2012-04-26

The Validity of the Group Questionnaire: Construct Clarity or Construct Drift?

Stephen D. Thayer
Brigham Young University - Provo

The Validity of the Group Questionnaire:

Construct Clarity or Construct Drift?

Stephen D. Thayer

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
Scott A. Baldwin
Patrick R. Steffen
Robert L. Gleave
Dawson W. Hedges

Department of Psychology

Brigham Young University

August 2012

ABSTRACT

The Validity of the Group Questionnaire:
Construct Clarity or Construct Drift?

Stephen D. Thayer
Department of Psychology, BYU
Doctor of Philosophy

The Group Questionnaire (GQ) is a recently developed measure of the quality of the therapeutic relationship in group treatment. Its 3 subscales—Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship—are taken from a 3-factor conceptualization of the group therapeutic relationship (Johnson et al., 2005). The purpose of the present study was to estimate the GQ's construct and criterion-related validity by 1) replicating the aforementioned factor structure with a similar sample and by 2) correlating the GQ with the measures from which is was derived (i.e., Working Alliance Inventory, Burns Empathy Scale, Therapeutic Factors Inventory, Group Climate Questionnaire) and to 3) explore the GQ's ability to measure relationship quality at member-member, member-leader, and member-group structural relationship levels using a sociometric test. Two hundred and ninety participants were recruited from 65 treatment groups at 4 university counseling centers and 1 community mental health clinic. Confirmatory factor analysis (CFA) fit indexes from both single- and multiple-level analyses met standards for acceptable model fit. Intraclass correlation coefficients (ICC) suggested the GQ is sensitive to group level processes. Therefore, the Johnson et al. (2005) 3-factor model was successfully replicated and the GQ's construct validity supported. Pearson product-moment (r) and Spearman's rank (ρ) correlation coefficients were sufficiently high to lend support for the GQ's criterion-related validity. Sociometric exploration yielded moderate support for the GQ's ability to access the structural parameters of group therapeutic relationships. The present study's findings suggest the GQ is an empirically valid, clinically useful measure of the quality of the group therapeutic relationship.

Keywords: group, therapeutic, relationship, validity, questionnaire.

ACKNOWLEDGEMENTS

Special thanks to Gary Burlingame, whose insightful counsel, tireless support, and ceaseless guidance has proffered me so much more than a completed dissertation. Warm appreciation to those who helped me with the nuts and bolts; recruiting participants, crunching numbers, and filling spreadsheets (I'm looking at you, Nick Patterson). Humble gratitude to my supportive, loving wife and my patient, awesome kids. You made this all possible.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

The Validity of the Group Questionnaire: Construct Clarity or Construct Drift?

In the face of increasing demands for inexpensive and accessible forms of mental health treatment, group psychotherapy has emerged as a cost-effective intervention that is comparable to individual psychotherapy (G. M. Burlingame, Fuhriman, & Mosier, 2003; G. M. Burlingame, MacKenzie, & Strauss, 2004; Fuhriman & Burlingame, 1999; K. Roy MacKenzie, 1995; McRoberts, Burlingame, & Hoag, 1998). But what makes group psychotherapy effective? In his seminal text on group psychotherapy (Yalom & Leszcz, 2005) Irvin Yalom declared:

> How does group therapy help clients? A naïve question, perhaps. But if we can answer it with some measure of precision and certainty, we will have at our disposal a central organizing principle with which to approach the most vexing and controversial problems of psychotherapy. (p. 1)

It is tempting to answer Yalom's "naïve question" with an equally naïve answer: Group therapy helps clients by effecting cognitive, behavioral, and emotional *changes.* This answer begs the question: What are the mechanisms by which group therapy facilitates such changes? Theory-specific treatment components that are common to individual therapy (e.g., cognitive restructuring, exposure, and unconditional positive regard) account for some of the change mechanisms active in group therapy, but a recent review of the last decade of group research noted that a large portion of patient improvement is not explained by theory-specific components (Burlingame et al., in press). There are therapeutic features unique to group therapy that transcend specific change theories (Fuhriman & Burlingame, 1990). One such feature, the group therapeutic relationship, has been identified as one of the most significant general mechanisms of change in group psychotherapy (Johnson, Burlingame, Olsen, Davies, & Gleave, 2005). Group relationships that both support and challenge group members have been consistently linked to

positive treatment outcome and low dropout rates (Burlingame, Fuhriman, & Johnson, 2002, 2004; Burlingame, McClendon, & Alonso, 2011; Castonguay, Pincus, Agras, & Hines, 1998; Marziali, Munroe-Blum, & McCleary, 1997, 1999).

While researchers like those cited above have succeeded in drawing an empirical connection between group therapeutic relationships and treatment outcome, the strength of their results is diminished by the nebulous nature of the construct they study. The term "therapeutic relationship" can mean several different things in the group therapy literature. For example, when Marziali et al. (1997) examined the link between the group therapeutic relationship and outcome in groups for borderline personality disorder, they focused on *therapeutic alliance* and *cohesion.* But when Beech and Hamilton-Giachritsis (2005) examined the same link in group-based sexual offender treatment programs, they were focused on the *therapeutic climate.* Without a comprehensive, standardized understanding of the group therapeutic relationship group researchers have had difficulty measuring this construct and demonstrating a clear relationship between it and other process/outcome variables. This confusion and disagreement about how to define the group therapeutic relationship has distilled in to two disparate definitions in the group therapy literature.

One of these definitions focuses on the interpersonal structure of relationships within groups. More precisely, relationships occur between members and leaders (member-leader), between members and other members (member-member), and between members and the group-as-a-whole (member-group). A review of the literature on the group therapeutic relationship found that two-thirds of relevant studies focused on the member-group relationship, with far fewer studies focusing on the other two structural parameters (G. M. Burlingame, Fuhriman, &

Johnson, 2002). This imbalance has complicated efforts to determine if these three structural parameters are theoretically and empirically distinct.

The other definition of the group therapeutic relationship focuses on the content of relationship constructs or the *quality* of the therapeutic relationship. Burlingame et al. (2002) identified four core group therapeutic relationship constructs that have been linked to group process and patient improvement: cohesion, alliance, climate, and empathy. Cohesion represents the esprit de corps or sense of belonging within a group and is often considered to be the essence of the group therapeutic relationship. Alliance within a group represents a fond working relationship between group members and facilitators. Group climate refers to the perceived tone of the group interaction and addresses experiences of safety, closeness, withdrawal, and conflict. Empathy among group members/facilitators represents a sense of being understood. In their seminal study, Johnson et al. (2005) outlined the variability with which these therapeutic relationship constructs are operationalized and measured and explained that without understanding the similarities and differences among these therapeutic relationship constructs, it is difficult to draw sound conclusions about the nature and function of the therapeutic relationship from the group literature.

In an effort to address this difficulty, Johnson et al. (2005) consolidated the aforementioned disparate views and provided an empirically refined, parsimonious representation of the therapeutic relationship in group treatment. This study, to date, contains the largest number of groups and group members in the published process literature—over 100 groups and nearly 700 group members participated nationally. Members from university counseling center groups and American Group Psychotherapy Association (AGPA) training groups responded to select items from four gold standard group process measures: Working

Alliance Inventory (WAI; Horvath & Greenberg, 1989); Group Climate Questionnaire (GCQ; MacKenzie, 1983); Therapeutic Factors Inventory (TFI; Lese & MacNair-Semands, 2000); and Empathy Scale (ES; Burns & Auerbach, 1996). From these data, Johnson et al. (2005) created three higher order relationship factors using exploratory factor analysis: Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship. Johnson and her colleagues empirically demonstrated that these higher order factors accounted for the overlap between the various group therapeutic relationship quality constructs (i.e., cohesion, alliance, empathy, and climate) and adequately described how these constructs are manifest across the three structural relationship units (i.e., member-member, member-group, and member-leader). In 2006 this consolidated definition of the group therapeutic relationship was incorporated by an international task force sponsored by AGPA to organize relationship measures to be recommended to group clinicians (G. Burlingame et al., 2006).

In 2007, Bormann and Strauss found that Johnson's revised definition of the group therapeutic relationship (hereafter referred to as the "Johnson Model") provided the same construct clarity when empirically tested on 453 patients in 67 Swiss and German inpatient therapy groups. Johnson's results were again successfully replicated again with 424 patients in Norwegian short- and long-term analytic groups (Bakali, Baldwin, & Lorentzen, 2009). Finally, Krogel (2009) took the preceding research a step further. Using a subset of the questionnaire items Johnson et al. (2005) used in their study, Krogel developed an empirically refined, short, and conceptually "tight" measure of the group therapeutic relationship based on the Johnson Model. Dubbed the Group Questionnaire (GQ), this tool was designed to measure the quality of the therapeutic relationship in groups (i.e., Johnson's Positive Bonding, Positive Working, and Negative Relationship factors) across the three structural parameters of the group therapeutic

relationship. The data collected with the GQ were shown to reproduce Johnson's 3-factor structure yet again, only this time with half the original items. The GQ has since been validated using several other group process measures with a population of 424 group members from nine psychiatric hospitals in Germany (Bormann & Strauss, 2009).

Now that 7 years of research has culminated in a refined definition of the group therapeutic relationship construct and the development of what appears to be a practical method for measuring that construct, it is time to reexamine what has been developed from a psychometric perspective. The multiple replications of the Johnson Model among a clinically diverse international population clarified the definition of the therapeutic relationship in group treatment. This clarity allowed for the creation of the GQ. Having been validated in a German population with German criterion measures (Bormann et al., 2011), validation in a US population using the original criterion measures is prudent to provide the final step in establishing the GQ's criterion-related validity. One goal of the present study was to meet this need. Based on the aforementioned research and observations, I hypothesized the following:

1. The Johnson et al. (2005) 3-factor structure will be successfully replicated in a US population using the GQ, establishing the GQ's construct validity.
2. The Group Questionnaire will be found to have significant correlations with the measures that it was derived from, establishing its criterion-related validity.

While addressing these hypotheses served as a capstone to the line of research that culminated in the creation of the GQ, a careful reading of Johnson et al. (2005) reveals that there may be limitations to the GQ's ability to tap the structural parameters of group relationships. In Johnson's study, group members did not tend to distinguish between the structural parameters of group relationships. Instead, group members tended to distinguish their relationships within

groups based on the quality of their relationships (i.e., bonding, working, and negative relationships). Johnson stated, "Given that the strongest distinctions were made in terms of relationship quality, results were mixed as to the degree to which role relationships added to the descriptive power afforded by quality variables" (2005, p. 319). Conversely, Krogel (2009) and Bakali et al. (2008) found support for retaining relationship structure as part of the Johnson Model. Due to the inconsistency of the findings within the GQ's research pedigree it is unclear whether including relationship structure in the Johnson Model adds to or departs from the construct clarity offered by the three quality variables. This issue is particularly consequential for the validity of the GQ, a questionnaire that was based on the Johnson Model and, therefore, designed with member-member, member-leader, and member-group relationships specifically in mind. Any drift from the clarifying parsimony provided by the Positive Bonding, Positive Working, and Negative Relationship factors would make the GQ a less valid measure of this newly defined group therapeutic relationship construct and, thereby, less useful to group clinicians than originally thought (Krogel, 2009).

Therefore, a secondary goal of the present study was to explore the extent to which the GQ accesses the structural parameters assumed to occur within group relationships. This exploration was conducted using a sociometric test. Pioneered by Jacob Moreno (1953), sociometry is a method of investigating networks of existing and preferred relationships within groups (Treadwell, Kumar, Stein, & Prosnick, 1998). Sociometry has been used in a variety of research and applied settings including developmental psychology, industrial psychology, sport psychology, education, government, public health, and group psychotherapy (e.g., Breen, 1994; Kumar & Treadwell, 2006; Remer, 1995). A sociometric exploration of a group reveals the interpersonal connections between group members and the reasons for those connections by

graphically depicting the impressions and preferences that group members have about one another (Hale, 1985). In so doing, sociometry reveals the "psychological geography" of a group. This makes sociometry an appropriate criterion to compare to the GQ.

A sociometric questionnaire (SQ) with three questions was designed for this study. Each question was devised to survey within-group interpersonal preferences thought to be related to one of the three GQ relationship quality subscales. SQ item 1 referenced Positive Working Relationship, item 2 referenced Positive Bonding Relationship, and item 3 referenced Negative Relationship. I began my exploration of the GQ's validity with the following hypotheses:

3. GQ Positive Working Relationship subscale score valence will correspond with the SQ item 1 score valence at the member-member structural relationship level.
4. GQ Positive Working Relationship subscale score valence will correspond with the SQ item 1 score valence at the member-leader structural relationship level.
5. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-member structural relationship level.
6. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-leader structural relationship level.
7. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-group structural relationship level.
8. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-member structural relationship level.
9. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-leader structural relationship level.

10. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-group structural relationship level.

Please see Table 1-1 for a matrix that depicts how each of these eight hypotheses addresses each GQ relationship quality subscale at each structural level.

Table 1-1

Hypothesis Matrix for Sociometric Exploration of the GQ

Relationship Quality	Relationship Structure		
	Member-Member	Member-Leader	Member-Group
Positive Working Relationship	Hypothesis 3	Hypothesis 4	–
Positive Bonding Relationship	Hypothesis 5	Hypothesis 6	Hypothesis 7
Negative Relationship	Hypothesis 8	Hypothesis 9	Hypothesis 10

In summary, the group therapeutic relationship has been identified as an important mechanism of therapeutic change. However, the group therapeutic relationship has been notoriously difficult to define due to overlapping constructs and disparate views as to whether the structure or quality of intra-group relationships is paramount. Johnson et al. (2005) consolidated these views and offered their 3-factor model as a clarifying framework for the group therapeutic relationship. Krogel (2009) used the Johnson Model as the basis for constructing the Group Questionnaire. The GQ has the potential to be a clinically useful measure of empirically valid therapeutic relationship constructs. However, its capacity for depicting relationship quality at the structural level is suspect.

The purpose of this study was to establish the validity of the GQ by 1) providing a final replication of the Johnson Model (Hypothesis 1), 2) comparing the GQ's psychometric

properties to those of the measures it was derived from (Hypothesis 2), and 3) exploring the GQ's ability to measure Positive Bonding, Positive Working, and Negative Relationship constructs at member-member, member-leader, and member-group structural levels using a sociometric test (Hypotheses 3-10). The overarching goal of this study was to establish the GQ as a valid and, therefore, useful tool for group clinicians to track the therapeutic relationships in their groups.

The following is a brief synopsis of contents of this dissertation. The Literature Review will review how the group therapeutic relationship has been approached in the empirical and theoretical literature, examine the empirical pedigree of the GQ, and review the history and uses of sociometry. The Method section provides an overview of the study methodology, participants, data collection procedures, data analysis techniques, and therapeutic relationship measures examined in this study. The Results section reviews the results of the construct validity analysis (i.e., model replication), criterion-related validly analysis, and sociometric exploration. The Discussion secion summarizes the study results, discusses the implications of the findings, reviews the limitations of the study methodology, and offers suggestions for future research.

Literature Review

The purpose of the following review is to establish the need for an exploration of the GQ's validity. I will first review how different researchers have approached the group therapeutic relationship in the recent empirical and theoretical literature. Second, I will review the empirical history behind the GQ's creation and discuss its clinical utility. Third and finally, I will review the history of sociometry in an effort to provide support for its use as a criterion for comparison in this study.

The Group Therapeutic Relationship

An examination of the extant group psychotherapy relationship literature reveals a nebulous collection of overlapping, underdeveloped, and ill-defined constructs competing for ideological supremacy (Burlingame et al., 2011; Johnson et al., 2005; McClendon & Burlingame, 2011). While the group therapeutic relationship has been called many things over the last several decades, four constructs have received the most empirical and theoretical attention: group cohesion, alliance, climate, and empathy. For an exhaustive review of these four constructs, the interested reader is directed to Jennifer Johnson's and JulieAnn Krogel's dissertations (Johnson, 2004; Krogel, 2009). Because a portion of my dissertation is a replication of Johnson's efforts to distil these four constructs, I will briefly discuss how each construct is thought to characterize the group therapeutic relationship. I will then discuss how the Johnson Model seeks to concretize and shore up the group therapeutic relationship construct.

Cohesion. Not only has cohesion been touted as one of the most important therapeutic factors in group therapy (Yalom & Leszcz, 2005), for some researchers it has become synonymous with the group therapeutic relationship (Burlingame et al., 2002). For all the credit it receives, cohesion has a notorious reputation for being difficult to define and measure. The more popular definitions include positive within-group atmosphere (Dion, 2000), a basic bond or uniting force (Piper, Marrache, Laroix, Richardsen, & Jones, 1983); attraction to the group (Frank, 1957); a sense of support, acceptance, and identification in the group (Bloch & Crouch, 1985); engagement (MacKenzie & Tschuschke, 1993), the quality of the patient's relationship to the group as a whole (Joyce, Piper, & Ogrodniczuk, 2007), and the total field of forces that influence members to stay in groups (Festinger, Schacter, & Back, 1950).

Critics of the cohesion construct look at these myriad definitions and observe that, ironically, there is precious little cohesion in the cohesion literature (Bednar & Kaul, 1994; Dion, 2000). This lack of cohesion in the cohesion literature led Hornsey, Dwyer, Oei, and Dingle (2009) to conclude that contrary to popular belief, there is no convincing and consistent link between cohesion and outcome. They contend that cohesion may actually have negative implications for members of group therapy (e.g., mutes personal expression) and advocate that cohesion be abandoned in favor of more concise, measurable constructs (i.e., identification, homogeneity, and interdependence; Hornsey, Dwyer, & Oei, 2007). However, recent meta-analytic research suggests that Hornsey and his colleagues may have spoken too soon regarding cohesion's spurious link to outcome. Burlingame et al. (2011) conducted a meta-analytic review of cohesion's relation to outcome in 40 studies spanning four decades of research. Accounting for the many different definitions of cohesion cited above, their overall conclusion was that there is a positive relation between cohesion and outcome (i.e., $r = .25$) that holds across theoretical orientation and treatment setting.

Alliance. The alliance is a well-researched construct in the individual therapy literature and has developed a reputation for being integral to positive treatment outcomes (Barber, Khalsa, & Sharpless, 2010, Horvrath, 1994; Martin, Garske, & Davis, 2000). It is typically characterized as the collaborative relationship between a patient and a therapist wherein the goals and tasks of treatment are mutually condoned in a context of warmth and respect (Barber et al., 2010; Bordin, 1979). Unsatisfied with this general definition, researchers have taken to pairing the alliance with various adjectives in an effort to give it conceptual specificity. Examples include ego alliance (Sterba, 1934), working alliance (Greenson, 1967), therapeutic alliance (Zetzel, 1956), and helping alliance (Luborsky, 1976). Like with the other therapeutic relationship constructs

discussed in this section, the varied approaches to the alliance construct make it difficult to understand and to measure.

The confusion only intensifies when examining the alliance construct in a group therapy context. Unlike individual therapy, group therapy is comprised of a nebulous web of multiple relationships occurring all at once. Some group researchers have attempted to navigate this interpersonal nebula by subdividing the alliance into the specific client/therapist relationship (i.e., individual alliance) and the relationship between the client and the entire group (i.e., group alliance; Johnson, Penn, Bauer, Meyer, & Evans, 2008). Others imply a direct transplantation of the individual alliance to the group context by suggesting that only the individual alliance between one group member and one group leader be considered as *the alliance*, while the other group relationships be relegated to the cohesion construct (Joyce et al., 2007; Marziali, Munroe-Blum, and McCleary, 1997; Piper & Ogrodniczuk, 2010).

Alliance and cohesion are frequently compared and contrasted in the group therapy literature, with several researchers contending that they are simply different names for the same phenomenon (Budman et al., 1989) or at least the same phenomenon occurring in different contexts (Fuhriman & Burlingame, 1990). Bakali, Wilberg, Hagtvet, and Lorentzen (2010) disagree, however, based on conclusions drawn from their own research on therapy group members' perceptions of alliance and cohesion across treatment stages. They analyzed 14 different sources of variation (therapist, patient, group, therapist x group, etc.) accounting for patient-rated alliance and cohesion. They ultimately concluded that because the sources accounting for the variability of alliance and cohesion in their sample differed, alliance and cohesion are likely distinct constructs.

Climate. Like alliance and cohesion, group climate has suffered from various definitions and measurement approaches. In their review of the group climate literature, McClendon and Burlingame (2011) observed that researchers generally conceptualize group climate in three different ways. First, climate is often described as the general emotional atmosphere of the group. This "emotional atmosphere" is thought to include anything from the group's consensually-perceived psychosocial environment to the group's feeling tone. Second, climate is sometimes viewed as the product of interacting relationship dynamics in a group. These dynamics can be oppositional (e.g., anger vs. harmony, engagement vs. avoidance, closeness vs. withdrawal) or complimentary (e.g., resistance and friction). Third and finally, climate is often conflated with what some would consider its conceptual alter ego, cohesion.

In addition to reviewing the various ways climate has been defined, McClendon and Burlingame (2011) reviewed previous efforts to operationalize and measure climate. They identified five different measures of group climate, each with its own approach, emphasis, and problems. For example, the Group Atmosphere Scale (GAS; Silbergeld, Koenig, Manderscheid, Meeker & Hornung, 1975) was designed to capture the group climate by measuring conformity, aggression, order, practicality, and variety in the group. However, the GAS also has a "Group Cohesion" subscale, which has led some researchers to argue that it is more a measure of cohesion than of climate. Of the five group climate measures McClendon and Burlingame (2011) reviewed, the Group Climate Questionnaire (GCQ; MacKenzie, 1983) was by far the most researched and most widely used measure. Touting it as the "one bright spot in the group climate literature" (p. 8), McClendon and Burlingame (2011) noted that most of what we know about group climate comes from research using the GCQ. Therefore, when group climate is associated with other small group processes or with outcome, more often than not these

associations are made using the GCQ's three subscales (Engaged, Avoiding, and Conflict) as the criterion for comparison.

Empathy. Rogers, Gendlin, Kiesler and Truax (1967) defined empathy as:

> The ability of the therapist [to] accurately and sensitively understand experiences and feelings and their meaning to the client during the moment to moment encounter of psychotherapy… It is a sensing of the client's inner world… A high level [of empathy] will indicate not only a sensitive understanding of the apparent feelings but… by its communication clarify and expand the patient's awareness of these feelings or experiences. (pp. 104-105)

Empathy has long been considered an important relationship construct for individual therapy (see Bohart & Greenberg, 1997). Yalom and Leszcz (2005) assert that empathy is a particularly important component to the group therapeutic relationship because it reduces the likelihood that members will harm one another and "…facilitates transfer of learning from the therapy group to the client's larger world" (p. 43). Ultimately, the role empathy takes on in the group therapeutic process depends on one's theoretical vantage point. For instance, psychodynamic, humanistic, and cognitive-behavioral schools of thought all assert that empathy is a therapeutic relationship variable in group treatment, but each school describes empathy differently.

In the psychodynamic tradition, empathy has been referred to as affective attunement (Schain, 1989), empathic attunement (Kleinberg, 1999), emotional availability (Shields, 2000), and tenderness (Livingston, 1999). Regardless of what they call it, psychodynamic group leaders rely on empathy to enable a transformative therapeutic relationship (Stone & Gustafson, 1982). Group leaders and other members who show empathy provide a safe environment for processing shame and guilt and for the exploration of unmet needs (Shapiro, 1991; Shields, 2000; Stone &

Whitman, 1977). Psychodynamic groups often rely on the tender responsiveness engendered by empathic bonds to promote emotional vulnerability. Once tenderness, understanding, and vulnerability are the relationship norms within a group, members are free to experiment with personality adjustment (Livinston, 1999).

For the humanistic group therapist, empathy is the primary vehicle for personality change and self-actualization (Rachman, 1981). Empathy is often invoked alongside the hallmark humanistic constructs of warmth, acceptance, genuineness, and unconditional positive regard. However, it is not synonymous with these constructs. Instead, empathy is the rich soil from which warmth, acceptance, genuineness, and unconditional positive regard sprout. As a group therapeutic relationship factor, humanistic thinkers believe empathy is curative when it increases clients' self-acceptance and helps them become more psychologically congruent (Trad, 1993).

When it comes to mechanisms of change in cognitive-behavioral groups, empathy takes a back seat, along with the other therapeutic relationship constructs, to specific technical interventions (Lambert, 2004). Empathy is a broader construct for the cognitive-behavioral group leader than it is for humanistic or psychodynamic group leaders. For example, cognitive-behavioral writers often subsume many other therapist-offered conditions into empathy, such as respect, warmth, and genuineness (see Burns & Auerbach, 1996). However, empathy is not considered curative in and of itself. Instead, empathy sets that stage for group member acceptance of a group leader's more active and incisive interventions.

Johnson Model. Given the overlap between, conflation of, and confusion about the aforementioned characterizations of the group therapeutic relationship, it is not surprising to find the research community bogged down in what some have called a "conceptual quagmire" (McClendon & Burlingame, 2011, p. 2). Fortunately, Jennifer Johnson's new model of the

group therapeutic relationship shows promise as a possible pathway to conceptual clarity. Praised for its large and diverse sample size, the Johnson et al. (2005) study has been called everything from a research "muse" (McClendon & Burlingame, 2011, p. 21) to "a harbinger for future cooperation and extension among complex group therapy research programs" (Burlingame, 2010, p. 3). As was explained in the Introduction, Johnson et al. (2005) produced a multi-level model that accounts for both the quality and structure of the group therapeutic relationship (see Figure 2-1). *Structure* refers to the direction of the relationship, or whom the relationship is with, and is comprised of member-to-member, member-to-leader, and member-to-group relationships. *Quality* refers to the content of the relationship (e.g., belonging, acceptance, collaboration, contention) and is comprised of the Positive Bonding, Positive Working, and Negative Relationship factors. The Positive Bonding Relationship describes the affective relationship group members have with their leader and with other members (i.e., cohesion, engagement, and empathy). The Positive Working Relationship captures the extent to which the group members feel like they are working toward mutually agreed upon goals by engaging in specific therapeutic tasks. The Negative Relationship represents empathic failures with the leader and conflict within the group. Johnson et al. (2005) and the subsequent empirical replications (i.e., Bakali et al., 2009; Bormann & Strauss, 2007; Bormann et al., 2011; Krogel, 2009) have all come to the same general conclusion regarding this distillation of the various group therapeutic relationship constructs: The Johnson Model provides an empirically robust and parsimonious framework for describing the therapeutic relationship in group treatment.

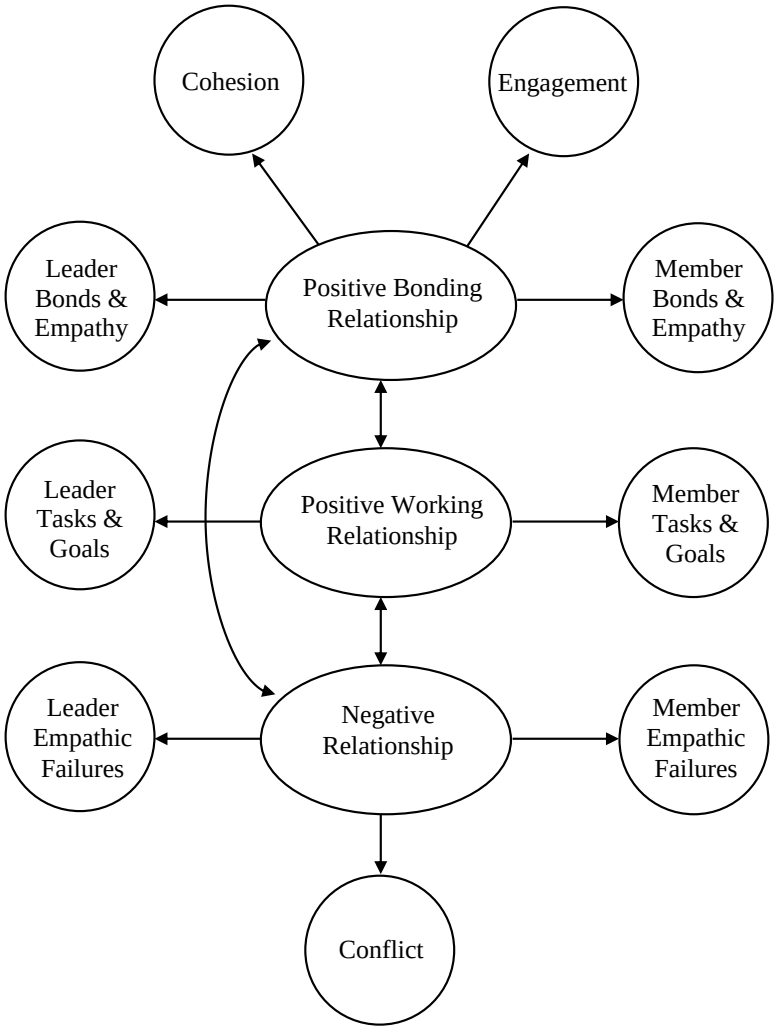


Figure 2-1. The Johnson Model of the group therapeutic relationship. Positive Bonding, Positive Working, and Negative Relationship exist between members and between members and leaders. Conflict, Cohesion, and Engagement represent the member-to-group relationship.

The Johnson Model is not without its critics, however. In their book chapter on the therapeutic alliance in group therapy, Piper and Ogrodniczuk (2010) cite several reasons why they believe the Johnson et al. (2005) study "did not produce compelling findings" (p. 267). First, they note that some of the psychometric properties of the measures, such as the internal consistency of the GCQ, were problematic. Second, they criticize the selection of "only a few constructs" that ultimately were "not reducible to a smaller number of basic factors" (p. 267).

Third, they claim that Johnson et al. (2005) did not "provide evidence that supported their hypothesized conceptual models into a structural model" (p. 267). Fourth and finally, they assert that the 3-factor structure of the Johnson Model should not be considered valid "until verified by confirmatory analysis or replicated with an independent sample" (p. 267). Since Piper and Ogrodniczuk (2010) provide the most recent and direct criticisms of the Johnson Model in the published literature, I will address each of their concerns in turn.

Regarding Piper and Ogrodniczuk's (2010) first criticism, Johnson et al. (2005) acknowledge the poor internal consistency of the Avoidance subscale of the GCQ (.36), stating that it consistently failed to load well with other subscales in their study. They also note other studies where its psychometric properties have also been problematic (e.g., Hurley & Brooks, 1987; Kay, 1996). They defend their use of the GCQ by indicating that the GCQ is one of the most commonly used measures in group psychotherapy research and is often considered to be the embodiment of the group climate construct. In their final analysis, Johnson et al. (2005) dropped the Avoidance subscale items from their model due to its poor internal consistency. Johnson et al. (2005) go on to encourage future research that examines the correlations between common scales and more psychometrically sound scales in order to reduce measurement difficulties and conceptual ambiguity.

Regarding their second criticism, it is unclear what Piper and Ogrodniczuk (2010) meant when they criticized Johnson et al. (2005) for addressing too few constructs. They are not explicit (or even implicit) about how many constructs would qualify as "enough" for a praiseworthy analysis. Johnson et al. (2005) identified four core constructs that had received the most empirical, psychometric, and theoretical attention in the group process and outcome literature at the time (Burlingame et al., 2002). Due to the conceptual overlap of other

therapeutic relationship constructs, it seems reasonable to assume that to include many more than the four identified by Johnson et al. (2005) would have been redundant. In addition, the Johnson et al. (2005) study was the first study at the time that endeavored to compare member-member, member-leader, and member-group relationships across more than one construct at a time, making Piper and Ogrodniczuk's (2010) desire for even more constructs seem greedy. Piper and Ogrodniczuk (2010) also asserted that Johnson et al. (2005) were unable to reduce their four constructs to a smaller number of basic factors. While their data did not fit well with a one- (Therapeutic Relationship) or two-factor (Bonding and Working) model of the group therapeutic relationship, Johnson et al. (2005) conducted a post hoc exploratory factor analysis that yielded a three-factor model with which the data fit nicely (i.e., the Johnson Model). Since they began with four constructs (i.e., group climate, cohesion, alliance, and empathy) and ended up with three factors (i.e., positive bonding, positive working, and negative relationship), it seems as if Piper and Ogrodniczuk (2010) were incorrect in their criticism.

Because Piper and Ogrodniczuk (2010) did not expound on what they were referring to, it is difficult to interpret their third criticism. If by "structural model" they are referring to the three, heretofore delineated, structural parameters of the group therapeutic relationship, then it is true that Johnson et al. (2005) concluded that group members in their sample tended to make the strongest distinctions in terms of relationship quality as opposed to whom they were relating to. When they did make structural distinctions, group members made them between member-member and member-leader, but could not distinguish member-member from member-group relationship parameters. Replications of the Johnson Model with independent and diverse samples have found support for all three structural components (e.g., Bormann & Strauss, 2007), but these mixed results make it unclear how much structural relationship distinctions add to the

conceptual clarity of group therapeutic relationship. Clarifying this issue is one of the primary aims of the present study.

The fourth and final criticism made by Piper and Ogrodniczuk (2010) suggests that they are unfamiliar with the many successful replications of Johnson et al.'s 3-factor structure (see Burlingame et al., 2011). Indeed, the present study seeks to provide the capstone to an ambitious and complex group therapy research program examining results from nearly 2200 group members in four countries with very diverse clinical populations (Burlingame, 2010).

The History and Utility of the Group Questionnaire

Conceptual ambiguity in the group therapy literature about how to define the therapeutic relationship construct has created a fertile atmosphere for measure proliferation. For example, Burlingame and his colleagues (2011) identified nine measures that claimed to capture cohesion, and these were only a sample of the most frequently used/studied measures. Observing that the Johnson Model offered conceptual parsimony and clarity to group therapy researchers, JulieAnn Krogel (2009) dedicated her dissertation research efforts to creating an empirically sound, clinician-friendly measure that would provide the same clarity to group clinicians: i.e., the GQ.

Krogel assembled a team of experienced group clinician/researchers to sort through the 60 items Johnson et al. (2005) used to create the Johnson Model. The team's analysis of these 60 items was conducted in two phases—an empirical analysis, followed by a thorough item review using clinical judgment. In the first phase, Krogel and her team examined the factor loadings and inter-item correlations of each item to determine its strength and uniqueness. In general, items with smaller factor loadings or high correlations with other items were considered for elimination. In the second phase, Krogel and her team established a clinically relevant definition for the content domain of each GQ subscale based on the items contained therein. They

considered items for inclusion based on their clinical relevance to the subscale definitions and on their empirical strength. Through this process they selected, eliminated, reworded, and combined items until they had 40 items that they thought captured the group therapeutic relationship as depicted by the Johnson Model.

After whittling 60 items to 40, Krogel and her team tested the GQ by administering it to 486 group members from university counseling center, American Group Psychotherapy Association, and Utah State Hospital groups. They used a confirmatory factor analysis to determine the goodness of fit between the sample data and the Johnson Model. Analysis results indicated that the data from the 40-item GQ did not fit the Johnson Model adequately. The reliability estimates of the Positive Bonding, Positive Working, and Negative Relationship subscales were .93, .90, and .84, respectively. After examining factor loadings in her own dataset, Krogel dropped 10 more items that possessed small regression weights and high inter-item correlations. A second confirmatory factor analysis revealed that the 30-item version of the GQ fit well with the Johnson Model ($\chi^2(381) = 775.4$; CFI = .96; RMSEA = .05; N = 486). The data did not fit the model, however, when relationship structure (i.e., member-member, member-leader, member-group) was removed ($\chi^2(403) = 3049.5$; CFI = .71; RMSEA = .12; N = 486), lending support for the retention of that factor in the Johnson Model. Reliability estimates for the Positive Bonding, Positive Working, and Negative Relationship subscales of the 30-item GQ were .92, .90, and .80, respectively. Krogel postulated that the slight decrease in reliability from the 40-item to the 30-item GQ was at least partly due to the shorter length of the 30-item version. They also observed large discrepancies in the reliability of the Negative Relationship subscale by population and suggested that restriction of range was responsible for its relatively low reliability. Krogel compensated for this restriction-of-range problem using a formula devised by

Ghisell, Campbell, and Zedeck (1981). This formula predicts a test's reliability for a sample population based on the reliability of another population and the variances and observed scores for both populations. Krogel's predicted reliability estimate for Negative Relationship was .90.

As part of the large international research consortium studying the measurement of group process variables, Bormann, Burlingame, and Strauss (2011) translated the GQ into German and compared it to four more criterion measures. They administered the German version of the GQ along with the Group Experience Scale (Strauss & Eckert, 1994), Stuttgart Questionnaire (Tschuschke, 1996), Helping Alliance Questionnaire (Bassler, Potratz, & Krauthauser, 1995), and Bonn Questionnaire for Therapy and Counseling (Fuchs, Sidiropoulou, Vennen, & Fisseni, 2003) to 424 group members in 9 different German hospitals. Confirmatory factor analysis revealed a good overall model fit for the German GQ ($\chi^2(376) = 818$; CFI = .97; RMSEA = .05; N = 424) and reliability estimates were almost identical to those found by Krogel (2009; Positive Bonding Relationship = .92, Positive Working Relationship = .89, and Negative Relationship = .79). Bormann et al. (2011) concluded that their data supported the construct validity of the German GQ.

Krogel and her team set out to construct an empirically sound and clinically useful measure of one of the most therapeutically relevant constructs in group psychotherapy. They used the Johnson Model as a template for the GQ. Recent praise and adoption of the Johnson Model bodes well for the relevance of the GQ to the research-conscious clinician who wants to make his or her practice evidence-based (Burlingame, 2010; Chapman, 2010; McClendon & Burlingame, 2011, Burlingame et al., 2011).

Sociometry

Given its widespread application in a number of fields (see Anshel, 1994; Breen, 1994; Buchanan, 1982; Gazda, 1982; Hinshaw & Melnick, 1995; Lee, 1991; Mouton, Blake, & Fruchter, 1960; Pareek & Singh, 1968), some view sociometry as one of the paramount contributions to the social and behavioral sciences (Treadwell et al., 1998). Jacob Moreno, the acclaimed father of sociometry, touted it as the conceptual progenitor to what we now know as group psychotherapy (1953). Broadly speaking, sociometry is the study and graphical representation of relatedness among people in a group (Treadwell et al., 1998). In sociometric terms, "relatedness" is operationalized as interpersonal preferences group members have regarding one another in reference to a particular criterion. These interpersonal preferences are examined using a sociometric test, which is a means for determining the varying degrees of acceptance, isolation, collaboration, sub-grouping, and overall social structure of a group by asking members to make hypothetical choices about the parameters of their in-group relationships (Northway, 1967). The criterion for investigation is determined by the investigator (e.g., group leader, researcher, teacher, coach) and typically takes the form of a question. For example, a group leader concerned about subgrouping might pose the question, "Whom in this group would you like to have lunch with?" Group member responses are then tabulated in a sociomatrix so that preferences for and of each member can be examined for mutual agreement, conflicting preferences, and total scores. These data are then used to construct a sociogram—a graphical representation of where each group member stands regarding how they are preferred by, and how they prefer, other members in relation to the criterion of interest (e.g., Figure 2-2).

In his comprehensive review and critique of sociometric techniques and measurement theory, Robert Terry (2000) discussed the various methodological issues relevant to conducting

sociometric tests, 3 of which are most pertinent to the present study. 1) The type of criterion presented to the group members; 2) The methods for soliciting sociometric choices; 3) Quantifying and analyzing sociometric information. I will review Terry's (2000) treatment of these issues and then discuss the appropriateness of sociometry as a psychometric validation criterion.

Type of criterion presented to the group. As was stated above, sociometric nominations are solicited using some type of question. These questions fall into four basic categories. First, there are *friendship questions.* An example of a friendship question would be, "Who are your best friends in the group?" or "Name your three best friends in the group."

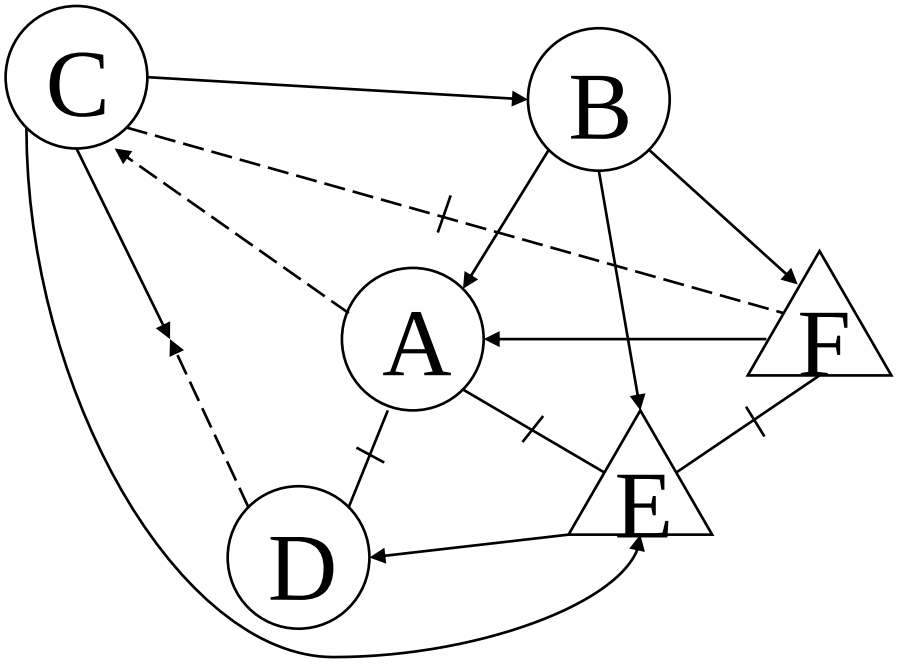


Figure 2-2. Example of a sociogram. This is one method of graphically representing interpersonal choices among group members. The meaning of each component (e.g., type of connecting line, nearness of each shape to the other shapes) depends on the particular sociometric method in use.

Friendship questions are some of the most commonly used criteria in sociometric investigations, especially in educational settings (e.g., Shoobs, 1947). Second, *direct preference*

questions are an explicit sampling of interpersonal likes and dislikes. For example, "Which group member do you like the best?" would be a direct preference question. Third, *acquaintance questions* like "Who in the class do you hang out with the most?" are often used to construct and evaluate peer social networks. Fourth, questions that require a group member to choose another member for a specific type of interaction are called *task-specific questions.* An example of a task-specific question would be, "Who in the group would you like to clean a house with?"

Moreno (1953) preferred task-specific questions because he thought they forced group members to make choices that were more concrete, comparably meaningful, and spontaneous than did the other types of questions. He argued that the variability in the criteria that different children use as the basis for selecting friends and acquaintances made the data produced by those types of questions too unreliable. He also thought that direct preference questions merely elicited psychological projections of an abstract and multidimensional construct (e.g., liking) that were ultimately too idiosyncratic to meaningfully compare sociometrically.

Methods of soliciting sociometric choices. Invariably connected to the four different types of criteria used for sociometric investigation are the four different methods of collecting sociometric choice information. First, the *peer-nomination* method asks groups members to nominate others according to the specific criterion (e.g., "Who is your best friend in the group?"). Second, the *rank-order* method asks each group member to rank other group members in relation to the criterion (e.g., "Rank-order everyone in the group based on how much you like them."). Third, the *peer rating* method asks group members to rate each group member on a Likert scale according to the criterion of interest ("On a scale of 1 to 5, rate how much you like each member of the group."). Fourth, the *paired comparisons* method asks group members to

make a preference choice from all possible dyadic pairings of the other group members ("Do you like James or John?" etc.).

While each method possesses its own strengths and weaknesses, Moreno (1953) preferred the use of peer-nomination because of the ease with which one can collect, quantify, and analyze sociometric data using this method. Moreno and others (e.g., Holland & Leinhardt, 1973) also preferred the unlimited-choice procedure for peer-nomination, meaning group members could nominate all of the group members instead of being limited to choosing just a few. Moreno (1953) argued that the unlimited-choice procedure allowed for the examination of certain important constructs like social expansivity (the tendency of group members to choose many others) and social isolation (the tendency for group members to receive no choices), whereas the limited-choice procedure would not. In addition, Holland and Leinhardt (1973) argued that using the limited-choice procedure would ultimately result in measurement error. For example, if a group member has exactly four good friends in a group and is only allowed to nominate three of them in response to the question "Who are your best friends in the group?" then the data elicited by that criterion are not as valid as they could be.

Quantitative analysis of sociometric data. There are several different methods for analyzing sociometric data. The most popular method is to sum or average the choices/rankings/ratings each group member receives (Hale, 1985; Treadwell et al., 1998). These data are then compared, correlated, and contrasted so as to gain insight into the nature of each group members' relationship to the rest of the group. Depending on the nature of the criterion question posed to the group, these data can be interpreted to represent social acceptance, rejection, status, isolation, and visibility. For example, by summing or averaging the number of positive or negative nominations a group member receives, one can infer how socially accepted

or rejected that group member is in the group. Indices like social isolation can be operationalized by examining how few nominations a group member receives in general, while socially visible group members can be identified by how often they elicit strong opinions in their peers (i.e., total number of extremely positive and extremely negative ratings they receive).

Not only do they believe it is informative to evaluate the choices received by each group member, Sociometrists also believe that is important to examine the choices *made* by each group member (Terry, 2000; Treadwell et al., 1998). In addition to revealing various social patterns, this allows group leaders to consider the reciprocated choices between group members, which are sometimes interpreted as an index of social support within the group (e.g., Bukowski & Hoza, 1989; Price & Ladd, 1986).

Sociometry and measure validation. Only two studies emerged from the extant sociometry and group process measurement literature that used a sociometric test as the principle validation criterion for another measurement tool. Corcoran (1989) sought to test the discriminant validity of the Sport Competition Anxiety Test (SCAT) by comparing it to sociometric test responses from 34 girls and 21 boys on 5 high school basketball teams. The SCAT is a 15-item self-report questionnaire used to measure competition-related trait anxiety (Martens, 1977). Corcoran's (1989) sociometric test was comprised of 6 unlimited-choice, rank-ordering, task-specific criterion questions (e.g., "Who on the team do you spend the most time with?") that represented different constructs he postulated were relevant to the SCAT (e.g., Nervousness). Corcoran (1989) was not explicit about his method for quantifying sociometric data, nor did he construct a sociogram to graphically represent the data. He used a heterotrait-heteromethod matrix to correlate SCAT scores with the each player's sociometric scores on each

sociometric question. He inferred based on the correlations that the SCAT was a valid instrument.

Wood, Kumar, Treadwell, and Leach (1998) studied the relationship between sociometric choices and cohesion in groups that were learning and applying psychodrama techniques in the exploration of interpersonal issues regarding intimacy. They measured cohesion using the Group Cohesiveness Scale (GCS; Veeraraghavan, Kellar, Gawlick, & Morein, 1996), a 26-item instrument that measures cohesion by surveying member retention, interaction among group members, and compatibility of individual and group goals. Wood et al. (1998) compared GCS scores with group member sociometric status (i.e., group popularity). Sociometric criteria included 6 limited-choice, rank-ordering, direct preference and task-specific questions. Sociometric status scores for each group member were determined by summing the number of choices they received. They did not construct sociograms as part of their analysis. Wood et al. (1998) hypothesized that the more popular group members (i.e., those who received the most sociometric choices) would perceive their groups to be more cohesive (i.e., have higher GCS scores). They found that only one group out of the six that they studied exhibited this correlation, offering only limited support for their hypothesis.

While both of these studies used a sociometric test to investigate the validity of a psychometric instrument, the Wood et al. (1998) study represents the closest approximation to the present study. Neither study utilized sociograms to interpret sociometric data, favoring the quantification of choices received instead. The limited precedent for scrutinizing group process measurement tools using sociometry is surprising given the depth and breadth of information one can obtain about relationships in a group from a sociometric test (Hale, 1985; Terry, 2000; Treadwell et al., 1998).

Method

Instruments

The GQ is a 30-item self-report questionnaire designed to measure the quality of the therapeutic relationship in group treatment. Items are scored on a 7-point Likert scale ranging from *not at all true* (1) to *very true* (7). Item scores are compiled to produce three subscale scores: Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship. The reliability estimates (Cronbach's alpha) of the three subscales are .92, .90, and .80, respectively (Krogel, 2009).

The Working Alliance Inventory (WAI; (Horvath & Greenberg, 1989) is a 20-item self-report questionnaire designed to capture three aspects of the working alliance between a client and a clinician. The Task subscale attempts to measure the extent to which the client perceives therapeutic activities as relevant and beneficial. The Goals subscale attempts to measure the extent to which the client believes that he/she and the clinician are united on the desired treatment outcome. The Bonds subscale attempts to measure the extent to which a client perceives mutual trust, acceptance, and confidence in the attachment between the client and clinician. In keeping with the procedure adopted by Johnson, et al. (2005), the current study will use all the items from the Bonds subscale and shortened versions of the Task and Goals subscales (Tracey & Kokotovic, 1989). These 20 items will be asked twice, once in reference to the working alliance between the group member and the leader and once in reference to other group members. Modifying the WAI for use in groups has some precedent in the group literature (e.g., Johnson et al., 2008). Items are scored on a 7-point Likert scale ranging from *not at all true* (1) to *very true* (7). Reliability estimates for the various subscales range from .56 to .90 (Johnson et al., 2005).

The Group Climate Questionnaire—Short Form (GCQ; (K. R. MacKenzie, 1983) is a self-report measure that purports to assess individual group members' perceptions of the closeness, withdrawal, and conflict within the group. The GCQ contains 12 items rated on a 7-point Likert scale, ranging from *not at all* (0) to *extremely* (6). Its subscales include Engagement, Avoidance, and Conflict. Construct validity of the GCQ has been found to be adequate (Johnson et al., 2005). Cronbach's alphas for the GCQ subscales in Johnson et al. (2005) are as follows: .70 for Engagement, .36 for Avoidance, and .69 for Conflict.

Therapeutic Factors Inventory (TFI; (Lese & MacNair-Semands, 2000) is a self-report measure designed to assess group members' perceptions of the degree to which the therapeutic factors described by Yalom (1995) are present. The nine items from the Cohesion subscale will be used. Each item is scored on a 7-point Likert scale that ranges from *strongly disagree* (1) to *strongly agree* (7). The Cohesion subscale has been used in a counseling center population, with a test–retest reliability of .93 over a 1-week period (Lese & MacNair-Semands, 2000). Cronbach's alpha in this sample was .90.

The Empathy Scale (ES; (Burns & Auerbach, 1996) is a self-report questionnaire that contains 10 items that reflect clients' perceptions of the therapist's warmth, genuineness, and caring during the most recent therapy session. Five items are worded so that agreement indicates a positive therapeutic relationship and five are worded so that agreement indicates a negative relationship. For this study, responses will be scored on a 7-point Likert scale, with response options ranging from *not at all* (0) to *extremely* (6). Like the WAI items, ES items will be asked twice, once referring to the group leader and again referring to the group members. Coefficient alpha in Johnson et al. (2005) was .83 for the both the Member–Member and the Member–Leader ES. For a summary of how all the measures are represented in the GQ see Table 3-1.

Table 3-1

Relationship Between Criterion Measures and GQ Subscales

Relationship Quality	Relationship Structure		
	Member-Member	Member-Leader	Member-Group
Positive Working Relationship	WAI-Tasks WAI-Goals	WAI-Tasks WAI-Goals	–
Positive Bonding Relationship	WAI-Bonds ES (positive)	WAI-Bonds ES (positive)	TFI-Cohesion GCQ-Engagement
Negative Relationship	ES (negative)	ES (negative)	GCQ-Conflict

Note. The GCQ Avoidance subscale is excluded as it was in the replicated Johnson et al. (2005) model due to poor model-fit.
"–" indicates that Positive Working Relationship is not measured at the member-group level.

In an effort to stay true to Moreno's (1953) original formulations and to be consistent with Terry's (2000) recent recommendations, group members were presented with three, task-specific, peer-nomination, unlimited-choice sociometric criterion questions:

SQ1: "Of the members in your group, with whom would you choose to work on an important project?" (Positive Working Relationship.)

SQ2: "Of the members in your group, from whom would you choose to receive direct support during a difficult life challenge?" (Positive Bonding Relationship.)

SQ3: "Of the members in your group, who would you most likely vote off the island? (i.e., Who would you least prefer to associate with or even distance yourself from?)" (Negative Relationship.)

These questions were formulated in accordance with the guidelines outlined in Treadwell (1998), Hale (1985), and Northway (1967) for valid sociometric investigation. Group members responded to SQ items by indicating their preference for every other member in attendance with

"yes," "no," or "no preference." Raw scores were compiled in a sociomatrix, after which average valence scores were calculated for each participating member.

Participants

Two hundred and ninety participants were recruited from 65 treatment groups: 12 (4.1%) participants from the Brigham Young University Comprehensive Clinic in Provo, Utah; 193 (66.6%) participants from the Brigham Young University Counseling and Career Center; 24 (8.3%) participants from Southern Utah University Counseling and Psychological Services in Cedar City, Utah; 33 (11.4%) participants from Utah State University Counseling and Psychological Services in Logan, Utah; and 28 (9.7%) participants from University of Utah Counseling Center in Salt Lake City, Utah. The sample included 148 (51%) males and 142 (49%) females. Two hundred and sixty-one (90%) participants identified themselves as Caucasian, 7 (2.4%) as Chicano/Latino, 6 (2.1%) as Asian/Pacific Islander, 1 (0.3%) as African American, and 15 (5.2%) as Other. The mean age of participants was 24 (SD = 6.04, range = 17-58). The mean number of sessions attended by participants at the time of data collection was 9.63 (mode = 5, SD = 13.8, range = 1-150). Participants were sought from a variety of group types including general process, disorder specific, and psychoeducational groups. Groups were led by licensed psychologists, predoctoral interns, social workers, and marriage and family therapists whose theoretical orientations varied.

Procedures

Data collection occured one time for each group during the productive working phase of group development when intimacy, engagement, and cohesion peak and questions about relationship quality become relevant (Yalom & Leszcz, 2005). There are various theories about how group relationships develop over the course of treatment (e.g., Tuckman & Jensen, 1977).

While such theories can help group leaders formulate expectations and time interventions, they are heuristics that ultimately fail to account for the complexity and unpredictability of group development (Yalom & Leszcz, 2005). Therefore, it is reasonable to estimate that if data are collected after the group has had several sessions, such data will represent the working phase. Group members were asked how many sessions they had attended in order to provide a mean and range.

Participants were recruited via a short recruitment script read by their group leader at the end of a session. Participating members signed an informed consent document and were invited to complete the measures after their session was over. Completing the entire assessment battery generally took participants between 20 and 40 minutes. Participants were compensated with a $20 Amazon.com gift card immediately after completing the assessment battery. Due to the rapid depletion of study support funds, compensation was reduced to $10 for 58 participants. The reduction in compensation did not appear to affect participation rates. As a courtesy to participating groups, each group leader was given the GQ subscales scores of the participating members from their groups.

Analysis

The statistical analyses used in this study were selected to address the research questions implicit in each hypothesis. Therefore, I will review the various analyses in terms of the research questions they address.

Hypothesis 1: Does the GQ support the Johnson Model? Construct validity refers to a test's ability to measure the phenomenon it purports to measure (Allen & Yen, 2002; Warner, 2008). Therefore, the construct validity of the GQ refers to its ability to actually measure the quality of the group therapeutic relationship. As is the case with many social science variables,

the group therapeutic relationship is not a directly observable construct (e.g., one cannot scrape the group therapeutic relationship into a beaker after it collects on the walls of the therapy room). Therefore, the nature of the group therapeutic relationship must be extrapolated from directly observable variables thought to be related to it. A frequently used statistical method for evaluating the relationships between directly observable variables and unobservable constructs is factor analysis. Using factor analysis, researchers attempt to make inferences about unobservable constructs (i.e., latent factors) by examining the covariation among a set of observed variables (Byrne, 2010). When a researcher already has an empirically or theoretically based hypothesis about how observed variables relate to latent factors, he or she can use confirmatory factor analysis (CFA) to test these hypothesized relationships statistically (Byrne, 2010; Warner, 2008). Known as the *measurement model*, the pre-determined framework of relationships in a CFA is constrained by the researcher in that he or she specifies the number of variables (both latent and observed) in the model and restricts the relationship or "loading" of observed variables to specific latent factors.

Confirmatory factor analysis was deemed appropriate for three reasons: 1) The GQ was designed to measure an unobservable construct (i.e., the group therapeutic relationship as defined by the Johnson Model) using observable variables (i.e., the 30 questionnaire items). 2) Krogel (2009) and Bormann et al. (2011) used CFA in their evaluation of the GQ's construct validity, providing a theoretically based and empirically verified measurement model as a precedent for the present study. 3) CFA is the most robust method available for analyzing multiple covariances among a set of latent and observable variables (Byrne, 2010).

Figure 3-1 represents the model used for the CFA in the present study. This model was constructed and analyzed using AMOS statistical software (Arbuckle, 2010).

Not only does AMOS allow one to graphically represent a structural equation model, it also provides a theoretically-based, consistent, efficient, and relatively unbiased method for handling missing data. Known as the full information maximum likelihood (FIML) estimation, this method uses a predictive distribution of scores to model the underlying pattern of missing data and then replaces missing scores with estimated values based on this model (Byrne, 2010). FIML was used for the single-level CFA in the present study.

In order to determine how well the study data match the hypothesized factor model, several "goodness-of-fit" indices were calculated. Of the many fit indices available, the χ^2 ("chi-squared") goodness-of-fit statistic is by far the most widely used (Hu & Bentler, 1995). The χ^2 value represents the magnitude of the discrepancy between the sample data and fitted covariance matrices. The χ^2 is used to test the null hypothesis that the sample covariance matrix (**S**) and the covariance matrix implied by the model (**Σ**) are equivalent. Therefore, failing to reject the null hypothesis (H_0: $\mathbf{S} = \mathbf{\Sigma}$) is desirable in this case because it suggests that the sample data have an overall good fit to the proposed model. In addition to a non-significant *p* value, it is desirable to have a small χ^2 relative to the model's degrees of freedom. The generally accepted guideline for this test of model fit is that a χ^2 value that is less than twice the model's degrees of freedom indicates acceptable overall model fit (Hu & Bentler, 1995).

The root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) were also examined. These specific statistics were selected per Brown's (2006) recommendations for evaluating CFA model fit. Brown (2006) suggests that in order to conduct a robust analysis of model fit in a CFA, researchers should examine at least one fit index from three different categories: absolute fit, fit adjusting for model parsimony, and comparative or incremental fit. The χ^2 represents the test of absolute fit for this study.

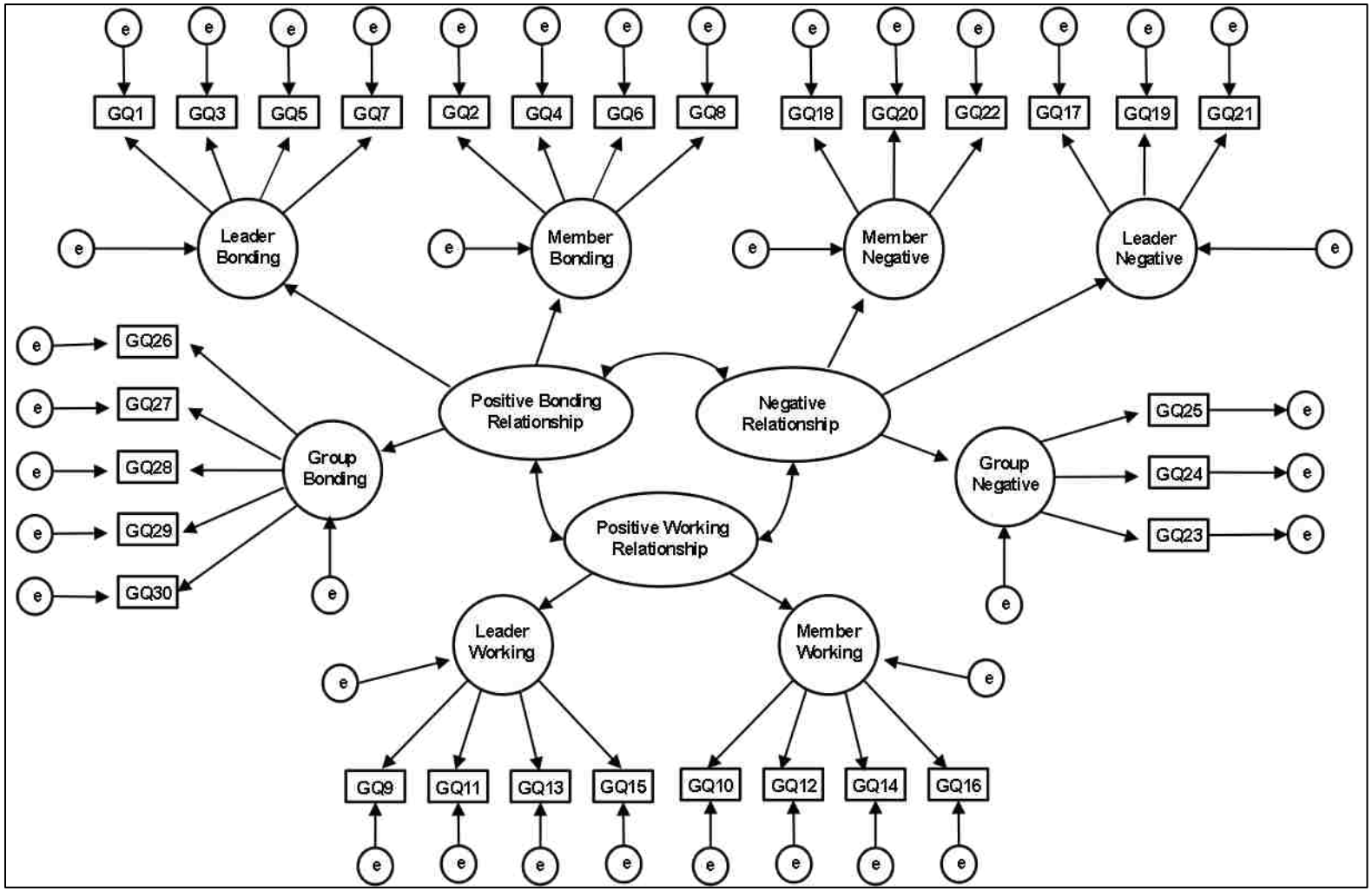


Figure 3-1. CFA model for the analysis of the GQ's construct validity. Ellipses and circles represent unobserved latent factors. Rectangles and squares represent observed variables. Single-headed arrows represent the impact of one variable on another. Double-headed arrows represent covariances or correlations between pairs of variables.

The RMSEA (parsimony adjustment) and the CFI/TLI (comparative fit) were specifically chosen because of the empirical research supporting their use and because they tend to be the "best behaved" indices available (Brown, 2006, p. 85). The RMSEA assesses whether the model holds *reasonably* well in the study sample, with values close to .06 or below indicating acceptable model fit (Hu & Bentler, 1999). The CFI and TLI offer a more liberal estimation than the χ^2 in that they evaluate model fit against a solution that posits no relationships among variables. Values close to .90 or greater are representative of acceptable model fit (Hu & Bentler, 1995).

Inherent in hierarchical data sets like the one from the present study (e.g., participants nested within groups) are risks of violating the "independence of observations" assumptions of most statistical tests (Baldwin, Wampold, & Imel, 2007; Burlingame, Kircher, & Honts, 1994). Ignoring intragroup dependency puts a study's analysis at risk for Type I error inflation (Baldwin, Murray, & Shadish, 2005, Baldwin et al., 2011; Baldwin, Stice, & Rohde, 2008). One method for evaluating the effects that intragroup dependency has on nested data is to compare statistical results at both the within-group and between-group levels (Baldwin et al., 2007). The intraclass correlations (ICC), which provide an effect size estimate for group-level effects, is one method by which to make such a comparison (Baldwin et al., 2011; Imel, Baldwin, Bonus, & Macoon, 2008). Although Johnson et al. (2005) addressed intragroup dependency in their original formulation of the Johnson Model, the two most recent validity studies on the GQ (i.e., Krogel, 2009; Bormann et al., 2011) did not. Because intragroup dependency poses a substantial threat to statistical integrity, it was accounted for in the present study.

Similar to Johnson et al. (2005), a multilevel CFA was conducted using the statistics software package Mplus (Muthen & Muthen, 2010). Mplus allows a researcher to obtain separate model fit statistics for the within-group and the between-group covariance structures. In

order to determine whether or not intragroup dependency affected model fit in the present study, ICCs were examined for valence and magnitude; positive ICCs were considered as evidence for within-group dependency (Baldwin et al., 2008).

Hypothesis 2: Does the GQ have criterion-related validity? Like construct validity, criterion-related validity refers to a test's ability to measure what it purports to measure. One assesses the criterion-related validity of a test by correlating it with other variables that the test should be related to (Warner, 2008). For example, if a researcher wanted to examine the criterion-related validity of the zombie depression scale (ZDS), he or she would need to examine the correlations between the ZDS and other measures of undead depression. In order to assess the GQ's criterion-related validity, GQ subscale scores were correlated with four other measures of the group therapeutic relationship: the WAI, GCQ, ES, and the Cohesion subscale of the TFI. (See Table 3-1 for a summary of which GQ subscale was correlated with which criterion measure.) The selection of these measures is particularly relevant to the problem of "construct drift" implied by in the title. Because these four measures were used as the template for the GQ's creation, low correlations between them and the GQ would suggest that the GQ has drifted off course in its efforts to measure what it purports to measure. Per the guidelines outlined in Reisinger and Burlingame (1997) for evaluating the validity of a psychotherapy questionnaire, validity coefficients at .50 and above were considered acceptable, while .75 and above were considered excellent. Both Pearson product-moment (r) and Spearman's rank (ρ) correlation coefficients were calculated. The addition of ρ was intended to offer a non-parametric correlation coefficient that would account for the skewedness of the data.

Crucial to a test's validity is its reliability. Reliability refers to how consistently a test measures what it purports to measure (Warner, 2008). For example, if a person were to step on

and off a bathroom five times in 30 seconds, a reliable scale would show the same weight each time. Because low reliability inflates the standard errors of estimates, it is a threat to a test's validity (Cook & Campbell, 1979). Therefore, reliability estimates were calculated and assessed. The most popular method of assessing the reliability of multiple-item tests like the GQ is by calculating the Cronbach alpha (Cronbach, 1951) statistic (Warner, 2008). Reliability estimates of .80 and above were considered acceptable, while .90 and above were considered excellent (Reisinger & Burlingame, 1997). In cases were reliability was attenuated by restriction of range, estimates were adjusted using a formula provided by Ghiselli et al. (1981) and reported separately.

Hypotheses 3-10: Does the GQ relationship structure? In order to explore the GQ's ability to measure the quality of the group therapeutic relationship at the three structural relationship levels (i.e., member-member, member-leader, and member group), GQ subscale scores were compared to SQ scores quantitatively using correlation coefficients and qualitatively using sociograms.

Quantitative sociometric analysis. As was demonstrated in Literature Review, there is very little precedent in the extant group process measurement literature for using sociometric data in measurement validity studies. Without the explicit guidance that more preceding research would provide, the method by which sociometric data were compiled, quantified, and correlated with GQ data in the present study was considered exploratory. Therefore, several different SQ variables were computed and compared to see which correlated with GQ item and subscale scores the best. I will now explain how each SQ variable was computed and the reasoning behind its composition.

Each of the three SQ items was designed to relate to one of the GQ subscales—item 1 for Positive Working Relationship, item 2 for Positive Bonding Relationship, and item 3 for Negative Relationship. In order to make direct comparisons at the level of relationship structure, SQ item total scores were calculated at the member-member, member-leader, and member-group levels. Total score calculations were based on a variety of response scaling schemes. At the member-member and member-leader levels, "yes," "no preference," and "no" were scaled as 1, 0, and -1, respectively. By scaling item responses this way, it was assumed that "yes" represented a completely positive preference (e.g., "Yes, I would like to work on an important project with Jack), that "no" represented a completely negative preference (e.g., "No, I would not like to work on an important project with Jack), and that "no preference" represented the complete absence of preference or apathy as opposed to ambivalence (e.g., "I have no preference about working on an important project with Jack). The mean of a group member's preferences for all the other members in his or her group was used as his or her member-member total score. The mean of a group member's preferences for the leaders of his or her group was used as his or her member-leader total score. The mean was used instead of the sum of choice scores so that item scores could be directly comparable between groups.

Quantifying the member-group relationship level was less straightforward. Unlike member-member and member-leader relationships, the member-group relationship has no clearly identifiable object of interest. In other words, it is unclear what a group member has a relationship with when we say he or she has a relationship with the "group-as-a-whole." Indeed, confusion and disagreement in the group therapy process literature about this issue is closely related to the lack of consensus about how to define the group therapeutic relationship (see Introduction and Literature Review). In spite of the many differing and sometimes contradictory

theories about the nature of group-as-a-whole relationships, at least one distinction seems consistent: the group-as-a-whole is more than the sum of its parts; it is the product of dynamic interacting relational forces (McClendon & Burlingame, 2011). It follows, then, that the member-group relationship as depicted by the SQ would need to be more than the sum of a member's unidirectional preferences about each member and leader in the group. This is precisely why sociometry was chosen for this study. It offers a level of measurement not achievable by self-report instruments like the GQ. That is, instead of being limited to unidirectional preferences and impressions about relationships, sociometry examines the dynamics of preferences and impressions. In other words, not only does it reveal what John's preferences are about receiving direct support from Jack, it reveals what Jack's preferences are about John, too. Based on the group-as-a-whole relationship theories outlined in McClendon and Burlingame (2011), these "dynamic choices" were considered the best approximation of the member-to-group-as-a-whole relationship that the SQ had to offer.

Dynamic choices on the SQ fall into two categories: mutual and incongruous. A mutual choice means that two group members gave each other the same SQ rating. An incongruous choice means that choices between group members differed. Eight different methods of calculating member-group dynamic choice scores were compared to see which ones correlated best with the GQ at the item and subscale levels (see Table 3-2). These methods fell under two categories: *single coding* and *multiple coding.* Single coding refers to scaling the dynamic choice of interest (e.g., *affirmative mutual* or when two members respond, "yes" regarding each other) as 1 and all other dynamic choices as 0. Multiple coding is the same as single coding except *negative mutual* (i.e., when two members respond, "no" regarding each other) choices are scaled as -1. Six of the eight methods were created specifically for use with the member-group

Negative Relationship subscale in response to uncertainty regarding which sociometric choice configuration best represented high Negative Relationship (i.e., affirmative mutual, incongruous, or neutral mutual choices).

Qualitative sociometric analysis. For the qualitative exploration of the GQ's validity using the sociometric test, a small sample of groups were extracted for the purpose of comparing sociographic composition to GQ subscale scores. Groups in this sample were selected based on the following 3 criteria: 1) All of the group members in attendance at the time of data collection participated in the study. 2) Each participating group member provided a complete sociometric test (i.e., rated every other participating member). 3) Each participating group member provided a complete GQ response set. These criteria were set to ensure that sociograms provided an exhaustive representation of the group's preferences. Forty-one total participants comprising 8 groups met the 3 criteria and were included in the qualitative analysis. A *social atom* (see Treadwell et al., 1998) sociogram was used, which places the member-of-interest at the center of several concentric rings upon which other group members are positioned to represent relational nearness and farness based on dynamic preferences (see Figure 3-2). Three concentric rings were employed to represent high, average, and low relationship quality. For SQ1 and SQ2, individuals with whom the member-of-interest had an affirmative mutual dynamic choice were placed on the first ring, negative mutual relationships were depicted using the third ring, and all other dynamic choice scores were depicted using the second ring. The inverse of this method was implemented for the negatively worded SQ3 item.

Table 3-2

Scaling Methods for Member-Group SQ Item Total Score Calculation

a	b	c	d
+/+ = 1	+/- = 1	+/+ = 1	o/o = 1
-/- = -1[†]	-/+ = 1	+/- = 1	-/- = -1[†]
	-/- = -1[†]	-/+ = 1	
		-/- = -1[†]	

Note. +/+ = affirmative mutual, -/- = negative mutual, o/o = neutral mutual.
[†] Only used for "multiple code" variables.

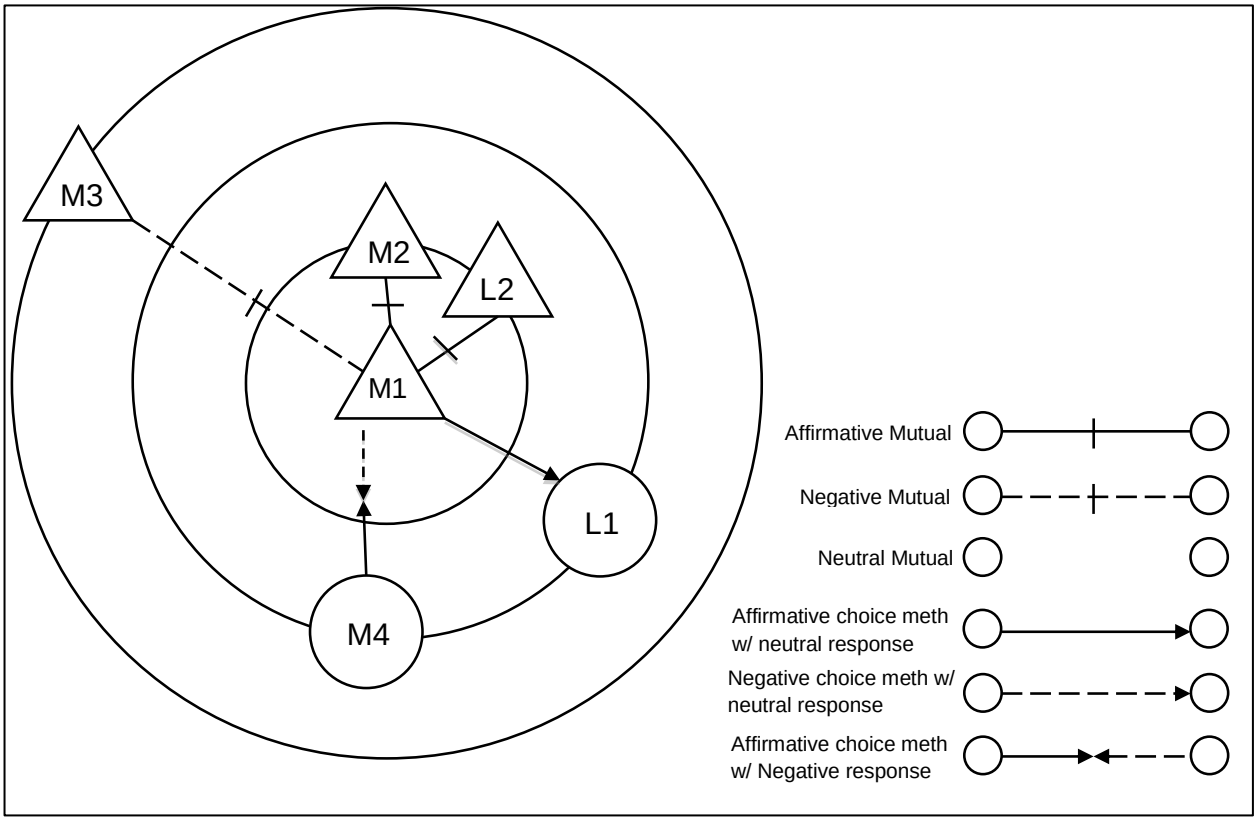


Figure 3-2. Example of the "social atom" sociogram used in the present study to graphically represent group member preferences regarding other members and group leaders. The member-of-interest is placed in the center and his or responses on the SQ determine where other group members/leaders are positioned. Circles represent females and triangles represent males. "M" refers to "member" and "L" refers to "leader."

Because dynamic choice scores were unavailable between members and leaders, unidirectional member-to-leader "yes," "no preference," and "no" choices were depicted on the first, second, and third rings, respectively.

Results

GQ item and subscale descriptive statistics are reported in Tables 4-1 and 4-2, respectively. All other analysis results will be presented in reference to the hypotheses they address. The reader will notice that half (16/30) of the GQ items are highly skewed (i.e., less than -1 or greater than +1; Bulmer, 1979). At the subscale level, Member-Group Positive Bonding, Member-Leader Positive Bonding, and Member-Leader Negative Relationship subscale scores were also highly skewed.

Table 4-1

Group Questionnaire Item Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	6.31	.92	6	-1.77	4.46
2	289	5.83	1.17	5	-.84	.05
3	289	6.35	.99	5	-1.82	3.53
4	290	5.99	1.12	5	-.97	.30
5	289	6.17	1.12	5	-1.57	2.46
6	289	5.52	1.41	6	-.79	-.15
7	290	6.34	1.01	6	-1.86	4.04
8	288	5.98	1.17	6	-1.28	1.63
9	289	5.68	1.30	6	-1.07	1.20
10	289	5.08	1.46	6	-.57	-.03
11	288	5.81	1.26	6	-1.24	1.57
12	287	5.13	1.52	6	-.74	.00

13	289	5.24	1.60	6	-.71	-.30
14	286	4.64	1.66	6	-.35	-.69
15	289	4.84	1.84	6	-.57	-.73
16	290	4.29	1.91	6	-.18	-1.09
17	288	1.72	1.19	6	2.30	5.56
18	289	2.42	1.48	6	1.09	.51
19	290	1.53	1.04	6	2.50	6.88
20	289	2.04	1.22	6	1.31	1.47
21	289	2.79	1.52	6	.71	-.22
22	290	3.22	1.62	6	.45	-.65
23	290	2.12	1.46	6	1.43	1.51
24	290	2.69	1.46	6	.89	.28
25	290	2.48	1.44	6	.88	.06
26	290	5.80	1.10	6	-1.33	2.54
27	290	5.6	1.21	6	-.80	.45
28	290	5.61	1.21	6	-1.00	1.03
29	290	5.86	1.38	6	-1.53	2.09
30	290	6.08	1.11	6	-1.84	4.30

Note. Missing values eliminated pairwise. Items scaled from 1-7.

Table 4-2

Group Questionnaire Subscale Descriptive Statistics

Subscale	N	Mean	SD	Range	Skewness	Kurtosis
Positive Bonding Relationship	287	77.45	10.21	47	-.83	.37
Member-Member	287	23.33	3.93	18	-.80	.05
Member-Leader	289	25.17	3.30	18	-1.62	2.89
Member-Group	290	29.00	4.99	30	-1.43	3.29
Positive Working Relationship	283	40.70	9.93	48	-.58	.23
Member-Member	285	19.14	5.59	24	-.47	-.12
Member-Leader	285	21.59	5.09	24	-.89	.57
Negative Relationship	288	21.04	7.65	39	.78	.37
Member-Member	288	7.67	3.26	15	.70	.02
Member-Leader	288	6.06	2.92	15	1.46	2.38
Member-Group	290	7.29	3.59	16	.99	.37

Note. Missing values eliminated pairwise.

Hypothesis 1

Model fit statistics for the single-level CFA were as follows: $\chi^2(380, N = 290) = 777.89$, $p = .00$; CFI = .93; TLI = .91; RMSEA = .05. Given that three of four indices reflect good fit and the fourth (χ^2) marginally failed (32%), we conclude the measurement model provides acceptable fit, supporting Hypothesis 1. Factor loadings and correlation coefficients for the single-level CFA (Figure 4-1) and within-group multilevel CFA (Figure 4-2) were similar. ICCs for each GQ item are listed in Table 4-3, with the average GQ item ICCs for member-member, member-leader, and member-group structural relationship parameters being .15, .11, and .27, respectively (ranged from .08 to .49).

Table 4-3

Intraclass Correlation (ICC) for GQ Items

GQ Item	ICC	GQ Item	ICC	GQ Item	ICC
1	.13	11	.10	21	.07
2	.19	12	.15	22	.13
3	.13	13	.12	23	.49
4	.15	14	.10	24	.33
5	.09	15	.12	25	.39
6	.17	16	.12	26	.19
7	.14	17	.09	27	.17
8	.18	18	.08	28	.20
9	.08	19	.11	29	.18
10	.13	20	.18	30	.23

Note. ICC = intraclass correlation for groups, or the proportion of variance accounted for by group.

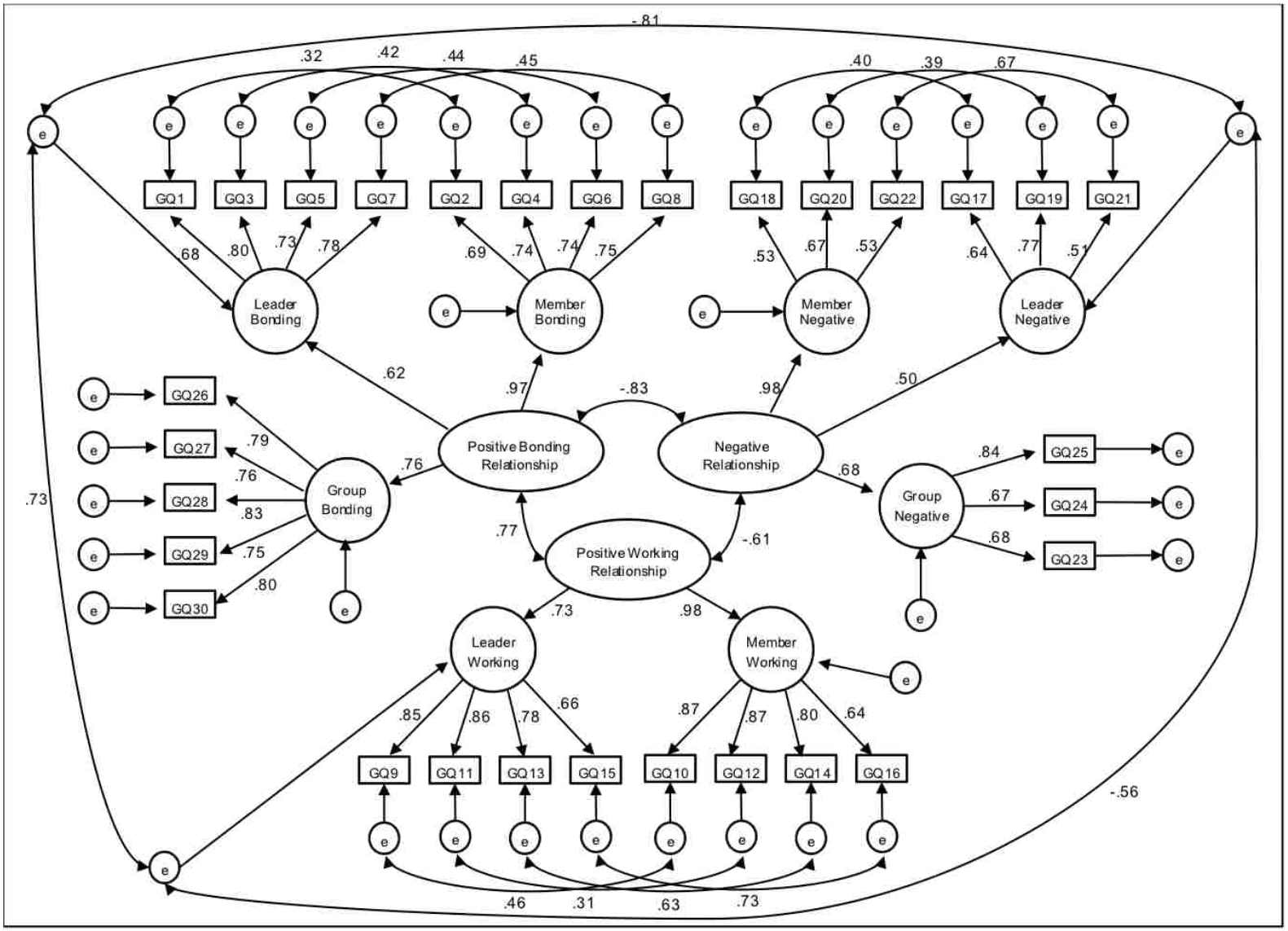


Figure 4-1. Single-level CFA for the GQ. The numbers by the single- and double-headed arrows represent factor loadings and correlations, respectively.

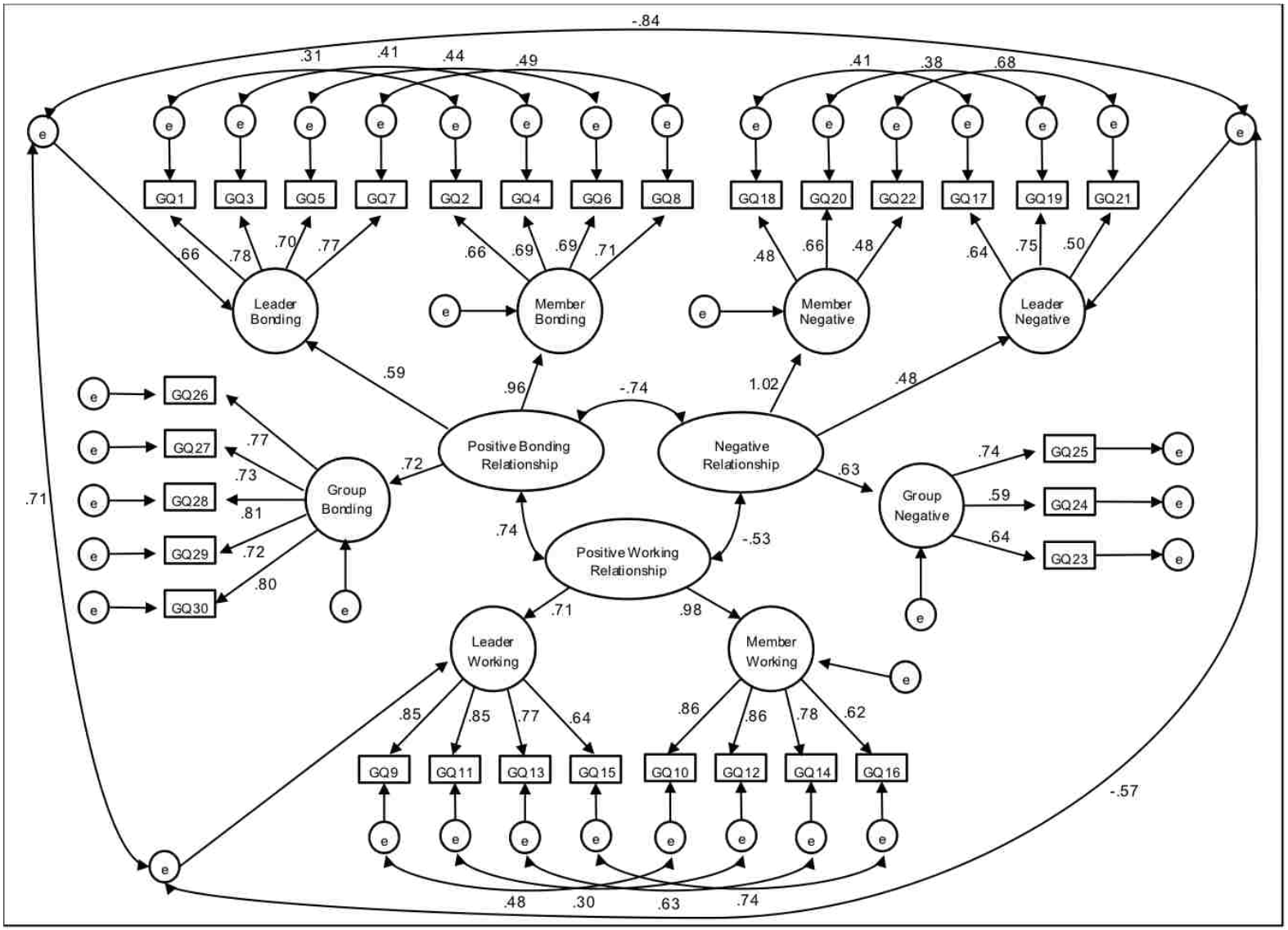


Figure 4-2. Within-group CFA for the GQ. The numbers by the single- and double-headed arrows represent factor loadings and correlations, respectively.

Hypothesis 2

Internal consistency reliability coefficients (Cronbach's alpha) for GQ subscales are presented in Table 4-4. Reliability coefficients indicated acceptable reliability for all three GQ subscales, with the Positive Bonding and Positive Working Relationship subscales meeting criteria for excellent reliability. These values are similar to those reported by Krogel (2009). Because the GQ data collected from the university counseling centers had a restricted range, as counseling center GQ data tend to have (Krogel, 2009), the Ghiselli et al. (1981) formula was employed to estimate what the GQ subscales' reliability coefficients would have been if range was not attenuated. Values from the Utah State Hospital population in Krogel (2009) were used to meet the conditions of the formula. The subsequent estimation yielded increased reliability for all three subscales (see Table 4-4).

Table 4-4

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).

Descriptive statistics for criterion measures are presented in Tables 4-5 through 4-11. Like the GQ, each measure produced highly skewed data for several items. At the subscale level, GCQ: Conflict, ES: Negative, and Member-Leader WAI: Bond were also highly skewed.

Table 4-5

Therapeutic Factors Inventory: Cohesion Subscale Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	5.69	1.12	6	-1.17	1.85
2	290	5.57	1.10	6	-.78	.80
3	290	5.72	1.15	6	-1.07	1.36
4	290	6.14	1.00	6	-2.10	6.59
5	289	5.30	1.38	6	-.68	-.06
6	286	5.85	1.19	5	-.99	.42
7	290	6.06	1.17	6	-1.81	4.15
8	290	5.45	1.24	6	-.86	.57
9	289	5.81	1.27	6	-1.25	1.33
Total Score	284	51.64	7.75	39	-.93	.63

Note. Missing values eliminated pairwise. Items 4 and 7 are reverse scored. Item responses are scaled from 1 to 7.

Table 4-6

Group Climate Questionnaire Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	5.71	1.07	4	-.63	-.09
2	290	5.67	1.10	5	-.61	-.19
3	288	2.44	1.32	6	1.16	1.40
4	290	5.50	1.14	5	-.58	-.06
5	290	3.86	1.45	6	.10	-.62
6	290	1.70	1.03	5	1.63	2.31
7	289	2.36	1.18	5	.95	.66
8	290	3.71	1.79	6	.03	-1.09
9	288	4.17	1.56	6	.07	-.82
10	290	1.35	.66	5	2.54	9.75
11	290	5.54	1.41	6	-.99	.45
12	290	3.18	1.47	6	.50	-.41
Engaged	290	5.23	.86	5	-.34	-.16
Conflict	289	2.15	.79	5	1.07	1.84
Avoidance	287	3.49	.93	6	.18	.16

Note. Missing values eliminated pairwise. Item responses are scaled from 1 to 7.

Table 4-7

Member-Member Empathy Scale Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	5.76	1.19	6	-.97	.59
2	290	5.43	1.41	6	-.85	.28
3	290	5.79	1.17	5	-.88	.25
4	290	5.41	1.63	6	-.75	.27
5	290	5.39	1.33	6	-.67	.06
6	290	1.97	1.10	5	1.35	1.89
7	289	1.65	1.06	6	2.03	4.49
8	290	1.67	1.00	6	1.91	4.63
9	290	2.68	1.24	6	.69	.28
10	290	1.15	.52	4	4.12	19.36
Positive	290	27.78	5.40	28	-.79	.42
Negative	289	9.12	3.47	20	1.41	2.60
Total	287	18.79	7.78	44	-.97	1.27

Note. Missing values eliminated pairwise. Item responses are scaled from 1 to 7. Items 1-5 are positively worded and 6-10 are negatively worded. Positive = positively-worded item total score. Negative = negatively-worded item total score. Total = Empathy scale total score calculated by subtracting negatively-worded item total score from positively-worded item total score.

Table 4-8

Member-Leader Empathy Scale Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	6.08	1.13	6	-1.45	2.08
2	290	5.75	1.33	6	-1.09	.80
3	290	6.02	1.12	6	-1.23	1.39
4	290	5.65	1.21	6	-.99	1.09
5	289	5.75	1.29	6	-.98	.40
6	290	1.54	1.02	6	2.79	9.26
7	289	1.49	.97	5	1.36	5.72
8	289	1.41	.88	6	2.99	10.81
9	290	2.39	1.22	6	.84	.50
10	290	1.10	.45	5	6.70	57.82
Positive	289	29.24	5.15	27	-1.05	.85
Negative	288	7.93	3.10	18	1.96	5.12
Total	285	21.46	7.28	44	-1.45	2.98

Note. Missing values eliminated pairwise. Item responses are scaled from 1 to 7. Items 1-5 are positively worded and 6-10 are negatively worded. Positive = positively-worded item total score. Negative = negatively-worded item total score. Total = Empathy scale total score calculated by subtracting negatively-worded item total score from positively-worded item total score.

Table 4-9

Member-Member Working Alliance Inventory Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	5.22	1.43	6	-.73	.08
2	290	4.75	1.39	6	-.64	.13
3	289	5.94	1.11	6	-1.35	2.16
4	289	5.24	1.16	6	-.81	.66
5	288	5.63	1.21	6	-1.03	1.34
6	289	5.51	1.45	6	-1.08	.64
7	290	5.72	1.22	6	-1.01	.87
8	289	6.20	1.00	6	-1.52	2.97
9	290	5.40	1.57	6	-.96	.06
10	289	5.04	1.40	6	-.60	.11
11	290	4.48	1.69	6	-.45	-.53
12	290	5.43	1.21	6	-.79	.60
13	290	4.84	1.51	6	-.57	-.10
14	290	5.60	1.25	6	-.87	.60
15	290	4.99	1.53	6	-.63	-.44
16	290	5.74	1.88	6	-1.31	1.49
17	290	5.67	1.62	6	-1.16	.44
18	290	4.31	1.69	6	-.34	-.70
19	290	5.16	1.48	6	-.80	.23
20	290	5.59	1.37	6	-1.23	1.37
Goal	289	4.82	1.14	6	-.45	.00
Task	289	5.17	1.13	6	-.78	.71
Bond	285	5.54	.92	5	-.81	.58

Note. Missing values eliminated pairwise. Item responses are scaled from 1 to 7. Items 1, 6, 9, 15, and 17 are reverse scored.Goal, Task, and Bond refer to the WAI's three subscales in the member-member context.

Table 4-10

Member-Leader Working Alliance Inventory Descriptive Statistics

Item	N	M	SD	Range	Skewness	Kurtosis
1	290	5.79	1.54	6	-1.57	1.75
2	290	5.11	1.37	6	-.99	.71
3	289	5.94	1.11	6	-1.35	2.16
4	290	5.50	1.17	5	-.85	.39
5	290	5.97	1.41	6	-1.49	2.51
6	288	5.98	1.26	6	-1.64	2.62
7	290	6.24	.99	5	-1.61	2.86
8	290	6.42	.84	4	-1.61	2.61
9	290	6.07	1.43	6	-1.69	1.90
10	290	5.62	1.35	6	-1.14	1.01
11	290	4.80	1.70	6	-.61	-.45
12	290	5.76	1.26	6	-1.28	1.68
13	290	5.20	1.50	6	-.87	.36
14	290	6.05	1.11	5	-1.43	2.07
15	290	5.22	1.62	6	-.92	-.14
16	289	5.49	1.45	6	-.94	.13
17	289	6.25	1.29	6	-1.90	3.21
18	290	4.67	1.72	6	-.51	-.61
19	290	5.37	1.42	6	-.84	.36
20	290	5.98	1.22	6	-1.62	3.01
Goal	288	5.17	1.17	6	-.69	.13
Task	289	5.41	1.13	6	-.97	.93
Bond	288	5.93	.84	4	-1.20	1.57

Note. Missing values eliminated pairwise. Item responses are scaled from 1 to 7. Items 1, 6, 9, 15, and 17 are reverse scored. Goal, Task, and Bond refer to the WAI's three subscales in the member-member context.

Table 4-11

Working Alliance Inventory Overall Subscale Descriptive Statistics

Subscale	N	M	SD	Range	Skewness	Kurtosis
Goal	288	4.99	1.09	6	-.57	.05
Task	289	5.20	1.14	6	-.83	.73
Bond	283	5.73	.80	4	-.88	.69

Note. Missing values eliminated pairwise. Goal, Task, and Bond refer to the WAI's three subscales.

Internal consistency reliability coefficients for criterion measures are presented in Table 4-12. Reliability coefficients for the Bond, Task, and Goal subscales of the WAI and for the Cohesion subscale of the TFI were comparable to those reported in Johnson et al. (2005). Reliability coefficients for the ES were slightly higher than those reported in Johnson et al. (2005). Reliability coefficients for the Engaged, Conflict, and Avoidance subscales of the GCQ were much lower than those reported in other studies (e.g., .94, .88, and 92, respectively; Kivlighan & Goldfine, 1991), falling closer to those reported in Johnson et al. (.70, .69, and .36, respectively; 2005).

Criterion-related validity coefficients for each GQ subscale are presented in Table 4-13. Positive Working Relationship correlated well with the WAI at both member-member and member-leader levels. All correlations are acceptable by the Reisinger and Burlingame (1997) standards, with several falling in the "excellent" category. The same can be said of Positive Bonding Relationship, with the exception of the GCQ Engaged correlations. While it is still acceptable, the GQ's correlation with this subscale was relatively low (e.g., $r = .56$). Correlations between Negative Relationship and GCQ: Conflict and the ES were also acceptable,

though not quite as high as those of the other two GQ subscales. Given that all correlations were greater than $r = .5$, Hypothesis 2 is supported.

Table 4-12

Criterion Measure Internal Consistency Reliability (Cronbach's Alpha)

Subscale	Overall	Member-Member	Member-Leader
TFI: Cohesion	.89	–	–
GCQ: Engaged	.65	–	–
GCQ: Conflict	.66	–	–
GCQ: Avoidance	.29	–	–
WAI: Bond	.93	.90	.89
WAI: Goal	.84	.69	.72
WAI: Task	.92	.84	.86
ES: Total	.92	.88	.87
ES: Positive	.94	.90	.90
ES: Negative	.81	.72	.69

Note. N = 290. The "–" indicates that this value was not calculated because it was not applicable.

Table 4-13

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.
** $p < .01$.

Hypotheses 3-10

The sociometric hypotheses for the present study were as follows:

3. GQ Positive Working Relationship subscale score valence will correspond with the SQ item 1 score valence at the member-member structural relationship level.
4. GQ Positive Working Relationship subscale score valence will correspond with the SQ item 1 score valence at the member-leader structural relationship level.

5. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-member structural relationship level.
6. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-leader structural relationship level.
7. GQ Positive Bonding Relationship subscale score valence will correspond with the SQ item 2 score valence at the member-group structural relationship level.
8. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-member structural relationship level.
9. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-leader structural relationship level.
10. GQ Negative Relationship subscale score valence will correspond with the SQ item 3 score valence at the member-group structural relationship level.

Quantitative sociometric analysis. Correlations between the SQ and GQ at the member-member and member-leader levels are presented in Table 4-14. All of the SQ1 and SQ2 correlations were positive and statistically significant ($p < .01$). Therefore, Hypotheses 3 through 7 were supported. Hypothesis 8 was unsupported due to statistical non-significance of the correlation between SQ3 and the GQ at the member-member level. SQ3's positive and statistically significant correlation with the GQ at the member-leader level lent support for Hypothesis 9.

Correlations between the SQ and GQ at the member-group level are presented in Tables 4-15 and 4-16. While they both yielded statistically significant correlations ($p < .01$, lending support to Hypotheses 7 and 10), the single code scaling method (see Table 3-2 for scaling methods) yielded higher correlations than the multiple code scaling method for SQ2. The

opposite was true for SQ3. The multiple code scaling methods consistently outperformed the single code methods for SQ3. Therefore, scaling negative mutual choices as -1 (i.e., multiple code) did not improve correlations between the SQ and the GQ Positive Bonding Relationship subscale at the member-group level. However, this scaling method did improve correlations between the SQ and the GQ Negative Relationship subscale. Multiple code scaling method "c" yielded the highest and only statistically significant correlation with Negative Relationship. Multiple code scaling method "c" represented the most inclusive approach to SQ item scoring, including both mutual and incongruous choices in its total score calculation.

Table 4-14

Correlations Between SQ and GQ at the Member-Member and Member-Leader Levels

SQ Item	GQ Subscale and Items				
	PWR Member	GQ 10	GQ 12	GQ 14	GQ 16
SQ1 Member	.23**	.21**	.22**	.19**	.16**
	PWR Leader	GQ 9	GQ 11	GQ 13	GQ 15
SQ1 Leader	.21**	.11	.18**	.19**	.18**
	PBR Member	GQ 2	GQ 4	GQ 6	GQ 8
SQ2 Member	.19**	.12**	.15**	.14**	.22**
	PBR Leader	GQ 1	GQ 3	GQ 5	GQ 7
SQ2 Leader	.30**	.28**	.26**	.21**	.26**
	NR Member	GQ 18	GQ 20	GQ 22	
SQ3 Member	.11	.12	.08	.05	
	NR Leader	GQ 17	GQ 19	GQ 21	
SQ3 Leader	.26**	.19**	.28**	.15*	

Note. PWR = Positive Working Relationship; PBR = Positive Bonding Relationship; NR = Negative Relationship. GQ = Group Questionnaire. SQ = Sociometric Questionnaire.
* $p < .05$. ** $p < .01$.

Table 4-15

Correlations Between SQ and GQ Positive Bonding Relationship at the Member-Group Level

SQ Item	GQ Subscale and Items					
	PBR	GQ 26	GQ 27	GQ 28	GQ 29	GQ 30
SQ2sa	.30**	.30**	.23**	.27**	.17**	.11
SQ2ma	.24**	.28**	.23**	.26**	.14*	.13

Note. PBR = Positive Bonding Relationship total score. GQ = Group Questionnaire. SQ = Sociometric Questionnaire. Refer to Table 3-2 for a description of the scaling methods.
s Single code method for calculating SQ item scores.
m Multiple code method for calculating SQ item scores.
a Scaling method a.
* $p < .05$. ** $p < .01$.

Table 4-16

Correlations Between SQ and GQ Negative Relationship at the Member-Group Level

SQ Item	GQ Subscale and Items			
	NR	GQ 23	GQ 24	GQ 25
SQ3sa	.01	-.02	.02	.04
SQ3ma	.08	.06	.03	.11
SQ3sb	.07	.03	.02	.12
SQ3mb	.10	.07	.04	.14*
SQ3sc	.08	.02	.04	.13*
SQ3mc	.12*	.08	.07	.15*
SQ3sd	-.07	.00	-.12*	-.07
SQ3md	.06	.07	-.01	.07

Note. NR = Negative Relationship total score. GQ = Group Questionnaire. SQ = Sociometric Questionnaire. Refer to Table 3-2 for a description of the scaling methods.
s Single code method for calculating SQ item scores.
m Multiple code method for calculating SQ item scores.

a Scaling method a.
b Scaling method b.
c Scaling method c.
d Scaling method d.
* $p < .05$.

Qualitative sociometric analysis. Using sociograms to investigate the GQ's ability to measure the group therapeutic relationship at member-member, member-leader, and member-group structural levels was intended to be a qualitative exploration. As such, only those groups with the most complete sociometric data (groups 24, 29, 34, 35, 55, 57, 58, and 63) were selected for analysis. An exhaustive analysis of all the sociometric data from each of the 41 selected group members is beyond the scope of the present study. Instead, sociograms were only created for those group members whose raw sociometric data suggested obvious support for or evidence against Hypotheses 3-10. A sociogram was deemed supportive if it contained affirmative or negative mutual choices that corresponded with above or below average GQ scores, respectively. Sociograms that depicted clear preferences of the target member but did not contain any mutual choices were still considered supportive of Hypotheses 3-10 if the incongruous choices matched the valence of the above/below average GQ scores. Sociograms whose configurations suggested the opposite relationship valence indicated by a participant's GQ scores were considered as evidence against Hypotheses 3-10. Some sociograms bore no obvious support for or evidence against Hypotheses 3-10. These sociograms either contained relationship dynamics that could be construed as either support for *or* evidence against Hypotheses 3-10 or a determination required too many assumptions about the participant's reasoning behind GQ item ratings. This was often the case when GQ scores fell in the average range.

Of the 123 possible sociograms of this subsample, 46 (37%) yielded support for and 17 (14%) yielded evidence against Hypotheses 3-10, leaving 60 (49%) "undetermined." The following are a sampling of several sociograms that exemplify these conflicting findings.

The sociogram for Member 3 of Group 29 (Figure 4-3) offered support for Hypotheses 8 and 9. Member 3 produced above average member-leader and member-member Negative Relationship GQ scores. As depicted in the sociogram, these high scores correspond with Member 3's desire to distance himself from the group leaders, with the reciprocated desire for distance between him and Member 5, and with the incongruous choice regarding Member 2.

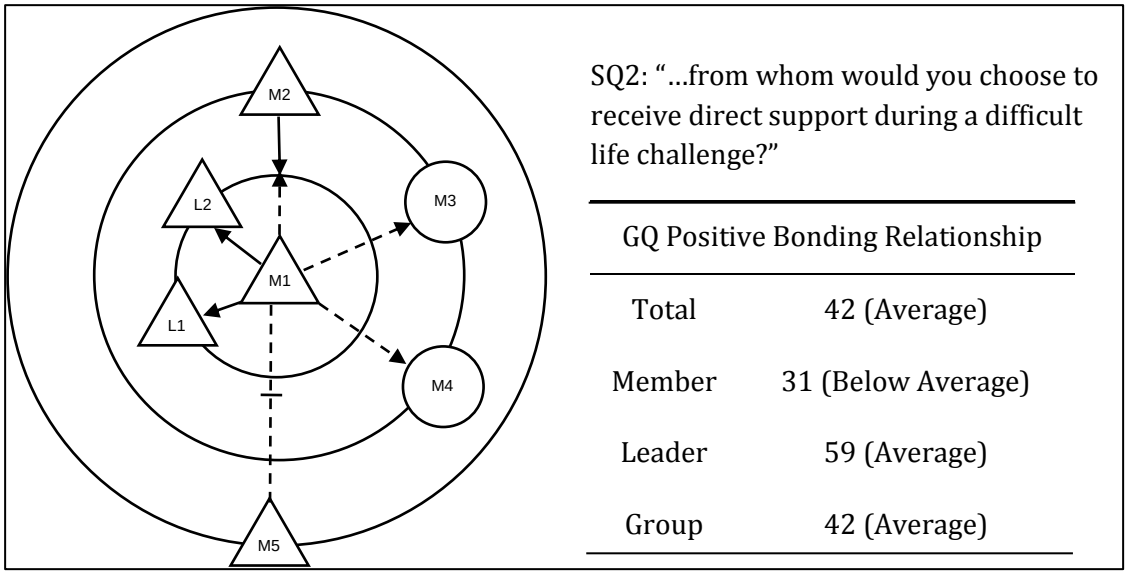
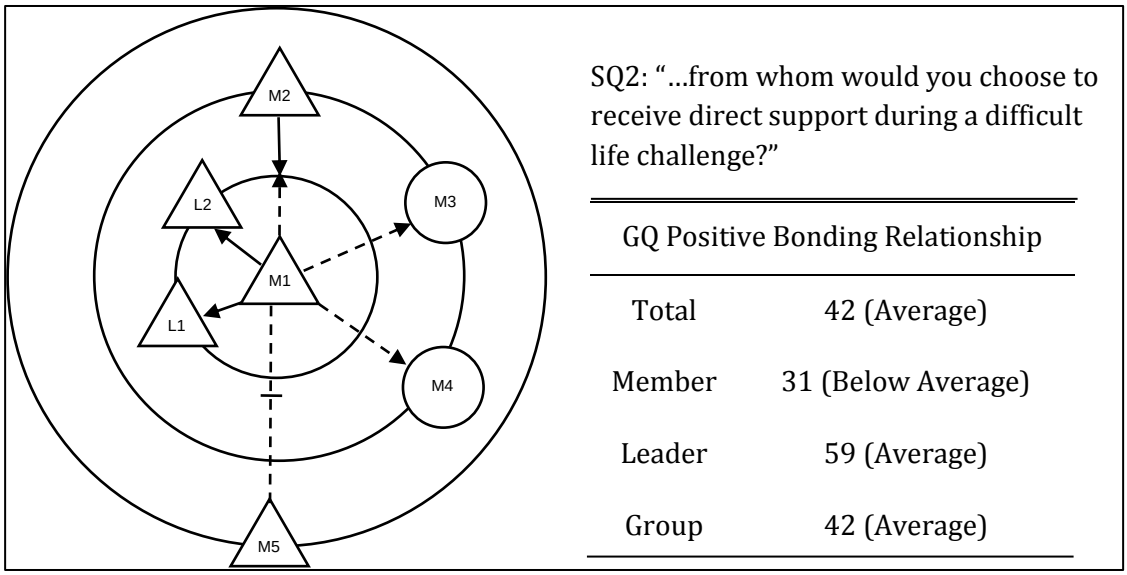


Figure 4-3. SQ3/Negative Relationship sociogram for Member 3 of Group 29. An example of support for Hypotheses 8 and 9.

The sociogram for Member 1 of Group 57 (Figure 4-4) offered support for Hypotheses 5 and 6. Member 1 produced high average member-leader and below average member-member Positive Bonding Relationship GQ scores. Member 1's high-quality bond with the group leaders is evidenced by his close proximity to them on the sociogram. His low-quality bond to other members of the group is made apparent by a negative mutual choice with Member 5 and by negatively valenced incongruities with the other group members. That the distinction between Member 1's good bond with the leaders and poor bond with the members is borne out in both his GQ scores and his sociogram gives credence to the GQ's ability to parcel out and measure the group therapeutic relationship at the structural level.



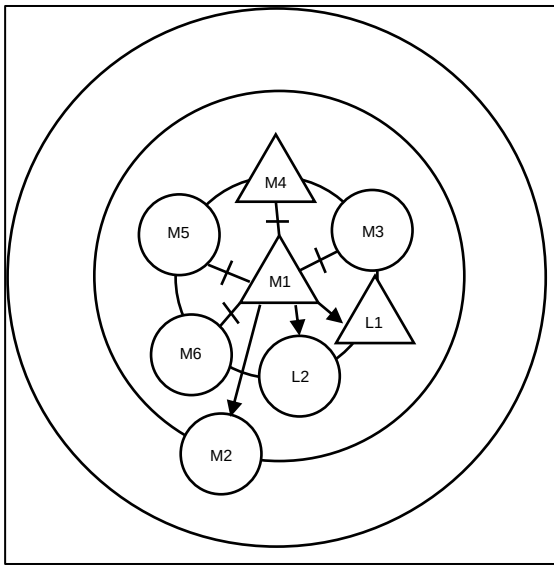
SQ2: "...from whom would you choose to receive direct support during a difficult life challenge?"

GQ Positive Bonding Relationship	
Total	42 (Average)
Member	31 (Below Average)
Leader	59 (Average)
Group	42 (Average)

Figure 4-4. SQ2/Positive Bonding Relationship sociogram for Member 1 of Group 57. An example of support for Hypotheses 5 and 6.

The disparity between the GQ and sociographic results for Member 1 of Group 24 (Figure 4-5) exemplifies the data that offer poor support for the present study's sociometric hypotheses. Member 1 produced a below average member-member Positive Working Relationship GQ score, but his sociogram depicts a high working alliance, with almost all of the other group members packed in closely around him. The only outlier is Member 2, and that incongruous relationship has a positive/neutral valence, not a negative one.

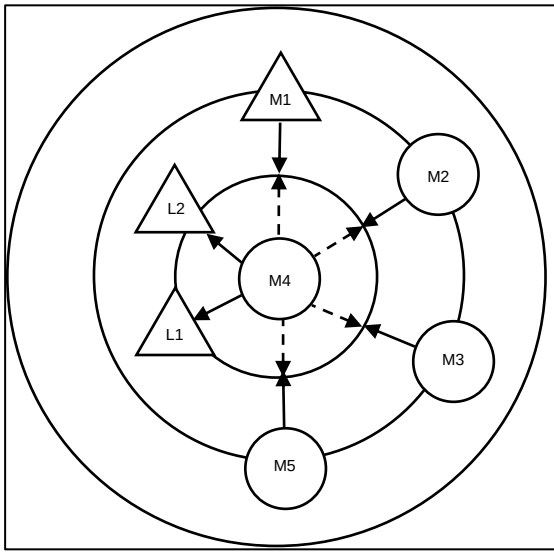
Figure 4-6 shows another example of poor support for the sociometric hypotheses. Member 4 of Group 58 produced an above average member-member Positive Bonding Relationship GQ score, but his sociogram reflects no desire to receive direct support from any of the other group members.



SQ1: "…with whom would you choose to work on an important project?"

GQ Positive Working Relationship	
Total	39 (Below Average)
Member	37 (Below Average)
Leader	50 (Average)

Figure 4-5. SQ1/Positive Working Relationship sociogram for Member 1 of Group 24. An example of data not supporting the sociometric hypotheses.



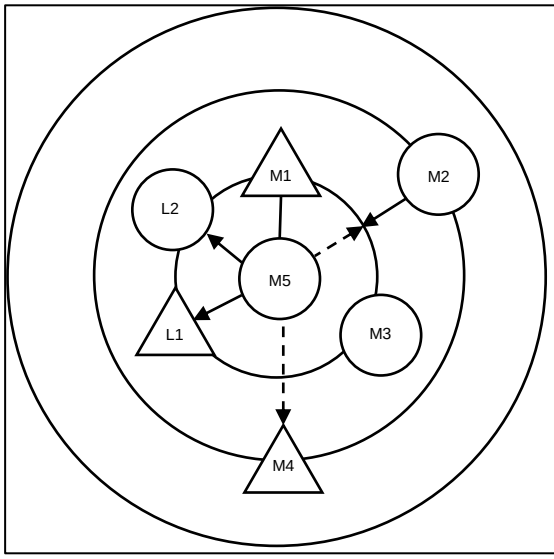
SQ2: "…from whom would you choose to receive direct support during a difficult life challenge?"

GQ Positive Bonding Relationship	
Total	57 (Average)
Member	62 (Above Average)
Leader	59 (Average)
Group	50 (Average)

Figure 4-6. SQ2/Positive Bonding Relationship sociogram for Member 4 of Group 58. An example of data not supporting the sociometric hypotheses.

As noted above, the majority of the sociograms produced by the 41 selected group members exhibited sociographic configurations that bore no obvious support for or evidence against Hypotheses 3-10. Figure 4-7 is an example of one such sociogram. Member 5 of Group 24 produced high average member-member and member-leader Positive Bonding Relationship

GQ scores. Her sociogram depicts a desire for direct support from the leaders and from Members 1 and 3, but not from Member 2 or 4. Does this split configuration correspond with high average member-member positive bonding? This is a difficult question to answer without knowing what the group member was thinking while rating the GQ. It is possible that Member 5 was considering her positive relationship with Members 1 and 3 when rating the GQ. But if she was, what kept her GQ score from climbing into the above average range? By the same token, what kept her GQ score from sinking into the below average range, considering her poor bond with Members 2 and 4? It is possible that Member 5 considered her relationship with all four group members while completing the GQ, each conflicting preference diminishing the power of the other. If this were the case, however, one might expect Member 5's member-group score to fall in the average range also. Instead, it fell in the above average range. Given that comparing sociograms and GQ scores like Member 5's generated more questions than answers, the degree to which such comparisons lent support for or provided evidence against Hypotheses 3-10 was deemed "undetermined."



SQ2: "...from whom would you choose to receive direct support during a difficult life challenge?"

GQ Positive Bonding Relationship	
Total	60 (Average)
Member	54 (Average)
Leader	59 (Average)
Group	62 (Above Average)

Figure 4-7. SQ2/Positive Bonding Relationship sociogram for Member 5 of Group 24. An example of data for which no determination could be made regarding their support for or refutation of Hypotheses 3-10.

Discussion

Since its inception in 2005, Johnson's 3-factor model of the group therapeutic relationship has been empirically supported by 5 separate studies that sampled nearly 2,200 individuals from over 260 groups in 4 different countries. The GQ was created to measure the group therapeutic relationship per the Johnson Model and, as such, has the potential to be a powerful tool in the hands of group clinicians. But has the GQ drifted from the firm foundation upon which it was based? Krogel (2009) noted that when she and her team altered and/or excluded many of the original questionnaire items used to formulate the Johnson Model such a drift might have occurred. In addition, Johnson et al. (2005) observed that group members have difficulty making meaningful distinctions at the level of relationship structure, which would make a self-report questionnaire like the GQ subject to the same difficulties. The present study was designed to address these concerns about the GQ's validity by 1) using GQ response data to replicate the Johnson 3-factor model, 2) correlating GQ items with the original measures used to create the Johnson Model, and 3) conducting a sociometric exploration of the GQ's ability to make relationship structure distinctions.

Model fit statistics for the single-level CFA of GQ response data met standards for acceptable model fit, thereby confirming Hypothesis 1. Not only do these findings support Johnson's 3-factor model as a theoretically sound and empirically robust definition of the group therapeutic relationship, they also suggest that the GQ is a valid measure of that construct. This empirical support responds to Krogel's (2009) concern that the GQ might have drifted from Johnson's original construct formulation.

One issue addressed by Johnson and her colleagues in their original construct formulation, but not accounted for by Krogel (2009) was intragroup dependency and the

corresponding risk of Type I error inflation. Like Johnson et al. (2005), the present study sought to address these concerns with a within-group, multilevel CFA and by examining the ICC for each GQ item. All 30 GQ item ICCs were positive, suggesting that participants within groups produced scores more similar to each other's than to other participants'. In other words, the positive ICCs indicate that at least some of the variance in GQ scores can be accounted for by group-dependent phenomena. This was especially true for the member-group GQ items (23-30), which produced the highest ICCs (from .17 to .49). These strong group effects are to be expected, given that the member-group items are the most group-dependent items by design. Johnson et al. (2005) found similar ICCs in their study (i.e., ranging from .13 to .42). The present study's replication of these findings lends support to the notion that the GQ is sensitive to group-level processes.

It is noteworthy that none of the GQ item ICCs exceeded .5. This suggests that while a notable amount of the variance in GQ scores can be accounted for by group-dependent phenomena, the majority of the variance remains unaccounted for. These unaccounted-for predictors of GQ variance could include various individual, member-dependent, and group-level phenomena. Given how difficult it is to measure an individual's relationship to a nebulous construct like the "group-as-a-whole," it might also be argued that the relatively high member-group ICCs are more than acceptable and indeed impressive; especially since the present findings replicate Johnson's earlier ICCs. However, more research on group-level relationship variables using multilevel models and ICCs is needed in order to make more definitive conclusions about measures designed to assess group dynamic properties. Two studies on a single group relationship measure are insufficient to ascertain normative expectations on acceptable intra-group dependency values. As it stands, model fit statistics for both the within-

group and between-group models indicate acceptable model fit, supporting the GQ's construct validity.

As hypothesized, the GQ subscales correlated well with their corresponding criterion measures at all three structural relationship levels. This finding implies that in spite of concerns expressed by Krogel (2009), the GQ has not drifted off target and is still a valid tool for measuring the group therapeutic relationship. The lowest correlation (r = .54) was between GQ Positive Bonding Relationship and the Engaged subscale of the GCQ. One reason for this lower correlation might be the low reliability of the Engaged subscale. At α = .65, the Engaged subscale had the lowest internal consistency reliability of any subscale correlated with the GQ Positive Bonding Relationship subscale.

Now that strong correlations have been discovered between the GQ and the original measures used to create it, dispelling doubts about construct drift, future research should compare the GQ with other measures of the group therapeutic relationship. Bormann et al. (2011) did this with German measures and found similar promising relationships. Additional English candidates might be the Group Atmosphere Scale (Silbergeld, Koenig, Manderscheid, Meeker, & Hornung, 1975) and the Harvard Group Cohesiveness Scale (Budman et al, 1987).

"Exploration" is the operative word when considering the use of sociometry in the present study. With very little precedent in the extant clinical group literature, using a sociometric test to validate a group process questionnaire like the GQ was a relatively novel approach. Indeed, sociometry was chosen because it illuminates group relationships from a different angle than traditional self-report questionnaires like the GQ. It was hoped that with its explicit depiction of dynamic relationship preferences, sociometry would provide a standard by

which to evaluate the GQ's ability to measure the group therapeutic relationship at the member-member, member-leader, and member-group structural levels.

The first attempt at using sociometric data to evaluate the GQ was quantitative. Member-member and member-leader GQ scores produced statistically significant correlations with SQ scores for nearly every GQ subscale. Member-member Negative Relationship was the only subscale that did not have a statistically significant correlation with the SQ. It is possible that this correlation was attenuated for the same reason already mentioned regarding the relatively low correlation between the GQ Positive Bonding Relationship and GCQ Engaged subscales: low reliability. The member-member Negative relationship subscale has the lowest internal consistency reliability of all the GQ subscales ($\alpha = .61$).

In the spirit of exploration, several different item scaling methods were employed to score the SQ for correlating with the GQ at the member-group level. The highest correlation between the SQ and the member-group Positive Bonding Relationship subscale was achieved by setting all affirmative mutual SQ scores to 1 and the remaining choice configurations to 0 (i.e., single code scaling method a). Setting negative mutual SQ scores to -1 (i.e., multiple code scaling method a) did not improve the correlation. This was a surprising result. The multiple code scaling method was designed under the assumption that while positively valenced affirmative mutual preferences between group members likely contributes to high member-group positive bonding, negative mutual preferences likely reduce it. This would make it important to account for both affirmative and negative mutual preferences. Such an accounting was apparently important for member-group correlations between the SQ and Negative Relationship, with the multiple code scaling method c producing the best correlation coefficient among its sister methods. The implications of these conflicting findings are unclear. There may be some

qualitative difference between positive and negative member-group therapeutic relationships, aside from their valence, that accounts for these differences. Further research is needed to explore this possibility.

It is important to note that while the aforementioned SQ-GQ correlations were statistically significant, none of them were greater than $r = .30$. Because the methods by which sociometric data were quantified were exploratory, they may have failed to offer the most robust numerical representation possible. Alternative quantification methods, such as those outlined in Robert Terry's *Latent Trait Model of Interpersonal Perception* (LaTRIPP; 2000), may offer more detailed statistical representations of sociometric data and should be explored in future research.

The second attempt at using sociometric data to evaluate the GQ was qualitative. It was hoped that comparing quantitative GQ data to graphical representations of group relational preferences (i.e., sociograms) would provide insight into the GQ's ability to measure the quality of the group therapeutic relationship at the heretofore-mentioned structural levels. Many of the insights gleaned from this comparison came in the form questions rather than answers. This was to be expected, given the aforementioned limited precedent for such a comparison and the exploratory intent behind it. As for answering the question regarding the GQ's ability to measure relationship quality at the structural level, the GQ-Sociogram comparisons offered moderate support for this ability, with 37% of the sociograms examined yielding support for the GQ. But this answer begs a very important question. What of the sociograms from which support for the GQ could not be determined? This group represented half of the examined sociograms. If conclusions about support for or evidence against the GQ's validity could be

garnered from these sociograms, the answer to the question about the GQ's measurement ability could change dramatically.

One barrier to getting more information from these sociograms was the lack of norms for what average-level group therapeutic relationships look like sociographically. Above average and below average group therapeutic relationships were more obvious in sociograms—group members were either reaching out to or distancing themselves from other group members, respectively. A mixture of the two preferences or a lack of preference was more difficult to interpret, however. Are such configurations representative of average-level relationship quality? Do they represent more complicated relationship dynamics and are thereby beyond simple "below average, average, above average" categorization? Perhaps the difficult-to-interpret sociographic configurations were an artifact of using the task-specific, peer-nomination, unlimited-choice sociometric criterion questions and could be remedied by using the other methods discussed in the Literature Review. Future research replicating and adjusting the present study's methodology is required to address these concerns.

GQ-sociogram comparisons also revealed limitations in how precisely the GQ measures the group therapeutic relationship. For example, there are myriad reasons a group member might produce an above average Positive Bonding Relationship score on the GQ. One might examine this group member's scores at the structural level in an effort to determine if his high score is attributable to member-member, member-leader, and/or member-group relationships. This, however, is where the GQ's precision stops. Perhaps this group member feels supported by only one or two members, producing a high Positive Bonding score in spite of negative relationships with other group members. Sociograms might offer this level of measurement detail, but the GQ does not.

Limitations/Conclusion

There are several limitations in this study including the lack of clinical diversity in the sample population (i.e., 96% of participants came from college counseling center groups). However, since the findings herein agree with studies using non-clinical process, inpatient therapy, and psychiatric hospital groups (Bormann et al., 2011; Krogel, 2009) it is less likely that homogeneity unduly affected results. The aforementioned restriction of range in GQ responses was also a limiting factor in the present study, attenuating the strength of validity coefficients. Restricted range is a common problem for all psychotherapy process measures that capture positive attributes of the therapist and/or group (Hill & Lambert, 2004).

Its limitations notwithstanding, the present study found that the GQ can provide group clinicians with valid measurements of the quality of their group member's therapeutic relationships at the member-member, member-leader, and member-group structural levels.

References

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press, Inc.

Anshel, M. H. (1994). *Sport psychology: From theory to practice* (2nd ed.). Scottsdale, AZ: Gorsuch Scarisbrick.

Arbuckle, J. L. (2010). Amos (Version 19.0) [Computer Program]. Chicago: SPSS.

Bakali, J. V., Baldwin, S. A., & Lorentzen, S. (2009). Modeling group process constructs at three stages in group psychotherapy. *Psychotherapy Research, 19*(3), 332-343.

Bakali, J. V., Wilberg, T., Hagtvet, K. A., & Lorentzen, S. (2010). Sources accounting for alliance and cohesion at three stages in group psychotherapy: Variance component analyses. *Group Dynamics: Theory, Research, and Practice, 14*(4), 368-383. doi: 10.1037/a0019170

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*(5), 924-935.

Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowitz, J. S., … Watson, J. (2011). Intraclass correlation associated with therapists: Estimates and applications in planning psychotherapy research. *Cognitive Behaviour Therapy, 40*(1), 15-33. doi: 10.1080/16506073.2010.520731

Baldwin, S. A., Stice, E., & Rohde, P. (2008). Statistical analysis of group-administered intervention data: Reanalysis of two randomized trials. *Psychotherapy Research, 18*(4), 365-376.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*(6), 842-852. doi: 10.1037/0022-006x.75.6.842

Barber, J. P., Khalsa, S. R., & Sharpless, B. A. (2010). The validity of the alliance as a predictor of psychotherapy outcome. In J. C. Muran & J. P. Barber (Eds.), *The therapeutic alliance: An evidence-based guide to practice.* (pp. 29-43). New York, NY US: Guilford Press.

Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der 'Helping Alliance Questionnaire' (HAQ) von Luborsky. Möglichkeiten zur Evaluation des therapeutischen Prozesses von stationärer Psychotherapie. *Psychotherapeut, 40*(1), 23-32.

Bednar, R. L., & Kaul, T. J. (1994). Experiential group research: Can the canon fire? In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 631-663). New York: Wiley & Sons.

Beech, A. R., & Hamilton-Giachritsis, C. E. (2005). Relationship between therapeutic climate and treatment outcome in group-based sexual offender treatment programs. *Sexual Abuse: Journal of Research and Treatment, 17*(2), 127-140. doi: 10.1177/107906320501700204

Bloch, S., & Crouch, E. (1985). *Therapeutic factors in group psychotherapy.* New York: Oxford University Press.

Bohart, A. C., & Greenberg, L. S. (1997). *Empathy reconsidered: New directions in psychotherapy*. Washington, DC: American Psychological Association.

Bordin, E. S. (1979). The generalizability of the psychoanayltic concept of the workiing alliance. *Psychotherapy: Theory, Research, and Practice, 16,* 252-260.

Bormann, B., & Straus, 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.

Bormann, B., Burlingame, G. M., & Strauss, B. (2011). Der gruppenfragebogen (GQ-D): Instrument zur messung von therapeutischen beziehungen in der gruppenpsychotherapie. *Psychotherapeut, 56*(4), 297-309. doi: 10.1007/s00278-011-0841-4Breen, M. D. (1994). Applied sociometry. *Journal of Group Psychotherapy, Psychodrama & Sociometry, 47*(2), 52.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Buchanan, D. R. (1982). Psychodrama: A humanistic approach to psychiatric treatment for the elderly. *Hospital and Community Psychiatry, 33*, 220-223.

Budman, S. H., Soldz., S., Demby, A., Feldstein, M., Springer, T., & Davies, M. S. (1989). Cohesion, alliance and outcome in group psychotherapy. *Psychiatry: Journal for the Study of Interpersonal Processes, 53,* 339-350.

Bukowski, W. M., & Hoza, B. (1989). Popularity and friendship: Issues in theory, measurement, and outcome. In T. J. Berndt, & G. W. Ladd (Eds.), *Peer relationships in child development* (pp. 15-45). New York: Wiley.

Bulmer, M. G. (1979). *Principles of statistics*. Mineola, NY: Dover Publications.

Burlingame, G. M. (2010). Small group treatments: Introduction to special section. *Psychotherapy Research, 20*(1), 1-7.

Burlingame, G. M., Fuhriman, A. J., & Johnson, J. E. (2002). Cohesion in group psychotherapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients.* (pp. 71-87). New York, NY US: Oxford University Press.

Burlingame, G. M., Fuhriman, A. J., & Johnson, J. E (2004). Process and Outcome in Group Counseling and Psychotherapy: A Perspective. In J. L. DeLucia-Waack, D. A. Gerrity, C. R. Kalodner & M. T. Riva (Eds.), *Handbook of group counseling and psychotherapy.* (pp. 49-61). Thousand Oaks, CA: Sage Publications Ltd.

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

Burlingame, G. M., Kircher, J. C., & Honts, C. R. (1994). Analysis of variance versus bootstrap procedures for analyzing dependent observations in small group research. *Small Group Research, 25*(4), 486-501. doi: 10.1177/1046496494254004

Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group psychotherapy. In J. C. Norcross (Ed.), *A Guide to Psychotherapy Relationships That Work* (2 ed.): Oxford University Press.

Burlingame, G., Strauss, B. & Joyce, A. (in press). 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: Wiley & Sons.

Burlingame, G. M., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K., Ogrodniczuk, J., & Taylor, S. (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 US: Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.

Castonguay, L. G., Pincus, A. L., Agras, W. S., & Hines, C. E., III. (1998). The role of emotion in group cognitive-behavioral therapy for binge eating disorder: When things have to feel worse before they get better. *Psychotherapy Research, 8*(2), 225-238. doi: 10.1093/ptr/8.2.225.

Chapman, C. (2010). Clinical judgement and prediction in group psychotherapy: Do assessments of group process, member outcome and therapeutic interventions provide useful feedback? Unpublished doctoral dissertaiton. Brigham Young University.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Corcoran, K. J. (1989). Is competitive anxiety an observable behavior?: A sociometric validity study of the SCAT. *Journal of Personality Assessment, 53*(4), 677-684.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi: 10.1007/bf02310555

Dion, K. L. (2000). Group cohesion: From "field of forces" to multidimensional construct. *Group Dyamics: Theory, Research, and Practice, 4*, 7-26.

Festinger, L., Schacter, S., & Back, K. (1950). *Social pressures in informal groups.* New York: Harper & Row.

Frank, J. D. (1957). Some determinants, manifestations, and effects of cohesiveness in therapy groups. *International Journal of Group Psychotherapy, 7*, 53-63.

Fuchs, T., Sidiropoulou, E., Vennen, D., Fisseni, H.-J. (2003). *BFTB. Bonner Fragebogen für Therapie und Beratung*. Göttingen: Hogrefe.

Fuhriman, A., & Burlingame, G. M. (1990). Consistency of matter: a comparative analysis of individual and group process variables. *Counseling Psychologist, 18*, 6-63.

Fuhriman, A., & Burlingame, G. M. (1999). Does group psychotherapy work? In J. R. Price, D. R. Hescheles & A. R. Price (Eds.), *A guide to starting psychotherapy groups.* (pp. 81-98). San Diego, CA US: Academic Press.

Gazda, G. (1982). *Basic approaches to group psychotherapy and group counseling*. Springfield, IL: Thomas.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Fransisco: W. H. Freeman and Company.

Greenson, R. R. (1967). The working alliance and the transference neuroses. *Psychoanalytic Quarterly, 34,* 155-181.

Hale, A. E. (1985). *Conducting Clinical Sociometric Explorations: A Manual for Psychodramatists and Sociometrists*. Roanoke VA: Royal Pub.

Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

Hinshaw, S. P., & Melnick, S. M. (1995). Peer relationships in boys with attention-deficit hyperactivity disorder with and without comorbid aggression. *Development and Psychopathology, 7*, 627-647.

Holland, P. W., & Leinhardt, S. (1973). The structural implications of measurement error in sociometry. *Journal of Mathematical Sociology, 3,* 85-111.

Hornsey, M. J., Dwyer, L., Oei, T. P. S. (2007). Beyond cohesiveness: Reconceptualizing the link between group processes and outcomes in group psychotherapy. *Small Group Research, 38*, 567-592.

Hornsey, M. J., Dwyer, L., Oei, T. P. S., & Dingle, G. A. (2009). Group Processes and Outcomes in Group Psychotherapy: Is It Time to Let Go of "Cohesiveness"? *International Journal of Group Psychotherapy, 59*(2), 267-278.

Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In Horvath, A. O., & Greenberg, L. S. (Eds.), *The working alliance: Theory, research, and practice* (pp. 109-130). New York: John Wiley & Sons, Inc.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223-233.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications.* (pp. 76-99). Thousand Oaks, CA US: Sage Publications, Inc.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hurley, J. R., & Brooks, L. J. (1987). Group climate's principal dimension: Affiliation. *International Journal of Group Psychotherapy, 37*, 441-448.

Imel, Z., Baldwin, S., Bonus, K., & MacCoon, D. (2008). Beyond the individual: Group effects in mindfulness-based stress reduction. *Psychotherapy Research, 18*(6), 735-742. doi: 10.1080/10503300802326038

Johnson, D. P., Penn, D. L., Bauer, D. J., Meyer, P., & Evans, E. (2008). Predictors of the therapeutic alliance in group therapy for individuals with treatment-resistant auditory hallucinations. *British Journal of Clinical Psychology, 47*(2), 171-183. doi: 10.1348/014466507x241604.

Johnson, J. E. (2004). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Dissertation Abstracts International: Section B. Sciences and Engineering, 65*(1-B), 440.

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group Climate, Cohesion, Alliance, and Empathy in Group Psychotherapy: Multilevel Structural Equation Models. *Journal of Counseling Psychology, 52*(3), 310-321.

Joyce, A. S., Piper, W. E., & Ogrodniczuk, J. S. (2007). Therapeutic alliance and cohesion variables as predictors of outcome in short-term group psychotherapy. *International Journal of Group Psychotherapy, 57*(3), 269-296. doi: 10.1521/ijgp.2007.57.3.269

Kay, S. (1996). *Effects of gender on small group cohesion: MacKenzie's Group Climate Questionnaire analyzed.* Unpublished doctoral dissertation, Brigham Young University.

Kivlighan, D. M., & Goldfine, D. C. (1991). Endorsement of therapeutic factors as a function of stage of group development and participant interpersonal attitudes. *Journal of Counseling Psychology, 38,* 150-158.

Kleinberg, J. L. (1999). The supervisory alliance and the training of psychodynamic group psychotherapists. *International Journal of Group Psychotherapy, 49*(2), 159-179.

Krogel, J. (2009). The Group Questionnaire: A new measure of the gorup relationship. Unpublished doctoral dissertation. Brigham Young Universtiy.

Kumar, V. K., & Treadwell, T. W. (2006). Introduction to the Special Issue on Using Group Action Techniques With Children and Adolescents. *Journal of Group Psychotherapy, Psychodrama & Sociometry, 58*(4), 155-156. doi: 10.3200/jgpp.58.4.155-156.

Lambert, M. J. (Ed.). (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed.). New York: John Wiley & Sons, Inc.

Lee, T. (1991). The sociodramatist and sociometrist in the primary school. *Journal of Group Psychotherapy, Psychodrama, & Sociometry, 44,* 191-196.

Lese, K. P., & MacNair-Semands, R. R. (2000). The Therapeutic Factors Inventory: Development of a scale. *Group, 24*(4), 303-317.

Livingston, M. S. (1999). Vulnerability, tenderness, and the experience of selfobject relationship: A self psychological view of deepening curative process in group psychotherapy. *International Journal of Group Psychotherapy, 49*(1), 19-40.

Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Cleghorn (Ed.), *Successful Psychotherapy* (pp. 92-116). New York: Brunner/Maze.

MacKenzie, K. R. (1983). The clinical application of a group climate measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159-170). New York: International Universities Press.

MacKenzie, K. R. (1995). *Effective use of group therapy in managed care*. Washington, DC US: American Psychiatric Association.

MacKenzie, K. R., & Tschuschke, V. (1993). Relatedness, group work, and outcome in long-term inpatient psychotherapy groups. *Journal of Psychotherapy Practice and Research, 2*, 147-156.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438-450. doi: 10.1037/0022-006x.68.3.438

Martens, R. (1977). Sport cometetion anxiety test. Champaign, IL: Human Kinetics.

Marziali, E., Munroe-Blum, H., & McCleary, L. (1997). The contribution of group cohesion and group alliance to the outcome of group psychotherapy. *International Journal of Group Psychotherapy, 47*(4), 475-497.

Marziali, E., Munroe-Blum, H., & McCleary, L. (1999). The effects of the therapeutic alliance on the outcomes of individual and group psychotherapy with borderline personality disorder. *Psychotherapy Research, 9*(4), 424-436. doi: 10.1093/ptr/9.4.424

McClendon, D. T., & Burlingame, G. M. (2011). Group climate: Construct in search of clarity. In R. Coyne (Ed.), *Oxford Handbook of Group Counseling*: Oxford University Press.

McClendon, D. T., & Burlingame, G. M. (2011). Has the magic of psychotherapy diappeared? Integrating evidence-based practic into therapist awareness and development. In R. H. Klein, H. S. Bernard, & V. L. Schermer (Eds.), *On becoming a psychotherapist: The*

personal and professional journey (pp. 190-211). New York, NY: Oxford University Press.

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.

Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama* (2nd ed.). Oxford England: Beacon House.

Mouton, J., Blake R., & Fruchter, B. (1960). The reliability of sociometric measures. In J. L. Moreno (Ed.), *The sociometry reader* (pp. 320-361). Glencoe, IL: Free Press.

Muthen, L., & Muthen, B. (2010). Mplus (Version 6.1) [Computer Program]. Los Angeles: Muthen & Muthen.

Northway, M. L. (1967). *A primer of sociometry* (2nd ed.). Toronto, Canada: University of Toronto Press.

Pareek, U., & Singh, Y. P. (1968). Sociometry and communication network in an Indian village. *International Journal of Psychology, 3*, 157-165.

Piper, W. E., Marrache, M., Lacroix, R., Richardsen, A. M., & Jones, B. D. (1983). Cohesion as a basic bond in groups. *Human Relations, 36*, 93–108.

Piper, W. E., & Ogrodniczuk, J. S. (2010). The therapeutic alliance in group therapy. In J. C. Muran & J. P. Barber (Eds.), *The therapeutic alliance: An evidence-based guide to practice.* (pp. 263-282). New York, NY US: Guilford Press.

Price, J. M., & Ladd, G. W. (1986). Assessment of children's friendships: Implications for social competence and social adjustment. In R. J. Prinz (Ed.), *Advances in behavioral assessment of children and families* (Vol. 2, pp. 121-150). Greenwich, CT: JAI Press.

Rachman, A. W. (1981). Humanistic analysis in groups. *Psychotherapy: Theory, Research & Practice, 18*(4), 457-477. doi: 10.1037/h0088399

Reisinger, C. W., & Burlingame, G. M. (1997). How to evaluate a mental health outcome measure. In K. M. Coughlin (Ed.), *The 1997 behavioral outcomes and guidelines sourcebook: A progress report and resource guide on mental health and substance abuse treatment* (pp. 370-374). New York, NY: Faulkner/Gray.

Remer, R. (1995). Using strong sociometry: Some guidelines and techniques. *Journal of Group Psychotherapy, Psychodrama & Sociometry, 48*(2), 79.

Rogers, C. R., Gendlin, G. T., Kiesler, D. V., & Truax, L. B. (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics*. Madison, Wisconsin, University of Wisconsin Press, 1967.

Schain, J. (1989). The new infant research: Some implications for group therapy. *Group, 13*, 112-121.

Shapiro, E. (1991). Empathy and safety in group: A self psychology perspective. *Group, 15*(4), 219-224. doi: 10.1007/bf01547355

Shields, W. (2000). Hope and the inclination to be troublesome: Winnicott and the treatment of character disorder in group therapy. *International Journal of Group Psychotherapy, 50(*1), 87-103. Retrieved February 12, 2011, from Research Library. (Document ID: 47759349).

Shoobs, N. E. (1947). Sociometry in the class room. *Sociometry, 10*, 154-164. doi: 10.2307/2785331

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.

Sterba, R. (1934). The fate of the ego in analytic therapy. *The International Journal of Psychoanalysis, 15*, 117-126.

Stone, W. N. & Gustafson, J. P. (1982). Technique in group psychotherapy of narcissistic and borderline patients. *International Journal of Group Psychotherapy, 32*(1), 29-47.

Stone, W. N., & Whitman, R. M. (1977). Contributions of the psychology of the self to group process and group therapy. *International Journal of Group Psychotherapy, 27*(3), 343-359.

Strauss, B. & Eckert, J. (1994). Dimensionen des Gruppenerlebens: Zur Skalenbildung des Gruppenerfahrungsbogens. *Zeitschrift für Klinische Psychologie, 23*, 188-201.

Terry, R. (2000). Recent advances in measurement theory and the use of sociometric techniques. *New Directions for Child & Adolescent Development, 88*, 27-53.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(3), 207-210.

Trad, P. V. (1993). Using the prospective approach as an adjunct to established models of group psychotherapy. *Group, 17*(1), 43-60.

Treadwell, T. W., Kumar, V. K., Stein, S. A., & Prosnick, K. (1998). Sociometry: Tools for research and practice. *International Journal of Action Methods, 51*(1), 23.

Tschuschke, V. (1996). Der Stuttgarter Bogen. In B. Strauß, Eckert, J., Tschuschke, V. (Ed.), *Methoden der empirischen Gruppenpsychotherapieforschung* (pp. 218-228). Opladen: Westdeutscher Verlag.

Tuckman, B. W., & Jensen, M. A. (1977). Stages of small-group development revisited. *Group & Organization Studies, 2*(4), 419-427.

Veeraraghavan, V., Kellar, H., Gawlick, M., & Morein, N. (1996). *A scale to measure group cohesion.* Unpublished manuscript, West Chester University, West Chester, PA.

Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage Publications.

Wood, D., Kumar, V. K., Treadwell, T. W., & Leach, E. (1998). Perceived cohesiveness and sociometric choice in ongoing groups. *International Journal of Action Methods, 51*(3), 122-137.

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy (5th ed.)*. New York, NY US: Basic Books.

Zetzel, E. (1956). Current concepts of transference. *International Journal of Psycho-Analysis, 37*, 369-376.

Appendix

The Validity of the Group Questionnaire:

Construct Clarity or Construct Drift?

Stephen D. Thayer and Gary M. Burlingame

Brigham Young University

Author Note

Stephen D. Thayer, Department of Psychology, Brigham Young University; Gary M. Burlingame, Department of Psychology, Brigham Young University.

Special thanks to Nick Patterson for his tireless assistance and enthusiasm.

Correspondence concerning this article should be addressed to Stephen D. Thayer, TLRB, Brigham Young University, Provo, UT 84602, stephen.thayer@gmail.com.

Abstract

The Group Questionnaire (GQ) is a recently developed measure of the quality of the therapeutic relationship in group treatment. Its 3 subscales—Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship—are taken from the 3-factor conceptualization of the group therapeutic relationship originally proposed by Johnson and her colleagues (2005). The GQ is designed to measure these 3 subscales/factors at member-member, member-leader, and member-group relationship levels. The purpose of the present study was to estimate the GQ's construct and criterion-related validity by 1) replicating the Johnson et al. (2005) factor structure with a similar sample and by 2) correlating the GQ with the measures from which is was derived (i.e., Working Alliance Inventory, Burns Empathy Scale, Therapeutic Factors Inventory, Group Climate Questionnaire). Two hundred and ninety participants were recruited from 65 treatment groups at 4 university counseling centers and 1 community mental health clinic. Confirmatory factor analysis (CFA) fit indexes from both single- and multiple-level analyses met standards for acceptable model fit. Intraclass correlation coefficients (ICC) suggested the GQ is sensitive to group level processes. Therefore, the Johnson et al. (2005) 3-factor model was successfully replicated and the GQ's construct validity supported. Pearson product-moment (r) and Spearman's rank (ρ) correlation coefficients between the GQ and the aforementioned criterion measures were sufficiently high to lend support for the GQ's criterion-related validity. The present study's findings suggest the GQ is an empirically valid, clinically useful measure of the quality of the group therapeutic relationship.

Key words: group, therapeutic, relationship, validity, questionnaire.

The Validity of the Group Questionnaire: Construct Clarity or Construct Drift?

In the face of increasing demands for inexpensive and accessible forms of mental health treatment, group psychotherapy has emerged as a cost-effective intervention that is comparable to individual psychotherapy (G. M. Burlingame, et al., 2003; G. M. Burlingame, et al., 2004; Fuhriman & Burlingame, 1999; K. Roy MacKenzie, 1995; McRoberts, et al., 1998). But what makes group psychotherapy effective? In his seminal text on group psychotherapy (Yalom & Leszcz, 2005) Irvin Yalom declared:

> How does group therapy help clients? A naïve question, perhaps. But if we can answer it with some measure of precision and certainty, we will have at our disposal a central organizing principle with which to approach the most vexing and controversial problems of psychotherapy. (p. 1)

It is tempting to answer Yalom's "naïve question" with an equally naïve answer: Group therapy helps clients by effecting cognitive, behavioral, and emotional *changes.* This answer begs the question: What are the mechanisms by which group therapy facilitates such changes? Theory-specific treatment components that are common to individual therapy (e.g., cognitive restructuring, exposure, and unconditional positive regard) account for some of the change mechanisms active in group therapy, but a recent review of the last decade of group research noted that a large portion of patient improvement is not explained by theory-specific components (Burlingame et al., in press). There are therapeutic features unique to group therapy that transcend specific change theories (Fuhriman & Burlingame, 1990). One such feature, the group therapeutic relationship, has been identified as one of the most significant general mechanisms of change in group psychotherapy (Johnson, et al., 2005). Group relationships that both support and challenge group members have been consistently linked to positive treatment outcome and

low dropout rates (Burlingame, Fuhriman, & Johnson, 2002, 2004; Burlingame, McClendon, & Alonso, 2011; Castonguay, Pincus, Agras, & Hines, 1998; Marziali, Munroe-Blum, & McCleary, 1997, 1999).

While researchers like those cited above have succeeded in drawing an empirical connection between group therapeutic relationships and treatment outcome, the strength of their results is diminished by the nebulous nature of the construct they study. The term "therapeutic relationship" can mean several different things in the group therapy literature. For example, when Marziali et al. (1997) examined the link between the group therapeutic relationship and outcome in groups for borderline personality disorder, they focused on *therapeutic alliance* and *cohesion.* But when Beech and Hamilton-Giachritsis (2005) examined the same link in group-based sexual offender treatment programs, they were focused on the *therapeutic climate.* Without a comprehensive, standardized understanding of the group therapeutic relationship group researchers have had difficulty measuring this construct and demonstrating a clear relationship between it and other process/outcome variables. This confusion and disagreement about how to define the group therapeutic relationship has distilled in to two disparate definitions in the group therapy literature.

One of these definitions focuses on the interpersonal structure of relationships within groups. More precisely, relationships occur between members and leaders (member-leader), between members and other members (member-member), and between members and the group-as-a-whole (member-group). A review of the literature on the group therapeutic relationship found that two-thirds of relevant studies focused on the member-group relationship, with far fewer studies focusing on the other two structural parameters (G. M. Burlingame, et al., 2002).

This imbalance has complicated efforts to determine if these three structural parameters are theoretically and empirically distinct.

The other definition of the group therapeutic relationship focuses on the content of relationship constructs or the *quality* of the therapeutic relationship. Burlingame et al. (2002) identified four core group therapeutic relationship constructs that have been linked to group process and patient improvement: cohesion, alliance, climate, and empathy. Cohesion represents the esprit de corps or sense of belonging within a group and is often considered to be the essence of the group therapeutic relationship. Alliance within a group represents a fond working relationship between group members and facilitators. Group climate refers to the perceived tone of the group interaction and addresses experiences of safety, closeness, withdrawal, and conflict. Empathy among group members/facilitators represents a sense of being understood. In their seminal study, Johnson et al. (2005) outlined the variability with which these therapeutic relationship constructs are operationalized and measured and explained that without understanding the similarities and differences among these therapeutic relationship constructs, it is difficult to draw sound conclusions about the nature and function of the therapeutic relationship from the group literature.

In an effort to address this difficulty, Johnson et al. (2005) consolidated the aforementioned disparate views and provided an empirically refined, parsimonious representation of the therapeutic relationship in group treatment. This study, to date, contains the largest number of groups and group members in the published process literature—over 100 groups and nearly 700 group members participated nationally. Members from university counseling center groups and American Group Psychotherapy Association (AGPA) training groups responded to select items from four gold standard group process measures: Working

Alliance Inventory (WAI; Horvath & Greenberg, 1989); Group Climate Questionnaire (GCQ; MacKenzie, 1983); Therapeutic Factors Inventory (TFI; Lese & MacNair-Semands, 2000); and Empathy Scale (ES; Burns & Auerbach, 1996). From these data, Johnson et al. (2005) created three higher order relationship factors using exploratory factor analysis: Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship. Johnson and her colleagues empirically demonstrated that these higher order factors accounted for the overlap between the various group therapeutic relationship quality constructs (i.e., cohesion, alliance, empathy, and climate) and adequately described how these constructs are manifest across the three structural relationship units (i.e., member-member, member-group, and member-leader). In 2006 this consolidated definition of the group therapeutic relationship was incorporated by an international task force sponsored by AGPA to organize relationship measures to be recommended to group clinicians (G. Burlingame, et al., 2006).

In 2007, Bormann and Strauss found that Johnson's revised definition of the group therapeutic relationship (hereafter referred to as the "Johnson Model") provided the same construct clarity when empirically tested on 453 patients in 67 Swiss and German inpatient therapy groups. Johnson's results were again successfully replicated again with 424 patients in Norwegian short- and long-term analytic groups (Bakali, et al., 2009). Finally, Krogel (2009) took the preceding research a step further. Using a subset of the questionnaire items Johnson et al. (2005) used in their study, Krogel developed an empirically refined, short, and conceptually "tight" measure of the group therapeutic relationship based on the Johnson Model. Dubbed the Group Questionnaire (GQ), this tool was designed to measure the quality of the therapeutic relationship in groups (i.e., Johnson's Positive Bonding, Positive Working, and Negative Relationship factors) across the three structural parameters of the group therapeutic relationship.

The data collected with the GQ were shown to reproduce Johnson's 3-factor structure yet again, only this time with half the original items. The GQ has since been validated using several other group process measures with a population of 424 group members from nine psychiatric hospitals in Germany (Bormann & Strauss, 2009).

Purpose of Study

Now that 7 years of research has culminated in a refined definition of the group therapeutic relationship construct and the development of what appears to be a practical method for measuring that construct, it is time to reexamine what has been developed from a psychometric perspective. The multiple replications of the Johnson Model among a clinically diverse international population clarified the definition of the therapeutic relationship in group treatment. This clarity allowed for the creation of the GQ. Having been validated in a German population with German criterion measures (Bormann et al., 2011), validation in a US population using the original criterion measures is prudent to provide the final step in establishing the GQ's criterion-related validity. The goal of the present study was to meet this need. Based on the aforementioned research and observations, the following hypotheses were generated:

1. The Johnson et al. (2005) 3-factor structure will be successfully replicated in a US population using the GQ, establishing the GQ's construct validity.
2. The Group Questionnaire will be found to have significant correlations with the measures that it was derived from, establishing its criterion-related validity.

Method

Instruments

The GQ is a 30-item self-report questionnaire designed to measure the quality of the therapeutic relationship in group treatment. Items are scored on a 7-point Likert scale ranging

from *not at all true* (1) to *very true* (7). Item scores are compiled to produce three subscale scores: Positive Bonding Relationship, Positive Working Relationship, and Negative Relationship. The reliability estimates (Cronbach's alpha) of the three subscales are .92, .90, and .80, respectively (Krogel, 2009).

The Working Alliance Inventory (WAI; (Horvath & Greenberg, 1989) is a 20-item self-report questionnaire designed to capture three aspects of the working alliance between a client and a clinician. The Task subscale attempts to measure the extent to which the client perceives therapeutic activities as relevant and beneficial. The Goals subscale attempts to measure the extent to which the client believes that he/she and the clinician are united on the desired treatment outcome. The Bonds subscale attempts to measure the extent to which a client perceives mutual trust, acceptance, and confidence in the attachment between the client and clinician. In keeping with the procedure adopted by Johnson, et al. (2005), the current study will use all the items from the Bonds subscale and shortened versions of the Task and Goals subscales (Tracey & Kokotovic, 1989). These 20 items will be asked twice, once in reference to the working alliance between the group member and the leader and once in reference to other group members. Modifying the WAI for use in groups has some precedent in the group literature (e.g., Johnson et al., 2008). Items are scored on a 7-point Likert scale ranging from *not at all true* (1) to *very true* (7). Reliability estimates for the various subscales range from .56 to .90 (Johnson et al., 2005).

The Group Climate Questionnaire—Short Form (GCQ; (K. R. MacKenzie, 1983) is a self-report measure that purports to assess individual group members' perceptions of the closeness, withdrawal, and conflict within the group. The GCQ contains 12 items rated on a 7-point Likert scale, ranging from *not at all* (0) to *extremely* (6). Its subscales include

Engagement, Avoidance, and Conflict. Construct validity of the GCQ has been found to be adequate (Johnson et al., 2005). Cronbach's alphas for the GCQ subscales in Johnson et al. (2005) are as follows: .70 for Engagement, .36 for Avoidance, and .69 for Conflict.

Therapeutic Factors Inventory (TFI; (Lese & MacNair-Semands, 2000) is a self-report measure designed to assess group members' perceptions of the degree to which the therapeutic factors described by Yalom (1995) are present. The nine items from the Cohesion subscale will be used. Each item is scored on a 7-point Likert scale that ranges from *strongly disagree* (1) to *strongly agree* (7). The Cohesion subscale has been used in a counseling center population, with a test–retest reliability of .93 over a 1-week period (Lese & MacNair-Semands, 2000). Cronbach's alpha in this sample was .90.

The Empathy Scale (ES; (Burns & Auerbach, 1996) is a self-report questionnaire that contains 10 items that reflect clients' perceptions of the therapist's warmth, genuineness, and caring during the most recent therapy session. Five items are worded so that agreement indicates a positive therapeutic relationship and five are worded so that agreement indicates a negative relationship. For this study, responses will be scored on a 7-point Likert scale, with response options ranging from *not at all* (0) to *extremely* (6). Like the WAI items, ES items will be asked twice, once referring to the group leader and again referring to the group members. Coefficient alpha in Johnson et al. (2005) was .83 for the both the Member–Member and the Member–Leader ES. For a summary of how all the measures are represented in the GQ see Table 1.

Participants

Two hundred and ninety participants were recruited from 65 treatment groups: 12 (4.1%) participants from the Brigham Young University Comprehensive Clinic in Provo, Utah; 193 (66.6%) participants from the Brigham Young University Counseling and Career Center; 24

(8.3%) participants from Southern Utah University Counseling and Psychological Services in Cedar City, Utah; 33 (11.4%) participants from Utah State University Counseling and Psychological Services in Logan, Utah; and 28 (9.7%) participants from University of Utah Counseling Center in Salt Lake City, Utah. The sample included 148 (51%) males and 142 (49%) females. Two hundred and sixty-one (90%) participants identified themselves as Caucasian, 7 (2.4%) as Chicano/Latino, 6 (2.1%) as Asian/Pacific Islander, 1 (0.3%) as African American, and 15 (5.2%) as Other. The mean age of participants was 24 (SD = 6.04, range = 17-58). The mean number of sessions attended by participants at the time of data collection was 9.63 (mode = 5, SD = 13.8, range = 1-150). Participants were sought from a variety of group types including general process, disorder specific, and psychoeducational groups. Groups were led by licensed psychologists, predoctoral interns, social workers, and marriage and family therapists whose theoretical orientations varied.

Procedures

Data collection occured one time for each group during the productive working phase of group development when intimacy, engagement, and cohesion peak and questions about relationship quality become relevant (Yalom & Leszcz, 2005). There are various theories about how group relationships develop over the course of treatment (e.g., Tuckman & Jensen, 1977). While such theories can help group leaders formulate expectations and time interventions, they are heuristics that ultimately fail to account for the complexity and unpredictability of group development (Yalom & Leszcz, 2005). Therefore, it is reasonable to estimate that if data are collected after the group has had several sessions, such data will represent the working phase. Group members were asked how many sessions they had attended in order to provide a mean and range.

Participants were recruited via a short recruitment script read by their group leader at the end of a session. Participating members signed an informed consent document and were invited to complete the measures after their session was over. Completing the entire assessment battery generally took participants between 20 and 40 minutes. Participants were compensated with a $20 Amazon.com gift card immediately after completing the assessment battery. Due to the rapid depletion of study support funds, compensation was reduced to $10 for 58 participants. The reduction in compensation did not appear to affect participation rates. As a courtesy to participating groups, each group leader was given the GQ subscales scores of the participating members from their groups.

Analysis

Hypothesis 1: Does the GQ support the Johnson Model? Confirmatory factor analysis was deemed appropriate for three reasons: 1) The GQ was designed to measure an unobservable construct (i.e., the group therapeutic relationship as defined by the Johnson Model) using observable variables (i.e., the 30 questionnaire items). 2) Krogel (2009) and Bormann and Strauss (2009) used CFA in their evaluation of the GQ's construct validity, providing a theoretically based and empirically verified measurement model as a precedent for the present study. 3) CFA is the most robust method available for analyzing multiple covariances among a set of latent and observable variables (Byrne, 2010).

Figure 1 represents the model used for the CFA in the present study. This model was constructed and analyzed using AMOS statistical software (Arbuckle, 2010). Not only does AMOS allow one to graphically represent a structural equation model, it also provides a theoretically-based, consistent, efficient, and relatively unbiased method for handling missing data. Known as the full information maximum likelihood (FIML) estimation, this method uses a

predictive distribution of scores to model the underlying pattern of missing data and then replaces missing scores with estimated values based on this model (Byrne, 2010). FIML was used for the single-level CFA in the present study.

In order to determine how well the study data match the hypothesized factor model, several "goodness-of-fit" indices were calculated. Of the many fit indices available, the χ^2 ("chi-squared") goodness-of-fit statistic is by far the most widely used (Hu & Bentler, 1995). The generally accepted guideline for this test of model fit is that a χ^2 value that is less than twice the model's degrees of freedom indicates acceptable overall model fit (Hu & Bentler, 1995).

The root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) were also examined. These specific statistics were selected per Brown's (2006) recommendations for evaluating CFA model fit. The RMSEA assesses whether the model holds *reasonably* well in the study sample, with values close to .06 or below indicating acceptable model fit (Hu & Bentler, 1999). The CFI and TLI offer a more liberal estimation than the χ^2 in that they evaluate model fit against a solution that posits no relationships among variables. Values close to .90 or greater are representative of acceptable model fit (Hu & Bentler, 1995).

Inherent in hierarchical data sets like the one from the present study (e.g., participants nested within groups) are risks of violating the "independence of observations" assumptions of most statistical tests (Baldwin, Wampold, & Imel, 2007; Burlingame, Kircher, & Honts, 1994). Ignoring intragroup dependency puts a study's analysis at risk for Type I error inflation (Baldwin, Murray, & Shadish, 2005, Baldwin et al., 2011; Baldwin, Stice, & Rohde, 2008). One method for evaluating the effects that intragroup dependency has on nested data is to compare statistical results at both the within-group and between-group levels (Baldwin et al., 2007). The

intraclass correlations (ICC), which provide an effect size estimate for group-level effects, is one method by which to make such a comparison (Baldwin et al., 2011; Imel, Baldwin, Bonus, & Macoon, 2008). Although Johnson et al. (2005) addressed intragroup dependency in their original formulation of the Johnson Model, the two most recent validity studies on the GQ (i.e., Krogel, 2009; Bormann & Strauss, 2009) did not. Because intragroup dependency poses a substantial threat to statistical integrity, it was accounted for in the present study.

Similar to Johnson et al. (2005), a multilevel CFA was conducted using the statistics software package Mplus (Muthen & Muthen, 2010). Mplus allows a researcher to obtain separate model fit statistics for the within-group and the between-group covariance structures. In order to determine whether or not intragroup dependency affected model fit in the present study, ICCs were examined for valence and magnitude; positive ICCs were considered as evidence for within-group dependency (Baldwin et al., 2008).

Hypothesis 2: Does the GQ have criterion-related validity? In order to assess the GQ's criterion-related validity, GQ subscale scores were correlated with four other measures of the group therapeutic relationship: the WAI, GCQ, ES, and the Cohesion subscale of the TFI. (See Table 1 for a summary of which GQ subscale was correlated with which criterion measure.) The selection of these measures is particularly relevant to the problem of "construct drift" implied in the title. Because these four measures were used as the template for the GQ's creation, low correlations between them and the GQ would suggest that the GQ has drifted off course in its efforts to measure what it purports to measure. Per the guidelines outlined in Reisinger and Burlingame (1997) for evaluating the validity of a psychotherapy questionnaire, validity coefficients at .50 and above were considered acceptable, while .75 and above were considered excellent. Both Pearson product-moment (r) and Spearman's rank (ρ) correlation

coefficients were calculated. The addition of ρ was intended to offer a non-parametric correlation coefficient that would account for the skewedness of the data.

Because low reliability inflates the standard errors of estimates, it is a threat to a test's validity (Cook & Campbell, 1979). Therefore, reliability estimates were calculated and assessed. The most popular method of assessing the reliability of multiple-item tests like the GQ is by calculating the Cronbach alpha (Cronbach, 1951) statistic (Warner, 2008). Reliability estimates of .80 and above were considered acceptable, while .90 and above were considered excellent (Reisinger & Burlingame, 1997). In cases were reliability was attenuated by restriction of range, estimates were adjusted using a formula provided by Ghiselli et al. (1981) and reported separately.

Results

Hypothesis 1

GQ item and subscale descriptive statistics revealed that half (16/30) of the GQ items are highly skewed (i.e., less than -1 or greater than +1; Bulmer, 1979). At the subscale level, Member-Group Positive Bonding and all Member-Leader subscale scores were highly skewed.

Model fit statistics for the single-level CFA were as follows: $\chi^2(380, N = 290) = 777.89$, $p = .00$; CFI = .93; TLI = .91; RMSEA = .05. Given that three of four indices reflect good fit and the fourth (χ^2) marginally failed (32%), we conclude the measurement model provides acceptable fit, supporting Hypothesis 1. Factor loadings and correlation coefficients for the single-level CFA (Figure 2) and within-group multilevel CFA (Figure 3) were similar. ICCs for each GQ item are listed in Table 2, with the average GQ item ICCs for member-member, member-leader, and member-group structural relationship parameters being .15, .11, and .27, respectively (ranged from .08 to .49).

Hypothesis 2

Internal consistency reliability coefficients (Cronbach's alpha) for GQ subscales are presented in Table 3. Reliability coefficients indicated acceptable reliability for all three GQ subscales, with the Positive Bonding and Positive Working Relationship subscales meeting criteria for excellent reliability. These values are similar to those reported by Krogel (2009). Because the GQ data collected from the university counseling centers had a restricted range, as counseling center GQ data tend to have (Krogel, 2009), the Ghiselli et al. (1981) formula was employed to estimate what the GQ subscales' reliability coefficients would have been if range was not attenuated. Values from the Utah State Hospital population in Krogel (2009) were used to meet the conditions of the formula. The subsequent estimation yielded increased reliability for all three subscales (see Table 3).

Internal consistency reliability coefficients for criterion measures are presented in Table 4. Reliability coefficients for the Bond, Task, and Goal subscales of the WAI and for the Cohesion subscale of the TFI were comparable to those reported in Johnson et al. (2005). Reliability coefficients for the ES were slightly higher than those reported in Johnson et al. (2005). Reliability coefficients for the Engaged, Conflict, and Avoidance subscales of the GCQ were much lower than those reported in other studies (e.g., .94, .88, and 92, respectively; Kivlighan & Goldfine, 1991), falling closer to those reported in Johnson et al. (.70, .69, and .36, respectively; 2005).

Criterion-related validity coefficients for each GQ subscale are presented in Table 5. Positive Working Relationship correlated well with the WAI at both member-member and member-leader levels. All correlations are acceptable by the Reisinger and Burlingame (1997) standards, with several falling in the "excellent" category. The same can be said of Positive

Bonding Relationship, with the exception of the GCQ Engaged correlations. While it is still acceptable, the GQ's correlation with this subscale was relatively low (e.g., $r = .56$). Correlations between Negative Relationship and GCQ: Conflict and the ES were also acceptable, though not quite as high as those of the other two GQ subscales. Given that all correlations were greater than $r = .5$, Hypothesis 2 is supported.

Discussion

Since its inception in 2005, Johnson's 3-factor model of the group therapeutic relationship has been empirically supported by 5 separate studies that sampled nearly 2,200 individuals from over 260 groups in 4 different countries. The GQ was created to measure the group therapeutic relationship per the Johnson Model and, as such, has the potential to be a powerful tool in the hands of group clinicians. But has the GQ drifted from the firm foundation upon which it was based? Krogel (2009) noted that when she and her team altered and/or excluded many of the original questionnaire items used to formulate the Johnson Model such a drift might have occurred. In addition, Johnson et al. (2005) observed that group members have difficulty making meaningful distinctions at the level of relationship structure, which would make a self-report questionnaire like the GQ subject to the same difficulties. The present study was designed to address these concerns about the GQ's validity by using GQ response data to replicate the Johnson 3-factor model and by correlating GQ items with the original measures used to create the Johnson Model.

Model fit statistics for the single-level CFA of GQ response data met standards for acceptable model fit, thereby confirming Hypothesis 1. Not only do these findings support Johnson's 3-factor model as a theoretically sound and empirically robust definition of the group therapeutic relationship, they also suggest that the GQ is a valid measure of that construct. This

empirical support responds to Krogel's (2009) concern that the GQ might have drifted from Johnson's original construct formulation.

One issue addressed by Johnson and her colleagues in their original construct formulation, but not accounted for by Krogel (2009) was intragroup dependency and the corresponding risk of Type I error inflation. Like Johnson et al. (2005), the present study sought to address these concerns with a within-group, multilevel CFA and by examining the ICC for each GQ item. All 30 GQ item ICCs were positive, suggesting that participants within groups produced scores more similar to each other's than to other participants'. In other words, the positive ICCs indicate that at least some of the variance in GQ scores can be accounted for by group-dependent phenomena. This was especially true for the member-group GQ items (23-30), which produced the highest ICCs (from .17 to .49). These strong group effects are to be expected, given that the member-group items are the most group-dependent items by design. Johnson et al. (2005) found similar ICCs in their study (i.e., ranging from .13 to .42). The present study's replication of these findings lends support to the notion that the GQ is sensitive to group-level processes.

It is noteworthy that none of the GQ item ICCs exceeded .5. This suggests that while a notable amount of the variance in GQ scores can be accounted for by group-dependent phenomena, the majority of the variance remains unaccounted for. These unaccounted-for predictors of GQ variance could include various individual, member-dependent, and group-level phenomena. Given how difficult it is to measure an individual's relationship to a nebulous construct like the "group-as-a-whole," it might also be argued that the relatively high member-group ICCs are more than acceptable and indeed impressive; especially since the present findings replicate Johnson's earlier ICCs. However, more research on group-level relationship

variables using multilevel models and ICCs is needed in order to make more definitive conclusions about measures designed to assess group dynamic properties. Two studies on a single group relationship measure are insufficient to ascertain normative expectations on acceptable intra-group dependency values. As it stands, model fit statistics for both the within-group and between-group models indicate acceptable model fit, supporting the GQ's construct validity.

As hypothesized, the GQ subscales correlated well with their corresponding criterion measures at all three structural relationship levels. This finding implies that in spite of concerns expressed by Krogel (2009), the GQ has not drifted off target and is still a valid tool for measuring the group therapeutic relationship. The lowest correlation (r = .54) was between GQ Positive Bonding Relationship and the Engaged subscale of the GCQ. One reason for this lower correlation might be the low reliability of the Engaged subscale. At α = .65, the Engaged subscale had the lowest internal consistency reliability of any subscale correlated with the GQ Positive Bonding Relationship subscale.

Now that strong correlations have been discovered between the GQ and the original measures used to create it, dispelling doubts about construct drift, future research should compare the GQ with other measures of the group therapeutic relationship. Bormann et al. (2011) did this with German measures and found similar promising relationships. Additional English candidates might be the Group Atmosphere Scale (Silbergeld, Koenig, Manderscheid, Meeker, & Hornung, 1975) and the Harvard Group Cohesiveness Scale (Budman et al, 1987).

Limitations/Conclusion

There are several limitations in this study including the lack of clinical diversity in the sample population (i.e., 96% of participants came from college counseling center groups).

However, since the findings herein agree with studies using non-clinical process, inpatient therapy, and psychiatric hospital groups (Bormann et al., 2011; Krogel, 2009) we do not believe the homogeneity unduly affected results. The aforementioned restriction of range in GQ responses was also a limiting factor in the present study, attenuating the strength of validity coefficients. Restricted range is a common problem for all psychotherapy process measures that capture positive attributes of the therapist and/or group (Hill & Lambert, 2004).

Its limitations notwithstanding, the present study found that the GQ can provide group clinicians with valid measurements of the quality of their group member's therapeutic relationships at the member-member, member-leader, and member-group structural levels.

References

Arbuckle, J. L. (2010). Amos (Version 19.0) [Computer Program]. Chicago: SPSS.

Bakali, J. V., Baldwin, S. A., & Lorentzen, S. (2009). Modeling group process constructs at three stages in group psychotherapy. [Article]. *Psychotherapy Research, 19*(3), 332-343.

Bormann, B., & Strauss, B. (2009). *The German Group Questionnaire: A multi-site validation study*. Paper presented at the Society for Psychotherapy Research.

Burlingame, G., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K., 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.

Burlingame, G. M., Fuhriman, A., & Johnson, J. E. (2002). Cohesion in group psychotherapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients.* (pp. 71-87). New York, NY US: 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, and Practice, 7*(1), 3-12.

Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.

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 US: Guilford Press.

Fuhriman, A., & Burlingame, G. M. (1990). Consistency of matter: a comparative analysis of individual and group process variables. [Feature]. *Counseling Psychologist, 18*, 6-63.

Fuhriman, A., & Burlingame, G. M. (1999). Does group psychotherapy work? In J. R. Price, D. R. Hescheles & A. R. Price (Eds.), *A guide to starting psychotherapy groups.* (pp. 81-98). San Diego, CA US: Academic Press.

Hale, A. E. (1985). *Conducting Clinical Sociometric Explorations: A Manual for Psychodramatists and Sociometrists*. Roanoke VA: Royal Pub.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223-233.

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group Climate, Cohesion, Alliance, and Empathy in Group Psychotherapy: Multilevel Structural Equation Models. *Journal of Counseling Psychology, 52*(3), 310-321.

Krogel, J. (2009). The Group Questionnaire: A new measure of the gorup relationship. Unpublished doctoral dissertation. Brigham Young Universtiy.

Lese, K. P., & MacNair-Semands, R. R. (2000). The Therapeutic Factors Inventory: Development of a scale. *Group, 24*(4), 303-317.

MacKenzie, K. R. (1983). The clinical application of a group climate measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159-170). New York: International Universities Press.

MacKenzie, K. R. (1995). *Effective use of group therapy in managed care*. Washington, DC US: American Psychiatric Association.

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.

Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama (2nd ed.)*. Oxford England: Beacon House.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(3), 207-210.

Treadwell, T. W., Kumar, V. K., Stein, S. A., & Prosnick, K. (1998). Sociometry: Tools for research and practice. [Article]. *International Journal of Action Methods, 51*(1), 23.

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy (5th ed.)*. New York, NY US: Basic Books.

Table 1

Hypothesis Matrix for Sociometric Exploration of the GQ

Relationship Quality	Relationship Structure		
	Member-Member	Member-Leader	Member-Group
Positive Working Relationship	Hypothesis 3	Hypothesis 4	–
Positive Bonding Relationship	Hypothesis 5	Hypothesis 6	Hypothesis 7
Negative Relationship	Hypothesis 8	Hypothesis 9	Hypothesis 10

Table 2

Intraclass Correlation (ICC) for GQ Items

GQ Item	ICC	GQ Item	ICC	GQ Item	ICC
1	.13	11	.10	21	.07
2	.19	12	.15	22	.13
3	.13	13	.12	23	.49
4	.15	14	.10	24	.33
5	.09	15	.12	25	.39
6	.17	16	.12	26	.19
7	.14	17	.09	27	.17
8	.18	18	.08	28	.20
9	.08	19	.11	29	.18
10	.13	20	.18	30	.23

Note. ICC = intraclass correlation for groups, or the proportion of variance accounted for by group.

Table 3

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).

Table 4

Criterion Measure Internal Consistency Reliability (Cronbach's Alpha)

Subscale	Overall	Member-Member	Member-Leader
TFI: Cohesion	.89	–	–
GCQ: Engaged	.65	–	–
GCQ: Conflict	.66	–	–
GCQ: Avoidance	.29	–	–
WAI: Bond	.93	.90	.89
WAI: Goal	.84	.69	.72
WAI: Task	.92	.84	.86
ES: Total	.92	.88	.87
ES: Positive	.94	.90	.90
ES: Negative	.81	.72	.69

Note. N = 290. The "–" indicates that this value was not calculated because it was not applicable.

Table 5

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.
** $p < .01$.

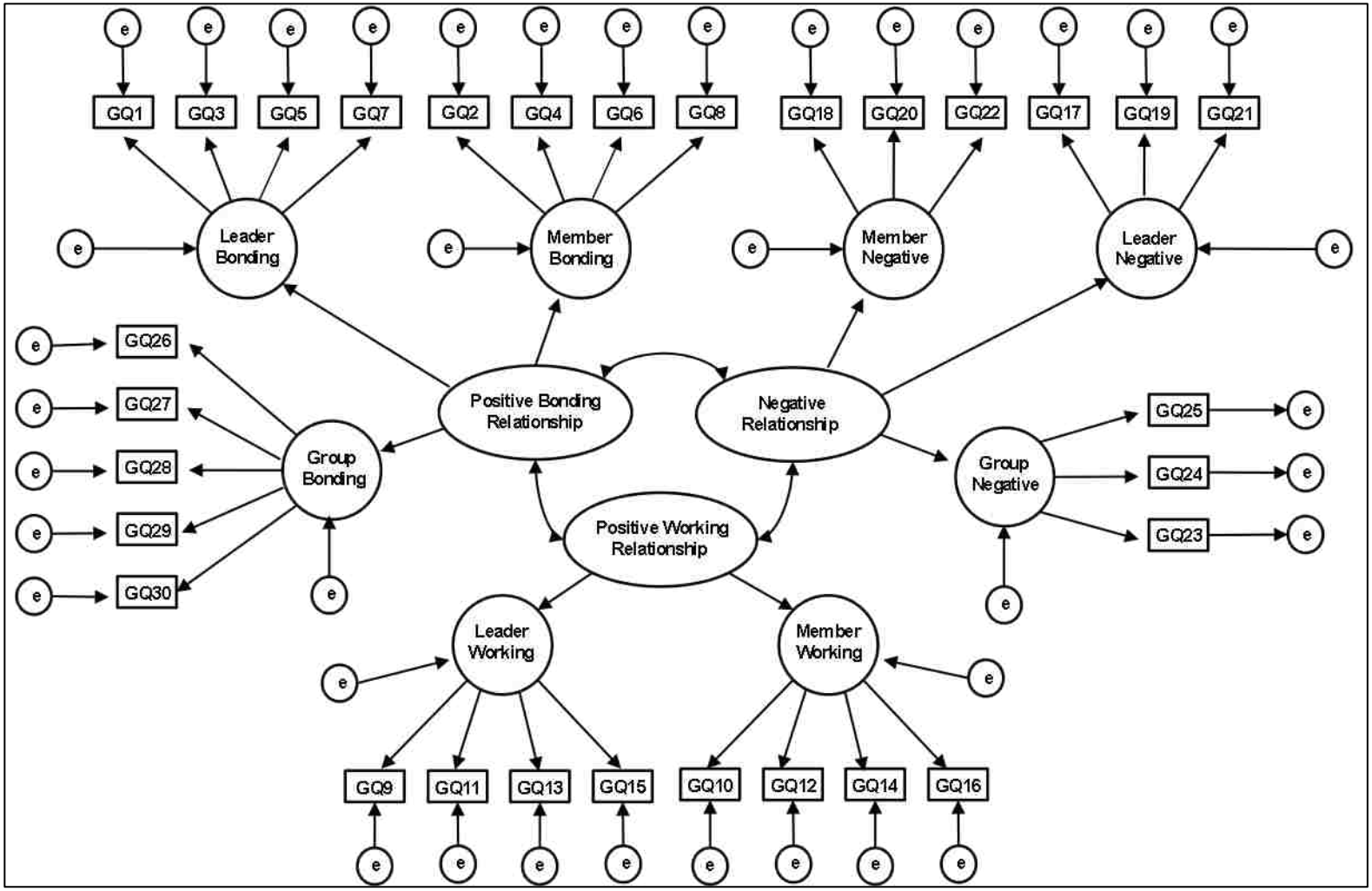


Figure 1. CFA model for the analysis of the GQ's construct validity. Ellipses and circles represent unobserved latent factors. Rectangles and squares represent observed variables. Single-headed arrows represent the impact of one variable on another. Double-headed arrows represent covariances or correlations between pairs of variables.

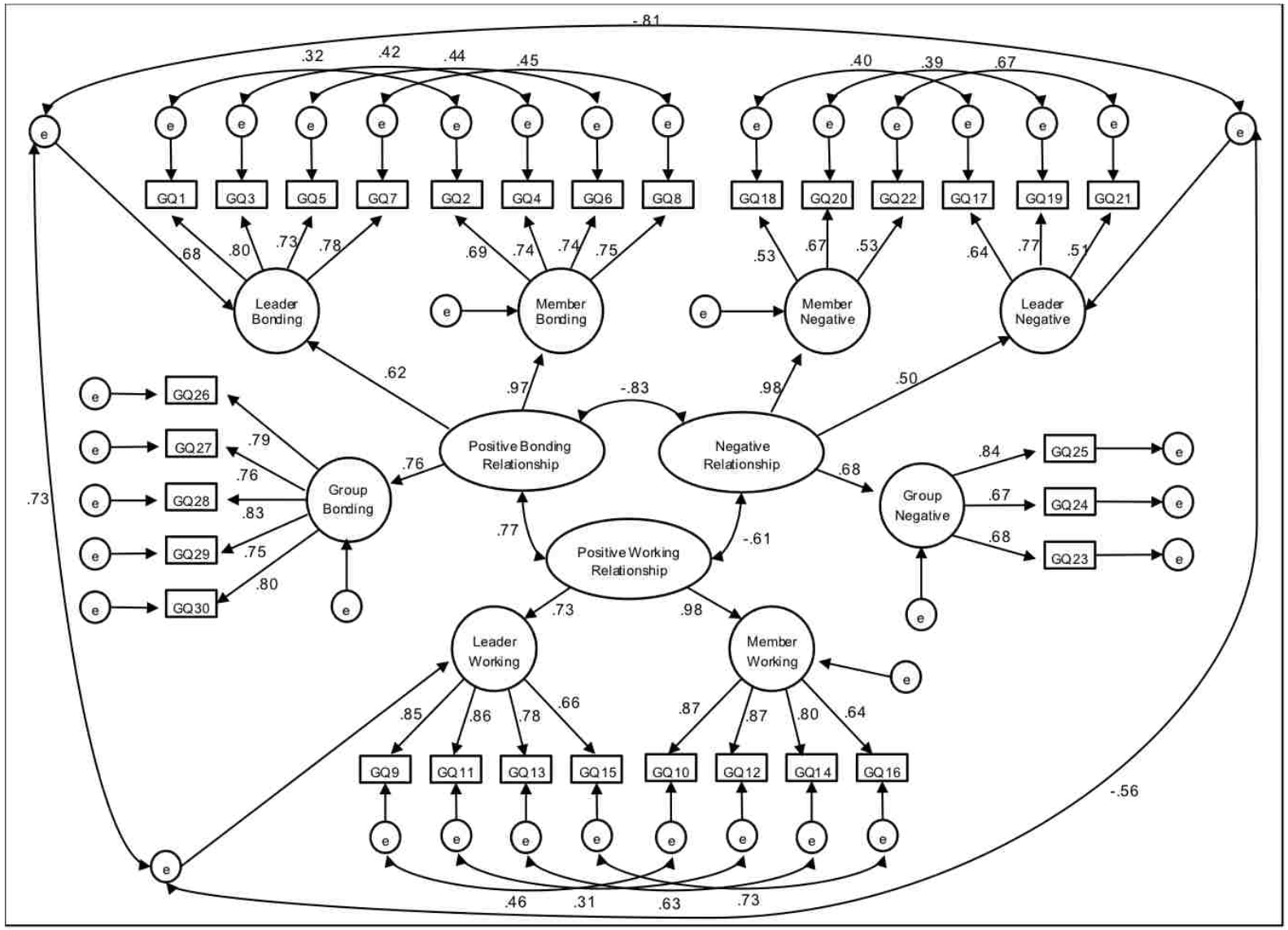


Figure 2. Single-level CFA for the GQ. The numbers by the single- and double-headed arrows represent factor loadings and correlations, respectively.

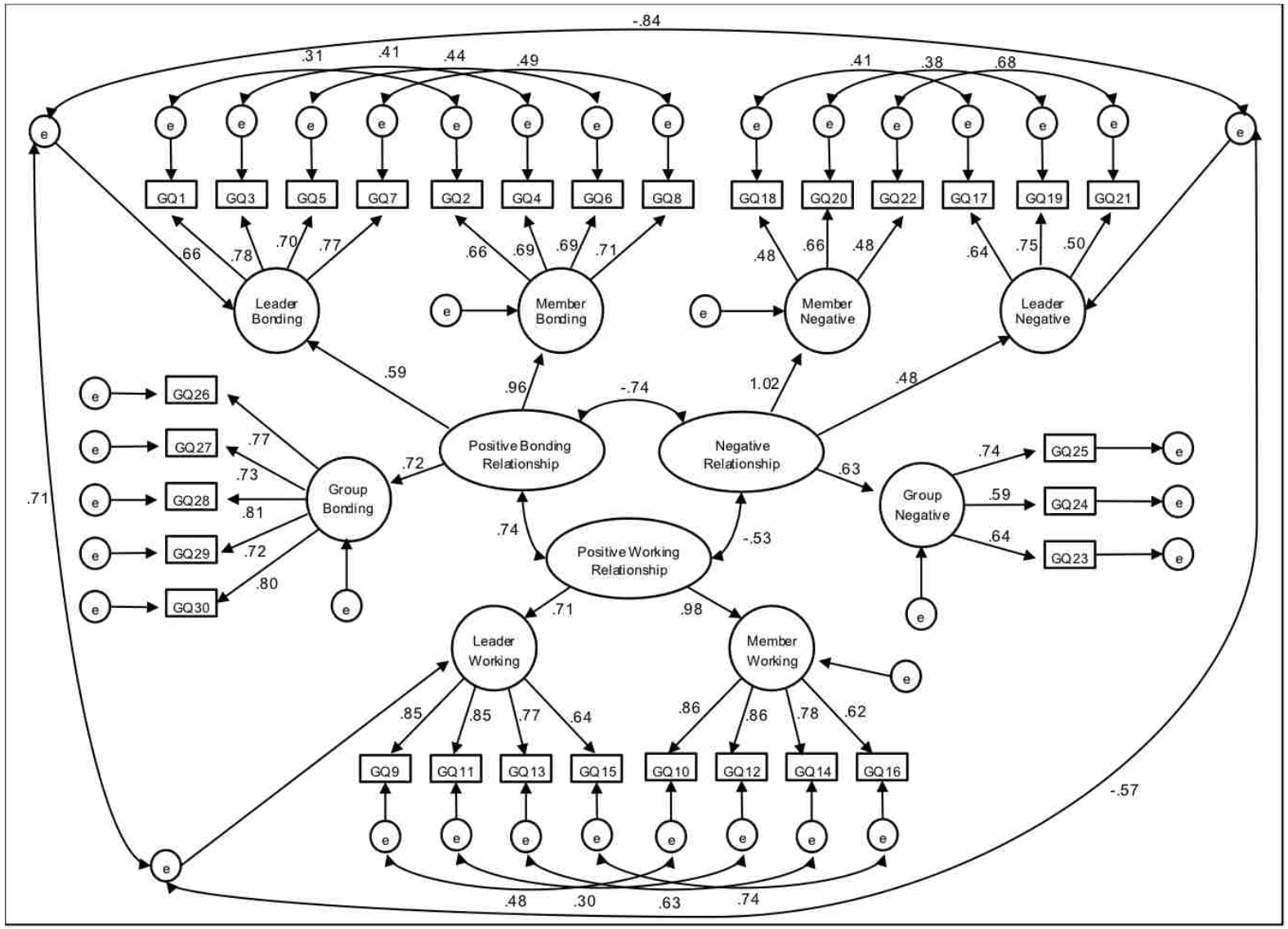


Figure 3. Within-group CFA for the GQ. The numbers by the single- and double-headed arrows represent factor loadings and correlations, respectively.