” or “off-track” for a successful treatment outcome. When both absolute and progress alerts are applied across multiple administrations of an outcome instrument, they can be used to identify clients who are improving as expected, neither improving nor deteriorating, or deteriorating and at risk for treatment failure (Finch, Lambert, & Schaajle,

2001). The use of absolute and progress alerts is defined by five colors in the OQ-Analyst system:

- White alerts: The client is functioning in the normal range. Termination ought to be considered.
- Green alerts: The rate of change the client is making is in the adequate range. No change in treatment plan is recommended.
- Blue Alerts: The rate of change the client is making is unusually rapid, and the client would be expected to maintain treatment gains. Termination should be considered.
- Yellow alerts: The rate of change the client is making is less than adequate. The clinician should consider altering the treatment plan by intensifying treatment, shifting intervention strategies, and monitoring progress especially carefully.
- Red alerts: The client is not making the expected level of progress. This client is most at risk for early drop-out or treatment failure. Steps should be taken to carefully decide a new course of action (Spielman, Masters, & Lambert, 2006; Whipple et al., 2003).

Effects of Using Outcome-Based Feedback

Lambert and colleagues have been the pioneers of testing the effects of progress feedback in nine randomized clinical trials (RCTs) that began in the late 1990's. They have shown through both meta-analyses and individual randomized clinical trials (Berking, Orth, & Lutz, 2006; Harmon et al., 2007; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert et al., 2002; Shimokawa, Lambert, & Smart, 2010; Simon, Lambert, Harris, Busath, & Vazquez, 2012) that providing measure-based outcome and process feedback to individual therapists helps improve client outcomes, especially when the client is at risk for treatment failure. They defined "at risk" clients by using algorithms that were based upon optimal recovery curves producing the

previously explained color-coded alerts. The most significant finding from the nine RCTs was that progress feedback improved outcomes and reduced treatment failures (i.e., defined by a client leaving treatment with more distress than when they began treatment) when compared to clients where progress feedback was withheld from the same therapist. These results are significant because the individual literature has shown that therapists are poor predictors of eventual client outcome (Burns & Auerbach, 1996; Hannan et al., 2005), thereby providing an empirical rationale for the use of outcome-based feedback. This is also significant clinically as it suggests that when therapists use measure-based feedback systems they can potentially see improved psychosocial functioning as compared to relying on clinical judgment alone.

Adding Process Measures to the Feedback Delivery

Though not as vast, there is a growing body of literature that suggests that adding feedback about the therapeutic process can also be a useful contributor to client's eventual outcome. In these studies, measures are used to assess therapeutic process domains that have been empirically linked to successful outcome. More specifically, these measures assess the quality of the therapeutic alliance, a client's motivation and readiness for change as well as their perceived social support. These assessments, often called clinical support tools (CSTs) or therapeutic process measures, provide specific "alerts" that inform the clinician about potential problems using rationally-derived absolute alerts. These alerts show when treatment is not progressing as expected, and may help the clinician to consider one or more variables that predict treatment success (i.e., alliance, motivation, or social support). The OQ-Analyst also provides the clinician in-depth suggestions about actions that may help the client in the next session.

A number of studies (Harmon et al., 2007; Hawkins et al., 2004; Simon, Lambert, Harris, Busath, & Vazquez, 2012; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Whipple et al.,

2003) have empirically tested the effect of utilizing clinical support tools for clients who are at-risk for treatment failure. For instance, Whipple and colleagues (2003) added CSTs to the aforementioned outcome progress feedback (change, absolute, and progress feedback) of 48 therapists who treated nearly 500 clients in a university counseling center. If the therapist received notification that the client was not progressing as expected on the OQ (i.e., receiving “yellow” or “red” alerts), three CST assessment instruments were administered. More specifically, assessments were conducted on three therapeutic domains (therapeutic alliance, motivation for change and social support) that might explain possible reasons for why the client might not be making the expected improvement in treatment. The results of this study empirically demonstrated that adding CST feedback (i.e., absolute alerts) for clients in danger of treatment failure resulted in longer treatment and better eventual outcomes when compared to individuals receiving treatment-as-usual (no feedback) or only outcome feedback. These results were replicated twice with similar results (Hawkins et al., 2004; Slade et al., 2008) and once where mixed findings were produced (Harmon et al., 2007). Collectively, the progress and CST feedback findings are encouraging for the group treatment regimen. Indeed, at face value, using outcome and progress feedback seems more essential in group than in individual treatment given the sheer number of patients that are being tracked as well as the fact that group process is one of the main drivers of individual patient change in group treatment (Yalom & Lescsz, 2005).

Outcome/Progress Feedback in Group Treatments

While the research on measure-based feedback for individual therapy is robust, there is little research testing the effect of outcome/progress or within-group-process feedback in group psychotherapy. Over the last two decades Burlingame and colleagues have made a case for outcomes in group therapy being equivalent to outcomes in individual therapy (Burlingame,

MacKenzie & Strauss, 2004; Burlingame, Strauss & Joyce, 2013; McRoberts, Burlingame & Hoag, 1998). More specifically, they have shown in both narrative and meta-analytic reviews that when the two formats are compared under equivalent conditions (randomized clinical trials) that statistically equivalent outcomes result. More recently, this team of researchers compared over 13,000 clients using change trajectories on the OQ-45 for clients who received either group or individual treatment in naturalistic settings; results demonstrated that outcomes were statistically equivalent (Burlingame, Gleave, Erekson, Nelson, Olson, Thayer, & Beecher, 2012). Combining the results of this study with past comparative RCTs provides ample rationale for empirically examining the OQ progress feedback system in a group format (Burlingame & Beecher, 2012).

I could only locate a single group treatment study that lends support to the assertion that progress feedback in group treatment should yield commensurate results to those found in the individual treatment literature. Chapman et al. (2012) studied group leaders in both state hospital and university counseling center settings who were asked to estimate whether members were improving, deteriorating or staying the same using OQ total distress scores at three points in treatment (3rd, 6th & 9th sessions) compared to baseline assessment of distress on the same measure. This research was a direct replication of past individual treatment research that had shown that individual therapists were unable to accurately predict a client's outcome status (Hannan, et al., 2005). Across the ten study groups ten different group members were classified as having deteriorated per scores on the OQ at the end of the group. However, group leaders were unable to accurately predict the outcome status of any of these clients. As in the Hannan et al. (2005) study, group leaders predicted more improvement in members than actually occurred. These results suggest that like individual therapists, group leaders in both inpatient and

university settings display a positive bias when asked about their clients' potential for improvement and are unable to accurately predict the outcome status of their group members; Given this research and the previously described comparative outcome research, Burlingame, Woodland, & Whitcomb (2014) proposed that progress feedback to group leaders should yield an effect that is commensurate to the effect found in individual therapy. The details of their predictions are provided in a subsequent section where I describe their randomized clinical trial that I'm drawing upon for the current study.

Feedback on Therapeutic Processes in Group Treatment

There has been a parallel emphasis in the group literature on the use of within-therapy process measures which has interestingly produced more group research than progress feedback. The therapeutic relationship is a harbinger of final outcome in group process because past studies have shown that it predicts the highest rates of patient improvement and lower drop-out rates than any other process variable studied (Burlingame, Fuhriman, & Johnson, 2002). For instance, Burlingame and colleagues' (2011) meta-analysis assessed the connection between outcome and the therapeutic relationship in group treatment across 40 studies. The authors identified nine instruments that were used at least twice in the group literature, concluding that 43% of the studies using these measures found positive relationships with outcomes ($r = 0.25, p < .05, SE = .04$), constituting a moderate effect size. Based on these and other findings (cf. Burlingame et al., 2004, 2013), the therapeutic relationship is arguably the process variable with the most empirical support in group treatment, and likely the most viable on which to provide therapists with feedback. However, a vexing issue is the large number of measures used to assess the therapeutic relationship in group treatment. Many group researchers across many settings (Evans & Dion, 1991; Gully, Devine, & Whitney, 1995; Fuhriman, Drescher, Hanson, Henrie, &

Rybicki, 1986; Kivlighan, Multon, & Brossart, 1996; Moos & Humphrey, 1974; Mullen & Copper, 1994; Silbergeld, Koenig, Manderscheid, Meeker, & Hornung, 1975) have tried their hand at defining and measuring the therapeutic relationship in the group, resulting in a general confusion as to how to best define this multidimensional relational construct.

In the late 1990's the leadership of the American Group Psychotherapy Association (AGPA) recognized this problem and created a taskforce to update its CORE battery of recommended outcome and process instruments. The taskforce was composed of noted group researchers and clinicians from North America and Europe. Their primary task was to identify and recommend the best empirically supported outcome and process instruments for group treatments. The work product—CORE-R (Burlingame et al., 2006; Strauss, Burlingame, & Bormann, 2008)—included measures tapping four key group relationship constructs (cohesion, group climate, therapist empathy, and working alliance). During the multi-year tenure of the taskforce, a training program for group clinicians was developed (e.g., MacKenzie, Burlingame & Fuhriman, 1999) to test the viability of the proposed measures in group practice. Indeed, this course continues to be offered each year under the sponsorship of the Group Registry (the credentialing arm of AGPA). The most consistent and strongest feedback from group clinicians was that the multiple measures being recommended were too lengthy for clinical practice. This information led to the creation of a subset of taskforce members who undertook the task of determining if there was an empirically viable reduction of the therapeutic relationship measures recommended in the CORE-R.

The first study that attempted to reduce the four CORE-R relationship measures was led by Johnson, Burlingame, Olsen, Davies, & Gleave (2005), who administered the four relationship measures to 662 group members from 111 clinical and non-clinical groups. Using

multilevel modeling, they proposed a three-factor latent structure (positive bond, positive work, negative relationship) that explained significant proportions of variance for all four constructs across both clinical and nonclinical samples (66% and 59%, respectively). These promising results were then replicated on 67 hospital-based groups in Germany and Switzerland, further supporting the three factor model, and also providing support for use of three common structural facets of relationship in group treatment: member-member, member-leader, and member-group (Bormann & Strauss, 2007). The model was further replicated in a Norwegian outpatient sample of short and long-term, psychodynamically-oriented groups (Bakali, Baldwin, & Lorentzen, 2009). Archival data reduction analyses led to a 30-item instrument called the Group Questionnaire (GQ; Krogel et al., 2013. See Table 1. and Table 2 for psychometric data) which was later replicated in hospital-based groups in Germany (Bormann, Burlingame & Strauss, 2011) and US university counseling centers (Thayer, 2012). In addition to the three-factor structure of the GQ being replicated, two criterion validity studies show high criterion-related validity with the original instruments from which it was drawn (Thayer, 2012) as well as separate instruments assessing the same constructs (Bormann et al., 2011). It has also been hypothesized that the GQ maps well on to specific leader interventions like group structuring, facilitating verbal interaction, and regulating the emotional climate of the group (Chapman et al., 2012; Burlingame et al., 2011). The aggregate of these results offers a solution to the current multitude of therapeutic relationship measures found in the literature and also offers a basis for using measure-based feedback in group psychotherapy.

Overview of the Parent Randomized Clinical Trial

As mentioned in my introduction, I relied upon data collected from a randomized clinical trial conducted by Burlingame and Beecher (2012) tested the effects of feedback from measures

of outcome and process in group treatments. The study was conducted at three Utah university counseling centers; Brigham Young University, Southern Utah University and Utah State University. Since my study is limited by the data collected in this RCT I will provide a brief overview of the study, hereafter referred to as the parent RCT study.

Rationale and Hypotheses

The parent RCT was a replication and extension of the individual and group RCTs reviewed above. More specifically, the parent RCT experimentally manipulated both progress and process feedback using three arms. The first arm was an archival no-feedback condition drawn directly from Lambert's prior work where the OQ-45 was administered to clients before each session but no progress feedback was provided to the therapist (cf. Shimokawa, Lambert, & Smart, 2010). This arm reflects treatment as usual when therapists are not using progress feedback to guide treatment but progress is monitored. The second arm was prospective where group members completed the OQ-45 before each group session and the group leader was provided with a progress feedback report (Appendix A) created for all group members. This progress feedback is identical to the Lambert and colleague RCTs described above (i.e., change, absolute, and progress alerts) with one important exception (Figure 1). The alerts generated by the OQ-A software program were collated into a single clinician GQ/OQ group report for all group members. Having OQ feedback on a single page eliminated the multiple pages of reports produced by the OQ-software for each group member (Appendix B).

Finally, the third arm tested the effect of combining both progress and process feedback using the OQ-45 and GQ clinical reports generated by the OQ-A software. This manipulation went beyond previous -process RCTs by developing and testing three types of alerts (change, absolute, and progress) for the GQ (Figure 1). Building upon the Chapman et al. (2012) findings

as well as several years of clinical use and training with the GQ, Burlingame, Woodland, & Whitcomb (2014) developed change, absolute, and progress alerts for the three GQ subscales. This is a significant departure from the single rational alert (absolute) that Lambert and colleagues tested in their RCTs. However, it reflected the type of feedback requested by group clinicians trained on the GQ by Burlingame and colleagues at AGPA for the past several years. The report that clinicians received, therefore, was more complex, with subscale change and absolute alerts appearing on the first page of the OQ/GQ clinician report while all process alerts were graphically depicted for all group members on the second page (Appendix A). Finally, item level absolute alerts appeared on later pages.

Table 1

GQ Subscale Internal Consistency Reliability (Cronbach's Alpha)

Subscale	Overall	Member-Member	Member-Leader	Member-Group
Positive Bonding Relationship	.90 (.97)	.82	.83	.88
Positive Working Relationship	.91 (.95)	.87	.86	–
Negative Relationship	.79 (.98)	.61	.66	.76

Note. $N = 290$. Values in parentheses represent adjusted reliability coefficients calculated using the Ghiselli et al. (1981) formula. Values necessary for the computation of these adjusted coefficients were taken from Krogel, Beecher, Presnell, Burlingame, & Simonsen (2009).

Study	Outcome Alerts	Clinical Support Alerts
Lambert research based upon individual treatment	<p>Administered before each session</p> <p>Change alerts reflecting reliable change in score from intake—RCI show whether client has a reliable improvement or deterioration in total score since intake</p> <p>Absolute Alerts based on passing clinical significance cutoff score into clinical or nonclinical range of distress</p> <p>Progress Alerts where client change is compared to normative change trajectory</p> <ul style="list-style-type: none"> • On track/normal=Green, White • Sudden gain=Blue • Off track=Red, Yellow 	<p>Administered when therapist receives an off-target progress alert</p> <p>Change alerts not relevant since scale administered only when progress alerts issued</p> <p>Absolute Alerts at subscale and item level based upon rational cut scores</p> <p>Progress Alerts not relevant since scale administered only when progress alerts issued</p>
Burlingame & Beecher RCT based upon group treatment	Identical to above	<p>Administered at the end of every session</p> <p>Change alerts reflecting reliable change in scores since last session—RCI shows whether client has a reliable improvement or deterioration in subscale score since last session</p> <p>Absolute alerts at subscale and item level based upon rational cut scores</p> <p>Progress alerts—descriptive only, client subscale scores plotted with all group members</p>

Figure 1. Comparison of parent RCT alerts with past individual therapy research.

Table 2

Criterion-Related Correlation Coefficients for GQ Subscales

GQ Subscale	Overall	Member-Member	Member-Leader	Member-Group
Positive Working Relationship				
WAI: Task	.79**/.77**	.74**/.73**	.78**/.75**	—
WAI: Goal	.71**/.71**	.67**/.66**	.70**/.70**	—
Positive Bonding Relationship				
WAI: Bond	.76**/.76**	.74**/.71**	.72**/.68**	—
ES: Positive	.77**/.76**	.72**/.70**	.70**/.67**	—
TFI: Cohesion	.81**/.80**	—	—	.72**/.76**
GCQ: Engaged	.56**/.53**	—	—	.54**/.58**
Negative Relationship				
GCQ: Conflict	.67**/.65**	—	—	.78**/.74**
ES: Negative	.66**/.64**	.66**/.64**	.69**/.62**	—

Note. Pearson/Spearman. $N = 290$. The “—” indicates that this value was not calculated because it was not applicable.

** $p < .01$.

The parent RCT was designed to have each group leader simultaneously run two groups—one where the leader received GQ/OQ feedback and the second where only OQ feedback would be received. Each group leader committed to run a minimum of 4 groups to provide a test of both leader and group effects. The study posed three primary hypotheses regarding groups where the leader received GQ/OQ feedback versus no feedback:

1. Leaders who receive GQ absolute and/or relative alerts and acted on these (assessed by weekly leader GQ reports) would have clients who show a quicker (i.e. slope) and larger

return (difference score) to subscale values that fall within the normative range on alerted subscale when compared to leaders who did not receive GQ alerts.

2. Leaders who receive GQ absolute and/or relative alerts and acted on these (assessed by weekly leader GQ reports) would have clients posting higher levels of group attendance (i.e., fewer dropouts).
3. Leaders who receive GQ absolute and/or relative alerts and acted on these (assessed by weekly leader GQ reports) would have clients who reported reduced symptom distress on the OQ-45.

In addition to these primary GQ hypotheses, the parent RCT sought to replicate progress feedback using the OQ-45 using the third archival arm. It is important to note that this hypothesis compared the second and third arms. More specifically, the hypothesis was:

4. Leaders who receive OQ alerts and acted on these (assessed by weekly leader GQ reports) would have clients who would show fewer treatment failures and reduced symptom distress than clients whose therapist received no OQ feedback alerts.

Participants

The parent RCT included individuals participating in group therapy at the student counseling centers at Brigham Young University, Southern Utah University, and Utah State University. The study groups were co-led by one licensed psychologist and typically one trainee or intern. Group therapists at college counseling centers generally espouse a number of theoretical orientations, including cognitive-behavioral, behavioral, rational-emotive, process-experiential, humanistic, interpersonal, existential, and psychodynamic. The average group size ranged from 6 to 10 members, and each group had members with presenting problems that are representative of college counseling center populations. Group leaders committed to participate

for two semesters over the course of 30 months and run a minimum of two pairs of GQ/OQ feedback/no GQ/OQ feedback groups to statistically model group and leader effects. As this is a naturalistic study, the groups they ran were a part of their normal group caseloads. Groups typically lasted one semester and were reconstituted at the beginning of the next semester. Clients in the fall semester groups were allowed to request additional treatment. All who did were accommodated but only those in a no-feedback group were eligible to be assigned to a study group for a second (winter) semester. Thus, it was possible that a client might participate in the study for two semesters, but this was the exception rather than the rule.

Those who opted in to the study were required to attend group therapy as primary treatment modality (i.e., receiving no more than one individual session for every three group sessions) to control for possible negative effects of group engagement as reported by Davies, Burlingame, Johnson, Gleave, & Barlow (2008). Other inclusion and exclusion criteria included:

Inclusion criteria:

- Willingness to commit to at least 4 sessions of group treatment
- Willingness to complete GQ and OQ on weekly basis, even if their group leaders do not receive GQ/OQ feedback
- Willingness to have group be their primary mode of treatment. This is to insure they are committed to the group as a primary vehicle for change

Exclusion criteria:

- Participation in a study group that received GQ/OQ feedback in an earlier semester to avoid the possibility of having a group member who received feedback one semester be placed in a no-feedback a second semester and thereby create a state of demoralization

- Participation in a group where the majority of members want to carry over to the next semester--we want less than 50% of members in a new study group to be carryover.
- No email address

All research participants were recruited or referred at intake at respective counseling centers. Group leaders were recruited through research meetings, and announcements made at faculty meetings. All group members received gift cards as compensation for research participation.

Procedures

Upon requesting services, potential clients complete an OQ-45 via the counseling center's online intake program. This program also included basic demographic information salient to therapeutic work, and this information was saved in an electronic database. Upon completing the OQ-45 potential clients were signed up for study groups as a part of normal clinic scheduling. Upon starting in a study group, members were explained the nature of the current study, including the benefits of group therapy, benefits of study participation, and potential risks. Potential therapy participants were also informed of monetary incentives for group participation (group leaders did not receive any compensation aside from the benefit of weekly feedback reports for their group assigned to the feedback condition). This included receiving \$10 upon consenting to participate, and receiving an additional \$5 for every set of weekly OQ and GQ they completed. At the group's conclusion clients received compensation for their contribution to the study in proportion to the number of OQ/GQ data sets they provided (e.g., if they provided 5 sessions of data they received \$25). For those who completed OQ/GQs for all possible sessions attended, a bonus of \$20 was awarded. Compensation was provided in cash.

Clients who showed interest in participation were explained study procedures in-depth by their group leader. The group leader followed a script (Appendix C) and determined eligibility based on the above inclusion and exclusion criteria and obtaining informed consent. Particular emphasis was given to completion of an OQ before the beginning of group and completion of a GQ at the end of each group. Members were informed that their group leader would use the OQ irrespective of group assignment to guide treatment based on their progress. The preferred method of completing the GQ was identified for each client by the group leader (CCC tablet, paper or online) clients were provided with a sample copy of the GQ.

After the group began, new members were assigned for up to four weeks after the initial session. At the end of each group session, the group leader reminded members to complete their GQ for that session. E-mail reminders were sent to those members who failed to complete a GQ one- and five-days after each group session.

Group leaders were oriented to the study before they ran their groups. The experimental manipulation was a double-sided GQ/OQ feedback report provided to the group leaders called the "Weekly GQ/OQ feedback Report" (Appendix A). The first page of this report provided the group leader with absolute alerts where all group member GQ scores are classified into one of three categories. Heretofore, this GQ information has not been available to clinicians using the OQ-A (see Appendix B for an example of the current GQ report). The second alert on page 1 of the "Weekly GQ/OQ feedback Report" reflects change alerts for each GQ subscale. Though this type of alert has been available on the OQ-Analyst for some time, this information has not as yet been available for clinicians in group therapy. Members who have shown statistically significant non-therapeutic or therapeutic change since the last session were flagged. Change alerts only identified group members who had significant subscale score change since their last session, in

this case clients who deteriorated or improved more than one RCI unit (see Figure 1). Group leaders were given group progress alerts, though there does not yet exist an algorithm that predicts whether a client is on track or not on track for the GQ. In lieu of this technology clinicians were provided with a graph of each client's progress for each subscale across the three GQ subscales (see Appendix B). Group leaders were also given subscale- and item- specific data for those clients who alert on specific subscales.

Current Study

The parent RCT proposed several hypotheses that pivot upon how a group leader uses GQ/OQ feedback. For example, the primary GQ hypotheses state that “Leaders who receive GQ absolute and/or relative alerts and **acted on these** (assessed by weekly leader GQ reports)...” More specifically, Burlingame and Beecher (2012) proposed that leaders who receive GQ/OQ feedback and then deliberately intervene in their groups using this feedback would be more likely to have clients who would meet the predictions of the hypotheses when compared to group leaders who either fail to review the GQ/OQ feedback or choose to ignore or disagree with it.

Indeed, powering the effect of GQ/OQ feedback must take into consideration the typical range and frequency of how group leaders act upon GQ/OQ feedback. For instance, if more group leaders fail to use GQ/OQ feedback, a larger number of groups would be needed to detect an effect (if one is present) and of course the opposite would be true. This may also be true, for example, if group leaders do not value the feedback they receive, even if they use it in some fashion; thus, describing and operationalizing what it means to “act upon” feedback is of central importance to understanding any result/non-result of GQ/OQ feedback, potentially ruling in or out the possibility of this “acting upon” as a moderator of feedback efficacy.

Method

One objective of the parent RCT was to collect open ended responses from each group leader for each session of their group that described how they used the feedback they received. A space for open ended responses was provided at the bottom on the first page of the GQ/OQ clinician report (Appendix B). Operationally, each group leader either printed this report and turned in handwritten comments or did the same in electronic form and emailed them to the researchers. These responses were then organized in an Excel spreadsheet for future qualitative analysis. Each leader who participated in the study was also administered a semi-structured debrief interview (Appendix D) at the end of the study period. The group leaders' responses were then transcribed.

In my study there were two main units of observation of feedback implementation: 1) each individual session and 2) the group episode over the course of the semester. Both manifest and latent content from the individual session and entire group episode were analyzed using combined methodology from both qualitative content analysis (QCA) and grounded theory (Krippendorff, 2012; Schreier, 2012; Strauss & Corbin, 1990). While both methods are similar and even at times interchangeable, the distinction made between the two methods is that grounded theory is often considered data-driven, while QCA is often considered to be concept-driven. Our content analysis established a coding framework using both session leader slips and debrief interview transcripts to operationalize how leaders "acted upon" the GQ/OQ feedback but it also had the flexibility to incorporate additional themes that emerged. Using both sources of leader input allowed for description of the themes within the data, as well as analyzing possible connections or relationships between themes, categories, subcategories, and dimensions.

All analyses were performed by two trained raters (Sean Woodland and Kait Whitcomb) whose ratings were monitored by an auditor (Gary M. Burlingame, PhD).

All content analytic procedures responded to the research question, “What does it mean to ‘act upon’ GQ/OQ feedback?” The two a priori categories hypothesized to appear in the data were leader use of feedback and leader value of feedback. These main categories were obtained after a pilot thematic analysis of a small portion of weekly leader responses ($n = 52$; see Appendix E). Informed by these two main categories, the content analysis then began with two main steps: unitizing and coding.

Unitizing

Many qualitative researchers have stressed the importance of dividing the content to be analyzed into codable units before performing the coding (Schrieier, 2012; Strauss & Corbin, 1990). Unitizing adds important structure to the material and enables easy indexing in preparation for developing a coding framework. Raters in the current study based their unitizing decisions on the rules established for unitizing by Stinchfield and Burlingame (1991), as well as a list of more specific rules adapted for the current qualitative dataset (see Appendix F). For example, in the previous set of rules it was established that “when in doubt, do not make a separate unit.” For the current study this rule was clarified further by introducing new terms such as simple processes/reflections, quick responses, and travelogues. For example, when the speaker appends a statement with a simple reflection (e.g., “I wish we had the GQ in there. That was one thing” , where “That was one thing” is considered a simple reflection) it is included with the previous statement as one unit because it does not provide additional meaning. Quick responses occurred when the group leader responded to the other person with one or two words

(e.g., stating “Yes” before going on to provide more detail). Travelogues occurred when group leaders gave a session narrative.

To begin unitizing, raters selected a random 20% of the data representative of each the leader slips and debrief interview transcripts; the purpose of this initial audit was for the raters to become familiar with the data in preparation for establishing a coding framework. After the trial audit, raters independently rated portions of the data in a graduated, iterative format. After each section was independently unitized, raters discussed agreements/disagreements, taking any outstanding disagreements to the auditor to be discussed and resolved. This process was undertaken with the goal of achieving stability between raters and with the established unitizing rulebook.

Open Coding of Qualitative Content

Phase I coding. The next step of the content analysis was to formally code the unitized content. Coding began with a pilot “trial coding” in which the raters independently coded another representative 20% of the data, paying close attention to the main categories of utilization and value, but also identifying additional subcategories and subsubcategories as they emerged. The purpose of the trial coding was again to provide initial exposure to the feel of the material, in this case for an understanding of what the unitized material might represent. After trial coding was completed, the raters began “Phase I” coding, independently applying descriptors to each unit. The raters then again discussed each other’s findings, making revisions to the descriptors to achieve consensual agreement. In Phase I the raters also began to rule out units that did not have a recognizable connection to use or value of the GQ/OQ feedback. For example, in Phase I clarification utterances such as “Mmhmm”, “Did I get that right?”, or “My

co-leader's name was David" were ruled out (For a full list of inclusion/exclusions rules, see Appendix F).

Phase II coding. After preliminary descriptors were established, raters performed "Phase II" coding in which they re-categorized the units described in Phase I. This step represented the beginning of a pattern in which the results alternated between becoming increasingly broad and increasingly specific, a phenomenon common and perhaps even essential to qualitative content analysis (Schreier, 2012). Phase II represented a broadening of the data as specific descriptors from Phase I were grouped under common labels; these labels were the precursors to the identification of main categories, or the main themes of the content analysis.

Phase III coding. In "Phase III" coding the raters returned to specificity. Using the units grouped under broad categories, subcategories were formed depending on the conceptual differences between units within a given category established in Phase II. As this was a third pass through the data, raters were familiar enough with general concepts to also transfer units from one category to another based on better "fit;" this also included excluding more units that in earlier phases could not be ruled out as unrelated to feedback use or value, but were deemed irrelevant in Phase III after further consideration and upon mutual agreement between raters.

Phase IV coding. In the final phase of coding, raters converted the qualitative categories into quantitative codes for eventual use in testing the hypothesis that the "acted upon" variable mediates the relationship between the receipt of GQ/OQ feedback and subsequent GQ scores (see Bryman, 2006, Jick, 1979, and Sandelowski, 2000 for descriptions of mixing qualitative and quantitative methods). Once a consensus was reached for main use and value categories and subcategories, the information was entered into an electronic database, where each unit was assigned a number for every applicable category and subcategory.

Results

The study sample included thirteen therapy groups across two academic semesters. The groups were run at three university counseling centers: Brigham Young University (7), Utah State University (4), and Southern Utah University (2). Of these groups, twelve were semester-long, while one was considered “longitudinal,” continuing after the semester break. Twelve of the thirteen were process-oriented groups, with one focused on psychoeducation for generalized anxiety (see Table 3). The mean number of sessions per group was 12.4 ($SD = 3.8$).

The thirteen feedback-only groups were led by eleven different leaders (two group leaders ran two groups apiece, while the remainder of the leaders only ran one group). Eight group leaders were male, and three were female. The mean age of the group leaders was 41.6 years old ($SD = 11.1$). Ten of the leaders identified as White/Caucasian, while one identified as Asian. Primary reported theoretical orientations included Humanistic/Existential (5), Interpersonal (1), Modern Gestalt (1), Psychodynamic (1), Integrative (1), Quantum (1), with one unreported. The mean number of years of group therapy experience per group leader was 12.3 ($SD = 11.9$) (see Table 4).

Group members included in the study ($n = 104$) were all students at the aforementioned universities, who had asked for services at their respective counseling centers. The average client age was 23.3 ($SD = 4.5$). Group members were predominantly female (65.1%) and single (62.3%). They either asked specifically for group-based services, or they were invited to attend group by their individual therapist. The most prevalent presenting problems were anxiety (23.6%), depression (18.9%), and relationship problems (18.9%). Most students were upperclassmen (54.7%) at the time. The vast majority of the students was Caucasian (70.8%), and affiliated with the Church of Jesus Christ of Latter-day Saints (61.3%).

Agreeing Upon Units for Rating

The first step before the content analysis was performed was achieving agreement about what a unit looked like. For the session-by-session leader slips, baseline independent agreement was 90%. Consensual agreement thereafter was 94%; that is, there was a 4% increase in agreement of leader slip units after raters consulted with one another. The final 6% was taken to the auditor to achieve 100% agreement. For the interview transcripts, initial independent agreement was lower than the leader slips, at 69.9%. This decrease in agreement was most attributable to a higher learning curve in unitizing verbal interview content when compared to the written content in the leader slips, which is typically more clear and succinct, and therefore easier to unitize. Raters then reached consensual agreement (97%) on transcript data after discussion. Outstanding disagreements were presented to the auditor to achieve 100% agreement.

Results of the Four Coding Phases

The four-phase content analysis was based upon two data sources: leader slips which were collected after each session and debrief interview transcripts that were conducted at the end of each semester-long group. In each phase raters passed through the data, developing an understanding of what it meant for group leaders to “act upon” GQ/OQ feedback. The original dataset of ratable units ($n = 481$ & $n = 701$ for leader slips and interview transcripts, respectively) became smaller after each coding phase, as shown in Table 5. This was a natural effect of going through the material and becoming more specific about the inclusion of units in the qualitative categories. As we understood better what each category entailed (i.e., as the categories fully emerged), more units were deemed unclassifiable for the purposes of the content analysis. There was a small increase in ratable units for the debrief interview transcripts after the

final audit of all units that were removed across the four coding phases. The final number of leader slip units was 341 and the final number of debrief interview transcript units was 390.

Table 3

Feedback Groups Included in the Content Analysis

Group Number*	Group Name	School	Orientation/Type	# of Sessions	Longitudinal?
2	General Process	BYU	Process	13	Yes
3	General Process	BYU	Process	13	No
44**	General Process	BYU	Process	12	No
5	General Process	BYU	Process	11	No
8	General Process	BYU	Process	11	No
10	GAP	BYU	Psychoed.	9	No
11	General Process	BYU	Process	9	No
13	Men's USO	SUU	Process	11	No
15	USO	SUU	Process	10	No
17	Interpersonal Process (IP)	USU	Process	12	No
19	Interpersonal Process (IP)	USU	Process	13	No
21	Interpersonal Process (IP)	USU	Process	13	No
23	Interpersonal Process (IP)	USU	Process	13	No

Note. GAP=Generalized Anxiety Psychoeducation; USO=Understanding Self and Others.

*Group numbers are not in order because they were taken from the larger RCT which contains both feedback and non-feedback groups

**This group was the continuation of group #4, which at the semester switched from the non-feedback to the feedback condition.

While it was much easier to arrive at a measure of agreement and disagreement for a unit of analysis, similar quantification for the agreement of the content of the analysis was difficult to arrive at because the fundamental goal was to iteratively achieve consensus between the two raters. Therefore, consensus was reached 100% of the time. However, of the 26 total content

categories that emerged from both the analysis of leader slips and interview transcripts, most categories had relatively high initial agreement between raters. The categories that generated

Table 4

Group Leaders who Provided Data for the Content Analysis

	Gender	Age	Identified Race/Ethnicity	Primary Theoretical Orientation	Years of Experience Running Groups
Leader #1	Male	34	Caucasian	Existential/Interpersonal	2
Leader #2	Male	62	Caucasian	Humanistic/Existential	37
Leader #3	Male	43	Caucasian	Existential/ACT	17
Leader #4	Male	36	Caucasian	Integrative	8
Leader #5	Female	33	Caucasian	Humanistic/Existential	6
Leader #6	Male	46	Caucasian	Modern Gestalt	4
Leader #7	Female	36	Caucasian	Psychodynamic/Interpersonal	7
Leader #8	Male	59	Caucasian	Quantum	30
Leader #9	Male	33	Caucasian	Humanistic/Existential	6
Leader #10	Female	38	Asian	Interpersonal	5
Leader #11	Male	30	Caucasian	Humanistic/Existential	3

more discussion before reaching 100% (about five or six categories) consensus portrayed more difficult or abstract concepts. This included difficulties with operationalizing categories and subcategories, finding appropriate cause-effect patterns to suggest inclusion or exclusion, and judging unit context for appropriate category inclusion or exclusion.

Table 5

Ratable Units Remaining After each Coding Phase

Data Type	Total	Phase I Coding	Phase II Coding	Phase III Coding	Audit of unclassified units
	Original Units				
Leader Slips	481	426	426	363	341
Debrief Interview Transcripts	701	537	406	386	390

Phase I-III Coding Results

As stated in the method section, the first three coding phases entailed a pattern of broadening and narrowing labels/potential categories to fit the content in preparation for the final product, which we have called the “Phase IV” coding results. Phase I for the leader slips and interview transcripts included all the units produced by the initial unitizing. Each unit was given a label to reflect how the rater believed it captured the “acted upon” construct. Raters also at this point decided if units did not fit the construct, and set them aside as unclassifiable. By the end of Phase I, raters had very long lists of labels/potential categories. There were 426 ratable leader slip units and 527 ratable interview transcript units at the end of Phase I coding.

Phase II coding involved grouping labels that were conceptually similar. The raters performed this step together, discussing the similarity and differences of the labels/potential categories. After Phase two there were 426 leader slip units and 406 interview transcript units. The non-uniformity in unit reduction across leader slip and interview transcripts was due to the fact that when raters analyzed the interview transcripts they already had an idea of what “acted

upon” looked like based upon their work in labeling leader slips. This resulted in a higher frequency of reductions for interview transcript units.

Phase III involved vetting existing categories and proposing subcategories. After this phase there were 363 leader slip units and 386 interview transcript units. In this phase there was a larger decrease in leader slip units. Nevertheless, Phase III ratings resulted in a decrease in units judged as ratable for both sources. The “finished” product for the content analysis of leader slips and interview transcripts was produced by Phase IV coding. This phase involved in-person vetting of both categories and subcategories. What follows is an in-depth description of these results, including language from actual units to illustrate categories and subcategories.

Phase IV Coding Results: Leader Slips

The final content analysis of the leader slip units ($n = 341$ across 116 leader slips) yielded 13 categories (see Table 6; categories are ordered by when they appeared in relation to receiving the feedback intervention). Of these 341 units, a very small fraction dealt specifically with how the group leader used the OQ feedback that was provided on the report, even though the both GQ and OQ alerts were provided. Accordingly, most of the content categories discussed below specifically address GQ/OQ feedback; comments about acting upon the OQ alone were infrequent. As noted above, each category in Table 6 was defined through rater consensus. To assist in greater understanding of each, specific definitions and coding rules can be found in Appendix G. When necessary, common content within a category was subdivided into subcategories to further define meaning.

Table 6

Leader Slip Categories Resulting from Content Analysis

Category	Freq.	% of total coded units
Review of Feedback	17	5
Reaction to Scores	77	22.6
Design Specific Intervention	20	5.9
Decision to Withhold Feedback	36	10.6
Explicit Feedback Use	41	12
Non-Explicit Feedback Use	20	5.9
Education About GQ	5	1.5
Downstream Effects	89	26.1
Attendance	8	2.3
Filling Out Measures	8	2.3
Group-Initiated Feedback Use	7	2
OQ Use	10	2.9
Looking Forward	3	.9
TOTAL	341	100

Review of feedback. The category “Review of Feedback” ($n = 17$) was created to represent the primary action of a group leader looking at the feedback in some way. Because a fraction of the units specified as “Review of Feedback” also made mention of a co-leader being involved in the review process, a subcategory “with co-leader” was also created (this process of subcategory creation was replicated for the majority of the main categories). For example, leader slip 112.2 states, “I reviewed this GQ/OQ feedback carefully by myself and with my co-therapist.” Other examples included leader slip 62.2 (“I glanced at it before the session, but didn’t spend a lot of time on it”), leader slip 94.1 (“I reviewed all of the data this week”), and leader slip 44.1

(“My co-leader and I reviewed the feedback prior to group”). All of these units show that upon receiving the feedback report, the group leader opened it, and visually scanned the data, without actually speaking to the content of the feedback report.

Reaction to scores. The second category that emerged from the content analysis was “Reaction to Scores ($n=77$)” capturing nearly a quarter of all ratable units. As the name suggests, often the group leaders reported going a step further than simply looking at the GQ/OQ feedback. Units coded as “Reaction to Scores” involved the group leader reporting the content, an interpretation, or a speculation stimulated by the results found on the feedback report. For example, leader slip 71.1 states, “Before the group I notice the lowered scores and wondered what had happened.”

The group leaders who reacted to scores also provided varying levels of added detail, which were coded into five subcategories; “with co-leader”, “direction of reaction”, “direction of change”, “reaction to GQ subscales,” and “reaction to GQ facets”. “Direction of reaction” refers to the group leader’s focus on whom the scores belong to; in particular, group leaders’ foci typically included a single group member, multiple members, or the group-as-whole. “Direction of change” captures the quality of those scores, namely positive change, negative change, no change, or mixed change. The “reaction to GQ subscales” subcategory represented group leader comments that specifically mention positive bond, positive work, or negative relationship subscales of the GQ. Similarly, the “reaction to GQ facets” represents leader comments regarding GQ scores on a member-member, member-leader, or member-group subscale.

At times, the leaders ($n = 20$) made process-like statements about GQ scores that either expressed surprise about a particular score, or a speculation as to why the scores occurred. Leader remarks such as these were coded in the surprise/speculation subcategory. To better understand these comments refer to an example from one leader slip that had three ratable units:

Looking at the GQ/OQ feedback it seemed that bond rose for a few members, but still went down for others. /Work also seemed to drop for several. / This feedback was surprising to both [my co-leader] and me.

The first phrase of this unit was coded as a basic reaction—“Looking at the GQ/OQ feedback” but since no co-leader is mentioned, this subcategory was not endorsed. The next phrase captures the focus (multiple members) leading to the use of the “direction of reaction” subcategory. This phrase also includes the direction of change (mixed change), and a reaction to GQ subscales (positive bond). In the second unit there is again no report of a co-leader, but we coded the direction of reaction (multiple members), direction of change (negative change), and one GQ subscale (positive work) subcategories. In the third unit the leader reports the involvement of a co-leader, and mutual surprise on their part in response to the scores. Using context from the previous two units, this unit was coded under “mixed change.”

Design specific interventions. The third category “Design Specific Interventions” ($n = 18$) was used any time a group leader reported taking actions that were meant to prepare for use or implementation of information from the GQ/OQ feedback report in the next session.

Subcategories included involvement of a co-leader, direction of planned intervention, direction of change upon which the intervention was based, and mention of GQ subscales and/or facets.

For example, the following leader slip had two units:

I wanted group members to experience increased positive bond and positive work and less negative relationship./I spent some time online and in my Yalom group text reviewing principles around helping groups resolve conflict.

In the first unit the group leader expresses a desire for improved scores but doesn't explicitly state a plan for intervention. However, in the next unit, they state specific actions with an intended outcome. In the content analysis both units were coded as being intended for the group-as-whole (specific members could not be inferred), and as reflecting direction of change, with all three GQ subscales mentioned.

Of the 18 units endorsed as Design Specific Interventions, eight belonged to one group leader (Leader #7). While this does not negate the validity of this category, it is interesting to note that nearly half of the units categorized into this category of “acted upon” came from a specific group leader. Interestingly, leader #7 reported a great deal of planning relative to other leaders, but did not report any interventions inside the group that could be attributable to GQ/OQ feedback.

Decision to withhold feedback. The next category was labeled “Decision to Withhold Feedback” ($n = 36$) which was invoked any time the leader reported choosing not to use the data from the feedback report. This was done both in and out of the presence of a co-leader. Our bias as raters leaned toward us wanting the group leaders to use the feedback, and yet understanding that sometimes clinical judgment trumped measure-based data intrigued us. Indeed, this was the fourth most frequently used category of acting upon feedback. We were interested in knowing why exactly the group leaders chose to withhold feedback which led to several subcategories that included (1) withholding due to lack of time to implement feedback, (2) insufficient information to create a meaningful intervention, (3) group member preference to not hear about the feedback, (4) decision to observe and be aware of alerting members, (5) extenuating circumstances, and (6) the fact that group members were showing only positive scores or improvement on the GQ—i.e., there was no need for an intervention. For example, one leader stated “We agreed to not take any

action, other than paying attention to how C was doing.” In this case the unit was coded as involving a co-leader with “observing/awareness” as the reason for withholding.

Explicit feedback use. This next category was “Explicit Feedback Use” ($n = 41$) which captured any time a group leader brought up scores, trends, or concepts directly from the GQ/OQ feedback report. The subcategories for Explicit Feedback Use paralleled those found in Reaction to Scores and included co-leader involvement, direction of intervention, direction of change upon which the intervention was based, and inclusion of GQ subscales and GQ facets. This category is the third most frequently endorsed in the leader slips.

As with all categories, units endorsed as Explicit Feedback Use employed both manifest and latent content. As an example that employs both manifest and latent content can be found in the following leader slip text:

We were still surprised by the scores, especially the member-to-member work alert./

Were not concerned about dropout, but felt it may be worth raising this info in group for clarification. /We may also wonder why others in the group didn't take action. /We raised the question a/ him (sic) and the group and it seemed to work out well.

In the excerpt above, the final unit (“We raised the question...”) is the unit endorsed as Explicit Feedback Use. The context created by the preceding units provides a more comprehensive story and demonstrates how subcategory coding was undertaken. In the first unit, we noted that the GQ was mentioned for both subscales (positive work) and facets (member-member); we also inferred that the intervention in the final unit reflects negative change because the leader is responding to an alert that seems surprising. Returning to the last unit, we also note that there was a co-leader involved (referring to “we”), and that the intervention was intended to affect an individual member (referring to “him”).

Non-explicit feedback use. The next category emerged when we noted group leaders using the GQ in the group but without explicitly mentioning results from the feedback report to group members. Hence, the creation of the category “Non-Explicit Feedback Use” ($n = 20$). Units rated under this category included the leader using the GQ to inform their notions about a client, to inform group process, or to inform treatment goals, again without making specific mention of the measures or the feedback. Non-Explicit Feedback Use also includes those units that could have been endorsed as Explicit Feedback Use, but there was not enough supporting context to rule it as explicit use. The example below illustrates a common pattern of non-explicit use:

For the first time in this group, we focused attention for a significant amount of the session on CP./CP reported during the session that her individual therapist had challenged her this week to share more of herself in group, so the timing was fitting./ Group leaders challenged CP to define and address her goals for her participation in group, and to help us and the other group members understand what she would like to get out of group.

In this sequence, both the first and last unit was identified as Non-Explicit Feedback Use. However, the middle unit could be classified as in-group use but we cannot explicitly infer that the group leader talked about positive work. Rather, we concluded that the group leader was informed by the GQ scores and used them to speak in group about constructs related to positive work (e.g., group goals). As with preceding categories (e.g. Explicit Use) similar subcategories emerged including the involvement of a co-leader (“Group leaders challenged...”) and the aforementioned direction of intervention (CP) and change (negative) as well as content related to specific GQ subscales (Positive Work).

Education about GQ. Aside from referring to or being informed by the GQ scores themselves, group leaders also intervened by talking to the group members generally about the GQ, its subscales, and the study. When leaders described these activities we categorized units in the “Education about GQ. ($n = 5$)” category. For example, one leader stated “We explained that the questions on the GQ load into 3 factors and described a little bit about what Positive Bond, Positive Work, and Negative Relationship mean.” Unlike other categories, no subcategories emerged for Education about GQ.

Downstream effects. The most frequently occurring phenomenon (over 25%) was the propensity for group leaders to report what happened after or as a result of previous GQ use. These effects were either directly or indirectly linked to an earlier point in time where the GQ was acted upon. We called this category “Downstream Effects” ($n = 89$) because often GQ use in the group seemed to set off a chain of events, often leading to resolutions and group cohesion, and sometimes leading to fractures in the group relationship. While these effects were not direct use of the GQ by the leader, they were certainly indicative of prior use. Also, they often provided an excellent context for understanding the group dynamic, and by extension understanding why the leader might use the GQ in subsequent sessions informed by the downstream effects. A good example is found in following leader slip:

Feedback today focused on one group member in particular- N, who [had] reported negative alerts in each category, and who has seemed to be quite detached from the rest of the group./The group agreed with her assessment that she does not feel connected to others,/and they invited her to play a more active role in the group, providing positive feedback for the times when she has contributed. /She appeared to be somewhat overwhelmed with the feedback.

The first unit of this segment (“Feedback today focused...”) constituted use of the GQ; the raters saw this as explicit use for this particular group leader. After that, Downstream Effects begin, starting with a group response to intervention, and followed by a response by the individual member whose alerts created the original leader intervention. The last three units were coded as dealing with an individual member who experienced negative change, with each subscale (“...alerts in each category...”) endorsed.

Another aspect about Downstream Effects that was observable was how immediate the effects were to the preceding intervention; we created three options for the “immediacy of effect” subcategory: immediate, secondary, and ultimate. Using the preceding leader example, the first unit in the sequence was classified as an immediate effect, while the final two were classified as secondary effects. Ultimate effect were coded when a group leader made a statement reflecting a final conclusion or effect. For example, raters coded ultimate effects when the leader described a group decision, or language providing closure or resolution to a previous alert. For example, one group leader reported making a prior intervention, and described how this intervention facilitated discussion about the current group dynamic in a subsequent session. The final unit in the sequence states, “Group leaders and members were validating of this,” referring to an individual member’s emotional expression in-group. This unit was judged as providing some resolution to a prior discussion, and therefore coded as an ultimate effect.

Attendance. On occasion the group leaders reflected upon the attendance of certain group members resulting in the creation of the “Attendance” category ($n = 8$). The majority of attendance units pertained to members who had alerted previously. Some of these alert cases may have otherwise been coded differently, had the group member actually come to group. For example, leader slip 21.8 states, “The group member whose scores were worse did not come to

the session”. It is possible in this example that if the group member had attended, the group leader would have used the feedback in another way. In another instance, the leader expressed surprise in response to the attendance of a member alerting: “One of the clients who triggered an alert did not return this week, although she had indicated that she was excited for group.” Given earlier work that suggested that member drop out was associated with a preceding alert, these leader comments seem understandable and reflect respect for their potential value.

Filling out measures. A small portion of the leader comments ($n = 8$) reflected their reaction to group members’ response to the GQ and/or the study. For example, one leader states “Noted at start of session that 5/10 group members completed the OQ/GQ and 5 had not.” There were also a couple units that addressed completion of measures and actions taken by the group leader. For example, one the leader stated, “Encouraged all group members to follow through,” indicating that the group leader brought up the GQ in the group and specifically invited members to consistently the tool.

Group-initiated feedback use. A fraction of leader comments ($n = 7$) involved group members reactions to the GQ. While this category does not directly address how the group leader “Acted Upon,” GQ/OQ feedback it was interesting to note how the group members reacted to reporting their group experience with the GQ. While some reported curiosity as to what the scores meant, some members reported feelings of irritation regarding a measure that evaluated their experience and group leaders who brought the scores into the group.

OQ use. As noted in the method section, both conditions in the RCT were offered OQ feedback. However, leaders infrequently talked about how they used the OQ, resulting in a separate category that pertained to the OQ alone. Ten units were classified under this category. One example of a unit rated in this category was “I looked at the OQ, but didn’t really do

anything with it other than to note that one of our members is feeling significant distress.”

Another example was “Curious about L (no OQ data for this week).”

Looking forward. Finally, three units in the leader slip data were classified as “Looking Forward.” The content that units in this category had in common was the group leader anticipating future GQ results. For example, “I’m looking forward to seeing next week’s feedback & hope that our group members are feeling more positively about their group work.” Another unit stated, “I will be interested to see this week’s scores.” These always appeared either near or at the end of the weekly submitted leader slip, providing finality the group leader’s entry.

Bottom-Up Consolidation: Added Dimensions

After completing the content analysis and generating thirteen somewhat discrete categories of how leaders “acted upon” or used GQ/OQ feedback, it became apparent that many of the categories that were generated could be viewed as following a temporal pattern. For instance group leaders tended to employ some categories before the group, others during the group, and others later in time. Indeed, this temporal pattern is reflects, in part, the ordering of categories in Table 6. When categories were temporally organized, the new rubrics were labeled “use dimensions”. More specifically, four use dimensions were generated: 1) Pre-Group Reaction, 2) Pre-Group Planning, 3) In-Group Intervention, and 4) In-Group Consequence. At least two of the thirteen content categories fell into each dimension; with In-Group Intervention holding the most categories (see Table 7).

The temporal progression captured by the use dimensions tells a comprehensive story of how the group leaders used the GQ/OQ feedback. For example, starting with Pre-Group Reaction, the leader reviews the feedback either alone or with their co-leader, and then reacts to

the scores that were reviewed. In the second use dimension (Pre-Group Planning), the leader then decides to not use the feedback (i.e., Decision to Withhold Feedback), or begins to make intervention plans that are informed by the GQ/OQ feedback received. Some group leaders went on to use the GQ/OQ feedback in the group itself; captured by the In-Group intervention dimension. Such interventions were either explicitly or implicitly related to the feedback results, or were related to encouraging members to fill out measures or educating them about the GQ

Table 7

Leader Slip Use Dimensions with Corresponding Categories From the Content Analysis

Dimensions	Categories
Dimension 1: Pre-Group Reaction	Review of Feedback
	Reaction to Scores
Dimension 2: Pre-Group Planning	Design Specific Intervention
	Decision to Withhold Feedback
	Attendance
Dimension 3: In-Group Intervention	Explicit In-Group Feedback Use
	Non-Explicit In-Group Feedback Use
	Education About GQ
	Filling Out Measures
Dimension 4: In-Group Consequence	Downstream Effects
	Looking Forward
Unclassified	Attendance
	Filling Out Measures
	Group-Initiated Feedback Use
	OQ Use

Finally, group leaders reported that their use of GQ/OQ feedback in the aforementioned categories, at times, had a consequence on the group itself. The In-Group Consequence dimension captures this with the categories of Downstream Effects and Looking Forward falling into this rubric.

While the four use dimensions bring a temporal order to the majority of the thirteen content categories, there were a few categories that it failed to capture, which were labeled as “Unclassified” (see Table 7). For example, Group-Initiated Feedback Use while interesting does not explicitly reflect how a *leader* used feedback. Similarly, the content categories of OQ Use, Attendance, and Filling Out Measures did not add to our understanding of how a group leader used feedback. With that said, there were a few units from both the Attendance and Filling Out Measures categories that did reflect the temporal progression, and thus were included in the four use dimensions.

Phase IV Content Analysis: Debrief Interview Transcripts

The second source of data for understanding how a group leader “acted upon” GQ/OQ feedback were transcripts from the debrief interviews. As a reminder, these leader interviews followed a structured protocol (Appendix D) and were conducted at the end of each group to supplement our understanding of how group leaders saw their use of GQ/OQ feedback when viewing the total group experience. Moreover, these interviews became a critical data source for understanding feedback use for leaders who failed to provide weekly leader slips on a regular basis. The raters followed the same multi-phase process used on the leader slips with resulting in a total of 390 ratable units classified across 12 use categories (see Table 8). While the majority of the content categories were identical to those generated from the content analysis of the leader slips, there were some minor differences described below. It is important to note that the

definitions and decision rules used in the content categories that carried over from the analysis of leader slips were identical for the content analysis of the interview transcripts.

There were four categories found in the content analysis of leader slips that were not found in the interview transcripts: Attendance, Group-Initiated Feedback Use, Looking Forward, and OQ Use. This is not surprising for the Group-Initiated Feedback Use and OQ Use categories because they did not fit the structured interview goal which was to solicit leader input on the question “What does it mean for group leaders to “act upon” GQ/OQ feedback?” The remaining two content categories—Attendance and Looking Forward—also were not present because both dealt with session-to-session phenomena (i.e., looking forward to receiving feedback in the next session, or reporting individual session-by-session attendance) rather than global impressions of use. There were three categories found in the content analysis of interview transcripts that were not found in the analysis of leader slips: Ambiguous Use, Awareness of Alerters, and Self-Awareness. These new categories are described below with specific examples.

Ambiguous use. There were a number of times in the interviews when a leader made a comment that was difficult to classify ($n = 13$). Thus, an ambiguous use category was created to capture content that implied use but was unclear. For example, Leader #8 stated:

I think there was definitely a learning curve. I don't exactly know at what point. I think I was paying attention to it right from the start but I think in terms of sharing it.

The raters agreed that the unit above reflected use of the GQ (referring to a “learning curve”, “paying attention,” and to “sharing it”), but could not discern the specific use the leader was referring to. The unit is similar to two content categories—Explicit Feedback Use and Review of

Feedback but does not meet the definitions of either (see Appendix G). Another couple units from Leader #6 give further illustration:

No I did use it.

I think I was getting better at using the feedback.

Again these units depict a non-descript use of the GQ. In the first unit, the leader plainly states the use, but the unit and surrounding context do not provide sufficient further information. In the second unit, the leader states a quality of the use, but use can be assumed, though further extrapolation was not possible.

A subcategory that emerged in this category was a proportion of use. One of the questions in the debrief interview asked groups leaders to give a proportional estimate of use, but did not pull for specific type of use, thus making the use ambiguous. For example, Leader #1 provided the following:

Well I would say, even though the group was ambivalent with how they used it, I would say we used it in one form or another every week so I think it was a constant presence in one way or another in the group, whether it was just brought up as circumstances dictated or whether they took feedback at the beginning of the group. I would say it was present almost every week.

The leader states that the feedback was used in “one form or another,” constituting ambiguous use. The proportion reported here is 100%; that is, the leader reported that for every week of group, the feedback was used in some way.

Awareness of alerters. A second new category that emerged from the content analysis of interview transcripts was Awareness of Alerters ($n = 12$). This category captured comments

from the group leader where GQ/OQ feedback resulted in an increased awareness of group members who might be struggling. For example, Leader #11 stated:

I think in each of the areas for which we received the data, I think for me personally, I felt much more attentive to ways in which I could connect members, ways in which I could draw out what they were hoping to get out of sharing with the group and what their goals were.

Table 8

Leader Slip and Debrief Interview Transcript Categories Resulting from the Content Analysis

Category	Leader Slip Freq.	% of total coded units	Debrief Interview Transcript Freq.	% of total coded units
Review of Feedback	17	5	24	6.2
Reaction to Scores	77	22.6	133	34.1
Design Specific Intervention	20	5.9	13	3.3
Decision to Withhold Feedback	36	10.6	31	7.9
Ambiguous Use	NA	NA	13	3.3
Awareness of Alerters	NA	NA	12	3.1
Explicit Feedback Use	41	12	30	7.7
Non-Explicit Feedback Use	20	5.9	30	7.7
Education About GQ	5	1.5	5	1.3
Downstream Effects	89	26.1	52	13.3
Attendance	8	2.3	NA	NA
Filling Out Measures	8	2.3	3	0.8
Group-Initiated Feedback Use	7	2	NA	NA
OQ Use	10	2.9	NA	NA
Self-Awareness	NA	NA	1	0.03
Looking Forward	3	.9	NA	NA
TOTAL	341	100	390	100

Note. "NA" signifies the category was not present.

So I think my overall level, not that I don't pay attention normally, but I think I found myself paying even more attention to members that I had additional information about. These two units were exemplary because they still showed a cause-effect pattern that was also present in other use categories (e.g., Explicit Feedback Use). However, the main action described here is the act of becoming and being aware or attuned. Another verb to describe this included in the category was "to be mindful." For example, #7 stated:

I felt like I did not get bored, didn't understand it better-I understood it and took it pretty seriously the whole time and was always excited to see it and tried to be mindful of in women's group.

In this unit, the leader's excitement led to being more mindful or attentive of the GQ/OQ feedback.

Self-awareness. The final category that emerged from the content analysis of interview transcripts was Self-Awareness ($n = 1$). As noted, there was only one unit that fit this description from Leader #10:

Maybe I was a little more self-conscious about assessing what I'm doing, how I'm doing. In this unit, the leader describes the GQ caused self-consciousness, perhaps anxiety about being evaluated negatively by the GQ results. While only one unit was included in this category, it still seemed to reflect use of the GQ, but did not seem to fit anywhere else. This was a departure from the general agreed-upon rule between the raters to not create a one-unit category, but is not outside the norm for content analysis in general, often occurring for "special cases" in the content.

Content Analysis of GQ Value

As noted above, we not only hypothesized that “Acted Upon” consisted of how the group leaders used the GQ, but also what value they ascribed to it. Value of GQ/OQ feedback was treated as a separate main construct, and as such was given its own content analysis. During the rating process that produced the aforementioned content categories of use the raters also identified items that targeted the general idea of value. Using both the leader slips and interview transcripts as the data source, 85 units were identified as ratable units that spoke to the general issue of value. Our working definition of value during the content analysis was that GQ value could be classified when any unit captured opinions, judgments, appraisals, or overall feelings about the GQ/OQ feedback. The content analysis yielded six main categories: 1) Judgment of the Utility of the GQ, 2) Preference for GQ vs. Clinical Judgment, 3) Preference for GQ over OQ, 4) Judgment of the Effect of the GQ, 5) Desire for GQ/OQ feedback, and 6) General Evaluations of the GQ (see Table 9 for frequencies). What follows below is a description of each of these main categories with their associated subcategories, along with example units.

Judgment of the utility of the GQ. This category of GQ value ($n = 20$) refers to any unit in which the statement of value reflects judgments/opinions/beliefs about the relative usefulness of the GQ. Additional information that formed subcategories dealt with why the GQ was deemed useful. These included its ability to add a different perspective to the clinician, its ability to cause the group leader to have greater focus on group members and associated interventions, validating the group leader’s notions or work, and improving the leader’s ability to predict future scores.

Table 9

GQ Value Categories from Content Analysis of Leader Slips and Debrief Interview Transcripts

Category	Freq.	% of total coded units
Judgment of the Utility of the GQ	20 (3)	23.5
Preference for GQ or Clinical Judgment	12 (0)	14.1
Desire for the GQ	33 (2)	38.8
Preference for GQ Over OQ	5 (0)	5.9
General Value of the GQ	6 (1)	7.1
Judgment of the Effect of the GQ	9 (0)	10.6
TOTAL	85 (6)	100

Note. Leader slip frequencies included in parentheses.

Preference for GQ vs. clinical judgment. The next category yielded by the content analysis was Preference for GQ vs. Clinical Judgment ($n = 12$). All units had in common that the statement of value was reflective of the group leader's tendency to more often trust the GQ, their own clinical judgment, or to balance the two. As such, subcategories were "deference to GQ," "deference to clinical judgment," and "balance of GQ and clinical judgment." Below are two examples of units that represent this category well:

Also I felt some 'Oh gosh, I'm not doing a good job in the room here'. I just felt some negativity from that. (Leader #6, debrief 18.2)

I will say that it was a weird balance between really coming to value it, but maybe having a piece that I was lacking but also being able to not let it override me. (Leader #1, debrief 11.2)

The first example was included in this category because the leader was expressing that the GQ knew much better the state of the therapy group, and the leader's resultant affective state. The unit was subcategorized as "deference to GQ." The second example was included as an expression of balance of the GQ with clinical judgment because the leader states that the GQ has novel information, but that the information it provides should not supplant clinical judgment.

Preference for GQ over OQ. Units included in this category ($n = 5$) were coded when the leader reported using the GQ more often than the OQ, or if the leader explicitly stated preferring the GQ over the OQ. Only one of the five units represented in this category provided a unique image of preference for the GQ, with the four remaining units double coded. The one unique unit came from the debrief interview for Leader #9, unit 22.4: "And I found the GQ/OQ feedback, at least for the group, to be more directly useful—something I can talk about and incorporate."

The above unit was an explicit statement of preference for the GQ; the other four double-coded units were less explicit. For example, Leader #7 stated in debrief unit 26.2, "...probably a little more closely in the USO because it was bundled in with our GQ and we were so attentive to the GQ." While the leader did not explicitly say, "I preferred the GQ over the OQ," this was assumed in the context of the conversation, which was about the OQ. In this portion the leader was stating that in the group which they received GQ/OQ feedback, they were also attentive to the OQ because it accompanied the GQ. This indicated preference for the GQ because the leader's reported action was to be more attentive to the GQ. This behavioral trend was corroborated in the other units, including "We spent the majority of the time with the GQ/OQ feedback" and "I didn't use it as much--I tend to go straight to the GQ".

Judgment of the effect of the GQ. This category ($n = 9$) was represented by any unit in which the statement of value showed a judgment of effect that the GQ had on the group leader's mood, motivation, or way of thinking about the group. For example, Leader #9 stated in debrief units 6.1 and 6.2:

The effect it had on me as a leader; I thought that it was empowering- it gave me an opening to intervene with people that I thought might be struggling. (6.1)

I think as the leader, it added to my sense of confidence. (6.2)

Referring to the GQ, it is not difficult to notice that this leader believed that the GQ created a personal change. This leader described be empowered and having a greater sense of confidence as specific effects.

Leader #4 also provided two examples of judgments of the effect of the GQ:

So by seeing the positive scores on these other group members, I think that it just made me relax and not worry so much about how they were doing with the group and we could focus more on the topic of the group because this was a kind of psycho-educational CBT-oriented group. (debrief, 4.1)

I would say that I might have gotten just a little bit bored, but that was only because the members were continuing to be in a positive state. (debrief, 6.1)

While these two units are not as explicit as the statements provided by Leader #9, they were still deemed judgments of a personal effect, as the leader described the effect of becoming relaxed after seeing the scores, and then by becoming bored after seeing repeated positive scores.

Desire for GQ/OQ feedback. Another category yielded by the content analysis of GQ value was "Desire for the GQ ($n = 33$).” Units in this category were statements reflecting either an explicit or implicit want to receive group members' GQ scores organized into the feedback

report. In addition to simple ‘wanting’ for the GQ, this category also captured expressions of curiosity toward and looking forward to receiving GQ/OQ feedback. More specifically, group leaders tended to talk about wanting feedback for their non-feedback-condition groups, which made “non-feedback” ($n = 18$) a fitting subcategory. The other subcategory was “feedback,” ($n = 14$) referring to instances in which group leaders expressed a desire for the GQ in their feedback-receiving groups. One unit reflected a lack of desire for feedback, while two were unclassified, unable to receive feedback-versus-non-feedback distinction.

Below are three examples of units categorized under Desire for GQ/OQ feedback:

I think that it did, especially in the early going when I had the effect of doubting my clinical instinct a bit I would walk into the group thinking ‘If I got feedback in this group, what it would be telling me?’ (Leader #1, debrief, 21.1)

There was one time, I think the following week after, I remember getting some feedback and going ‘I don’t want to open this-I don’t know if I want to know. I’d rather go on in my blissful ignorant place’. (Leader #6, debrief, 19.2)

I’m looking forward to seeing next week’s feedback & hope that our group members are feeling more positively about their group work. (Leader #7, leader slip 112.10)

The first unit above from Leader #1 was considered Desire for GQ/OQ feedback because the leader expressed curiosity about the GQ/OQ feedback in the non-feedback group, which was considered closely related to a desire for the feedback. Naturally, this unit was subcategorized as non-feedback. The second example above was included in the main category because Leader #6 expressed a *lack* of desire for the GQ/OQ feedback. While this was not placed under either established subcategory, the unit was ultimately given a “0” to indicate the absence of the category. Finally, the third example from Leader #7 was included in the category because it was

determined that the leader was looking forward to or hoping for future feedback reports. This was subcategorized as feedback because the leader had already received reports, and would receive more in the future. This report was corroborated by records indicating assignment to the feedback condition.

General evaluations of the GQ. The final value category yielded in the content analysis was General Evaluations of the GQ ($n = 6$). This category was endorsed when there was a clear value statement, yet that did not seem to fit in any other existing category, or did not seem fit for a new category. For example, Leader #6 stated in the interview, “So I think I used it more valuably- with more value towards the end. I would lean towards the finishing side of it” (unit 20.4). This unit was clearly a statement of value given that the leader’s language was explicit in its indication of value. However, there did not seem to be any type of value indicated. One might state that the leader’s valuing changed over time, but this was not subcategorized because it was the only value statement that explicitly included a time effect.

Another example of a general evaluation came from Leader #7: “I attended a conference last week with Scott D. Miller, PhD and left with a renewed desire to be intentional about the feedback I receive” (leader slip 112.1). This was not an explicit statement of value (i.e., the leader did not say “I valued the GQ”), but it was judged a statement of value because the group leader was describing being more engaged in the feedback process.

Double-Coded Units

We hypothesized that value of the GQ/OQ feedback could serve as a potential answer for what it means to “Act Upon” the GQ/OQ feedback. Our a priori assumption was that value of the GQ would represent content separate from the content which represented GQ use, and therefore GQ value received its own content analysis. With that said, much of the same content

that was classified in the content analysis for GQ use was also classified under the categories for GQ value (see Table 10). When units were coded for both use and value they were considered “double-coded.” This is observable even at a word-by-word level, where the same portions of units were used to decide classification. Of the total 85 units classified under GG value, 17 were double-coded. An additional 27 units were classified in both content analyses, but different pieces of the same unit were interpreted as the evidence for the category classification. An example of double-coding that occurred is in the following unit:

I’m looking forward to seeing next week’s feedback and hope that our group members are feeling more positively about their group work.

This unit originally was coded in the content analysis of GQ use under the category Looking Forward. The same content in this unit was used as evidence in the content analysis of GQ value, and the unit was categorized under Desire for the GQ. An example of separated use and value content:

So by seeing the positive scores on these other group members, I think that it just made me relax and not worry so much about how they were doing with the group and we could focus more on the topic of the group because this was a kind of psychoeducational CBT-oriented group.

For this unit, the beginning and the end of the unit (from “So by seeing” to “group members” and from “could focus more” to the end of the unit) were coded as GQ use, specifically under the category Decision to Withhold Feedback. The content most reflective of GQ value was the portion when the leader described being made to relax because of the GQ/OQ feedback.

Table 10

Value-Use Overlap Totals by Value Category

Category	Unit Content Double-Coded for Use and Value	Separate Pieces of Same Unit Coded for Use and Value	Coded for Value, but Not for Use	% Overlap with GQ Use
Judgment of the Utility of the GQ	5	11	4	80
Preference for GQ vs Clinical Judgment	3	5	4	67
Preference for GQ over OQ	4	0	1	80
Judgment of the Effect of the GQ	2	3	4	56
Desire for the GQ	3	6	24	27
General Evaluations of the GQ	0	2	4	33
TOTAL	17	27	41	52

Quantification of “Acted Upon”

A fundamental goal of my dissertation was to first get a sense of how group leaders used the GQ. As shown above, group leaders start off by assessing GQ results and making plans based upon those results. Then, if they decide to use the GQ in the group, they do so either explicitly or non-explicitly. Finally, they also tend to report consequences of in-group use. While on a high level this pattern is fairly evident, group leaders tended to vary quite a bit in how they approached the GQ individually; some used the GQ religiously while others treated the measure as an afterthought. This variation could create a challenge when attempting to create a quantifiable mediator variable to represent “acting upon” feedback as we have defined it. More specifically, if the mediator variable relied too heavily on the categories from the content analysis, it would result in treating the content categories as more definitive than exploratory.

After going through the content analysis, it became apparent that we needed to come up with a quantification scheme. The purpose would not only be to create a variable (s) that in the future could be manipulated in mediation analyses, but to do so as correctly as possible; the quantification scheme was also meant to give credit to leaders who used it in varied ways, and to keep from mistakenly giving credit to those who did not use it as dynamically. While there is likely a myriad of ways to create a quantification scheme, I present one method below. It is also important to note that this scheme was extracted from the data collected from the start of the parent study. When the full set of qualitative data is coded (i.e., all two years) the scheme for quantification is likely to change due to shifts in content categories (e.g., addition of categories, re-naming of categories, etc.) or shifts in our understanding of “Acted Upon.” Hence, while the categories from the current analysis accurately portray “Acted Upon,” the current quantification scheme is purely exploratory.

To begin illustrating the process we undertook to arrive at the current scheme, see Table 11 below. This table shows sums for each of the four dimensions found in the content analysis, stratified by group leader. As noted above, there was a great deal of variability across both leaders and dimensions in how the feedback was used. The “Total Units” column is a sum across dimensions for each leader, The “Total Units per Group” column shows how much each leader used the feedback when controlling for number of groups led, with per-group corrections affecting the first three leader sums, as they were the only leaders who ran more than one therapy group. We went through the same process for the debrief interview transcripts, shown in Table 12. For this table there is no “Total Units per Group” column because the interviews were held only once for each leader at the end of the study period, not per-session as with the interview transcripts.

These two sums were a good start to creating the quantification scheme, but there are problems inherent in the count data that make sums inadequate for capturing the entire picture. First, summing units from qualitative categories rewards verbosity; that is, group leaders who simply write or speak more accumulate more ratable units. Second, qualitative categories are subject to confirmation bias in coding, which may have caused raters to provide more units to certain leaders, even after a rigorous checking process. Finally, for the leader slips, the raters were dependent also on the number of slips that were actually turned in; thus, group leaders who were more diligent with providing data were favored for higher sums. Because of these reasons, it seemed that a definition of “Acted Upon” based solely on sums for leader slips and debrief interview transcripts was not fully accurate.

The content analysis revealed a couple facts aside from main categories described above. First, sometimes leaders did not act upon the GQ/OQ feedback because they did not have any alerts (especially negative alerts) to prompt them. For example, Leaders #4, #8, and #10 received far fewer negative alerts than the other eight leaders, making their “chances” for a higher sum less because of fewer baseline issues (see Table 13). This is illustrated in the debrief interview of Leader #4, who stated, “So they were doing well in all of those areas, so we didn’t feel the need to bring [the feedback] up or to discuss it with them.”

This phenomenon led us to attempt to quantify how leaders responded to alerts given. Fundamentally, negative absolute and relative alerts from the GQ represent a direct parallel to not-on-track alerts given in preceding feedback studies (e.g., absolute negative alerts on positive bond are analogous to red alerts on the OQ). This calls the therapist to some sort of action. As shown in Table 13, it is possible that leaders who over the course of the semester received more negative alerts also responded to those alerts more often. Accordingly, we correlated the four

leader slip use dimensions yielded by the content analysis with the sum of negative alerts per group per session. Of the four Pearson correlations there was one resultant significant relationship; when more alerts occurred leaders tended to use the “In-Group Consequence” use dimension more often ($r = .13, p < .05$; see Appendix H).

Table 11

Leader Slip Dimension Totals by Leader

Leader	Pre-Group Reaction	Pre-Group Planning	In-Group Intervention	In-Group Consequence	Total Units	Total Units per Group
1	5	5	19	41	70	35
2	27	18	8	8	61	30.5
3	23	7	14	23	67	33.5
4	2	3	0	0	5	5
5	9	3	11	9	32	32
6	1	2	1	0	4	4
7	21	14	1	3	39	39
8	4	2	8	5	19	19
9	0	1	3	3	7	7
10	2	0	3	0	5	5
11	0	3	2	0	5	5
MEAN (SD)	8.54 (10.13)	5.27 (5.69)	6.36 (6.20)	8.36 (12.76)	28.54 (26.81)	19.54 (14.56)

Table 12

Debrief Interview Transcript Dimension Totals by Leader

Leader	Pre-Group Reaction	Pre-Group Planning	In-Group Intervention	In-Group Consequence	Total Units
1	25	2	6	12	45
2	22	0	7	10	39
3	20	3	8	10	41
4	11	7	0	0	18
5	12	3	13	3	31
6	9	6	16	1	32
7	9	3	1	1	14
8	8	0	11	3	22
9	13	3	9	10	35
10	10	5	0	0	15
11	18	3	6	1	28
MEAN (SD)	14.27 (5.93)	3.18 (2.18)	7.00 (5.23)	4.64 (4.78)	29.09 (10.68)

Another way of quantifying “Acted Upon” is by examining what happens at the session level. Looking at the leader slips only (the debrief transcripts are not linked to session), we may be able to understand if alerts had a moment-specific impact. It is possible that when alerts occurred at a higher frequency in a given session that this base rate may have influenced how the leader used the feedback. In order to better understand this, we looked at the total number of negative alerts per session and number of leader slips responding to negative change. Five of the categories from the content analysis had a subcategory specification for “direction of change,” coded dichotomously as either positive or negative change (see Appendix H). Using Pearson

point-biserial correlation, only one use category yielded a significant correlation with number of per-session alerts. A higher frequency of alerts in a given session was significantly related with the frequency of units coded on the Explicit Feedback Use category ($r = -.39, p < .05$).

Table 13

Comparing Alerts and “Acted Upon” for Leader Slips and Debrief Interview Transcripts

Leader	NEG Alerts per Group	# of leader slip units responding to NEG change	# of debrief units in response to NEG change
1	26.5	20	23
2	18	12	21
3	28.5	18.5	11
4	13	0	4
5	56	27	10
6	35	1	9
7	54	14	0
8	12	10	6
9	35	3	20
10	6	5	0
11	41	2	8

Finally, it is also possible that characteristics of both the group and session are related to how group leaders act upon the feedback. To assess this, we decided to look at the number of members alerting per session, and the proportion of members alerting to the total number of members attending the session. These statistics may help us understand how “severe” the group was in terms of negative alerts, and if having a higher proportion of members alerting makes a difference to how leaders act upon the feedback. As was used above, per-session counts of the four use dimensions (Pre-Group Reaction, Pre-Group Planning, In-Group Intervention, and In-

Group Consequence) were used to portray “Acted Upon.” We first correlated the count of members alerting per session with each of the four leader slip use dimensions. Similar results were found with the correlations above; two of the four associations were significant, though small. Namely, an increase in the number of members alerting per session was significantly associated with increased use of the In-Group Use ($r = .16, p < .001$) and In-Group Consequence ($r = .18, p < .001$) categories. More specifically, as the frequency of members alerting increased, therapists produced more ratable units describing use in the group and its consequences. The proportion of members alerting yielded similar results; the same two use dimensions showed small but significant correlations ($r = .09, p < .01$ for In-Group Use and $r = .12, p < .001$ for In-Group Consequence, respectively).

Overall, the majority of correlations explored were non-significant. This suggests that the frequency of negative alerts per session, the absolute, and proportional number of group members are not associated with the categories generated by the content analysis. However, there were a few consistent yet small correlations that emerged. Leaders who experienced a higher number of alerts over the entire group episode were more likely to have acted upon feedback using the “In-Group Consequence” dimension. Taking a more granular view, when the number of alerts increase in a session leaders are also more likely to report their use of feedback with the “Explicit Use” category. And as both the absolute number and proportion of members who are alerting increases, leaders also act upon feedback more frequently with the “In-Group Consequence” and “In-Group Use” dimensions.

Discussion

The purpose of this study was to supplement and hopefully explain the findings from a larger study that is looking at the effect of GQ and OQ feedback in group therapy. As shown in

Figure 2, the original study was a randomized controlled trial in which 179 participants were assigned to therapy groups in which the group leader received feedback from the GQ and OQ; 185 participants were assigned to groups in which the group leader only received OQ feedback. In the entire study there were 59 unique groups, with 69 semester-long group episodes from which data was collected (the difference was attributable to groups that carried over from one semester to the next).

One preliminary result from the parent study that is important to note has to do with covariates that explain variability in GQ subscale scores (Burlingame, Woodland, & Whitcomb, 2014). First, researchers found that differences in both slope (change) and intercept (absolute value) for all GQ subscales can be explained by group membership; i.e., the group that a member attended. Second, the effect of feedback had less of an effect over time due to the fact that group members in the feedback condition alerted less frequently in later compared to earlier sessions. In fact, the data showed that three out of four clients were no longer alerting on negative relationship by the end of the group episode (Burlingame et al., 2014). For groups that did not receive feedback, group members were more likely to remain alerting over the course of the group without a temporal shift downward. Upon talking to group leaders in my study, it seems that group leaders who received the feedback intervention were in essence able to “stamp out” the alerts later in the semester, which may help explain this time effect. In early sessions group leaders acted upon the feedback—with particular attention given to the negative relationship subscale—and these actions were associated with a reduction in alerts at both an aggregate and session level. These findings make sense clinically, as therapists are more likely to respond when something is amiss in the group, placing positive gains as a lower priority.

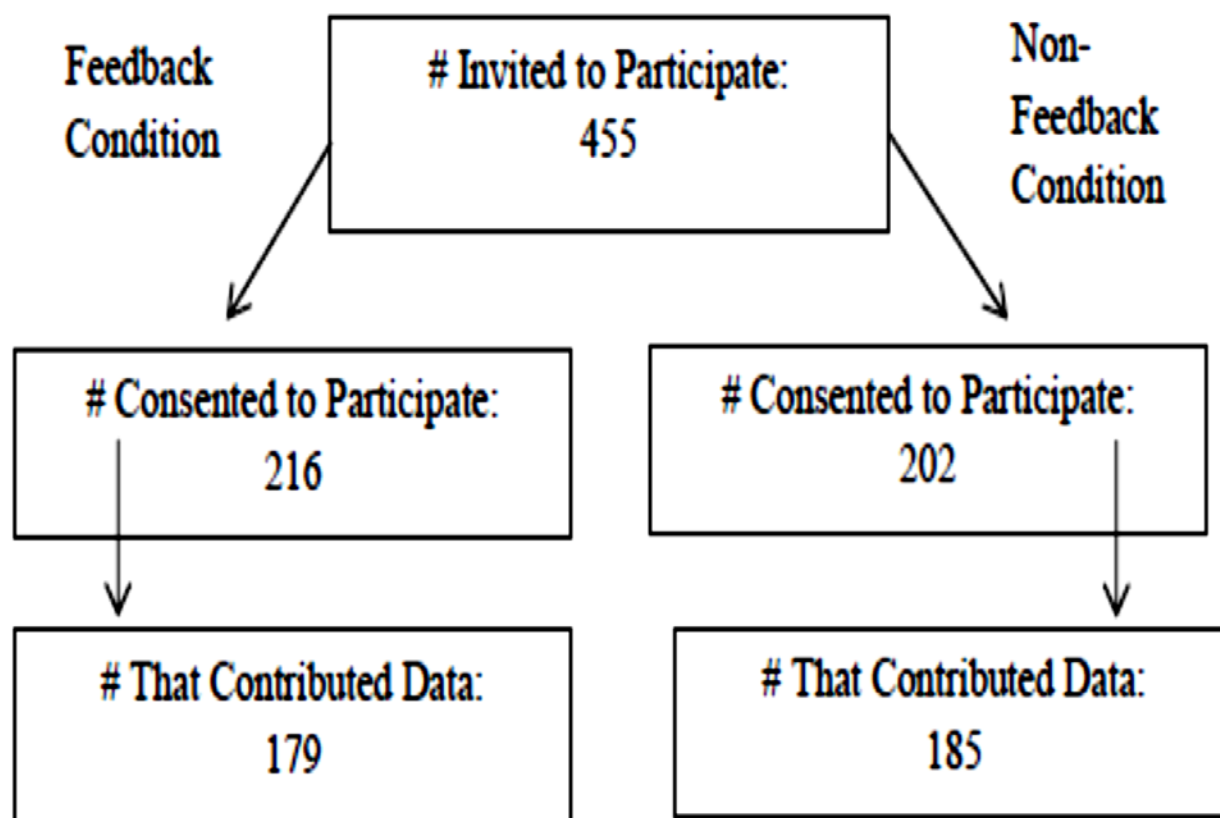


Figure 2. Flow of study participants for the parent RCT

Quantification of “Acted Upon”

The second goal of this study was to develop a preliminary idea of how to quantify the “Acted Upon” construct that was operationalized through the content analysis. At a high level, this idea of quantifying from qualitative is not only the first step toward statistical analysis, but also represents the first step toward creating an accurate and reliable measurement. Each measure in the pantheon of psychological testing has undergone the same process that the “acted upon” construct has undergone. In creating an operational definition of “acted upon” future studies will be able to attempt to replicate the findings, ultimately arriving at a construct that is measurable, reliable, and valid. With that said, operationalization of new phenomena requires a great deal of trial and error, as was represented in the current study.

The first issue to note is that the quantification scheme largely relied upon the data from the weekly leader slips, because the data from the interview transcripts could not be tied to each session. This fact suggests that future studies collect session level data on use of feedback. As we first looked at simple sums for each of the use dimensions by session it clearly provided useful information, but these sums seemed insufficient to adequately capture “Acted Upon” quantitatively. For instance, a simple sum rewards verbose group leaders, or group leaders who turned in more leader slip data which may not accurately reflect the true “acted upon” construct. Stated differently, it’s entirely possible to have a less verbose group leader who assiduously uses GQ/OQ feedback but simply doesn’t talk much about it. An equally and perhaps more important limitation of using simple sums is that it might unduly quantitatively punish group leaders who received fewer negative alerts and thus had fewer chances to provide data regarding how they used the feedback.

We felt that this last reason was particularly important so we several post hoc analyses were conducted to explore its viability. We chose to use Pearson correlations to assess these hypotheses because at this point we had count variables (i.e., (1) sums by leader, session, group, and member from the four use dimensions and (2) sums of negative alerts by leader, session, group, and member) which could be assessed linearly. Because these hypotheses were exploratory, it also made sense to use Pearson’s r because this coefficient represents a basic linear relationship. If the Pearson correlations showed significant relationships, then they would provide greater justification for quantification in future studies.

We first examined four correlations to see if there was a significant relationship between number of negative alerts received and the number of units across the four use dimension sums at a group episode level; i.e., frequency of alerts and acted upon units were summed across the

entire group episode. This analysis tests if the frequency of use in the four dimensions of acted upon was at all related to the receipt of negative alerts. Two significant but small significant relationships were found. Increases in two use dimensions—In-Group Use and In-Group Consequence—were positively related to the number and proportion of alerters in a given session. One way to understand these relationships is that when a leader is faced with more individuals who are report conflict or relational dissonance in the group, this becomes more salient leading to a higher number of interventions. It is also possible that group leaders are simply able to remember and report more on feedback interventions when the perceived group environment is conflictual or relationally problematic.

Noting a relationship between the frequencies of comments between the four use dimensions the number of alerts across the entire group episode led to exploration of similar differences at the session level. In short, we wanted to understand whether more individuals alerting in a given session might lead to group leaders reporting a higher number of acted upon units for specific use categories. We examined the correlation between the total number and proportion of group members alerting in a given session with the frequency of interventions reported in the four use dimensions. Once again, there was a significant relationship between the number of members alerting and the same two use dimensions; however, these relationships were quite small. When squaring these correlations one can estimate the variance in the use dimensions explainable by members alerting and proportion of alerters to non-alerters, and vice versa. For example, 3.3% of the variance in the In-Group Consequence dimension can be explained by the absolute number of members alerting in a particular session. Clearly, there are other variables that might explain how and why group leaders use the feedback.

There are several possible explanations for differences in variability in frequencies of interventions within and between the four use dimensions. For instance, this was a naturalistic study and group leaders clearly had a routine for how they prepared for their groups and delivered in session interventions. The parent study added to this clinical demand by asking group leaders to report on GQ/OQ use and it is highly likely that differences in reporting frequencies is simply related to a to group leaders' clinical load. Some group leaders may simply have had insufficient time to report if and how they used GQ/OQ feedback. They likely did not have time to provide in-depth reporting on how they used the feedback. A second explanation for variability might include varying attitudes toward providing data and toward the feedback in general. It is possible that leaders vary on the continuum of favorable and unfavorable attitudes toward participating in research, utility of feedback, or beliefs that feedback adds to value to improving client outcomes. Future research might explore attitudinal and belief differences to establish if they are associated with differences in use of GQ/OQ feedback.

Content Analysis of GQ Value

Another aspect of "Acted Upon" sought to tap into attitudes toward the GQ/OQ feedback, conceptualized as the value that group leaders ascribed to the feedback. A separate content analysis over the entire data set was performed which yielded six value categories. While this content analysis provided some useful information regarding attitudes toward the feedback, making claims about its validity beyond that is tenuous. First, because the same units were used from the original content analysis for GQ Use, there was a high degree of overlap between units coded as "Use" and units coded as "Value." This led to double-coding, which could potentially dilute the meaning extractible from both the use and value constructs. Also, there was a great

deal of overlap conceptually between the categories themselves. Another reason why the value construct was difficult to interpret was because the categories themselves dealt with attitudes, which are much more difficult to reliably code in content analysis. Perhaps future research should more directly assess value by way of direct self-report tapping specific aspects identified a priori. For these reasons, the results from the content analysis of GQ Value should be interpreted cautiously. Future studies that may like to revisit the idea of ascribing value to the feedback should do so with the understanding that the construct is more difficult to interpret, and may really only be represented by one or two subcategories.

General Limitations/Biases

Several measures were taken in this study to reduce bias. These included blind, independent rating of the qualitative content on the first pass, as well as raters challenging one another on their perception of the content in subsequent passes through the data. Furthermore, the raters also sought out consultation from a trained supervisor when disagreements about the content arose. With that said, it is important to note potential limitations of the study as a whole.

First, because the study was only a piece of a larger study that has yet to be completed, the results should be considered exploratory, and are subject to change once the full dataset is analyzed. This is applicable to both the results of the content analysis and of the preliminary quantification scheme. With a full dataset future researchers will be able to produce a replication that likely has more statistical power and likely better represents the “Acted Upon” construct.

Another limitation dealt with the raters’ relationship to the larger parent study. Both raters were involved in the process through the data were gathered and managed. They were also directly involved in the creation of the feedback intervention. And while the raters independently rated the data upon the first pass, they did not do so for subsequent passes through

the content (all later passes were performed using consensus coding). The raters' level of involvement in the study, familiarity with the group leaders, and use of consensus coding potentially could have biased their perception of the content and the leaders who provided the content. On the other hand, this greater familiarity may have allowed the raters to better understand the meaning inherent in the group leader units. Finally, it is possible that they missed some aspect of the "Acted Upon" construct that was present in the current dataset. A replication should employ raters who are naïve to the purposes of the larger RCT, and preferably have more established experience with content analysis. The raters should also be naïve to the results from this preliminary study.

For the quantification scheme, there exists a myriad of alternatives to quantitatively reflect the results of the qualitative study. Whichever scheme is presented in the future should be done with the understanding that subjectivity when converting qualitative categories into quantitative codes will likely still exist and be an important issue when considering validity of sums of content units. One help with this issue will be greater number of total units. A larger sample size will provide greater power to better detect whether an effect of "Acted Upon" exists. Also, obtaining a full dataset of both qualitative categories and data from the parent study will allow for exploring other questions. For example, researchers might look at "successful" vs. "unsuccessful" group leaders (i.e., group leaders who saw significant positive or negative changes in alert status), and whether acting upon the feedback made a difference in either contributing to or inhibiting that success.

A final limitation to the study lies in its novelty. This study is one of very few in the entire feedback literature (both in individual and in group psychotherapy) that looked at what takes place between the intervention of given feedback and the outcome. As such, there is not

much precedent for how to conduct a study that attempts to operationalize what it means to act upon feedback. Nevertheless, future studies will likely add to our current understanding of what it means to act upon feedback. Understanding the use of feedback from both a scientific as well as applied practice perspective could have an impact on how clinicians and researchers alike will advance practice-based evidence in the coming years.

References

- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77*(4), 693-704.
- Bakali, J., Baldwin, S., & Lorentzen, S. (2009). Modeling group process constructs at three stages of group psychotherapy. *Psychotherapy Research, 19* (3), 332 – 343.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C.,...McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*(2), 184.
- Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology, 69*(2), 197.
- Bormann, B., & Strauss, B. (2007). Gruppenklima, Kohäsion, Allianz und Empathie als Komponenten der therapeutischen Beziehung in Gruppenpsychotherapien — Überprüfung eines Mehrebenen-Modells. [Group climate, cohesion, alliance, and empathy as components of the therapeutic relationship within group psychotherapy — Test of a multilevel model]. *Gruppenpsychotherapie und Gruppendynamik, 43*, 1-20.
- Bryman, A. (2006). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology, 9*(2), 111-126.
- Burlingame, G. M., & Beecher, M. (2012). *The effects of therapeutic relationship feedback on group leader behavior and group member outcomes*. Unpublished manuscript, Department of Psychology, Brigham Young University, Provo, Utah.
- Burlingame, G. M., Fuhrman, A., & Johnson, J. (2002). Cohesion in group psychotherapy. In J.

- Norcross (Ed.), *A guide to psychotherapy relationships that work* (pp. 71-88). Oxford: Oxford University Press.
- Burlingame, G., Gleave, R., Erekson, D., Nelson, P. Olson, J., Thayer, S. & Beecher, M. (in press). Differential effectiveness of group, individual and conjoint treatment: An archival analysis. *Psychotherapy Research*.
- Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (5th ed.) (pp. 646-697). New York: Wiley & Sons.
- Burlingame, G., Strauss, B., & Joyce, A. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th ed.) (pp. 640-689). New York: Wiley & Sons.
- Burlingame, G. M., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K.R., & Ogrodniczuk, J., (2006). *CORE Battery-Revised: An assessment tool kit for promoting optimal group selection, process, and outcome*. New York: American Group Psychotherapy Association.
- Burns, D. D., & Auerbach, A. (1996). Therapeutic empathy in cognitive-behavioral therapy: Does it really make a difference? In P. M. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 135-164). New York: Guilford.
- Chapman, C. L., Burlingame, G. M., Gleave, R., Rees, F., Beecher, M., & Porter, G. S. (2012). Clinical prediction in group psychotherapy. *Psychotherapy Research*, 22(6), 673-681.

- Davies, D. R., Burlingame, G. M., Johnson, J. E., Gleave, R. L., & Barlow, S. H. (2008). The effects of a feedback intervention on group process and outcome. *Group Dynamics: Theory, Research, and Practice, 12*(2), 141.
- Evans, C. R. & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research, 22*, 175-186.
- Finch, A. E., Lambert, M. J., & Schallje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy, 8*, 231-242.
- Fuhriman, A., Drescher, S., Hanson, E., Henrie, R., & Rybicki, W. (1986). Refining the measurement of curativeness: An empirical approach. *Small Group Behavior, 17*, 186-201.
- Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. *Small Group Research, 26*(4), 497-520.
- Hannan, C. Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059-1064.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.

- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602-611.
- Johnson , J. E. , Burlingame , G. M. , Olsen , J. A. , Davies , D. R. , & Gleave , R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology*, 52, 310-321.
- Kivlighan, D. M., Jr., Multon, K. D., & Brossart, D. F. (1996). Helpful impacts in group counseling: Development of a multidimensional rating system. *Journal of Counseling Psychology*, 43, 347-355.
- Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. New York: Sage.
- Krogel, J., Burlingame, G., Chapman, C., Renshaw, T., Gleave, R., Beecher, M., & MacNair-Semands, R. (2013). The Group Questionnaire: A clinical and empirically derived measure of group relationship. *Psychotherapy Research*, 23(3), 344-354.
- Krogel, J., Beecher, M., Presnell, J., Burlingame, G., & Simonsen, C. (2009). The Group Selection Questionnaire: A qualitative analysis of potential group members. *International Journal of Group Psychotherapy*, 59, 529-542.
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69(2), 159.
- MacKenzie, K. R., Burlingame, G. M., & Fuhriman, A. (1999, February). *Psychotherapy group and group program evaluation: Clinical use of brief process and outcome measures*. Workshop presented at the annual convention of the American Group Psychotherapy Association, Houston, TX.

- Moos, R. H., & Humphrey, B. (1974). *Group environment scale*. Palo Alto, CA: Consulting Psychologists Press.
- Mullen, B. & Copper, C. (1994). The relationship between group cohesiveness and performance: An integration. *Psychological Bulletin*, 115(2), 210-227.
- OQ–Analyst [Computer software]. Salt Lake City, UT: OQ Measures.
- Sandelowski, M. (2000). Focus on research methods: Whatever happened to qualitative description? *Research in Nursing and Health*, 23(4), 334-340.
- Schreier, M. (2012). *Qualitative content analysis in practice*. New York: Sage.
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298-311.
- Silbergeld, S., Koenig, G. R., Manderscheid, R. W., Meeker, B. F., & Hornung, C. A. (1975). Assessment of environment-therapy systems: The Group Atmosphere Scale. *Journal of Consulting and Clinical Psychology*, 43, 460-469.
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research*, 22(6), 638-647.
- Spielmanns, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology & Psychotherapy*, 13(3), 202-214.
- Stinchfield, R. D., & Burlingame, G. M. (1991). Development and use of the directives rating system in group therapy. *Journal of Counseling Psychology*, 38(3), 251.

- Strauss, B., Burlingame, G. M., & Bormann, B. (2008). Using the CORE-R battery in group psychotherapy. *Journal of Clinical Psychology: In Session*, 64(11), 1225-1237.
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. New York: Sage.
- Substance Abuse and Mental Health Services Administration (2012, August). OQ-Analyst. Retrieved from the National Registry of Evidence-based Programs and Practices Web site, <http://www.nrepp.samhsa.gov/ViewIntervention.aspx?id=22>
- Thayer, S. D. (2012). *The validity of the group questionnaire: Construct clarity or construct drift?* Unpublished dissertation, Brigham Young University, Provo, UT.
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem solving strategies in routine practice. *Journal of Counseling Psychology*, 58, 59-68.
- Yalom, I., & Lescsz, M. (2005). *The theory and practice of group psychotherapy* (5th Ed.). New York: Basic Books.

Appendix A

Weekly GQ/OQ feedback Report All alerts are from your last group session

Leader Name: J.M. Barrie

Group ID: 2

Date of Group: 12/5/1902 (Session #11)

Group Members who completed a GQ: Captain Hook, Smee, Wendy, Peter Pan, Tinkerbell, Michael, Lost Boy #1

DID NOT COMPLETE GQ:

DID NOT ATTEND:

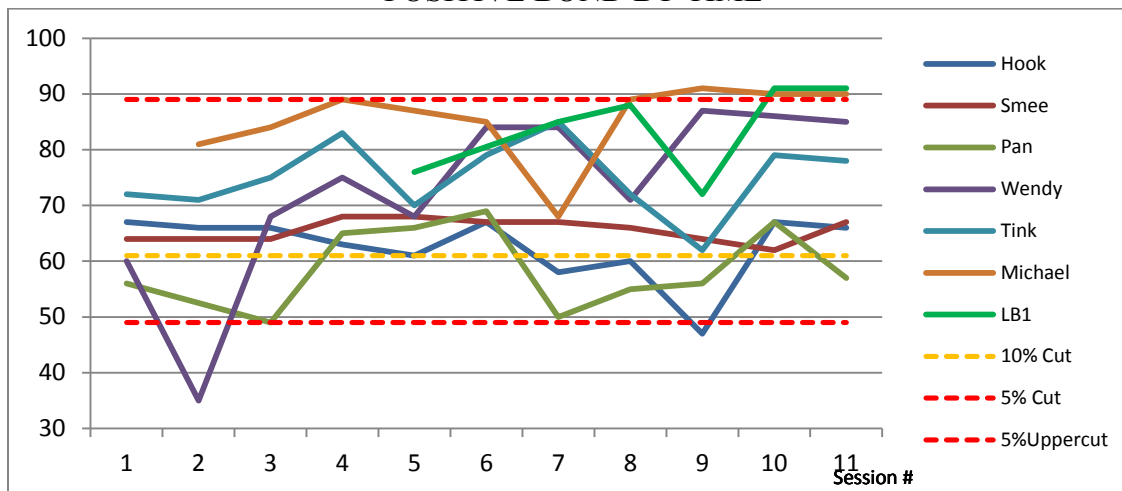
ABSOLUTE ALERTS—based on last group session GQ		
	Clients at or below the 10 th percentile ☹	Clients at or above the 95 th percentile ☺
Positive Bond	Peter Pan	Lost Boy #1, Tinkerbell
Positive Work	Peter Pan	Tinkerbell
Negative Relationship	None	Captain Hook, Michael, Wendy, Lost Boy #1, Tinkerbell

RELATIVE ALERTS—based on last group session GQ		
	Clients reporting reliable negative change	Clients reporting reliable positive change
Positive Bond	Peter Pan	None
Positive Work	None	None
Negative Relationship	None	None

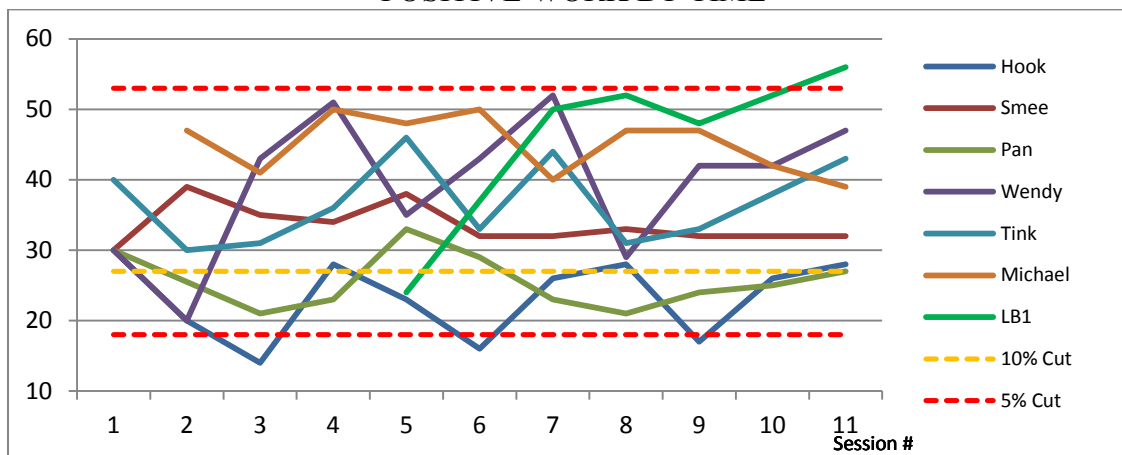
OQ ALERTS				
	Alert Status	Change From Initial	Initial Score	Most Recent Score
Captain Hook	Red	Reliably Worse	79	129
Smee	Yellow	Reliably Worse	65	81
Wendy	Green	Reliably Improved	65	44
Peter Pan	Yellow	Reliably Worse	72	89
Tinkerbell	White	No Reliable Change	44	47
Michael	Green	Reliably Improved	60	44
Lost Boy #1	Blue	Reliably Improved	86	45

List actions (if any) that you took based upon last week's GQ/OQ feedback. List any specific member targeted.
Leader Name _____ Date of session where feedback was implemented:

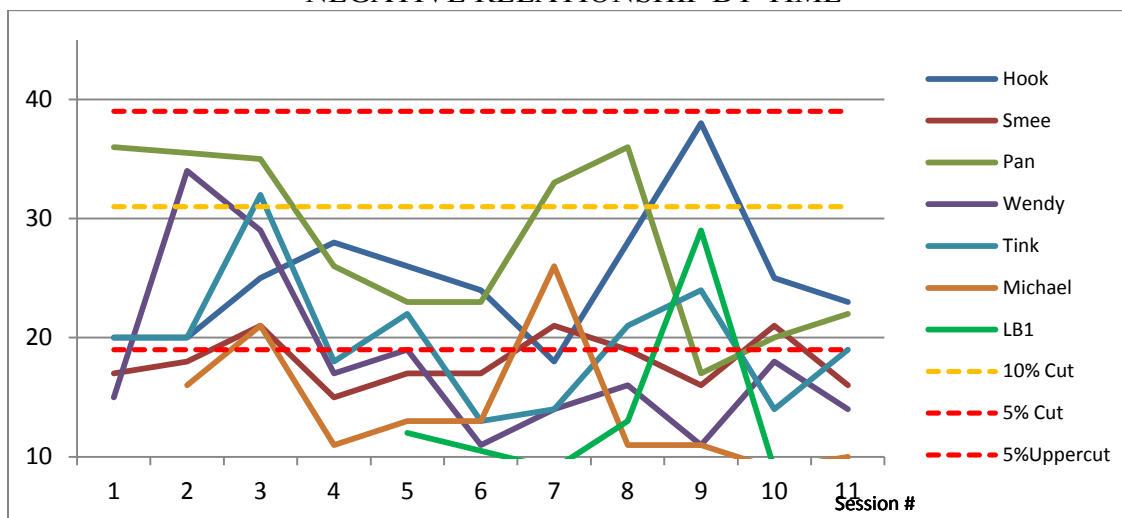
POSITIVE BOND BY TIME



POSITIVE WORK BY TIME



NEGATIVE RELATIONSHIP BY TIME



FACET-LEVEL DATA FOR ALERTERS

Name	Facets			
Peter Pan	Trait	Member-Member	Member-Leader	Member-Group
	Positive Bond	Weak (19)	Weak (19)	Weak (19)
	Positive Work	Weak (10)	Average (17)	
	Negative Relationship	Average (8)	Weak (9)	Weak (5)

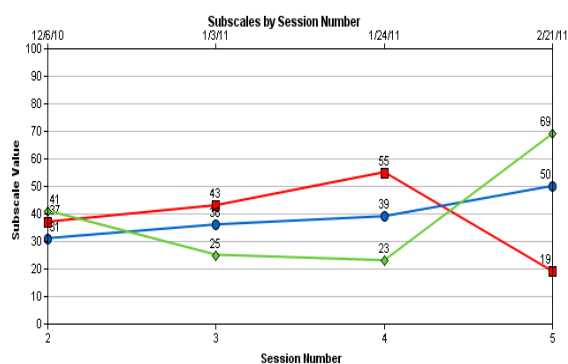
ITEM-LEVEL DATA FOR ALERTERS

Name	Facets			
Peter Pan	Positive Bond	2-I felt that I could trust the other group members during today's session. (Moderately true)	1-I felt that I could trust the group leaders during today's session. (Moderately true)	26-The members liked and cared about each other. (Moderately true)
		4-The other group members and I respect each other. (Somewhat true)	3-The group leaders and I respect each other. (Moderately true)	27-The members felt what was happening was important and there was a sense of participation. (Slightly true)
		6-I feel the other group members care about me even when I do things that they do not approve of. (Moderately true)	5-I feel the group leaders care about me even when I do things that they do not approve of. (Somewhat true)	28-We cooperate and work together in group. (A little true)
		8-The other group members were friendly and warm toward me. (Moderately true)	7-The group leaders were friendly and warm toward me. (Moderately true)	29-Even though we have differences, our group feels secure to me. (Somewhat true)
	Positive Work	10-The other group members and I agree about the things I will need to do in therapy. (Slightly true)	9-The group leaders and I agree about the things I will need to do in therapy. (Moderately true)	30-The group members accept one another. (Moderately true)
		12-The other group members and I agree on what is important to work on. (Slightly true)	11-The group leaders and I agree on what is important to work on. (Somewhat true)	
		14-The other group members and I have established a good understanding of the kind of changes that would be good for me. (A little true)	13-The group leaders and I have established a good understanding of the kind of changes that would be good for me. (Moderately true)	
		16-The other group members and I are working together toward mutually agreed upon goals. (A little true)	15-The group leaders and I are working together toward mutually agreed upon goals. (Slightly true)	

Appendix B

Sample GQ/OQ feedback Report Currently Found on OQ-Analyst

Name: C-OQ45, GEORGE, R ID: MRN0101 Session Date: 2/21/2011 Session: 5 Clinician: Clinician, Bob Clinic: UT Sandy Clinic Diagnosis: Unknown Diagnosis Instrument: Group Questionnaire Questionnaire Status: Valid Unanswered Questions: 0	Most Recent Score: 138 Initial Score: 109																																				
Comparative Group: Counseling Center <table border="1"> <thead> <tr> <th>Trait</th> <th>Member-Member</th> <th>Member-Leader</th> <th>Member-Group</th> </tr> </thead> <tbody> <tr> <td>Positive Bond</td> <td>Weak (19)</td> <td>Average (23)</td> <td>Average (27)</td> </tr> <tr> <td>Positive Work</td> <td>Strong (24)</td> <td>Strong (26)</td> <td></td> </tr> <tr> <td>Negative Relationship</td> <td>Average (8)</td> <td>Weak (5)</td> <td>Weak (6)</td> </tr> </tbody> </table>	Trait	Member-Member	Member-Leader	Member-Group	Positive Bond	Weak (19)	Average (23)	Average (27)	Positive Work	Strong (24)	Strong (26)		Negative Relationship	Average (8)	Weak (5)	Weak (6)	<table border="1"> <thead> <tr> <th>Subscales</th> <th>Current</th> <th>CC</th> <th>SMI</th> </tr> </thead> <tbody> <tr> <td>Positive Bond:</td> <td>69</td> <td>67-88</td> <td>48-82</td> </tr> <tr> <td>Positive Work:</td> <td>50</td> <td>31-51</td> <td>25-51</td> </tr> <tr> <td>Negative Relationship:</td> <td>19</td> <td>12-29</td> <td>12-33</td> </tr> <tr> <td>Total:</td> <td>138</td> <td>110-168</td> <td>85-166</td> </tr> </tbody> </table>	Subscales	Current	CC	SMI	Positive Bond:	69	67-88	48-82	Positive Work:	50	31-51	25-51	Negative Relationship:	19	12-29	12-33	Total:	138	110-168	85-166
Trait	Member-Member	Member-Leader	Member-Group																																		
Positive Bond	Weak (19)	Average (23)	Average (27)																																		
Positive Work	Strong (24)	Strong (26)																																			
Negative Relationship	Average (8)	Weak (5)	Weak (6)																																		
Subscales	Current	CC	SMI																																		
Positive Bond:	69	67-88	48-82																																		
Positive Work:	50	31-51	25-51																																		
Negative Relationship:	19	12-29	12-33																																		
Total:	138	110-168	85-166																																		



Graph Label Legend:

Green Diamond: Positive Bond

Blue Circle: Positive Work

Red Square: Negative Relationship

Appendix C

GQ/OQ feedback Study Script

- Research in individual therapy has shown that when the therapist receives questionnaire-based feedback they are better able to meet the needs of clients, which in turn helps clients achieve more improvement. We are extending this research to group treatment. Thus, you will be invited to complete one questionnaire called the GQ or Group Questionnaire that measures how you are experiencing the helping environment of the group as well as me as a group leader and one questionnaire called the OQ which measures your current level of distress. The OQ is completed by all CCC clients before each service appointment and the GQ is the focus of this study. We will want to get you to complete the GQ after each group.
- I will be running two groups this semester. In one group I will be receiving GQ/OQ feedback and in the other I will not be receiving feedback. I don't know which group you will be in because it is based upon the toss of a coin (random assignment). In either case, it is important that you complete the GQ because we're interested in determining if and how this helps to improve the group relationship and client outcomes.
- **We understand that this will take up about 5 minutes of your time and the supervising faculty member (Dr. Burlingame and Beecher) has agreed to compensate you for this time. You will be given \$10 after you sign up for the study and complete the initial two measures (OQ & GQ). After that you will earn \$5 for each group session you attend where you complete both measures. The \$10 will be given to you immediately and the remaining amount will be sent to you at the end of the study as a cumulative sum. If you complete measures for all the group sessions you attend, you will receive an extra \$20 as a thank you. If you attended the maximum number of sessions possible (12) during a semester you could earn as much as \$90 for your participation in the study.**
- Your participation in the study will be completely voluntary. Any and all information that you provide will be kept secure and confidential. Please read and sign the attached informed consent form if you would like to participate.

Appendix D

Interview Questions for GQ/OQ feedback study leaders

1. To what extent, if any, did you verbally introduce feedback to your group?
 - a. When you introduced feedback, what was the effect of receiving feedback upon:
 - i. You as a group leader; if yes have them explain their view on causal path
 - ii. specific group members (can you give a typical example about how group members responded to the feedback); if yes have them explain their view on causal path
 - iii. members responded to the feedback); if yes have them explain their view on causal path
 - iv. the group process as a whole; if yes have them explain their view on causal path
2. Did you notice that you were affected by receiving feedback even if you did not explicitly share the feedback with your group members? If so, please explain.
3. What if any effect did you notice from the process of receiving GQ/OQ feedback over time?
4. If you were to give a percentage of agreement between your own perceptions as a leader and the GQ/OQ feedback, how often were your perceptions aligned with:
 - a. Positive bond feedback
 - b. Positive work feedback
 - c. Negative relationship feedback
5. Did you ever purposely withhold bringing feedback into the group (i.e., per clinical judgment)? If so, why?
6. To what extent (if any) has your feedback experience influenced you or your clients' behavior in your **non-feedback groups**?
7. To what extent, if any, did you use the OQ feedback in your two groups?

Appendix E

Preliminary Themes from Pilot Content Analysis of Leader Feedback Slips

Yes/No Items	Count	Percent of Total Themes
Includes Narrative	48	92.3
Includes cause and effect?	33	63.5
Report of eliciting further group process/interaction (loosening effect)	21	40.4
Verbal Implementation		
Education about GQ subscales/purpose	5	9.6
Monitored those who were alerting	7	13.5
Elicited member feedback regarding specific subscales on individual level	14	26.9
Elicited member feedback regarding specific subscales at group-as-a-whole level	14	26.9
Discussed feedback with co-leader	14	26.9
Invited members to work on a specific subscale	1	1.9
Only reported results at group-as-a-whole level	2	3.8
Elicited member feedback at a member-group level	1	1.9
Asked members how they would like to receive feedback they give	2	3.8
Focused on general group process/cohesion	1	1.9
Focused on individual member without discussing feedback	6	11.5
Invitation to fill out measures	2	3.8
Elicited group-level feedback without mentioning feedback	1	1.9
Non-Verbal Implementation		
Personal Reaction to scores	15	28.8
Confirmation/disconfirmation of impressions	3	5.8
Withheld feedback from group per clinical judgment	3	5.8
Other		

Reported process, but no connection to feedback reported	5	9.6
Issues with work items	2	3.8
Perception of increased Cohesion	2	3.8
Misinterpretation of scales	2	3.8
Attrition Prevention	1	1.9
Member-initiated process	1	1.9
Members rejecting feedback use	2	3.8

Appendix F

Rules for Unitizing Leader Slips and Interview Transcripts

From Stinchfield & Burlingame (1991):

1. The scoring unit consists of an independent clause, standing by itself or occurring with one more dependent clauses.
2. An independent clause is a statement containing a subject and a verb and which grammatically can stand alone as a sentence.
3. In compound and complex sentences, an independent clause can often be distinguished from a dependent clause by the facts that (a) when two independent clauses are connected, the second may be introduced by a coordinating conjunction or a conjunctive adverb (e.g., but, and, for, or, nor) and (b) dependent clauses, which are always used as part of speech, are introduced by subordinating conjunctions or by pronouns such as who, which, or what.
A dependent clause cannot stand alone as a simple sentence. If the meaning of a clause is not clear without reference to another clause, it should be treated as a dependent clause that is part of the independent clause that completes its meaning. The general rule is that when in doubt, do not make a separate independent unit.
4. Some combinations of words without an expressed subject and verb can make complete sentences. These are called elliptical sentences. Examples:
 - a. "Speak." (a command)
 - b. "Good!" (an exclamatory sentence)
 - c. "What?" (a supplement question)
 - d. Therapist: "What room did they give you?" Patient: "The same one I had before."
(patient's utterance is a complete sentence)
 When the unitizer is unsure about the meaning of an elliptical clause and cannot reliably expand it (as with many exclamations and maintenance responses), the clause should not be treated as a distinct scoring unit)
5. False starts do not count as separate units. Example: "And Wednesday night uh I more or less--I didn't high pressure him" (one unit).
6. Utterances lacking some essential feature of a complete sentence because of interruption by the other speaker or a lapsing into silence are considered separate units whenever the meaning is clear. When the speaker has not said enough to make meaning clear, we do not consider his utterance a unit, and we bracket the phrase.
7. Affirmations and negations are not counted as separate units if the speaker goes on to amplify or explain. Example: "Yes, I was happy at home" (one unit). But if the affirmation stands alone, it is separately unitized. Example: "Uh huh./ I was, I was strictly on an ulcer diet."/ (Two units).

8. Phrases like “you know” or “I guess”, when added on to sentences are not considered separate units. Example: “Some very serious things may be happening, you know.”
9. If one independent clause is interrupted parenthetically by another independent clause, each is scored as a separate unit. For example: “And the uh--again I didn’t uh go to any frenzy or have any all-out emotional exhibition on my part, except that I enjoyed it./ But it wasn’t too obvious, I don’t imagine./ Enjoyed it in a passive way, I guess you’d say./” The false start at the beginning is not considered a unit. One unit is: “But it wasn’t too obvious, I don’t imagine.” A second unit is: “Again I didn’t ug go to any frenzy or have an all-out emotional exhibition on my part, except that I enjoyed it...enjoyed it in a passive way, I guess you’d say.” As explained in Rule 8, the phrases, “ I don’t imagine” and “I guess you’d say” are not considered separate units. When two parts of a single independent clause are separated by client speech (usually an interruption or talk over), the unitizer should indicate with arrows and marginal notations that the disconnected parts comprise a single unit.
10. Do not unitize if the speaker is quoting or reading a text, i.e., all material within a reading or a quote will be considered one response unit.

Adapted Rules for the Current Project:

1. If there exist two sentences and the meaning of the second sentence reflects the same meaning of the first sentence, then it is coded as one unit. Moreover, they are coded as one unit if the second statement does not add new meaning. This includes but is not limited to statements of clarification such as “Does that make sense?”
2. When separating two units within a complex sentence, the conjunction (but, and, which, etc.) is included in the second unit.
3. Statements from which no meaning can be pulled are subsumed in adjacent statements that do have interpretable meaning.
4. Statements that serve as context to or a subset of another statement are coded within that statement as one unit unless a portion of that story provides different meaning in relation to the research question.
5. Stories/travelogues are coded as one whole unit as long as they reflect unitary meaning. If separable themes can be pulled from the narrative, then those segments are separated into units.
6. Statements in which the person in essence is saying “on second thought” or “on the other hand” are always separated from the previous thought. Example: “It was still in the positive ranges, it just wasn’t in those really high ranges. /Well, let me take it back-I attended to how deeply they were processing.” The phrase, “let me take that back” is inherently referencing something other than what was just said.

7. Quick responses that are specific to the question just asked are coded as their own unit. Example:

Interviewer: “So I’m just curious if any of your feedback experience you think influenced you or client’s behavior in your non-feedback group?”

Group Leader: “I think that it did./ Especially in the early going when I had the effect of doubting my clinical instinct./”

When a short statement similar to “quick responses” is found at the end of a paragraph, it is almost always included in the previous unit because it is now reflective of the statement preceding it. Example:

Interviewer: “Did you verbally use the GQ/OQ feedback?”

Group Leader: “I did only in times when I really needed to, when the members were in trouble. So, I did use it verbally.” (One unit)

8. Simple processing/reflecting is considered part of the previous statement as one unit, while complex processing/reflections are considered separate units. For example, a simple reflection may be “I wish we had the GQ to help us know what’s going on here. That was one thing. Sort of frustrating to not have it a lot of times” (One unit). A complex process/reflection includes a simple reflection, but adds expansion on the thought, or provides an alternative perspective. For example:

“I wish we had the GQ to help us know what’s going on here. That was one thing./ Sort of frustrating to not have it a lot of times, especially because the group seemed to experience a lot of conflict..” (Two units: expansion)

“I wish we had the GQ to help us know what’s going on here. That was one thing./ Sort of frustrating to not have it a lot of times, although I do accept that the study needs to structure it this way to find an effect of feedback.” (Two units: alternative perspective)

“I wish we had the GQ to help us know what’s going on here. That was one thing./ Sort of frustrating to not have it a lot of times, and made me feel helpless as a leader. (Two units: repeated simple reflections, new meaning added)

9. If the speaker is explaining why, the reason “why” is always included in the previous statement, even if it is a standalone sentence. Example: “Just to reiterate before, I really liked having it. It became this nice way to have discussions with my co-leader too, so that’s almost like a conceptualization” In this unit “because” can be inferred, and therefore is one unit, not two.

10. When the rater sees a quick response followed by a simple reflection, it is coded as one unit. Example:

Interviewer: "Would you be willing to do another set of groups for us?"

Group Leader: "Sure. That'd be fun" (One unit)

11. Sandwich Rule: Disjointed units of the same meaning: separate units that may reflect greater value in that unit of meaning because it is said twice. Example:

"In the group where we did get more GQ/OQ feedback, I don't know, I think most of our focus was on the GQ. /I know we acknowledged the OQ feedback and if there was anything that was really significant, we again would make it a point to bring it up. /We spent the majority of the time with the GQ/OQ feedback." (Three separate units)

Appendix G

Coding Glossaries

GQ USE

Review of Feedback: Any time that the leader (alone or with co-leader) reports looking at the feedback report.

Example: "my co-leader and I reviewed the feedback before group."

Reaction to Scores: Any time that the leader (alone or with co-leader) reports the content, an interpretation, or a speculation of the results of the feedback report. Subcategories include co-leader involvement, direction of reaction (single member, multiple members, group-as-whole), direction of change (positive, negative, no change, mixed change), reaction to scores on all three GQ subscales, and reaction to GQ facet-level scores.

Example: "Looking at the GQ/OQ feedback it seemed that bond rose for a few members, but still went down for others. This feedback was surprising to both Jon and me."

Design Specific Interventions: Any time that the leader (alone or with co-leader) reports actions taken in response to the feedback report that are meant to prepare for use or implementation in the next session. Subcategories include co-leader involvement, direction of intervention (single member, multiple members, group-as-whole), direction of change upon which the intervention is based (positive, negative, no change, mixed change), intervention intended for all three GQ subscales, and intervention based on GQ facet-level scores.

Example: "I wanted group members to experience increased positive bond and positive work and less negative relationship. I spent some time online and in my Yalom group test reviewing principles around helping groups resolve conflict."

Decision to Withhold Feedback: Any time that the leader (alone or with co-leader) reports choosing to not use the data from the feedback report. Subcategories include the involvement of a co-leader and reasons for withholding feedback. Reasons include 1) lack of time available to review or implement feedback, 2) not enough information in the report to create a meaningful intervention, 3) group member preference to not hear about the feedback report, 4) decision to observe or be aware of alerters without actively implementing feedback, 5) withholding based on any number of other extenuating circumstances, and 6) withholding based on only positive scores/improvement.

Example: "We agreed to not take any action, other than paying attention to how C was doing."

Ambiguous Use: Any time the leader reports using the feedback in some fashion, but without specifically stating how it was used.

Example: "I'd say I used the feedback in some fashion every week."

Explicit In-group Feedback Use: Any time that the group leader (alone or with co-leader) brings up scores, trends, or concepts pulled directly from the GQ/OQ feedback

report. Subcategories include co-leader involvement, direction of intervention (single member, multiple members, group-as-whole), direction of change upon which the intervention is based (positive, negative, no change, mixed change), intervention intended for all three GQ subscales, and intervention based on GQ facet-level scores.

Example: “Group leaders noticed JB’s Positive Bond, Positive Work, and Negative Relationship scores were deteriorating. We addressed this directly in group by sharing these data with the group (in paraphrased, summary form (not the actual graphs or numbers)).”

Non-Explicit In-group Feedback Use: any time that the leader (alone or with co-leader) reports using the feedback in the group without specifically stating GQ scores, subscales, or constructs. This might include using the GQ to inform notions about a client, to inform group process, or to inform treatment goals, but without making specific mention of measures or feedback.

Example: “Group members SF’s Pos Work score plummeted, so when she brought up content, group leaders were especially devoted to her using time in the group to address her concerns and do her work.”

Downstream Effects: any time that the leader (alone or with co-leader) reports in-group effects that were either a direct or indirect result of implementing feedback. Subcategories include co-leader involvement, direction of intervention (single member, multiple members, group-as-whole), direction of change upon which the intervention is based (positive, negative, no change, mixed change), intervention intended for all three GQ subscales, and primacy of the effect (immediate, secondary, ultimate).

Example: “The group appeared satisfied with this, and the conversation led to some good exchanges of feedback between Ja and other members.”

Group-Initiated Feedback Use: any time that the leader reports that group members during the session brought up the feedback report. This can be either in the members asking about results, or bringing up how they responded to the GQ, which would then be corroborated by the group leader’s viewing of the feedback report.

Example: “Without group leader intervention, JB interacted with other group members to address his desire to be understood and the barriers to his participation in group.”

OQ: Any time that the leader reports reacting to or using information about the Outcome Questionnaire (OQ).

Example: “I looked at the OQ, but didn’t really do anything with it other than to note that one of our members is feeling significant distress

Attendance: any time that the leader reports attendance of group members, including those who reported alerts on the GQ/OQ feedback report.

Example: “One of the clients who triggered an alert did not return this week, although she had indicated that she was excited for group.”

Filling Out Measures: any time the leader reports in-group discussion about or invitation to fill out the GQ or OQ.

Example: “We talked with the group at the beginning of group and asked them to fill out the GQ as accurately as possible.”

Education: any time that the leader reports educating group members about the GQ, OQ, or related subscales.

Example: “We explained that the questions on the GQ load into 3 factors and described a little bit about what Positive Bond, Positive Work, and Negative Relationship mean.”

Awareness of Alerters: any time that the leader reports being made more aware of group members who are struggling via receipt and review of the GQ/OQ feedback.

Example: “I think in each of the areas for which we received the data, I think for me personally, I felt much more attentive to ways in which I could connect members, ways in which I could draw out what they were hoping to get out of sharing with the group and what their goals were.”

Self-Awareness: any time the leader reports being more self-aware by the feedback.

Example: “Maybe I was a little more self-conscious about assessing what I’m doing, how I’m doing.”

Looking Forward: any time that the group leader reports anticipation of receiving GQ results in a future session.

Example: “I’m looking forward to seeing next week’s feedback & hope that our group members are feeling more positively about their group work.”

Semi-Relevant: any unit that seems to not relate to GQ/OQ feedback, but that cannot be completely ruled out as unrelated. This may include units that use language related to the GQ, but cannot clearly be interpreted as feedback-related.

Example: “In our previous session (10-22), it felt like the members were more bonded with one another and that they were headed in a good direction. “

Irrelevant: units that have no recognizable connection to either the use or value of GQ/OQ feedback. These units often take the form of reporting events in the group session, or other comments on the general group dynamic.

Example: “A left for a significant portion of last session because she felt herself beginning to have a panic attack.” able connection and there is no mention of leader behavior in the unit or in adjacent units where use may be interpretable in context of member behavior.

GQ VALUE

Judgment of the Utility of the GQ: any unit in which the statement of value reflects judgments/opinions/beliefs about the relative usefulness of the GQ. Subcategories are all reasons why the GQ is judged as useful, and include added perspective, enabling focus/engagement, validation, and improved prediction.

Example: “It’s so helpful to see the GQ to be able to gauge how in sync I am with what they are feeling and how they are tolerating what is happening.”

Preference for GQ vs OQ: any unit in which the statement of value reflects a judgment of preference for either the GQ or OQ as a clinical support tool when compared to the other.

Example: “I don’t use the OQ as much. I tend to go straight to the GQ.”

Preference for GQ vs Clinical Judgment: any unit in which the statement of value reflects a judgment of preference for either the GQ or the clinician’s own clinical judgment in conceptualization of the group dynamic. Subcategories include preference for GQ, preference for clinical judgment, and balancing GQ and clinical judgment.

Example: “The GQ for me ended up being this nice way to gauge whether my clinical judgment was truly in sync. Near the beginning I only trusted the instrument, but then later on I realized that there were things about the group that it wasn’t telling me, and I could take it with a grain of salt as I balanced it with what I saw was going on.”

Desire for GQ/OQ feedback: any unit in which the statement of value reflects explicitly or implicitly the group leader’s desire to receive the GQ/OQ feedback. Subcategories include desire for feedback in the feedback study condition and non-feedback study condition.

Example” “One thing I took away from the study is that I was always wanting the feedback in the non-feedback group...like, I was always wondering what so-and-so or so-and-so’s scores were.”

Judgment of Effect of the GQ: any unit in which the statement of value is represented by a judgment of effect that the GQ had on the group leader’s mood, motivation, or way of thinking about the group.

Example: “I think as a group leader, it made me more self-conscious about kind of nurturing group relationships.”

General Evaluations of the GQ: any unit that has a clear value statement, but does not fit within any of the other value categories.

Example: “ So I think I used it more valuably--with more value towards the end. I would lean toward the finishing side of it.”

Appendix H

Additional Analyses

Group Descriptive Statistics

Preliminary Descriptive Statistics for BYU Study Groups

	Absolute Alerts	Progress Alerts	#total/#members producing these alerts	Average attendance per week	Average # of sessions attended
Group 1	33	24	9.50	6.72	8.56
Group 2	46	23	8.63	6.27	7.78
Group 3	9	17	4.33	5	7.11
Group 4	4	10	2.80	3.42	8.20
Group 5	15	22	5.29	5.64	8.86
Group 6	9	7	2.67	5.73	9.14
Group 7	11	2	2.60	7.22	9.00
Group 8	37	20	8.14	7.00	7.00

Preliminary Descriptive Statistics for SUU Study Groups

	Absolute Alerts	Progress Alerts	#total/#members producing these alerts	Average attendance per week	Average # of sessions attended
Group 9	22	15	9.25	4.36	7.33
Group 10	2	5	3.50	5.27	10.67

Preliminary Descriptive Statistics for USU Study Groups

	Absolute Alerts	Progress Alerts	#total/#members producing these alerts	Average attendance per week	Average # of sessions attended
Group 11	1	11	2.40	8.58	11.22
Group 12	23	15	6.33	5.92	9.63
Group 13	2	4	2.00	5.92	12.67

Group 14 17 24 5.13 7.08 11.25

Correlations for Supplemental Variables to Quantify “Acted Upon”

Variable 1: Negative alerts and Use Dimensions

	Sum of negative alerts per session	Dimension 1: Pre-Group Reaction	Dimension 2: Pre-Group Planning	Dimension 3: In-Group Intervention	Dimension 4: In-Group Consequence
Sum of negative alerts per session					
Dimension 1: Pre- Group Reaction	.045				
Dimension 2: Pre- Group Planning	.038	.803**			
Dimension 3: In- Group Intervention	.087	-.219**	-.367**		
Dimension 4: In- Group Consequence	.128*	-.274**	-.320**	.933**	

Variable 2: Negative Alerts and Leader Slip Categories Involving Report of Change

	Sum of negative alerts per session	Change on Reaction to Scores	Change on Design Specific Interventio n	Change on Explicit Feedback Use	Change on Non- Explicit Feedba ck Use	Change on Downstrea m Effects
Sum of negative alerts per session	1	-.153	.266	-.393*	-.056	-.260
Change on Reaction to Scores	-.153	1	.b	.b	.b	.b
Change on Design Specific Intervention	.266	.b	1	.b	.b	.b
Change on Explicit Feedback Use	-.393*	.b	.b	1	.b	.b
Change on Non-Explicit Feedback Use	-.056	.b	.b	.b	1	.b
Change on Downstream Effects	-.260	.b	.b	.b	.b	1

* $p < 0.05$

b. Cannot be computed because at least one of the variables is constant.

** $p < 0.01$

Variable 3: Number of members displaying alerts in a given session and Use Dimensions

Variable 4: Proportion of alerters in a given session to total members and Use Dimensions

	Dimension 1: Pre-Group Reaction	Dimension 2: Pre-Group Planning	Dimension 3: In-Group Intervention	Dimension 4: In-Group Consequence	# of members displaying alerts in a given session	Proportion of alerters in a given session to total members
Dimension 1: Pre-Group Reaction						
Dimension 2: Pre-Group Planning	.858**					
Dimension 3: In-Group Intervention	.422**	.175**				
Dimension 4: In-Group Consequence	.342**	.166**	.928**			
# of members displaying alerts in a given session	-.011	.000	.156**	.183**		
Proportion of alerters in a given session to total members	.006	.007	.094**	.122**	.945**	

** p < 0.01