

Boise State University

ScholarWorks

Social Work Faculty Publications and
Presentations

School of Social Work

2017

Measurement Issues

Randy Magen

Boise State University

Publication Information

Magen, R. (2017). "Measurement Issues". In C.D. Garvin, L.M. Gutiérrez, and M.J. Galinsky (Eds.), *Handbook of Social Work with Groups* (pp. 535- 548). The Guilford Press.

This document:

Magen, R. (2017). Measurement Issues. In C.D. Garvin, L.M. Gutiérrez & M.J. Galinsky (Eds.), *Handbook of Social Work with Groups* (pp. 535- 548). The Guilford Press, was originally published in *Handbook of Social Work with Groups* by Guilford Press. Copyright 2017, Guilford Press. Reprinted with permission of The Guilford Press. Copyright restrictions may apply.

CHAPTER 30

Measurement Issues

RANDY MAGEN

Examining groups from a systems perspective may be natural for contemporary social workers. The systems perspective, or group-as-a-whole orientation, provides a solid conceptual foundation for understanding groups (Agazarian, 1992). However, to engage in evidence-based group work practice or to conduct research on groups, abstract group concepts must be conceptualized and operationalized (Engel & Schutt, 2009). In qualitative approaches the conceptualization occurs after data have been collected to extract meaning from the observations. In quantitative approaches, operationalization is the process of assigning values to observable indicators of constructs prior to collecting data (Bonito, Ruppel, & Keyton, 2012). Whether inductive or deductive, the collection of observations is the process of measurement. This chapter discusses measurement issues in groups.

GROUP WORK IN THE CONTEXT OF MEASUREMENT

It is an ethical mandate for social workers to “protect clients from harm” (National Association of Social Workers, 2008, 1.04(c)). Research has pointed out the potentially harmful effects of participating in a group (Coyne, 1999; Lieberman, Yalom, & Miles, 1973; Smokowski, Rose, Todar, & Reardon, 1999). It is only through systematic evaluation that practitioners can understand the effects of their groups and be able to, in the least, refute the claim that the group caused harm. Measurement generates the data necessary to systematically evaluate a group. Garvin (1997) also points out that evaluation is necessary in our current political climate, which stresses accountability.

Critics of the use of measures in social work practice argue that measurement is reductionistic and has little relevance to the complexities of social work practice (Goldstein, 1992; Witkin, 1996). Challengers of evidence-based practice argue that the focus on data devalues clinical experience (Williams & Garner, 2002). However, Gibbs (2003, p. 16) writes that measurement is not a substitute for practitioner judgment but rather lies at the “intersection between experience, the client’s preferences and current best evidence.”

Measures are tools for the group work practitioner to utilize in collecting data to understand the effects of a group. There are two general types of effects that can be measured. The first are the effects of the intervention on members' goals—measurement of the outcome of the group. Generally, the question being asked of the data collected through the use of outcome measures is, Did the client change?

The second type of effects are measures of the group conditions, often referred to as measures of group processes. The main question being asked of data collected from group process measures is, Do the group conditions exist that are necessary for clients to make progress toward their goals? The general failure in the clinical literature to link group processes with outcomes is a long-standing and serious problem. Methodologically, group processes can confound outcome. Rose, Tolman, and Tallant (1985) point out that differences between two interventions may be misattributed to content differences when in fact the differences were due to a failure to achieve optimal group processes in one condition.

In individually based practice, what can be measured is limited to the client, the practitioner, and the interactions between the practitioner and the client. In groups, what can be measured expands geometrically with every group member because each individual, dyad, triad, and so on can be the target of measurement. For example, in a group of six people in which one of the members is the designated leader, there are six individuals, 15 dyads, 20 triads, 15 subgroups of four members, six subgroups of five members, and the entire group of six that can be the focus of measurement. This assumes, of course, that measurement of the dyads, triads, or subgroups is meaningful and useful.

Like any tool, measures can be used carelessly and foolishly. Using measures that are a burden because of their complexity, length, or relevance is foolish. Similarly, using measures without a clear idea of how they will be used to assist clients or leaders is careless.

INTERVENTION

The use of measures in groups is not limited to any particular approach to group work. As is shown below, qualitative and quantitative measures have been utilized in psychodynamic and behaviorally oriented groups, in remedial and reciprocal model groups, and in leader-centered and self-help groups. However, some methodologies and types of measures are more syntonic, with specific approaches to group work. For example, Kurtz (1997) advises that qualitative methods and participatory action research are best suited to self-help groups.

Empirical Evidence and Theoretical Base

Five issues need to be addressed before using any measure in groups. These issues are related to the unit of analysis, definition of concepts, individualization of measurement, reliability, and validity.

Unit of Analysis

The unit-of-analysis issue may be the most commonly ignored methodological quagmire in group work research. The basic idea is that the unit measured should be the unit analyzed. However, it is common for group work researchers to collect data on one level (e.g., individuals) and infer something about a different system level (e.g., dyads, triads, or a group). There are many dangers with this practice. First is the reductionist fallacy (Engel & Schutt,

2009), a common but suspect practice in which individual-level data are collected yet analyzed and reported in terms of the group. Another problem with using individual data to describe the group as a whole relates to errors in aggregating the data. Group workers often report the mean of the group members' individual scores as the score for the group. Glisson (1986) pointed out the dangers of this approach: Depending on the distribution of individual scores, this procedure can either overestimate or underestimate the group score.

Analyzing individual scores to draw conclusions about the group rests on the assumption that individual data values are independent. However, part of the therapeutic impact of a group is that members influence each other; thus scores on measurement instruments are affected by what other group members say and do—the data are not independent observations. The violation of the assumption of independence can lead to bias in standard error estimates and erroneous results in statistical tests (Barcikowski, 1981; Pollack, 1998).

Similar is the ecological fallacy, less common in the literature than the reductionist fallacy but equally suspect, in which group-level data are used to make statements about individuals (Engel & Schutt, 2009). Conceptually, based on systems theory, the group is understood as more than the sum of the individuals; therefore, we need data at multiple levels and/or analytic approaches that accurately describe the different levels. Although statistical analyses to deal with this problem are beyond the scope of this chapter, two solutions that have been suggested are using nested designs (Morran, Robison, & Hulse-Killacky, 1990) and hierarchical linear modeling (Pollack, 1998). Bonito and colleagues (2012) point out that analyses at both the group and individual levels are precisely what are needed to understand the interaction between individuals and groups.

In some group work research, the individual level of analysis may be less important than other levels. For example, Maton (1993) proposed an ecological paradigm for understanding self-help groups within the context of communities and other large systems. In this model, if self-help groups are conceptualized as community organizations (Kurtz, 1997), then community-level variables such as human service budgets, prevalence of focal problems, and population stability affect group effectiveness and need to be measured.

Finally, the practice of aggregating individual scores as measures of group process is also suspect on conceptual grounds, given that almost all definitions of group processes are linked to the group rather than to individuals.

Definitions of Concepts

Measurement does not start with the selection of an instrument but rather with clear definitions of the concepts to be measured. Bednar and Kaul (1994, p. 633), in reviewing 50 years of group research, were highly critical of the state of measurement in group work research. Instead of more rigorous research designs, they called on researchers and practitioners to carefully observe and describe groups so that specific measurement tools could be created. Bednar and Kaul's critique should be seen as a warning to anyone measuring aspects of group work. The usefulness of any measure is dependent on the clarity of the concept it is describing.

In group work practice and research, the difficulty in defining concepts is most pronounced in measures of group process. Although there have been rich descriptions of various group processes, little agreement exists about even the most basic phenomena. For example, on the one hand, there is widespread agreement that cohesion is a necessary group process for the functioning of effective groups. High cohesiveness has been linked to therapeutic change (Dies & Teleska, 1985, p. 120), whereas low cohesiveness has been shown to

correlate highly with members dropping out of groups (Lieberman et al., 1973). A meta-analysis of group cohesion research by Burlingame, McClendon, and Alonso (2011) found, across 40 studies, a positive, medium-size effect between cohesion and outcome. On the other hand, there is little agreement about how to define cohesion, as illustrated by the following definitions:

- The total field of forces which act on members to remain in the group (Festinger, Schachter, & Back, 1950, p. 164).
- Attraction to the group (Lieberman et al., 1973, p. 302).
- The composite of member–member, member–therapist, and member–group relations (Fuhriman & Barlow, 1983, p. 263).
- The degree of members' involvement in and commitment to the group and the concern and friendship they show for one another (Moos, 1986, p. 2).

For group work researchers and practitioners, the first step in measurement is a clear conceptual definition of the concept. This leads directly to the operationalization of the concept. Given the variation in conceptual definitions of cohesion, it should be no surprise that Burlingame and colleagues (2011) identified nine commonly used measures of group cohesion in the published research literature.

Individualization of Measurement

A common research design for investigating the effect of a group intervention tests the hypothesis that the group members' mean score on an outcome measure is significantly different from pretest to posttest or is significantly different from scores of members in control or comparison groups. For example, to test whether a cognitive-behavioral group reduces members' levels of depression, a researcher could administer the Beck Depression Inventory (BDI) prior to the beginning of the group and again at the end of the group. The researcher could then perform a matched-pairs *t*-test, assuming adequate sample size, to reject the null hypothesis that there was no difference in scores on the BDI from pretest to posttest. The purpose of this research is to show that a cognitive-behavioral group intervention can reduce depression and develop knowledge generalizable beyond the group being studied.

Utilizing a norm-based instrument such as the BDI allows researchers and practitioners to compare their samples to various published norms. The practitioner reading the report about the cognitive-behavioral intervention has a standard against which to evaluate the level of group members' depression. Measures that allow comparison to norms are referred to as *nomothetic* measures.

Nomothetic measures also have disadvantages. First, norm-based measures are often indirect proxy measures of client troubles and, as a result, are difficult to use in evaluating individual reactions and the importance participants attach to their difficulties. Norm-based measures are often insensitive to small changes. Nomothetic explanations that rely on tests of sample means provide us with little or no information about what happened to an individual client. Average scores can underestimate clinically significant changes when the effect of treatment is variable within treatments (i.e., when there is large within-subject variability). Furthermore, the imposed significance levels associated with statistical tests often have little relevance to clinical practice (Jacobson, Follette, & Ravenstorf, 1984). In fact, with a large number of participants, results can be statistically significant even when the effectiveness of an intervention is "weak" (Barlow, Hayes, & Nelson, 1984). Adjuncts to the use of nomothetic measures are ideographic measures.

Ideographic measures are precise and sensitive indicators of individual change. Although not commonly reported in group work research, ideographic measures can be found in the behavioral literature associated with single-system research designs. Ideographic measures allow clients to select indicators of important problems and determine how to measure a successful outcome. Furthermore, some ideographic measures can measure progress, as well as success. Ideographic measures of outcome, discussed later, include goal attainment scaling and individualized rating scales (Bloom, Fischer, & Orme, 1999).

Researchers and practitioners often want to know not only whether a group intervention is effective but also for whom it was most effective. Clinical significance tests following statistical significance tests are a method for obtaining ideographic explanations of outcome. Several statistical tests have been suggested for evaluating clinical significance (Christensen & Mendoza, 1986; Jacobson et al., 1984, 1986; Jayaratne, 1990). The most useful of these statistics is the *SC* suggested by Christensen and Mendoza (1986), which yields a score, in standard deviation units, that indicates the magnitude of an individual subject's change on a measure. The *SC* statistic, used in conjunction with nomothetic explanations, can tell us not only whether the group was effective but also which individual members benefitted most from the intervention.

Reliability

A synonym for reliability is consistency. There are many different ways to assess the reliability of a measure; the most common of these are test-retest, split-half, internal consistency, and multiple forms. For measures involving observation, the assessment of reliability refers to the consistency between raters, known as interrater reliability. Three methods for computing interrater reliability are Cohen's kappa, the interclass correlation coefficient, and the percentage of agreement.

By convention, the agreed-on standard for a reliable measure is .80, or 80%. Clearly, the higher the reliability, the better, but researchers commonly use measures with reliabilities considerably lower than .80. It is important for researchers and practitioners to understand the reliability of any instrument utilized. At a minimum, basic reliability statistics should be calculated and reported.

Validity

Valid measures are accurate measures. There are two questions associated with evaluating the validity of a measure. First, is the instrument measuring what it is supposed to measure? Second, how well does the instrument measure what is intended? Valid measures are reliable, but reliable measures are not necessarily valid. As with reliability, there are many methods for assessing the validity of a measure. These methods include face validity, content validity, criterion validity (concurrent and predictive), and construct validity (convergent, discriminate, and convergent). Assessing the validity of a measure is a complex and difficult task involving both clear concepts and careful empirical research. Measures are never completely valid nor completely invalid. Generally, the more information that is available about the validity of a measure, the more faith the researcher or practitioner can have in its use.

Measures of Outcome

To select an outcome measure, the group practitioner or researcher needs to decide (1) what to measure, (2) how to measure it, and (3) who will collect the measurement data (or where

they will be collected). This process of selecting outcome measures can be conceptualized as a three-dimensional cube, with each question corresponding to one dimension of the cube. Within each dimension are specific domains to be measured, methods of measurement, and source of the data (see Figure 30.1). There are many ways to conceptualize the elements within each dimension; the elements identified in Figure 30.1 are one example. The conceptualization of elements within each dimension is influenced, at a minimum, by the purpose of the group and the clients' needs. Other considerations in conceptualizing what to measure, how to measure, and who will collect the measurement data include theoretical orientation, practicality, and feasibility.

The metaphor of the cube is also helpful in identifying whether data collection is multimethod and multisourced, what qualitative researchers refer to as *triangulation* (Berg, 1998). Multiple blocks within the cube containing data will indicate measures that are triangulated. However, not all measures can be triangulated; for example, some outcomes are covert, such as a change in cognitions. As a result, the cells in Figure 30.1 that correspond to cognitions can be collected only from the client. However, the measure of cognitions can be done by means of self-monitoring, via tests and checklists; through analog observations; or by way of interviews. Furthermore, the measure of cognitions can be either quantitative or qualitative.

In reality, practitioners and researchers rely on a few common methods for measuring client outcomes. Perhaps the most common of these tools are tests and checklists, primarily

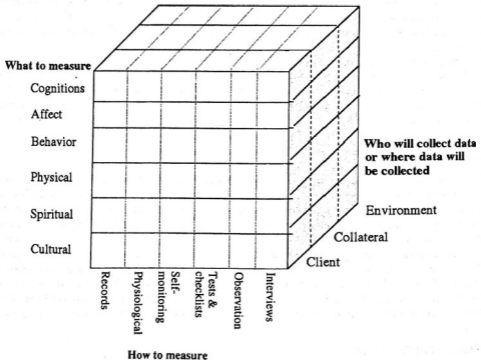


FIGURE 30.1. Triangulation matrix for selecting outcome measures.

norm-based instruments. Empirical research studies, in particular, rely on norm-based measures. Highly touted and extremely practical, but less utilized, are rapid assessment instruments such as those developed by Hudson (1990a, 1990b).

Another type of instrument, commonly used in children's social skills training groups, is sociometric ratings (see, e.g., Hepler, 1994). Originally developed by J. L. Moreno, the father of psychodrama, these checklists have high face validity and some evidence of reliability. In a school setting, children would be asked to rate, on a Likert scale, the likelihood of their playing with every other member of the classroom. The scores from the classroom would then be used to identify those children most isolated, as well as those most popular. These sociometric ratings are an assessment tool that can be used to compose groups that are heterogeneous. The sociometric ratings can also be repeated after the end of the group to test either ideographically or nomothetically for changes.

Two methods for obtaining ideographic measures of change are individualized rating scales (IRS) and goal attainment scaling (GAS). IRS, as described by Bloom and colleagues (1999), use a single scale with descriptive anchors to operationally define target problems. For example, a client in a support group for people who have breast cancer might use such a scale to rate the intensity of her fears about chemotherapy. The practitioner and client would work together to develop the labels for the 5-9 points found on the typical IRS. The client would then use this scale on a regular basis (e.g., weekly) to measure fears about chemotherapy.

Similar to IRS, GAS is frequently discussed in texts on measurement but rarely reported in the group work literature. GAS and IRS share many similarities. Both involve individualized ratings of a particular target. However, GAS is usually constructed on a scale from -2 to +2, with the level 0 being the client's current level of functioning. GAS, like IRS, can be used to measure not only outcome but also progress toward outcome (see, e.g., Magen & Rose, 1994).

Observation of group members' behavior within the group has long been used as material for social workers' case notes. One of the more common applications of systematic observation has been parent-child interactions (see, e.g., Whipple, 1999), in which structured observational systems are available. Observation of classroom behavior as reported by a teacher has also been used to assess outcome in some skills training groups. Finally, analog measures of behavior through role playing have been used both during group sessions and outside of the group as measures of outcome.

In the biopsychosocial perspective of social work, the biological realm often receives less attention than the psychological or social realms. As a result, scant attention has been given to the use of physiological measures of outcome in the group work literature. The one exception to this has been in the field of substance abuse, in which urinalysis is employed to supplement self-reports of abstinence. Several other simple physiological measures can be used to measure agitation or stress, such as heart and respiration rates. More complex measures of stress, such as cardiac output, have been used in social support and stress management groups by O'Brien and his colleagues (see, e.g., Anthony & O'Brien, 2002).

Self-monitoring as a method for obtaining outcome measurement is easily implemented but can have low reliability due to problems with adherence. It is useful for practitioners and researchers to follow guidelines for enhancing adherence when implementing self-monitored outcome measurements. For example, to improve self-reports of sexual behavior in HIV research, Weinhardt, Forsyth, Carey, Jaworski, and Durant (1998) suggested building in techniques that improve the recall of behavior, such as appointment books or calendars, and placing the burden of denial on the client. Rather than having participants record "if" a behavior occurred, the researchers asked them to record how many times a behavior

took place. From a research perspective, the reactive nature of self-monitoring can present a problem; however, in practice, this reactivity has been taken advantage of to strengthen an intervention. For example, in smoking cessation groups, monitoring the number of cigarettes smoked can help to reduce smoking.

Social workers are well trained in techniques of interviewing. Systematic collection of data during an interview can be used as a baseline for assessing outcome when compared with postgroup interview data. Interviews allow the collection of rich qualitative data that cannot be obtained from other measurement methods. For example, to understand the effect of a self-help group in preventing drug use and improving client's lives, Felix-Ortiz, Salazar, Gonzalez, Sorensen, and Plock (2000) interviewed 7 of 14 members participating in a self-help group at a methadone maintenance clinic. Whereas drug use and quality of life can be measured using other methods (e.g., tests and checklists), these interviews provided data on what the clients perceived to be the value of the group in contributing to their personal changes, a link that would be difficult to make using written measures.

Existing records may contain all the types of data listed here, as well as time-series data. Intake forms and other routinely collected agency records may be the only empirical data that exist to evaluate change for some group members. There are problems with the reliability of data in records, particularly if multiple practitioners wrote the records. However, practitioners and researchers should not neglect this relatively easily obtained source of data.

Research indicates that group workers (Chapman, 2010), as well as other clinicians (Hannan et al., 2005) have difficulty making accurate predictions of client outcomes without the aid of systematic measures. In addition, there is evidence that monitoring (i.e. measuring) treatment outcomes can not only reduce adverse outcomes but also prevent deterioration (Burlingame, Fuhriman, & Johnson, 2004; Kazdin, 2005; Lambert, Harmon, Slade, Whipple, & Hawkins, 2005). Systematic and ongoing measurement provides the group worker with the opportunity to readjust interventions should the data indicate changes are needed (Hannan et al., 2005).

Measures of Group Process

Measures of group process are designed to illuminate what is occurring within the group and to identify and assess the factors that are necessary for the functioning of an effective small group. A number of variables have been identified in the literature as belonging to the phenomenon of group process, but no one constellation of variables has received universal acceptance. Thus each measure of group process described in this section constitutes a slightly different subset of group process variables. Furthermore, a definitive list of group processes is probably not possible, given that group processes are influenced by the group work approach, by the stage of the group's development, by forces outside of the group, and by individual differences within the group.

A number of different tools have been developed to measure group processes or, more generally, group conditions. The tools discussed here were selected based on the following criteria: (1) their use has been discussed in peer-reviewed professional literature; (2) they have utility for both researchers and practitioners; (3) they are relatively low in cost and require little training to use. The reader is also referred to the book by Beck and Lewis (2000) that offers a comprehensive review of nine approaches developed over the past 50 years to study small groups. The systems of analysis presented focus on observations of group interaction from a psychodynamic perspective. Two of the instruments presented

by Beck and Lewis, which may be of most interest to readers of this chapter, are the Hill Interaction Matrix (HIM) and the Individual Group Member Interpersonal Process Scale (IGIPS). Readers are also referred to the annotated review of 26 group process instruments by Fuhriman and Packard (1986).

Perhaps the most widely known conceptualization of group process is Yalom's (1995) therapeutic factors. To empirically explore these 12 therapeutic factors, which were developed based on his and others' experience with group psychotherapy, Yalom developed 60 statements, 5 for each of the 12 therapeutic factors. Patients, and in some cases group psychotherapists, were asked to perform a Q-sort on these 60 statements using seven categories, from "most helpful" to "least helpful." The original Q-sort has been replaced by a self-administered checklist, on which Yalom's 60 statements are ranked on a 4-point Likert scale, with 0 corresponding to "not helpful" and 3 to "very helpful" (Butler & Fuhriman, 1983).

A variety of studies have demonstrated that clients view the presence of these factors as important to their group experience. However, the relative importance assigned to each factor has varied by the group studied. For example, research has consistently shown that cohesion and universality are perceived as important in the group experience of clients in self-help groups (Heil, 1992; Lieberman & Borman, 1979), psychotherapy groups (Bednar & Kaul, 1994), and cancer support groups (Weinberg, Uken, Schmale, & Adamek, 1995).

More than 40 separate studies have reported empirically on the use of the Yalom therapeutic factors questionnaire. Peer-reviewed data on reliability and validity have been published, including a factor analysis of the questionnaire; Yalom (1995) lists references to these in his book (see also Fuhriman, Drescher, Hanson, Henrie, & Rybicki, 1986). A critique of the therapeutic factors questionnaire and an alternative instrument based largely on the same concepts has been offered by MacNair-Semands and Lese (2000). However, this new questionnaire has not been as widely adopted as the original Yalom questionnaire.

Clearly, the advantage of the Yalom therapeutic factors questionnaire is its widespread adoption by both the practice community and researchers. Furthermore, the 60-item questionnaire is relatively easy to administer to adults. One disadvantage of the therapeutic factors questionnaire is that it measures participants' perceptions of the *importance or value* of each of the therapeutic factors but not whether the factor was actually present in the group (MacNair-Semands & Lese, 2000).

Another self-report instrument for measuring group conditions, developed by Rudolph Moos, is the Group Environment Scale (GES; Moos, 1986). The GES is a 90-item, true-false instrument. Conceptually, the GES consists of three domains: relationship, personal growth, and system maintenance/system change. Each domain is measured by 1 of the 10 subscales that constitute the GES. In the Relationship domain, subscales measure Cohesion, Leader Support, and Expressiveness. The Personal Growth domain is measured by the Independence, Task Orientation, Self-Discovery, and Anger/Aggression subscales. The System Maintenance/System Change domain is measured by the Order and Organization, Leader Control, and Innovation subscales. The manual for the GES contains a conceptual and operational definition for each of the subscales, as well as normative data from 148 groups. The normative data are further broken down by type of group: task-oriented, social-recreational, and psychotherapy/mutual-support groups. Data are provided on the reliability and validity of the scale. Finally, the manual for the GES contains references to more than 25 studies that have used one or more of the subscales.

Unlike Yalom's therapeutic factors questionnaire, which grew out of practitioners' experiences, the GES's development was guided by the application of systematic psychometric

methods. Although the GES has been used with a wide variety of groups, from committees to support and psychotherapy groups, most of the current research appears to be focused on task-oriented groups and is found in the organizational management literature.

A number of relatively simple methods for collecting data on group conditions have been suggested by Rose (1984). Relevant to the discussion of tests and checklists, Rose proposed the use of a postsession questionnaire (PSQ) administered at the end of every group session. The questionnaire collects both qualitative and quantitative data on group member's perceptions of the usefulness of the group, satisfaction with the group, cohesion, and self-disclosure. There are no published data on the reliability or validity of the PSQ, and no norms exist; thus the PSQ should be used ideographically. Although the postsession questionnaire has been used in a variety of studies by Rose and his colleagues (e.g., Magen & Rose, 1994; Rose & Edleson, 1987; Whitney & Rose, 1989), as well as by others (e.g., Anthony & O'Brien, 2002), its utility is as a simple, easy-to-use tool for practitioners.

A qualitative instrument for collecting data on group conditions is the critical incident report. Like the postsession questionnaire, the critical incident methodology asks group members at the end of a group session to:

Please describe briefly the event that was most personally important to you during today's session. This might be something that involved you directly, or something that happened between other members, but made you think about yourself. Explain what it was about the event that made it important for you personally. (MacKenzie, 1987, p. 31)

MacKenzie (1987) had clinicians categorize the critical incidents into categories that roughly corresponded to Yalom's therapeutic factors. A similar critical incident methodology, but one with a different conceptual coding system, has been used in the study of task group behavior (see, e.g., Taggar & Brown, 2001). The advantage of the critical incident report is that group members' words are not filtered through preconceived questions or limited to specific response categories. A disadvantage of this approach is that participants must possess the cognitive and writing skills to clearly explain the critical incident.

An instrument developed by Macgowan (2000) is the Groupwork Engagement Measure (GEM). In this instrument, engagement is conceptualized as a multidimensional construct across seven dimensions: attendance, contributing, relating to worker, relating with members, contracting, working on own problems, and working on others' problems (Macgowan & Levenson, 2003). The GEM has 37 items scored on a 5-point Likert scale from 1, "rarely or none of the time," to 5, "most or all of the time." Macgowan has subjected the GEM to tests of validity and reliability with various populations and types of groups (see also Macgowan, 1997). The GEM is a relatively short instrument, which makes it easy to use both in practice and research. There is a growing body of evidence to support the psychometric properties of this instrument. Furthermore, the multidimensional construct of engagement emerges both from practice with groups and from research on group conditions.

The Clinical Outcome Results Standardized Measures (CORE-R) battery (Strauss, Burlingame, & Bormann, 2008) is a toolkit of measures promoted by the American Group Psychotherapy Association (AGPA). The CORE-R measures are grouped into three categories, each of which contains multiple measures: group selection and pregroup preparation measures, process measures, and outcome measures. The group selection and pregroup preparation category has two self-report instruments, the group therapy questionnaire and the group selection questionnaire, which are designed to aid group workers in making decisions about group composition. Six instruments are included in the CORE-R's group

process section: Working Alliance Inventory, Empathy Scale, Group Climate Questionnaire, Cohesiveness scale of the Therapeutic Factors, and the Critical Incidents Questionnaire (discussed earlier). The outcome portion of the CORE-R has three scales—Target Complaints, the Rosenberg Self-Esteem Scale, and the Inventory of Interpersonal Problems—and the Outcome Questionnaire 45. An advantage of the CORE-R is that instruments were selected using criteria that included being psychometrically sound, short, and easy to use. In addition, the advocacy of the AGPA may mean that these instruments will become widely used.

Rose (1984) also suggests several easy-to-implement observational coding systems for use in groups. A simple measure, attendance, is logically related to outcome, although, as Feldman, Caplinger, and Wodarski (1983) indicated, group members who drop out, who successfully complete treatment, and who unsuccessfully complete treatment are three distinct groups. Similarly, records of promptness to group sessions can be used to measure time in the group. The marketing wisdom that people vote with their feet also applies to groups; these simple measures of attendance and promptness can indicate to what degree people are voting with their feet.

Observational measures of participation in the group can be as straightforward as tally marks to record the frequency of group member participation. Rose (1984) recommends recording who in the group is speaking every 10 seconds. Somewhat more complex observational systems record the patterns of speaking, such as whether members are addressing the group, the leader, or one particular member. These measures of participation, much easier to learn and use than other observational systems such as the HIM, systematize what many group leaders and researchers treat as anecdotal data.

FUTURE DIRECTIONS

The knowledge and tools to measure outcomes for members participating in groups have advanced, both in the research and in the practice arenas. Although a number of tools exist for assessing group processes, they have not been subjected to the same rigorous use or testing as measures of outcome. The larger problem, however, is that the conceptual definitions of group conditions are still amorphous. Progress has been made; for example, Bonito (2002) tackled the methodological and conceptual issues involved in understanding and analyzing participation in groups. This is a good first step for one of the multitude of variables that have been identified as relevant to understanding groups. Researchers too often conduct rigorous studies in which a group is the context for the intervention, but no measures are made of the group condition. This ignores both the context of the intervention and the possible confounding nature of group variables on outcome. Finally, the relatively new and complex statistical techniques for analyzing nonlinear and dependent data offer another avenue for understanding the multiple relationships with groups, as well as the connections between process and outcome. While these statistical techniques offer researchers tools for analysis, their complexity may create a further divide between research and practice.

The *Oxford English Dictionary* (2003; <http://dictionary.oed.com>) lists one of the earliest uses of the word *measurement* in the 17th century. One of the definitions is "a magnitude, quantity, or extent calculated by the application of an instrument or device marked in standard units." This chapter has discussed the instruments as applied to groups. Measurement need not be a complex undertaking, but it needs to be systematic, with a set of associated standards. Measurement is necessary for understanding what is occurring in groups and with group members. The choice is what to measure and how, not whether to measure.

REFERENCES

- Agazarian, Y. M. (1992). Contemporary theories of group psychotherapy: A systems approach to the group-as-a-whole. *International Journal of Group Psychotherapy*, 42, 177-203.
- Anthony, J. L., & O'Brien, W. H. (2002). The effects of a group-based social support intervention on cardiovascular reactivity. *Small Group Research*, 33(2), 155-180.
- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6, 267-285.
- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist practitioner*. New York: Pergamon Press.
- Beck, A. P., & Lewis, C. M. (2000). *The process of group psychotherapy: Systems for analyzing change*. Washington, DC: American Psychological Association.
- Bednar, R. L., & Kaul, T. J. (1994). Experiential group research: Can the canon fire? In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 631-663). New York: Wiley.
- Berg, B. L. (1998). *Qualitative research methods for the social sciences* (3rd ed.). Boston: Allyn & Bacon.
- Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating practice: Guidelines for the accountable professional*. Boston: Allyn & Bacon.
- Bonito, J. A. (2002). The analysis of participation in small groups: Methodological and conceptual issues related to interdependence. *Small Group Research*, 33(4), 412-438.
- Bonito, J. A., Ruppel, E. K., & Keyton, J. (2012). Reliability estimates for multilevel designs in group research. *Small Group Research*, 43(4), 443-467.
- Burlingame, G. M., Fuhrman, A. J., & Johnson, J. (2004). Current status and future directions of group psychotherapy research. In J. L. DeLucia, D. A. Gerrity, C. R. Kalodner, & M. T. Riva (Eds.), *Handbook of group counseling and psychotherapy* (pp. 651-660). Thousand Oaks, CA: Sage.
- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy*, 48(1), 34-42.
- Butler, T., & Fuhrman, A. (1983). Level of functioning and length of time in treatment variables influencing patients' therapeutic experience in group psychotherapy. *International Journal of Group Psychotherapy*, 33, 489-505.
- Chapman, C. (2010). *Clinical prediction in group psychotherapy*. Unpublished doctoral dissertation. Available at www.researchgate.net/profile/Gary_Burlingame/publication/229007313_Clinical_prediction_in_group_pschotherapy/links/004635162f21d8c648000000.pdf.
- Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, 17, 305-308.
- Coyne, R. K. (1999). *Failures in group work: How we can learn from our mistakes*. Thousand Oaks, CA: Sage.
- Dies, R. R., & Teleska, P. A. (1985). Negative outcome in group psychotherapy. In D. T. Mays & C. M. Franks (Eds.), *Negative outcome in psychotherapy and what to do about it* (pp. 118-141). New York: Springer.
- Engel, R. J., & Schutt, R. K. (2009). *The practice of research in social work* (2nd ed.). Thousand Oaks, CA: Sage.
- Feldman, R. A., Caplinger, T. E., & Wodarski, J. S. (1983). *The St. Louis conundrum: The effective treatment of antisocial youths*. Englewood Cliffs, NJ: Prentice-Hall.
- Felix-Ortiz, M., Salazar, M. R., Gonzalez, J. R., Sorensen, J. L., & Plock, D. (2000). Addictions services: A qualitative evaluation of an assisted self-help group for drug-addicted clients in a structured outpatient treatment setting. *Community Mental Health Journal*, 36, 339-350.
- Festinger, L., Schachter, S., & Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing*. Palo Alto, CA: Stanford University Press.
- Fuhrman, A., & Barlow, S. H. (1983). Cohesion: Relationship in group therapy. In M. J. Lambert

- (Ed.), *A guide to psychotherapy and patient relationships* (pp. 263–289). Homewood, IL: Dow Jones, Irwin.
- Fuhriman, A., Drescher, S., Hanson, E., Henrie, R., & Rybicki, W. (1986). Refining the measurement of curativeness: An empirical approach. *Small Group Behavior, 77*, 186–201.
- Fuhriman, A., & Packard, T. (1986). Group process instruments: Therapeutic themes and issues. *International Journal of Group Psychotherapy, 36*(3), 399–425.
- Garvin, C. D. (1997). *Contemporary group work* (3rd ed.). Boston: Allyn & Bacon.
- Gibbs, L. E. (2003). *Evidence-based practice for the helping professions: A practical guide with integrated multimedia*. Pacific Grove, CA: Brooks/Cole.
- Glisson, C. (1986). The group versus the individual as the unit of analysis in small group research. *Social Work with Groups, 9*, 15–30.
- Goldstein, H. (1992). Should social workers base practice decisions on empirical research? No. In E. Gambrell & R. Pruger (Eds.), *Controversial issues in social work* (pp. 114–120). Boston: Allyn & Bacon.
- Hannan, C., Lambert, M., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 1–9.
- Heil, R. A. (1992). A comparative analysis of therapeutic factors in self-help groups (Alcoholics Anonymous, Overeaters Anonymous). *Dissertation Abstracts International, 53*(8), 4373B. (UMI No. AAG9226528)
- Hepler, J. B. (1994). Evaluating the effectiveness of a social skills program for preadolescents. *Research on Social Work Practice, 4*(4), 411–435.
- Hudson, W. W. (1990a). *MPSI starter kit*. Tallahassee, FL: WALMYR.
- Hudson, W. W. (1990b). *WALMYR classroom training package*. Tallahassee, FL: WALMYR.
- Jacobson, N. A., Follette, W. C., & Ravenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352.
- Jacobson, N. A., Follette, W. C., & Ravenstorf, D. (1986). Toward a standard definition of clinically significant change. *Behavior Therapy, 17*, 308–311.
- Jayarathne, S. (1990). Clinical significance: Problems and new developments. In L. Videka-Sherman & W. J. Reid (Eds.), *Advances in clinical social work research* (pp. 271–285). Silver Springs, MD: National Association of Social Workers.
- Kazdin, A. E. (2005). Evidence-based assessment for children and adolescents: Issues in measurement development and clinical application. *Journal of Clinical Child and Adolescent Psychology, 34*, 548–558.
- Kurtz, L. F. (1997). *Self-help and support groups: A handbook for practitioners*. Thousand Oaks, CA: Sage.
- Lambert, M. J., Harmon, C., Slade, K., Whipple, J., & Hawkins, E. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61*(2), 165–174.
- Lieberman, M. A., & Borman, L. D. (1979). *Self-help groups for coping with crisis: Origins, members, processes, and impact*. San Francisco: Jossey-Bass.
- Lieberman, M. A., Yalom, I. D., & Miles, M. B. (1973). *Encounter groups: First facts*. New York: Basic Books.
- Macgowan, M. J. (1997). A measure of engagement for social groupwork: The Groupwork Engagement Measure (GEM). *Journal of Social Service Research, 23*(2), 17–37.
- Macgowan, M. J. (2000). Evaluation of a measure of engagement for group work. *Research on Social Work Practice, 10*(3), 348–362.
- Macgowan, M. J., & Levenson, J. S. (2003). Psychometrics of the group engagement measure with male sex offenders. *Small Group Research, 34*(2), 155–160.
- MacKenzie, K. R. (1987). Therapeutic factors in group psychotherapy: A contemporary view. *Group, 11*(1), 26–34.

- MacNair-Semands, R. R., & Lese, K. P. (2000). Interpersonal problems and the perception of therapeutic factors in group therapy. *Small Group Research, 31*(2), 158-174.
- Magen, R. H., & Rose, S. D. (1994). Parents in groups: Problem solving versus behavioral skill training. *Research on Social Work Practice, 4*(2), 172-191.
- Maton, K. I. (1993). Moving beyond the individual level of analysis in mutual help group research: An ecological paradigm. *Journal of Applied Behavioral Science, 29*, 272-286.
- Moos, R. H. (1986). *Group Environment Scale manual* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Morran, D. K., Robison, F. F., & Hulse-Killacky, D. (1990). Group research and the unit of analysis problem: The use of ANOVA designs with nested factors. *Journal for Specialists in Group Work, 15*, 10-14.
- National Association of Social Workers. (2008). Code of ethics. Retrieved from www.socialworkers.org/pubs/code/code.asp.
- Pollack, B. N. (1998). Hierarchical linear modeling and the "unit of analysis" problem: A solution for analyzing responses of intact group members. *Group Dynamics: Theory, Research and Practice, 2*, 299-312.
- Rose, S. D. (1984). Use of data in identifying and resolving group problems in goal-oriented treatment groups. *Social Work with Groups, 7*, 23-36.
- Rose, S. D., & Edleson, J. L. (1987). *Working with children and adolescents in groups: A multi-method approach*. San Francisco: Jossey-Bass.
- Rose, S. D., Tolman, R., & Tallant, S. (1985). Group process in cognitive-behavioral therapy. *Behavior Therapist, 8*(4), 71-75.
- Smokowski, P. R., Rose, S., Todar, K., & Reardon, K. (1999). Postgroup-casualty status, group events, and leader behavior: An early look into the dynamics of damaging group experiences. *Research on Social Work Practice, 9*, 555-574.
- Strauss, B., Burlingame, G. M., & Bormann, B. (2008). Using the CORE-R battery in group psychotherapy. *Journal of Clinical Psychology: In Session, 64*(1), 1225-1237.
- Taggar, S., & Brown, T. C. (2001). Problem-solving team behaviors: Development and validation of BOS and a hierarchical factor structure. *Small Group Research, 32*(6), 698-726.
- Weinberg, N., Uken, J. S., Schmale, J., & Adamek, M. (1995). Therapeutic factors: Their presence in a computer-mediated support group. *Social Work with Groups, 18*(4), 57-69.
- Weinhardt, L. S., Forsyth, A. D., Carey, M. P., Jaworski, B. C., & Durant, L. E. (1998). Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Archives of Sexual Behavior, 27*(2), 155-180.
- Whipple, E. E. (1999). Reaching families with preschoolers at risk for physical child abuse: What works? *Families in Society, 80*, 148-160.
- Whitney, D., & Rose, S. D. (1989). The effect of process and structural content on outcome in stress management groups. *Journal of Social Service Research, 13*, 89-104.
- Williams, D. D. R., & Garner, J. (2002). The case against "the evidence": A different perspective on evidence-based medicine. *British Journal of Psychiatry, 180*, 8-12.
- Witkin, S. (1996). If empirical practice is the answer, then what is the question? *Social Work, 46*, 109-119.
- Yalom, I. D. (1995). *The theory and practice of group psychotherapy* (4th ed.). New York: Basic Books.