2016-08-01

Process Feedback in Group Psychotherapy: A Second Look at Leader Implementation of GQ Feedback

Kaitlyn Elizabeth Whitcomb
Brigham Young University

Process Feedback in Group Psychotherapy: A Second Look at Leader Implementation of GQ Feedback

Kaitlyn Elizabeth Whitcomb

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

August 2016

ABSTRACT

Process Feedback in Group Psychotherapy: A Second Look at
Leader Implementation of GQ Feedback

Kaitlyn Elizabeth Whitcomb
Department of Psychology, BYU
Doctor of Philosophy

The current dissertation is a replication of a pilot study and aims to define what it means for group leaders to "act on" feedback from a group therapy process measure called the Group Questionnaire (GQ). Twelve leaders received feedback reports based on group member responses to the GQ after each session. Leaders submitted two sources of qualitative data: brief written session-by-session explanations of feedback use and end-of-semester debrief interviews to describe their experience with the measure. Researchers conducted a qualitative content analysis that yielded 15 categories of leader GQ use summarized by three temporal dimensions. Quantitative analyses were performed to test for variability in leader use. Categories common to both the pilot study and the current study were established, and the two data sets were merged to create one complete data set. A brief questionnaire designed to summarize leader use is introduced, and quantitative analyses were performed to test the relationship between this measure and qualitative findings. Finally, implications of these findings are discussed.

Keywords: process measure, measure-based feedback, Group Questionnaire (GQ)

ACKNOWLEDGEMENTS

This dissertation was an enormous undertaking and I could not have done it without the help and support of a few key individuals. First, I would like to thank my research mentor Gary Burlingame for the countless hours he spent in research meetings helping me plan and execute this project. His support and insight was absolutely vital both to the success of this dissertation, and to my own personal growth. I will be forever grateful for the influence he has had in my life. Second, I would like to thank Kevin Baer and Thomas Childs for agreeing to be my research assistants, and then for sticking with me when they realized just how much time and mental effort it would take to do this work. Third, I would like to thank Joseph Olsen for the guidance he provided for the statistical analyses in this dissertation. Finally, I would like to thank my husband for his constant support throughout this long process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

A growing body of research emphasizes the benefit of using increased measurement and accountability in psychotherapy (Miller, Hubble, Duncan, & Wampold, 2010). This is part of a greater movement toward practice-based evidence (Barkham, Stiles, Lambert, & Mellor-Clark, 2010), a paradigm that emphasizes informing everyday therapy practice with measurement systems. Research has paid significant attention to measures of both outcome and process in order to help therapists evaluate the success of their work and subsequently cater interventions to their clients' needs (Burlingame & Beecher, 2008; Lambert, 2013; Lambert et al., 2003). Outcome measures assess specific symptoms or behaviors expected to change through the course of treatment, such as client distress. On the other hand, process measures assess the therapy process itself, including domains such as quality of interpersonal interactions and client experience of therapist empathy (Burlingame et al., 2006). Providing therapists with questionnaire-based outcome and process feedback is intended to highlight potential obstacles to a positive outcome. This feedback is presented to clinicians in the form of alerts that call attention to various outcome and process domains, including distress levels, social support, the therapeutic relationship, and motivation for change (Lambert et al., 2004). The effect of receiving this feedback has been studied in individual therapy (described below), but similar studies have not yet taken place in group therapy. As a backdrop for this dissertation, I will first outline why research on feedback in group therapy is the essential next step by describing outcome- and process-based feedback research in the individual therapy literature and the state of the feedback research in the group therapy literature. Next, I will review a randomized controlled trial conducted by Burlingame and Beecher that addresses this next step in the group research, and provides context for the current dissertation. Then I will describe findings from a

pilot study conducted before the current dissertation, followed finally by a description of the current study and associated findings.

Measure-Based Feedback in Individual Therapy

Studies using feedback to inform treatment progress and outcome are found much more frequently in the individual therapy literature as compared with the group literature. Several of these measurement systems have been developed starting in the 1990s. The first was an outcomes management system called COMPASS that reported on current well-being, symptoms, and life functioning (Howard, Moras, Brill, Martinovich, & Lutz, 1996). Other systems that attempt to provide continual immediate feedback to clinicians were later created (Barkham, Mellor-Clark, et al., 2010; Kopta & Lowry, 2002; Kraus & Castonguay, 2010; Lambert, Hansen, & Harmon, 2010; Miller, Duncan, Brown, Sparks, & Claud, 2003), allowing therapists to monitor patient outcomes throughout treatment. The Outcome Questionnaire-45 (OQ-45) is an example of an evidence-based feedback measure. It is one of many measures found on the OQ-Analyst (OQ-A), the online software program from which clinicians can administer questionnaires and monitor client progress. The OQ-A is the outcome management system with the most empirical support (Shimokawa, Lambert, & Smart, 2010), and is the only system included in the Substance Abuse and Mental Health Services Administration's (SAMHSA) National Registry of Evidence-based Programs and Practices.

One way that the OQ-A provides feedback to clinicians is in the form of "alerts." There are three main types of alerts: change alerts, absolute alerts, and progress alerts. Change alerts are alerts related to statistically significant score change from the first session to subsequent sessions. They are determined by a reliable change index (RCI), which implies that the change noted is statistically reliable and not merely due to measurement error (Jacobson & Truax, 1991).

Absolute alerts compare client scores with normative sample values that are determined by empirically established cut-off scores. They are triggered when a score changes from one normative range to another (e.g., clinical range, normal range). Each normative population is separated by a cutoff score that reflects the 50th percentile, evenly separating the two normative samples. Progress alerts are related to the progress that is expected to occur in therapy. They are triggered when outcome scores change beyond what can be expected from established normative change, or in other words, optimal recovery curves (Howard et al., 1996). Progress alerts help the therapist know whether a client is progressing toward or deviating from a successful treatment outcome.

The OQ-A system uses a color coding system to identify different types of absolute and progress alerts. White alerts signify clients functioning in the normal range, where additional symptom change is not expected, and thus termination should be considered. Green alerts suggest client change or improvement is within the predicted range, and treatment can continue as is. Blue alerts indicate positive client change that is more rapid than usual. Treatment gains are expected to be maintained, and thus termination should be considered. Yellow alerts signify the rate of change is less than expected and treatment plan alterations should be considered. Red alerts indicate that client change is not consistent with expected progress and that the client is at risk for early termination or treatment failure. This alert indicates that a change in treatment plan is warranted (Spielmans, Masters, & Lambert, 2006; Whipple et al., 2003).

Effects of using outcome feedback provided by OQ-A have been tested in several RCTs and meta-analytic studies, starting in the late 1990s (Berking, Orth, & Lutz, 2006; Harmon et al., 2007; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Lambert, Hansen, & Finch, 2001; Shimokawa et al., 2010). These studies indicate that therapists who receive outcome feedback

have clients that achieve better outcomes, particularly when they are at risk (based on algorithms of optimal recovery curves) for treatment failure (a client leaving treatment with more distress than when they began treatment). Additionally, findings suggest clients of therapists receiving feedback were less likely to experience treatment failure. This is especially important because research has shown that individual therapists are poor predictors of client outcome (Burns & Auerbach, 1996; Hannan et al., 2005).

There is also a growing body of research indicating that process feedback in addition to outcome feedback can influence eventual client outcomes. Again, process feedback includes feedback about therapeutic domains that have been empirically linked to successful outcome, such as therapeutic alliance, client motivation and readiness for change, and client-perceived social support. Tools have also been created to measure this type of feedback. These types of measures provide alerts that inform clinicians about the quality of their therapeutic relationship with rationally derived absolute alerts. Several studies (Harmon et al., 2007; Hawkins et al., 2004; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Whipple et al., 2003) have empirically tested the effect of providing therapeutic relationship feedback to therapists who have clients at risk for treatment failure. For example, in addition to outcome feedback, Whipple and colleagues (2003) provided process feedback regarding therapeutic alliance, motivation for change, and social support to therapists. When therapists received a yellow or red outcome progress alert (an indication that a client was not making expected treatment gains), they administered process measures as a way of investigating why the client was not achieving expected improvement. Results indicated that the addition of process feedback for at-risk clients resulted in longer treatment and better eventual outcomes when compared with clients who received treatment-as-usual (no feedback to therapists) or only outcome feedback. Similar results were found in later

studies (Hawkins et al., 2004; Slade et al., 2008), though one subsequent study produced mixed findings (Harmon et al., 2007). These outcome- and process-feedback findings indicate a helpful addition to routine individual therapy practice that can decrease drop-out rates, improve outcomes, and help therapists inform clinical judgment.

Measure-Based Feedback in Group Therapy

Although there is a large and growing body of research about outcome and process feedback in individual therapy, there is little research testing the effect of this feedback in group therapy. However, given the trend toward including more group-based interventions in the literature, applying this research to group therapy seems especially important (Gallagher, 2009, 2010, 2011, 2012). This is true given that outcome feedback may be even more essential for therapists in group treatment based on the number of patients being treated and the importance of group process factors in individual client change in group treatment (Yalom, 2005).

Over the last two decades research has shown that outcomes in group therapy are equivalent to outcomes in individual therapy (Burlingame, MacKenzie & Strauss, 2002; Burlingame, Strauss & Joyce, 2013; Fuhriman & Burlingame, 1994; McRoberts, Burlingame & Hoag, 1998). Additionally, in a study comparing over 13,000 clients on the OQ-45 in either group or individual treatment, outcomes were statistically equivalent (Burlingame et al., 2015). These results support an empirical rationale for testing the OQ progress feedback system in a group format (Burlingame & Beecher, 2008).

Additional rationale for this research is gained through a study that indicates that group leaders, like individual therapists, are poor predictors of client outcome. Chapman et al. (2012) replicated research that has shown individual therapists are unable to predict a client's outcome status accurately (Hannan et al., 2005). They asked group leaders to estimate whether members

were improving, deteriorating, or not experiencing change on the OQ-45 as measured by the total distress score at three points in treatment (3rd, 6th, & 9th sessions) compared with baseline assessments of distress on the same measure. As is true with individual therapists, group leaders predicted their clients would improve more than they actually did. Taking this research together, Burlingame and Beecher proposed that feedback to group leaders should have the same effect in group therapy that individual therapists see in individual therapy with feedback.

There has also been significant growth in the group literature on process measures. This is not surprising given the research suggesting that the therapeutic relationship in group, or cohesion, is a predictor of patient improvement and decreased drop-out rates (Burlingame, Fuhriman, & Johnson, 2002). Further, in a recent meta-analysis (Burlingame, McClendon, & Alonso, 2011), authors included 40 studies to test the association between cohesion and outcome and found a significant positive relationship ($r = 0.25$, $p < .05$, $SE = .04$), which is a medium effect. Given the correlation between group therapeutic relationship and outcome, providing information to group leaders about that relationship in the form of a group process measure is important. However, the large number of measures developed to define and operationalize the group therapeutic relationship (Evans & Dion, 1991; Fuhriman, Drescher, Hanson, Henrie, & Rybicki, 1986; Gully, Devine, & Whitney, 1995; Kivlighan, Multon, & Brossart, 1996; Moos & Humphrey, 1974; Mullen & Copper, 1994; Silbergeld, Koenig, Manderscheid, Meeker, & Hornung, 1975) has resulted in confusion about the best way to capture this construct.

In order to address the confusion in the literature, leaders of the American Group Psychotherapy Association (AGPA) created a taskforce of top group researchers to update its Clinical Outcome Results Standardized Measures (CORE) battery of recommended outcome and process instruments. The taskforce was asked to review the literature and determine which

outcome and process instruments had the most empirical support for group treatments. They identified four measures, each of which assesses a different group relationship construct (cohesion, group climate, therapist empathy, and working alliance), and called the combination of the measures the revised CORE battery, or the CORE-R (Burlingame et al., 2006; Strauss, Burlingame, & Bormann, 2008). However, with all four measures included in the battery, it was determined that the length of the CORE-R was too long to be practical for clinical practice. This led to a subset of taskforce members who set out to determine if they could empirically reduce the length of the measures recommended in the CORE-R, while still capturing therapeutic relationship.

In 2005, Johnson and colleagues successfully reduced the length of the measures by using structural equation modeling on a large sample of CORE-R data (responses from 662 group members from 111 clinical and nonclinical groups were included). Authors found that the four constructs captured by the CORE-R measures (cohesion, group climate, therapist empathy, and working alliance) were indeed correlated, but that they did not all load onto one construct. Instead, using exploratory factor analysis, they found a three-factor latent structure (positive bond, positive work, negative relationship) fit the data best in both clinical (66%) and nonclinical samples (59%). Since then, these results have been replicated in different settings such as hospital-based groups in Germany and Switzerland and a Norwegian psychodynamic outpatient setting, in which the three-factor model was also found to be a good fit (Bakali, Baldwin, & Lorentzen, 2009; Bormann & Strauss, 2007). In addition to the three-factor model, Johnson also found support for structural or directional aspects of the group relationship: member-member, member-leader, and member-group. The items remaining were then reduced further to a 30-item instrument called the Group Questionnaire (GQ) with good internal consistency reliability (see

Table 1.1; Krogel et al., 2013) which was later replicated in hospital-based groups in Germany (Bormann, Burlingame, & Strauss, 2011) and U.S. university counseling centers (Thayer & Burlingame, 2014). Additionally, the validity of the constructs being measured by the three-factor model is addressed in two studies that show high criterion-related validity with instruments that assess the same constructs (Bormann et al., 2011; Thayer & Burlingame, 2014; see Table 1.2). This measure and the research discussed provide a solution to the confusion in the research regarding the group therapeutic relationship, which in turn allows researchers to begin using this measure-based feedback in group psychotherapy.

Table 1.1

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 (2009).

Outline of Burlingame and Beecher Randomized Clinical Trial

This dissertation relies on data collected from an RCT conducted by Burlingame and Beecher that aims to fill the gap in the group feedback literature. This RCT is a multi-site study that tests the effects of outcome and process feedback in group treatments at three Utah university counseling centers: Brigham Young University (BYU), Southern Utah University (SUU), and Utah State University (USU). To provide context for my dissertation, I will briefly describe this study, hereafter referred to as the parent RCT study.

Table 1.2

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.
**$p < .01$.

Rationale and Hypotheses

Burlingame and Beecher's study intended to replicate the individual therapy feedback RCTs reviewed previously in a group setting. It aimed to experimentally manipulate outcome and process feedback using three conditions. The first was an archival no-feedback condition taken from Lambert's study in which the OQ-45 was administered to clients before each group session but no progress feedback was provided to the therapist (Shimokawa et al., 2010). This arm of the study reflected treatment as usual in that therapists were not receiving progress feedback.

In the second condition, group leaders received outcome feedback alone. Group members completed the OQ-45 before each group session. The following week, consistent with Lambert's procedures in his RCTs, the group leader was provided with a progress feedback report (see Appendix A) created for all group members before their next group session. That is, the group leader received progress feedback that was delayed one week because of the logistics of preparing a report for the leaders. The feedback delivered was the same as the feedback used in Lambert's RCTs described above (i.e., change, absolute, and progress alerts) with one exception (see Table 1.3). The alerts generated by the OQ-A software program were manually combined into a single report including feedback for all group members. Providing the OQ feedback on a single page was done for simplicity's sake, so that leaders would not have to look at multiple pages of reports produced by the OQ-A software for each group member (see Appendix C).

Lastly, the third condition of the study tested the effect of combined outcome and process feedback using the OQ-45 and GQ clinical reports generated by the OQ-A software, combined into one report (see Appendix B). The therapeutic relationship feedback was provided in the form of three newly developed alerts for the GQ (relative, absolute, and progress; see Table 1.3). These alerts were developed for all three subscales of the GQ by Burlingame and Beecher and are a product of findings from Chapman et al. (2012) combined with years of clinical use and training with the GQ. Using multiple alerts was different from the single absolute alert provided by Lambert and colleagues in their RCTs. However, this type of feedback better reflects the requests of group clinicians that Burlingame and colleagues trained on the GQ at the AGPA conference for the past several years. The subscale relative and absolute alerts were on the first page of the OQ/GQ clinician report while the progress alerts were graphically depicted for all

group members on the second page (see Appendix B). Lastly, item level and facet level information for negatively alerting clients was included on later pages of the report.

Table 1.3

Comparison of Parent RCT Alerts with Past Individual Therapy Research

Study	Outcome Alerts	Clinical Support Alerts
Lambert research based upon individual treatment	Administered before each session	Administered when therapist receives an off-target progress alert
	Change alerts reflecting reliable change in score from intake—RCI show whether client has a reliable improvement or deterioration in total score since intake	**Change alerts** not relevant since scale administered only when progress alerts issued
	Absolute Alerts based on passing clinical significance cutoff score into clinical or nonclinical range of distress	**Absolute Alerts** at subscale and item level based upon rational cut scores
	Progress Alerts where client change is compared to normative change trajectory • On track/normal = Green, White • Sudden gain = Blue • Off track = Red, Yellow	**Progress Alerts** not relevant since scale administered only when progress alerts issued
Burlingame & Beecher RCT based upon group treatment	Identical to above	Administered at the end of every session
		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
		Absolute alerts at subscale and item level based upon rational cut scores
		Progress alerts—descriptive only, client subscale scores plotted with all group members

In the parent RCT, each group leader simultaneously ran two groups—in one group the leader received GQ/OQ feedback, and in the second group they only received OQ feedback. Each group leader committed to run a minimum of four groups to provide a test of both leader and group effects, though six of the 16 leaders were unable to comply with this commitment. However, every leader ran at least two groups. The study posed three primary hypotheses regarding groups in which the leader received GQ feedback versus no GQ feedback:

1. Leaders who receive GQ absolute and/or relative alerts and act on these (assessed by weekly leader GQ reports) will 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 do not receive GQ alerts.
2. Leaders who receive GQ absolute and/or relative alerts and act on these (assessed by weekly leader GQ reports) will have clients posting higher levels of group attendance (i.e., fewer dropouts).
3. Leaders who receive GQ absolute and/or relative alerts and act on these (assessed by weekly leader GQ reports) will have clients who report greater symptom distress reduction on the OQ-45.

In addition to these three main GQ hypotheses, the parent RCT sought to replicate the outcome feedback findings using the OQ-45 as the third archival arm. The corresponding hypothesis utilized the second and third arms of the study, and states:

4. Leaders who receive OQ alerts and act on these (assessed by weekly leader GQ reports) will have clients with fewer treatment failures and greater symptom distress reduction than clients whose therapist received no OQ alerts.

Parent Study Method

Participants. The parent RCT included individuals participating in group therapy at the student counseling centers at BYU, SUU, and USU. There were a total of 58 groups included in the study, including process groups, psychoeducational groups (i.e., generalized anxiety, autism spectrum disorder) and specific focus groups (i.e., trauma, sexual concerns). Forty-nine of the 58 groups lasted only one semester while nine of the groups included in the study were longitudinal and ran more than one semester. A group was considered longitudinal if it lasted for two semesters or more and if the majority of the members attending group in later semesters were part of the group the previous semester. Eight of the nine longitudinal groups met for two semesters, while one met for four semesters. Each group was co-led by one licensed psychologist and typically one trainee or intern.

Sixteen group therapists were included in the study as primary group leaders. Group therapists at college counseling centers generally espouse a number of theoretical orientations. Those included in our study claimed to practice cognitive-behavioral (including acceptance and commitment therapy and mindfulness therapies), humanistic (including existential, interpersonal, emotion focused, constructivist, and modern gestalt therapy), psychodynamic, systems, and integrative therapeutic approaches. Group leaders committed to participate for two semesters over the course of 30 months and run a minimum of two pairs of GQ-feedback/no-GQ-feedback groups to statistically model group and leader effects, though six leaders were unable to keep this commitment and only contributed one pair of GQ-feedback/no-GQ-feedback groups. As this was a naturalistic study, the groups they ran were a part of their normal group caseloads.

There were 430 total participating group members. Each group had between five and 12 members with an average of 7.7 members per group with presenting problems that were

representative of college counseling center populations. These included relationship issues, anxiety, depression, self-esteem, adjustment, impulse control, stress, substance abuse, social skills, eating disorders, self-mutilation, and pornography. Several group members ended up contributing data for more than one semester as they sought group treatment multiple times during the study.

Those who agreed to participate in the study were required to attend group therapy as their 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, and Barlow (2008). Other inclusion and exclusion criteria for participants are described below.

Inclusion criteria:

- Willingness to commit to at least four sessions of group treatment
    - Of the 430 total clients enrolled in the study and who originally agreed to attend at least four sessions, 56 only attended three or fewer sessions
- Willingness to complete GQ and OQ on weekly basis, even if their group leaders did not receive GQ feedback
- Willingness to have group be their primary mode of treatment; this was to ensure they were committed to the group as a primary vehicle for change

Exclusion criteria:

- No e-mail address

All research participants were recruited during their first group session. Group leaders were recruited through research meetings and announcements made at faculty meetings. All group members received cash as compensation for research participation based on the number of

measures they completed. Group leaders did not receive any compensation aside from the benefit of weekly feedback reports for their group assigned to the feedback condition.

Procedures. Upon requesting services, potential clients completed an OQ-45 and reported basic demographic information via the counseling center's online intake program. Upon completing the OQ-45, potential clients were referred to potential study groups as a part of normal clinic scheduling. During their first session of the group, members were informed of the nature of the current study, including the benefits of group therapy, benefits of study participation, and potential risks. Group members were also informed of the monetary incentive they would receive for their participation. Those who agreed received $10 upon consenting to participate, as well as an additional $5 for every set of weekly OQ and GQ they completed thereafter. Members were paid the first $10 up front, and the remainder was paid in cash at the end of the group based on the number of OQ/GQ data sets they provided (e.g., if they provided five additional sessions of data, they received $25). Members who completed an OQ *and* GQ for all possible sessions attended received a bonus of $20.

Clients who indicated an interest during intake were referred to a study group with a participating leader. During the first session of the group, either the group leader or a research assistant involved in data collection described the study using a script (see Appendix D), determined eligibility based on the above inclusion and exclusion criteria, and obtained informed consent. The person explaining the study made sure to emphasize the importance of completing both an OQ before the beginning of group *and* a GQ at the end of each group. Members were told that their group leader would use the OQ irrespective of group assignment to guide treatment. Of the 455 group members who were invited to participate in the study, 430 opted to participate. Members were informed that they could complete the GQ in one of two ways: using

a tablet or computer at the counseling center immediately after the group, or online within the week after the group session.

After the group began, new members were allowed to join the group and the study for up to four weeks after the initial session. Group leaders were instructed to remind their group members to complete their GQ at the end of each session. As an extra precaution, e-mail reminders were also sent to those members who failed to complete a GQ three days after the most recent group session.

Group leaders were oriented to the study by a member of the research team before they ran their groups. They were also trained in how to interpret the information on the GQ feedback report, called the "Weekly GQ Feedback Report" (see Appendix B). The first page of this report provided leaders with absolute alerts on each of the GQ subscales (positive bond, positive work, and negative relationship) based on percentile-based cut scores. The second alert on the first page was a relative alert that reflected change alerts for each GQ subscale from the previous week's score. Members who showed statistically significant nontherapeutic or therapeutic change since the last session were flagged. Change alerts identified group members who had significant subscale-score change since their last session (clients who deteriorated or improved more than one Reliable Change Index unit; see Figure 1). Group leaders were also given group progress alerts. Because there is not currently an algorithm to predict whether or not a client is on track, clinicians were provided with a graph of each client's progress for each subscale across the three GQ subscales instead (see Appendix A). Group leaders were also given subscale- and item-specific data for those clients who had negative alerts. Finally, as part of the GQ report, leaders were asked to report how they used the feedback each session by responding to a short prompt included in the GQ feedback report (i.e., leader slips). Leader responses were guided by the

following instructions: “List actions (if any) that you took based upon last week’s GQ feedback. List any specific member targeted.” Group leaders then either printed the report, filled out the slip, and returned it to the researchers, or they sent their responses via e-mail. This type of GQ report is not currently available to clinicians using the OQ-A (see Appendix C for an example of the current GQ report as produced by the OQ-A), so the reports were compiled manually each week by four graduate students (Sean Woodland, Kaitlyn Whitcomb, Michael Williams, and Elyssa Hunsaker) and one undergraduate student (Jordan Rands).

Defining “Acted Upon”

The parent RCT included hypotheses that intentionally distinguished between leaders who “act upon” GQ feedback and those who do not. (The main hypotheses each started with the phrase: “Leaders who receive GQ absolute and/or relative alerts **and *act on* these**…”) This was done because it was believed that in order for the GQ to influence the outcomes being studied, it had to be utilized. Burlingame and Beecher proposed that leaders who received GQ feedback and then deliberately intervened 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 failed to review the GQ feedback or chose not to act on the feedback provided. Thus, in order to accurately interpret these hypotheses, it is essential to first understand what it means to *act upon* feedback, and then use this understanding to create a fidelity check to be used in the interpretation of the hypotheses. Thus, during the course of the study, we asked the group leaders to report on the way that they acted on the GQ on a session-by-session basis (weekly leader slips) and at the end of the semester by reflecting on their actions over the entire course of the group (leader debrief interviews). Two dissertations aim to qualitatively analyze this information in order to determine what it means to use GQ feedback and how best to use this information as a

fidelity check. The first is a pilot study, designed to analyze a portion of the data and establish initial impressions of what it means to use GQ feedback (Woodland, 2014). The second is the current study, which aims to utilize the same methods employed in Woodland's study, address challenges identified during the pilot, build upon findings, and use a new sample of the data to qualitatively describe what it means to act upon GQ feedback. I will first summarize the findings from the pilot study to give context for the current study, and then provide a detailed description of the methods and procedures used in the current study.

The Pilot Study

The pilot study was led by Sean Woodland, with Kaitlyn Whitcomb as a rater for the qualitative analysis and Gary Burlingame acting as auditor and supervisor of the project. The pilot included a little less than half of the total data from the parent RCT (13 out of 30 groups that received GQ feedback; see Figure 1 for population details). The procedures used to collect leader slips and debrief interviews in the pilot are the same as the procedures used in the current study. As such, the details will be included later on in the description of the methods of the current dissertation. The same is true for the details of the qualitative content analysis. However, in order to give context for understanding the findings of the pilot, I will briefly describe the steps of the qualitative content analysis here, with further detail included later on.

In order to gain an understanding of what it means to act on GQ feedback from the leader responses we collected, we followed steps for qualitative content analysis outlined by Margrit Schreier in her book *Qualitative Content Analysis in Practice*. Briefly, the process starts with unitizing the data, or separating the text into short sections that only include one thought or concept. Next, each individual unit is then coded, or in other words, the meaning of the unit is

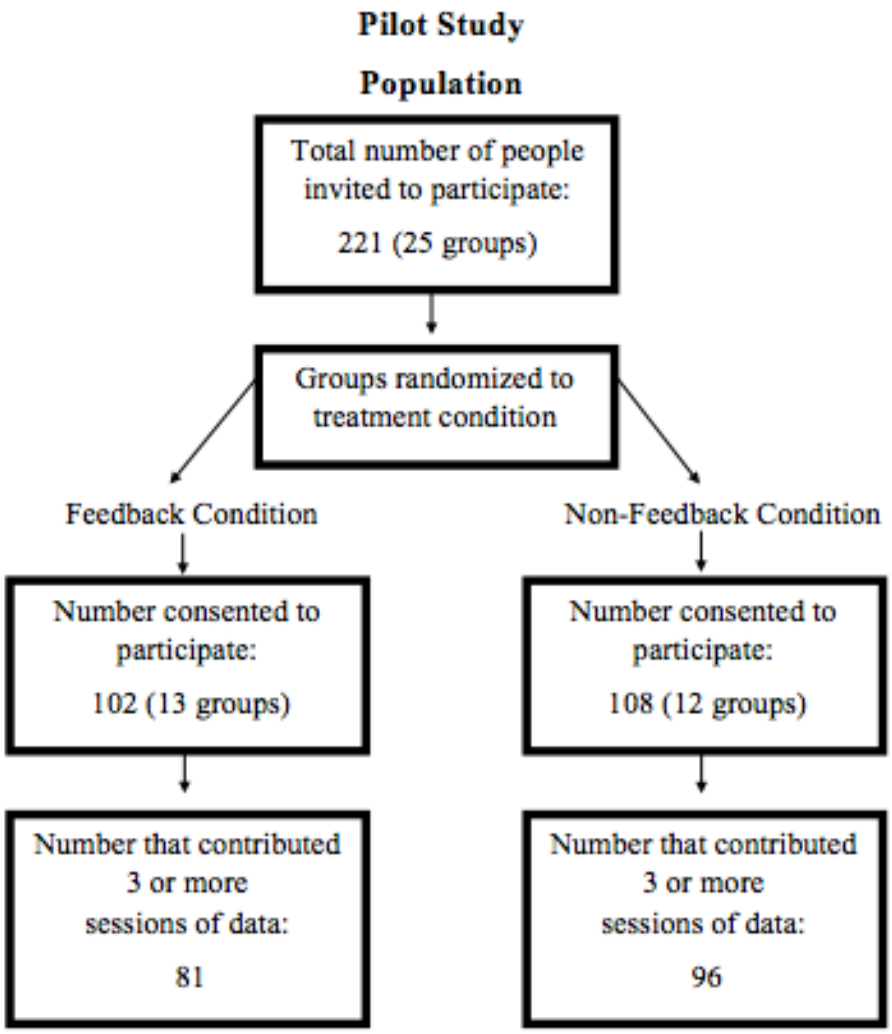


Figure 1. Population details for the pilot study.

interpreted and named. Individual units are then combined based on similar meaning into categories, and differences within those categories are identified and included as subcategories. This coding was directed by an a priori definition of "acted upon" established through a pilot thematic analysis of a small sample of data. Specifically, Woodland examined 52 leader slips prior to the formal research study and found two general themes. The majority of leader slips described how therapists *used* GQ feedback, and a smaller portion addressed its overall *value*.

Findings

Findings from the Woodland pilot were based on 116 leader slips and 11 debrief interviews. Categories were identified within the a priori themes of *use* and *value*. The data was coded first for use, followed by a second and separate pass through the data coding for value (see Table 1.4 and 1.5 for frequencies; see Appendix H for glossary).

Table 1.4

GQ Use Categories from Content Analysis

Category	Leader Slip Freq.	% of Total Units	Debrief Interview Freq.	% of Total Units
Review of Feedback	17	5.0	24	6.2
Reaction to Scores	77	22.6	133	34.1
Design Specific Interventions	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 Use	41	12.0	30	7.7
Non-Explicit 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	2.0	NA	NA
OQ Use	10	2.9	NA	NA
Self-Awareness	NA	NA	1	0.03
Looking Forward	3	0.9	NA	NA
TOTAL	341	100	390	100

Table 1.5

GQ Value Categories from Content Analysis

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

Use.

Initial Use. Findings from Woodland's study identified four main categories that define group leaders' GQ-feedback use. The first category identified was **Initial Use**, which describes the first basic steps of acting upon feedback. The first subcategory within Initial Use was named **Review of Feedback** and was made up of units in which leaders simply described looking at the feedback. For example, one leader said, "My coleader and I reviewed the feedback before group." While simple and perhaps a given when considering feedback use, reviewing feedback was an action that leaders frequently reported. More importantly, some group leaders in the GQ feedback condition indicated that they seldom or never engaged in a "review of feedback" thereby creating variability in this most basic act of using feedback. The next subcategory identified within the Initial Use category was called **Reaction to Scores**. This category captured units in which leaders reported their thoughts or interpretations of the data after they reviewed it. For example, one leader said, "Looking at the GQ 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."

Pre-Group Decisions. The second category of use created in the pilot study was **Pre-Group Decisions**, which referred to feedback-informed decisions made about the group and/or its members and leaders *before* the next group session. The first subcategory within Pre-Group Decisions was **Designing Specific Interventions**, which was made up of units in which leaders described specific things they planned to do in group based on the feedback. For example, one leader said, "We planned to be on alert for opportunities to intervene with group member JB, particularly in ways that might improve his positive bond." The second subcategory of Pre-Group Decisions was named **Decisions to Withhold Feedback**, which included leader descriptions of their choice to leave the feedback out of group based on either clinical judgment or other reasons. For example, one leader said, "We agreed to not take any action, other than paying attention to how C was doing," while another wrote, "The data was spotty this week, so we didn't do much with it."

In-Group Use. The third category of use leaders reported was **In-Group Use** of feedback, which included units in which leaders described the way that they used the GQ feedback in session after reviewing the feedback report. The first subcategory identified in this category was called **Explicit Use**. Units in this subcategory described in-session instances in which leaders referred directly to feedback observed in the feedback report. For example, after noticing that one member's scores were deteriorating, a group leader said, "We addressed this *directly in group* by sharing these data with the group in paraphrased summary form—not the actual graphs or numbers." Another leader wrote, "I decided to bring a copy of the feedback with me to the session. . . . I shared that most members seemed to feel more bond and work and less negative relationship." In other words, leaders used the feedback explicitly and members were made aware that leaders were referring to findings from the feedback. The second subcategory of

In-Group Use was **Non-Explicit Use** in which leaders based in-session interventions on GQ feedback, but did not directly refer to the findings with the group or specific members. In these cases, the members may not have known that the leader was acting on feedback. For example, one leader said, "Noting low member-member and member-group to be the source of their alert status, I focused particularly on group cohesion." Another wrote, "I was aware that one member's scores were dropping and made effort today to help her express more emotion and engage with others."

Evaluation of Effects. The final category of use from the Woodland pilot was **Evaluation of Effects** of using feedback. This included leader reflections about member reactions, group discussions, subsequent leader actions, and evaluations of the effects of the feedback use. More specifically, this category encompassed reactions to one of the first three categories of use. For example, one leader used Explicit-Use with his group, noting variability in scores, then noted, "This seemed to enable the group members to jump right into a discussion of the previous session and what they wanted from each other."

Value.

Findings from the Woodland pilot study that spoke to the "value" of GQ feedback were organized into six subcategories. First, **Judgment of the Utility of the GQ** included units that evaluated the usefulness of the GQ as a measure of the therapeutic relationship, as opposed to describing how it was used by group leaders. For example, "[GQ feedback] seemed extremely helpful as a starting point for a discussion of feelings in the group that I'm not sure I would have been aware of otherwise." The second subcategory of value was **Preference for GQ vs. Clinical Judgment**. This category captured units in which leaders compared their own clinical judgment with what they were getting out of the GQ. In some cases, clinicians would indicate valuing their

own clinical judgment above what the GQ could offer, and some would report feeling their own judgment was less accurate than the GQ. For example, one leader stated, "I will say that it was a weird balance between really coming to value [GQ feedback], but maybe having a piece that I was lacking but also being able to not let it override me."

The third value subcategory was **Preference for GQ Over OQ**. As OQ feedback was included in the feedback report provided to leaders, some provided feedback comparing the two. For example, "I didn't use the [OQ] much—I tend to go straight to the GQ." The fourth value subcategory was called **Judgment of the Effect of the GQ** and consisted of units like, "The effect it had on me as a leader; I thought it was empowering—it gave me an opening to intervene with people that I thought might be struggling." The fifth value subcategory was **Desire for GQ Feedback** and consisted of units in which the leaders described wanting the feedback. For example, one leader said, "I'm looking forward to seeing next week's feedback and hope that our group members are feeling more positively about their group work." Lastly, the sixth subcategory of value was simply **General Evaluations of the GQ**. This category was made up of units that clearly included an element of value, but did not fit in any of the other categories. For example, "My group was going through a lot of conflict and I sat down with the feedback and got more serious about it and realized if we're still getting alerts, I need to take that more seriously and do something based on that."

Quantification Scheme

As evidenced by the brief summary of the pilot study above, the analysis provided rich information and new insight into what it means to act on GQ feedback. However, in order to use this information to create a mediation variable for the parent RCT, a remaining challenge was to transform the qualitative data into a quantifiable metric. Woodland piloted one quantification

method by averaging three rankings into a total leader "acted upon" score. More specifically, he used a leader rank based on (a) the total "use" units identified, (b) response to negative absolute and relative alerts on the GQ (alerts that are considered a red flag to leaders that something has gone wrong and action is suggested), (c) the number of "ratable units" provided by each group leader, or in other words, the units that were relevant to acting upon the GQ. However, there were some limitations and challenges to this way of measuring acting on feedback. Specifically, the first rank used total use units for each leader identified in the qualitative analysis. This can be problematic because in the debrief interviews, this strategy might indicate that more verbose leaders were higher users of the GQ because increased words increase the likelihood of accumulating units that represent use. Similarly, with the leader slips, leaders who wrote more would be counted as higher users, when this might not actually be the case. The third "ratable units" ranking was included to address the limitation of the first ranking. However, taking into account how relevant the content was to use or value does not necessarily indicate a higher level of acting on the GQ.

The Current Dissertation

The purpose of the current dissertation is twofold. First, it is a systematic replication of the Woodland study, with some alteration, conducted to increase construct validity of what it means to act on GQ feedback. Replications verify what has been found in past research in order to provide additional credibility to those research findings. Further, as is the case with all systematic replications, conducting the same study with limited alteration helps to enhance the validity established in initial studies, or in this case, construct validity. This replication implemented the same methods as the pilot, but on a different sample from the parent RCT, and with new raters. These changes were essential to establishing additional construct validity by

considering input from different leaders on the same topic, and by gaining new insight from blind raters. Additionally, basing conclusions on multiple samples with distinct pairs of raters will help to reduce the possibility that findings are established in error (Kazdin, 2003). In other words, if we find the same findings twice using a different sample and new raters, it is more likely that we will have accurately captured the construct of GQ use than if we had based our findings on only one study.

The second purpose of the current dissertation, in conjunction with the Woodland pilot, is to provide a fidelity check to verify that the manipulation (the GQ feedback report) being tested in the parent RCT was actually being implemented. While there are several studies that use questionnaire-based feedback (Harmon et al., 2007; Hawkins et al., 2004; Lambert et al., 2001; Slade et al., 2008; Whipple et al., 2003), none systematically check to see whether or not the feedback was actually used, or how it was used. This assumption that the feedback is being used, or that it's being used the same way, introduces noise into the results of the study because of the differences between the way people implement or use the feedback. We believe that in order to truly test the effects of a feedback system, it is essential to know the different ways it is being used and take that into account when interpreting findings, which is why the Woodland pilot and the current study are so important.

As shown in Table 1.6, there are several similarities and differences between the methods of the pilot study and the proposed study. More specifically, similarities include the source of the data, the guiding questions, the qualitative research strategy, and the data collection methods, and the number of leaders included. Differences between the two studies include the amount of data used, the number of groups from which data was collected, the different types of groups

included, the level of insight raters had going into the study, and the information used to develop the quantification scheme.

Table 1.6

Similarities and Differences Between the Woodland Study and the Current Study

	Woodland	Whitcomb
	Similarities	
Data Source	Burlingame and Beecher RCT	
Guiding Questions	What does it mean to "act upon" GQ feedback	
Research Strategy	Qualitative content analysis (Margrit Schreier, 2012)	
Data Collection	Open ended responses through leader slips and debrief interviews	
Number of Leaders	11 Leaders	12 Leaders
Number of Leaders that Were the Same Across Studies	7 Leaders	
	Differences	
Data Sample	116 Leader Slips, 11 Interviews	144 Leader Slips, 20 Interviews
Number of Groups	13 Feedback Groups	17 Feedback Groups
Group Types	2 (General Process, Generalized Anxiety Psychoeducation)	4 (General Process, Trauma, Autism Spectrum Group, Sexual Concerns Group)
Number of Leaders that Were Different	4 Leaders	5 Leaders
Raters	Two blind raters for unitization and coding	Three raters for unitization, two blind raters for coding
Quantification	Used qualitative data to develop leader rankings	Used questionnaire findings based on qualitative categories

An important difference between the current study and the Woodland pilot is the inclusion of value as an a priori definition of "acted upon." Woodland included value because in the small sample he looked at before conducting the formal analysis, it seemed like an important theme. However, after analyzing a larger portion of the data in a more structured way, we found weak support for including value as a major definition. Also, it became evident that including value as a main definition of "acted upon" does not help us achieve our main goal of formulating a parsimonious quantification scheme. First of all, after analyzing the 1,182 units included in the Woodland pilot, only 85 units were found to explicitly describe value, while 731 were found to describe use. Also, of the 85 value units, 44 of those 85 were also coded for use, or in other words expressed both value and use in one unit. This leaves only 41 units that were unique to value. Given this information, use accounted for the bulk of findings of what it means to act on GQ feedback, and adding value did not add significantly to our understanding. Accordingly, we did not include value as a main a priori definition of "acted upon," but rather decided to let any themes related to value emerge organically from the data without imposing it as an a priori definition.

Hypotheses

We hypothesized that categories and subcategories similar to those identified in the pilot study would be identified in a new sample of GQ feedback groups. That is, the four categories identified in the pilot will be similar to the categories identified in the current dissertation. However, we also hypothesized that unbiased raters who have not been exposed to the Woodland pilot may identify new "acted upon" categories and subcategories.

Method

Participants

In the parent RCT, there were 58 groups, 30 of which were in the feedback condition. Thirteen of those 30 were included in the pilot study, and 17 were included in the present study, 16 of which provided qualitative data (see Figure 2 for population details). These groups met for approximately 8–14 sessions with between five and 10 members. The type of groups included were 12 general process groups (one of which was a couples process group, and one of which was a men's process group), two trauma groups, one psychoeducation group specifically for autism spectrum disorder, and two sexual concerns groups. The Woodland pilot also had 12 general process groups, but only had one other type of group (generalized anxiety psychoeducation). There were 12 group therapists that led these groups, seven of which also participated in the pilot study, and five of which only contributed groups for the current study. The members participating in these groups were 101 college students.

Procedures

There were two data sources that were used to qualitatively understand how leaders acted upon GQ feedback. The first source was from the open ended responses (i.e., leader slips) that were solicited from each group leader as part of their GQ feedback report each week (see Appendix B). Leaders were prompted to "List actions (if any) that you took based upon last week's GQ feedback. List any specific member targeted." Group leaders then either printed the report and handed it in to research assistants, or filled out their responses electronically and sent them via e-mail. These slips were then organized in an excel spreadsheet for future qualitative analysis. There were 116 leader slips collected for the pilot study, and 144 in the current study.

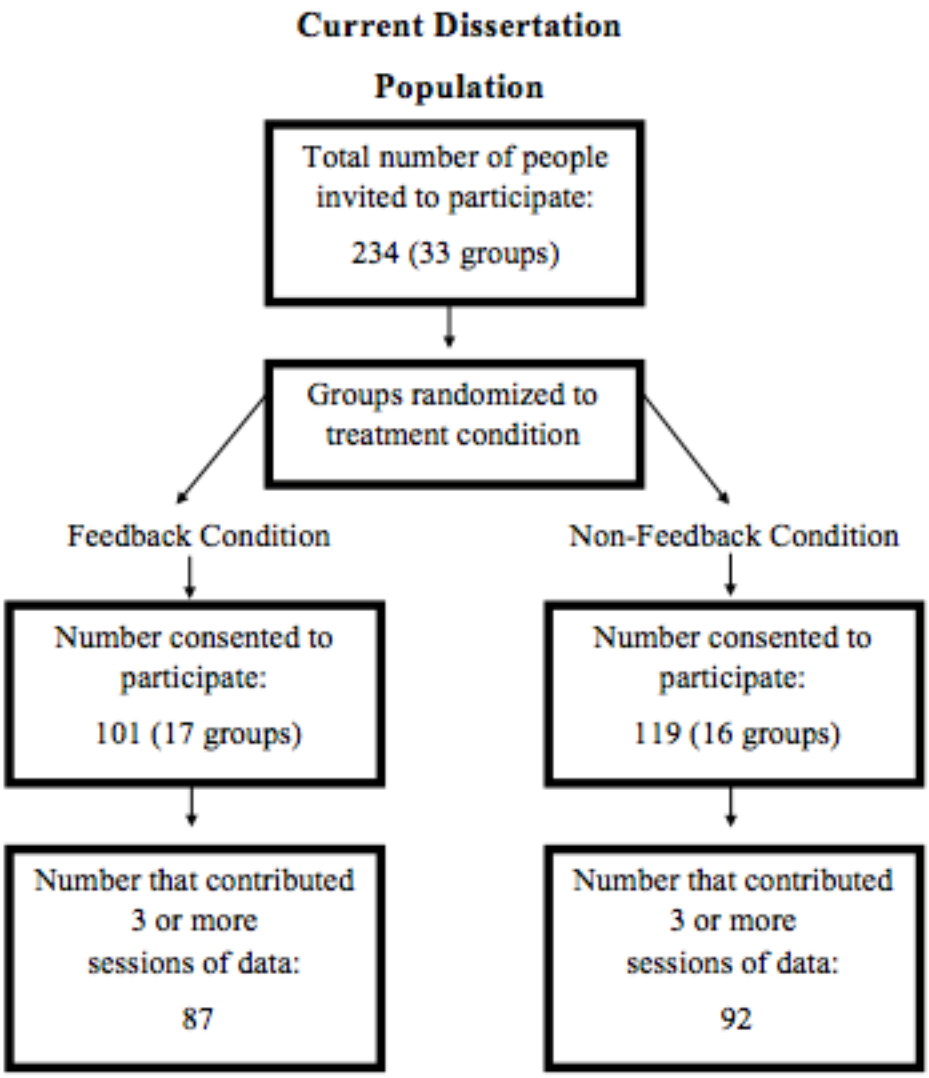


Figure 2. Population details for the current dissertation.

During the data-collection phase of the RCT, we noticed that there was significant variability in leader consistency in turning in leader slips. Given this variability, we wanted to include a second source of qualitative information so that each leader's habits of acting on GQ feedback could be captured. Thus, we included a semi-structured debrief interview (see Appendix E) that leaders participated in after they concluded their semester-long group. These debrief interviews were recorded and transcribed for later analysis. There were 11 interviews included in the pilot study. The current study included 20 debrief interviews, 14 of which were

completed verbally, and six were completed in written format following the same structured interview.

In addition to the two qualitative data sources, a third and new source of data was collected at the end of the parent study to help determine how leaders acted on GQ feedback. Based on the main categories of use identified in the pilot, a brief questionnaire (the GQ Use Questionnaire) was developed to assess the degree to which each leader implemented each area of use established in the pilot (see Appendix F). Specifically, the questionnaire asks leaders to rate on a six-point Likert scale the percentage of sessions in which they implemented each type of use for each group they ran. The areas of use assessed included reviewing feedback, reacting to or noticing scores, educating members about the GQ itself, using the GQ to guide specific interventions, specifically mentioning the GQ data in the group, using the GQ to inform in-group interventions *without* specifically mentioning the GQ scores or subscales, and deliberately choosing not to bring the feedback in to the group. This questionnaire was developed and administered as an attempt to capture a global "acted upon" score without the effects of certain limitations present in the qualitative data. For example, variability between leaders in how many leader slips were turned in and the volume of text contributed in the debrief interviews could skew the representation of a leader's level of feedback use, while the questionnaire is not subject to the same limitations.

Qualitative Content Analysis Method

Content from the leader slips and debrief interviews were qualitatively analyzed by trained raters (Kaitlyn Whitcomb and undergraduate research assistants, Thomas Childs and Kevin Baer) whose ratings and work were vetted by a trained auditor (Gary M. Burlingame, PhD). In order to attempt to keep my previous experience of rating in the pilot study from

impacting the results of the current analysis, I participated only in the unitizing process and none of the coding. Instead, I trained both research assistants how to unitize and code data (process explained below), without revealing findings from the pilot study that would bias their take on the new data. I also oversaw the coding process in order to answer questions about the process of coding, about the parent study, and about the GQ itself. Throughout the course of coding, however, I did not offer any feedback or input on the codes the research assistants established. Because the meaning identified in the units was subjective, we took as many steps as possible to limit the influence of previous findings on the blind raters; however, it is possible that I may have influenced them unintentionally in some way while I oversaw the coding process. However, because we never discussed any of the previous findings, and they had no access to previous findings, the influence is assumed to be minimal.

Analysis followed the same structured process that was used in the pilot study, outlined in *Qualitative Content Analysis in Practice*, by Margrit Schreier (2012). Schreier describes qualitative content analysis as a systematic way to describe meaning found in qualitative data by grouping material of similar meaning into categories within an area of interest. She outlines eight steps to follow in order to achieve this task (see Table 2.1).

Table 2.1

Steps of Qualitative Content Analysis (As Outlined by Margrit Schreier, 2012)

	Steps of Analysis
1	Decide on your research question
2	Select your material
3	Build a coding frame
4	Divide your material into units of coding
5	Initial coding based on coding frame
6	Evaluating and modifying your coding frame
7	Main analysis
8	Interpreting and presenting findings

Step 1: Decide on Your Research Question. First, a research question is established. My analysis was guided by the research question, "What does it mean to 'act upon' GQ feedback?" Again, this is the same question that directed the Woodland pilot.

Step 2: Select Your Material. Next, material that is conducive to answering the research question is selected. As was stated above, the leader slips and debrief interviews served as my qualitative material.

Step 3: Build a Coding Frame. Building a coding frame is a way of determining what specific aspects of the research area will be focused on in the analysis. The pilot study included both *use* and *value* as guiding definitions of acting on feedback based on a pilot thematic analysis of 52 units of data. The current study focused on use due to the limited support for value found in the pilot. Some qualitative researchers choose to approach their data with several a priori categories as part of the coding frame. The alternative, and the method we chose, is to adopt a general focus (feedback use) and allow the categories to emerge from the data based on similarities identified in the content under the general coding frame of "use." In the pilot, we decided to take this approach because of our limited understanding of what it meant to act on feedback when first approaching the data. We adopted this same approach in the current study instead of adopting the categories identified in the pilot as our a priori coding frame. This was done in order to let the new data sample guide our categorical findings, and so that our conceptions of "acted upon" established in the pilot would not dominate the new data sample without providing opportunity for alternative categories.

Step 4: Divide Your Material into Units of Coding. Dividing material into units of coding, or unitizing, consists of separating the qualitative content into small parts, or units, that include only one thought or concept. We decided to establish unitization decisions based on rules

established by Stinchfield and Burlingame (1991) because the rules described therein are more specific than the guidelines described in Schreier's work (see Appendix G). This added specificity was intended to increase the reliability of unitizing between raters. Additionally, throughout the pilot study, additional rules were established specific to our data set that helped increase reliability, and therefore were also included in the current study (specific adaptations are found in Appendix G).

Unitizing was split into two separate rounds. For the first round, I unitized the entire data set and assigned each research assistant half of the leader slips and half of the debrief interviews in order to cut down the time it took to unitize. First, all raters would independently unitize a subsample of the data. We unitized in sections or subsamples in order to increase coding efficiency and stability between raters as we went, instead of unitizing the entire data set without discussing any of our work during the process. Next, I met with each research assistant to compare my unitizing decisions with theirs. We discussed and resolved disagreements and took remaining disagreements to the auditor for a final decision. This first round of unitizing functioned to educate raters on the unitizing process and to establish increased stability between raters, but decisions made in the first round of coding were not considered final.

The second round of unitizing was intended to produce the final units that would be coded in the subsequent step of the qualitative analysis. As such, 10% of the leader slips and 10% of the debrief interviews were randomly selected using a random numbers generator, and were unitized by all three raters to establish reliable unitizing across raters. We again initially unitized the material independently, then all met together to discuss disagreements, with any remaining disagreements being taken to our auditor. Next, half of the remaining 90% of each data source was randomly assigned to each research assistant to be unitized, again, to increase

the rate at which we unitized. This followed the same procedure of unitizing independently, then meeting together to discuss and resolve disagreements. The number of disagreements between raters was recorded throughout the unitizing process.

Step 5: Initial Coding Based on Coding Frame. The next step in the analysis is to start to code the data, or in other words identify meaning within the units of data. It was at this point that my role shifted from participatory to exclusively observational. During the pilot, this initial coding was split into three phases that consisted of describing the meaning of individual units. Accordingly, the same process was used in the current study. These three phases took place over the course of one full pass through each unit in the data set.

During the first phase, raters started with 10% of the data that was randomly selected for initial analysis. They described the meaning in each unit, watching for common themes as potential main categories and distinct ideas within those themes as potential subcategories. As was the case with unitizing, raters first independently described each unit, and then came together to discuss findings. They determined together which descriptors best captured the meaning of the units and took any outstanding disagreements to the auditor. Raters also watched for units that had no recognizable connection to GQ use and coded them as irrelevant.

As raters continued to move through subsamples of data, the second phase of coding was initiated. This consisted of grouping similar units under common labels established in the initial phase of coding. This grouping was the first step toward establishing main categories. Raters worked together during this phase of coding; however, consensus was still reached on every decision.

The third phase of coding occurred while raters completed the first pass through the data. During this phase, raters identified distinct ideas within each common group in order to identify

meaningful differences and establish subcategories. While they had been mindful of potential subcategories throughout the first two phases, during this phase raters formally evaluated the distinctions within each common grouping and identified labels for these subcategories.

Step 6: Evaluating and Modifying Your Coding Frame. After the first three phases of coding, raters evaluated the categories and subcategories identified in the initial phases of coding. This was done by considering how well each category fit or provided understanding of the main guiding definitions (i.e., use) of "acted upon," how well the category names represented the units grouped within those categories, and whether or not there were enough units within each category and subcategory to warrant their own grouping.

Step 7: Main Analysis. After establishing more finalized categories and subcategories, each individual unit was considered again or recoded to ensure it was categorized in the most accurate category in the coding frame. This included re-considering units deemed irrelevant to GQ use to ensure that there was in fact no identifiable relationship between the content of the unit and use of GQ feedback. During this main analysis, main categories and subcategories were finalized and the coding frame was completed.

Step 8: Interpreting and Presenting Findings. In the last step of coding, each unit was given a numeric code based on its qualitative category and subcategories. This served to organize the data and put it in a format for eventual use in testing mediation in the parent RCT (testing whether or not the way leaders act on GQ feedback mediates the relationship between receiving GQ feedback and subsequent member GQ scores, OQ scores, and dropout rates).

Quantitative Analysis

After completing the qualitative analysis, I used multinomial logistic regression analyses to better understand the relationship between the qualitative category results and other variables

that might predict variability in use, including data source (whether the data came from the leader slips or the debrief interviews), group type (process or specific focus), leader, and group. I also used multinomial logistic regression to better understand the relationship between qualitative categories and alerts provided to the leaders, including the total number of negative alerts per session, total number of people alerting, and the proportion of people alerting. Post-estimation tests were used to estimate whether the variables studied accounted for a significant portion of the observed variance. Because the leader and group variables each had several possible values (12 leaders and 16 groups), pairwise comparisons were used to investigate whether individual leaders or groups seemed to account for the observed variability, or if findings were more evenly distributed. Given the high number of comparisons conducted, the Benjamini-Hochberg Procedure was used to limit the familywise Type I error rate (Thissen, Steinberg, & Kuang, 2002). Finally, to determine the relationship between the qualitative category findings and the GQ Use Questionnaire, Spearman and Pearson correlations were used.

Results

Part I: Qualitative Findings

Unitization and rater agreement. Throughout the course of unitizing, the percentage of agreement was recorded between each undergraduate research assistant and myself. The first round of unitizing was conducted to educate raters on the process and establish stability in unitizing. With the first research assistant, initial agreement on leader slips was 89%. After discussing discrepancies, raters established 100% agreement and did not have any remaining disagreements to take to the auditor. Initial agreement for debrief interviews was lower, at 80%. Raters established 99% agreement after discussing discrepancies, and reached 100% agreement after taking the remaining 1% to the auditor. First-round unitizing with the second research

assistant yielded 85% initial agreement on leader slips, with 99% agreement established after discussion, and 100% established after taking the remaining 1% of disagreements to the auditor. Agreement on debrief interviews was also lower for this research assistant, with 77% initial agreement, followed again by 99% agreement after discussion, and 100% agreement after meeting with the auditor.

The second round of unitizing was conducted to finalize unitization decisions. With each research assistant and each data source, initial agreement was higher in the second round than in the first round. The 10% of leader slips unitized by all raters yielded 91% agreement between each dyad and 87% total agreement between all three raters. With the remaining leader slips, the first research assistant achieved 91% initial agreement, with 100% agreement after discussing discrepancies. The second research assistant had 89% initial agreement, and also reached 100% agreement after discussing differences. For debrief interviews, the 10% unitized together yielded 80% total initial agreement between all three raters, and 86%, 87%, and 88% initial agreement between the separate dyads. No discrepancies remained after discussion. With the remaining debrief interviews, the first research assistant established 86% initial agreement, and 100% agreement after discussion. Lastly, the second research assistant achieved 82% initial agreement with debrief interviews, and 100% agreement after discussion.

The completed unitization process yielded 2,182 total units to be coded. Throughout the process of coding, certain units were determined to be irrelevant to use, reflecting no significant meaning to be captured. For example, unit 77.1 reads, “12/5/13 was our final group session of Fall Semester.” Similarly, some units were only relevant to OQ use, which falls outside the scope of the current study. For example, unit 12.16.2 states, “And then, I think we also talked with them about their OQ scores.” After these units were excluded, a total of 1,467 units were

identified to include information pertinent to GQ use, 378 of which came from leader slips, and 1,089 came from debrief interviews.

Coding. The process of coding yielded rich insight into what it means to act on GQ feedback. Broadly speaking, research assistants identified three overarching dimensions of use, including Pre-Group use, In-Group use, and Post-Group/Effect. Within each dimension, they identified specific categories of use. Raters took some categories and further reduced the information to subcategories to add detail about GQ use, with a handful of subcategories also organized by sub-subcategories that added one final layer of detail **(**see Table 3.1 for a list of all categories, subcategories, sub-subcateogires, and frequencies). I will describe each of these starting with the most highly populated to those with the least endorsement.

Dimensions and main categories that research assistants established are mutually exclusive, but subcategories within categories are not. That is to say a unit could only be coded within one main category, but multiple subcategories and sub-subcategories could be used to describe a single unit. For example, unit 16.3.5 states, "For example if a group member experienced a major change in negative relationship and work and bond, like say it went poorly last week, I'd be kind of curious about what that was about." This unit was coded in only one category, called Analysis of GQ Feedback, but included two subcategories to describe the type of analysis. The first was Leader Reaction, which captures the leader's described curiosity. The second was Observation about Member, which captures the first portion of the unit, or what the leader was curious about. Further, one sub-subcategory of Observation about Member was included, namely Member Score, which indicates that the leader's observation about the group member related to their score, as opposed to in group behavior.

Table 3.1

Complete List of Categories Identified in Current Qualitative Analysis

Dimensions, Main Categories, *Subcategories,* and Sub-subcategories	*n*
Pre-Group	**620**
Analysis of GQ Feedback	504
Comparison Between Perception and Feedback	130
Planning Interventions	128
No Action	42
Withhold	40
Observation about Group Member	94
Member Score	77
Member Demeanor	10
Leader Reactions to GQ Feedback	90
Observing Group Trends	48
Alerts	24
Speculation about Group Feedback	20
With Coleader	17
Just Coleader	1
Comparison Between Groups	16
Member Self-Analysis	6
Review of Feedback	50
With Coleader	6
Just Coleader	1
Alerts	4
Disruption in Data	26
Member Failure to Complete GQ	9
Failure to Receive Feedback	6
Leader Absence	4

(continued)

Table 3.1 Complete List of Categories Identified in Current Qualitative Analysis (continued)

Proficiency with GQ	20
Member Understanding	8
At Beginning	4
At End	4
Sharing Feedback with Others (Coleader)	20
In-Group	**408**
GQ in Group	301
Discussion in Group about GQ	176
Specific Member(s)	59
Direct	59
Indirect	28
Group Trends	8
No GQ in Group	60
Reason for Not Using	36
Intentions to Use GQ	37
Members	32
Review of Feedback	4
Mindfulness of GQ Feedback	20
Explanation of Use	15
Importance of GQ	5
Review of Feedback	3
Initiated by Member	13
With Coleader	2
Just Coleader	2
Member Reaction	55
Reaction to Leader Use	46
Feelings	7

(continued)

Table 3.1 Complete List of Categories Identified in Current Qualitative Analysis (continued)

Reaction to Individual Report	8
Sharing Feedback with Others (Members)	22
Member Preference	8
Follow-Up Outside of Group	19
Leader Confusion	12
Post-Group	**434**
Effect on Group	321
Effect on Leader	224
Change in GQ Analysis	41
Change in GQ Implementation	40
Change in Leader Approach	36
Heightened Awareness	29
Confirmation	14
Desire for Feedback	14
Anticipation	10
Effect on Members	71
Specific Members	21
Willingness to Express Themselves	18
Effect on Coleader	11
Change in Leader Approach	2
Influence of GQ Feedback Group on Non-Feedback Group	48
Opinion of GQ	71
Usefulness	44
Value	22
Commitment	8
Member(s) Opinion	1

(continued)

Table 3.1 Complete List of Categories Identified in Current Qualitative Analysis (continued)

Hindsight Opinion	36
Suggestions for Improvement	27
Changes in GQ Implementation	6
Expectations of Future GQ Scores	5
Overall GQ Use[a]	5

[a]Not included in any dimension.

Dimension 1: Pre-Group.

Analysis of GQ Feedback. Analysis of GQ Feedback is the category containing the largest number of units *(n* = 504). It captures units in which leaders describe studying, scrutinizing, or interpreting the GQ feedback. It has several subcategories and sub-subcategories that offer more specific detail about the type of analysis the leader reported.

The first subcategory is called Comparison Between Perception and Feedback (*n* = 130). It is comprised of units in which leaders compare their own opinions of the therapeutic relationship to the information they receive on the GQ feedback report. For example, leader slip 97.1 reads, "W's positive work score doesn't make sense since he worked really hard last week and was the primary focus of the group." Research assistants included instances of surprise due to GQ scores in this category.

The second subcategory is called Planning Interventions (*n* = 128) and includes units in which leaders (and/or coleaders) discuss, organize, or strategize interventions based on the GQ feedback. For example, leader slip 47.2 stated, "We planned to provide feedback about these differences using the GQ feedback with the group." Within this subcategory, research assistants noticed two sub-subcategories emerge, including instances in which the leader reports doing nothing with the feedback (No Action*; n* = 42), and instances in which the leader intentionally withheld the feedback from the group (Withhold; *n* = 40). The expressed intentionality of the

leader is what distinguishes these two categories, with purposeful decisions to keep feedback from the group falling under the Withhold sub-subcategory (for example, debrief interview unit 5.11.2 states, "I withheld feedback in hopes that the group would address on their own the issues that prompted certain responses to the GQ questions."), and simple statements in which leaders decide not to do anything at all with feedback falling under No Action (for example, leader slip 91.2 states, "No specific action was taken in group this week based on feedback.").

The third subcategory of Analysis of GQ Feedback was named Observation about Group Member (n = 94) and refers to units in which leaders make note of something about a member or group of members. Two sub-subcategories were identified within this subcategory, including Member Score (n = 77), which refers to observations relating to member scores on the GQ (for example, leader slip 43.1 reads, "Group leaders noticed that while most group members' scores for Positive Bond decreased and Negative Relationship increased, C's scores did so more steeply than her peers.") and Member Demeanor (n = 10), which refers to observations relating to in-group observations of member behavior or tendencies. For example, leader slip 71.7 states, "…we related that to our own observations that she seemed a bit withdrawn at points during group."

The fourth subcategory is Leader Reactions to GQ Feedback (n = 90). It was created to capture units in which leaders express their thoughts or feelings about what they observed about group members and/or feedback. One of these units (leader slip 119.3) reads, "This told me that the conflict was meaningful for him."

Observing Group Trends (n = 48) is the fifth subcategory established and includes units in which leaders identify patterns in the GQ feedback scores. For example, leader slip 13.4

states, “We would see the growth of the group under positive bonds, we’d see everybody kind of, like, inching up on that scale.”

The sixth subcategory is called Alerts (n = 24) and was created to capture units in which leaders specifically mention studying, investigating, or interpreting a GQ alert. For example, leader slip 79.3 reads, “We were pleased to see A’s positive relative alert.”

Speculation about Group Feedback (n = 20) is the seventh subcategory identified in Analysis of GQ Feedback and includes units in which leaders identify possible explanations for member scores or in-group behavior. For example, in debrief interview unit 4.11.6, the leader explained a possible reason for scores he was seeing by stating, “We hypothesized that maybe he was just having a bad day.”

As is true with some other categories as well, research assistants included a subcategory to indicate instances in which coleaders were involved in analysis. For example, the following two units (leader slip 16.7 and 16.8) indicate, “We discussed it as coleaders/ and planned our response (to some degree).” The first unit was coded under the With Coleader subcategory of Analysis of GQ Feedback (n = 17), and the second was coded both to indicate coleader involvement and under the Planning Interventions subcategory to capture that planning occurred with the coleader as well. Research assistants also differentiated units in which the coleader was reported to do something independent of the group leader (Just Coleader, n = 1). For example, “[My coleader] did not take any specific actions based on the feedback.”

The ninth subcategory of Analysis of GQ Feedback is Comparison Between Groups (n = 16) and is made up of units in which the leader compares the current feedback group with a previous feedback group or with a group of a different therapeutic style. For example, leader slip

25.9 reads, "I feel that I would have been more inclined naturally to bring the GQ feedback in more often in a general process group."

The final subcategory is called Member Self-Analysis (n = 6) and refers to units in which leaders describe a member's awareness of his/her own score and/or that member offers an interpretation of the score. For example, leader slip 126.3 states, "One group member commented that she saw her decline in positive bond as a reflection of her pattern of feeling more discomfort the closer she feels to others."

Review of Feedback. Review of Feedback (n = 50) was a category established to capture units in which leaders describe looking at and making simple observations about the GQ feedback, but do not offer any interpretation or analysis of the data. For example, leader slip 33.1 read, "My coleader and I looked at our results." Three subcategories were identified within this category. The first subcategories captured coleader involvement, or whether the coleader reviewed the feedback with the leader (as in the example just mentioned; n = 6) or alone (n = 1). The third subcategory (Alerts) identified units in which leaders acknowledged the presence of an alert during their review of feedback, without offering any further analysis (n = 4). For example, leader slip 73.3 read, "I noticed Ch's alerts."

Sharing Feedback with Others. This category was established to capture instances in which the leader describes distributing the actual feedback report itself to someone else (n = 42). It is the only category that is split between two dimensions (Pre-Group and In-Group Use). The first subcategory (Group Members; n = 22) refers to times when the leaders indicated that they brought copies of the feedback report to their group members in group. For example, leader slip 31.4 states, "At the beginning of group we gave each member their individual feedback form for review." Therefore, units in this subcategory fit best in the In-Group dimension. There is one

sub-subcategory of Group Members which captures the instances in which leaders describe members offering their preference for how they receive physical copies of the feedback (Member Preference; *n* = 8). For example, debrief interview unit 16.1.4 states, "They didn't want everybody to see their scores so I printed off their individual responses from the analyst."

The second subcategory (Coleader; *n* = 20) refers to the times leaders report delivering the feedback to their coleaders. For example, unit 83.1 reads, "I emailed this feedback to my coleader prior to group." This always occurred before the group started, and therefore fits in the Pre-Group dimension.

Proficiency with GQ. The next main category research assistants identified captures instances in which leaders describe their own familiarity or expertise in using and understanding the GQ (*n* = 20). For example, debrief interview unit 15.14.2 explains, "I kind of jumped in not knowing much about the instrument." There are three subcategories that research assistants identified within this category. The first captures units in which leaders report a member's understanding of the GQ or lack thereof (Member Understanding; *n* = 8). For example, leader slip 111.6 reads, "Members also expressed concern about not understanding the question about the group having a common goal." The second and third subcategories capture change over time by differentiating between instances in which leaders report their proficiency at the beginning of the study (At Beginning, *n* = 4) and at the end of the study (At End, *n* = 4). For example, debrief interview unit 9.27.2 reads, "I feel like I was, from the beginning was really familiar with it," and unit 15.15.2 from another leader read, "And now I would say that I'm much more—not totally, totally familiar about how to use it best, but more aware of what, for me, I tend to focus on more as I'm looking at it."

Disruption in Data. The final main category in the Pre-Group dimension captures leaders' descriptions of interruptions in the delivery of the feedback report (n = 26). For example, one leader reported not receiving their feedback report because "apparently we encountered some email snafus" (leader slip 54.3). Research assistants identified three subcategories, including Member Failure to Complete GQ (n = 9), Failure to Receive Feedback (n = 6), and Leader Absence (n = 4), with each subcategory capturing a specific reason for the interruption.

Dimension 2: In-Group.

GQ in Group. The first main category broadly describes any use of the GQ during a group session (n = 301). Research assistants identified seven subcategories that describe that in group use. The first subcategory is called Discussion in Group about GQ (n = 176) and captures any verbal use, interventions, or questions about the GQ during the group session. For example, leader slip 126.4 states, "We used the overall feedback to discuss differing dynamics in the group when different members are present." This subcategory includes four sub-subcategories. The first is called Specific Member(s) (n = 59) and refers to discussions that are directed by the leader toward a specific member or subgroup of members. One of these units reads, "Interventions made in response to feedback included exploring with one member her lack of engagement as this seems to be related to her work in group" (leader slip 132.4). The next sub-subcategory is called Direct (n = 59) and refers to discussions in which the leader directly and explicitly refers to the GQ feedback in such a way that group members are aware that the information is based on GQ feedback. For example, debrief interview unit 8.1.1 states, "As part of the 'go around' during the first 10 minutes of group, we shared brief observations from the data, sometimes suggesting a group member seemed ready to work." In contrast, the third sub-

subcategory is called Indirect (n = 28) and includes units in which leaders introduce the feedback verbally, but in a way that members are not aware that the statement is based on GQ feedback. For example, debrief interview 2.2.8 states, "So, occasionally through the semester, I would make an observation to someone about, 'you seem to be less engaged the last couple of weeks.'" The final sub-subcategory is called Group Trends (n = 8) and describes discussions in which GQ feedback patterns are brought up. Leader slip 48.2, for example, reads, "Group leaders brought up the group members' general trend of decreased Positive Work and increased Negative Relationship following the conflict as a way to propose options for how to spend this semester's last session."

Research assistants also identified units in which leaders described *not* using the feedback in group, and coded these units under the GQ in Group category with a 0 to indicate lack of use (n = 60). For example, leader slip 117.1 states, "We did not directly address feedback with the group." Within this portion of units, there was a subcategory of units identified in which leaders provide a reason for their decision not to use the feedback (Reason for Not Using; n = 36). For example, debrief interview unit 12.30.3 states, "We didn't have enough time to sort of process it, so we didn't bring it on that time."

The third subcategory is called Intentions to Use GQ (n = 37) and includes units in which leaders or members talk about their plans to use the GQ in the future. For example, leader slip 122.6 reads, "We discussed whether they would want to look at the results every week or not." There are two sub-subcategories in this subcategory, including Members (n = 32; includes units in which members are quoted as expressing an opinion about how the GQ feedback should be used in the future), and Review of Feedback (n = 4; includes units in which leaders discuss the way leaders intend to review the feedback in the future). For example, when describing

members' intentions to use feedback, one leader said, "They decided they didn't want it very specific. They were concerned they might 'teach to the test' (or try to match performance to influence the results)." When describing reviewing feedback in the future, another leader stated, "And then filling out consent forms, I let them know that I would be looking at it and reading it" (debrief interview unit 9.1.3).

Mindfulness of GQ Feedback is the fourth subcategory (n = 20) of GQ in Group. It captures units in which leaders report keeping information from the GQ feedback in their minds during group. For example, leader slip 117.2 states, "We are mindful of the members who are experiencing negative bonds with either members or leaders."

The fifth subcategory of GQ in Group is called Explanation of Use (n = 15) and refers to instances in which leaders spend time in group explaining to group members some aspect about the GQ. For example, leader slip 122.5 reads, "We talked with [group members] about how to read the results." There are two sub-subcategories within this subcategory, the first of which is called Importance of GQ (n = 5) and contains units in which the leader explains that the GQ is important to him or her. For example, debrief interview 2.1.1 states, "I informed the group that the study was important to me." The second sub-subcategory is called Review of Feedback (n = 3). These units are characterized by an explanation to group members that group leaders review the feedback. For example, debrief interview unit 9.2.3 said, "So then I said well, as you know we've been looking at this."

The next subcategory was called Initiated by Member (n = 13) and describes instances in which group members brought up the GQ without being prompted by an intervention from the group leader. For example, leader slip 119.6 reads, "Also of note, another member whose bond

went down and negative relationship went up challenged us to be aware of the challenges he finds in connecting with the group."

As with other categories, research assistants also included a Coleader subcategory to indicate in group interventions that were attributed both to the coleader and group leader (n = 2) or just the coleader (n = 2). For example, leader slip 119.5 reads, "[My coleader] and I challenged this member to sit with conflict again with the same member to allow continued resolution of conflict and cohesion."

Member Reaction. The second category describes member thoughts, feelings, or behaviors that occur in group in response to GQ feedback or related interventions (n = 55). The first subcategory is called Reaction to Leader Use (n = 46), and refers to units in which members react to any other form of leader use. Debrief interview unit 2.3.3 illustrates an example of this, stating, "I got much better response when I would just say it without referring back to the scores." There is one sub-subcategory of Reaction to Leader Use which research assistants called Feelings (n = 7) and captures instances in which members express personal feelings toward the leader's use of the GQ feedback. For example, debrief interview unit 12.13.4 states, "But I think that they really appreciated the fact that we were looking at what was going on." The other subcategory identified is called Reaction to Individual Report (n = 8) and includes units in which members are given a copy of the actual report and respond to the report itself. For example, debrief interview unit 15.8.1 states, "Most of them would take a look at it and just nod."

Follow-Up Outside of Group. The next category illustrates instances in which leaders have a conversation with a group member based on GQ feedback immediately following the group session (n = 19). For example, leader slip 33 units 3 and 4 state, "We asked the group member to stay after group/ and we talked with him about the results." While these interventions

did not technically occur during the course of the group, this category was still included in the In-Group dimension because the interventions occurred in conjunction with the group and reflect direct interventions, while categories in the Post-Group dimension do not capture interventions specifically.

Leader Confusion. The second main category of the In-Group Dimension is called Leader Confusion and captures units in which leaders express confusion about the GQ, how to use it, or about the study itself (n = 12). For example, debrief interview 12.54.1 states, "Yeah I just thought, it's like 'What does that mean? I don't know even how to answer. No, we don't have mutual group goals, so are we bad?'"

Dimension 3: Post-Group/Effect.

Effect on Group. The first and most frequently identified category in the Post-Group dimension is called Effect on Group (n = 321). The first three subcategories describe the party affected, with Effect on Leader (n = 224), Effect on Members (n = 71), Effect on Coleader (n = 11). Effect on Leader is characterized by seven sub-subcategories. The first is called Change in GQ Analysis (n = 41) and includes units in which leaders report altering or adjusting the way they studied, scrutinized, or interpreted the GQ feedback. For example, debrief interview unit 15.13.3 states, "And so I noticed that I would look for things like that, like who was in the alerts going one way or another." The second sub-subcategory is Change in GQ Implementation (n = 40), which includes units in which the leader altered or adjusted the way he or she applied the GQ feedback group. For example, debrief interview unit 9.22.2 states, "From the previous semester I thought it would be a good idea to use it indirectly." The third sub-subcategory is called Change in Leader Approach (n = 36) and refers to instances in which the leader reported changing demeanor, attitude, or thought processes toward using the GQ feedback. For example,

debrief interview unit 15.6.1 states, “If anything I would say it gave me more patience with the silence as they struggled with how to respond because I knew that it was feedback that they had given.” The fourth sub-subcategory is called Heightened Awareness (n = 29) and includes units in which leaders report feeling mindful, alert, or attentive to the members or of events happening in group as a result of receiving the GQ feedback. For example, debrief interview unit 4.6.1 reads, “It made me watch some of the members more closely.” Next, Confirmation (n = 14) captures units in which leaders interventions, use, or observations about the GQ feedback are validated or affirmed by subsequent feedback. Debrief interview unit 8.4.1 provides an example for this, “I think for me it affirms things that I was observing either in terms of attendance or participation in the group.” The sixth sub-subcategory of Effect on Leader is called Desire for Feedback (n = 14) and includes units in which leaders express wishing they could have feedback in another group. For example, debrief interview 15.26.1 states, “I knew that I wanted it in the non-feedback group.” The final sub-subcategory is called Anticipation (n = 10) and refers to units in which the leader reports feeling eager, excited, or anxious to see future GQ scores as a result of using the feedback. Leader slip 25.6 illustrates this, stating, “Now, I’m very curious what the scores will show after this session.”

Effect on Members has two sub-subcategories including Specific Members (n = 21) and Willingness to Express Themselves (n = 18). The former refers to units in which the specific impact or outcome of a GQ intervention affects an individual member or subgroup of members. Leader slip 10.3 illustrates this by stating, “The scores for other members have recovered as we’ve attended to them.” The latter refers to instances in which members felt more inclined to share thoughts or feelings as a result of the leader’s GQ use. For example, debrief interview 2.3.1

states, “One member whose scores indicated he had been hurt and was withdrawing, was willing to talk about his feelings.”

Effect on Coleader has one sub-subcategory, which is common to Effect on Leader (Change in Leader Approach; *n* = 2). The meaning captured is identical to that in the Effect on Leader subcategory, except that the observations are about the Coleader.

Finally, there is one additional subcategory in Effect on Group which is called Influence of GQ Feedback Group on Non-Feedback Group (*n* = 48). It includes units in which leaders describe the effect of having the GQ in their feedback group on their experience in the non-feedback group. For example, debrief interview unit 15.20.1 states, “Mostly with the non-feedback group I just find myself wondering, like wondering what else is going on.”

Opinion of GQ. The second category is called Opinion of GQ (*n* = 71) and includes units in which leaders or members express an evaluation of or attitude toward the GQ. For example, debrief interview unit 7.22.3 states, “I’ve enjoyed having the feedback.” The first subcategory is called Usefulness (*n* = 44) and includes descriptions of leader evaluations of the utility of the GQ feedback. Debrief interview unit 18.18.2 speaks to this point by stating, “Getting that kind of hard data on the clients is just helpful, I think, over all.” The second subcategory researchers identified was named Value (*n* = 22) and describes units in which leaders express their appreciation or value of the GQ feedback, as shown in debrief interview unit 7.3.2, “I really appreciated the feedback.” The next subcategory is called Commitment (*n* = 8) and describes units in which leaders express their commitment to using the GQ feedback; for example, debrief interview 18.25.2 states, “I was saying that I really believe in using OQ, GQ feedback.” Lastly, Member(s) Opinion (*n* = 1) captures units in which members express their evaluation or attitude

toward the GQ feedback. Interview unit 16.6.6 states, "And that's when a couple members talked about how they found it valuable."

Hindsight Opinion. The next category in the Post-Group dimension was identified to capture units in which leaders make comments in retrospect about the GQ or its use (n = 36). The first of two subcategories is called Suggestions for Improvement (n = 27) and includes leader proposals for changes in the GQ or the study itself. For example, debrief interview unit 12.63.4 states, "So maybe getting [the report] a day or two before the group would be helpful." The second subcategory is called Changes in GQ Implementation (n = 6) and includes units in which leaders retrospectively express what they wish they would have done with the GQ. Debrief interview unit 18.29.4 states, "So that, indirectly, affected the process, but I didn't ever make it explicit, which is a bummer, because I didn't think about it."

Expectations of Future GQ Scores. The final category in the Post-Group dimension is called Expectations of Future GQ Scores (n = 5). It captures leader thoughts, assumptions, and predictions for future GQ feedback scores. For example, leader slip 21.6 reads, "I'll be interested to see if working hard and challenging each other have lower scores this week also."

Overall GQ Use. Finally, there was one category that does not fit in the three dimension model. It is called Overall GQ Use (n = 5) and includes units in which a leader states that they used the feedback, but does not offer specific details about how they used it. For example, one leader noted, "I used just the GQ," in debrief interview unit 16.20.5.

Part II: Quantitative Findings

After completing the qualitative analysis, frequency tables were generated to illustrate the distribution of units across data source, group type, leader, and group (see Tables 3.2–3.5 for frequency percentages by dimension; see Appendix K for frequency percentages by main

category). Chi-square analysis demonstrates significant observed variability in the way leaders reported their use by source (debrief or leader slip) and type of group (process or specific focus), as well as differences between leaders and groups.

I used multinomial logistic regression to explore the differences in observed variability (see Appendix L). More specifically, I examined data source, group type, leader, and group as potential predictors of variability. Again, there were two data sources (leader slips and debrief interviews), two group types (process and specific focus), 12 leaders, and 16 groups. Omnibus post-test estimation significance tests indicate that all four variables significantly predict differences in the frequency of units observed across the qualitative use dimensions ($p < .01$).

Table 3.2

Dimension Frequency Percentages by Data Source

Dimension	Leader Slips	Interviews	Total
Pre-Group	64.3	34.6	42.3
In-Group	27.8	27.8	27.8
Post-Group	7.7	37.2	29.6
Unclassified	0.3	0.4	0.3
Total Units	378	1,089	1,467

Note. $\chi^2(3) = 141.1649$, $p < 0.01$

Table 3.3

Dimension Frequency Percentages by Group Type

Dimension	Process	Specific Focus	Total
Pre-Group	38.9	50.2	42.3
In-Group	29.5	23.9	27.8
Post-Group	31.3	25.7	29.6
Unclassified	0.4	0.2	0.3
Total Units	1,027	440	1,467

Note. $\chi^2(3) = 16.4284$, $p < 0.01$

Table 3.4

Dimension Frequency Percentages by Leader

Dimension	2	4	5	6	7	8	9	12	14	15	16	18	Total
Pre-Group	47.1	53.6	38.8	37.9	65.1	44.2	35.9	39.6	30.0	36.1	28.4	25.9	42.2
In-Group	23.3	29.6	32.2	17.2	13.0	27.9	31.4	31.5	40.0	26.2	43.3	24.1	27.8
Post-Group	29.2	16.8	29.0	44.8	21.2	26.9	32.1	28.8	30.0	37.6	27.6	50.0	30.9
Unclassified	0.4	0.0	0.0	0.0	0.7	1.0	0.6	0.0	0.0	0.0	0.8	0.0	0.3
Total Units	240	125	152	29	146	104	156	111	10	202	134	58	1,467

Note. $\chi^2(33) = 97.9354$, $p < 0.01$

Table 3.5

Dimension Frequency Percentages by Group

Dimension	25	27	29	32	33	34	36	38	40	42	50	52	53	57	60	61	Total
Pre-Group	42.3	34.0	37.9	65.1	47.1	53.6	37.7	39.6	41.7	30.0	28.4	39.6	25.9	28.1	44.9	39.6	42.2
In-Group	38.5	40.8	17.2	13.0	23.3	29.6	36.1	31.5	31.7	40.0	43.3	29.7	24.1	18.3	24.4	13.2	26.6
Post-Group	19.2	24.3	44.8	21.2	29.2	16.8	26.2	28.8	26.7	30.0	27.6	30.8	50.0	53.7	29.5	47.2	30.9
Unclassified	0.0	1.0	0.0	0.7	0.4	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	1.3	0.0	0.3
Total Units	26	103	29	146	240	125	61	111	120	10	134	91	58	82	78	53	1,467

Note. $\chi^2(45) = 136.0783$, $p < 0.01$

For data source, units categorized in the In-Group and Post-Group/Effect dimension were more likely to come from debrief interviews than leader slips. The same pattern was found when process groups were compared to specific focus groups.

Pairwise comparisons were conducted to explore the significant multinomial logistic regression for leader and group. This was necessary given the large number of leaders and groups being compared to determine if specific leaders or groups were significantly different from each other. The Benjamini-Hochberg procedure resulted in 23 significant comparisons out of 198 for the leader variable, and 33 significant comparisons out of 360 for the group variable. For leaders, it appears that one leader reported using the feedback significantly less than other leaders. For group, no one group stood out as significantly different in the Pre-Group dimension; however, the In-Group dimension had one group with significantly fewer units and one with significantly more units. For the Post-Group/Effect dimension, one group had significantly less units than several others.

I included alert data from the parent RCT to shed light on potential explanations for the observed differences in leader use by group. In other words, I wanted to know if differences in rates of alerts across leaders and groups might explain the above differences in GQ use by group. I performed separate multinomial logistic regressions to determine the relationship between total negative alerts per session, total alerts per person per session, and the proportion of people alerting per session (see Appendix L). Results show a significant relationship between leader use (the number of units in each dimension) and the number of people alerting per session, as well as the proportion of people alerting per session (see Table 3.6). Specifically, the more people that alert and the higher the proportion of people alerting, the less likely leaders are to report using

In-Group use as compared with Pre-Group use. The total negative alerts per session was not associated with a significant relationship ($p = .5$).

Table 3.6

Alert Data Multinomial Logistic Regression Results

	Point Estimate	95% Confidence Interval
Number of People Alerting		
Pre-Group	(base outcome)	--
In-Group	-.45***	-0.61— -0.29
Post-Group	.24	-0.07— 0.55
Proportion of People Alerting		
Pre-Group	(base outcome)	--
In-Group	-3.63***	-4.78— -2.48
Post-Group	.77	-1.40— 2.93

***$p < .001$

Part III: Merging Results

An added analysis in my dissertation that was not included in my prospectus was a careful merging of my data with the Woodland pilot data. There are three reasons that merging the data is advantageous. First, doing so provides a unified set of qualitative information that can be used to create a mediator variable for the parent RCT. Second, considering the overlap between the two studies bolsters the results of the current study and the Woodland study by enhancing the construct validity of what it means to use the GQ. Third, pooling data from all participating leaders provides a more powerful test of the GQ Use questionnaire's viability as an alternate index of leader GQ use.

The first step in merging the two datasets was to establish the overlap between studies. To establish common categories, a consensus process similar to unitizing and coding was followed. Specifically, a member from each coding team (Sean Woodland and Kevin Baer) and I

examined the complete list of categories, subcategories, and sub-subcategories from both studies. Initial agreement on common categories between each dyad was 80%, 79%, and 69%. After discussing differing opinions, 100% consensus was reached. Thirteen common categories with corresponding definitions were identified (see Table 3.7). Interestingly, the common categories reflect the majority of units across studies. Those that did not overlap were all infrequently endorsed.

Table 3.7

Comparison of Common Categories

Woodland Study Definition	Freq. %	Common Categories	Freq. %	Current Study Definition
Any time that the leader (alone or with coleader) reports looking at the feedback report.	6.30	Review of Feedback	4.43	When a leader/coleader makes a quick, basic observation about the GQ feedback, but does not interpret or analyze data.
Any time that the leader (alone or with coleader) reports the content, an interpretation, or a speculation of the results of the feedback report.	31.19	Analysis of GQ Feedback	24.54	Any time someone studies, scrutinizes, or interprets GQ feedback.
Any time that the leader (alone or with coleader) reports actions taken in response to the feedback report that are meant to prepare for use or implementation in the next session.	4.54	Planning Interventions	3.14	When group leaders (or coleaders) discuss, organize, or strategize interventions based on GQ feedback.
Any time that the leader (alone or with coleader) reports choosing to not use the data from the feedback report.	9.08	Withholding Feedback	9.13	When group leaders intentionally did not bring GQ feedback into the group.
Any time that the group leader (alone or with coleader) brings up scores, trends, or concepts pulled directly from the GQ feedback report.	9.96	GQ in Group-Explicit	13.02	When the leader/coleader verbally and explicitly refers to GQ feedback making the member aware that it's from the GQ.

(continued)

Table 3.7 Comparison of Common Categories (continued)

Any time that the leader (alone or with coleader) 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, group process, or treatment goals, but without making specific mention of measures or feedback.	7.32	GQ in Group-Nonexplicit	1.70	When leader/coleader verbally introduces feedback in a way that the member isn't aware that the statement is based off of the scores.
Any time that the leader reports educating group members about the GQ, OQ, or related subscales.	2.20	Education	0.48	When a leader/coleader spends time in group explaining to group members some aspect about the GQ.
Any time that the leader reports being made more aware of group members who are struggling via receipt and review of the GQ feedback.	1.61	Awareness of Alerters	2.59	When the leader is more mindful, alert, or attentive of events happening in group or to the group members that can be linked to GQ feedback.
Any time that the group leader reports anticipation of receiving GQ results in a future session.	0.44	Looking Forward	1.02	Leaders' thoughts, assumptions, or predictions for GQ feedback in the future.
Any time that the leader (alone or with coleader) reports in-group effects that were either a direct or indirect result of implementing feedback.	19.77	Effects of GQ Use	24.54	The impact, outcome, or consequence that using the GQ feedback has on groups members and leaders.
Any time the leader reports using the feedback in some fashion, but without specifically stating how it was used	3.07	Ambiguous Use	1.02	When a leader/coleader states that they used the GQ feedback, but did not offer more information or was not specific about how they used it.
Any unit that has a clear value statement	0.0[a]	Value/Opinion	6.68	When a group leader expresses their appreciation or value of the GQ feedback.

(continued)

Table 3.7 Comparison of Common Categories (continued)

Any time that the leader reports that group members during the session brought up the feedback report.	1.02	Group-Initiated	0.68	When group members bring up the GQ in group.
Indicates categories that were not included in both analyses. Categories from the Woodland study that were not found in the Whitcomb study include: Attendance, Filling Out Measures, Self-Awareness, and specific value categories.	3.51	Not Captured	7.02	Indicates categories that were not included in both analyses. Categories from the Whitcomb study that were not found in the Woodland study include: Proficiency with GQ, Disruption in Data, Leader Confusion. Subcategories included Member Preference (from Sharing Feedback with others), Suggestions for Improvement (Hindsight Opinion).
Total	100.00	Total	100.00	Total

[a]Note: Value was included in Woodland analysis, but units cannot be included here because of double coding in the Woodland study.

While the qualitative meaning of the categories identified in each study was remarkably similar, the structure of the coding frames was different. For example, the Woodland study conducted an entirely separate analysis on GQ value while the current study focused only on use. In the Woodland study, some units were coded twice, once for value and once for use, while other units were only coded for value. This inconsistency led to the Woodland value units not being included in the current analysis.

A second difference between studies is a discrepancy between the structure of the categories and subcategories. For example, one category in the Woodland study was called Design Specific Interventions. In the current study, the same content was classified as Planning Interventions, but it was included as a subcategory of Analysis of GQ Feedback. In order to code units under a common coding frame, I examined all of my subcategories that overlapped with

main categories from the Woodland study and assigned units with multiple subcategory codes to a single subcategory. I did this by examining the unit and choosing the subcategory that most clearly identified the use described in the unit.

Finally, the Woodland study and the current study used slightly different dimensions to summarize qualitative categories. While the current study utilized three dimensions (Pre-Group, In-Group, and Post-Group/Effect), the Woodland study utilized four dimensions (Pre-Group Reaction, Pre-Group Planning, In-Group, and In-Group Consequence).

To help determine which dimension model to follow (three versus four), I ran a second multinomial logistic regression on the qualitative data from the current study using the Woodland four-dimension model (see Appendix M). As was true with the three-dimension model, omnibus post-test estimation significance tests showed that data source, group type, leader, and group all significantly predicted the number of units in the qualitative use dimensions ($p < .01$). Specific results for data source and group type were also similar. Lastly, pairwise comparisons revealed slightly more differentiation between leaders and groups than the three-dimension model. Because the four-dimension model shows increased differentiation and provides more detail about leader use before group, I decided to use a four-dimension model for the combined data (Pre-Group Review, Pre-Group Planning, In-Group, and Effect).

All of the data from both the Woodland study and the current study was coded according to a common coding scheme with 13 common categories and four common dimensions. Frequency tables illustrate that, again, leader use is variable across different leaders, and some categories are much more densely populated than others (see Tables 3.8-3.12), with chi-square tests again showing significant variability.

A final series of multinomial logistic regression was conducted to determine if the data source, group type, data set (Woodland or Whitcomb), leader, or group predicted how many units exist in each of the common categories from the combined data (see Appendix N).

Table 3.8

Combined Dimension Frequency Percentages by Data Source

Dimension	Leader Slips	Debrief Int.	Total
Pre-Group Review	40.06	34.42	36.44
Pre-Group Planning	20.48	11.33	14.61
In-Group	21.38	22.00	21.78
Effect	18.09	32.25	27.18
Total Units	669	1,200	1,869

Note. $\chi^2(3) = 60.1250$, $p < 0.01$

Table 3.9

Combined Dimension Frequency Percentages by Group Type

Dimension	Process	Specific Focus	Total
Pre-Group Review	34.91	42.20	36.44
Pre-Group Planning	11.71	25.58	14.61
In-Group	23.68	14.58	21.78
Effect	29.70	17.65	27.18
Total Units	1,478	391	1,869

Note. $\chi^2(3) = 73.5299$, $p < 0.01$

Table 3.10

Combined Dimension Frequency Percentages by Data Set

Dimension	Woodland	Whitcomb	Total
Pre-Group Review	40.57	34.25	36.38
Pre-Group Planning	14.74	14.50	14.6
In-Group	22.82	21.19	21.6
Effect	21.87	30.06	27.4
Total Units	631	1,241	1,872

Note. $\chi^2(3) = 15.3943$, $p < 0.01$

Table 3.11

Combined Dimension Frequency Percentages by Leader

Dimension	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	18	Total
Pre-Group Review	26.3	46.2	41.0	39.0	28.8	23.7	55.9	40.0	29.1	57.1	52.9	34.4	22.2	25.9	21.5	30.0	36.4
Pre-Group Planning	6.1	9.1	8.6	36.8	14.7	25.4	25.4	8.4	9.5	23.8	17.7	13.3	44.4	12.4	5.0	25.0	14.6
In-Group	21.1	14.9	20.0	11.0	31.3	30.5	10.2	36.8	25.7	14.3	26.5	20.0	0.0	21.6	38.0	20.0	21.8
Effect	46.5	29.8	30.5	13.2	25.3	20.3	8.5	14.7	35.8	4.8	2.9	32.2	33.3	40.1	35.5	25.0	27.2
Total Units	114	329	105	136	198	59	177	95	179	21	34	90	9	162	121	40	1,869

Note. $\chi^2(45) = 312.8590$, $p < 0.01$

Table 3.12

Combined Dimension Frequency Percentages by Group

Dimension	2	3	5	8	10	11	13	15	17	19	21	24	25	27	29
Pre-Group Review	26.3	51.4	38.5	42.4	54.2	32.3	31.3	57.7	32.4	31.0	57.1	52.9	33.3	23.7	14.8
Pre-Group Planning	6.1	31.4	7.7	9.1	45.8	9.2	25.0	32.7	5.4	9.5	23.8	17.7	14.3	10.8	25.9
In-Group	21.1	8.6	20.5	19.7	0.0	38.5	40.6	3.9	54.1	28.6	14.3	26.5	28.6	31.2	18.5
Effect	46.5	8.6	33.3	28.8	0.0	20.0	3.1	5.8	8.1	31.0	4.8	2.9	23.8	34.4	40.7
Total	114	35	39	66	24	65	32	52	37	42	21	34	21	93	27
Groups cont.	32	33	34	36	38	40	42	44	50	52	53	57	60	61	Total
Pre-Group Review	55.2	43.2	35.7	30.2	34.4	28.3	22.2	53.7	21.5	25.0	30.0	22.2	51.4	38.6	36.4
Pre-Group Planning	22.4	5.3	34.8	22.6	13.3	14.1	44.4	10.5	5.0	13.8	25.0	9.5	8.1	6.8	14.6
In-Group	12.8	15.0	13.4	24.5	20.0	19.2	0.0	17.9	38.0	30.0	20.0	25.4	24.3	11.4	21.8
Effect	9.6	36.6	16.1	22.6	32.2	38.4	33.3	17.9	35.5	31.3	25.0	42.9	16.2	43.2	27.2
Total Units	125	227	112	53	90	99	9	67	121	80	40	63	37	44	1,869

Note. $\chi^2(84) = 398.9822$, $p < 0.01$

Omnibus post-test estimation tests showed that group type was no longer found to be a significant predictor ($p = .26$), but that data source, leader, group, and data set all significantly predicted the number of units present in each dimension ($p < .01$). For data source, results indicate that when compared with Pre-Group Review, units are significantly more likely to come from leader slips in the Pre-Group Planning dimension and debrief interviews in the Effect dimension. For data set, results show units are significantly more likely to come from the Whitcomb data set for the Effect dimension when compared with the Pre-Group Review dimension.

Pairwise comparisons for combined data produced 480 within-dimension comparisons with 157 significant differences for leader, and 1,624 within-dimension comparisons with 414 significant differences for group. Because there was a higher number of leaders and groups in the combined data, we had more power to detect variability. This explains why the combined data revealed several leaders and groups that used feedback significantly differently from the majority, while previous pairwise comparisons produced only one or two leaders and groups that stood out in each dimension. The pattern of variability in leader use, however, was observed in both the combined data and the previous analysis. Variability was also noted in groups led by the same leader, with the majority of leaders who led multiple groups showing significant variability in the way feedback was used from one group to another.

Quantification

A central goal of the current dissertation is to provide rationale for a method to quantify leader GQ use in a way that could be used as a mediator variable in the parent RCT. The GQ Use Questionnaire was developed as a potential means of accomplishing this goal in the present study. As a reminder, the items for this instrument were developed using categories from the

Woodland pilot study. Of the 16 leaders that participated in the parent study, 15 completed a GQ Use questionnaire for each semester-long feedback group they led, totaling 33 GQ Use Questionnaire responses. The questionnaire as it was administered to leaders has eight items (see Appendix F), but because the last item referred to OQ use (for purposes specific to the Woodland study), only the first seven items capture GQ use. Chronbach's alpha for these seven items is .74, which is considered to be in the acceptable range for research purposes given the early stages of the questionnaire's development (Nunnally, 1978).

The seven items included in the GQ Use Questionnaire correspond with seven of the common categories of GQ use from the combined qualitative data set. Six of these items fell into the seven most frequently endorsed categories (see Table 3.7), providing support for their importance in the minds of group leaders. These include Review of Feedback, Analysis of Feedback, Education about Feedback, Planning Interventions, Explicit Use, Non-Explicit Use, and Decisions to Withhold Feedback. When the GQ Use Questionnaire was developed, five categories were not included in the questionnaire because they did not directly assess leader use and they were not highly endorsed by group leaders. The primary reasons for eliminating each were as follows:

- Group-Initiated Feedback Use was excluded because it refers to instances in which members, not leaders, bring up feedback, and therefore cannot be considered as leader use.
- Looking Forward refers to units in which leaders express anticipation for future results, but it does not describe implementing feedback in any way.
- Value/Opinion was excluded because we determined that while value may impact a leader's appreciation for or feelings toward the GQ, it does not in itself capture

behavior and thus cannot be directly included as part of leader use.

- Ambiguous Use was excluded because while it reflects some action, it provides no description about how the GQ was used.
- Effect on Group was a frequently endorsed category, but was excluded because effects of feedback use were determined only to be related to feedback use, but not indicative of use. For example, debrief interview unit 2.12.1 states, "I think the effect on the group process as a whole was positive." Additionally, five of the seven questions in the debrief interview asked explicitly about the effect of use, so the high proportion of Effect units may be related to the prompt the leaders were given. Ultimately we determined that developing an item that asks leaders if they evaluated the effect of their use did not help us better measure the use itself.
- Awareness of Alerter does capture use, but was not included in the GQ Use Questionnaire because at the time of development, it only made up a little over 1% of the units from the pilot study. However, after finding more support for this category in the current study, consideration for its inclusion in the future seems warranted and is explored in the discussion section.

Given the promising alpha and the fact that the GQ Use Questionnaire items capture the common categories most relevant to use and most frequently endorsed (with one exception), I conducted an analysis to determine how well the GQ Use Questionnaire correlated with frequency data collected in the qualitative study. In other words, I wanted to know if leader responses to a brief questionnaire assessing global GQ Use would correlate with the time-intensive qualitative assessments of GQ Use that relied upon data generated in session-by-session and global reports. I computed a Spearman correlation using a rank for each leader based

on the total GQ Use Questionnaire score and the total number of qualitative units produced by the group leader. I calculated a second rank using only the common categories in Table 3.7 that paralleled the seven items from the GQ Use Questionnaire. Finally, I computed a Pearson correlation that used the raw data from which the aforementioned ranks were created. The results of these analyses are reported in Table 3.13. The Spearman rank order correlations are high—ranging from .78 to .84—with comparable values reported by the Pearson estimates.

Table 3.13

Correlation Between GQ Use Questionnaire and Qualitative Units

n = 15	Total Qualitative Units	7 Main Qualitative Categories
Spearman	0.78***	0.84***
Pearson	0.88***	0.90***

***$p < .01$

After completing all of the main analyses for my dissertation, I decided to conduct two exploratory analyses to better understand my results and potentially guide future research. First, I wanted to know how the correlations between the GQ Use Questionnaire responses and the qualitative data would be impacted if I disaggregated the qualitative data and examined leader slips separately from debrief interviews. The goal of this analysis was to understand the degree to which the correlations were impacted by considering session-by-session explanations of behavior (leader slips) and episode-long summaries of use (debrief interviews) separately. Both Spearman and Pearson analyses for leader slips produced strong correlations, but the correlations for debrief interviews were lower than the leader slip session-by-session ratings which is promising since the latter are closer to real-time reporting of GQ use (see Table 3.14).

Table 3.14

Correlation Between GQ Use Questionnaire and Disaggregated Qualitative Units

n = 15	Leader Slips		Debrief Interviews	
	Total Units	7 Main Qualitative Categories	Total Units	7 Main Qualitative Categories
Spearman	0.76**	0.80**	0.56*	0.68**
Pearson	0.79**	0.78**	0.69**	0.74**

*$p < .05$. **$p < .01$

Given the promising Chronbach's alpha and strong correlations between leader responses to the GQ Use Questionnaire and qualitative findings, the second and final exploratory analysis was performed to determine each item's contribution to the total Chronbach's alpha (see Table 3.15). This was intended to provide future researchers with directions for improving the measure and also helped me better understand what was contributing to the above correlations. It appears that the full scale alpha would decrease if items 4 or 6 were taken out of the scale, and that it would increase if item 7 were removed. Descriptive statistics show similar endorsement across possible responses for these items (see Table 3.16).

Table 3.15

Individual-Item Analysis of the GQ Use Questionnaire

Item	Qualitative Category Assessed	Average Inter-item Covariance	Alpha if Item Deleted
Item1	Review	0.53	0.71
Item2	Analysis	0.58	0.74
Item3	Education	0.56	0.72
Item4	Planning Interventions	0.45	0.68
Item5	Explicit Use	0.51	0.71
Item6	Indirect Use	0.36	0.65
Item7	Withhold	0.52	0.77

Note. Full scale alpha = .74

Table 3.16

Frequency Percentages by Response Value for GQ Use Questionnaire Items

Response	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
0	2.94	2.94	17.65	2.94	11.76	11.76	29.41
1	0.00	0.00	50.00	5.88	44.12	20.59	14.71
2	0.00	2.94	14.71	44.12	26.47	11.76	17.65
3	0.00	0.00	14.71	14.71	5.88	17.65	17.65
4	23.53	23.53	0.00	17.65	2.94	23.53	0.00
5	70.59	67.65	0.00	11.76	5.88	11.76	17.65
Missing	2.94	2.94	2.94	2.94	2.94	2.94	2.94

Note. N = 34

Discussion

The current dissertation was born out of a need to understand an important construct in our randomized controlled trial of GQ and OQ feedback: acting on GQ feedback. This parent study tests the effects of using outcome and process feedback in group treatment. The study's authors hypothesized that leaders who received and acted on process feedback (GQ) would have clients who showed faster and larger returns to normative ranges on the GQ, higher rates of attendance, and greater symptom reduction. In order to accurately interpret these hypotheses, it is essential to understand what it means to act on feedback. Thus, the purpose of the current dissertation was first to define what it means to act on GQ feedback by systematically replicating a companion pilot study (Woodland, 2015), and second, to provide a potential fidelity check for the parent RCT for eventual mediation analysis. The first purpose was achieved through the qualitative content analysis, and the second was achieved through the initial development of the GQ Use Questionnaire. Collectively, the results from the current dissertation produced three main findings. First, a strikingly similar qualitative definition of what it means to act on GQ feedback was established using the same method on two separate sets of data, and with two

separate sets of raters. Second, there is significant variability in the way leaders use feedback, both compared with other leaders, and compared with their own feedback use in different groups. Third, there is surprisingly high correspondence between leader's qualitative description of their feedback use (both session-by-session and by group episode) and their responses to the GQ Use Questionnaire, a seven-item questionnaire that summarizes use by group episode.

The content analysis from the current dissertation mirrored findings from the Woodland pilot, yielding 15 main categories that capture different ways that leaders used the feedback, as well as content related to leader use. Raters identified a temporal pattern of use with three dimensions, Pre-Group Use, In-Group Use, and Post Group Use/Effect, into which the 15 main categories were grouped. One noteworthy finding is that slightly more than half of the units fell into one of two categories: Analysis of Feedback (n = 504) and Effect on Group (n = 321). This supports the notion that much of what leaders do with feedback happens before or after an actual intervention is implemented. But of course, the third most highly populated category was GQ in Group (n = 301), which shows that for many leaders, what happens in the group accounts for a large portion of use.

When comparing the Woodland and Whitcomb studies,13 common categories were established that capture 94.1% of the combined 2,150 units analyzed in the two studies. These common categories include Review of Feedback, Analysis of Feedback, Planning Interventions, Decision to Withhold Feedback, In-Group Use (Explicit), In-Group Use (Non-Explicit), Education about GQ, Awareness of Alerters, Looking Forward, Effects of Feedback Use, Ambiguous Use, Value/Opinion, and Group-Initiated Feedback Use. These common categories are summarized by four temporal dimensions: Pre-Group Review, Pre-Group Planning, In-Group Use, and Effect. The units from each study that were not included in these common categories

were sparsely populated and did not directly reflect use: Attendance, Filling Out Measures, Self-Awareness, Proficiency with GQ, Disruption in Data, and Leader Confusion. The substantial overlap between the Whitcomb and Woodland results strengthens the validity of our final definition of what it means to act on GQ feedback using 16 group leaders who ran 30 feedback groups.

The second main finding is the observed variability between leaders in their reported GQ use, based on the total number of units they provided. Frequency tables showed substantial variability between individual leaders and groups. Results showed that some leaders use the GQ much more than others, and some leaders seem to have one or two categories in which the majority of their units fall, but these categories of use are not the same across leaders. This indicates that leaders often favor a specific way of using the feedback, and that there is no universal pattern across leaders. Frequency tables also show a difference in use by Data Source. Specifically, about 60% of use in the leader slips was identified in the Pre-Group Review and Pre-Group Planning dimensions, while 40% was identified in the In-Group and Effect dimensions. The opposite pattern was found for debrief interviews, with about 45% of use identified in the Pre-Group use dimensions and about 55% in the In-Group and Effect dimensions. This indicates a differential focus in leader responses depending on how they were reporting. This may have been influenced by the open ended prompt in the leader slip compared with the more directive format of the debrief interview.

When considering the data from the Whitcomb data alone, multinomial logistic regression findings showed significant variability in leader use between different data sources, group types, leaders, and groups. Direct comparisons between individual leaders and groups

provide more evidence for this variability with a few individual leaders and groups standing out as using feedback significantly differently from other leaders and groups.

Even more variability was observed when considering the combined data, with data source, data set, leader, and group predicting the number of units by dimensions. Group type was no longer found to be a significant predictor, which was not surprising given that all but one of the groups from the Woodland were process groups. Direct comparisons between individual leaders and groups showed substantial variability across dimensions with several leaders and groups standing out with either high- or low-use profiles. Additionally, the majority of leaders who led more than one group were found to report using feedback differently across groups and categories of use. This finding is important in that it supports the idea that a specific group can impact the way leaders use GQ feedback. Leaders often alluded to this phenomenon in the debrief interviews. For example, some mentioned that they might have used the feedback more if they had been leading a process group, but found fewer opportunities in their structured specific-focus groups. Others reported that their groups seemed to be doing well based on low number of alerts, and they therefore felt that intervention was unnecessary. Finally, some leaders said that their members indicated a preference about how they wanted leaders to use the GQ in that particular group, which leaders reported shaped their feedback use.

An additional multinomial logistic regression testing the relationship between leader use and the number of negative alerts per session, the number of people alerting, and the proportion of people alerting provides more evidence for variability in leader use by group. Findings show that there was not a significant relationship between use and the total number of negative alerts per session, but that the number of people alerting and the proportion of people alerting both predict more Pre-Group Use than In-Group Use. This may indicate an increased attempt to

meaningfully analyze data and ponder how to use it most effectively before the group begins. This analysis only speaks to In-Group Use relative to Pre-Group use, and does not speak to whether or not there was an absolute change in In-Group Use in the presence of an increased number and proportion of people alerting. In other words, the results do not necessarily indicate a reduction in In-Group use when more alerts are present, it just shows a greater increase in Pre-Group use. The fact that leaders use the GQ differently from each other and in different groups they run suggests a leader-by-group interaction for GQ use. In the current study, we did not have enough groups to be able to test this interaction, but if this proves to be the case, differences between leaders and groups may be found to mediate outcome and dropout in group therapy.

The last main finding is the strong correlation between leader's qualitative description of their feedback use and their responses to the GQ Use Questionnaire. This questionnaire summarizes the central definition of GQ Feedback use by capturing each of the seven main common categories from the qualitative analysis (Review of Feedback, Analysis of Feedback, Education about Feedback, Planning Interventions, Explicit Use, Non-Explicit Use, and Decisions to Withhold Feedback) in an individual question. It quantifies this use by attaching a Likert scale to each question for each leader to rate their own use. It also provides some support for summing the items in the GQ Use Questionnaire to obtain a total use score for every leader. This is one way of capturing what it means to use GQ feedback.

In order to add support for the use of the GQ Use Questionnaire as a potential mediator variable for the parent RCT, I used Spearman rank and Pearson correlations to determine how well the questionnaire summarized the information from the qualitative analysis. A strong correlation between the total leader GQ Use Questionnaire scores and (a) the total qualitative units produced by leader, and (b) the total qualitative units from the seven categories that are

explicitly covered in the GQ Use Questionnaire was found ($p < .001$). These correlations indicate a significant overlap between information generated by the GQ Use Questionnaire and the qualitative data. In short, the questionnaire appears to summarize the qualitative findings well. Additionally, using the questionnaire eliminates some of the limitations of using qualitative findings alone. For example, using unit counts exclusively can award leaders a higher unit-total-use score simply because they turned in more leader slips or used more verbose speech during the debrief interviews to describe the same use another leader could describe in fewer words. Therefore, considering the questionnaire's inclusion of the most salient qualitative categories, the correlation findings, and the limitations of the qualitative data that are avoided with the questionnaire, the GQ Use Questionnaire may be a parsimonious way to quantify GQ use.

After establishing initial support for the reliability and validity of the GQ Use Questionnaire, two exploratory analyses were performed to guide future research. The first analysis disaggregated the qualitative data in the correlation with the GQ Use Questionnaire to examine leader slips and debrief interviews separately. Results showed a higher correlation for leader slips than debrief interviews. This provides evidence for the criterion validity of the GQ Use Questionnaire and its potential use as a fidelity check for leader feedback use in future studies. The stronger relationship between the GQ Use Questionnaire and session-by-session reports of behavior as opposed to episode-long summaries provided in the debrief interviews is particularly promising since the former reflects ratings that are closer in time to actual use.

The second exploratory analysis was conducted to guide future research on the GQ Use Questionnaire itself. The secondary benefit of the development of the GQ Use Questionnaire is that it has potential for use as a fidelity check in future GQ studies. If additional psychometric testing yields similarly promising results, using the GQ Use Questionnaire would provide a quick

and simple way to measure acting on GQ feedback, avoiding years to conduct a qualitative analysis. Thus, I performed an analysis to test the inter-item variance of the GQ Use Questionnaire to direct future psychometric research. It appears one item (item 7) may be decreasing the overall Chronbach's alpha, and that removing it may increase the full scale alpha, and thereby represent a more unified construct. Therefore, future researchers may consider deleting this item. Another suggestion on how to improve the measure is to add a question to assess the Awareness of Alerter category, thereby providing coverage for every common use category established in the Woodland and Whitcomb studies. Additionally, researchers and authors of the parent RCT might consider weighting certain items more than others. This is because it is likely that certain areas of use as measured by the GQ Use Questionnaire are more salient to acting on GQ feedback than others. This prediction is based on logical supposition. For example, it seems likely that a leader's use of the GQ in group will affect outcomes (OQ scores, GQ scores, and dropout) more than a leader simply opening and reviewing the feedback report. Accordingly, items that reflect more use could be given a higher point value toward the items' sum total. Finally, past research using feedback systems has identified therapist awareness of feedback as the mechanism of change in client outcomes (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012). Future researchers may consider comparing GQUQ items that assess therapist awareness (Item 1- Review of Feedback, and a new item that assesses Awareness of Alerters if it is added to the scale) with the total GQUQ to determine if there is an additional improvement in client outcome associated with other forms of use over and beyond simple awareness.

In conclusion, the objectives of the current dissertation, to define what it means to act on GQ feedback and to provide a fidelity check for the parent RCT, were both successfully completed. The findings that resulted from this study are key to defining GQ Use, interpreting

the hypotheses of the parent RCT, and potentially contributing to a means of measuring feedback use that can be used for GQ studies conducted in the future. If this feedback use is determined to be a mediator of outcome and dropout in the parent study, measuring this construct may become increasingly important.

Limitations

There are a number of limitations to the current study. First, not all leaders submitted a leader slip after every session. As such, we had missing data that affected our impression of leaders' use. An attempt to address this limitation was the use of leader debrief interviews in addition to leader slips in order to get information from leaders even when they did not consistently report their use on a session-by-session basis. Another attempt to address this limitation was using the GQ Use Questionnaire which provided an opportunity for leaders to summarize their overall group use without the session-by-session reports. However, one leader did not submit a response to the questionnaire, and therefore we still have missing data for one leader. Another potential limitation that some might wonder about is experimenter expectancies. However, by using blind raters who do not have any association with the lab that created the GQ (outside of their participation in this study), they did not report, nor were they likely to have, any significant bias that would alter what is identified in the slips, so this is not likely to be a significant threat to the validity of the study.

Lastly, a potential limitation is the self-report nature of our data. This applies both to the qualitative data and the data acquired from the GQ Use Questionnaire. Leaders themselves reported their use of the feedback, which may not perfectly represent the reality of their use. In the future, researchers might consider recording sessions and having trained raters code sessions for use based on the categories established in this study. Again, our current resources did not

permit this, nor did it fit the scope of the study. Also, without an understanding of what it means to use feedback gained through the qualitative analysis, this type of coding would not be possible, so the source of our data seems like a necessary first step. Additionally, several leaders have blatantly reported their lack of use of the measure, so it does not seem as though leaders feel the need to exaggerate their use. Similarly, underreporting use is not likely as the experimenter developed the measure. As such, this limitation also does not appear to be a significant threat to the validity of the study.

References

Auld, F., & White, A. M. (1956). Rules for dividing interviews into sentences. *Journal of Psychology, 42,* 273–281.

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., Mellor-Clark, J., Connell, J., Evans, R., Evans, C., & Margison, F. (2010). The CORE measures & CORE system: Measuring monitoring and managing quality evaluation in the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 175–219). Chichester, United Kingdom: Wiley.

Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a rigorous and relevant knowledge-base for the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 21–61). Chichester, United Kingdom: Wiley.

Berking, M., Orth, U., & Lutz, W. (2006). Wie effektiv sind systematische rückmeldungen des therapieverlaufs an den therapeuten? Eine empirische studie in einem stationärverhaltenstherapeutischen setting. [How effective is systematic feedback of treatment progress to the therapist? An empirical study in a cognitive-behavioral-oriented inpatient setting.]. *Zeitschrift Für Klinische Psychologie Und Psychotherapie: Forschung Und Praxis, 35*(1), 21–29.

Bormann, B., Burlingame, G., & Strauss, B. (2011). Der Gruppenfragebogen (GQ-D). Instrument zur Messung von therapeutischen Beziehungen in der

Gruppenpsychotherapie. [The group questionnaire (GQ-D). Instrument to measure therapeutic relationships in group psychotherapy.]. *Psychotherapeut*, *56,* 297–309.

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.

Burlingame, G. M., & Beecher, M. E. (2008). New directions and resources in group psychotherapy: Introduction to the issue. *Journal of Clinical Psychology: In Session, 64,* 1197–1205.

Burlingame, G. M., Cox, J. C., Davies, D. R., Layne, C. M., & Gleave, R. (2011). The Group Selection Questionnaire: Further refinements in group member selection. *Group Dynamics: Theory, Research, and Practice, 15*(1), 60–74.

Burlingame, G. M., Fuhriman, A., & Johnson, J. (2002). Cohesion in group psychotherapy. In J. Norcross (Ed.), *A guide to psychotherapy relationships that work* (pp. 71–88). New York, NY: Oxford University Press.

Burlingame, G. M., Fuhriman, A., & Mosier, J. (2003). The differential effectiveness of group psychotherapy: A meta-analytic perspective. *Group Dynamics: Theory, Research, & Practice, 7*(1), 3–12.

Burlingame, G., Gleave, R., Erekson, D., Nelson, P., Olson, J., Thayer, S. & Beecher, M. (2015). Differential effectiveness of group, individual, and conjoint treatments: An archival analysis of OQ-45 change trajectories. *Psychotherapy Research.* doi:10.1080/10503307.2015.1044583.

Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2002). Zum aktuellen Stand der Gruppenpsychotherapieforschung. II. Effekte von Gruppenbehandlungen als Bestandteil Kompleaer Be handlung sangebote. *Gruppenpsychotherapie und Gruppendynamik, 38,* 5–32.

Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy, 48*(1), 34–42.

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.). New York, NY: John Wiley & Sons.

Burlingame, G. M., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K. R., Ogrodniczuk, J., et al. (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, NY: Guilford Press.

Chapman, C., Burlingame, G., Rees, F., Gleave, R., Beecher, M., & Porter, G. (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.

de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22,* 464-474.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research, 22,* 175–186. doi: 10.1177/1046496491222002

Fuhriman, A., & Burlingame, G. M. (1994). Group psychotherapy: Research and practice. In A. Fuhriman & G. M. Burlingame (Eds*.), Handbook of group psychotherapy: An empirical and clinical synthesis* (pp. 3–40). New York, NY: Wiley.

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.

Gallagher, R. P. (2009). *National survey of counseling center directors.* Alexandria, VA: International Association of Counseling Services.

Gallagher, R. P. (2010). *National survey of counseling center directors.* Alexandria, VA: International Association of Counseling Services.

Gallagher, R. P. (2011). *National survey of counseling center directors.* Alexandria, VA: International Association of Counseling Services.

Gallagher, R. P. (2012). *National survey of counseling center directors.* Alexandria, VA: International Association of Counseling Services.

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.

Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychotherapy Research, 17*(4), 379–392.

Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L., & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychotherapy Research, 14*(3), 308–327.

Hill, C. E., Greenwald, C., Reed, K. G., Charles, D., O'Farrell, M. K., & Carter, J. A. (1981). *Manual for the counselor and client verbal response category systems.* Columbus, OH: Marathon.

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.

Johnson, J., Burlingame, G. M., Olsen, D., Davies, 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.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.

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.

Kopta, S. M., & Lowry, J. L. (2002). Psychometric evaluation of the behavioral health questionnaire-20: A brief instrument for assessing global mental health and the three phases of psychotherapy outcome. *Psychotherapy Research, 12,* 413–426.

Kraus, D., & Castonguay, L. G. (2010). Treatment outcome package (TOP)—Development and use in naturalistic settings. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 155–174). Chichester, United Kingdom: Wiley.

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.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York, NY: Wiley.

Lambert, M. J., Burlingame, G. M., Umphress, V. J., Hansen, N. B., Vermeersch, D., Clouse, G., & Yanchar, S. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 3,* 106–116.

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.

Lambert, M. J., Hansen, N. B., & Harmon, S. C. (2010). Outcome questionnaire system (the OQ system): Development and practical applications in health-care settings. In M. Barkham, G E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 141–154). Chichester, United Kingdom: Wiley.

Lambert M. J., Whipple J. L., Harmon C., Shimokawa K., Slade K., & Christofferson, C. (2004). *Clinical support tools manual.* Provo, UT: Brigham Young University.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science & Practice, 10,* 288–301.

McRoberts, C., Burlingame, G. M., & Hoag, M. J. (1998). Comparative efficacy of individual and group psychotherapy: A meta-analytic perspective. *Group Dynamics: Theory, Research, and Practice, 2*(2), 101–117.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy, 2,* 91–100.

Miller, S. D., Hubble, M. A., Duncan, B. L., & Wampold, B. E. (2010). Delivering what works. In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (2nd ed., pp. 421–429). Washington, DC: American Psychological Association.

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.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Pinsof, W. M. (1986). The process of family therapy: The development of the family therapist coding system. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process* (pp. 201–284). New York, NY: Guilford.

Schreier, M. (2012). *Qualitative content analysis in practice.* Los Angeles, CA: SAGE Publications.

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.

Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy, 15,* 287–303.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). Comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology & Psychotherapy*, *13*(3), 202–214.

Stinchfield, R. (1988). *Development and application of a measure of therapist directives* (Unpublished doctoral dissertation). Brigham Young University, Utah.

Stinchfield, R. D., & Burlingame, G. M. (1991). Development and use of the directives rating system in group therapy. *Journal of Counseling Psychology, 38*(3), 251–257.

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.

Thayer, S. D., & Burlingame, G. M. (2014). The validity of the group questionnaire: Construct clarity or construct drift? *Group Dynamics: Theory, Research, and Practice, 18*(4), 318–332.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77–83.

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.

Woodland, S. (2014). *Process feedback in group psychotherapy: A qualitative inquiry into leader implementation of GQ feedback* (Unpublished doctoral dissertation). Brigham Young University, Utah.

Yalom (2005). *The theory and practice of group psychotherapy* (5th ed.). New York, NY: Basic Books.

Appendix A

Weekly 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 OQ: None
DID NOT ATTEND: 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 OQ feedback. List any specific member targeted.
*Leader Name*_____ *Date of session where feedback was implemented:* _____

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Appendix B

Weekly GQ 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 10th percentile ☹	Clients at or above the 95th 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

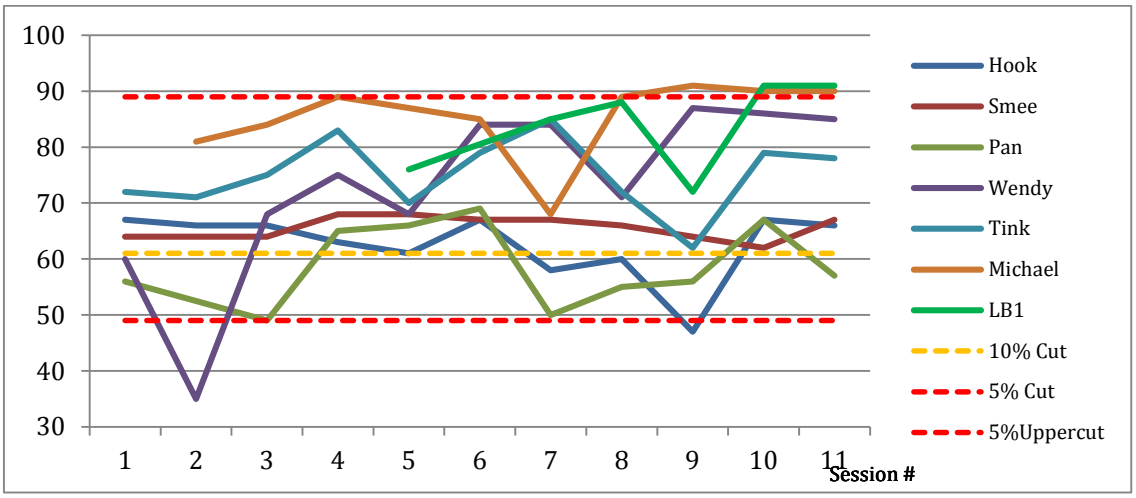
RELATIVE ALERTS—based on last group session GQ		
	Clients reporting reliable negative change	Clients reporting reliable positive change
Positive Bond	Peter Pan	None
Positive Work	None	None
Negative Relationship	None	None

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

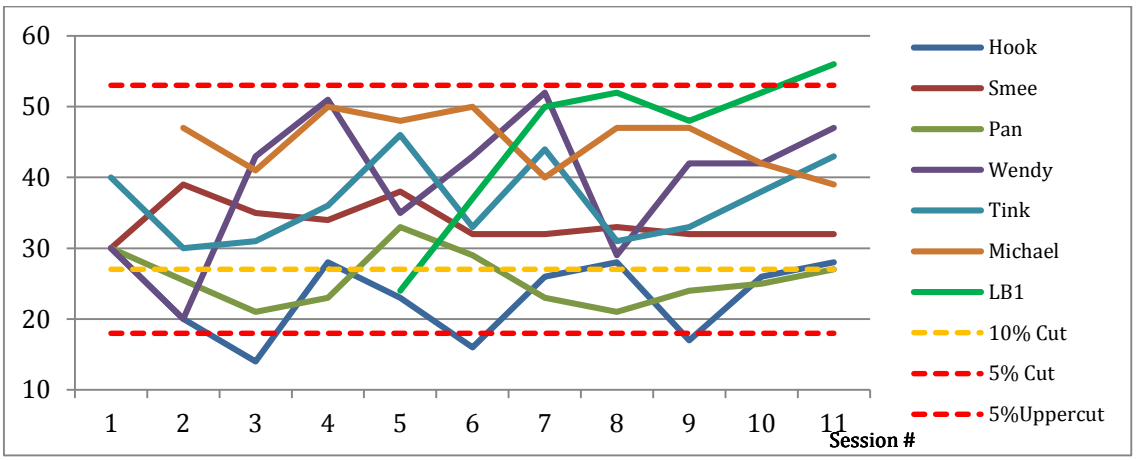
List actions (if any) that you took based upon last week's GQ feedback. List any specific member targeted.
*Leader Name*_____ *Date of session where feedback was implemented:* _____

_____
_____
_____
_____
_____
_____
_____
_____

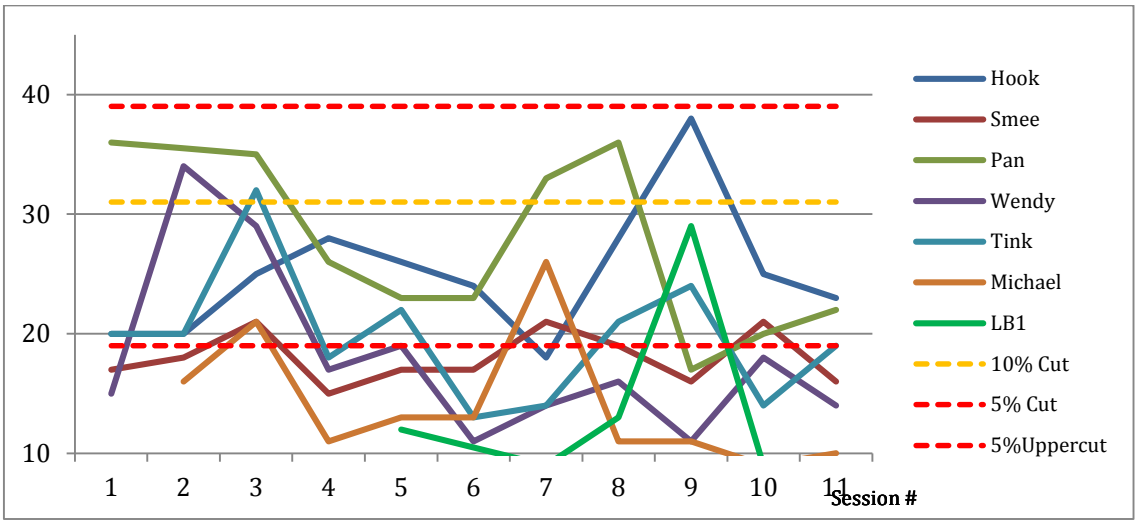
POSITIVE BOND BY TIME



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

Appendix C

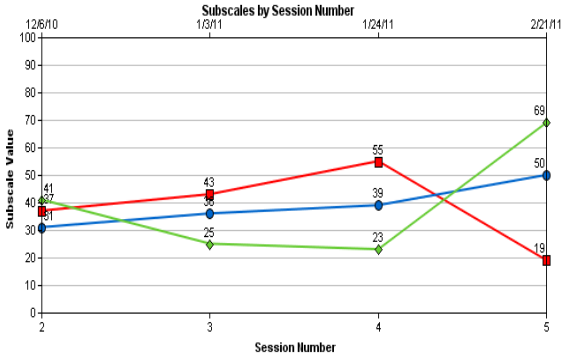
Sample GQ Feedback Report Currently Found on OQ-Analyst

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

Comparitive Group: Counseling Center

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 D

GQ 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 feedback and in the other I will not be receiving feedback. A coin toss (random assignment) is used to assign your group as either a GQ feedback group, or a non-feedback group. In either case, it is important that you complete the GQ every time you attend 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 your for this time. You will be given $10 after you sign up for the study and complete the initial two measures (OQ & GQ). For every session after that, you will earn $5 for each group session you attend where you complete both measures. The OQ can be taken any time within the week before the session, and the GQ should be completed immediately after**

group. If for some reason you need to leave immediately after, you can take the GQ online, but it must be completed by 12:00pm the day before your next group session. The $10 will be given to you immediately and the remaining amount earned in future sessions 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 and extra $20 as a thank you. The total amount you could earn depends on how many sessions you attend, and how many sets of measures you complete.

- 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 E

Interview Questions for GQ 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. 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 feedback over time?
4. If you were to give a percentage of agreement between your own perceptions as a leader and the GQ 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 F

GQ Use Questionnaire (Example)

The following questions are related to feedback you received as part of the Burlingame & Beecher GQ Feedback Study you participated in from 2012–2014.

You received feedback for:

9 sessions for your GAP Group from 2/7/13–4/4/13
11 sessions for your ASD Group from 9/20/13–12/13/13
13 sessions for your ASD Group (longitudinal) from 1/17/1–4/11/14

Using the following Likert scale, please indicate what percentage of the sessions you implemented the following types of use in each group you ran.

0	1	2	3	4	5
0%	1-19%	20-39%	40-59%	60-79%	80-100%

	What percentage of the sessions did you…	**GAP**	**ASD**	**ASD (longitudinal)**
1.	**… review the GQ feedback you received?**			
2.	**… notice changes in client scores when looking at the report?**			
3.	**… educate members about the GQ itself?**			
4.	**…use GQ feedback to inform and/or guide specific interventions?**			
5.	**…specifically mention the GQ results to the group?**			
6.	**…use the GQ to inform in-group interventions *without* specifically mentioning the GQ scores or subscales?**			
7.	**…deliberately choose not to bring the feedback in to the group?**			
8.	**…review the OQ alert feedback you received?**			

Appendix G

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 completive 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 uh 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 Woodland and Whitcomb Study:

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 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 coleader 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 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 feedback." (Three separate units)

Appendix H

Woodland Coding Glossary—Use

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

Example: "my coleader and I reviewed the feedback before group."

Reaction to Scores: Any time that the leader (alone or with coleader) reports the content, an interpretation, or a speculation of the results of the feedback report. Subcategories include coleader 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 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 coleader) reports actions taken in response to the feedback report that are meant to prepare for use or implementation in the next session. Subcategories include coleader 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 coleader) reports choosing to not use the data from the feedback report. Subcategories include the involvement of

a coleader and reasons for withholding feedback. Reasons include (a) lack of time available to review or implement feedback, (b) not enough information in the report to create a meaningful intervention, (c) group member preference to not hear about the feedback report, (d) decision to observe or be aware of alerters without actively implementing feedback, (e) withholding based on any number of other extenuating circumstances, and (f) 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 coleader) brings up scores, trends, or concepts pulled directly from the GQ feedback report. Subcategories include coleader 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 coleader) 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 Positive 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 coleader) reports in-group effects that were either a direct or indirect result of implementing feedback. Subcategories include coleader 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 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 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.

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."

Undecided: Any unit that does not fit into the above-described categories, but still seems relevant to use or value of the GQ feedback.

Example: "Jon and I both commented after the session that it felt like our bringing the feedback from the GQ into the session had the effect of distancing us from the members of the group."

Semi-Relevant: Any unit that seems to not relate to GQ 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 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." There is no mention of leader behavior in the unit or in adjacent units where use may be interpretable in context of member behavior.

Appendix I

Whitcomb Codebook

Dimension 1: Pre-Group

Review of Feedback: Any time a leader/coleader makes a quick, basic observation about the GQ feedback, but does not interpret or analyze the data.

Example: "My coleader and I looked at our results."

Sharing Feedback with Others (Coleader): Any time a leader describes distributing the actual feedback report itself to his/her coleader (n = 42).

Example: "I emailed this feedback to my coleader prior to group."

Analysis of GQ Feedback: Any time leaders describe studying, scrutinizing, or interpreting the GQ feedback.

Example: "We were pleased to see A's positive relative alert."

Proficiency with GQ: Any time leaders describe their own familiarity or expertise in using and understanding the GQ.

Example: "I kind of jumped in not knowing much about the instrument."

Disruption in Data: Any time leaders describe interruptions in the delivery of the feedback report.

Example: "Apparently we encountered some email snafus."

Dimension 2: In-Group

GQ in Group: Any time leaders describe a specific time they used the GQ feedback during a group session.

Example: "We used the overall feedback to discuss differing dynamics in the group when different members are present."

Leader Confusion: Any time leaders express confusion about the GQ, how to use it, or about the study itself.

Example: "Yeah I just thought, it's like 'What does that mean? I don't know even how to answer. No we don't have mutual group goals so, are we bad?'"

Follow-Up Outside of Group: Any time leaders have a conversation with a group member based on GQ feedback immediately following the group session.

Example: "We asked the group member to stay after group/ and we talked with him about the results."

Member Reaction: Any time leaders describe member thoughts, feelings, or behaviors that occur in group in response to GQ feedback or related interventions.

Example: "Most of them would take a look at it and just nod."

Sharing Feedback with Others (Members): Any time a leader brought copies of the feedback report to group members in group.

Example: "At the beginning of group we gave each member their individual feedback form for review."

Dimension 3: Post-Group/Effect

Effect on Group: Any time leaders describe the impact, outcome, or consequence of using the GQ feedback on group members or leaders.

Example: "If anything I would say it gave me more patience with the silence as they struggled with how to respond because I knew that it was feedback that they had given."

Expectations of Future GQ Scores: Any time leaders express thoughts, assumptions, and predictions for future GQ feedback scores.

Example: “I’ll be interested to see if working hard and challenging each other have lower scores this week also.”

Opinion of GQ: Any time leaders or members express an evaluation of or attitude toward the GQ.

Example: “I’ve enjoyed having the feedback.”

Hindsight Opinion. Any time leaders comment in retrospect about the GQ or its use.

Example: “So maybe getting [the report] a day or two before the group would be helpful.”

Overall GQ Use. Any time a leader states that they used the feedback, but does not offer specific details about how they used it.

Example: “I used just the GQ.”

Irrelevant: Any unit that contains information that does not relate in any way to the GQ feedback.

Example: “So I don’t know.”

Appendix J

Common Codebook (Both Woodland and Whitcomb)

Review of Feedback: Any time a leader/coleader reports looking at the GQ feedback report.

Example: "My coleader and I looked at our results."

Analysis of Feedback: Any time a leader reports the content of the feedback, studies, scrutinizes, or interprets the feedback.

Example: "We were pleased to see A's positive relative alert."

Planning Interventions: Any time a leader/coleader reports preparing interventions based on the GQ feedback for a future session.

Example: "We planned to provide feedback about these differences using the GQ feedback with the group."

Withholding Feedback: Any time a group leader reports intentionally not using data from the GQ feedback report in group.

Example: "We didn't have enough time to sort of process it, so we didn't bring it on that time."

GQ In-Group (Explicit): Any time a leader/coleader verbally and directly refers to the GQ feedback in a way that makes members aware that the information comes from the feedback.

Example: "As part of the 'go around' during the first 10 minutes of group, we shared brief observations from the data, sometimes suggesting a group member seemed ready to work."

GQ In-Group (Non-Explicit): Any time a leader/coleader reports using the feedback verbally in the group without directly referring to GQ scores or alerts, and members are not aware that the intervention is based on feedback.

Example: "So, occasionally through the semester, I would make an observation to someone about, 'you seem to be less engaged the last couple of weeks.'"

Education: Any time a leader/coleader spends time in group explaining something about the GQ to members.

Example: "We talked with [group members] about how to read the results."

Awareness of Alerters: Any time a leader is more aware of or attentive to group members because of GQ scores they received informing them of a member's alerts.

Example: "We are mindful of the members who are experiencing negative bonds with either members or leaders."

Looking Forward: Any time a leader reports predictions about or anticipation for receiving GQ results in the future.

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

Effects of GQ Use: Any time a leader reports the impact, outcome, or consequence that GQ use had on members and/or leaders.

Example: "If anything I would say it gave me more patience with the silence as they struggled with how to respond because I knew that it was feedback that they had given."

Ambiguous Use: Any time a leader reports using the feedback without specifically explaining how it was used.

Example: "I used just the GQ."

Value/Opinion: Any time a leader expresses their appreciation for, value of, or opinion of the GQ.

Example: "I've enjoyed having the feedback."

Group Initiated: Any time leaders report group members bringing up the GQ feedback in group.

Example: "Also of note, another member whose bond went down and negative relationship went up challenged us to be aware of the challenges he finds in connecting with the group."

Appendix K

Category Frequency Percentages by Main Category

Whitcomb Category Frequency Percentages by Data Source

Categories	Leader Slips	Interviews	Total
Overall GQ Use	0.26	0.37	0.34
Review of Feedback	10.58	0.92	3.41
Sharing Feedback with Others	8.47	0.92	2.86
Analysis of GQ Feedback	46.56	30.12	34.36
Proficiency with GQ	1.32	1.38	1.36
Disruption in Data	1.32	1.93	1.77
GQ in Group	15.61	16.71	16.43
Leader Confusion	0.00	1.10	0.82
Member Reaction	2.12	4.32	3.75
Follow-Up Outside of Group	2.12	1.01	1.30
Expectations for Future GQ Scores	1.32	0.00	0.34
Effect on Group	5.29	27.61	21.88
Opinion of GQ	0.53	6.34	4.84
Hindsight Opinion	0.26	3.21	2.45
No GQ in Group	4.23	4.04	4.09
Total Units	378	1,089	1,467

Note. $\chi^2(14) = 274.0913$, $p < 0.01$

Whitcomb Category Frequency Percentages by Group Type

Categories	Process	Specific Focus	Total
Overall GQ Use	0.39	0.23	0.34
Review of Feedback	2.04	6.59	3.41
Sharing Feedback with Others	2.14	4.55	2.86
Analysis of GQ Feedback	33.01	37.5	34.36
Proficiency with GQ	1.17	1.82	1.36
Disruption in Data	2.34	0.45	1.77
GQ in Group	19.67	8.86	16.43
Leader Confusion	0.78	0.91	0.82
Member Reaction	4.48	2.05	3.75
Follow-Up Outside of Group	0.49	3.18	1.3
Expectations for Future GQ Scores	0.49	0.00	0.34
Effect on Group	24.05	16.82	21.88
Opinion of GQ	4.97	4.55	4.84
Hindsight Opinion	1.75	4.09	2.45
No GQ in Group	2.24	8.41	4.09
Total Units	1,027	440	1,467

Note. χ^2 (14) = 123.2932, $p < 0.01$

Whitcomb Category Frequency Percentages by Leader

Categories	2	4	5	6	7	8	9	12	14	15	16	18	Total
Overall GQ Use	0.4	0.0	0.0	0.0	0.7	1.0	0.6	0.0	0.0	0.0	0.8	0.0	0.3
Review of Feedback	2.5	11.2	4.0	0.0	8.2	0.0	0.6	2.7	10.0	0.5	4.5	0.0	3.4
Sharing Feedback with others	0.8	0.8	1.3	0.0	10.3	0.0	0.0	2.7	0.0	4.0	7.5	1.7	2.9
Analysis of GQ Feedback	42.5	39.2	31.6	37.9	45.9	33.7	33.3	32.4	20.0	30.7	20.2	22.4	34.4
Proficiency with GQ	0.0	0.8	0.0	0.0	0.7	0.0	1.3	4.5	0.0	4.0	1.5	1.7	1.4
Disruption in Data	1.3	1.6	3.3	0.0	0.0	10.6	0.6	0.0	0.0	1.0	1.5	0.0	1.8
GQ in Group	17.9	5.6	24.3	13.8	5.5	14.4	21.2	18.0	0.0	13.4	32.1	6.9	16.4
Leader Confusion	0.0	0.0	0.0	0.0	0.0	7.7	0.0	3.6	0.0	0.0	0.0	0.0	0.8
Member Reaction	4.2	4.0	3.3	0.0	0.7	2.9	7.1	2.7	0.0	6.4	3.0	0.0	3.8
Follow-Up Outside of Group	0.8	8.0	0.0	0.0	2.1	0.0	1.3	0.9	0.0	0.5	0.0	0.0	1.3
Expectations for Future Scores	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
Effect on Group	25.0	14.4	21.7	41.4	15.1	11.5	26.3	19.8	30.0	26.7	23.9	20.7	21.9
Opinion of GQ	1.7	0.8	7.2	0.0	5.5	15.4	5.1	3.6	0.0	3.5	3.7	12.1	4.8
Hindsight Opinion	0.4	0.8	0.0	3.5	0.7	0.0	0.6	5.4	0.0	7.4	0.0	17.2	2.5
No GQ in Group	0.4	12.8	3.3	3.5	4.8	2.9	1.9	3.6	40.0	2.0	1.5	17.2	4.1
Total Units	240	125	152	29	146	104	156	111	10	202	134	58	1,467

Note. $\chi^2(154) = 700.1619$, $p < 0.01$

Whitcomb Category Frequency Percentages by Group

Categories	25	27	29	32	33	34	36	38	40	42	50	52	53	57	60	61	Total
Overall GQ Use	0.0	1.0	0.0	0.7	0.4	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	1.3	0.0	0.3
Review of Feedback	0.0	1.0	0.0	8.2	2.5	11.2	1.6	2.7	0.8	10.0	4.5	5.5	0.0	0.0	0.0	0.0	3.4
Sharing Feedback with Others	0.0	0.0	0.0	10.3	0.8	0.8	3.3	2.7	6.7	0.0	7.5	0.0	1.7	0.0	0.0	0.0	2.9
Analysis of GQ Feedback	34.6	31.1	37.9	45.9	42.5	39.2	36.1	32.4	32.5	20.0	20.2	28.6	22.4	28.1	33.3	37.7	34.4
Proficiency with GQ	0.0	1.9	0.0	0.7	0.0	0.8	0.0	4.5	6.7	0.0	1.5	0.0	1.7	0.0	0.0	0.0	1.4
Disruption in Data	7.7	0.0	0.0	0.0	1.3	1.6	0.0	0.0	1.7	0.0	1.5	5.5	0.0	0.0	11.5	1.9	1.8
GQ in Group	23.1	26.2	13.8	5.5	17.9	5.6	18.0	18.0	10.0	0.0	32.1	28.6	6.9	18.3	11.5	11.3	16.4
Leader Confusion	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	0.0	10.3	0.0	0.8
Member Reaction	11.5	10.7	0.0	0.7	4.2	4.0	6.6	2.7	10.8	0.0	3.0	1.1	0.0	0.0	0.0	0.0	3.8
Follow-Up Outside of Group	0.0	1.9	0.0	2.1	0.8	8.0	0.0	0.9	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3
Expectations for Future Scores	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
Effect on Group	7.7	19.4	41.4	15.1	25.0	14.4	16.4	19.8	20.0	30.0	23.9	25.3	20.7	36.6	12.8	39.6	21.9
Opinion of GQ	11.5	3.9	0.0	5.5	1.7	0.8	9.8	3.6	0.0	0.0	3.7	5.5	12.1	8.5	16.7	7.6	4.8
Hindsight Opinion	0.0	1.0	3.5	0.7	0.4	0.8	0.0	5.4	6.7	0.0	0.0	0.0	17.2	8.5	0.0	0.0	2.5
No GQ in Group	3.9	1.9	3.5	4.8	0.4	12.8	8.2	3.6	3.3	40.0	1.5	0.0	17.2	0.0	2.6	1.9	4.1
Total Units	26	103	29	146	240	125	61	111	120	10	134	91	58	82	78	53	1,467

Note. $\chi^2(210) = 846.4060$, $p < 0.01$

Combined Category Frequency Percentages by Data Source

Categories	Leader Slips	Debrief Int.	Total
Review of Feedback	10.27	2.44	5.03
Analysis of Feedback	27.43	26.32	26.69
Planning Interventions	5.77	2.51	3.59
Withholding Feedback	13.50	6.96	9.13
In-Group—Explicit	14.49	10.93	12.11
In-Group—Non-Explicit	2.95	3.76	3.49
Education	1.83	0.70	1.07
Awareness of Alerters	0.84	2.99	2.28
Looking Forward	1.69	0.21	0.70
Effects of GQ Use	15.33	26.74	22.96
Ambiguous Use	0.14	2.44	1.68
Value/Opinion	0.42	6.62	4.56
Group-Initiated	1.27	0.56	0.79
Not Captured	4.08	6.82	5.92
Total Units	711	1,436	2,147

Note. $\chi^2(13) = 221.5554$, $p < 0.01$

Combined Category Frequency Percentages by Group Type

Categories	Process	Specific Focus	Total
Review of Feedback	3.62	10.13	5.03
Analysis of Feedback	27.04	25.43	26.69
Planning Interventions	3.80	2.80	3.59
Withholding Feedback	6.48	18.75	9.13
In-Group—Explicit	13.73	6.25	12.11
In-Group—Non-Explicit	4.40	0.22	3.49
Education	1.19	0.65	1.07
Awareness of Alerters	1.49	5.17	2.28
Looking Forward	0.71	0.65	0.70
Effects of GQ Use	25.37	14.22	22.96
Ambiguous Use	1.84	1.08	1.68
Value/Opinion	4.10	6.25	4.56
Group-Initiated	0.59	1.51	0.79
Not Captured	5.64	6.90	5.92
Total Units	1,683	464	2,147

Note. $\chi^2(13) = 179.0920$, $p < 0.01$

Combined Category Frequency Percentages by Data Set

Categories	Woodland	Whitcomb	Total
Review of Feedback	6.30	4.43	5.02
Analysis of Feedback	31.19	24.54	26.65
Planning Interventions	4.54	3.14	3.58
Withholding Feedback	9.08	9.13	9.12
In-Group—Explicit	9.96	13.09	12.09
In-Group—Non-Explicit	7.32	1.70	3.49
Education	2.20	0.55	1.07
Awareness of Alerters	1.61	2.59	2.28
Looking Forward	0.44	1.02	0.84
Effects of GQ Use	19.77	24.40	22.93
Ambiguous Use	3.07	1.02	1.67
Value/Opinion	0.00	6.68	4.56
Group-Initiated	1.02	0.68	0.79
Not Captured	3.51	7.02	5.91
Total Units	683	1,467	2,150

Note. $\chi^2(13) = 147.2835$, $p < 0.01$

Combined Category Frequency Percentages by Leader

Categories	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	18	Total
Review of Feedback	0.9	3.7	5.5	12.8	5.0	0.0	18.5	0.7	1.0	4.6	7.5	2.7	10.0	0.5	5.2	1.7	5.0
Analysis of Feedback	24.6	39.7	33.6	22.8	20.9	20.9	29.8	24.3	25.1	50.0	37.5	25.2	10.0	20.3	14.2	19.0	26.7
Planning Interventions	1.7	2.9	1.8	3.4	3.2	6.0	5.9	2.0	6.0	0.0	5.0	5.4	0.0	5.0	1.5	0.0	3.6
Withholding Feedback	4.2	5.7	6.4	30.2	10.0	16.4	16.1	3.3	2.5	22.7	10.0	5.4	40.0	5.0	3.0	17.2	9.1
In-Group—Explicit	12.7	7.7	16.4	6.7	19.1	4.5	2.0	14.5	17.1	4.6	2.5	13.5	0.0	15.4	27.6	0.0	12.1
In-Group—Non-Explicit	4.2	2.9	0.9	0.0	5.5	20.9	0.0	4.0	6.0	9.1	10.0	0.9	0.0	0.5	5.2	0.0	3.5
Education	3.4	1.7	0.9	0.7	0.9	0.0	0.0	3.3	0.0	0.0	2.5	1.8	0.0	0.5	0.0	0.0	1.1
Awareness of Alerters	0.0	1.7	0.9	2.7	2.7	1.5	6.8	1.3	0.0	0.0	7.5	0.0	0.0	1.0	1.5	13.8	2.3
Looking Forward	0.0	2.6	0.0	0.0	0.5	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.7
Effects of GQ Use	44.9	25.4	29.1	12.1	22.3	17.9	5.9	9.2	32.2	4.6	2.5	26.1	30.0	31.2	32.1	17.2	23.0
Ambiguous Use	0.9	1.4	1.8	0.7	0.5	10.5	2.0	2.6	1.0	0.0	10.0	0.0	10.0	0.5	1.5	1.7	1.7
Value/Opinion	0.0	2.0	0.0	3.4	6.4	1.5	5.9	13.2	5.5	0.0	0.0	3.6	0.0	5.9	3.0	13.8	4.6
Group-Initiated	0.9	0.0	1.8	2.0	0.9	0.0	2.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	2.2	0.0	0.8
Not Captured	1.7	2.6	0.9	2.7	2.3	0.0	3.9	21.7	3.5	4.6	5.0	13.5	0.0	13.4	3.0	15.5	5.9
Total Units	118	350	110	149	220	67	205	152	199	22	40	111	10	202	134	58	2,147

Note. $\chi^2(195) = 1.0e+03$, $p < 0.01$

Combined Category Frequency Percentages by Group

Categories	2	3	5	8	10	11	13	15	17	19	21	24	25	27	29
Review of Feedback	0.9	2.6	0.0	8.8	20.8	8.8	0.0	23.7	2.1	2.3	4.6	7.5	0.0	1.0	0.0
Analysis of Feedback	24.6	44.7	35.7	32.4	33.3	22.1	26.3	27.1	22.9	27.9	50.0	37.5	26.9	20.4	13.8
Planning Interventions	1.7	5.3	2.4	1.5	8.3	2.9	7.9	17.0	2.1	9.3	0.0	5.0	7.7	6.8	3.5
Withholding Feedback	4.2	23.7	4.8	7.4	37.5	5.9	13.2	11.9	2.1	0.0	22.7	10.0	3.9	2.9	20.7
In-Group—Explicit	12.7	7.9	16.7	16.2	0.0	22.1	0.0	0.0	14.6	9.3	4.6	2.5	23.1	25.2	10.3
In-Group—Non-Exp.	4.2	0.0	0.0	1.5	0.0	10.3	34.2	0.0	12.5	18.6	9.1	10.0	0.0	2.9	3.5
Education	3.4	0.0	2.4	0.0	0.0	2.9	0.0	0.0	10.4	0.0	0.0	2.5	0.0	0.0	0.0
Awareness of Alerters	0.0	0.0	0.0	1.5	0.0	1.5	0.0	3.4	4.2	0.0	0.0	7.5	0.0	0.0	3.5
Looking Forward	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Effects of GQ Use	44.9	7.9	31.0	27.9	0.0	19.1	2.6	5.1	6.3	30.2	4.6	2.5	19.2	31.1	37.9
Ambiguous Use	0.9	0.0	0.0	2.9	0.0	1.5	15.8	1.7	4.2	2.3	0.0	10.0	0.0	1.0	3.5
Value/Opinion	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.5	3.9	3.5
Group-Initiated	0.9	0.0	4.8	0.0	0.0	2.9	0.0	3.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Not Captured	1.7	7.9	2.4	0.0	0.0	0.0	0.0	6.8	18.8	0.0	4.6	5.0	7.7	4.9	0.0
Total	118	38	42	68	24	68	38	59	48	43	22	40	26	103	29
Groups cont.	32	33	34	36	38	40	42	44	50	52	53	57	60	61	Total
Review of Feedback	16.4	3.3	11.2	0.0	2.7	0.8	10.0	5.6	5.2	5.5	1.7	0.0	0.0	0.0	5.0
Analysis of Feedback	30.8	37.5	20.8	26.2	25.2	22.5	10.0	44.4	14.2	16.5	19.0	17.1	24.4	32.1	26.7
Planning Interventions	1.4	2.9	2.4	6.6	5.4	3.3	0.0	1.4	1.5	1.1	0.0	7.3	0.0	1.9	3.6
Withholding Feedback	17.8	2.1	28.8	13.1	5.4	8.3	40.0	8.3	3.0	11.0	17.2	0.0	3.9	3.8	9.1
In-Group—Explicit	2.7	8.3	8.0	21.3	13.5	14.2	0.0	5.6	27.6	15.4	0.0	17.1	11.5	7.6	12.1
In-Group—Non-Exp.	0.0	2.5	0.0	0.0	0.9	0.0	0.0	5.6	5.2	5.5	0.0	1.2	0.0	1.9	3.5
Education	0.0	1.7	0.8	0.0	1.8	0.8	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	1.1
Awareness of Alerters	8.2	1.7	3.2	0.0	0.0	0.8	0.0	2.8	1.5	5.5	13.8	1.2	0.0	0.0	2.3
Looking Forward	2.1	3.8	0.0	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.7
Effects of GQ Use	6.2	30.8	14.4	18.0	26.1	31.7	30.0	16.7	32.1	27.5	17.2	30.5	7.7	35.9	23.0
Ambiguous Use	2.1	0.8	0.8	0.0	0.0	0.0	10.0	4.2	1.5	0.0	1.7	1.2	2.6	0.0	1.7
Value/Opinion	8.2	2.9	4.0	13.1	3.6	2.5	0.0	0.0	3.0	6.6	13.8	11.0	21.8	13.2	4.6
Group-Initiated	1.4	0.0	2.4	0.0	1.8	0.0	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.8
Not Captured	2.7	1.7	3.2	0.0	13.5	15.0	0.0	2.8	3.0	5.5	15.5	11.0	28.2	3.8	5.9
Total	146	240	125	61	111	120	10	72	134	91	58	82	78	53	2147

Note. $X^2(364) = 1.5e+03$, $p < 0.01$

Appendix L

Multinomial Logistic Regression Findings

From Whitcomb Data—Three-Dimension Model

Leader and Data Source
Number of Observations—1,462

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
Leader ID			
4	0.03	0.26	0.91
5	0.54	0.25	0.04
6	-0.05	0.57	0.93
7	-0.79	0.30	0.01
8	0.10	0.29	0.74
9	0.42	0.26	0.11
12	0.38	0.28	0.17
14	0.91	0.79	0.24
15	0.25	0.25	0.31
16	1.11	0.27	0.00
18	0.43	0.41	0.30
Data Source			
Debrief Int.	0.48	0.15	0.00
Intercept	-0.98	0.19	0.00
Post-Group/Effect Compared with Pre-Group			
Leader ID			
4	-0.92	0.30	0.00
5	0.26	0.27	0.34
6	0.77	0.48	0.11
7	-0.16	0.28	0.57
8	-0.43	0.29	0.15
9	-0.06	0.26	0.82
12	-0.11	0.29	0.71
14	0.25	0.86	0.77
15	0.14	0.24	0.54
16	0.40	0.29	0.17
18	0.57	0.36	0.11
Data Source			
Debrief Int.	2.21	0.22	0.00
Intercept	-2.13	0.25	0.00

Group and Data Source
Number of Observations—1,462

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
Group			
27	0.09	0.50	0.86
29	-0.69	0.70	0.32
32	-1.41	0.51	0.01
33	-0.65	0.47	0.17
34	-0.63	0.49	0.20
36	0.13	0.53	0.80
38	-0.27	0.50	0.58
40	-0.32	0.49	0.52
42	0.26	0.89	0.77
50	0.47	0.49	0.34
52	-0.28	0.51	0.59
53	-0.25	0.58	0.67
57	-0.60	0.56	0.28
60	-0.79	0.53	0.14
61	-1.25	0.62	0.05
Data Source			
Debrief Int.	0.55	0.16	0.00
Intercept	-0.38	0.45	0.40
Post-Group/Effect Compared with Pre-Group			
Group			
27	-0.06	0.63	0.93
29	0.97	0.73	0.18
32	0.03	0.61	0.96
33	0.19	0.59	0.75
34	-0.73	0.63	0.24
36	0.71	0.67	0.29
38	0.08	0.62	0.90
40	-0.05	0.62	0.93
42	0.44	1.02	0.67
50	0.60	0.62	0.33
52	0.29	0.63	0.64
53	0.76	0.65	0.25
57	0.75	0.63	0.24
60	-0.32	0.63	0.61
61	0.34	0.65	0.60
Data Source			
Debrief Int.	2.22	0.22	0.00
Intercept	-2.31	0.60	0.00

Group Type and Data Source
Number of Observations—1,462

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
Group Type			
Specific Focus	-0.43	0.14	0.00
Data Source			
Debrief Int.	0.59	0.14	0.00
Intercept	-0.69	0.13	0.00
Post-Group/Effect Compared with Pre-Group			
Group Type			
Specific Focus	-0.35	0.15	0.02
Data Source			
Debrief Int.	2.17	0.21	0.00
Intercept	-2.00	0.20	0.00

Total Number of Negative Alerts per Session.
Number of Observations—366

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
Neg. Sessions	0.05	0.04	0.29
Intercept	-0.98	0.16	0.00
Post-Group/Effect Compared with Pre-Group			
Neg. Sessions	0.05	0.07	0.48
Intercept	-2.26	0.27	0.00

Total Number of People Alerting per Session.
Number of Observations—366

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
# of Alerters	-0.45	0.08	0.00
Intercept	1.25	0.40	0.00
Post-Group/Effect Compared with Pre-Group			
# of Alerters	0.24	0.16	0.13
Intercept	-3.45	0.92	0.00

Proportion of People Alerting per Session.
Number of Observations—366

Variable	Coefficient	Standard Error	*p* Value
In-Group Compared with Pre-Group			
Proportion of Alerters	-3.63	0.59	0.00
Intercept	1.62	0.41	0.00
Post-Group/Effect Compared with Pre-Group			
Proportion of Alerters	0.77	1.10	0.49
Intercept	-2.74	0.90	0.00

Appendix M

Multinomial Logistic Regression Findings

From Whitcomb Data—Four-Dimension Model

Leader and Data Source
Number of Observations—1,241

Variable	Coefficient	Standard Error	*p* Value
Pre-Group Planning Compared with Pre-Group Review			
Leader ID			
4	2.16	0.38	0.00
5	1.64	0.41	0.00
6	2.65	0.70	0.00
7	1.11	0.38	0.00
8	0.75	0.55	0.17
9	1.14	0.45	0.01
12	1.23	0.46	0.01
14	2.89	0.92	0.00
15	1.47	0.41	0.00
16	0.66	0.55	0.23
18	2.12	0.54	0.00
Data Source			
Debrief Int.	-0.39	0.21	0.06
Intercept	-1.92	0.32	0.00
In-Group Compared with Pre-Group			
Leader ID			
4	0.05	0.37	0.89
5	1.15	0.31	0.00
6	1.35	0.70	0.06
7	-0.24	0.35	0.48
8	0.45	0.39	0.24
9	0.85	0.32	0.01
12	0.50	0.36	0.17
14	-11.52	375.23	0.98
15	0.83	0.31	0.01
16	1.67	0.32	0.00
18	0.53	0.51	0.30
Data Source			
Debrief Int.	0.42	0.19	0.02
Intercept	-1.36	0.23	0.00
Post-Group/Effect Compared with Pre-Group			
Leader ID			
4	-0.91	0.33	0.01

5	0.20	0.29	0.50
6	1.18	0.62	0.06
7	-1.25	0.36	0.00
8	-1.05	0.40	0.01
9	0.04	0.27	0.88
12	-0.15	0.31	0.63
14	0.27	0.94	0.77
15	0.27	0.26	0.30
16	0.57	0.30	0.06
18	-0.54	0.46	0.24
Data Source			
Debrief Int.	1.66	0.21	0.00
Intercept	-1.30	0.22	0.00

Group and Data Source
Number of Observations—1,241

Variable	Coefficient	Standard Error	*p* Value
Pre-Group Planning Compared with Pre-Group Review			
Group			
27	0.18	0.79	0.82
29	1.41	0.93	0.13
32	-0.11	0.73	0.88
33	-1.24	0.76	0.10
34	0.91	0.73	0.21
36	0.51	0.79	0.52
38	-0.02	0.77	0.98
40	0.22	0.77	0.77
42	1.64	1.11	0.14
50	-0.59	0.83	0.48
52	0.30	0.79	0.70
53	0.86	0.82	0.30
57	0.19	0.86	0.82
60	-0.81	0.94	0.39
61	-0.72	0.94	0.44
Data Source			
Debrief Int.	-0.33	0.21	0.11
Intercept	-0.70	0.70	0.31
In-Group Compared with Pre-Group Review			
Group			
27	0.29	0.63	0.64
29	0.37	0.87	0.67
32	-1.22	0.62	0.05
33	-0.98	0.59	0.10
34	-0.93	0.64	0.14
36	0.02	0.67	0.98
38	-0.49	0.63	0.44
40	-0.32	0.63	0.62
42	-12.51	375.42	0.97
50	0.68	0.61	0.26
52	0.27	0.64	0.67
53	-0.46	0.73	0.53
57	0.08	0.67	0.91
60	-0.80	0.70	0.25
61	-1.25	0.76	0.10
Data Source			
Debrief Int.	0.44	0.19	0.02
Intercept	-0.38	0.57	0.50
Post-Group/Effect Compared with Pre-Group Review			

Group			
27	0.31	0.67	0.64
29	1.32	0.86	0.13
32	-1.11	0.69	0.11
33	0.13	0.63	0.83
34	-0.78	0.67	0.25
36	0.29	0.73	0.69
38	-0.02	0.66	0.98
40	0.39	0.66	0.56
42	0.41	1.11	0.72
50	0.71	0.66	0.29
52	0.36	0.68	0.60
53	-0.42	0.75	0.58
57	0.42	0.70	0.54
60	-1.39	0.77	0.07
61	-0.06	0.70	0.93
Data Source			
Debrief Int.	1.69	0.21	0.00
Intercept	-1.45	0.63	0.02

Group Type and Data Source
Number of Observations—1,241

Variable	Coefficient	Standard Error	*p* Value
Pre-Group Planning Compared with Pre-Group Review			
Group Type			
Specific Focus	0.55	0.18	0.00
Data Source			
Debrief Int.	-0.16	0.18	0.37
Intercept	-1.00	0.16	0.00
In-Group Compared with Pre-Group Review			
Group Type			
Specific Focus	-0.66	0.18	0.00
Data Source			
Debrief Int.	0.52	0.17	0.00
Intercept	-0.64	0.15	0.00
Post-Group/Effect Compared with Pre-Group Review			
Group Type			
Specific Focus	-0.84	0.17	0.00
Data Source			
Debrief Int.	1.70	0.20	0.00
Intercept	-1.22	0.18	0.00

Appendix N

Multinomial Logistic Regression Findings

From Combined Data—Four-Dimension Model

Leader, Data Source, Group Type, and Data Set
Number of Observations—1,869

Variable	Coefficient	Standard Error	*p* Value
Pre-Group Planning Compared with Pre-Group Review			
Leader			
2	-0.01	0.49	0.98
3	-0.09	0.56	0.87
4	2.10	0.61	0.00
5	0.96	0.50	0.05
6	1.90	0.58	0.00
7	0.89	0.52	0.09
8	0.23	0.59	0.70
9	0.80	0.54	0.14
10	0.96	0.69	0.17
11	0.81	0.64	0.21
12	1.18	0.67	0.08
14	2.56	1.00	0.01
15	1.16	0.55	0.04
16	0.28	0.66	0.68
18	2.19	0.72	0.00
Data Source			
Debrief Int.	-0.79	0.17	0.00
Group Type			
Specific Focus	-0.31	0.44	0.49
Data Set			
Whitcomb	-0.02	0.22	0.93
Intercept	-1.26	0.42	0.00
In-Group Compared with Pre-Group Review			
Leader			
2	-0.85	0.34	0.01
3	-0.49	0.38	0.20
4	-2.29	0.88	0.01
5	0.37	0.35	0.29
6	0.53	0.46	0.25
7	-2.48	0.78	0.00
8	0.21	0.38	0.59
9	0.18	0.37	0.62
10	-1.15	0.70	0.10
11	-0.45	0.50	0.36

12	-1.55	0.88	0.08
14	-12.80	501.43	0.98
15	0.15	0.41	0.72
16	0.90	0.41	0.03
18	-1.41	0.95	0.14
Data Source			
Debrief Int.	-0.03	0.15	0.83
Group Type			
Specific Focus	1.34	0.80	0.10
Data Set			
Whitcomb	-0.09	0.18	0.62
Intercept	-0.21	0.28	0.44
Post-Group/Effect Compared with Pre-Group Review			
Leader			
2	-1.77	0.31	0.00
3	-0.88	0.33	0.01
4	-2.37	0.78	0.00
5	-1.47	0.34	0.00
6	-1.41	0.48	0.00
7	-2.87	0.65	0.00
8	-2.42	0.42	0.00
9	-1.36	0.35	0.00
10	-3.34	1.07	0.00
11	-3.79	1.06	0.00
12	-1.46	0.77	0.06
14	-1.30	0.97	0.18
15	-1.28	0.37	0.00
16	-1.12	0.40	0.01
18	-1.69	0.84	0.05
Data Source			
Debrief Int.	0.66	0.15	0.00
Group Type			
Specific Focus	-0.28	0.71	0.69
Data Set			
Whitcomb	0.83	0.20	0.00
Intercept	0.31	0.24	0.20

Group and Data Source
Number of Observations—1,869

Variable	Coefficient	Standard Error	*p* Value
Pre-Group Planning Compared with Pre-Group Review			
Group			
3	0.78	0.57	0.17
5	-0.33	0.76	0.66
8	0.09	0.62	0.88
10	1.69	0.60	0.01
11	0.32	0.63	0.61
13	1.68	0.65	0.01
15	0.89	0.52	0.09
17	-0.17	0.87	0.85
19	0.65	0.72	0.36
21	0.93	0.69	0.18
24	0.78	0.64	0.23
25	0.77	0.81	0.34
27	1.08	0.58	0.06
29	2.22	0.76	0.00
32	0.57	0.48	0.24
33	-0.45	0.52	0.39
34	1.77	0.49	0.00
36	1.23	0.57	0.03
38	0.83	0.55	0.13
40	1.07	0.54	0.05
42	2.51	0.97	0.01
44	-0.01	0.59	0.99
50	0.24	0.62	0.70
52	1.14	0.57	0.05
53	1.82	0.62	0.00
57	1.16	0.66	0.08
60	0.16	0.76	0.84
61	0.21	0.77	0.78
Data Source			
Debrief Int.	-0.73	0.18	0.00
Intercept	-1.28	0.42	0.00
In-Group Compared with Pre-Group Review			
Group			
3	-1.58	0.68	0.02
5	-0.41	0.52	0.43
8	-0.54	0.44	0.22
10	-16.59	1248.36	0.99
11	0.40	0.40	0.32
13	0.50	0.51	0.33
15	-2.48	0.78	0.00

17	0.74	0.46	0.11
19	0.16	0.49	0.75
21	-1.15	0.71	0.10
24	-0.46	0.50	0.36
25	0.08	0.62	0.90
27	0.51	0.40	0.20
29	0.45	0.73	0.53
32	-1.24	0.39	0.00
33	-0.83	0.34	0.02
34	-0.75	0.41	0.07
36	0.02	0.46	0.97
38	-0.31	0.41	0.45
40	-0.15	0.41	0.71
42	-15.52	1866.21	0.99
44	-0.87	0.43	0.05
50	0.80	0.37	0.03
52	0.42	0.41	0.31
53	-0.16	0.54	0.76
57	0.37	0.47	0.43
60	-0.51	0.50	0.31
61	-0.98	0.59	0.09
Data Source			
Debrief Int.	-0.03	0.15	0.87
Intercept	-0.22	0.28	0.44
Post-Group/Effect Compared with Pre-Group Review			
Group			
3	-2.07	0.67	0.00
5	-0.42	0.45	0.35
8	-1.15	0.38	0.00
10	-17.47	1088.03	0.99
11	-1.18	0.43	0.01
13	-3.25	1.08	0.00
15	-2.88	0.65	0.00
17	-2.14	0.69	0.00
19	-0.91	0.46	0.05
21	-3.38	1.07	0.00
24	-3.83	1.06	0.00
25	-1.08	0.64	0.09
27	-0.56	0.37	0.13
29	0.23	0.63	0.72
32	-2.33	0.39	0.00
33	-0.94	0.28	0.00
34	-1.68	0.37	0.00
36	-0.93	0.45	0.04
38	-0.94	0.35	0.01
40	-0.56	0.35	0.11

42	-0.50	0.95	0.60
44	-1.85	0.41	0.00
50	-0.32	0.34	0.36
52	-0.62	0.39	0.11
53	-1.19	0.50	0.02
57	-0.35	0.41	0.40
60	-2.16	0.53	0.00
61	-0.87	0.42	0.04
Data Source			
Debrief Int.	0.73	0.15	0.00
Intercept	0.28	0.24	0.24