

**A systematic review and quality assessment of therapeutic group process
questionnaires**

^aStavros Orfanos, PhD

^bErin Burn, MSc

^bStefan Priebe, Dipl.-Psych, Dr. med. habil, FRCPsych

^aAimee Spector, PhD, DCLinPsy

^aUniversity College London, Clinical, Education and Health Psychology, Division of Psychology and Language Sciences, London, United Kingdom

^bUnit for Social and Community Psychiatry, Queen Mary University of London, London, United Kingdom

*Requests for reprints should be addressed to Stavros Orfanos, University College London, Clinical, Education and Health Psychology, Division of Psychology and Language Sciences, 1-19 Torrington Place, London, United Kingdom, WC1E 7HB, stavros.orfanos.16@ucl.ac.uk

Abstract

The aims of this study were to systematically review the literature on questionnaire-based group process measures and assess the psychometric scale development properties of these measures. A systematic search of questionnaire-based measures of group process was conducted from four databases: PsycINFO, Medline, EMBASE and AMED. A quality assessment was undertaken using established criteria. Only studies that described the development of questionnaire-based scales were included. Seventeen studies which described thirteen measures were included in the analysis. The developed list highlights a range of questionnaires that can be used by clinicians and researchers seeking to measure either 'specific therapeutic group mechanisms' or 'overall group experiences'. Further research should focus on developing the content validity of these scales by involving the target population during item selection and reduction.

Keywords: Group therapy, group process, measures, psychometric assessment, questionnaire

A Systematic Review and Quality Assessment of Therapeutic Group Process Questionnaires

Psychological therapies are often delivered in groups involving two or more individuals. Due to increasing demands for inexpensive and accessible therapies (McCrone, 2008), groups are becoming an increasingly common approach in community and acute inpatient mental health settings (Burlingame, 2014; Lo Coco et al., 2019). In addition to the pragmatic benefits of groups, factors inherent to a group environment are also valued on a clinical level, where social aspects of the group setting are utilised as an agent of clinical change (Lo Coco et al., 2015). The mechanisms of therapeutic change inherent to the interaction between group members in group therapies are broadly described as ‘group processes’ (Garcia-Cabeza et al., 2011).

Burlingame and colleagues (2013) developed an organisational framework to conceptualise the role of group processes in small group treatments. In their model of group therapies, they distinguished between the group form (structure) and group function (processes). Burlingame and colleagues refer to group *form* as actions imposed on a group by the leader which influences the structure of the group, including the selection of group members and the composition of the group. Group *processes*, on the other hand, are considered as primary mechanisms of change associated with interactions between members. More specifically, Burlingame and colleagues (2013) highlight five components of ‘emergent’ processes describing member/leader interactions (including interpersonal feedback and self-disclosure, leader interventions) or by products of interactions (including cohesion and therapeutic factors) within their model. They also highlight the role of social psychological processes from organisational psychology principles that impact groups, including conformity, power, and management of conflict.

The measurement of group processes has important implications for both researchers and clinicians interested in monitoring and/or evaluating interactions between people in a

therapeutic group setting (Delucia-Waack, 1997; Jensen et al., 2012). In terms of research, the Medical Research Council Framework (Campbell et al., 2007) guidance recommends that randomised controlled trials (RCTs) of complex interventions map mechanisms of change linked to desired outcomes. RCTs of group therapies are therefore increasingly seeking to measure unaccounted outcome variance related to group processes in addition to theory-specific mechanisms common to individual therapy (for example, cognitive restructuring) (Priebe et al., 2013). In terms of clinical practice, therapists delivering routine therapy groups are encouraged to monitor group processes for their own audit and/or evaluative purposes (Marmarosh, 2018). This is in accordance with the literature highlighting the relationship between group processes and therapeutic outcomes (Chapman et al., 2012). A number of meta-analyses have concluded that group processes can be identified independently from the method-specific tasks implemented (Burlingame et al., 2004; Burlingame et al., 2013). However, despite the importance of measuring group processes, selecting an appropriate measure of group process is widely recognised as a challenging task (Thayer & Burlingame, 2014).

Within the group therapy literature, two broad categories of group process measures exist. The first category of measures are questionnaire-based tools that seek to measure group processes as phenomena, where “some aspect or characteristic of the member, leader or group behaviour” are described (Fuhriman & Burlingame, 1994b, p. 503). In doing so, the degree or intensity to which pre-defined group mechanisms are present or experienced is rated, typically on a Likert-type scale. Questionnaire-based measures of group phenomena are quick, inexpensive, and easy to use (Burlingame et al., 2006). They can either be self-reported, rated by the group leader, or rated by an independent observer. They can be used across multiple group sessions to either regularly track the progress of therapy or explore how group processes develop over time (Bakali et al., 2013).

Yalom's theoretical contributions on curative, or helpful, group factors have been highly influential within the field of group phenomena research, (Yalom & Leszcz, 2005; Yalom, 1985). Yalom described eleven beneficial mechanisms of change associated with clinical outcomes including: universality, altruism, instillation of hope, imparting information, corrective recapitulation of primary family experience, development of socialising techniques, imitative behaviour, cohesiveness, existential factors, catharsis, interpersonal learning-input, interpersonal learning-output/input, and self-understanding (Yalom & Leszcz, 2005). Measures developed to assess Yalom's therapeutic curative factors include the Therapeutic Factor Inventory (TFI) (Lese & MacNair-Semands, 2000) and TFI-Short (TFI-S) (MacNair-Semands, Ogrodniczuk, & Joyce, 2010). MacKenzie's construct of *group climate* (MacKenzie, 1983) is also a commonly assessed group process phenomena (Johnson et al., 2006) measured by the Group Climate Questionnaire. Unlike Yalom's curative factors, which describe individual helpful group experiences, the group climate refers to the group environment as a whole - including the degree to which members are engaged with each other, avoidant, or in conflict.

The second category of measures are behavioural coding tools, which are used to describe and annotate discrete interactive transactions between group members (Bales, 1950; Cahill et al., 2008; Delucia-Waack, 1997). The aim of this approach, also referenced in the literature as in-session coding (Delucia-Waack, 1997) and process analysis of interaction (Beck & Lewis, 2000), is to reliably measure classifications of behaviour occurring in a natural environment (Heath et al., 2010). Behavioural coding tools have the advantage of being exploratory and theory building (Fuhrman & Burlingame, 1994a). However, they are often laborious and time-intensive for clinicians and/or researchers seeking to monitor group processes or link group processes with outcomes. New technologies have been used to automatically annotate interactions in a clinical population, for example 3D motion caption

techniques (Lavelle et al., 2013). Whilst new technologies have much potential, they are currently limited to interaction laboratories. The present study therefore focused on questionnaire-based measures, which assess pre-defined group process phenomena. This includes questionnaires that can either be used by therapists, group members, or an independent observer.

Few attempts have been made to collate information on group process questionnaires for clinicians and/or researchers interested in selecting an appropriate measure. The American Group Psychotherapy Association compiled an evidence-based ‘self-evaluation kit,’ the CORE-R, aimed at aiding group therapists in their clinical practice (Strauss, Burlingame, & Bormann, 2008). This included a list of questionnaires that focused on measuring group therapeutic relationships (Johnson, Burlingame, Olsen, Davies, & Gleave, 2005). Measures included within the CORE-R assess both group members’ relationship with the therapist and with the group (Jensen et al., 2012). More specifically, the measures in the CORE-R were chosen to represent one of three basic components of group phenomena outlined by Johnson and colleague’s model of group therapeutic relationships (Johnson et al., 2005), including Positive Bond Relationship, Positive Working Relationship and Negative Working Relationship (Johnson et al., 2005). This model has been supported by six separate studies, including over 3000 participants from over 500 groups, across four countries (Burlingame, 2010).

More recently, Sodano and colleagues conducted a comprehensive review of group process, dynamics, and climate measures (Sodano et al., 2014). In their book chapter, they highlight a range of methodological approaches and tools available for measuring group processes. They also described the levels of reliability and validity for each approach. However, there are at least two important limitations within this literature. First, to the authors’ knowledge, there has been no attempt to review the literature on group process

measures using systematic methods to search and collect data. Hence, existing non-systematic reviews are potentially subject to bias and errors in the study selection process (Egger, Davey-Smith, & Altman, 2008). Second, no attempt has been made to assess the psychometric properties of studies describing the scale development of group process measures using standardised criterion. For example, Sodano and colleagues (2014) stated that measures included in their study must have at least a moderate level of reliability and construct validity. However, without explicit criterion definitions, what constitutes a moderate level of psychometric quality is not clear.

This study therefore sought to address these limitations by systematically reviewing and comprehensively appraising existing group process measures. In doing so, only articles that reported and/or evaluated the psychometric development of measures were included. Overall, the specific aims of the present study were to: develop an up-to-date summary of questionnaire-based group process measures used in a therapeutic group setting, and assess the psychometric properties of the identified questionnaire-based group process measures using an established set of quality criteria. This study sought to develop a list of measures that can be used by clinicians and researchers seeking to assess mechanisms of group process occurring in in group therapies.

Methods

Literature Search Strategy

The search process is summarised in Figure 1. A study protocol was developed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement guidance (Moher et al. 2009). A systematic search of published questionnaires, which sought to measure group processes used within a group therapy setting, was conducted. Systematic principles were followed for searching and screening. Three search term categories were used, including ‘measure’ AND ‘group process’ AND ‘group therapy.’ The search terms

included in the literature are outlined in the online supplementary Table 1. Search terms related to definitions of ‘group process’ sought to be as broad as possible. Whilst the group process search terms used included specific constructs related to the quality of therapeutic relationships—including cohesion alliance, climate and empathy—broader terms such as nonverbal behaviour, group experience, therapeutic relationship, and social interaction were also used. Four electronic databases were searched, including PsycINFO (1806 to December 2018), Medline (1946 to December 2018), EMBASE (1974 to December 2018), and AMED (1806 to December 2018). As part of the literature search a number of related journal articles were hand searched, including the *International Journal of Group Psychotherapy* and *Group Therapy*.

Studies were *included* if: (a) group process was measured in the context of a therapeutic intervention or treatment involving more than two people other than the group leader/co-facilitator; (b) reference was made to scale development or the evaluation of psychometric properties, including reliability and /or validity, in the title or abstract; (c) they were published in English; and (d) they were published in a peer-reviewed journal. Studies were *excluded* if: (a) they were a review article (Potentially relevant studies within identified reviews were included in the grey literature hand search.); (b) group process was measured in a specific context of family or couples therapy; (c) personal attributes of an individual in a group were measured (e.g., group expectations or group leadership); (d) moment-to-moment interactions from segments of group sessions were measured using behavioural coding tools (This included any studies that involved the use of digital technologies, including video recordings equipment and/or other annotation software.); (e) there was a sample size of less than 50 participants, in accordance to Terwee and colleagues’ (2007) established criteria on the psychometric properties of health outcome measures; and (f) they were only abstract publications or study protocols.

[INSERT FIGURE 1]

Data Extraction Procedure

Titles and abstracts of studies were retrieved using the stated search strategy and inclusion/exclusion criteria. The lead researcher (the first author) first screened the titles and abstracts. Following this, a research assistant (the second author) independently re-screened 50% of all the identified studies. Any ambiguity was resolved with the lead researcher's primary supervisor (the fourth author). The first and second authors then both independently reviewed all studies included in the full paper review stage using a structured format developed by the first author to extract relevant information. Again, any ambiguity was resolved with the fourth author. Extracted information included details on the group process measure, including type of measure, number of items, construct being measured, how to use measure, and details on the development of the measure. Any missing data were requested from study authors.

Appraisal of psychometric properties

The quality assessment was undertaken using established criteria which assessed the development of questionnaire-based measures used in health (Terwee et al., 2007). In particular, these criteria assess the procedures used to develop questionnaires. Eight criteria outlined by Terwee and colleagues (2007) were used to determine the methodological quality of studies aimed at developing and evaluating group process measures. This included: (a) content validity (i.e., the extent to which concepts of interest are comprehensively represented by items in the questionnaire); (b) internal consistency (i.e., the assessment of whether items in scale of questionnaire are inter-correlated); (c) construct validity (i.e., the extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses); (d) reproducibility-agreement (i.e., the extent to which the scores on repeated measures are close to each other); (e) reproducibility-

reliability (i.e., the extent to which patients can be distinguished from each other, despite relative measurement error); (f) responsiveness (i.e., the ability of a questionnaire to detect clinically important changes over time); (g) floor and ceiling effects (i.e., number of respondents who achieved the lowest or highest possible score); (h) interpretability (i.e., degree to which one can assign qualitative meaning to quantitative scores). Terwee and colleagues' (2007) criteria guidance on assessing development procedures are outlined in the online supplementary Table 2.

Each item within the criterion was scored depending on whether there was an adequate design and appropriate reporting of statistics. If criteria for a given psychometric property were met, a "positive rating" of 2 was given. If criteria were doubtful in design, where a clear description of the outlined aspects was lacking, or only partial criteria were met for a given psychometric property, an "intermediate rating" of 1 was given. If there was no information provided on a given criteria, or criteria were explicitly not met, a "no rating" of 0 was given. Total scores for each measure and each quality criteria category across all measures are given.

Two raters (the first and second authors) rated the quality based on the operationalized definitions outlined in the online supplementary Table 2. Each rater independently conducted a quality appraisal of all studies which met the relevant inclusion/exclusion criteria in the following four steps: first, both raters independently conducted a quality appraisal from three randomly selected studies; second, the two raters reviewed and discussed their ratings together to ensure 100% rater agreement; third, the two raters both independently reviewed all the remaining papers using Terwee and colleague's (2007) quality criteria; finally the two raters compared their ratings to ensure 100% agreement. Any ambiguity was resolved with the fourth author.

Results

A total of seventeen studies, which outlined twelve group process measures, met the inclusion criteria. An additional study/measure was identified during the hand searching stage. Therefore, a total of thirteen group process measures were reviewed and included in the psychometric assessment.

Six questionnaires focused on measuring helpful group processes associated with clinical outcomes and were categorised *Therapeutic Group process* measures. These measures included the Group Cohesiveness Scale (Wongpakaran et al., 2013), Therapeutic Factor Inventory (Joyce et al., 2011; Lese & MacNair-Semands, 2000; MacNair-Semands et al., 2010), Group Cohesion Scale (Treadwell et al., 2001), Group Attitude Scale (Evans & Jarvis, 1986), Curative Climate Instrument (Fuhrman et al., 1986) and the Scale for Evaluation of Group Counselling (Murillo et al., 1981). The characteristics of the six Therapeutic Group Process questionnaires are summarised in Table 1. Three versions of the Therapeutic Factor Inventory were identified and were, therefore, separately described in Table 1.

[INSERT TABLE 1]

The remaining seven questionnaires measured group processes more broadly, including both positive and negative group experiences, as well as individual and group level mechanisms. Hence these questionnaires were categorised as *Overall Group Process* measures. This included the Group Observational Measurement of Engagement (Cohen-Mansfield et al., 2017), Factors Aspecific and Specific in Group Therapy (Marogna & Caccamo, 2014), Social Exchange Scale (Brown et al., 2014), Ferrara Group Experiences Scale (Caruso et al., 2013), Group Questionnaire (Krogel et al., 2013), Group Sessions Rating Scale (Quirk et al., 2013) and Group Climate Questionnaire (MacKenzie, 1981; MacKenzie, 1983). The characteristics of the seven Overall Group Process questionnaires are summarised

in Table 2. Two versions of the Group Climate Questionnaire were identified and were therefore separately described in Table 2.

[INSERT TABLE 2]

Sixty-nine percent of studies described self-report questionnaires. The majority of questionnaires (61%) were rated on either a five-point or seven-point Likert-type scale. The identified questionnaires were developed across a diverse range of therapeutic orientations and clinical and non-clinical populations. The number of items ranged from four items (Group Cohesiveness Scale) to 79 items (Scale of the Evaluation of Group Counselling).

The assessment of the psychometric properties of the thirteen included measures are summarised in Table 3. Scores for each of the eight categories outlined by Terwee and colleagues (2007) are given, in addition to total ‘measure’ and total ‘category’ scores. Based on the total measure scores, the Group Cohesiveness Scale (Wongpakaran et al., 2013), Therapeutic Factor Inventory-Short (MacNair-Semands et al., 2010), Therapeutic Factor Inventory-19 (Joyce et al., 2011), Group Questionnaire (Krogel et al., 2013) and Group Sessions Rating Scale scored highest in terms of their scale development quality. All measures except for the Social Exchange Scale (Brown et al., 2014) demonstrated good internal consistency.

[INSERT TABLE 3]

Discussion

To the authors’ knowledge, the present study was the first to systematically review the literature on group process measures. Furthermore, this study was the first to use a clearly defined set of quality criteria to assess the psychometric properties of questionnaire-based group process measures. A list of measures that can be used by clinicians and/or researchers seeking to assess mechanisms of group process in group psychological therapies are presented. A total of thirteen questionnaire-based tools were organised into two categories.

The first category outlined measures which assessed beneficial mechanisms of change relevant to clinical outcomes. This category included six measures and was titled Therapeutic Group Process measures. The second category of measures assessed overall group processes beyond helpful mechanisms, including positive and negative group experiences. This category included seven measures and was titled Overall Group Process measures.

The analysis of the psychometric properties highlighted several categories of quality criteria that were not reported. Only one measure, the Social Exchange Scale (Brown et al., 2014), was given a positive rating for content validity, regarded by Terwee and colleagues (2007) as the most important psychometric property. Twelve measures were given a ‘no rating’ for the reproducibility agreement criterion, where no information was given, or criteria were explicitly not met. Therefore the ability for the identified measures to assess change over time and longitudinal validity is limited or unclear. This is important when considering the applicability of these measures when assessing therapeutic change. Furthermore, all measures scored a ‘no rating’ for the ‘floor and ceiling effects’ criterion, making interpretation of the content validity further difficult.

Measuring Therapeutic Group Processes

The findings from this study give insight into six measures that assess mechanisms of change inherent to beneficial, or helpful, group interactions. Three out of six measures within this category measured aspects of Yalom’s proposed curative factors (Yalom & Leszcz, 2005; Yalom, 1985). Therapeutic Factor Inventory-Short (TFI-S) (MacNair-Semands et al., 2010) and Therapeutic Factor Inventory-19 (Joyce et al., 2011) appear to be suitable instruments available for evaluating what group members find helpful about their group experience. Compared to the Group Cohesiveness Scale and the Curative Climate Instrument, the TFI-S and TFI-19 measure Yalom’s curative factors more comprehensively—i.e., they seek to measure a broader range of curative factors. In terms of the reported psychometric

properties, the TFI-S and TFI-19-item scored on par with the other measures identified. The TFI-S demonstrated good internal consistency, construct validity, responsiveness and interpretability and the TFI-19 demonstrated some evidence of reproducibility and responsiveness.

Within the group therapy literature, the TFI-S is valued as a reliable and simple tool (Strauss et al., 2008) and features in the AGPA's CORE-R battery (Burlingame et al., 2006; Strauss et al., 2008). However, the TFI-S has more items than several other questionnaires in this category. The shorter TFI-19 is, therefore, arguably a more accessible choice of measure. Furthermore, the TFI-19 (Joyce et al., 2011) has been validated cross-culturally in a Japanese population (Kageyama et al., 2016).

The Group Cohesiveness Scale (Wongpakaran et al., 2013) also appears to be a suitable measure of helpful group processes. It had the highest total psychometric quality score and was the only measure which had a positive criterion validity score. Furthermore, it consists of only seven items. However, this measure is specifically focused on measuring group engagement and is therefore possibly too limited to be used alone.

Measuring Overall Group Process

The findings from this study highlight seven questionnaires which measure Overall Group Process, including both supportive and challenging group experiences. The Group Questionnaire (GQ) (Krogel et al., 2013) is arguably the most suitable questionnaire available for researchers seeking to measure group processes beyond those experienced as curative. An important strength of this measure is that it was developed to measure Johnson and colleagues' empirically robust theoretical model of group relationships (Johnson et al., 2005). The GQ was designed to combine concepts of "interpersonal structure of relationships," including member-leader, member-member, and member-group; and the "quality of the therapeutic relationship" in groups, including group cohesion, alliance, group climate, and

empathy. Findings from the present study suggest that the GQ has good internal consistency and construct validity. These findings are also consistent with analyses from Thayer and Burlingame (2014), who also replicated the GQ's factor structure and correlated the GQ with measures from which it was originally developed.

The GQ is referenced in a previous review of measures related to group process, dynamics and interventions (Sodano et al., 2014). In this review, it is noted that population-specific norms from Krogel and colleagues' sample (Krogel et al., 2013)—including inpatient, non-clinical and counselling centre groups—allow for the comparison of GQ scores with relevant normative group scores. A more recent study further supported the GQ factor structure (Janis et al., 2018). Janis and colleagues (2018) used results from six GQ studies conducted across a diverse clinical and nonclinical sample of 2479 participants further supporting the implementation of unique GQ norms.

Findings from this current study highlight that popular measures are not necessarily psychometrically robust. The Group Climate Questionnaire-Short (GCQ-S) (MacKenzie, 1983) has been cited as the most commonly used and extensively validated group process measure in the group psychotherapy literature (Burlingame et al., 2004; Johnson et al., 2006; Strauss et al., 2008). Furthermore, like the TFI, it is also recommended in the AGPA's CORE-R battery of group process measures (Burlingame et al., 2006; Strauss et al., 2008). However findings from the present study highlight that the Group Climate Questionnaire (MacKenzie, 1981) and the Group Climate Questionnaire-Short (MacKenzie, 1983) have relatively low scale development quality scores compared to other available tools. The discrepancy between the GCQ-S' low psychometric properties and high popularity might be because the GCQ-S is an extremely user-friendly tool, consisting of only 12 items and can be flexibly used by either a group member, group leader, or observer (DeChavez et al., 2000).

Limitations

This present study is limited to questionnaire-based measures of group process phenomena and does not include moment-to-moment behavioural coding tools (Davis et al., 2000) or open-ended qualitative measures of group process (Strauss et al., 2008). Questionnaires are likely to be the most appropriate approach for clinicians who have busy practices, or researchers seeking to incorporate feedback on group processes during ongoing groups (Burlingame et al., 2015; Slone et al., 2015). However, behavioural coding tools (Bales, 1950; Cahill et al., 2008; Delucia-waack, 1997) have the advantage of reliably measuring fine-grain interactive behaviours often too subtle to the human eye. Furthermore, they can be measured and analysed using automated digital technologies (Heath et al., 2007). Compared to questionnaires, behavioural coding tools are a more bottom-up exploratory approach, where simple behavioural categories and/or statements can be combined or aggregated to form multiple variations of interactive behaviours (Davis et al., 2000). Hence this approach might be more suitable for researchers and/or clinicians seeking to identify less understood group processes (Fuhriman & Burlingame, 1994a).

A second limitation of the present study is that the appraisal of psychometric properties was limited to studies that reported scale development. To address this limitation, quality appraisal scores were given for newer revisions of scales, where reference was made to scale development or the evaluation of psychometric properties. This included separate psychometric quality scores for the GCQ and GCQ-Short, and the TFI, TFI-S, and TFI-19. In addition to this, the present study reported detailed separate quality assessment ratings for each of the eight quality criteria for all included measures, as recommended by Terwee and colleagues (2007).

A third methodological limitation to be considered is that a high number of studies did not meet inclusion criteria and were excluded during the screening stages. Whilst

systematic principles were followed for searching and screening potentially relevant studies, the search term strategy may have been too comprehensive. Identifying measures that met inclusion criteria was therefore a challenging task. To address this limitation, stringent inter-rater procedures were followed. A second rater re-extracted 50% of the studies initially identified and independently reviewed all studies included in the full-paper review stage. Furthermore, both the lead researcher and an independent researcher completed all quality ratings. Future researchers may benefit from refining the search strategy to reduce the number of irrelevant studies and minimise risks of error at the screening stages.

Implications

As noted by Lo Coco and colleagues (2015) there is a need to “bridge the gap between science and practice in group psychotherapy” (p.288). The results from the current study help address this need, by providing researchers and clinical practitioners with a systematic evaluative collation of group process measures for research and clinical use. To aid researchers and clinicians in their decision making process, the identified questionnaires are organised into Therapeutic Group Process measures and Overall Group Process categories. The Therapeutic Factor Inventory-Short (MacNair-Semands et al., 2010) and the Group Questionnaire (Krogel et al., 2013) are recommended questionnaires from these respective categories.

The results from this study demonstrate the importance of testing the psychometric properties of existing group process measures. To date, none of the identified studies referenced absolute measurement error—i.e., the degree to which repeated scores across similar measures are in line with each other. Furthermore, all included studies failed to define a minimal important change. Therefore evaluating the interpretability of the identified measures was a challenge—in particular, their applicability to assessing therapeutic change.

If these psychometric properties were evaluated, it is recommended that researchers report these important criteria.

It is also recommended that when selecting a group process measure, clinicians and researchers consider issues related to content validity—described as the degree to which group process concepts of interest are adequately represented in the chosen questionnaire (Terwee et al., 2007). For a positive rating of this criterion, Terwee and colleagues (2007) state that authors should highlight the involvement of the target population during item selection, in addition to providing a clear description of the measurement aim, target population and concepts. Whilst most authors gave a clear description of the measurement aim and concepts of the questionnaire, Brown and colleagues' (2014) were the only authors who explicitly reported that they involved the target population during the development of the Social Exchange Scale. Hence most studies scored zero for this quality criterion. It is not clear whether authors failed to involve the target population in the item selection/reduction phase of questionnaire development, or failed to report this. Nonetheless, future researchers seeking to use or adapt a given group process measure would benefit from ensuring that items are piloted with the clinical population of interest first.

Current guidance for measuring and monitoring group processes in group therapies recommend questionnaire-based instruments (Bernard et al., 2008). However, as noted, future researchers would benefit from reviewing relevant behavioural coding tools, which allow for a bottom-up exploration of less defined group processes. The present study gives insight into how behavioural coding tools can be systematically identified and evaluated in future studies—including strategies for searching, screening, and appraising results from the group therapy literature. Whilst qualitative methods are less commonly reported within the literature, this approach also has the advantage of being theory-building and is also recommended within group therapy guidelines (Strauss et al., 2008). In the CORE-R it is

argued that it is important to gain a “qualitative perspective on group experiences” (p. 1233) when measuring group processes (Strauss et al., 2008). In line with this guidance, a *critical incidence* approach, where group members are asked about what incidents in the group were critical to them using a series of open-ended questions may also be a helpful exploratory approach (Strauss et al., 2008). Depending on the resources available, we argue that future researchers seeking to identify and measure group process mechanisms would benefit from using a combination of questionnaire-based, behavioural coding and qualitative approaches.

Conclusion

A list of thirteen questionnaire-based group process measures were described and evaluated, providing a valuable source of reference for clinicians and researchers. The Therapeutic Factor Inventory-Short (MacNair-Semands et al., 2010) and the Therapeutic Factor Inventory-19 (Joyce et al., 2011) are recommended for those seeking to assess beneficial mechanisms of change relevant to clinical outcomes. However, if one is interested in measuring overall group processes, then Group Questionnaire (Krogel et al., 2013) is recommended as the most appropriate measure. This might be more relevant in an exploratory study seeking to identify both beneficial and challenging aspects of interactive group mechanisms relevant to clinical outcomes. Future studies should consider validating these measures further. In particular, studies should focus on developing the content validity of these scales. Furthermore, future research would benefit from considering behavioural coding tools, as well as qualitative methods when measuring the therapeutic group process.

Acknowledgements: This project was supported by University College London as part of a Doctorate in Clinical Psychology thesis project.

Declaration of interest statement: None to declare.

References

- Bakali, J., Wilberg, T., Klungsøyr, O., & Lorentzen, S. (2013). Development of group climate in short-and long-term psychodynamic group psychotherapy. *International Journal of Group Psychotherapy, 63*(3), 366-393. <https://doi.org/10.1007/s10464-013-9621-3>
- Bales, R. F. (1950). *Interaction process analysis; a method for the study of small groups*. Addison-Wesley.
- Bernard, H., Burlingame, G., Flores, P., Greene, L., Joyce, A., Kobos, J.C., Leszcz, M., MacNair-Semands, Piper, W.E., McEaney, A.M., & Feirman D. (2008). Clinical Practice Guidelines for Group Psychotherapy. *International Journal of Group Psychotherapy, 58*(4), 445-542.
- Brown, L.D., Tang, X., & Hollman, R.L. (2014). The structure of social exchange in self-help support groups: Development of a measure. *American journal of community psychology, 53*(2), 83-95. <https://doi.org/10.1007/s10464-013-9621-3>
- Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. J. Lambert (Ed.), *Handbook of Psychotherapy and Behavior Change* (5th ed., pp. 647–696). Wileys & Sons.
- Burlingame, G., Strauss, B., Joyce, A., MacKenzie, K., Ogrodniczuk, J., & Taylor, S. (2006). CORE battery: A revision and update. American Group Psychotherapy Association.
- Burlingame, G., Gleave, R., Erekson, D., Nelson, P., Olsen, J., Thayer, S., & Beecher, M. (2015). Differential effectiveness of group, individual, and conjoint treatments: An archival analysis of OQ-45 change trajectories. *Psychotherapy Research, 26*(5), 556-572. <https://doi.org/10.1080/10503307.2015.1044583>

- Burlingame, G. M., Strauss, B., & Joyce, A. S. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Ed.), *Handbook of Psychotherapy and Behavior Change* (6th ed., pp. 640–689). Wileys & Sons.
- Burlingame, G. M. (2010). Small group treatments: Introduction to special section. *Psychotherapy Research, 20*(1), 1-7.
<https://doi.org/10.1080/10503301003596551>
- Burlingame, G. M. (2014). Some Observations on How the Practice of Small Group Treatments Can Be Built Upon the Foundation of Research: A 30+ Year Perspective. *International Journal of Group Psychotherapy, 64*(4), 567-583. <https://doi.org/10.1521/ijgp.2014.64.4.567>
- Cahill, J., Barkham, M., Hardy, G., Gilbody, S., & Richards, D. (2008). A review and critical appraisal of measures of therapist-patient interactions in mental health settings. *Health Technology Assessment, 12*(24), 68. <https://doi.org/10.1108/cgij.2009.24814aaf.003>
- Campbell, N.C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., G., Frances, . . . Kinmonth, Louise, A. (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal, 334*(7591), 455. <https://doi.org/10.1136/bmj.39108.379965.be>
- Caruso, R., Grassi, L., Biancosino, B., Marmai, L., Bonatti, L., Moscara, M., . . . Priebe, S. (2013). Exploration of experiences in therapeutic groups for patients with severe mental illness: development of the Ferrara group experiences scale (FE-GES). *BMC Psychiatry, 13*(242).
<https://doi.org/10.1186/1471-244x-13-242>
- Chapman, L., Burlingame, G., Gleave, R., Rees, F., Beecher, M., & Porter, G. (2012). Clinical prediction in group psychotherapy. *Psychotherapy Research, 22*(6), 673-681. <https://doi.org/10.1080/10503307.2012.702512>

Cohen-Mansfield, J., Hai, T., & Comishen, M. (2017). Group engagement in persons with dementia: the concept and its measurement.

Psychiatry Research, 251, 237-243. <https://doi.org/10.1016/j.psychres.2017.02.013>

Davis, M., Budman, S.H., & Soldz, S. (2000). The individual group member interpersonal process scale. In A. P. Beck & C. M. Lewis (Eds.),

The process (pp. 283-308). American Psychological Association.

DeChavez, M. G., Gutierrez, M., Ducaju, M., & Fraile, J. C. (2000). Comparative Study of the Therapeutic Factors of Group Therapy in

Schizophrenic Inpatients and Outpatients. *Group Analysis*, 33(2), 251-264. <https://doi.org/10.1177/0533316400332006>

Delucia-waack, J. L. (1997). Measuring the effectiveness of group work: A review and analysis of process and outcome measures. *The Journal*

for Specialists in Group Work, 22(4), 277-293. <https://doi.org/10.1080/01933929708415531>

Egger, M., Davey-Smith, G., & Altman, D. (2008). *Systematic reviews in health care: meta-analysis in context* (2nd ed.). John Wiley & Sons.

Evans, N. J., & Jarvis, P. A. (1986). The group attitude scale: A measure of attraction to group. *Small group behavior*, 17(2), 203-216.

<https://doi.org/10.1177/104649648601700205>

Fuhriman, A., Drescher, S., Hanson, E., Henrie, R., & Rybicki, W. (1986). Refining the Measurement of Curativeness: An Empirical Approach.

Small Group Research, 17(2), 186-201. <https://doi.org/10.1177/104649648601700204>

Fuhriman, A., & Burlingame, G. M. (1994a). *Handbook of group psychotherapy: An empirical and clinical synthesis*. Wiley.

Fuhriman, A., & Burlingame, G. M. (1994b). Measuring small group process a methodological application of chaos theory. *Small Group*

Research, 25(4), 502-519. <https://doi.org/10.1177/1046496494254005>

- Garcia-Cabeza, I., Ducaju, M., Chapela, E., & Gonzalez de Chavez, M. (2011). Therapeutic Factors in Patient Groups with Psychosis. *Group Analysis*, 44(4), 421-438. <https://doi.org/10.1177/0533316411413522>
- Heath, C., Hindmarsh, J., & Luff, P. (2010). Video in qualitative research: analysing social interaction in everyday life. *Introducing qualitative methods*. California, USA: Sage.
- Heath, C., Luff, P., & Sanchez S. M. (2007). Video and qualitative research: analysing medical practice and interaction. *Medical education*, 41(1), 109-116. <https://doi.org/10.1111/j.1365-2929.2006.02641.x>
- Janis, R. A, Burlingame, G. M., & Olsen, J. A. (2018). Developing a therapeutic relationship monitoring system for group treatment. *Psychotherapy*, 55(2), 105. <https://doi.org/10.1037/pst0000139>
- Jensen, D. R., Abbott, M. K., Beecher, M. E., Griner, D., Golightly, T. R., & Cannon, J. N. (2012). Taking the pulse of the group: The utilization of practice-based evidence in group psychotherapy. *Professional Psychology: Research and Practice*, 43(4), 388. <https://doi.org/10.1037/a0029033>
- Johnson, J., Pulsipher, D., Ferrin, S., Burlingame, G. M., Davies, R., & Gleave, R. (2006). Measuring group processes: A comparison of the GCQ and CCI. *Group Dynamics: Theory, Research, and Practice*, 10(2), 136-145. <https://doi.org/10.1037/1089-2699.10.2.136>
- 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. <https://doi.org/10.1037/0022-0167.52.3.310>

- Joyce, A. S., MacNair-Semands, R., Tasca, G. A., & Ogrodniczuk, J. S. (2011). Factor structure and validity of the Therapeutic Factors Inventory–Short Form. *Group Dynamics: Theory, Research, and Practice, 15*(3), 201-219. <https://doi.org/10.1037/a0024677>
- Kageyama, M., Nakamura, Y., Kobayashi, S., & Yokoyama, K. (2016). Validity and reliability of the Japanese version of the Therapeutic Factors Inventory–19: A study of family peer education self-help groups. *Japan Journal of Nursing Science, 13*(1), 135-146. <https://doi.org/10.1111/jjns.12098>
- Kroegel, J., Burlingame, G. M., 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>
- Lavelle, M., Healey, P. G., & McCabe, R. (2013). Is nonverbal communication disrupted in interactions involving patients with schizophrenia? *Schizophrenia Bulletin, 39*(5), 1150-1158. <https://doi.org/10.1093/schbul/sbs091>
- Lese, K., & MacNair-Semands, R. R. (2000). The Therapeutic Factors Inventory: Development of a Scale. *Group, 24*(4), 303-317.
- Lo Coco, G., Salvatore, G., Prestano, C., & Burlingame, G.M. (2015). Current issues on group psychotherapy research: An overview. In: *Psychotherapy Research* (O.C.G Gelo et al., eds.), 279-292. Springer-Verlag. https://doi.org/10.1007/978-3-7091-1382-0_14
- Lo Coco, G., Tasca, G. A., Hewitt, P. L., Mikail, S. F., & Kivlighan Jr, D. M. (2019). Ruptures and repairs of group therapy alliance. An untold story in psychotherapy research. *Research in Psychotherapy: Psychopathology, Process and Outcome, 22*(1), 58-70. <https://doi.org/10.4081/ripppo.2019.352>

MacKenzie, R. (1981). Measurement of group climate. *International Journal of Group Psychotherapy*, 31(3), 287-295.

<https://doi.org/10.1080/00207284.1981.11491708>

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

MacNair-Semands, R.R., Ogrodniczuk, J., & Joyce, A. (2010). Structure and Initial Validation of a Short Form of the Therapeutic Factors Inventory. *International Journal of Group Psychotherapy*, 60(2), 245-281. <https://doi.org/10.1521/ijgp.2010.60.2.245>

Marmarosh, C. L. (2018). Introduction to special issue: Feedback in group psychotherapy. *Psychotherapy*, 55(2), 101.

<https://doi.org/10.1037/pst0000178>

Marogna, C., & Caccamo, F. (2014). Analysis of the process in brief psychotherapy group: the role of therapeutic factors. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 43-51. <https://doi.org/10.4081/ripppo.2014.161>

McCrone, P. (2008). Paying the price: the cost of mental health care in England to 2026.

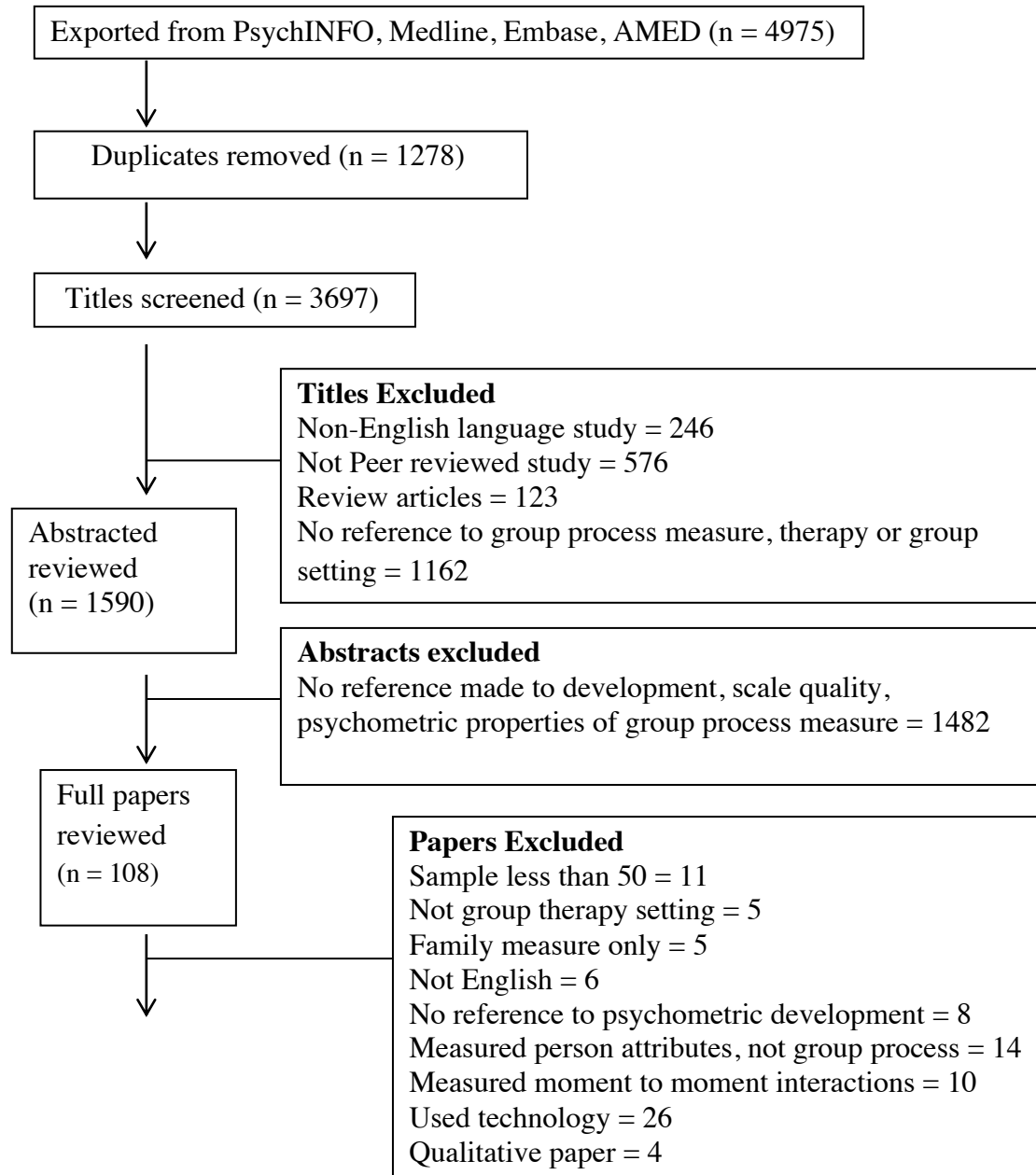
https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf

Moher, D., A. Liberati, J. Tetzlaff and D. G. Altman (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4). <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

- Murillo, N., Shaffer, P., & Michael, W. B. (1981). The development and validation of a preliminary measure for student evaluation of group counseling experiences. *Educational and Psychological Measurement*, 41(2), 463-472. <https://doi.org/10.1177/001316448104100225>
- Priebe, S., Savill, M., Reininghaus, U., Wykes, T., Bentall, R., Lauber, C., . . . Eldridge, S. (2013). Effectiveness and cost-effectiveness of body psychotherapy in the treatment of negative symptoms of schizophrenia – a multi-centre randomised controlled trial. *BioMed Central*, 13(26), 1-8. <https://doi.org/10.1186/1471-244x-13-26>
- Quirk, K., Miller, S., Duncan, B., & Owen, J. (2013). Group Session Rating Scale: Preliminary psychometrics in substance abuse group interventions. *Counselling and Psychotherapy Research*, 13(3), 194-200. <https://doi.org/10.1080/14733145.2012.744425>
- Slone, N., Reese, R., Mathews-Duvall, S., & Kodet, J. (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice*, 19(2), 122. <https://doi.org/10.1037/e533762014-001>
- Sodano, S. M., Guyker, W. M., DeLucia-Waack, J. L., Cosgrove, H. D., Altabef, D. T., & Amos, B. S. (2014). Measures of group process, dynamics, climate behaviour, and outcome: A review. In J. L. DeLucia-Waack, C. Kalodner, & M. T. Riva (Eds.), *The handbook of group counselling and psychotherapy* (2nd ed., pp. 159-177). Sage.
- Strauss, B., Burlingame, G. M., & Bormann, B. (2008). Using the CORE-R battery in group psychotherapy. *Journal of Clinical Psychology*, 64(11), 1225-1237. <https://doi.org/10.1002/jclp.20535>

- Terwee, C. B., Bot, S., de Boer, M., van der W., Daniëlle A., Knol, D., Dekker, J., . . . de Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, *60*(1), 34-42.
<https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Thayer, S. D., & Burlingame, G. M. (2014). The validity of the Group Questionnaire: Construct clarity or construct drift? *Group Dynamics: Theory, Research, and Practice*, *18*(4), 318. <https://doi.org/10.1037/gdn0000015>
- Treadwell, T., Lavertue, N., Kumar, V. K., & Veeraraghavan, V. (2001). The group cohesion scale-revised: reliability and validity. *Journal of Group Psychotherapy, Psychodrama and Sociometry*, *54*(1), 3.
- Wongpakaran, T., Wongpakaran, N., Intachote-Sakamoto, R., & Boripuntakul, T. (2013). The group cohesiveness scale (GCS) for psychiatric inpatients. *Perspectives in psychiatric care*, *49*(1), 58-64. <https://doi.org/10.1111/j.1744-6163.2012.00342.x>
- Yalom, I., & Leszcz, M. (2005). *The Theory and Practice of Group Psychotherapy* (5th edition ed.). Harper Collins.
- Yalom, I. (1985). *The theory and practice of group psychotherapy*. Basic Books.

Figure 1. Search process of literature



Studies included from
literature search = 17

Studies identified from
hand search = 1

Table 1. Characteristics of measures in the ‘therapeutic group process’ category

Measure	Study	Measurement details	Number of items / delivery	Group characteristics, sample size, location
Group Cohesiveness Scale	(Wongpakaran et al., 2013)	Measures the degree to which members of the group feel engaged with each other as a whole. This includes whether members accepted, trusted, liked/cared for each other, revealed personal information and felt a sense of participation.	Seven-items / self-report, observer-rated, five-point Likert scale.	Psychiatric inpatient group psychotherapy setting (N=96, 44% female), Thailand.
	(Lese & MacNair-Semands, 2000)	<i>Therapeutic Factor Inventory</i> : Measures Yalom’s 11 curative group factors; including instillation of hope, universality, imparting information, altruism, corrective recapitulation of the primary family group, development of socialising techniques, imitative behaviour, interpersonal learning, group cohesiveness, catharsis, existential learning.	99-items / self-report, seven-point Likert scale	Counselling and support groups for undergraduate/graduate college students, USA (N=77, 76% female)
Therapeutic Factor Inventory	(MacNair-Semands et al., 2010)	<i>Therapeutic Factor Inventory-Short</i> : Measures four broad therapeutic factors more globally; including instillation of hope, secure emotional expression, awareness of relational impact and social learning.	44-item / self-report, seven-point Likert scale.	Self-awareness groups (N=174, 65% female) at a day treatment program, Canada.
	(Joyce et al., 2011)	<i>Therapeutic Factor Inventory-19</i> : Also measures four broad therapeutic factors more globally; including instillation of hope, secure emotional expression, awareness of relational impact and social learning.	19-item / self-report, seven-point Likert scale.	University counselling groups, outpatient and inpatient groups. Total of 52 groups, (N=379, 61% female), USA and Canada.

Table 1 continued.

Measure	Study	Measurement details	Number of items / delivery	Group characteristics, sample size, location
The Group Cohesion scale-Revised	(Treadwell et al., 2001)	Measures cohesion among group members in terms of interaction and communication. Items include whether group members felt free to share information, feelings of unity and togetherness and receptiveness to feedback.	25-items / self-report, four-point scale	Eight experiential training courses (N=110) cognitive and psychodramatic techniques, USA.
Group attitude scale*	(Evans & Jarvis, 1986)	Measures attraction to group, defined as an individual's desire to identify with and be an accepted member of the group. Focuses on evaluating group member feelings about the group rather than behaviour in the group.	20 items / self-report, nine-point Likert Scale	26 groups (N=178); unstructured/ semi-structured growth, assertion-training, Gestalt, community health groups, USA.
Curative Climate Instrument	(Fuhriman et al., 1986)	Measures four of Yalom's curative group factors. Subscales include: 1) cohesion (forces within a group that draw it together); 2) catharsis (emotional expression of self); 3) interpersonal learning (receiving information about behaviour from other group members) and 4) insight (seeing and experiencing self in a new way).	20 items / self-report, five-point Likert scale	Outpatient therapy or personal growth groups (N=161); veteran, community mental, counselling, centres, USA
Scale for the Evaluation of Group Counselling	(Murillo et al., 1981)	Measures group counselling experiences. Based on Yalom's curative group factors, including 1) satisfying experience (catharsis, cohesiveness, group satisfaction, counsellor effectiveness, openness); 2) interpersonal learning (altruism, interpersonal learning, cohesiveness); 3) increasing self-confidence (instillation of hope, universality)	79 items / self-report, five-point Likert scale	18 counselling groups for personal growth, weight reduction, assertion training and stress management, USA (N=99)

* Reference made to three earlier scale development studies although these were not published as peer-reviewed articles, hence not included.

Table 2. Characteristics of measure in the ‘overall group process’ category

Measure	Study	Measurement details	Number of items / delivery	Group characteristics, sample size, location
Group Observational Measurement of Engagement	(Cohen-Mansfield, Hai & Comishen, 2017)	Measures engagement on both individual level (in terms of duration, participation and attitude) and group level (in terms of total group participation, positive interactions and negative interactions)	Eight items / observer-rated, including five, six and seven-point rating scales	Activity groups (N=105, 65% female) delivered in nursing units, senior day centres and independent living facilities, Canada.
Factors Aspecific and Specific in Group Therapy	(Marogna & Caccamo, 2014)	Measures therapeutic group factors that are specific (related to a theoretical model used by therapist) and non-specific (not related to a theoretical model/specific technique used by therapy).	41-items / self-report, five-point Likert scale	52 group therapies (N=167, 39% female); range of mental health diagnoses from mental health centres, family clinics psychiatric hospitals, Italy
Social exchange scale*	(Brown et al., 2014)	Measures social exchange during self-help support group meetings. Seven subscales assess: 1) emotional support provided; 2) experiential information provided; 3) emotional support received; 4) experiential information received; 5) unwanted behaviours received; 6) humour exchanged; and 7) exchanges outside meetings.	29 items / self-report	18 parenting self-help groups (N=194, 99% female); all parent-led, USA.
Ferrara Group Experiences scale	(Caruso et al., 2013)	Measures core experiences of patients (therapeutic or not) during a single or course of group therapy session(s) within a community mental health context. Five subscales assess: 1) sharing of emotions and experience; 2) cognitive improvement; 3) group learning; 4) difficulties in open expression; and 5) relationships	20 items / self-report, five-point Likert scale	Community setting; psychodynamic, psychosocial rehabilitation, psycho-educational, expressive, body-oriented groups (N = 166 / 65% female), Italy

Table 2 continued.

Measure	Study	Measurement details	Number of items / delivery	Group characteristics, sample size, location
Group Questionnaire	(Krogel et al., 2013)	Measures the quality of therapeutic relationship in groups across member-member, member-leader and member-group relationships Consists of three subscales: positive bonding (cohesion, engagement and emotional bond), positive working (agreement on therapeutic tasks/goals), and negative relationship (conflict and empathic failure).	30-item / self-report, seven-point Likert scale	Outpatient university counselling centre, non-patient process groups (N=486, 65% female) delivered in the USA.
Group Sessions Rating Scale	(Quirk et al., 2013)	Measures group alliance in terms of: 1) relationships (feeling respected and understood or not); 2) goals (working topics of interest or not); 3) acceptability of approach (whether group approach/leadership was adequate enough or not); 4) overall fit (whether group was suitable or not)	Four item visual analogue scale.	Five open therapy groups for issues related to substance abuse (N=105, 58% female)
Group Climate Questionnaire	(MacKenzie, 1981)	<i>Group Climate Questionnaire</i> : Measures group environment. Eight subscales assess; 1) engagement, 2) disclosure, 3) support, 4) conflict, 5) challenge, 6) practicality, 7) cognition, and 8) control.	32 items / self-report, observer-rated, seven-point Likert scale	Psychiatric outpatient psychotherapy, supportive, activity and social groups, (N=119, 56% female)
	(MacKenzie, 1983)	<i>Group Climate Questionnaire-Short</i> : Measures group environment. Three subscales assessed; 1) engagement, 2) conflict and 3) avoidance.	12 items / self-report, observer-rated, seven-point Likert scale	12 therapy groups (N=75, 60% female), USA.

*Consisted of two studies, part two are reported in this review.

Table 3. Psychometric quality scores for the 13 included measures

Group construct measured	Measure	Content Validity	Internal consistency	Criterion Validity	Construct Validity	Reproducibility		Responsiveness	Floor /ceiling effects	Interpretability	Total scores
						a.	b.				
Therapeutic group process measures	Group Cohesiveness Scale	0	2	2	1	0	2	0	0	0	7
	TFI	0	1	0	0	0	0	0	0	0	1
	TFI-Short	0	2	0	2	0	0	1	0	1	6
	TFI-19-item	0	0	0	1	1	1	1	0	2	6
	Group Cohesion Scale-Revised	0	1	0	0	0	0	1	0	1	3
	Group attitude scale	0	1	1	2	0	0	1	0	0	5
	Curative Climate Instrument	1	2	0	0	0	0	0	0	1	4
	Scale for the Evaluation of Group Counselling	0	1	0	0	0	0	0	0	0	1

NB: a = Responsiveness Agreement; b = Responsiveness Reliability; TFI = Therapeutic Factor Inventory; GCQ = Group Climate Questionnaire

Table 3 continued.

Group construct measured	Measure	Content Validity	Internal consistency	Criterion Validity	Construct Validity	Reproducibility a. b.		Responsiveness	Floor /ceiling effects	Interpretability	Total scores
Overall group process measures	Group Observational Measurement of Engagement	0	1	0	0	0	0	1	0	0	2
	Factors aspecific and specific in group therapy	0	2	0	2	0	0	0	0	0	4
	Social exchange scale	2	0	0	1	0	0	0	0	0	3
	Ferrara group experiences scale	1	2	0	0	0	0	0	0	0	3
	Group Questionnaire	0	2	0	2	0	0	0	0	1	5
	Group Sessions Rating Scale	0	2	0	2	0	0	1	0	1	6
	GCQ	0	1	0	1	0	0	0	0	0	2
	GCQ-Short	0	1	0	0	0	0	1	0	1	3
Total category scores		4	21	3	14	1	3	7	0	8	

NB: a = Responsiveness Agreement; b = Responsiveness Reliability; TFI = Therapeutic Factor Inventory; GCQ = Group Climate Questionnaire

Supplementary Material

Table 1: Search terms included in the literature review

Measure	Group therapeutic process	Group therapy
Measure	Group climate	Group therapy
Instrument	Cohesion	Group psychotherapy
Questionnaire	Alliance	Group cognitive behavioural
Scale	Empathy	therapy
Assessment	Therapeutic relationship	Dynamic groups
Reliability	Therapeutic factors	Group analytic therapy
Validity	Engagement	Psychoanalytic therapy
Psychometric	Curative climate	Tavistock groups
Patient health	Social interaction	Interpersonal group therapy
Questionnaire/ or	Interpersonal relationships	Group analysis
psychometrics/	Synchrony	Psychodynamic Psychotherapy
	Motion analysis	Counselling groups
	Nonverbal behaviour	Group counselling
	Gestures behaviours	Support groups
	Interpersonal	Couples therapy
	Group process	Family therapy
	Group experience	Marital therapy

Table 2: Terwee and colleagues' (2007) criteria guidance on assessing development procedures

Property	Definition	Quality Criteria (Score / Definition)	
Content validity	Extent to which concepts of interest are comprehensively represented by items in the questionnaire.	2	Clear description of a) measurement aim of the questionnaire; b) target population; c) concepts intended to measure; d) item selection and e) target population AND (investigators OR experts) involved in item selection; f) clear interpretability of the items.
		1	A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method
		0	No target population involvement OR no information found on target population.
Internal consistency	Assessment of whether items in scale of questionnaire are inter-correlated, i.e. are measuring the same construct	2	Factor analyses performed on adequate sample size ($7 * \# \text{ items}$ and ≥ 100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95.
		1	No factor analysis OR doubtful design or method OR Cronbach's alpha(s) < 0.70 or > 0.95 , despite adequate design and method.
		0	No information found on internal consistency.
Criterion validity	The extent to which scores on a particular questionnaire relate to a gold standard	2	Convincing arguments gold standard is "gold" AND correlation with gold standard ≥ 0.70
		1	No convincing arguments that gold standard is "gold" OR doubtful design or method.
		0	Correlation with gold standard < 0.70 , despite adequate design and method OR no information found on criterion validity.
Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses	2	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses.
		1	Doubtful design or method (e.g., no hypotheses)
		0	Less than 75% of hypotheses were confirmed, despite adequate design and methods OR no information found on construct validity

Table 1. continued.

Property	Definition	Quality Criteria (Score / Definition)
Reproducibility (Agreement)	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	2 MIC<SDC OR MIC outside the LOA OR convincing arguments agreement is acceptable;
		1 Doubtful design or method OR (MIC not defined AND no convincing arguments that agreement is acceptable)
		0 MIC≥SDC OR MIC equals or inside LOA, despite adequate design and method OR no information found on agreement.
Reproducibility (Reliability)	The extent patients can be distinguished from each other, despite measurement errors (relative measurement error)	2 ICC or weighted Kapp≥0.70
		1 Doubtful design or method (e.g., time interval not mentioned)
		0 ICC or weighted Kappa<0.70, despite adequate design and method; OR No information found on reliability.
Responsiveness	The ability of a questionnaire to detect clinically important changes over time	2 SDC or SDC<MIC OR MIC outside the LOA OR RRO1.96 OR AUC≥0.70;
		1 Doubtful design or method;
		0 SDC or SDC≥MIC OR MIC equals or inside LOA OR RR≤1.96 OR AUC<0.70, despite adequate design and methods; OR No information found on responsiveness.
Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	2 ≤15% of the respondents achieved the highest or lowest possible scores;
		1 Doubtful design or method;
		0 >15% of the respondents achieved the highest or lowest possible scores, despite adequate design and methods OR No information found on interpretation.
Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	2 Mean/SD scores presented of at least four relevant subgroups of patients and MIC defined;
		1 Doubtful design or method OR less than four subgroups OR no MIC defined.
		0 No information found on interpretation

NB: MIC=minimal important change; SDC=smallest detectable change; LOA=limits of agreement; ICC=Intraclass correlation; SD=standard deviation. Doubtful design or method = lacking of a clear description of the design or methods of the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis), or any important methodological weakness in the design or execution of the study