

Group Dynamics: Theory, Research, and Practice

The Group Questionnaire (GQ)—Psychometric Properties Among Outpatients With Personality Disorders

Geir Pedersen, Elfrida H. Kvarstein, Theresa Wilberg, Espen J. Folmo, Gary M. Burlingame, and Steinar Lorentzen
Online First Publication, November 4, 2021. <http://dx.doi.org/10.1037/gdn0000176>

CITATION

Pedersen, G., Kvarstein, E. H., Wilberg, T., Folmo, E. J., Burlingame, G. M., & Lorentzen, S. (2021, November 4). The Group Questionnaire (GQ)—Psychometric Properties Among Outpatients With Personality Disorders. *Group Dynamics: Theory, Research, and Practice*. Advance online publication. <http://dx.doi.org/10.1037/gdn0000176>

The Group Questionnaire (GQ)—Psychometric Properties Among Outpatients With Personality Disorders

Geir Pedersen^{1, 2}, Elfrida H. Kvarstein^{1, 3}, Theresa Wilberg^{3, 5}, Espen J. Folmo¹,
Gary M. Burlingame⁴, and Steinar Lorentzen³

¹ Network for Personality Disorders, Section for Personality Psychiatry and Specialized Treatments,
Department for National and Regional Functions, Division of Mental Health and Addiction, Oslo
University Hospital, Oslo, Norway

² The Norwegian Centre of Mental Disorders Research (NORMENT),
Institute of Clinical Medicine, University of Oslo

³ Institute of Clinical Medicine, University of Oslo

⁴ Department of Psychology, Brigham Young University

⁵ Department for Research and Innovation, Clinic for Mental Health and Addiction,
Oslo University Hospital, Oslo, Norway

Objective: Relational and emotional vulnerability represent essential problems among patients with personality disorder (PD). Group psychotherapy is a central component of evidence-based PD treatments. Generally, patient and therapist interrelationships predict improvement in therapy. However, although treatment of patients with PD is a more complex process, group processes are poorly elaborated in PD research. Documentation of the psychometric quality of group process measures in PD samples is an important precursor of such research. The Group Questionnaire—GQ is based on concepts of group cohesion and climate, empathy, and alliance and aims to capture the quality of member–member, member–group, and member–leader relationships in group therapy. A three latent factor structure (positive bond, positive work, and negative relationship) has generally been supported. This study aimed to perform a psychometric analysis of GQ administered in a clinical population of patients with PD. **Method:** The study included

Geir Pedersen  <https://orcid.org/0000-0002-8599-8871>
Theresa Wilberg  <https://orcid.org/0000-0002-4700-0388>

None of the above authors have any financial disclosure or other conflicts of interest related to this study. Gary M. Burlingame is one of the co-authors of the GQ but was not involved in the conceptualization or analysis of this article.

We wish to thank Ole Klungsøyr at Department for Research and Innovation, Clinic for Mental Health and Addiction, Oslo University Hospital for statistical advice and assistance, and the patients and staff from the following units of the Norwegian Network for Personality Disorders for their contribution to this study: Unit for Group Therapy, Øvre Romerike District Psychiatric Center, Akershus University Hospital, Jessheim; Group Therapy Unit, Nedre Romerike District Psychiatric Center, Akershus University Hospital, Lillestrøm; Group Therapy Unit, Follo District Psychiatric Center, Akershus University Hospital, Ski; Group Therapy Unit, Groruddalen District Psychiatric Center, Akershus University Hospital, Oslo; Clinic for Personality disorders, Outpatient Clinic for Specialized Treatment of Personality Disorders, Section for Personality psychiatry

and specialized treatments, Oslo University Hospital, Oslo; Group Therapy Unit, Lovisenberg District Psychiatric Center, Lovisenberg Hospital, Oslo; Group Therapy Team, Vinderen Psychiatric Center, Diakonhjemmet Hospital, Oslo; Unit of Personality psychiatry, Vestfold District Psychiatric Center, Sandefjord; Unit for Intensive Group Therapy, Aust-Agder District Psychiatric Center, Sørlandet Hospital, Arendal; Unit for Group Therapy, District Psychiatric Center, Strømme, Sørlandet Hospital, Kristiansand; Group Therapy Unit, Stavanger District Psychiatric Center, Stavanger University Hospital, Stavanger; Section for group treatment, Kronstad District Psychiatric Center, Haukeland University Hospital, Bergen; MBT Team, Department of Substance Abuse Medicine, Haukeland universitetssjukehus, Bergen; and MBT Team, Outpatient Clinic, Rogaland A-senter, Stavanger.

Correspondence concerning this article should be addressed to Geir Pedersen, Network for Personality Disorders, Section for Personality Psychiatry and Specialized Treatments, Department for National and Regional Functions, Division of Mental Health and Addiction, Oslo University Hospital, Ullevaal, P.O. Box 4956, Nydalen, Oslo 0424, Norway. Email: geir.pedersen@medisin.uio.no

369 patients with PDs attending group psychotherapy in 14 outpatient treatment units on a specialist mental health service level within the Norwegian Network of Personality Disorders in the period 2017–2020. Psychometric analyses included three latent factors and eight subfacets. **Results:** The three latent factors, positive bond, positive work, and negative relationship, were replicated. Psychometric integrity of the tripartite relationship structure—member–member, member–group, and member–leader, with eight subfacets, was supported. **Conclusion:** The GQ with three latent factors and eight subfacets can be recommended for future research and clinical practice in patient populations with PD.

Highlights and Implications

- The study highlights that Group Questionnaire (GQ) can satisfactorily capture the group psychotherapy process for patients with personality disorder (PD) in its current form with three latent factors and eight subfacets.
- Results replicate former research on the psychometric qualities of GQ.
- Results implicate that the GQ can be useful in further research on PD treatment processes and mechanisms and mediators of change.

Keywords: factor analysis, GQ, Group Questionnaire, group relationship, personality disorder

Personality disorders (PDs) are frequent mental disorders (Winsper et al., 2020) with a reported prevalence of 40% within mental health services (Newton-Howes et al., 2010). Features of PD can be conceptualized within dimensions of self and interpersonal personality functioning with core interpersonal problems of social insecurity and communication (Bender et al., 2011; Fonagy et al., 2017; Hutsebaut et al., 2016). Both self and interpersonal dimensions are likely to have considerable implications for the group therapy treatment process. PD patients may, in particular, have profound difficulties trusting and connecting to other people, or conversely, describe extreme dependency and easily activated fears of separation. Understanding and managing responses in a complex, social arena such as group psychotherapy may be particularly challenging (Kvarstein et al., 2020). However, only few studies have studied the significance of aspects of group relationships, for example, therapeutic alliance, in specialized group-based PD treatments (Euler et al., 2018; Folmo et al., 2021).

Both efficacy and effectiveness of group psychotherapy are generally well established for a variety of mental disorders (Burlingame et al., 2013; Burlingame & Strauss, 2021). Research has indicated that individual and group formats are equally effective in promoting change (Burlingame, Seebeck, et al., 2016). Within specialist mental health services, group psychotherapy is a frequent

treatment approach (Lorentzen & Ruud, 2014) and in specialized PD treatments, groups often represent core components (Storebø et al., 2020). Although results revealed considerable heterogeneity, a meta-analysis of 24 randomized controlled trials of group-based treatments for borderline PD concluded that the group therapy format can be recommended for these patients (McLaughlin et al., 2019).

According to an integrative model, components that may explain the outcome of group treatments consist of elements from formal change theory, aspects of small group processes, therapist and patient characteristics, and structural aspects, such as duration of treatment, session frequency, and group size, or interactions across these domains (Burlingame et al., 2013). Characteristic and complex aspects of groups are the multiple and different relationships that develop between group participants. In clinical and empirical group literature, group cohesion is the most popular of several group relationships constructs (e.g., therapeutic alliance, group climate, and group atmosphere). Group cohesion has, over time, become synonymous with the therapeutic relationship in group psychotherapy (Burlingame et al., 2002). However, an array of partly overlapping cohesion instruments exist, and consensual definitions between the different measures are lacking (Alldredge et al., 2021; Hornsey et al., 2007, 2009). In spite of this

diversity of measures and definitions of group relationships, ample evidence supports dimensions of cohesion incorporating member–leader relationships, member to member relationships, and the group as a whole (Dion, 2000). Two meta-analyses of 40 and 55 studies have demonstrated that group cohesion significantly predicts patient improvement across a wide range of theory orientations and patient populations (Burlingame et al., 2011; Burlingame, McClendon, & Yang, 2018). Within the PD research field, literature on the role of group process factors is scarce (McLaughlin et al., 2019). Despite emphasis on group dynamics in several treatment manuals for borderline PD, few have investigated change processes, therapeutic alliance and group cohesion. Nonetheless, several qualitative studies of patients with borderline PD in specialized PD treatment have emphasized participation in group therapies often offers strong learning experiences (Morken et al., 2019).

Development of the Group Questionnaire

As a part of a programmatic research conducted by the Consortium for Group Research and Practice (C-GRP; <http://crgp.byu.edu>) in the early 2000s (Burlingame, Gleave, et al., 2016), Johnson et al. (2005) sought to achieve a conceptual clarification of the conundrum of competing cohesion measures and carried out factor analyses including items from the four most commonly used measures: Group Climate Questionnaire (GCQ; MacKenzie, 1983), the Therapeutic Factors Inventory, Subscale Cohesiveness (TFI-Coh; Lese & MacNair-Semands, 2000), Working Alliance Inventory (WAI; Horvath & Greenberg, 1989), and the Empathy Scale (ES; Burns & Auerbach, 1996). The resulting model identified a latent relationship quality dimension with three factors, two affective aspects (positive emotional bonds and negative relationships) and one work-based facet of group relationships (positive work), and a latent relationship structure dimension (member–leader, member–member, member–group). Participants from two sets of groups were used: The first was a clinical sample of 326 participants from 81 groups at 14 university counseling centers across the U.S. and the second was nonclinical consisting of 336 professional group therapists who participated in 30 2-day training groups.

This model was later replicated in a sample of 453 members from inpatient groups from 15 hospitals in Germany and Switzerland (Bormann & Strauß, 2007), and it was partially replicated in a sample of 145 outpatients with heterogeneous diagnoses, treated in nine short-term and nine long-term groups in Norway (Bakali et al., 2009).

Eventually, a more user-friendly measure of group relationships, the Group Questionnaire (GQ), was developed in two further C-GRP studies, by reducing the number of underlying items from 80+ items used in the Johnson model to 30 items in the GQ. The data reduction relied heavily on both empirical data analysis and clinical judgment, and in general, items with smaller factor loadings or high correlations with other items were eliminated, without information loss. The 30-item GQ model replicated the Johnson model with 485 members participating in nonclinical outpatient and inpatient groups in the U.S. (GQ; Krogel, 2008; Krogel et al., 2013). A German version of the GQ has also been validated by Bormann et al. (2011), who analyzed 424 members from 67 inpatient groups from 15 hospitals in Germany and Switzerland, confirming the Johnson model. Moreover, the GQ criterion validity was assessed to be high with the original four therapeutic relationships measures in U.S. data from 290 patients from 65 treatment groups at 4 university counseling centers and 1 community mental health clinic (Thayer & Burlingame, 2014). A meta-analysis provided further support for the factorial validity of GQ. By including data from six former studies, comprising a total of 2,479 participants, Janis et al. (2018) concluded that the constructs are operating in the same way across different populations. Finally, a recently developed Italian version of the GQ has been validated in a sample of 536 group members from 32 nonclinical groups of undergraduate students, and the results from different structural equation models in multilevel confirmatory factor analysis (CFA) confirmed the three-factor structure of the GQ, defined by Positive bond, Positive work, and Negative relationship (Giannone et al., 2020). It should also be mentioned that a systematic review and quality assessment of group process questionnaires (Orfanos et al., 2020) including 13 different measures rated GQ (Krogel et al., 2013) as the most appropriate measure if one is interested in measuring overall group processes.

Thus, Krogel et al.'s (2013) 30-item questionnaire (GQ), based on Johnson et al.'s model (2005) of group relationship organized in Positive

Bonding, Positive Working, and Negative Relationship, has been validated in various samples: professionals in the field of group psychotherapy, undergraduate students, outpatients from university counseling and community mental health centers, chronic inpatients with severe mental illness, and inpatients hospitalized in German and Swiss clinics and psychiatric hospitals. All findings have supported the factor structure initially proposed by Johnson et al. (2005). However, none of these numerous studies of GQ are, to the best of our knowledge, based on study samples representative of patients with PD. A psychometrically sound and valid measure of central group relationships is essential to further research aiming to explore the associations between aspects of these processes and outcome, for patients with PD treated in groups.

So far, the GQ has been used to monitor the course of therapy and help therapists to detect and attend to problems when group relationships worsen or patients deteriorate, and studies show that such feedback improves outcomes and reduces attritions from therapy in more heterogeneous patient samples (Burlingame, Whitcomb, et al., 2018). More recent articles demonstrate how central therapeutic alliance that mostly has been studied in individual psychotherapy also is for group therapy (Alldredge et al., 2020) and how ruptured alliances are detected and coupled with the therapists repair attempts in group therapy, using the GQ (Burlingame et al., 2021).

Although several studies on GQ have either validated the factor structure, that is, with reference to the Johnson et al.'s model (2005), or compared GQ scores with other measures of alliance, we do not know of any study that has focused on possible associations between GQ scores and concurrent measures of subjective distress related to depression, anxiety, social functioning, or overall patient satisfaction. One exception to this is perhaps process studies using the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2001), such as Obeid et al. (2018), but they are outcome studies showing the impact of group relationship on treatment outcome.

Aims of the Present Study

The aim of the present study is to validate the proposed factor structure of the Norwegian version of GQ in a sample of patients with PD or severe personality problems, analyze the relationship

between the different subfacets of GQ, as well as its relations to measures of depression, anxiety, social and occupational functioning, and overall satisfaction with treatment. Our hypotheses will be the following: (a) The composition of GQ reveals the same factor and conceptual structure as proposed by the Johnson model. (b) Concurrent levels of subjective distress have minor associations with the perceived group relationship, and (c) Group relationship is substantially associated with overall patient satisfaction.

Material and Method

The Context for Data Collection

In this multisite, naturalistic study, the material comprised data from 14 outpatient units within the Norwegian Network for Personality Disorders (the Network; Karterud et al., 2003). All units are outpatient services on a specialist mental health service level, providing treatment for a broad range of patients with significant personality problems and PD. Group therapies constitute a central treatment format. The most frequent formats are long-term (psychodynamic) psychotherapy groups in a stand-alone format or incorporated in a comprehensive outpatient program, most often mentalization-based treatment (MBT). More structured/focused short-term groups using psychodynamic, cognitive, or pedagogical techniques (psychoeducation) are also frequent. In addition, several units offer art therapy and body awareness in a group format. Among multicomponent PD treatment programs combining group and individual formats, schema-focused therapy and dialectical behavioral therapy are also represented in addition to MBT (see Table 1). Treatment duration is most often time limited, for short-term groups approximately 6 months, and for long-term groups, 2–3 years. The therapist teams are mainly multidisciplinary including psychiatrists, psychologists, psychiatric nurses, and social workers. All group leaders are formally trained in their respective discipline (i.e., group analytic psychotherapy, MBT, dialectical behavioral therapy [DBT]).

All units within the Network have the same assessment procedures, including initial clinical and diagnostic evaluations, and follow-up

Table 1*Overview of Treatment During the Last 6 Months Prior to 1-Year Assessment*

Treatment category	Specification of treatment	% ^a
Manual-based PD treatment programs combining group and individual modalities	Mentalization-based treatment (MBT)	40
	Schema-focused therapy	9
	Dialectical behavioral therapy	1
Stand-alone or preparatory before MBT	MBT psychoeducation	2
Other group-based therapies, stand-alone and individualized combinations with other groups/individual therapy	Psychodynamic group therapy	31
	Physical activity group	2
	Body awareness group	2
	Art therapy	2
	Stabilization group (trauma therapy)	1
	Metacognitive interpersonal therapy	3
	Individual psychodynamic therapy	13
Individual therapies stand-alone or in individualized combination with groups	Individual cognitive therapy	3
	Metacognitive interpersonal therapy	1
	Supportive therapy	1
	Trauma exposure therapy	1
	Other approaches	6

Note. PD = personality disorder.

^a Valid percent ($n = 288$).

evaluations every 6 months during treatment. Clinical data from these assessments are collected in an anonymous, data registry for research purposes, thus aggregating data across units. Evaluation includes both patient self-report and therapist reported information. The GQ was incorporated in this collection of evaluation instruments used within the treatment units as a regular routine.

The Study Sample

Within units, the initial evaluation process can last from 3 weeks to 3 months, and some patients are allotted to a waiting list before treatment can start. Thus, the patients vary as to how long they have been in treatment at the time of the first 6-month evaluation. Consequently, to ensure that all patients had some experience of the group treatment, data in this study represent the second 6-month evaluation, 1 year after the initial assessment. An implication of this choice is that data mainly represent attendance in long-term psychotherapy groups. Of a total sample of 1,329 psychiatric outpatients admitted to treatment in the period between August 2017 and April 2019, 852 participants had been in treatment long enough for the 1-year assessment to take place. Among these, the final study sample included 369 participants (43%) who had completed the 1-year assessment.

The Participants

In the final study sample ($N = 369$), 78% were females, mean age was 30.6 years ($SD = 8.6$), 56% was either in work or study prior to treatment, 63% were single. Twenty-three percent reported incidents of self-mutilation and 5% had attempted suicide in the 6-month period before the 1-year assessment. Seventy-six percent had one or more PD diagnoses, borderline and avoidant PD being the two dominating categories, and 93% had one or more symptom disorders. See Table 2 for a more detailed description of diagnostic prevalence.

The average number of participants per treatment unit was 26 ($Mdn = 23$, range 8–41). In the inclusion period (2017–2019), a total of 109 long-term groups were active across the 14 units. Data on the nature of treatment given in the 6-month period prior to the 1-year assessment (therapist reported) were available for 80% of the study sample ($N = 288$; Table 1).

Group Questionnaire

The GQ is a 30-item self-report questionnaire measuring the multifaceted therapeutic relationships in groups, capturing aspects of alliance, cohesion, and climate. The questionnaire is organized around *three qualitative aspects* of relationships, each operationalized to characterize the

Table 2
Diagnostic Prevalence

Diagnoses	<i>n</i>	% ^a
Symptom disorders:		
Current major depressive disorder	136	48
Major depressive disorder in remission	25	9
Dysthymic disorder	38	13
Bipolar II	12	4
Panic without agoraphobia	22	8
Panic with agoraphobia	25	9
Agoraphobia without panic	10	4
Social phobia	68	24
Obsessive–compulsive disorder	9	3
Generalized anxiety disorder	45	16
PTSD	27	10
Alcohol dependence/abuse	25	9
Substance dependence/abuse	17	6
Eating disorders	27	10
No symptom disorder	25	7
Personality disorders:		
Paranoid	19	6
Borderline	116	34
Avoidant	146	43
Dependent	18	5
Obsessive–compulsive	15	4
PD NOS	18	5
No PD disorder	83	24

Note. PTSD = posttraumatic stress disorder. PD = personality disorder; PD NOS = personality disorder not otherwise specified.

^a Diagnoses occurring in less than 3% omitted. Valid percent reported (*n* = 369).

perceived relationships between two to three, *structural subcomponents*, that is, leader, group, and other group members (Krogel et al., 2013): *Positive Bond*, toward the group (climate), the leader (alliance), and the members (cohesion); *Positive Work*, with the group (Task/Goals), and leader (Task/Goals); and *Negative Relationship*, with the group (conflict), leader (alliance rupture), and members (empathic failure). Thus, the GQ comprises eight different subcomponents. See Appendix. GQ is scored on a 7-point response format (1–7) from 1 = *not true* to 7 = *very true*, and the current GQ measures are scored with reference to the respondents' last attended group meeting. The Norwegian translation of the GQ was created in accordance with the guidelines of Wild et al. (2005) by a group of five persons, in which four had a PhD and four had extensive experience in group therapy, and one was a native English speaker.

The organization of questions (items) addressing member–group, member–leader, and member–member relations is theoretically sound but gives

no meaning psychometrically. The reason is that within each of these groups of items, there are three different subgroups, addressing positive bond, positive work, and negative relations. Combining these into one single score to operationalize the three structural relationships would make such a score hard to interpret. Psychometrically, there are two possible ways to combine the items of GQ. The first is to organize and combine them into three constructs addressing positive bond, positive work, and negative relationship. The second is to organize the items into eight subcomponents. In the first procedure, a model of three first-order latent constructs can be tested. In the second procedure, a model of eight first-order latent constructs can be tested, as well as a second-order conceptual model with eight first-order constructs, accounted for by three second- or higher order constructs.

In the present study, subfacets are computed by taking the average of responses on their respective items, and in that way make the subfacets comparable to each other. Due to different numbers of items within each domain and subfacet, computing measures by adding up item responses into sum scores makes comparisons hard. In the present study, the domain scores are computed as sum scores to enable comparisons with findings from former studies.

Diagnostic Assessment

Clinical diagnosis was assessed according to the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association [APA], 2013) using the Structured Clinical Interview for DSM-5 Personality Disorders for PD (SCID-5-PD; First et al., 2015) and the Mini-International Neuropsychiatric Interview (M.I.N.I.; Sheehan & Lecrubier, 1994) for symptom disorders. Diagnostic reliability was not investigated. However, diagnostic assessments were performed in each unit by clinical staff who had received systematic training in diagnostic interviews and principles of the Longitudinal, Expert, All Data procedure (LEAD; Pedersen et al., 2013; Spitzer, 1983). This means that diagnoses were based on all available information including referral letters, self-reported history and complaints, and overall clinical impression, in addition to diagnostic interviews.

Clinical Measures

Prior to the collection of data to the present study, a measure of client satisfaction was constructed comprising the five following questions: *Do you feel that your problems have been taken seriously?; Have you gained a better understanding of what you are struggling with?; Have you become more able to cope with your problems in everyday life?; Have you felt well taken care of throughout the treatment period?; and Has the treatment been useful to you?* The questions are rated on a 4-point scale (1–4) from 1 = *absolutely not*, 2 = *no, not very much*, 3 = *yes, mostly*, to 4 = *yes absolutely*. A formal psychometric evaluation of the scale has not yet been reported, but scale reliability based on the present study is reported in Table 3. Mean interitem correlation was 0.455 (range 0.299–0.533). The scoring instruction to the scale, called *Client Satisfaction Scale* (CSS), was “*Below you will find some questions about the treatment you have received. Please circle the number that best describes your opinion.*” The CSS was administered, along with the GQ and other clinical measures, every 6 months after the initial assessment.

Psychosocial functioning was assessed by the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002), a self-report 5-item scale of functional impairment that measures the level of impairment on a scale from 0 to 8, with 0 = *no impairment at all* and 8 = *very severe impairment*. The scores on the five different items are summarized by a total score of 0–40. The WSAS constitutes a reliable instrument, measuring the individual variation in a clinically important aspect of impairment (Pedersen et al., 2017). According to Mataix-Cols et al. (2005), WSAS scores above 30 denote severe disability, scores between 15 and 30 denote moderate impairment, and scores below 15 can be regarded as mild impairment or disability. This categorization has been supported in a similar clinical sample as the current one (Pedersen et al., 2017).

Self-reported level of depression was measured by the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), a 9-item questionnaire with a 4-point response format (0–3) from 0 = *not at all* to 3 = *nearly every day*. PHQ-9 scores are computed as the sum score of all nine items, ranging from 0 to 27. According to Kroenke and Spitzer (2002) cut points of PHQ-9 represent the following levels of depression:

None (0–4); Mild (5–9); Moderate (10–14); Moderately Severe (15–19); and Severe (20–27).

Self-reported level of anxiety was measured by the PHQ-9 (Generalized Anxiety Disorder [GAD-7]; Spitzer et al., 2006), a 7-item questionnaire with a 4-point response format (0–3) from 0 = *not at all* to 3 = *nearly every day*. GAD-7 scores are computed as the sum score of all seven items, ranging from 0 to 21. According to Löwe et al. (2008), cut points of GAD-7 represent the following levels of anxiety: None (0–4); Mild (5–9); Moderate (10–15); and Severe (>15).

Statistics

CFA was done in three steps. First, we conducted a first-order CFA of the three global dimensions of GQ. Second, we conducted a first-order CFA of the eight subfacets of GQ. Third, we conducted a second-order (full) model of GQ with the subfacets on the first level and the global dimensions on the second higher order level. Finally, we analyzed mean scores and scale reliability of the GQ measures, as well as inter-correlations between them.

Descriptive and correlational statistics, as well as chi-square (χ^2) and *t*-test statistics were conducted using IBM SPSS Statistics for Windows, Release 26 (IBM, 2019). The CFA was conducted with Mplus 7.4 (Muthén & Muthén, 2012) with estimations based on the maximum likelihood mean adjusted (MLM) function (Curran et al., 1996). Goodness of fit was estimated by root-mean-square error of approximation (RMSEA; Steiger, 1990), the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), and the comparative fit index (CFI; Bentler, 1990). Scale reliability was estimated by McDonald’s Omega (ω ; McDonald, 1999; Trizano-Hermosilla & Alvarado, 2016).

A RMSEA value of 0.05 or below indicates a good model fit, values between 0.05 and 0.08 indicate a reasonable fit, values between 0.08 and 0.10 a mediocre fit, and values greater than 0.10 an unacceptable fit (MacCallum et al., 1996). The general consensus is to use a cutoff value close to 0.06 (Hu & Bentler, 1999) or a stringent upper limit of 0.07 (Steiger, 2007). Values greater than 0.90 on CFI and TLI are normally required for good model fit, although Hu and Bentler (1999) have suggested a TLI \geq 0.95 as the threshold.

Table 3
GQ and Clinical Measures—Descriptives, Scale Reliability, and Zero-Order Intercorrelations

Measure	Descriptives				Reliability		Correlations										
	Item	N	M	(SE)	SD	Omega (90% CI)	PBG	PBL	PBM	PWL	PWM	NRG	NRL	NRM	Positive bond	Positive work	Negative relation
GQ positive bond:																	
Group (PBG)	5	369	5.63	0.06	1.07	0.846 (0.825–0.867)	—										
Leader (PBL)	4	369	6.10	0.05	1.05	0.892 (0.877–0.908)	0.540	—									
Member (PBM)	4	369	5.62	0.06	1.16	0.850 (0.829–0.871)	0.818	0.526	—								
GQ positive work:																	
Leader (PWL)	4	368	5.17	0.08	1.52	0.924 (0.913–0.935)	0.479	0.584	0.442	—							
Member (PWM)	4	360	4.04	0.09	1.70	0.948 (0.940–0.955)	0.621	0.429	0.619	0.646	—						
GQ negative relation:																	
Group (NRG)	3	369	2.93	0.06	1.19	0.640 (0.586–0.694)	–0.446	–0.203	–0.399	–0.185	–0.299	—					
Leader (NRL)	3	369	2.60	0.07	1.37	0.762 (0.727–0.798)	–0.460	–0.652	–0.442	–0.461	–0.396	0.321	—				
Member (NRM)	3	367	3.04	0.07	1.44	0.767 (0.732–0.802)	–0.559	–0.435	–0.608	–0.304	–0.432	0.410	0.591	—			
GQ global scores ^a :																	
Positive bond	13	369	75.05	0.64	12.38	0.924 (0.914–0.933)	0.908	0.785	0.908	0.575	0.641	–0.405	–0.593	–0.618	—		
Positive work	8	369	36.96	0.61	11.80	0.933 (0.924–0.942)	0.604	0.542	0.580	0.897	0.917	–0.262	–0.467	–0.401	0.662	—	
Negative relation	9	369	25.68	0.50	9.56	0.824 (0.801–0.847)	–0.618	–0.553	–0.614	–0.404	–0.478	0.698	0.818	0.860	–0.686	–0.481	
Clinical measures:																	
CSS	5	367	3.20	0.03	0.53	0.811 (0.785–0.837)	0.469	0.496	0.410	0.588	0.463	–0.149	–0.418	–0.324	0.525	0.527	–0.382
WSAS	5	368	21.15	0.44	8.46	0.849 (0.829–0.870)	–0.244	–0.282	–0.296	–0.215	–0.231	0.171	0.242	0.236	–0.316	–0.245	0.273
PHQ-9	9	368	15.53	0.31	5.95	0.852 (0.833–0.871)	–0.192	–0.297	–0.276	–0.198	–0.201	0.143	0.268	0.223	–0.294	–0.217	0.268
GAD-7	7	368	12.41	0.26	4.91	0.859 (0.841–0.878)	–0.173	–0.247	–0.220	–0.196	–0.140	0.157	0.259	0.170	–0.246	–0.180	0.247

Note. Correlations above ± 0.120 are significant at the 0.01 level (two-tailed). The correlations, in italics, between subscales and the global scores are expected to be high due to the computations, for example, *Positive Bond* is operationalized by PBG, PBL, and PBM. PBG = positive bond group; PBL = positive bond leaders; PBM = positive bond members; PWL = positive work leaders; PWM = positive work members; NRG = negative relationship group; NRL = negative relationship leaders; NRM = negative relationship members; GQ= Group Questionnaire; CSS = Client Satisfaction Scale; WSAS = Work and Social Adjustment Scale; PHQ-9 = Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder.

^a Mean of sum scores (not mean scores) of *Positive Bond*, *Positive Work*, and *Negative Relationship*.

Missing GQ Items

Among those who had completed GQ, some had missing GQ items. In all, 52 patients had missing items (14%), among whom 32 (61.5%) had one missing item and 5 (9.6%) had two missing items. Two patients had the maximum number of six missing items on GQ. Due to the low frequency and lack of a systematic pattern (no specific item had more missing response than others), they were considered random and of no threat to the validity of inferences from the present study. The estimation method for the CFA (MLM) requires exclusion of cases with missing items. Accordingly, the total number of observations for the CFA was 317. For the computation of scale scores, mean scores of items within each scale were computed regardless of missing items. In this way, the value of missing items is estimated to the mean of item responses within the rest of the scale.

Missing 1-Year Assessments

Among the 57% with missing 1-year assessments, 84% lacked both patient and therapist reports, indicating that the main reason for missing data was administrative, failing routines at the 1-year assessment.

In order to investigate possible sample bias, chi-square tests based on 2×2 tables were conducted to compare missing diagnosis between those with ($n = 369$) and those without a valid GQ at the 1-year assessment ($n = 483$). Among participants without GQ, 15% ($n = 72$) lacked all information on PD diagnosis, compared to 8% ($n = 29$) of participants with GQ ($p < .01$). Further investigation of possible differences in diagnostic distribution of PD (chi-square tests) revealed that paranoid PD (PaPD) was diagnosed more frequently among participants who lacked GQ (11%, $n = 44$) compared to those with GQ (6%, $n = 19$, $p < .05$), and likewise, that obsessive-compulsive PD (OCPD) was also more frequent among those who lacked GQ (10%, $n = 39$) compared to those with GQ (4%, $n = 15$, $p < .01$). Diagnostic information on symptom disorders was lacking for 37% ($n = 178$) of the participants without GQ, compared to 23% ($n = 84$) with GQ ($p < .001$). Among participants with information on symptom disorders, differences between the two groups were not significant ($p > .05$). Thus,

the present study sample with GQ data comprises a generally larger number of participants with information on diagnostic assessment, but more specifically, somewhat fewer patients with PaPD and OCPD than its comparable counterpart (those without GQ). Neither PD represented a dominating PD in this sample.

To further analyze possible impact of some bias in the frequency of PaPD and OCPD, independent t -tests were conducted in the study sample in order to compare mean scores on the eight GQ subfacets for participants with and without these diagnoses. No differences were found for seven of the subfacets, but participants with PaPD had higher scores on the subfacet *Negative Relationship* to the leader (NRL; 3.40, SD 1.48 vs. 2.59, SD 1.37, $p < .05$). Comparison of those with and without OCPD revealed no significant differences in subfacet scores ($p > .05$).

Finally, comparison of baseline data in the study sample ($n = 369$) with those who lacked 1-year assessment ($n = 483$) revealed no significant differences in the frequency of self-mutilation, suicide attempts, or levels of self-reported depression, anxiety, or social functioning ($p < .05$). We conclude that such bias in the present study sample is therefore minimal.

Nested Data

Data from the present study are from multiple sites, with patients nested within groups, and groups nested within sites. A limitation in the design is that the nesting structure is not registered. The factor analysis is not adjusted for the nesting structure and must be interpreted with this in mind. An analysis of measurement invariance (MI) between the 14 different units was initiated, the model did not converge as the number of parameters to be estimated (means, variances, and covariances) is greater than the number of observations within most of the groups of units. The intraclass correlation coefficients (ICCs) for all 30 items of GQ across the 14 units varied between 0.002 and 0.100 ($M = 0.021$, $SD = 0.022$). In other words, site differences contributed only a small part of the total variance.

Ethics

All participating patients from each treatment unit gave their written consent to use anonymous, clinical data for research purposes. Anonymized

data from each treatment unit were collected and transferred to a common research database. The collection procedures were approved by local data protection officer for each contributing unit. Data security procedures for the research database were approved by data protection officer at the responsible center for the research (Network for Personality disorders, Section for Personality Psychiatry and specialized treatments, Oslo University Hospital). Since the data are anonymous, formal approval from the Norwegian State Data Inspectorate and Regional Committee for Medical Research and Ethics is not required.

Results

First-Order Three-Factor CFA of GQ

A confirmatory first-order factor analysis testing model fit of three domains, *Positive Bond*, *Positive Work*, and *Negative Relationship*, yielded poor fit: $\chi^2 = 1891.689$ ($df = 402$), RMSEA = .108 (90% CI [0.103–0.113]), CFI = .740, and TLI = .718. Poor fit was mostly due to an uneven covariance structure indicating multidimensionality and residual covariance of items within and between the three constructs. No cross-loading of items were causing misfit. Inter-correlations between the three latent variables were as follows—*Positive Bond* with *Positive Work*: .721 ($p < .001$), *Negative Relationship* with *Positive Bond* and *Positive Work*: $-.796$ ($p < .001$) and $-.598$ ($p < .001$), respectively.

First-Order Eight-Factor CFA of GQ

A confirmatory first-order factor analysis testing model fit of the eight subfacets yielded good model fit: $\chi^2 = 732.957$ ($df = 377$), RMSEA = .055 (90% CI [0.049–0.060]), CFI = .938, and TLI = .928, with no critical indications of model modifications (i.e., critical cross-loadings or residual covariance).

Second-Order CFA of GQ

In the second-order CFA, the eight first-order subfacets were modeled under three higher order latent constructs in the following way: *Positive Bond*—accounting for variance of positive bond related to the group, the leader, and the members; *Positive Work*—accounting for variance of

positive work related to the group and leader; and *Negative Relationship*—accounting for variance of negative relationships related to the group, leader, and members. Each of the eight subfacets are modeled to account for variance of their respective indicators (items). See Appendix and Figure 1 for the operationalization and conceptual model, respectively.

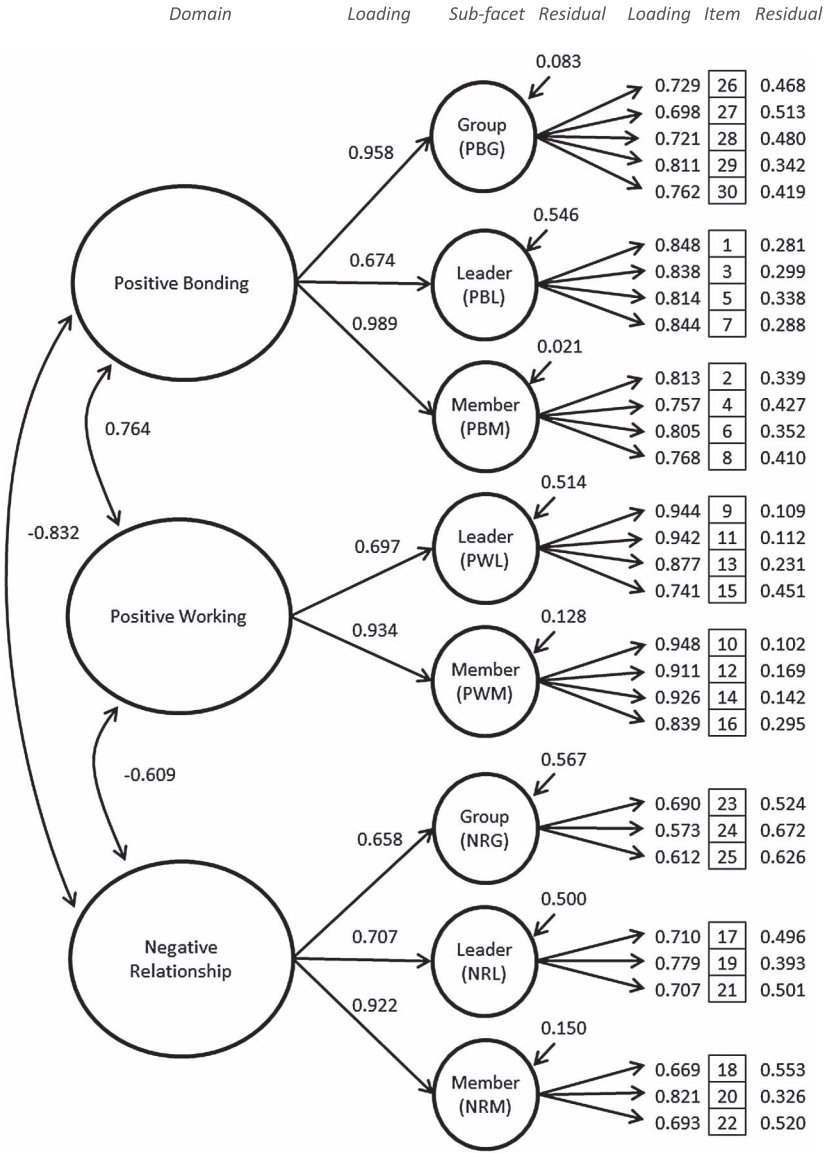
The initial analysis revealed a somewhat acceptable model fit: $\chi^2 = 904.446$ ($df = 394$), RMSEA = .064 (90% CI [0.058–0.069]), CFI = .911, and TLI = .901. The most prominent reason for model weakness was due to covariance between the latent first-order factors addressing the leader aspects within the three higher order construct: negative relation (NRL) with positive bond (PBL) and positive work (PWL; $r = -.741$ and $-.424$, respectively) and between positive bond (PBL) and positive work (PWL; $r = .764$). Second, model fit was weakened by residual covariance between Item 21 and 22: “*The group leaders did not always understand the way I felt inside*” and “*The other group members did not always understand the way I felt inside*” ($r = .493$) and between Item 4 and 30: “*The other group members and I respect each other*” and “*The group members accept one another*” ($r = .334$). Model modifications due to these constraints yielded good model fit: $\chi^2 = 688.426$ ($df = 389$), RMSEA = .049 (90% CI [0.043–0.055]), CFI = .948, and TLI = .941. See Figure 1 for conceptual model with parameter estimates.

Descriptive Statistics, Scale Reliability, and Correlations

Table 3 gives descriptive statistics, scale reliability of the GQ sub- and global scales, as well as correlations between the GQ scales. The sub-scales of GQ reveal highest mean scores for scales addressing positive bond, whereas scales addressing negative relationships show the lowest mean scores. Scale reliability was in the acceptable range for most scales, except for the GQ subscale addressing negative relationships to the group (NRG), with an Omega of 0.64 (Table 3).

As to zero-order correlations between the GQ scales, these were in the low to moderate level with the highest relation between positive bond between members and the group as a whole ($r = .818$, $p < .001$). The weakest relation was found between positive work with leader and

Figure 1
Conceptual GQ Model With CFA Parameters



Note. GQ = Group Questionnaire; CFA = confirmatory factor analysis.

negative relations with the group as a whole ($r = -.203, p < .001$). The zero-order correlations between the three global dimensions of GQ were in the moderate to high range (Table 3). Measures of patients satisfaction with the treatment (CSS) was at large moderately associated with the measures of GQ, with strongest relation to Positive Work with leaders and generally Positive

Bond. Weakest association was found between CSS and Negative Relationship with the group (Table 3). Associations between measures of depression (PHQ-9), anxiety (GAD-7), and social functioning (WSAS) were generally in the low range, but in expected direction, that is, higher levels of distress correlated negatively with the positive facets of GQ, and vice versa (Table 3).

Discussion

Factor Analysis

The first CFA testing a model of the three higher order dimensions, *Positive Bond*, *Positive Work*, and *Negative Relationship*, yielded poor fit according to conventional standards. These findings could be used to argue against the incremental validity of individual subscale interpretations. Further support for this argument comes from the moderate to high correlations between the latent constructs, a finding which partially replicates findings from Janis et al. (2018). Stated differently, moderate to high subscale correlations might lower the incremental subscale validity given high common variance.

However, a novel contribution of this article, compared to past GQ psychometric explorations, was the ordering of GQ items with subfacets. More specifically, the subfacets take into consideration the complex relationship structure found in group treatment (leader, members, and group) and fully cross this relationship structure by the three GQ quality subscales (*Positive Bond*, *Positive Work*, and *Negative Relationship*). The second CFA based upon these eight subfacets yielded acceptable CFA model fit producing a more parsimonious model. Finally, the third CFA (full model) constituted an even more parsimonious model in support of the theoretical conceptualization behind GQ, that is, replicating the three higher order dimensions when by considering the eight subfacets.

The residual covariance between Item 21 of NRL and Item 22 of negative relationship members (NRM) is a typical method effect stemming from content overlap. Both items contain the same formulation “... *did not always understand the way I felt inside*,” but refer to different targets, that is, group leader and group members, respectively. Since these items are organized under different subfacets, it would have no impact on the individual subscale interpretation but might somewhat increase the observed covariance. Both Items 21 and 22 are included in the *Negative Relationship* subscale and their residual covariance had minor impact for model fit compared to other item residual covariance ($r = 0.269, p < .001$). The same content overlap occurred with Items 4 and 30 by the formulations: “... *group members and I respect each other*”

and “*The group members accept one another*,” respectively. Like Items 21 and 22, these two items are organized into different subfacets (positive bond members [PBM] and positive bond group [PBG], respectively) but are included in the *Positive Bond* subscale and their residual variance of 0.319 ($p = .001$) and an insignificant impact on overall model fit.

The results of the second-order CFA (full model) was highly comparable to model fit indices and parameter estimates from former studies of GQ (e.g., Bormann et al., 2011; Giannone et al., 2020; Janis et al., 2018; Krogel et al., 2013; Thayer & Burlingame, 2014). Some model modifications were applied to improve model fit, but the nature and magnitude of these does not threaten the validity of the conceptual model. After all, the model fit indices from the initial analysis yielded estimates within a less strict acceptable range. Covariance of item residuals is usually a method effect and depending on the magnitude, can be viewed as a necessary scale improvement. Strong associations between latent constructs within a single model, as seen between the “leader” aspects of GQ, can be explained by characteristics of the study sample, commonality among item contents, or both. Findings like this are nearly unavoidable in the assessment of clinical or social constructs addressing opinions, thoughts, and feelings. Thus, stringent constraints on CFA (i.e., no residual covariance or cross-loadings) can lead to misplaced critique or even rejection of a theoretically sound model (see Cole et al., 2007; Hopwood & Donnellan, 2010; Sellbom & Tellegen, 2019).

All in all, the present study supports the conceptual model proposed by Johnson et al. (2005) in a sample representative of patients with PD or severe personality problems. Although the second-order CFA of GQ supports the theoretical model of Johnson et al. (2005), an operationalization of these three dimensions should be interpreted with caution due to multidimensionality. The application of the eight facets should be preferable until the global scores are further validated.

Scale Levels of GQ and Their Intercorrelations

Mean scores of the three higher order dimensions of GQ were in the same range as found in former studies (e.g., Giannone et al., 2020; Janis

et al., 2018; Krogel et al., 2013). The measure of *Positive Bond* is, on average, around 75, *Negative Relationship* around 25, and *Positive Work* somewhere in between. This is also reflected by scores on the eight subfacets of GQ. There are reasons to believe that these estimates are from group therapies that are functioning well, and that these estimates can serve as basis of comparisons at an individual level. In this respect, it would be appropriate to reach an agreement on how to compute the dimensions and subfacets of GQ.

Each of the eight subfacets of GQ is operationalized by quite few indicators (items), and especially the three aspects of *Negative Relationship*, with only three items each. Traditionally, a scale with three items and high interitem correlations would normally indicate that the items are addressing much of the same, and thus measure some narrow aspect only. However, by evaluating the item content of the scales of GQ (Appendix), this is not so obviously the case. Moreover, compared to the other subfacets, the three aspects of *Negative Relationship* had considerably less variance. Thus, the scale reliability of these might be underestimated due to range restrictions (Fife et al., 2012), as also found by Krogel et al. (2013).

GQ and Associations to Clinical Distress and Overall Patient Satisfaction

As assumed, concurrent subjective distress had only minor, although statistically significant, associations with measures of group relationship. However, it adds to the validity of GQ that higher levels of distress were positively associated with negative group relationships, and negatively associated with positive group relationship. Furthermore, positive group relation was associated with higher levels of patients' concurrent overall satisfaction with their treatment, and negative group relations were associated with dissatisfaction of treatment. Although some of these associations are somewhat general, the results taken together, add to the validity of GQ. For more nuanced studies, application of the eight facets of GQ is preferable.

GQ as a Relevant Instrument for PD Research and Treatment Monitoring

The currently supported factorial validity and psychometric properties of GQ in this study may

facilitate future research on change processes in treatments for patients with PD. The group component is often considered a cornerstone in PD treatment but is yet poorly documented (Storebø et al., 2020). A study of outpatient group psychotherapy for patients with different PDs suggested satisfactory group processes for the majority of patients, with the exception of unstable patients with borderline PD (Kvarstein et al., 2017). Other studies have likewise indicated problems in the group process for borderline PD patients (Cloitre & Koenen, 2001; Hummelen et al., 2007). A recent study suggested high sensitivity to social exclusion as a predictor of poorer group alliance (Euler et al., 2018). Although not specifying group processes, a recent PD study highlighted interactions between the mutual agreement on aims and tasks in therapy and positive outcomes (Folmo et al., 2021). Further investigation of therapist interventions, specific patient, and alliance factors has indeed been recommended in this field of research (Barnicot et al., 2012). In a recent study, information on group therapy alliance using GQ was tested as a feature of clinical feedback during the treatment process (Burlingame, Whitcomb, et al., 2018). In this study, therapeutic relationship predicted improvement in outcome and the feedback during the process reversed relationship deterioration. Such positive results are of expanded interest if replicated in studies of PD treatment.

Strengths and Limitations

A strength of this study is the data come from a large and representative sample of patients with PDs and PD traits, where borderline and avoidant PD are the most prevalent diagnoses. A reasonable methodological concern is our inability to test the within- and between-group levels due to missing information on group membership. However, since the analyses only address relationships between measures at a specific point in time, we do not consider this a major threat to the validity of the findings. In a longitudinal or effect study, this would be more serious. A substantial amount of missing data might represent a possible bias affecting the results and inferences from this study, but our analysis of this possible bias did not confirm this concern. Further support is found in the replication of factor analyses reported by former studies leading us to conclude that the findings from this study may be

generalized to other comparable patient samples. However, the parameter novel use of subfacet estimates call for replication in future research since the findings may be unique to PD samples or the methods employed by this study. Furthermore, there is a possibility that estimated subfacet scores on NRL might be underestimated due to fewer patients with PaPD in the present study sample.

Conclusion

From the overall findings of this study, it is reasonable to conclude that GQ constitutes a theoretically coherent and reliable measure of group relations, also applicable in a sample of PD patients. By this, the conditions are laid for further studies of validity and clinical utility of GQ in group-based treatment of patients with PD. A reliable measure enables further and much needed process–outcome research in group-based treatment of PD. Such research may importantly engender further understanding of appropriate treatment durations, frameworks and structuring elements in group therapies, and comprehensive treatments combining modalities for a population of poorly functioning patients. In particular, process–outcome research in sufficiently powered studies may increase the evidence base concerning treatment differentiation according to type of personality problems, overall levels of personality functioning, capacity for social cognition, and maladaptive styles of attachment.

References

- Allredge, C., Burlingame, G., Olsen, J., & Van Epps, J. (2020). Outcome Questionnaire-45 (OQ-45) progress alert rates in group versus individual treatment: An archival replication. *Group Dynamics*, 24(4), 247–260. <https://doi.org/10.1037/gdn0000121>
- Allredge, C. T., Burlingame, G. M., Yang, C., & Rosendahl, J. (2021). Alliance in group therapy: A meta-analysis. *Group Dynamics*, 25(1), 13–28. <https://doi.org/10.1037/gdn0000135>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
- 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. <https://doi.org/10.1080/10503300902894430>
- Barnicot, K., Katsakou, C., Bhatti, N., Savill, M., Fearn, N., & Priebe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder: A systematic review. *Clinical Psychology Review*, 32(5), 400–412. <https://doi.org/10.1016/j.cpr.2012.04.004>
- Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a model for assessing level of personality functioning in DSM-5, part I: A review of theory and methods. *Journal of Personality Assessment*, 93(4), 332–346. <https://doi.org/10.1080/00223891.2011.583808>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bormann, B., Burlingame, G. M., & Strauß, B., (2011). Der gruppenfragebogen (GQ-D): Instrument zur messung von therapeutischen beziehungen in der gruppenpsychotherapie [The group questionnaire (GQ-D). Instrument to measure therapeutic relationships in group psychotherapy]. *Psychotherapeut*, 56(4), 297–309. <https://doi.org/10.1007/s00278-011-0841-4>
- Bormann, B., & Strauß, B. (2007). Gruppenklima, Kohäsion, Allianz und Empathie als Komponenten der therapeutischen Beziehung in Gruppenpsychotherapien - Überprüfung eines Mehrebenen-Modells. *Gruppenpsychotherapie und Gruppendynamik*, 43(1), 3–22. <https://doi.org/10.13109/grup.2007.43.1.3>
- Burlingame, G. M., Allredge, C. T., & Arnold, R. A. (2021). Alliance Rupture detection and repair in Group Therapy: Using the Group Questionnaire-GQ. *International Journal of Group Psychotherapy*, 71(2), 338–370. <https://doi.org/10.1080/00207284.2020.1844010>
- Burlingame, G. M., Furhman, A., & Johnson, J. (2002). Cohesion in group psychotherapy. In J. Norcross (Ed.), *A guide to psychotherapy relationships that work* (pp. 71–88). Oxford University Press.
- Burlingame, G. M., Gleave, R., Beecher, M., Griner, D., Hansen, K., & Jensen, J. (2016). *Administration and scoring manual for the Group Questionnaire-GQ*. OQ Measures.
- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 110–131). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199737208.003.0005>
- Burlingame, G. M., McClendon, D. T., & Yang, C. (2018). Cohesion in group therapy: A meta-analysis. *Psychotherapy: Theory, Research, & Practice*, 55(4), 384–398. <https://doi.org/10.1037/pst0000173>
- Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., & Strauss, B. (2016). Outcome differences between individual and group formats when identical and

- nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy: Theory, Research, & Practice*, 53(4), 446–461. <https://doi.org/10.1037/pst0000090>
- Burlingame, G. M., & Strauss, B. (2021). Efficacy of small group treatments: Foundation for evidence-based practice. In M. Barkham, L. Castonguay, & W. Lutz (Eds.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (7th ed., pp. 583–624). Wiley.
- Burlingame, G. M., Strauss, B., & Joyce, A. S. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (pp. 640–690). John Wiley.
- Burlingame, G. M., Whitcomb, K. E., Woodland, S. C., Olsen, J. A., Beecher, M., & Gleave, R. (2018). The effects of relationship and progress feedback in group psychotherapy using the Group Questionnaire and Outcome Questionnaire-45: A randomized clinical trial. *Psychotherapy: Theory, Research, & Practice*, 55(2), 116–131. <https://doi.org/10.1037/pst0000133>
- 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). Guilford Press.
- Cloitre, M., & Koenen, K. C. (2001). The impact of borderline personality disorder on process group outcome among women with posttraumatic stress disorder related to childhood abuse. *International Journal of Group Psychotherapy*, 51(3), 379–398. <https://doi.org/10.1521/ijgp.51.3.379.49886>
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12(4), 381–398. <https://doi.org/10.1037/1082-989X.12.4.381>
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Dion, K. L. (2000). Group cohesion: From “field of forces” to multidimensional construct. *Group Dynamics*, 4(1), 7–26. <https://doi.org/10.1037/1089-2699.4.1.7>
- Euler, S., Wrege, J., Busmann, M., Lindenmeyer, H. J., Sollberger, D., Lang, U. E., Gaab, J., & Walter, M. (2018). Exclusion-proneness in borderline personality disorder inpatients impairs alliance in mentalization-based group therapy. *Frontiers in Psychology*, 9, Article 824. <https://doi.org/10.3389/fpsyg.2018.00824>
- Fife, D. A., Mendoza, J. L., & Terry, R. (2012). The assessment of reliability under range restriction: A comparison of α , ω , and Test–Retest Reliability for dichotomous data. *Educational and Psychological Measurement*, 72(5), 862–888. <https://doi.org/10.1177/0013164411430225>
- First, M. B., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (2015). *Structured clinical interview for DSM-5 (R) personality disorders (SCID-5-PD)*. American Psychiatric Publishing.
- Folmo, E. J., Stänicke, E., Johansen, M. S., Pedersen, G., & Kvarstein, E. H. (2021). Development of therapeutic alliance in mentalization-based treatment—Goals, Bonds, and Tasks in a specialized treatment for borderline personality disorder. *Psychotherapy Research*, 31(5), 604–618. <https://doi.org/10.1080/10503307.2020.1831097>
- Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personality Disorder and Emotion Dysregulation*, 4(1), Article 9. <https://doi.org/10.1186/s40479-017-0062-8>
- Giannone, F., Guarnaccia, C., Gullo, S., Di Blasi, M., Giordano, C., Lo Coco, G., & Burlingame, G. (2020). Italian adaptation of the Group Questionnaire: Validity and factorial structure. *Research in Psychotherapy*, 23(2), Article 443. <https://doi.org/10.4081/ripppo.2020.443>
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14(3), 332–346. <https://doi.org/10.1177/1088868310361240>
- Hornsey, M. J., Dwyer, L., & Oei, T. P. (2007). Beyond cohesiveness: Reconceptualizing the link between group processes and outcomes in group psychotherapy. *Small Group Research*, 38(5), 567–592. <https://doi.org/10.1177/1046496407304336>
- Hornsey, M. J., Dwyer, L., Oei, T. P., & 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. <https://doi.org/10.1521/ijgp.2009.59.2.267>
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 36(2), 223–233. <https://doi.org/10.1037/0022-0167.36.2.223>
- 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), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hummelen, B., Wilberg, T., & Karterud, S. (2007). Interviews of female patients with borderline personality disorder who dropped out of group psychotherapy. *International Journal of Group Psychotherapy*, 57(1), 67–92. <https://doi.org/10.1521/ijgp.2007.57.1.67>
- Hutsebaut, J., Feenstra, D. J., & Kamphuis, J. H. (2016). Development and preliminary psychometric

- evaluation of a brief self-report Questionnaire for the assessment of the DSM-5 level of personality functioning scale: The LPFS Brief Form (LPFS-BF). *Personality Disorders*, 7(2), 192–197. <https://doi.org/10.1037/per0000159>
- IBM. (2019). *IBM SPSS statistics for windows* (Version 26.0).
- Janis, R. A., Burlingame, G. M., & Olsen, J. A. (2018). Developing a therapeutic relationship monitoring system for group treatment. *Psychotherapy: Theory, Research, & Practice*, 55(2), 105–115. <https://doi.org/10.1037/pst0000139>
- Johnson, J., Burlingame, G. M., Olsen, D., Davies, R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology*, 52(3), 310–321. <https://doi.org/10.1037/0022-0167.52.3.310>
- Karterud, S., Pedersen, G., Bjordal, E., Brabrand, J., Friis, S., Haaseth, O., Haavaldsen, G., Irion, T., Leirvåg, H., Tørum, E., & Urnes, O. (2003). Day treatment of patients with personality disorders: Experiences from a Norwegian treatment research network. *Journal of Personality Disorders*, 17(3), 243–262. <https://doi.org/10.1521/pe di.17.3.243.22151>
- Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32(9), 509–515. <https://doi.org/10.3928/0048-5713-20020901-06>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Krogel, J. (2008). *The Group Questionnaire: A new measure of the group relationship*. Brigham Young University, BYU ScholarsArchive, 07-16. All Theses and Dissertations, Article 1732. <https://scholarsarchive.byu.edu/etd/1732>
- Krogel, J., Burlingame, G., Chapman, C., Renshaw, T., Gleave, R., Beecher, M., & Macnair-Semands, R. (2013). The Group Questionnaire: A clinical and empirically derived measure of group relationship. *Psychotherapy Research*, 23(3), 344–354. <https://doi.org/10.1080/10503307.2012.729868>
- Kvarstein, E. H., Folmo, E., Antonsen, B. T., Normann-Eide, E., Pedersen, G., & Wilberg, T. (2020). Social cognition capacities as predictors of outcome in mentalization-based treatment (MBT). *Frontiers in Psychiatry*, 11, Article 691. <https://doi.org/10.3389/fpsy.2020.00691>
- Kvarstein, E. H., Nordviste, O., Dragland, L., & Wilberg, T. (2017). Outpatient psychodynamic group psychotherapy-outcomes related to personality disorder, severity, age and gender. *Personality and Mental Health*, 11(1), 37–50. <https://doi.org/10.1002/pmh.1352>
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69(2), 159–172. <https://doi.org/10.1037/0022-006X.69.2.159>
- Lese, K. P., & MacNair-Semands, R. R. (2000). The therapeutic factors inventory: Development of a Scale. *Group*, 24(4), 303–317. <https://doi.org/10.1023/A:1026616626780>
- Lorentzen, S., & Ruud, T. (2014). Group therapy in public mental health services: Approaches, patients and group therapists. *Journal of Psychiatric and Mental Health Nursing*, 21(3), 219–225. <https://doi.org/10.1111/jpm.12072>
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 46(3), 266–274. <https://doi.org/10.1097/MLR.0b013e318160d093>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- MacKenzie, K. R. (1983). The clinical application of group measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159–170). International Universities Press.
- Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., Gega, L., Cameron, R., & Marks, I. M. (2005). Reliability and validity of the work and social adjustment scale in phobic disorders. *Comprehensive Psychiatry*, 46(3), 223–228. <https://doi.org/10.1016/j.comppsy.2004.08.007>
- McDonald, R. (1999). *Treat theory: A unified treatment*. Lawrence Erlbaum.
- McLaughlin, S. P. B., Barkowski, S., Burlingame, G. M., Strauss, B., & Rosendahl, J. (2019). Group psychotherapy for borderline personality disorder: A meta-analysis of randomized-controlled trials. *Psychotherapy: Theory, Research, & Practice*, 56(2), 260–273. <https://doi.org/10.1037/pst0000211>
- Morken, K. T. E., Binder, P. E., Arefjord, N. M., & Karterud, S. W. (2019). Mentalization-based treatment from the patients' perspective—What ingredients do they emphasize? *Frontiers in Psychology*, 10, Article 1327. <https://doi.org/10.3389/fpsyg.2019.01327>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180(5), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed., pp. 1998–2012). Muthén & Muthén.
- Newton-Howes, G., Tyrer, P., Anagnostakis, K., Cooper, S., Bowden-Jones, O., & Weaver, T.,

- (2010). The prevalence of personality disorder, its comorbidity with mental state disorders, and its clinical significance in community mental health teams. *Social Psychiatry and Psychiatric Epidemiology*, 45(4), 453–460. <https://doi.org/10.1007/s00127-009-0084-7>
- Obeid, N., Carlucci, S., Brugnera, A., Compare, A., Proulx, G., Bissada, H., & Tasca, G. A. (2018). Reciprocal influence of distress and group therapeutic factors in day treatment for eating disorders: A progress and process monitoring study. *Psychotherapy: Theory, Research, & Practice*, 55(2), 170–178. <https://doi.org/10.1037/pst0000138>
- Orfanos, S., Burn, E., Priebe, S., & Spector, A. (2020). A systematic review and quality assessment of therapeutic group process Questionnaires. *International Journal of Group Psychotherapy*, 70(3), 425–454. <https://doi.org/10.1080/00207284.2020.1755292>
- Pedersen, G., Karterud, S., Hummelen, B., & Wilberg, T. (2013). The impact of extended longitudinal observation on the assessment of personality disorders. *Personality and Mental Health*, 7(4), 277–287. <https://doi.org/10.1002/pmh.1234>
- Pedersen, G., Kvarstein, E. H., & Wilberg, T. (2017). The Work and Social Adjustment Scale: Psychometric properties and validity among males and females, and outpatients with and without personality disorders. *Personality and Mental Health*, 11(4), 215–228. <https://doi.org/10.1002/pmh.1382>
- Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428–1441. <https://doi.org/10.1037/pas0000623>
- Sheehan, D. V., & Lecrubier, Y. (1994). *Mini international neuropsychiatric interview (M.I.N.I.)*. University of South Florida Institute for Research in Psychiatry/INSERM-Hôpital de la Salpêtrière.
- Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*, 24(5), 399–411. [https://doi.org/10.1016/0010-440X\(83\)90032-9](https://doi.org/10.1016/0010-440X(83)90032-9)
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Steiger, J. H. (1990). Structural model evaluation and modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, 25(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. <https://doi.org/10.1016/j.paid.2006.09.017>
- Storebø, O. J., Stoffers-Winterling, J. M., Völlm, B. A., Kongerslev, M. T., Mattivi, J. T., Jørgensen, M. S., Faltinsen, E., Todorovac, A., Sales, C. P., Callesen, H. E., Lieb, K., & Simonsen, E. (2020). Psychological therapies for people with borderline personality disorder. *Cochrane Database of Systematic Reviews*, 5(5), Article CD012955. <https://doi.org/10.1002/14651858.CD012955.pub2>
- Thayer, S. D., & Burlingame, G. M. (2014). The validity of the group Questionnaire: Construct clarity or construct drift? *Group Dynamics*, 18(4), 318–332. <https://doi.org/10.1037/gdn0000015>
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, Article 769. <https://doi.org/10.3389/fpsyg.2016.00769>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. <https://doi.org/10.1007/BF02291170>
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation Process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>
- Winsper, C., Bilgin, A., Thompson, A., Marwaha, S., Chanen, A. M., Singh, S. P., Wang, A., & Furtado, V. (2020). The prevalence of personality disorders in the community: A global systematic review and meta-analysis. *The British Journal of Psychiatry*, 216(2), 69–78. <https://doi.org/10.1192/bjp.2019.166>

Appendix

Organization of Group Questionnaire (GQ) Items

Positive Bond

Group (PBG)—“Climate”

26. The members liked and cared about each other.

27. The members felt what was happening was important and there was a sense of participation.

28. We cooperate and work together in group.

29. Even though we have differences, our group feels secure to me.

30. The group members accept one another.

(Appendix continues)

Leaders (PBL)—“Alliance”

1. I felt that I could trust the group leaders during today’s session.
3. The group leaders and I respect each other.
5. I feel the group leaders care about me even when I do things that they do not approve of.
7. The group leaders were friendly and warm toward me.

Members (PBM)—“Cohesion”

2. I felt that I could trust the other group members during today’s session.
4. The other group members and I respect each other.
6. I feel the other group members care about me even when I do things that they do not approve of.
8. The other group members were friendly and warm toward me.

Positive Work***Leaders (PWL)—“Task/Goals”***

9. The group leaders and I agree about the things I will need to do in therapy.
11. The group leaders and I agree on what is important to work on.
13. The group leaders and I have established an understanding of the kind of changes that would be good for me.
15. The group leaders and I are working together toward mutually agreed upon goals.

Members (PWM)—“Task/Goals”

10. The other group members and I agree about the things I will need to do in therapy.
12. The other group members and I agree on what is important to work on.

14. The other group members and I have established a good understanding of the kind of changes that would be good for me.

16. The other group members and I are working together toward mutually agreed upon goals.

Negative Relationship***Group (NRG)—“Conflict”***

23. There was friction and anger between the members.
24. The members were distant and withdrawn from each other.
25. There was tension and anxiety between the members.

Leaders (NRL)—“Alliance Rupture”

17. Sometimes the group leaders did not seem to be completely genuine.
19. The group leaders did not always seem to care about me.
21. The group leaders did not always understand the way I felt inside.

Members (NRM)—“Empathic Failure”

18. Sometimes the other group members did not seem to be completely genuine.
20. The other group members did not always seem to care about me.
22. The other group members did not always understand the way I felt inside.

Received April 13, 2021

Revision received September 7, 2021

Accepted September 8, 2021 ■