

12/11/87

Harvard Community Health Plan
Group Cohesiveness Scale -- Version II

Stephen Soldz, Ph.D.

Elizabeth Bernstein, B.A.

Phyllis Rothberg, M.S.W.

Simon Budman, Ph.D.

Annette Demby, M.S.W.

Michael Feldstein, Ph.D.

Tamar Springer, B.A.

Lois Keithly, Ph.D.

Ted Powers, Ph.D.

Michael Davis, B.A.

HCHP - GCS II RATER MANUAL

Rationale

Much recent research on psychotherapy has focused on the relationship between process and outcome. Due to the complexity of factors affecting process and outcome, such research has often been difficult and elusive. This has been true especially of investigations in group psychotherapy. The multiplicity of interactions in group treatment make it particularly difficult to isolate relevant dimensions for study, and to develop valid and reliable measures of those dimensions.

The HCHP GCS-II is intended to measure a construct which has frequently been cited as critical to the successful functioning of therapy groups, namely group cohesiveness. Previous research on group cohesiveness has usually defined cohesiveness as the average attraction of group members toward each other (Stokes, 1983); cohesiveness as so defined is measured by subjective member ratings. In contrast, the GCS-II views cohesiveness as an observable characteristic of group functioning, which can therefore be rated by non-participant observers (and perhaps participant observers, such as the therapist as well). This observable phenomenon is defined by the Global Cohesiveness scale of the GCS-II as "group connectedness, demonstrated by working together toward a common therapeutic goal, constructive engagement around common themes, and openness to sharing personal material." Key components of cohesiveness are further specified in the GCS-II subscales as group processes of interest/involvement, trust, focus, facilitative behavior and bonding.

The authors of the GCS-II believe that a moderate amount of cohesiveness is a prerequisite for the successful functioning of any interactional therapy group. That is, group cohesiveness is a necessary (but not sufficient) prerequisite for beneficial therapeutic outcome. This hypothesis is based on clinical experience, on the clinical literature (Yalom, 1985), and on the rapidly growing literature on the therapeutic alliance in individual therapy (e.g. Marziali, Marmar and Krupnick, 1981; Hartley and Strupp, 1983). It is important to note that this hypothesis mainly applies to interactional therapy groups. That is, the GCS-II may not be applicable to other types of groups, or even to other types of therapy groups, such as those based on Gestalt therapy or behavioral principles.

Development of the GCS

Since 1982 a group of clinicians at the Harvard Community Health Plan have been developing a scale for measuring group cohesiveness. This group (the Process Research Action Group, PRAG) aimed to develop a measure of overt behavior which would be sensitive enough to capture critical aspects of members' interactions. A synthesis of contemporary professional views on the subject of curative therapeutic factors, cohesiveness in groups and therapeutic alliance in individual therapy were carefully considered. Extensive discussions by experienced clinicians in the PRAG group generated a list of factors related to cohesiveness that could be operationalized and rated by trained observers. Preliminary versions of the scales were piloted by the PRAG group, by a peer supervision group of experienced clinicians, and by trained clinician raters familiar with the Vanderbilt individual therapy process scales (O'Malley, Suh and Strupp, 1983; Sachs, 1983).

Clinical observations of videotaped group psychotherapy sessions were used throughout the development of the scale.

Revision of the Instrument

Several concerns led to revision of the GCS-I: (1) Despite initially satisfactory inter-rater reliabilities (subscale intraclass correlations from .68 to .85) our experience showed that new raters often had difficulty learning the scale. (2) Preliminary data indicated some lack of sensitivity of the instrument in the 0 to +2 range. This range is particularly important when rating early sessions. With a more sensitive instrument it was felt that important differentiation between "good" groups and "poor" groups could be made earlier in the process. (3) The revisions also aimed at decreasing the conceptual overlap amongst the subscales. (4) Additional concerns chiefly centered around problems with the scale's bipolar construction and ability to discriminate subtler forms of conflict and the expression of negative affect.

To make the scale clearer, more sensitive and more reliable the following major changes were made:

- (1) Scales which were originally bipolar were transformed into unipolar scales.
- (2) The original -5 to +5 scale was changed to a 1 to 9 scale.
- (3) The names of some of the subscales were modified, one scale was dropped (abusiveness vs. caring) due to difficulties in operationally defining it and overlap with the other scales, and one scale (bonding) was added.

- (4) Scales were provided with explicit definitions of the basic concept underlying each scale, and scale level descriptions were clarified.

Components of GCS-II

The GCS-II consists of one global scale and five subscales. The global scale measures cohesiveness as a whole, while the subscales measure four identified aspects of group cohesiveness; trust, focus, interest/involvement facilitative behavior, and bonding. Each scale ranges from Level 1 (very slight) to Level 9 (very strong), with explicit descriptions for odd-numbered levels. The definition of each subscale is meant to serve as a guide for inference when observed behavior does not clearly match scale-level descriptors.

Length of Segments

This scale was designed for use with 30-minute segments of 90-minute group psychotherapy sessions. Its applicability to longer or shorter segments is not known.

Choice of Media

Trial and error convinced the project staff that videotaped segments were far superior to audiotaped segments. For the purpose of individual therapy session ratings, audiotapes may be satisfactory. For ratings of group psychotherapy however, videotaped sessions are necessary. (Our group has not worked with transcripts of group psychotherapy sessions. It is possible that transcribed group psychotherapy sessions might be sufficient for obtaining reliable ratings.)

Training Procedures

The GCS-II is utilized by a trainer rater. The reliability and validity of the information obtained depend largely on the rater's initial training. This section will present the training methods which have been developed and will caution against known pitfalls.

The GCS-II rater should have had some prior exposure to clinical and research techniques. Exposure to group psychotherapy (as a patient or a therapist) is also helpful. People with M.S.W.'s, M.S.N.'s or Ph.D.'s have used the scale in the past.

There are five steps suggested for training: familiarization; discussion of the subscales and anchor point descriptors; discussion and practice of techniques; reliability testing; and periodic review. These can be accomplished with or without the aid of a previously trained rater although the presence of a trained rater helps assure the validity of the scale.

The beginning rater should first become familiar with all available material on the scale which includes this Manual and the Scale itself. A line-by-line discussion of the subscales and anchor points is useful initially to insure the raters understand the items. The raters should consider the purpose of each question and the area of behavior it covers. The anchor points should be discussed individually and examples of each agreed upon.

Agreement at this stage will help preclude misunderstanding after the 'real' rating phase has begun. Rating is an objective function. Personal opinions on what groups should be like are unavoidable, but must be kept separated from the behaviors being rated if reliability is to be achieved.

It is suggested that each rater make notes in his/her copy of the scale wherever personal experience might lead him/her to code in a manner other than agreed upon by the group.

Raters should be trained in groups, optimally; such training permits full discussion of the anchor points. Initially (stage 1) raters should view segments together, rate them individually, and then discuss their ratings. When adequate reliability is obtained at this stage, the raters should view segments separately (stage 2) and have the trainer (or designated leader) compare the ratings. This stage should continue until adequate reliability (intraclass coefficients of .7 minimum) can be consistently maintained. Reliability should continue to be assessed periodically during the rating process. Periodic reliability checks (duplicate ratings) should be performed.

During the first stage (in which raters view segments together), segments viewed should include examples of low, medium and high cohesiveness segments. At least two or three of each (six or nine in total) are suggested. In addition, a broad range of examples of group stage (early, middle, late) should be included.

Overall Rating Instructions

1. Focus on Group Interaction: Ratings should be made of group interactions, not of separate individual behaviors. For each dimension of group behavior being rated, the rater should ask, "What is the overall manifestation of this behavior by the group during this particular segment?"

2. Using the Scale level Descriptors: Group therapy segments present a virtually endless variety of patterns of group interaction, and patterns of variation over time. Since scale-level descriptors cannot possibly capture this variation across members and overtime, descriptors instead attempt to describe the overall pattern of behavior during a particular segment. In some instances, alternative ways of manifesting a behavior are specified. In attempting to match the unique patterns in a particular segment with a particular scale-level descriptor, the rater should ask, "Which descriptor most closely approximates the overall pattern for this segment, both across members and over time?".
3. Interpolation: Descriptors are only provided for alternating levels of the nine-level scales (1, 3, 5, 7, 9). If the overall pattern for a segment seems to clearly fall between described levels, the intermediate (undescribed) rating should be applied. Fractional scores may not be used.
4. Average among members: The more consistent group behavior is among members, the easier it is to match with a scale-level descriptor. The more inconsistent or "split" group behavior is among members, the more difficult it may be to determine which descriptor approximates the behavior displayed in a segment. In such instances of inconsistency, the rater must average the in consistent behaviors, or weigh their relative weight in determining an intermediate rating. Averaging should be done by weighing the relative intensity and duration of inconsistent

behaviors in determining a rating. That is, the rater should consider the extent to which the intensity or duration of behavior by certain group members offsets the intensity or duration of behavior by other members. For example, in a group where two members are intensively engaged for virtually the whole session while others are silent, the quality of silence or non-participation will be critical: if non-contributing members show non-verbal signs of interest and involvement and occasionally comment, an interest rating may not be significantly lowered; in contrast, if the majority of silent members seem utterly bored and disengaged, the interest rating will be somewhat lowered. When in doubt, unless there are clearly observable indicators (such as yawning or slumping) of disinterest, give members the benefit of the doubt.

5. Averaging Over Time and Recency Errors: Similarly, the more consistent group behavior is over time, the easier it is to match with a scale-level descriptor. Since cohesiveness subscale levels often fluctuate over a 30-minute period, the rater should consider actual minutes viewed at different levels. It is advisable to note the actual time at which one begins to rate a segment, and then note the time(s) at which changes in levels appear to occur. Some raters find it helpful to make preliminary or intermediate ratings on all subscales at 10-minute intervals. This also serves to reduce recency errors.

Experience indicates that raters tend to assign higher ratings to segments in which cohesiveness is rising rather than falling.

Raters should carefully average the amount of time spent at each level of a particular subscale. For example, for a segment in which 2/3 is a level 5, and the last 1/3 a level 8, the segment would receive a 6.

6. Rules of Inference: As far as possible, ratings should be based on observable evidence for a particular concept, not on inferences about that concept. That is, raters should be able to justify their ratings from observable behavior in the taped segments. Ratings should not be based on stereotypes or preconceived ideas about what groups should be like.

a) Ratings should be made independent of ideas about levels of pathology: Scale levels are not meant to correspond with different levels of pathology, or with different patient populations, but rather are behavioral descriptors. That is, while strong evidence of mistrust may appear more commonly in an acutely psychotic inpatient group, some poorly functioning outpatient groups may display such behavior as well. Similarly, while high levels of facilitating behaviors may occur more typically in higher functioning groups, quite disturbed groups may at times display clear focus, bonding or cohesiveness. Ideas about population-specific behaviors should not influence ratings.

b) Ratings should be independent of ideas about group stage: Raters should be blind to the stage of the group; any preconceived notions of stage-related

behavior should not bias ratings. E.g., the rater should not assign a rating that reflects the sense of the group functioning extremely well although it is a first session, but should simply rate what is seen.

c) Ratings should be independent of ideas about the potential of the session: E.g., a group may appear to be on the verge of a breakthrough. The rater should keep in mind that subsequent segments will reflect this change.

d) Prior knowledge of the group: Raters may become familiar with certain groups. An effort should be made to rate only the segment currently being viewed.

7. Non-Verbal Behavior: Non-verbal behavior should be taken into account in determining ratings as well as verbal behavior, even when scales do not specifically describe forms of non-verbal behavior accompanying a scale-level. In particular, quiet members often need to be rated on the basis of their nonverbal behavior.

Assume that a silent member who leans forward, maintaining eye contact, is interested and involved. As previously described, unless there are clearly observable indicators (e.g., yawning or slumping), give quiet members the benefit of the doubt.

8. Therapist Behavior: Although it is assumed that the group therapist may facilitate or hinder the development of group cohesiveness, as far as possible therapist behaviors should be excluded from ratings. That is, ratings may reflect member

behavior that results from therapist interventions but should not reflect therapist interventions themselves.

9. Rating Conflict and Negative Affect: In certain instances (i.e., in the Global and Facilitative Behavior scales), raters are asked to distinguish between constructive and destructive confrontation or expressions of negative affect. Ratings of conflictual material may range from high to low, depending upon the context and the intent. This distinction should be based on clinical judgment.
10. Referring to the Scales: Scales should be actively consulted in forming ratings. When observed behavior does not closely match the descriptors provided, scale definitions should be used as guidelines for forming ratings.
11. Number of Members Present: To the degree possible, rate segments independently of the number of members present. Again, "rate what you see." While low attendance may have implications for a group's cohesiveness, ~~the number of members present can be taken~~ into account in other ways.
12. Rate Process as Well as Content: In rating the subscales, the "how" (process) as well as the "what" (content) must be considered. E.g., an intellectualized discussion of fears regarding self-disclosure, with little affective involvement, would merit a lower rating on the Trust subscale than one in which appropriate affective involvement was observed.

13. Rate Each Subscale Independently: To the degree possible, view each subscale independently. A segment may be low on some subscales and high on others.

Specific Subscale Rating Instructions

1. Global Cohesiveness Scale: The global scale is intended to measure the rater's overall sense of the cohesiveness of the session. It is believed that cohesiveness may be more than the sum of the subscales. Thus, raters should not simply average their subscale scores in rating the global. In addition, while there are level descriptors in the global scale, there may be other dimensions of cohesiveness (e.g., depth, affective level or caring) that are not captured in the descriptors. These dimensions should be taken into account in rating this scale.
2. Focus: The following should be noted:
 - a) A higher rating would reflect both a strong sustained discussion of one topic, or a discussion of a variety of perspectives.
 - b) It is important to consider the nature of the transitions. Reaching closure before a transition to a different topic calls for a higher score.
 - c) The therapeutic effectiveness of the segment should not alter the focus ratings.

3. Interest Involvement: Ratings of interest and involvement should be independent of the topic discussed. For example, an animated, sustained discussion of sports, involving all members, should receive a high rating.
4. Trust: It is important to average the degree of risk taking of main actor with the responsiveness of other members. In some instances personal material may be discordant with members' responses. E.g., a member may reveal personal material despite others' discomfort or distancing.
5. Facilitative Behaviors: It is important to refer back to the list of facilitative behaviors in the definition, when arriving at ratings. Also, raters should be aware that this list is not exhaustive, and other similar facilitative behaviors may occur.

It is extremely important to rate ATTEMPTS to promote change and affective explorations, rather than the success of such efforts. The intent at the facilitative behaviors scale is to measure the extent to which members engage in behaviors aimed at promoting constructive affective exploration and personal changes in OTHER members; it is not a measure of efforts made by an individual member to work on their own issues.

6. Bonding: The lower pole of this scale may reflect EITHER indifference OR active disliking. Also, it should be noted that a group can be very friendly and warm WITHOUT therapeutic work taking place. It is important that ONLY the bonding be rated.

GCS: ADDITIONAL NOTES AND RULES

~~AUGUST 12, 1988~~

8-26-88

GENERAL ISSUES

1. The need for more precise definitions. The vagueness of descriptors leads to individual interpretation of the scale items. At times there can be equally valid but conflicting ways of understanding group behavior leading to poor reliability. One of the purposes of meeting together will be to more precisely define the concepts. These changes will be incorporated into the manual and/or scale descriptors.
2. Raters are using different levels of clinical inference. For example one rater gives a high score for focus because of a subtle underlying theme of loss, while another rater gives a low score because the group jumps from topic to topic. We need to be more explicit about what level of inference should be employed.
3. Difficulty in interpreting silence. Without verbal cues it is very difficult to judge the behavior of quiet members. Poor video quality and patients being off the screen make the task even more difficult. We need to be more precise in the manual about how to deal with this problem.
4. How to rate groups with low attendance. Raters are not to take attendance into consideration in cohesiveness ratings; we have other means of taking this into account. In making judgments one should consider the proportion of the group engaging in the behavior rather than the absolute number.
5. Poor audio quality. We decided that we should eliminate tape segments of poor audio quality from the dataset. Raters felt they would be able to make the following distinctions and that the last category would be considered "unrateable." We should add this check-off to the GCS Scoresheet.
OK _____ Questionable _____ Too Poor to Rate _____
6. A group with highs and lows on a given scale may end up with a mid-level score, just as a group that was mid-level all the way through might. There is a feeling that something isn't being captured in the rating, about the unevenness of certain dimensions.
7. Here-and-now interactions often merit higher ratings (especially on Trust, Facilitative Behavior and Global Scales). It is important to differentiate if it's constructive or not.
8. Universe of segments rules is still part of the process.

9. People are feeling more confident in their ratings recently as a result of the clarifications in the scale. However, they noted that it is taking longer to do ratings. This is due to the need to read the manual additions each time.

10. A rater expressed interest in using any additional time at COH meetings to discuss clinical issues that come up with the groups. There is concern that a discussion of this kind might bias raters when watching other segments of the same group. We decided to postpone further discussion of this until a time when more people are present.

11. Raters were given the most recent reliability data which consisted of interclass correlations done last week by 4 raters on 13 segments. Data was also presented comparing each rater with the mean of the other raters, and each rater paired up with every other rater.

12. While the data looks promising, Steve emphasized that we cannot draw too many conclusions because the n is so small. Also, the range of scores given for some scales is very small, so that in the statistical test we used even a one point difference is great.

13. The plan for the immediate future is to re-rate a subset of the data. If we receive more grant money we will be able to re-rate the entire data set.

14. The "Moments of Despair" form was distributed to raters for use in highlighting segments where ambiguity still occurs.

FOCUS

1. People are using different levels of abstraction. (see item 2, general issues). At some level there is always a connecting theme. We decided that from now on the connections between themes need to be obvious or articulated by group members in order for a group to be focused. If in doubt ask yourself if the group members would be able to see the connection. An underlying theme apparent only to a clinician would not count as focus.

2. How to judge transitions. Can a group with several topics be focused? We decided it could as long as the transition is smooth and closure is reached on the previous topic.

3. How much does depth factor into focus? Some raters are considering this more than others. We decided upon the following guidelines: At the lower levels depth is not a consideration. in order to score above a 5, however, there does need to be a logical buildup of material; a sense of the topic "going somewhere". At the 7 and 9 levels depth is a major factor.

4. When there are two different themes and abrupt transitions, the segment gets a lower rating. If there are two themes and closure is reached on the first topic, it could merit an 8 (not a 9) if there is a logical buildup and some depth.

5. If Therapist actively structures change (e.g. goes around the room and asks people to give feedback) how should this be handled? Clearly, the therapist often influences focus. If we really discount the therapist we still look at the amount of closure, the quality of the transition and the extent of connectedness that group members make or could reasonably make between the topics.

If the therapist introduces a focus and the change seems abrupt, with few or no connections to previous material, that lowers it. This does give the therapist more "power" in a sense. If the therapist is "cutting off" therapeutic work, this scale will capture the choppiness. (GTIS may pick this up also.)

6. If a discussion is focused in general, but an individual member is scattered and superficial, (style of presentation isn't going anywhere), this would lower the focus.

7. "Umbrellas" and "Laundry lists": In the beginnings of groups, typically members take turns and list their issues (often therapist provides "umbrella" of focus). If the connections between the presentations are not made, a rating of 4 is appropriate. If connections or other indicators of higher focus are present, then the rating can be higher.

Similarly, if a member lists "why I came to group," and there is an "umbrella" and connections made, but little logical development, buildup and little sense of "going anywhere," this is a 5.

If just a list is presented with an "umbrella" but connections between items on the list are not made (a connection would be "I think my self-esteem problem is related to my difficulty with relationships, another problem I have) the appropriate rating is a 4.

8. "Phony connections": If a member articulates a transition and the connection seems lacking in authenticity, (e.g. a member runs off on their own thing), raters must still judge the logical buildup, sustainment, abruptness, etc. and not consider the authenticity/phoniness of the articulated connection.

9. In case of several distinct topics, with virtually no connection between them and some coherence within a topic, but with very choppy transitions, the rating is a 4. If there is strong coherence within topics, the rating is a 5.

In a case where the transition is the result of an interpretive comment on the process e.g. "We're being very chatty today. I wonder if we're avoiding talking about termination." The transition is appropriate and doesn't reduce focus.

10. The words "theme" and "topic" have been used interchangeably in the latest additions to the manual (see pg. 3, #9). In both cases we are referring to the subject under discussion. The term "theme" does not mean the rater can use a higher level of inference, e.g. as in "an underlying theme of loss."

11. Focus, to nobody's surprise, is the scale with the lowest reliability. At least a third of the recent segments had a range covering 4 points in focus. Possible reasons for this might be:
A. Most people feel that despite the clarifications the concept is still the most nebulous.

B. Other scales have more explicit scoring guidelines at each level. With focus it is hard to be more specific than we already have been.

C. The judgements called for in this scale pertain to concepts where opinions vary greatly. An example is the direction to give higher scores to a topic that is "going somewhere". Everyone has different ideas about "where" the topic should "go".

D. It is harder to discount the therapist's effect on focus than it is with the other scales.

INTEREST/INVOLVEMENT

1. People had difficulty rating groups with low attendance. (see item 4-general issues). We will change the scale descriptors to read proportion of group participating, not absolute numbers.

2. Several people found the term "animation" ambiguous. What we mean by this term is a sense of liveliness or excitement. If many people are participating but with no apparent emotional investment the group would not get a high interest rating.

3. How to rate a group with several silent members. (see item 3-general issues). A rather extreme example was given where one member did all the talking for 30 minutes. Some raters averaged and gave a medium rating while others gave a high rating. We decided upon the following guidelines. In order to receive a rating of 8 or 9 the majority of members need to say something. 7 is the highest rating that can be given for non-verbal involvement. High involvement connotes verbal participation. If silent members are on the edge of their seats (non-verbal involvement), a 7 is the highest rating that can be given. To rate an 8 or 9, the majority of the members should be talking.

4. Split in group: If 4 people are very animated and three are quiet and uninvolved for most of the time, the segment must be averaged.
5. In the 7 descriptor, "most" means more than just a majority.
6. If the therapist implores the group to participate (in one segment therapist repeated this behavior several times), just rate what you see. Take the therapist out.
7. In a situation where one person (Main Actor) is intensely involved, (rating = 7), and other members' interest and involvement is slight (rating = 3) (in other words there are pieces from the 7 descriptor and the 3 descriptor that apply to the segment), averaging is called for. Since it is a GROUP scale, we decided to weigh the group feeling more and rate a 4 rather than a 5.
8. IMPORTANT: Eliminate "now and then" phrase from the Interest/Involvement #7 descriptor.

TRUST

1. "No Gut Feelings!": Some raters are following their gut feelings on this while others are using the scale descriptors. We said that while there can be many valid definitions of trust, for the purposes of this study we need to stick to the definition "openness to sharing personal material, and quality of responsiveness to this sharing."
2. It is hard to say what is trusting for different people. What may be easy for one person to reveal, may be difficult for another. We decided to take individual differences in difficulty with self-disclosure (SD) into account. Do consider the extent to which someone is having difficulty with SD.
3. The question of how to anchor responsiveness came up, especially the issue of whether to consider quantity vs quality of responsiveness. The scale defines 5 as limited responsiveness, 7 as "some" responsiveness and 9 as "highly" responsive. The rule we agreed upon is: If the quality of the responsiveness doesn't lower the MA trust level, then rate the MA trust level. If the responsiveness is, however, clearly limited (e.g. only factual questions, no risk-taking, intellectualized style, or clearly "out of sync" with MA) then the rating is lowered and qualifies as "limited." If the quality of the responsiveness is higher, that would raise the trust score.

4. Quantity issue: If, for example the MA bares his/her soul, and there is one clear, powerful, in-sync response but nothing else verbal....If the trust level isn't lowered, don't lower the rating just because of low quantity of responses, as long as the trust level is sustained. The rule is: The quality is weighed more than the quantity, especially if the responsiveness is limited in quantity.

5. Averaging: On the trust scale: We confirmed that we do indeed average parts of segments on the trust scale. If for example, there is a low level of trust in the first two thirds and a high level in the final third, these are averaged. Same is true if the trust level is high in the first third and low in the second and third.

6. Confrontation only raises trust when it is constructive, either in owning ones (the confronters) own likings or constructively encouraging a self-examination on the part of the other. Attacking or blaming confrontation is not constructive.

7. Moments of trust that are not sustained get a lower rating than sustained trust.

8. On 5 ratings "some" personal material involves both frequency and depth, not just frequency.

TRUST:

1. A patient's style of presenting should be taken into consideration in rating. For example, if a patient discusses deep personal issues in an intellectualized manner, the score would be lowered.

2. We discussed how to rate segments where the quality of responsiveness would merit a higher score than the level of trust shown by the main actor. An example was given where the group was extremely open and responsive to a main actor who had great difficulty in sharing. It was decided that as a general rule one should not rate a group more than one point higher on trust than the rating given to the main actor. However, there might be a few exceptional cases (we couldn't think of any offhand) where the group could be given a higher score.

TRUST AND FACILITATION:

1. Clarification of "ruthless confrontation" (see pg 6, #3). In instances where a group member confronts more than the target patient is clearly ready to hear, or confronts in an aggressive manner, the rater must judge whether this confrontation would be helpful to the "average" member. If you are not sure, give the situation the benefit of the doubt and assign a rating as if the behavior were facilitative and trusting. This item is still very

subjective and it was felt that there might continue to be problems with it.

INTEREST:

1. Under no circumstances can a rating higher than 7 be given if the majority of members are not participating on a verbal level. (see pg. 4, #3)

FACILITATIVE BEHAVIOR

1. One question raised regarded how to rate silence. Is silence which provides an opportunity for the Main Actor to "get things out" facilitative? The decision was a rule: Silence cannot increase the facilitation level. It can, however, lower it if silence indicates lack of responsiveness.
2. Another issue concerned how to rate support such as "It must be very hard for you" The decision was that, support is judged to be facilitation to the degree to which it appears intended to encourage further exploitation or personal change. (A point was made that the level descriptors are oriented toward questions.)
3. As regards "ruthless confrontation" the rule is the rater must judge whether the confrontation would be expected to lead to a constructive response in the average expectable group situation. If, in most situations, an unclear behavior would be facilitative, then all judge it, in the absence of clear evidence to the contrary. Otherwise, it is not facilitation. In ambiguous situations, give the benefit of the doubt.
4. Averaging issues were also raised. In particular, what about a segment where only a couple of people are very facilitative. In such a group, the rating depends on the quality of involvement of less active members. If they are clearly involved, the rating can go up to 7 (but not 9). If silent members are listless or emotionally absent, facilitation will be lower.
5. Sometimes advice giving is deep, insightful, personalized, not geared toward affective exploration but does promote behavioral or personal change. (e.g. "Perhaps you should confront your parents because its a serious issue for you.") This is more than factual/concrete.
6. We discussed a segment which exhibited extreme confrontation/attack: This segment had been accidentally rated by two raters with poor reliability. The issue raised was whether the ruthless confrontation should actually detract from the facilitation score or whether one should simply count the facilitation. The answer is to simply not count it and bring the facilitation down through averaging.

BONDING

1. One issue concerned the influence of conflict on the rating. It was decided that a conflictual group can be bonded if and only if there is a basic sense of mutual attraction, liking and warmth. For our purposes, pure hostile attraction is not bonding.
2. In one segment, all but one member exhibited strong attraction, but this one member was totally excluded. Lower the bonding rating in this case.
3. If the exclusion of member(s) is not as extreme, but there is disconnectedness between members--this also lowers the rating but not as much.
4. We discussed further how to rate a segment where one of more members are actively excluded, while the rest of the group appears highly bonded. (see pg. 7, #2). While everyone agreed that the bonding score should be lowered, how much should depend on the degree of exclusion, the number of people present, and how bonded the remainder of the group appears.

GLOBAL COHESIVENESS

1. Main issue: how to weigh different factors. Cohesiveness is more than the sum of the other subscales. (See manual p.13)
2. How to handle focus when good work is being done on one member's issues. If the exclusive focus on one member's issues reflects avoiding other members issues, this will lower the score. If it seems an authentic, legitimate, shared enterprise, this would raise the score. If Main Actor is defensive, difficult, unreceptive, this lowers the score. Cohesiveness does take into account the success of the efforts.
3. Example of members being sarcastic toward a member--this is destructive behavior if unprocessed.
4. Cohesiveness is explicitly cohesiveness in carrying out the task of a therapeutic group. That is, therapeutic work is required for a high rating.
5. If there is no therapeutic work but considerable engagement between members, the highest a rating can be is a 4.

Therapeutic work usually involves identification of a problem, elaboration of how the problem manifests itself, exploration of causes, reasons for the problem (external or internal), making connections between problem areas, and interpersonal learning through feedback between members or between therapist and a member.

6. Elements to be considered in rating Global Cohesiveness are many; some are: level of therapeutic work, extent to which the group is working as a unit, proportion of members involved or displaying a particular dimension.

7. Some concrete examples of situations and appropriate ratings:

A: Example of considerable engagement around superficial topic (low level identification of a problem), with predominantly advice giving interactions, and a high level of connectedness merits a 4.

B: Identification of a problem by one member, little connectedness between members merits a 3.

C: Group members make substantial attempts to connect with MA, who deflects all efforts; efforts reflect depth but MA is avoiding or denying. Merits a 6.

8. Higher levels of Global Cohesiveness require high level of work and connectedness.

9. Proportion of members displaying a particular quality or behavior must always be taken into account. Can't get a 9 with the "benefit of the doubt" rule.

10. Elaboration of definition of "therapeutic work" (see pg. 8 of old manual). We felt that the definition given in the old manual is too limiting, in that it implies that working on an individual's problem is the only work that can be counted. This led to a big discrepancy in rating a termination session where members were summing up and taking stock of their progress. We are adding a second paragraph to the definition to read "other instances of therapeutic work include a group discussion of how the group is functioning, discussion of members' concerns and expectations of group therapy and taking stock of progress within the group." In the first paragraph the word "usually" should be inserted.

2. (see pg. 8, #7). Omit 7B. This merely repeats the information in 7A, and adds confusion.

GLOBAL QUALITY

1. Everyone agreed that this is a more subjective "gut level" scale, which may capture other things than the Global Cohesiveness scale such as therapist's interventions and effectiveness, or unusual events. We noted that there is usually no difference, however, between Global Cohesiveness vs. Global Quality ratings, or perhaps a point difference, rarely more.

AFFECTIVE INTENSITY

1. An extended discussion of the Affect scale occurred. Among the questions discussed were: Is tension an affect? Yes, a very tense group would receive some score on Affect. How about depression? In this case one must distinguish sadness (an Affect) from listlessness and hopelessness, which are not rated on this scale. Regarding laughter, this would be rated if it truly represented positive feelings, as opposed to anxiety. In general the content of the group discussion should not affect the rating on this subscale. The scale measures explicit manifest emotion.

2. We agreed on a rule that a "baseline" normal conversational tone receives a score of 3 on the affect scale. If conversation is more intense (voice level, etc.) and/or animated, = 4, but still at low level.

3. "Intensely" involved means obvious intensity (e.g. higher affect, or display or internal struggle.)

4. The reliability on "affect" is poor despite the fact that most ratings are within one point. The reason is the constricted range. The descriptors as they now stand make scores other than 3 or 4 unlikely. To merit even a 5 rating, members need to cry or show other strong emotion that is not typically seen in these groups. Modifications to the "affect" scale are as follows:

Level 5: (baseline) should be given to groups where there is a moderate amount of animation and group member's appear engaged. No strong emotion needs to be present at this level.

Level 6: A brief episode of crying or overt anger would merit a rating of 6.

Level 7: Strong: A few instances of crying, anger or other intense emotion would qualify.

Level 8 and 9: Sustained intense emotion such as crying or anger.

OTHER NOTES REGARDING AFFECT

1. We discussed the problem of how to rate depressed affect, which can sometimes be hard to distinguish from boredom. We decided that explicit expression of sadness, wavering or cracking voice, or teary eyed expression are indications of strong emotion and should be scored as such. In contrast a low voice, monotone, without other clear indications of sadness would merit a score of less than 3.

2. Another area of difficulty with the affect scale occurs when the group's affect is not synchronous with that of the main actor. In general, one should use the affect level of the main actor if the rest of the group is not out of synch. Note: In synch does not necessarily mean displaying the same emotion, but does take into consideration receptivity, empathy and not hindering the emotionality of the group.³ It is sometimes difficult to distinguish between nervous and convivial laughter. No resolution was reached on this problem.

CONFLICT

1. We decided that when the therapist introduces conflict (e.g. with a challenging, tension-producing question), with the effect of increased tension in the group, this is to be taken into account, NOT excluded, from the rating.

2. We confirmed that a challenging question and some tension is often a 3.

FOCUS

Measures the degree to which the presentations reflect a common agenda with thematic coherence. Focus implies that the content of members' contributions reflects a logical development and build up of material, the consideration of various points of view, with understandable transitions. A lower score would reflect a topic left hanging or clearly unfinished.

1) Very Slight

Discussion reflects no common agenda. Silence, unconnected or individualized presentations.

3) Slight

Tangential relationship among themes.

5) Moderate

Some associations between themes with moderate coherence. or a confused presentation by one member with others making clear attempts to focus the presentation.

7) Strong

Discussion of a common theme and logical buildup of material, but with brief or slight digressions, or some unevenness. A variety of perspectives may be present.

9) Very Strong

Sustained discussion of a topic. The session has clear thematic coherence. A variety of perspectives are attended to and developed.

INTEREST/INVOLVEMENT

Measures the degree to which members exhibit verbal and non-verbal interest in and involvement with the group discussions.

1) Very Slight

Predominant silence with non-verbal withdrawal (e.g. no eye contact, listless posture).

or explicit statements of total disinterest.

or negativistic avoidance of discussion.

3) Slight

Several members are slightly involved.

or One to two members are involved with the rest of the members apathetic or uninvolved.

5) Moderate

Most members paying attention, with some observable signs of interest.

or 1 or 2 members involved with some unevenness in interest exhibited by others.

7) Strong

Discussion somewhat animated.

Most members interested and involved, in an animated way.

or one to two members intensely involved with most others clearing interested and participating now and then.

9) Very Strong

Virtually all members intensely involved, speaking frequently and/or listening attentively.

TRUST

Measures the degree to which members are open to sharing personal material. Trust involves risk taking and/or self disclosure which may result in vulnerability or being seen in a negative light. Trust is also reflected in the quality of responsiveness to members presentation of issues. Constructive confrontation for example, involves risk taking.

1) Very Slight

Total inability to share personal material demonstrated by blocked silence.
or explicit statements of acute fearfulness. A paranoid quality may be present.

3) Slight

Some interchange about impersonal issues (e.g. sports, dieting techniques).
or a few instances of revealing personal material in an atmosphere of discomfort or guardedness.
or explicit discussion of fears of sharing with little exploration of underlying reasons.
(e.g., career choice discussion that centers on concrete advice and problem solving.)

5) Moderate

Some disclosure: Most of the discussion involves an issue of some substance where moderate risk is involved.
or some personal material brought up with limited responsiveness from others.
or group discussion of difficulty with self disclosure with some exploration of underlying reasons.
(e.g., career choice discussion acknowledging feelings and/or beginning to search for underlying causes.)

7) Strong

Deep personal material is discussed with some risk taking and some responsiveness.
or fear of sharing discussed with extensive exploration of reasons.
(e.g., career choice discussion centers around deep seated feelings of insecurity.)

9) Very Strong

Members very open and responsive to sharing deep personal material.
or all members highly responsive to one or two members sharing of deep personal material.
(e.g., career choice reveals clear vulnerability and/or negative aspects of self.)

FACILITATIVE BEHAVIOR SCALE

Measures the extent to which members engage in behaviors aimed at promoting constructive affective exploration and personal change, in other members (e.g. supporting, helping, constructively confronting, nonverbally responsive, or exploring personal material.)

This scale measures attempts to promote affective exploration; not the success of such efforts.

1) Very Slight

No facilitation of therapeutic work shown by no engagement. or active resistance, e.g. subgroup conversations, or hostile attacking behavior (e.g., talk about sports.)

3) Slight

Slight efforts are made to facilitate therapeutic work. (e.g., solely asking factual questions.)

5) Moderate

Some effort to facilitate therapeutic work (some attempts to examine underlying causes or elicif feelings.)

7) Strong

Significant efforts are made to facilitate therapeutic work.

9) Very Strong

Strong efforts to further therapeutic work with attempts to deepen affective exploration.

BONDING

Measures the extent to which members appear connected to each other, characterized by mutual attraction liking and warmth. This rating is made without reference to the therapeutic work, or depth of material being discussed.

1) Very slight

A strong sense of indifference, separateness; members may appear repelled by each other.

3) Slight

Tentative presentations reflecting slight engagement with or responsiveness from others; a cool aloof quality.

5) Moderate

Some sense of mutual liking, and mutual interest: A lukewarm quality.

7) Strong

Clear sense of mutual attraction, liking and warmth.

9) Very Strong

Very strong, consistent, mutual attraction; much warmth is present.

NB: Manual should contain last two sentences from original bonding definition: A group can be very friendly, warm and so on without therapeutic work taking place. It is central that for this scale only the bonding is rated.

11/30/87

HCHP - GCS II

GLOBAL COHESIVENESS

Measures group connectedness demonstrated by working together toward a common therapeutic goal, constructive engagement around common themes and openness to sharing personal material.

1) Very Slight

No efforts to connect with others toward a common therapeutic goal demonstrated by total inactivity or withdrawal.
or extreme fragmentation.
or extreme destructive conflict.

3) Slight

Predominantly individualistic interactions without extreme hostility, or active disruption, but with very slight attempts to connect with others and work toward a common therapeutic goal.

5) Moderate

Some attempts to connect with others and work together toward a common therapeutic goal but with limited constructive responsiveness, openness and depth.

7) Strong

Substantive attempts to connect with others around common themes with constructive responsiveness, openness and depth.

9) Very Strong

Strong efforts to connect with others and work together toward a common therapeutic goal with consistent empathic responsiveness, marked openness and depth.

Global Quality of Sessions

Rate how good this segment was as a group therapy session. That is, how closely does it match your ideal of group therapy process?

- 1) TERRIBLE SEGMENT
- 3) NOT VERY GOOD SEGMENT
- 5) MODERATELY GOOD SEGMENT
- 7) VERY GOOD SEGMENT
- 9) IDEAL SEGMENT

Affective Intensity

- 1) VERY SLIGHT INTENSITY (Virtually no emotions expressed as visible. A sense of boredom or listlessness.)
- 3) SLIGHT INTENSITY (A slight amount of emotional involvement. May talk with some interest on a topic, but no emotional depth.)
- 5) MODERATE INTENSITY (One or two members show a fair amount of emotion without others exhibiting much affect or most members exhibit moderate affect.)
- 7) STRONG INTENSITY (A general emotional tone. Many statements of members are expressed emotionally, but with little expression of very strong emotion.)
- 9) VERY STRONG INTENSITY (A general tone of very strong emotion, e.g. crying, anger, very fond feelings.)

Conflict

Disagreement or challenging behavior which also involves increased tension. Both the elements of disagreement and negativistic, nonverbal behaviors on the one hand, and tension on the other hand, must be present. A disagreement that involves no discomfort does not constitute conflict. Similarly, a tense group without disagreement or negativistic nonverbal behaviors, does not constitute conflict. The rater is not asked to judge whether the conflict is destructive or constructive.

- 1) VERY SLIGHT CONFLICT (Members are either polite and cooperative or animatedly discussing a topic. No negativistic nonverbal behaviors.)
- 3) SLIGHT CONFLICT (Mild disagreement or a little challenging behavior or some negativistic nonverbal behavior, such as bored, uninterested looks, with slight tension.)
- 5) MODERATE CONFLICT (Several instances of challenging behavior with some sense of intermember tension in the group.)
- 7) STRONG CONFLICT (Most members disagreeing and arguing with others. A sense of tension in the group.)
- 9) VERY STRONG CONFLICT (Very argumentative/angry. Yelling or physical acting out such as throwing something or storming out.)

HCHP GCS-II SCORESHEET

Rater _____

Date _____

Master Tape # _____

Segment # _____

1) Focus

1 2 3 4 5 6 7 8 9

2) Interest/Involvement

1 2 3 4 5 6 7 8 9

3) Trust

1 2 3 4 5 6 7 8 9

4) Facilitative Behavior

1 2 3 4 5 6 7 8 9

5) Bonding

1 2 3 4 5 6 7 8 9

6) Global Cohesiveness

1 2 3 4 5 6 7 8 9

7) Global Quality

1 2 3 4 5 6 7 8 9

8) Affective Intensity

1 2 3 4 5 6 7 8 9

9) Conflict

1 2 3 4 5 6 7 8 9

