Scientific interest in groups started in the late 1800s with the burgeoning of the social science disciplines, although scholars had been observing groups for centuries. Pioneering scientists from the fields of economics, political science, anthropology, sociology, and psychology each uniquely incorporated a particular slant of group examination in their study of human behavior and institutions (Forsyth & Burnette, 2005). Their efforts, though, were not without theoretical and methodological controversies. Some felt that the study of group behavior was too ambitious when scientists had such limited understanding of the individual. Understanding of human behavior at the time was centered on stimulus–response learning and conditioning, and scientists had little understanding of the relationship of mental states to nervous system functioning. Others questioned the existence of group-level processes and argued that group behavior could be understood as the sum of its individual components (Cartwright & Zander, 1968). Only after decades of research demonstrated the existence and powerful influence of group-level processes such as norms, leadership, and cohesiveness did respected skeptics such as Floyd Allport rescind their protests (Allport, 1924, 1961). The study
of how and why groups change over time is little more than a half century old.

In this chapter, we first discuss methodological issues affecting research in group development. The acquisition of knowledge in the area of group process has followed the typical pattern of other areas of social science and has been intricately intertwined with technological advances. To appreciate the current state of the knowledge base, a brief examination of some of the questions and issues facing group development researchers is essential.

Second, we describe the early research efforts on group development. The work of Bales (1950; Bales & Strodbeck, 1951) and Bion (1961), who proposed a cyclical model of group process, is contrasted with the work of others who supported a progressive model (Bennis & Shepard, 1956; Stock & Thelen, 1958). Tuckman's (1965; Tuckman & Jensen, 1977) elaboration of a stage model based in part on his empirical review of the literature marked a transition from this early period to the contemporary involvement of a more rigorous empirical methodology.

Third, we present three types of empirical evidence that support a stage theory of development. The study of therapeutic factors looks at how individual group member’s perceptions of the importance and presence of these factors change over the life of the group. The examination of leadership patterns involves the centrality of specific types of group leaders and their systematic changes over the course of the group. In a third group of studies, the maturity of the group (number of sessions) is linked to characteristics of the group climate in student and organization groups with evidence for a model of progressive development. This section is followed by the evidence and writing that does not support a progressive stage model.

A progressive stage model is only useful if it has some relationship to the success of the group and its members. In our final question, we tackle the issue of whether outcome, individual or group success, is related to a group that is functioning at a more advanced level of development. We present evidence from inpatient and outpatient groups, process and training groups, and organizational groups. We conclude that there is a resoundingly positive relationship between stage of development and outcome.

THE CONTRIBUTION OF RESEARCH TO THE STUDY OF GROUP DEVELOPMENT

In any field of study, a summary research chapter often is presented simply and elegantly. It rarely captures the uncertainty, tentativeness, and downright messiness on which the scaffolding was built. To do so would mire the reader in too many details of the empirical problems common to all areas of study and prevent researchers from making any meaningful statements about the subject matter as they continuously diluted the strength of their
assertions. It is important to know that although researchers in the area of group development have worked hard to advance their knowledge base, they have and are continuing to overcome many obstacles to their study. We briefly discuss the conceptual issues, problems in methodology, technological advances, and interpretation of findings.

Conceptual Issues

Each theory of group development that is presented is a coherent story of how a group changes over time. Take, for instance, this description (with the text references removed) of the second of five stages in a theory of group development:

The second stage is described as a period of counterdependency and conflict. Issues of power, authority, and competition are debated at this stage . . . these early struggles regarding authority and status are prerequisites for subsequent increases in cohesion and cooperation. (Wheelan & Williams, 2003, p. 444; emphasis added)

This description illustrates the difficulties a researcher has in providing empirical support for that statement. Each italicized term requires considerable deliberation.

First, consider the conceptual question highlighted by the use of stage. A developmental process has the potential to be continuous or composed of stages or steps. A belief in a discrete change versus a continuous one will determine how one goes about studying the process. Does the group progress through discrete stages (e.g., MacKenzie, 1990), or does it cycle back (e.g., Bales, 1950; Bion, 1961)? If the latter is the case, under what circumstances does that phenomenon occur? If a particular stage is not manifest, what does this absence mean for the group process? Can other stages occur or is the group’s growth stunted? Is there a way a group can compensate for a lost stage?

Second, testing the veracity of a theory requires that each term be operationalized, so that it is observable and measurable. The references to time (e.g., “period of counterdependency,” “subsequent”) require specification of a time frame. Should time be measured in hours or by sessions or purely by stage characteristics? How does one decide when to take the measurements? Is each of the stages approximately equal in time? Terms such as power, authority, and competition can be captured in a measure of thematic content, but they can also refer to an interaction between individuals. How does one measure a power struggle between individuals? How does one decide what a debated issue is or when it is necessary to label such an issue as a struggle for status? Should content or process be measured? Are both counterdependency and conflict needed, or is one enough to label the interaction as Stage 2?
Another important conceptual issue concerns how one defines the purpose of the group (MacKenzie, 1994a). When a laboratory study uses a clearly defined task (e.g., designing a solution to a complex problem), success can be determined by some external measure, such as the time to reach a solution or the number of solutions found. The outcome is a group product. In a psychotherapy group, however, success is based on how many members feel better (e.g., 5 out of 8 patients no longer meet criteria for depression). The outcome is the composite of each member's health, which is basically an individual outcome. Is the same process of development required for successful completion of a group task versus an individual one?

Problems in Methodology

Methods of observation, correlation, and experimental intervention are all integral parts of the empirical method. As discussed in the section later in this chapter, "Early Study of Group Development," early work relied primarily on observation and personal experience for theory construction (e.g., Bennis & Shepard, 1956; Bion, 1961). Authors' findings were impressionistic, subject to personal bias and a priori conceptualization. Later studies tended to use empirical methodology, focusing on more objective measures of the strength of a relationship between two variables. Empirical methodology requires that constructs be operationally defined, and this definitional enterprise may involve tackling important but thorny conceptual issues. For example, can one track group process by looking at individuals only, or must there be some measure of the group climate? Should investigators rely on members' perceptions or require independent behavioral analysis?

Instruments measuring these concepts require creative thought, cross-validation, and proper establishment of psychometric properties (e.g., reliabilities and validities) to enable the professional community to communicate deftly and precisely about the same phenomenon. To bring such an instrument to fruition can take a decade of a researcher's life, as has been the case for Beck's sociometric measure of leadership roles (A. P. Beck, 1974, 1981b). Once the instrument is available, how does one determine when to implement its use? Is it sufficient to administer it once or are several equally spaced administrations necessary? If the data do not yield the hypothesized findings, is it because the theory is incorrect or has the timing of instrument administration been incorrectly determined? The well-standardized instrument often mandates a standardized procedure of data collection. However, the often unacknowledged dark side of any methodological paradigm is that it can only report on the questions it asks and cannot speak to the procedure it does not use or the data that it does not collect, thus potentially limiting what researchers "see."

Studies using a true experimental design are still rare in the area of group development. By far the most common design is to examine an already...
organized group in progress (Wheelan, Burchill, & Tilin, 2003) created for a purpose other than the pursuit of a scientific question. The lab groups created by Bales (1950) and Kivlighan and colleagues (Kivlighan & Lilly, 1997; Kivlighan & Mullison, 1988; Kivlighan, Multon, & Patton, 2000), enabled a focus on variables of significance.

Technological advances such as the widespread use and acceptance of videotaping have allowed for improved methods of recording data. The use of tapes and other electronic products is less subject to human error than is note-taking. Tapes permit independent review, enable establishment of interrater reliability, and entail less intrusiveness than a group observer who makes notes. Permanent recordings also facilitate the recoding of new variables at some future time. Regardless of the technological advances, group development research is labor intensive; considerable time is consumed identifying units of behavior and categorizing them. For example, in the Wheelan, Davidson, and Tilin study (2003), transcripts from 26 groups were transcribed, and 31,782 units of communication were identified and catalogued into one of eight categories.

Statistical Advances

The widespread use of computers and availability of statistical packages has greatly enhanced researchers’ ability to track and analyze a great many more variables than was previously possible when all statistical analyses were calculated by paper and pencil. Bales and Strodtbeck (1951) had little statistical help available to them. Their chi-squares and creative use of transpositions were cumbersome compared with today’s more technologically driven methods. They concluded that their groups were not following a random pattern but could only hypothesize judiciously about phases in problem-solving groups. Later multivariate analyses supported the study of a greater number of variables but could only do so if the relationships were linear. These methods limited researchers in tracking complicated systems.

Within the past few years, growth curve analysis has made possible the examination of variable relationships that may be curvilinear in nature (Kivlighan & Lilly, 1997) and enables researchers to track individual members over time (see the methodological discussion by Burlingame, Fuhriman, & Johnson, 2004). Nascent efforts of the application of chaos or complexity theory (see chap. 3, this volume) to the area of group development have allowed investigation of the usefulness of multidimensional visual representations of the group system (Wheelan & Williams, 2003). These new paradigms of analysis have facilitated the greatest amount of growth and knowledge in this area because they have encouraged researchers to consider group change as a much more complex process.
Interpretation of Findings

Thus far, each research team has developed its own model and methodological paradigm to study group development. In addition, group therapy researchers have been for the most part unaware of their counterparts in social psychology and organizational development (and vice versa). As a result, little cross-fertilization has occurred in terms of theory and empirical work.

Although awareness of others investigating group process has increased, the question still remains as to whether the findings of one set of investigators on one type of group are applicable to another type. Is the development similar in work groups, training groups, therapy groups, and laboratory task groups? Does the length of time allotted for the task affect stage development? Does the number of members in the small group influence the ability of the group to progress through stages? How does leadership impact development? Is group process affected by demographics such as gender, age, intelligence, status, and ethnicity? How does group process affect outcome? With little intergroup exchange, many of these questions remain insufficiently (and in some cases barely) explored. In the remainder of this chapter, we examine what knowledge has been empirically acquired in these areas. We hope that different research groups will begin to recognize the value of each other's contributions.

EARLY STUDY OF GROUP DEVELOPMENT

Robert Bales (1950) is credited with the earliest empirical work in group development. Using single-session laboratory groups run as problem-solving exercises, Bales developed a system to code interaction patterns in small groups. His interactional process analysis was composed of 12 categories of behavior, which collapsed into four broader areas: attempted answers (gives suggestion, opinion, or orientation), questions (asks for orientation, opinion, or suggestion), positive reactions (shows solidarity, shows tension release, or agrees), and negative reactions (shows tension, shows antagonism). The attempted answers and questions areas pertain to member activities that are instrumental to the assigned task. Positive and negative reactions areas are expressive functions that regulate emotional tension. Bales observed that as the group attempts to work on the task, tensions build. These tensions begin to interfere with task resolution unless dealt with by emotional activity. Such activity then permits the group to return to the task with renewed energy. Bales identified two leaders in this process: the task and the socioemotional leaders. The instrumental and emotional functions oscillate over time, an observation that led Bales to conclude that there are recurring cyclical patterns in small groups. He felt all groups balanced both task and socioemotional...
activity. Given the brief span of Bales’s groups, MacKenzie (1994a) averred that the cyclical pattern observed by Bales was not two different stages of group development but represented a micropattern that could be expected in all stages of group.

Bales and Strodtbeck (1951) observed 22 single-session problem-solving groups occurring in the natural environment. Using their 12-category system to code interaction, they concluded that “when considered as an aggregate [they] show a significant departure from a random distribution of acts between phases” (p. 494). Their descriptive statistics suggest that orientation statements are highest in the first phase and then progressively drop. Evaluation statements remain quite high throughout the phases. Positive and negative reactions and control statements increase over time with positive statements always remaining proportionately higher than the other two. On the basis of these findings, Bales and Strodtbeck offered a three-phase developmental progression. In the initial stage, members are concerned with orientation and the giving of information. This is followed by a period of information evaluation. A decision is made in a phase that involves control, during which there is support for some opinions and rejection of others. These findings and conclusions form the foundation on which most small group studies rest. MacKenzie (1994a) has indicated the seminal nature of Bales’s work.

Bales’s recognition of the importance of the group roles of task and socioemotional leadership links group development to the characteristics of the members. It deepens the understanding of the relationship between the individual and the group that was first identified by Lewin. The two theoretical approaches of group development and group roles have formed the principle axes for understanding group level process. (p. 229)

For the next 2 decades, a number of authors, such as Bennis and Shepard (1956) and Bion (1961), primarily clinicians, wrote of their observations and experiences concerning group development (for a review of their work, see chap. 3, this volume). Thelen and colleagues, using Bales’s instrumental and expressive functions, devised observational categories based on Bion’s theory (Stock & Thelen, 1958). They rated natural units of group interaction1 (referred to as the Behavioral Rating System) and collected group members’ perceptions of the group and members (Reactions to Group Situations Test). Although their interests were primarily in finding empirical support for some of Bion’s concepts, such as valency and basic assumption groups, they concluded that the group balanced its efforts between basic assumption states and work, and they came to believe that as each group assumption state emerged, the group dealt with it in an increasingly mature way (MacKenzie, 1994a, 1994b). In their final published efforts, they did find

---

1Units were determined by observers’ judgment of a natural break in the content and participation of the members.
preliminary evidence for stages of group development and furthermore pro-
vided some evidence to support their view of the interplay between the group-
advancing stages and individual and subgroup obstructions (Stock & Thelen,
1958).

In summary, each of these three clinician-scholars provided thoughtful
observations about group process. Bennis and Shepard (1956) supported the
notion that groups move through successive stages. Bion (1961) appeared to
reject this model of developmental progress in favor of one that is more cy-
clic, underscoring that the group periodically regresses from the task to struggle
with socioemotional issues. Stock and Thelen (1958), although initially sup-
porting Bion’s approach, did find some support for a stage model.

Tuckman (1965) conceptualized a developmental model of group pro-
cess by organizing more than 50 articles that dealt with either research data
or theory. He separated his review into therapy (26 studies), training (11
studies), and natural and laboratory groups (collapsed because there were too
few studies in each) and summarized the literature for each type of group. He
identified four stages of development that could be discerned from each of
these types of groups. In keeping with Bales’s division of socioemotional and
task functions, he characterized each stage by a description of the interper-
sonal or group structure realm and the task realm. These stages—forming,
storming, norming, and performing—are described in chapter 3. Although
Tuckman’s review included therapy, training, and natural and lab groups,
the last group is underrepresented, so it is difficult to assert that his postu-
lated stages are characteristic of all types of groups. In addition to the limited
number of studies, evaluation of the experimental rigor of those studies in-
cluded is lacking. Most of those reviewed were observations based on single
groups. These observations were more descriptive and subject to the impres-
sionistic biases of the observer, who was usually the therapist. The standards
for qualitative studies had not developed to the rigorousness is the case today
(Josselson & Lieblich, 2003). Tuckman used the equivalent of the box tally-
ing method of analysis to support his assertions, counting the number of pub-
lished studies to support his theory. This method gives Bales’s seminal work
the same weight as therapists’ observations of their group. The use of meta-
analysis was not to occur for another 15 years. At the time it was written,
Tuckman’s article was an important summation of a disjointed field, even
though by his own admission his conceptual system was merely suggested by
the data rather than supported by it.

Although Tuckman is most known for his stage sequencing, he at-
ttempted to find commonalities among the different types of groups (labora-
tory, natural, psychotherapy, and training) and to identify some ways in which
these types of groups might manifest differently across the developmental
stages. He also discussed the importance of addressing separately the task
aspect and the interpersonal aspect (or group structure) for the continuity of
development to be recognized. He noted that particularly in the therapy and
training groups, this separation is difficult. Tuckman's thoughts in this regard are helpful because they set the stage for an examination of how developmental phases might appear in group process groups versus more structured groups such as cognitive-behavioral groups (for further discussion, see chap. 7, this volume). For example, speaking of the second stage of development, Tuckman wrote, "The task stage will be most evident when the task has as its goal self-understanding and self-change, namely, the therapy- and training-group tasks, and will be considerably less visible in groups working on impersonal, intellectual tasks" (1965, p. 386).

A decade later, Tuckman followed up on his original research review (Tuckman & Jensen, 1977). In his survey of 22 new empirical papers, he found only one that actually attempted to test his theory. Runkel, Lawrence, Oldfield, Rider, and Clark (1971) investigated three groups of 15 to 20 college students. Each group was to decide on a project, collect data, interpret it, and write a final report. Unfortunately, observers were given only descriptions of Tuckman's four stages and were instructed to fit what they observed into one of the stages in the model. Results did support Tuckman's theory of group development. However, rather than being given Tuckman's model and being asked to fit their observations into that system, the observers should have been requested to record their observations independently.

Most of the other studies reviewed supported at least some aspects of Tuckman's theory. Only two studies substantially deviated from the four-stage Tuckman model. One was a 9-month study of two sections of a social relations course. Groups were self-analytic. Using a computer content analysis, Dunphy (1968) identified six stages, three of which could be folded into Tuckman's second phase. The biggest discrepancy was that there was no stage similar to performing. The other study involved psychiatric patients in an 18-session group (Heckel, Holmes, & Rosecrans, 1971; Heckel, Holmes, & Salzberg, 1967). The data pointed to a two-stage group process. However, significant methodological flaws were present, including the attrition of one half of the members. Tuckman concluded that what little empirical work had been published generally supported his proposed model.

CONTEMPORARY EFFORTS TO TEST PROGRESSIVE STAGE MODELS

In the 1980s, additional efforts were made to formulate models of group development (A. P. Beck, 1974; MacKenzie & Livesley, 1983; Wheelan, 2005). Each group, conscientious in its efforts to concretize its models, worked to develop methodology to allow empirical investigation. Each theory predicts particular changes in group process over time, and observations ensconced within that contextual paradigm mark whether these shifts actually occur. These researchers' collective contributions to this effort have resulted
in three areas of inquiry that generally provide support for a progressive developmental model of group process. These areas, the study of therapeutic factors, leadership roles, and group climate, are examined in the following sections.

**Therapeutic Factors**

Although not designed to track stage development, changes in members' perceptions of the importance of various therapeutic factors over the life of the group provides support for group progression.

**Measurement of Therapeutic Factors**

One empirical approach to tracking changes in groups is through an examination of how member perceptions of their significant experiences may change over time. The first method, referred to as the *direct approach* (Bloch, Reibstein, Crouch, Holroyd, & Themen, 1979), descended from the therapeutic factors devised by Corsini and Rosenberg (1955) to delineate a taxonomy of potential therapeutic processes. The technique was modified and extended to have terminating group members complete a questionnaire or Q sort to determine retrospectively those aspects of the experience perceived as important (Bloch et al., 1979; Yalom, 1995). Yalom's (1970, 1995) Q sort requires the placement of 60 statements representing 12 therapeutic factors into piles from most to least helpful. This procedure allows for the comparison of relative importance among factors. Yalom (1995) proposed that the salience of particular therapeutic factors would vary as a function of developmental stage (i.e., the instillation of hope would be critical at the commencement of a group but less prominent as the group progressed). Procedurally, this is a time-consuming and difficult task for the individual. Yalom has provided little psychometric information, and the research that is available suggests that items sometimes do not statistically correlate highly with their assigned category.

A second approach, the *critical-incident technique*, requires group members to describe the most important or critical event of the session. This more indirect method infers a therapeutic factor from the incident, and then either categories are developed to capture the sample or the incident is placed into a predetermined classificatory scheme. The critical-incident technique is quickly and easily administered, allowing for multiple administrations. It has the potential advantage of examining a change process over time and can accommodate a revision of the classification system as incoming data may support. However, it requires a minimum level of writing skill, engagement in the task, and adequate intrater reliability of critical incident placement into categories, and it assumes that a series of single incidents for each individual will capture the complexity of the group change process (for an excellent review, see DeLucia-Waack & Bridbord, 2004).
A third more recent and promising method of measuring is the Therapeutic Factors Inventory (Lese & MacNair-Semands, 2000). It uses a 7-point Likert scale and has participants rate 99 questions (9 questions for each 11 factors). It has good internal consistency and test-retest psychometrics. Although based on Yalom's theory and factors, the Therapeutic Factors Inventory asks members to rate the presence of a factor rather than to follow Yalom's original instructions to rate the relative helpfulness of a factor.

Evidence of Group Progression

An initial study by Freedman and Hurley (1980) involving 28 undergraduates in a 9-week course (50 hours) used Yalom's Q sort at the initial, middle, and final meeting. They found that with increasing group experience, catharsis and interpersonal output (improving interpersonal skills) were perceived as more helpful. Similarly, a study by Butler and Fuhriman (1983) found that members' interactions deepened with continued group involvement. They used the Curative Factors Questionnaire (variation of the Q sort) to sample a cross section of 23 therapy groups (91 patients), which had been meeting for a minimum of 6 months prior to their study. Members in therapy groups meeting for a longer time were more likely to value acceptance (cohesiveness), self-understanding (insight), and learning from interpersonal actions (interpersonal learning output). Finally, MacNair-Semands and Lese (2000) had members of 15 therapy and support groups complete the Therapeutic Factors Inventory on Sessions 3 through 6 and 8 through 12. All therapeutic factors increased with time in the group; universality, instillation of hope, imparting information, recapitulation of family, cohesiveness, and catharsis increased significantly. Although these studies do not provide specific support for a stage model, they did observe changes consistent with most progressive stage models.

Support for a Stage Model

The Butler and Fuhriman (1983) study assumed that data obtained in a single session from a cross section of groups meeting for varying lengths of time were equivalent to data collected from a single group over time. In an effort to correct for this difficulty, MacKenzie (1987) used a longitudinal design with four outpatient psychotherapy groups. In this study, 34 members completed a critical-incident form at the end of the first 20 sessions. Five sessions equally spaced across that time period were chosen for analysis. Twelve categories of therapeutic factors were collapsed into three groups: morale (acceptance, instillation of hope, and universality), self-revelation (self-disclosure, catharsis), and psychological work (self-understanding, learning from interpersonal actions, and vicarious learning). MacKenzie found psychological work increased and morale decreased with advancing sessions. Self-revelation initially dropped and then increased to its highest levels with more advanced sessions. This pattern was interpreted to mean that the initial self-
disclosures, which were likely superficial, dropped, and the later revelations occurred with deeper self-exploration (MacKenzie, 1987). The MacKenzie study was the first to find evidence for a stage model rather than simply a progression as the previously mentioned studies had found. These trends suggest the importance of common factors initially and the growth of deeper self-reflection later (MacKenzie, 1997). However, the patterns were reported in percentage endorsed and were not subject to statistical analysis. The interpretation of the phasic nature of self-disclosure, however, is a demonstration of how this methodology alone is not able to capture the complexities of this phenomenon. Another problem with this study is that interrater reliabilities reached only modest levels even when categories collapsed into three groups (kappa coefficients ranging from .28 to .35).

Using a longitudinal design over 11 sessions, Kivlighan and Mullison (1988) studied three groups (18 students) with the critical-incident technique. They found that universality was an important factor early in the life of the group, and interpersonal learning was more highly valued later in group life. It should be noted that groups were analyzed using a split-half design (Sessions 1 to 5 early, 6 to 11 late) that may have masked sensitive changes in the group climate (MacKenzie, 1994a).

In the previous studies (Butler & Furhiman, 1983; Freedman & Hurley, 1980; Kivlighan & Mullison, 1988; MacKenzie, 1987; MacNair-Semands & Lese, 2000), length of time in therapy is equated with advancing stage of development, which is not necessarily the case. In an attempt to rectify this problem, Kivlighan and Goldfine (1991) used the Critical Incident Questionnaire to study six process groups (36 students completing a group psychotherapy course). Groups met for 1 1/2 hours twice a week for 13 weeks. Six judges using the framework of MacKenzie's model of group development and the Group Climate Questionnaire (GCQ; MacKenzie, 1983) made determinations as to when each group entered each of the first three stages. (The model is described in detail in the section on group climate later in this chapter.) Critical incidents were classified into 1 of 10 therapeutic factors (Bloch et al., 1979). Results provided partial support for the predictions that Yalom (1995) made, derived from a developmental model.

As predicted, groups showed significant decreases in universality and hope, suggesting that feelings of hopefulness and universality are important in early group development but are less so later. Unexpectedly, the importance of guidance increased over the course of development (consistent with Kivlighan & Mullison, 1988). The authors hypothesized that the absolute levels of advice and suggestion (the main constituents of guidance) did not change, but group members recorded them increasingly as critical events when they were articulated in a more personal way (later stage) in contrast to a general, less specific, and personal manner (initial stage). Also as predicted, an increase in catharsis occurred over the first three stages of development. This finding is consistent with the notion that as participants explored more
personal issues (a characteristic of a later stage of development), they experienced an affective release. Somewhat contrary to expectation, acceptance was an important factor both at the initial and later phases of group development, with it being less so during the middle phase; acceptance is essential when initially engaging with the group as well as when examining more personal concerns.

In this study, interpersonal learning, self-understanding, vicarious learning, altruism, and self-disclosure were not related to group development. Because these were not therapy groups, it is possible that certain mechanisms did not come into play or that a larger sample would have revealed these differences. In addition, these findings may reflect that individuals with different interpersonal styles may value different aspects of the group experience. The lack of change in interpersonal learning over the duration of the group is especially surprising. The authors explained this latter finding in terms of the possibility that the group did not move into the later stages in which interpersonal learning may be especially strong. The possibility of this outcome, that the group may not progress, shows the difficulty of creating tests for the existence of developmental stages.

The studies using therapeutic factors as the dependent variable to measure changes in groups provide some support for the notion that groups change in definable and measurable ways over the life of the group. However in this context, three limitations exist. The first problem is conceptual. By definition, therapeutic factors measure individuals’ perceptions about what is important to them about the group experience. Developmental theory would predict that these perceptions would change in a systematic way; by and large the research supports the theory. At the same time, developmental theory predicts that the processes themselves would change over the life of the group. Unfortunately, the prominence of these processes can be ascertained merely through inference when the only data collected are members’ self-reports. Second, members may be influenced by the therapist’s view of what is important. The therapist’s developmental model may be imposed on members’ perceptions (Sandahl, Lindgren, & Herlitz, 2000). The third problem is the lack of methodological rigor and small sample sizes characterizing most of the studies. Only the Kivlighan and Goldfine (1991) study subjects the data to statistical hypothesis testing. Despite these limitations, this collection of studies represents a significant advancement in terms of positive empirical support for the concept of group development when compared with the observational and anecdotal work of the earlier decades.

Leadership Patterns

With underpinnings in general systems theory (J. E. Durkin, 1981), client-centered therapy (Rogers, 1970), and developmental psychology (Piaget, 1960), Ariadne Beck and her collaborators developed and elabo-
rated on a theory of group development spanning several decades beginning in the late 1960s (A. P. Beck, 1974, 1981a, 1981b; Brusa, Stone, Beck, Dugo, & Peters, 1994). This theory was based on both informal observation and participation in groups as well as careful study of small (6 to 9 members), time-limited therapy (15 to 20 sessions), training, and encounter groups (A. P. Beck, 1974). These closed groups were relatively unstructured from the onset, and the client-centered leader intervened at both individual and group levels. This context enabled the evolution of a group-as-a-whole dialectical process over the life of these short-term groups, out of which emerged a consistent, discernible, and invariant sequence of interactional patterns.

Beck hypothesized the potential for nine phases in which certain unique structures, tasks, and leaders emerge, which may aid or thwart individual and interpersonal changes (A. P. Beck & Lewis, 2000c). Beck and colleagues’ nine-phase model is described in chapter 3 of this volume. Beck posited that the appearance of the nine phases as well as the patterns within each were necessary for the group to function successfully and achieve its stated goals. In her theory, Beck focused on group level issues, the process, and emerging leadership roles as ways of differentiating each of the phases of development.

Beck articulated four leadership functions, each of which is usually fulfilled by a single individual for the life of the group but in their absence would be taken up by another individual. The designated or task leader is usually the official leader or senior therapist of the group. He or she must be interpersonally skilled and emotionally supportive. The emotional leader, important throughout the life of the group, helps manage the emotional life of the group and is often most ready for significant personal growth. The scapegoat leader, often insensitive to the more subtle cues in communication, becomes the target for negative affect from the group. His or her struggle with self-assertion versus group conformity crystallizes for the members their own conflicts in this area and in doing so helps the group clarify norms and goals, particularly during the second phase. His or her style forces the group during other phases to communicate more directly. The defiant leader, usually an individual who is experiencing psychological and social alienation in his or her outside life, expresses conflict about dependence versus independence. Demanding to be cared for beyond reasonable limits of the group, he or she is reluctant to be involved in a mutual relationship with peers to satisfy these needs. As noted in chapter 3, different leaders emerge at different phases. For example, the emotional leader emerges in the first phase of development and is critical to members’ forging initial connections with one another. In Phase 5, the defiant leader, a member who is threatened by intimacy and dependency, becomes prominent.

This detailed theory is potentially empirically testable with A. P. Beck’s (1974) enormous and rich data set of approximately 36 groups. To date, however, only a handful of studies have been published primarily around emerging leadership functions; some of these are demonstrations of how the tenets
of her theory could be empirically tested but would not meet the standards of hypothesis testing (A. P. Beck, Eng, & Brusa, 1989; A. P. Beck & Lewis, 2000a, 2000b, 2000c). Her reports have often captured an inside view of the triumphs and disappointments that occur in research. Tenacious, tedious, intensive data analysis has yielded tentative, but promising findings.

Central to Beck and colleagues’ theory of group development is the hypothesis that different leadership structures emerge over the nine phases. To test this hypothesis, A. P. Beck and Peters (1981) devised a brief instrument that could differentiate each of the leadership roles. The development of this instrument illustrates the typical manner in which empirical work proceeds in a clinical domain—a maneuvering back and forth between clinical observation and sociometric data occurs. A series of questions using a sociometric technique was compared with clinical judgments of leadership. Using 6 of the original 18 questions, Peters and Beck (1982) could clearly distinguish the designated and emotional leaders from all others, but the defiant and scapegoat leaders could not be differentiated from each other. If Beck was correct that different phases require differentially active roles by the four leaders, then completing these questions only once at the end of the group may have obscured changing perceptions. In an attempt to rectify this problem, A. P. Beck et al. (1989) analyzed the questions by phase. Each group was given 7 to 10 questions once. Of the 36 groups analyzed, the use of 3 questions (rank-order group members according to the pleasure derived from each, in terms of the risks they take, and according to how ambivalent they are about being in group) nicely differentiated each of the leaders over the span of the nine phases proposed by Beck. However, only question means for each leader were graphed by phase because the questions were administered only once per group. This sociometric effort has undergone further development. In a study by Brusa, Stone, Beck, Dugo, and Peters (1994), 273 group members (in 31 psychotherapy groups) answered a subset of the standard 18 questions. Groups were clinically evaluated to be in Phases 2 through 6 and 9. A discriminant function analysis of the four leadership role assignments yielded a 97% hit rate with clinically identified roles. The scapegoat leader was the least accurate with a hit rate of 81% (Brusa et al., 1994). Although promising, this work has not yet been replicated or cross-validated with another sample of group members.

Beck’s work not only supports her particular model of group development but also the more general idea that the life of the group is organized into a series of stages.2 It should also be underscored that Beck’s model, in

---

2C. M. Lewis and Beck (1983) also explored the use of the Experiencing scale developed by Gendlin (1967). They were interested in the relationship of individual self-exploration and the process of group development. Preliminary findings indicated that the scale reflects changes from one stage to another. For example, turn-taking of self-exploration occurs in Phase 3. Although promising, no further publications have appeared on this project; it is discussed briefly in the “Leadership Patterns” section in this chapter.
Group Climate

The therapeutic factors approach asks the question of what individual members have found helpful. Systematic changes in what has been helpful can point to group development. The leadership pattern approach examines the emergence of certain leadership structures at different times in the group. Who fulfills these functions and when this is perceived by its members has been offered as evidence of group development. Examining the atmosphere in a group over time shifts the focus of investigation away from the individual onto the group as a whole.

Three research groups led by A. P. Beck, Kivlghan, and Wheelan, respectively, have attempted to investigate group development by using the concept of group climate. Each group, whose work is explored in some detail, has a slightly different theory of group development, and as needed, these theories are summarized to make the findings understandable.

Studies of Beck's Model

A. P. Beck and her colleagues, whose theory was delineated in the previous section, characterized each of the nine phases by identifying leadership functions, interpersonal processes, and intrapersonal processes and group-as-a-whole processes (A. P. Beck, 1974). Although Beck has not directly attempted to study group climate, one of her goals was to operationalize the phases and transitions between them (A. P. Beck, 1983; C. M. Lewis, Beck, Dugo, & Eng, 2000). However, even demonstrating the transition from one phase to the next is a formidable task. In preparation for analyzing these transitions, the research team needed to articulate behavioral units and validate instruments to measure unfolding processes. The initial focus of the Beck group was on the transition from Phase 2 to Phase 3, which was concretized as three criteria: movement from competitiveness to cooperation (interpersonal); movement from defensive behavior to mutual exploration (intrapersonal); and movement from struggle around organizational and norm development to concentration on personal concern (group as a whole). Measures of interpersonal hostility and support (Hostility/Support scale) and group-as-a-whole transition from norm concerns to personal concerns (Normative–Organizational/Personal Exploration Scale) were developed. The Experiencing Scale (Gendlin, 1967), a measure of intrapsychic exploratory behavior, was adapted (A. P. Beck, 1983; Dugo & Beck, 1984; C. M. Lewis & Beck, 1983). Each of these scales engaged outside raters who required training and acceptable interrater reliability. The laborious process of tape or transcript review then began.
The results to date have been mixed. For the three groups in which this method was applied, two demonstrated clear transition with all criteria met. The third group was reported to have difficulties with development, but the specifics were not presented (A. P. Beck, 1983; C. M. Lewis et al., 2000). Without a larger sample, it is difficult to know whether the criteria of phase transition would be supported. Beck and colleagues have expressed belief that the application of empirically validated instruments will catalyze researchers’ understanding and aid in refinement of how group structures are created and evolve. It appears, however, that the labor-intensive nature of the specific instruments used will significantly limit how quickly this research group will be able to share their findings.

Another exciting aspect of what the Beck group has done is to collaborate with those who have other systems of measuring process and demonstrate how each system would analyze the same session (A. P. Beck & Lewis, 2000a, 2000b, 2000c, 2000d; C. M. Lewis & Beck, 2000). This particular demonstration highlights the confluence of various systems to provide layers of information about a single session. However, its application could readily be instituted to explore the development of group process over time.

Studies of the MacKenzie and Livesley Model

On the basis of their observations and data, MacKenzie and Livesley (1983, 1984) articulated a group development theory that includes five stages plus termination. MacKenzie (1983) also developed a methodology to study these group development changes, the GCQ. The short version of this questionnaire (GCQ–S), with 12 items, uses a 7-point Likert scale. This questionnaire contains three factor-analytic subscales—Engaged, Avoiding, and Conflict. The Engaged scale encompasses engagement, support, self-disclosure, challenge, and cognition related to cohesion. The Avoiding scale captures the avoidance of responsibility in the change process. The Conflict scale represents interpersonal conflict and distrust. Its reliability and validity are adequate (MacKenzie, 1983, 1990). It is the most frequently used instrument to measure group process in growth and psychotherapy groups.

Kivlihan and colleagues used the GCQ–S to test the MacKenzie and Livesley model. Kivlihan and Goldfine (1991) studied changes in therapeutic factors with student growth groups. In that study, the GCQ–S was also administered after every session. It was then used by judges to determine the stage of each group as well as the trends of each of its three component factors. For the six groups, 12.3 sessions was the mean for Stage 2, and 17.7 sessions was the mean for Stage 3. Using multivariate analysis, researchers found engagement increased linearly over the three stages; avoidance decreased linearly over the three stages; and conflict, which was initially low to moderate, increased to higher levels and then decreased close to initial levels. No differences among the six groups in these trends were obtained. These
findings are consistent with MacKenzie’s (1990) description of the first three stages of development.

During Stage 1, scores on the Engagement and Conflict subscales are relatively low, whereas Avoidance is relatively high. During Stage 2, Conflict is high, whereas Engagement and Avoidance are at a moderate level. In Stage 3 in which members begin to explore more personal concerns, Engagement is relatively high, whereas Avoidance and Conflict are relatively low. Although the groups in this study were growth groups composed of college students, an analysis of group climate data revealed no significant differences between this sample and psychotherapy groups reported on by MacKenzie (1983).

In a later study with a similar but larger sample size (83 participants, 14 groups) and format, Kivlighan and Lilly (1997) used growth curve analysis, a more powerful alternative to the earlier multivariate analysis of variance for analyzing repeated measures. Some of their findings were consistent with their earlier study. However, the larger sample size and the ability to take into account therapeutic gain enabled a more sophisticated analysis. Unlike the earlier study, which was not able to distinguish any significant differences in climate patterns among groups, this study found that there were no consistent growth patterns for all groups, although there was high agreement among group members as to their climate. Rather, patterns of engagement, conflict, and avoidance were related to therapeutic gain. Also groups varied in total sessions completed (14 to 26). Number of sessions did not influence group pattern. Rather, within limits, groups of varying lengths followed similar developmental patterns. Other findings in this study are discussed in the next section.

Using an advanced statistical procedure known as Tuckerizing, Brossart, Patton, and Wood (1998) analyzed a preexisting data set similar to the one used by Kivlighan and Lilly (1997). This method preserves individual input while at the same time allowing for group similarities to emerge. Brossart et al. also attempted to evaluate the degree of congruence between the MacKenzie and Livesley (1983) model and the members’ reported experiences. Using four groups of students that met for 28 sessions, investigators had participants complete the GCQ-S after each session and also complete questionnaires addressing their goals and issues. Only the Conflict subscale was analyzed. The results revealed high levels of conflict occurring early in group life and then dropping during the individuation stage. Although the MacKenzie and Livesley theory predicts an increase in conflict, they suggested that it is likely to occur more gradually than was manifested in this study. Inconsistent with the theory, Brossart et al. found high levels of con-

---

3One problematic aspect of this methodology was its circularity: It appears that the GCQ-S was used by judges to determine stage of the group, and these stages were the groupings that were then used to determine the relative impact of each of the three factors at each stage.
Conflict during the later stages of group development (mutuality), which surpassed Stage 2 (differentiation) levels. The use of three different growth curves allowed for the interpretation that only some members may have been experiencing high levels of conflict, whereas others may have reached resolution in the early stages of the group.

**Studies Using Wheelan's Model**

Wheelan (2005) formulated her own scheme of how groups change and develop across time using Tuckman's (1965; Tuckman & Jensen, 1977) review and her own empirical work with both nonprofit and for-profit organizations. In her five-stage model, Stage 1 deals with issues of inclusion and dependency as members endeavor to recognize behavior acceptable to leader and other members. During Stage 2, identified as the counterdependency stage, issues of power, authority, and competition are focused on. Struggles with authority and status are viewed as essential requirements for ensuing increases in cooperation and cohesion. Stage 3 has an atmosphere of trust, in which there is more open and mature negotiation of goals, roles, group structure, and division of labor. In Stage 4, designated as the work stage, an increase in task orientation, more open exchange of ideas and feedback, and enhanced goal achievement occur. When a distinct ending to the group is planned, Stage 5 is peppered with disruption and conflict as positive feelings are expressed and issues of separation are explored. Wheelan offered supporting preliminary evidence for the validation of these stages (Verdi & Wheelan, 1992; Wheelan & McKeage, 1993; Wheelan & Verdi, 1992).

Wheelan created the Group Developmental Observation System (GDOS) to code members' verbalizations. Eight categories were derived from the work of Bion, Thelen, and Stock. Each complete articulated statement is placed into one of the categories—dependency statements, counterdependency statements, fight statements, flight statements, pairing statements, counterpairing statements (which are avoidance of intimacy), work statements, and unscorable statements. Agreement between independent coders was 85% to 95%. Wheelan also designed the Group Development Questionnaire (GDQ) to measure group change or development. Consisting of 60 statements, it measures her first four stages of group development. Validation of scale included establishment of retest reliability and internal consistency as well as concurrent, construct, and criterion-related validity (Wheelan & Hochberger, 1996). Higher scales (representing higher stages) are associated with higher productivity and effectiveness ratio. Likewise, a higher productivity and effectiveness ratio is correlated with lower scores on the earlier scales (Wheelan & Tilin, 1999).

Wheelan used these two instruments to study ongoing for-profit and not-for-profit organizational groups in their natural settings. To address the question of whether a relationship exists between members' perceptions of their groups' development and length of time that the groups have been
meeting, Wheelan examined a cross section of groups meeting for 1 to 15 months prior to the study (Wheelan et al., 2003). In the first sample, 26 work groups (180 individuals) from 12 organizations completed the GDQ. Two groups were judged to be in Stage 1, two groups in Stage 2, 12 groups in Stage 3 (meeting an average of 5.2 months), and 12 in Stage 4 (meeting an average of 8.5 months). In the second sample of 88 groups (639 participants) meeting for 1 to 7 months prior to the study, each member completed the GDQ at the end of the meeting. Of 88 groups, 27 were judged to be in Stage 1 (meeting an average of 2.6 months), 14 in Stage 2 (meeting an average of 3.7 months), 28 in Stage 3, and 19 in Stage 4. Both Stage 3 and 4 groups had been meeting for more than 4.6 months. The investigators obtained a significant difference in both samples with regard to the age of the group and members’ perceptions of group development. Statistically, Stage 1 groups met significantly less time than Stage 3 and 4 groups (Wheelan et al., 2003). The investigators also found that as the size of the group increased, members were more likely to see the group as having the characteristics of the earlier stages of development but that individual factors such as age, gender, education, and length of service were not related to perceptions of group development. The authors had no explanation as to why in the first sample Stage 4 groups had been meeting for an average of 8.5 months whereas in the second sample, Stage 4 groups had been meeting on average 4.6 months.

All 26 groups in the first sample were audiotaped, and their statements were coded using the GDOS. Group stage was determined by members’ perceptions (GDQ). Members were also asked how productive they felt that their group was on a scale of 1 to 4. Individual and group level analyses were performed. Results revealed that the age of the group was correlated with the number of dependency, fight, and work statements that individuals made in the group. Groups that met for less time were more likely to have individuals who made more dependency and fight statements and fewer work statements. A group-level analysis produced consistent results. Groups with higher proportions of dependency and fight statements had met for less time. Conversely, groups with higher percentages of work statements tended to have met for more time. Members of older groups also perceived themselves to be more productive than those who had been meeting for less time. Analysis of members’ perceptions of the group development revealed similar findings. Members of groups meeting for longer durations tended to see their groups as having fewer Stage 2 characteristics and more Stage 3 and 4 characteristics. Members who saw themselves to be in Stage 3 made more counterdependent and fight statements than members who saw themselves in Stage 4 of group development. The longer a group had been meeting, the greater number of Stage 4 characteristics members perceived in the group and the fewer characteristics of Stage 2 groups.

This study is an important one in that it not only used much larger sample sizes than had earlier studies but it also combined both independent
and objective assessments of group members' behavior (verbalizations) and the more subjective measure of members' perceptions of group interaction. The study also expanded the applicability of the notion of group development beyond the clinical or college volunteer population to include functioning adults in a work setting. It provides support for the traditional models of group development, demonstrating that both verbal behaviors and members' perceptions of the group climate change in predicted directions across time.

With a sample similar to that in the previously described groups, Wheelan and Williams (2003) used a promising and exciting new technology for data analysis to examine group development. These investigators used nonlinear methods derived from complexity theory (see chap. 3, this volume) and research to visually map the communication dynamics of groups during a single meeting. A sample of 108 participants involved in 16 ongoing company groups (group age 1 to 30 months) were audiotaped and completed the GDQ. Transcriptions of the audiotapes were coded on the basis of the GDOS. One group was perceived by its members to be in Stage 1, 8 groups in Stage 3, and 7 groups in Stage 4. The coded statements from each group were recorded in sequence into the Chaos Data Analyzer software. A wavelet transform test constructed a color visual representation of the dynamic patterns of each group. The final product was a color image of the verbal exchange sequence that transpired during the group meeting. Using complexity–chaos theory as a model, six individuals inspected the images visually for similarities and differences. Shape, color, overall tone, and visual clarity were compared.

The patterns that emerged could be placed into three tiers. The only characteristic that differentiated Tier 3 groups from the others is that 80% of them had been meeting for 4 months or less. In Tier 1, communication patterns were longer than in the other two tiers. There were also fewer patterns because work and pairing statements composed 93% of the verbalizations. The authors interpreted this finding as “members of these groups were working and supporting each other's ideas and suggestions with regard to the group's task” (Wheelan & Williams, 2003, p. 458). For Tier 2 groups, the overall tone was more active and erratic and less symmetrical than that of Tier 1 groups. Tier 2 groups manifested more communications patterns and a greater number of shifts in those patterns than Tier 1 groups. Work and pairing statements still constituted the vast majority (82%), but they occurred in more attenuated bursts interspersed with brief recurring patterns of flight or focus on members' reactions to the leader. Tier 3 groups have the most erratic patterns. Of their statements, 66% were still work and supportive statements, but they oscillated with frequent patterns of flight and conflict or patterns that emphasized members' reactions to the leader. The researchers saw this pattern as the group's struggle to preserve the direction of the discussion and stay on task. Statistically, significant differences in percentage of work state-
ments between each tier were obtained. Tier 1 and 2 groups also had a lower percentage of fight and flight statements compared with Tier 3 groups. Tier 1 group members had proportionately fewer dependency statements than Tier 3 group members. These objective findings are consistent with members' perceptions of group functioning. Members of Tier 1 groups perceived themselves to be functioning at a higher level of group development than members of Tier 3 groups. This perception speaks to a major assumption of group developmental theory that group-level phenomena exist.

Group development theory predicts that groups occupying different stages of group development would communicate differently; the results of this study support this hypothesis. Although all groups began with a work and pairing or supportive communication configuration, members of Tier 1 groups were able to maintain a work focus for a longer period of time. For Tier 2 groups, pattern shifts were more frequent. Work and pairing were interspersed with periods of flight. Neither the work and pairing patterns nor the flight patterns increased over the life of the session. In Tier 3 groups, pattern changes were frequent. Unlike the other two tiers, work and pairing patterns were disrupted by conflict. Each pattern of flight increased the duration of the successive pattern of flight, thus exacerbating difficulties for group members to return to the task.

Summary

Until recently, analysis of group climate was limited by linear research methods. The Wheelan and Williams study and that of Kivlighan have allowed for the analysis of more complex relationships in the group. Kivlighan used a growth curve analysis that could examine effects at both individual member and overall group levels. This study introduced chaos theory to the study of group development (McGrath, Arrow, & Berdahl, 2000; Wheelan, 1996b). This approach more accurately allows groups to be seen as complex adaptive systems. These constellations are not the result of individuals, but rather they are unique patterns created by their joint interactions. The GDOS system enables the generation of visual representations of group development process, permitting the possibility of using constructs of chaos theory. Early findings seem to suggest that although commonalities exist from group to group with respect to process, each group has its own unique process trajectory as well.

The three lines of evidence (therapeutic factors, leadership, and group climate) provide support for the model of progressive stages. The weakest support is provided by the therapeutic factors studies both because the research findings are inconsistent and because the predictions from the progressive models are not direct and unambiguous. Measures of therapeutic factors were not designed to delineate stages of group development, although future researchers may provide a rationale for the prediction of particular patterns of factors for each stage. The data on leadership patterns hold prom-
ise, particularly with support by other research groups using different venues to provide further validation. The group climate studies are especially useful because the methods used (especially the GCQ-S and the GDQ) were devised specifically to examine developmental phenomena. Consequently, predictions of how groups should perform on these measures over time can be made with greater directness than with measures that have evolved outside of a group development framework.

NEGATIVE EVIDENCE FOR PROGRESSIVE DEVELOPMENTAL STAGES

The evidence for progressive developmental stages is not entirely positive. Two lines of research give those supporting a stage theory of group development some pause. The first questions were raised in the classic studies of Bales and Strodtbeck (1951). Seeger (1983), in his review of problem-solving groups, distinguished between the process of problem solving and the group formation process. He argued that problem-solving groups generally did not follow sequential stages unless they had not previously met. He saw phase development as a reflection of formative group process development and not what is required to complete the group’s task.

Cissna (1984) reviewed the literature and found that 13 studies failed to find developmental trends. He reviewed each of the studies that he was able to locate and found that in addition to methodological flaws, many of them assumed that particular sessions (often equally spaced) would reveal certain stages. Many also measured aspects of group life that researchers now know may not be revealing of changes in developmental phenomena (e.g., source of communication; relative frequency of speaking; whether the utterance was an assertion, request, or proposal). Cissna concluded that groups have a unique developmental time frame of development and that it is possible that not all groups are successful in proceeding through the identified stages. These conclusions do not discredit a theory of group developmental phenomena.

The second line has led to the development of the punctuated equilibrium theory as an alternative to sequential group development. Gersick (1988) studied eight field teams that were working on short-term projects. Teams met for varying lengths of time from 7 days to 6 months. Every meeting was observed, audiotaped, and transcribed. The transcribed text of the meeting was analyzed using qualitative methodology without preconceived theoretical dimensions. The analysis of the data did not support a progressive stage model and led to an alternative model. Gersick observed that each group has a distinctive approach to its task, which it maintains through a period of inertia that extends to about half the time allotted. Members of the group maintain a “habitual routine,” and little is accomplished until external pres-
sures change it. This event is followed by a transition in which goals and objectives of the team are completed. Gersick proposed to call what she saw the punctuated equilibrium model (see chap. 3, this volume). She admitted that her study design may have influenced the process. Team members could have been affected by the project-imposed time constraints, which could modify the normal process of group development.

In a second study, Gersick (1989) observed eight master of business administration student groups as they worked on creative projects. All of the groups generated similar patterns. Each group developed an interactional style, which continued until the midpoint. At the midpoint, each group underwent a transition in which goals and methods were examined and evaluated. This new interactional pattern was maintained until the end, concluding with a high level of activity in the last session. The author interpreted these studies to indicate that the framework in which the group took place, particularly the time context, was a significant factor in molding the group change process and asserted that individual, group, and organizational levels of analysis might have been bridged by this model (Gersick, 1991). Small sample size, narrow participant type, and open-ended observational analysis are possible criticisms of this study. It has also been suggested that level of observation affects whether change is seen as incremental or punctuated (Wollin, 1999). Seers and Woodruff (1997) amplified this distinction of levels of observation. They distinguished between the Tuckman (1965) tradition of sequential development as representing the evolution of psychosocial dynamics and Gersick’s studies, which focused on task progress. They contended that Gersick’s studies potentially confounded group development and task completion, delineating little about group behavioral patterns. They supported their contentions with two studies that demonstrated that task pacing and group development are distinguishable phenomena, although sometimes they may co-occur; studying groups that are task driven with deadlines may muddle researchers’ understanding of group development.

These studies are significant and should give pause to what otherwise appears to be almost an unopposed view. They suggest that in certain contexts alternate models to the progressive stage model may provide a more descriptive account of how groups change over time.

IS PROCESS LINKED TO OUTCOME?
DOES DEVELOPMENT MAKE A DIFFERENCE?

Thus far, we have concentrated our focus on the evidence for a progressive model of development in a group. Our review suggests that although the findings of some studies are at odds with the existence of progressive developmental stages, the preponderance of the evidence supports the descriptive value of this concept. With that point relatively well established, we need to
ask whether development will make a difference in members' outcomes. If group development is a major phenomenon in defining the group, then a group's development should have some important linkage to outcomes. Until recently, this question presented a formidable challenge because to examine it directly required a dynamic rather than static analysis (Kivlighan & Lilly, 1997). The only methods previously available were pre–post difference scores or repeated measures of multivariate analysis. The first method estimates amount of change rather than actually measuring the process of change, and the second is plagued by a host of statistical problems (Kivlighan & Lilly, 1997). Now the availability of hierarchical linear modeling allows for one way of appropriately examining the change process in a study using repeated measures. We review the evidence that development is indeed linked to outcome for psychotherapy groups, training and process groups, and organizational groups in their natural setting.

In the review that follows, two types of studies are presented. One type links process variables with outcomes without specifically determining the stages through which the group proceeded. These studies are relevant because the processes studied have developmental implications. If engagement increases, the presumption is that the group is showing greater maturity. However, a linear increase in engagement is actually more consistent with a continuous change model (unless the group has failed to go beyond Stage 1). Hence, a given study may support development–outcome relationships, but not a stage development–outcome model. The second type of study explicitly examines developmental stages in relation to outcome.

**Psychotherapy Groups**

A series of studies have measured process and outcome relationships (see Table 4.1). Although all fall short of being able to unequivocally demonstrate that a particular sequence influences outcome, they generally support the notion that to the extent that a group develops, members are able to accomplish their goals. The studies ranged in length of time in group, meetings per week, closed versus open membership, theoretical orientation, and the country in which the study took place. We review both inpatient and outpatient group psychotherapy studies.

**Inpatient Studies**

Groups conducted in inpatient settings are of particular interest because outcomes of members of the group are likely to be affected by many factors and therapeutic involvements that are different from those that affect the outpatient psychotherapy group.

Tschuschke, MacKenzie, and colleagues conducted two studies examining process and outcome relationships in an inpatient setting. MacKenzie and Tschuschke (1993) took process measures of two psychoanalytically ori-
## TABLE 4.1
Studies Investigating Developmental Process–Outcome Relationships

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Format</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inpatient therapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacKenzie and Tschuschke</td>
<td>1993</td>
<td>Two psychoanalytic inpatient groups, 93 sessions</td>
<td>Early engagement scores were not related to group outcome.</td>
</tr>
<tr>
<td>Tschuschke, MacKenzie, Haaser, and Janke</td>
<td>1996</td>
<td>Used data from MacKenzie and Tschuschke (1993)</td>
<td>At 18-month follow-up, most successful members were involved with early interpersonal work.</td>
</tr>
<tr>
<td><strong>Outpatient therapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phipps and Zastowney</td>
<td>1988</td>
<td>Nine community mental health groups, 56 patients</td>
<td>Successful groups showed a pattern consistent with more advanced developmental stages. In unsuccessful groups, members’ lack of improvement was associated with groups in which engagement was not higher than avoidance or conflict.</td>
</tr>
<tr>
<td>Tschuschke and MacKenzie</td>
<td>1989</td>
<td>Two weekly psychoanalytic groups (Germany)</td>
<td>Positive outcome was associated with pattern of painful disclosures followed by hostility. Poor outcome was associated with being unable to move through differentiation phase.</td>
</tr>
<tr>
<td>Sexton</td>
<td>1993</td>
<td>One group, 12 sessions of verbal therapy and exercise (Norway)</td>
<td>Early positive ratings by therapist were associated with later symptom improvement in patients.</td>
</tr>
<tr>
<td><strong>Process and training groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacKenzie, Dies, Coché, Rutan, and Stone</td>
<td>1987</td>
<td>53 groups (4 sessions, 4 hours each), 555 group therapists</td>
<td>Successful groups had decreased conflict over time and shifted to deeper psychological exploration. An increase occurred in realistic, here-and-now, and interpersonal dimensions of goals with increasing sessions.</td>
</tr>
<tr>
<td>Kivlighan and Jauquet</td>
<td>1990</td>
<td>Six personal growth groups, 26 sessions</td>
<td>Positive outcome was associated with the leader, early task orientation, and later relationship connection with the leader.</td>
</tr>
<tr>
<td>Kivlighan</td>
<td>1997</td>
<td>72 groups, 372 students, 28 sessions</td>
<td></td>
</tr>
</tbody>
</table>
Kivlighan and Lilly 1997 14 groups, 84 students, 14–26 sessions Positive outcome was associated with stage development as proposed.
Kivlighan and Tarrant 2001 233 adolescents, 8 sessions, 2 hours each Positive relationship was evident between participant outcome and stage development. Group climate mediated outcome.

<table>
<thead>
<tr>
<th>Organizational groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheelan, Murphy, Tsumura, and Kline 1998 44 work groups of financial teams Group productivity was positively related to group development as measured by member perceptions of internal dynamics.</td>
</tr>
<tr>
<td>Wheelan and Lisk 2000 263 undergraduate students Positive correlation was evident between grade point average and stage of group development.</td>
</tr>
<tr>
<td>Wheelan, Burchill, and Tilin 2003 17 intensive care units, 394 staff Perceived higher stage of group development was associated with lower mortality rate of patients.</td>
</tr>
<tr>
<td>Wheelan and Kesselring 2005 61 elementary school faculty groups Faculty perceived higher stage of group development was related to fourth-grade children's greater proficiency in reading, science, and citizenship on standardized tests.</td>
</tr>
</tbody>
</table>

ented inpatient groups that met for 93 sessions over a 6 1/2-month period. The investigators used the GCQ–S subscale Engagement to Measure Group Work, a questionnaire using semantic differential items to tap relatedness, and outcome measures including the Revised Symptom Checklist and target goal ratings. Consistent with developmental theory, the investigators hypothesized that early engagement scores would predict outcome as would overall levels of engagement. However, this hypothesis was not supported at termination, 12-month follow-up, or 18-month follow-up. The authors considered that perhaps group members (consistent with the directions) may have been describing the groups' engagement rather than their own. If a disparity exists between their own engagement and collective engagement, a relationship would not be expected between their perception of collective engagement and their personal outcome. The investigators did find that participants obtaining higher relatedness scores had better outcomes.

A follow-up study (Tschuschke, MacKenzie, Haaser, & Janke, 1996) on these same groups focused on members' actual behaviors. Self-reports and observer process ratings concerning five therapeutic factors (Emotional Relatedness to the Group, Self-Disclosure, Feedback, Interpersonal Learning-Output, and Family Reenactment) were collected. The process ratings were
based on videotapes of every second session of the 93 sessions. The outcome measures were the same as in the prior study. There was general concordance of outcome for the patient, the therapist, and an independent observer. The Systematic and Multiple Level Observation of Groups method (Bales & Cohen, 1979) was used to assess feedback subcategories of task versus emotional behavior, dominance versus submission, and positive versus negative behavior as well as self-disclosure. At 18-month follow-up, the most successful members had exhibited higher levels of interpersonal work in the early sessions (first and second quarters of the group) than the least successful members. Although the least successful members performed more interpersonal work in the latter portion of the group, it did not help them (or as Tschuschke et al. [1996] said, “late bloomers do not blossom,” p. 43). No pattern was observed between feedback subcategories and outcome. Two groups is a small sample size, and the authors acknowledged that the patterns may have been particular to groups held within a larger therapeutic milieu. Yet, their finding of the importance of early participation is consistent with a larger corpus of research from a variety of settings.

Outpatient Studies

Inpatient studies have difficulty isolating the impact of the group from that of other therapeutic treatments. In contrast, process–outcome relationships of outpatient groups may be somewhat easier to discern. In an early but particularly strong study, Phipps and Zastowny (1988) used outpatient groups meeting 6 to 24 months prior to the commencement of the study. Using a repeated measures design over a 6-month period that assessed patient outcome and level of functioning (Brief Symptom Inventory) and group leadership style (Group Leader Behavior Instrument) and used the GCQ, they found two patterns of group process. The first pattern had highest engagement and then lower conflict and avoidance. Avoidance was higher than conflict and both would periodically increase. These groups appeared to manage conflict with a relatively stable climate of engagement and the best outcomes. In the other pattern, engagement was not invariably higher than others, and there was not a large difference between engagement and conflict and avoidance scores. This latter group pattern appeared unable to master the postdifferentiation phase. This study is perhaps the best demonstration of the relationship between developmental stage progression and outcome.

Tschuschke and MacKenzie (1989) studied two weekly psychoanalytic outpatient groups in Germany using verbal content analysis scales. One group had a successful outcome; the other did not. The verbal content analysis identified group climate. The more successful group (outcomes judged by therapist ratings of individuals) spent prolonged periods first in painful self-disclosure followed by periods of greater hostility directed toward others. The pattern repeated itself and eventually resulted in a lasting state of reduced affect. The second group with poorer outcomes failed to show any consistent
pattern and appeared unable to move through the differentiation stage. Although the sample size was small and the outcome measures were weak, the findings are promising (MacKenzie, 1994b).

Sexton (1993) examined a group of Norwegian outpatients who participated in 30 minutes of verbal therapy followed by aerobic exercise, group exercises, and brief discussion. Both therapist and patients completed symptom scales at each session. Although no overall measure of group climate or process was used, changes in patient symptoms and therapist ratings allow for the inference of a change process. The author found that therapists’ ratings of the session importance for a patient and positive feelings for the patient in the early sessions were more relevant to outcome then the same ratings in later sessions. This suggests that with successful therapy, there is mutual positive regard and an increase in patient positive alliance and therapist positive feelings; the therapist records some improvement, and some sessions later, the patient reports insight into his or her problems. Likewise, the patient’s lack of perceived improvement appears to result in more negative feelings in the therapist toward patient a few sessions later. Of course, because these relationships are correlational, no direction of causation can be determined, only hypothesized.

The most recent study, conducted by Castonguay, Pincus, Agras, and Hines (1998), is especially important because it is unique in studying process–outcome relationships in the context of a structured cognitive–behavioral outpatient group. Symptomatic ratings were taken before and after treatment. Members were asked to complete a group version of the Therapy Session Report (Orlinsky & Howard, 1966). Castonguay et al. found that overall, engagement increased across the sessions, whereas negative affective climate increased, reaching a peak in the middle sessions, and then decreased. The authors noted that these patterns were consistent with progressive stage models such as MacKenzie’s (1994b). Process and outcome measures were significantly related. The researchers found that the client’s experience of positive affectivity in the beginning of treatment and the perception of a negative group climate in the middle of treatment were predictive of a favorable treatment response. Significant limitations of this study are its correlational nature, the significant attrition of members (10 out of the original 75), and the specificity of the population (binge-eating patients).

In summary, the inpatient and outpatient studies generally showed that group development is a good thing for group members. Future studies would provide a clearer picture of relationships between stages and outcome by taking a broader view of group development as was done in the Phipps and Zastowny (1988) study. Rather than looking merely at individual scores such as the Engagement subscale of the GCQ–S, investigators should consider the overall pattern of scores to determine whether groups that progress beyond the formative stages show outcomes superior to groups that remain within the formative stages. Also, like the Tschuschke et al. (1996) study, future
studies should include measures of process beyond the individual member's self-report. Therapy group studies are plagued by human subject concerns, balancing known effective treatments with questions of science and design.

**Process and Training Groups**

Investigations with process and training situations allow greater potential for more rigorous methodologies, although outcome may assume a different meaning when participants are in therapy versus training groups. Consequently, the studies done on the relationship between process and outcome have offered the clearest view of the relationships between these domains. Five studies using a group format have yielded robust evidence for the existence of a model of group development. Sample sizes were much larger than the psychotherapy groups studied. Although outcome was assessed differently for each study, all but one used the GCQ–S with multiple administrations. Statistical methods were superior, and results were comparable.

MacKenzie, Dies, Coché, Rutan, and Stone (1987) studied 53 psychodynamic and special interest groups (impressively, 555 participants) at the American Group Psychotherapy Association Institute in 1982 who completed the GCQ–S and the Global Outcome Form. A stepwise regression analysis revealed that engagement was highly predictive of outcome for each session, and in the last sessions, conflict was negatively correlated with outcome. Both successful and unsuccessful groups (determined by the Global Outcome Form) began with similar levels of conflict. In the successful groups conflict decreased, suggesting that they mastered intragroup tension and could move on, whereas the less successful ones were unable to move beyond a conflictual process. In the final sessions, the more successful groups shifted to a focus on deeper psychological exploration. All groups increased on the Engaged scale throughout the sessions, but the most successful groups were significantly higher on engagement for all four sessions than the least successful groups. One more finding of interest is that the psychodynamic groups and the special interest groups, which were presumably more structured, did not differ from one another. Perhaps the weakest part of the study was the outcome measure, which was a self-report measure. The authors concluded that the patterns on the GCQ–S supported a group development hypothesis and demonstrated that the greater the progress through the stages, the better the outcome.

Kivlighan and colleagues, in a series of investigations of students involved in personal growth groups, carried out a series of programmatic studies that related group process to outcome. They sought patterns of leadership in relation to outcome with 372 students in 72 groups (Kivlighan, 1997). Outcome was measured by identification of problem areas and the level of distress experienced with these participants before and after group. Members had better outcomes when leaders initially focused on reinforcing therapeut-
tic norms through task-oriented interventions and later emphasized their personal connection with members.

In another study, Kivlighan and Jauquet (1990) had group members complete the GCQ-S and a problem (agenda) card. Three judges rated agendas on realism, here-and-now focus, and emphasis on interpersonal issues. Sessions were grouped into early, middle, and late for comparison. Consistent with the MacKenzie et al. (1987) study, engagement increased over time, whereas avoidance decreased over time. Conflict increased and then decreased. A significant increase in realistic, here-and-now, and interpersonal dimensions of goals occurred over the sessions. In the early sessions, when goals were more realistic, group members engaged in more productive work. In the middle and late sessions, when focusing on interpersonal agendas in the here and now, members were engaged and less avoidant of work. Middle sessions are also a time to deal with conflict and power. The authors suggested that in the early phase of a group, leaders may want to encourage the development of realistic agendas and then shift members to setting agendas with significant interpersonal components and here-and-now dimensions as the sessions progress. Caution should be exercised when applying these findings to a clinical population, as these group members were nonpatient volunteers.

The Kivlighan and Lilly (1997) study is of interest because of the variation in the number of sessions completed (14 to 26). Groups meeting for differing numbers of sessions followed similar developmental patterns with no relationship between number of sessions and group outcome. As in the Kivlighan (1997) study, therapeutic gain is related to a high level of engagement throughout. A low–high–low pattern for conflict correlated with positive outcome. It seems that groups need to begin with little conflict to foster cohesion. Conflict later permits the challenge of defenses and the deepening of intimacy. However, conflict must be resolved by the end to have a good outcome. A quadratic relationship with avoidance was related to therapeutic outcome (high–low–high–low). Although the authors offered no explanation for this variation, it is a relationship that makes sense for progressive group development; avoidance is high at the commencement as group members are getting to know each other, then low when issues of authority are being addressed, high as the group attempts to focus on issues of intimacy, and low during work.

Kivlighan and Tarrant (2001) expanded their study of group development to the adolescent population, using structured group sessions, didactic instruction, and discussion. Leaders with little formal group training were given 16 hours of instruction emphasizing member-to-member interaction. In an additional twist, leaders were given a questionnaire that measured their intentions in the sessions. In addition to the Group Climate Questionnaire, they completed symptom inventories. Findings indicated that adolescents liked their leaders and benefited from the sessions. The relationship between
group climate and participant outcome was consistent with previous findings. Increases in active engagement across sessions were related to member benefit. Decreases in conflict with time were related to participants liking their therapist. The adolescents felt more positive about leaders who structured the discussions, solicited elaborations, and gave feedback. They felt more negatively about leaders' attempts to do individual therapy in the group. The relation between leaders' intentions and member outcome was mediated by group climate; efforts to do therapy were negatively related, and intentions to provide a safe environment were positively related to an increasingly active and engaged climate. Even in these semistructured groups, group climate is an important mediator of outcome.

We present a final study by Kivlighan, McGovern, and Corazzini (1984) not because it supports a process–outcome relationship but because it is one of the few examples of an actual experiment that attempts to manipulate (although unsuccessfully) group climate. A small number of students were given sheets on the expression of anger and intimacy either at an appropriate time (anger sheet during storming or intimacy sheet during norming) or an inappropriate time (anger sheet during norming and intimacy sheet during storming) in development. For intimacy, the results were as predicted, but for anger, the sheet helped no matter what the timing. The authors interpreted this result to indicate that expressing anger is generally not comfortable. Aside from the small number of participants, one problem with interpreting the results is that there was no measure of what the therapists might have naturally done in the group to neutralize the incongruous directions.

Groups in Work Settings

Using the GDQ, Wheelan and associates have established a relationship between productivity and stage of group development in many field settings. In one study, financial teams functioning at higher levels of group development were able to earn more income in less time and were rated higher in customer service than teams functioning at lower levels of group development (Wheelan, Murphy, Tsumura, & Kline, 1998). In a second study, undergraduate groups functioning at higher stages of development obtained higher grade point averages than did participants in cohort groups functioning at lower stages of development (Wheelan & Lisk, 2000). These findings are consistent with those from the therapy and process group studies.

With the aforementioned studies, the outcome measure is of the individual in the group. When the outcome affects an individual who is one step removed from group, the power of the group process can be seen as even more impressive. Using the GDQ, Wheelan, Burchill, and Tilin (2003) found that employees of intensive care units (primarily White female nurses, working an average of 12 years in the unit) who perceived themselves to be functioning at higher stages of group development had lower than predicted mor-
tality rates than did other intensive care unit staff who perceived themselves to be functioning at lower levels. Similarly, the students of faculty whose peer groups were functioning at higher levels of development scored higher on a standardized proficiency test in citizenship, reading, and science than the students of faculty groups functioning at lower levels of group development (Wheelan & Kesselring, 2005).

Summary

Strong evidence exists for the link between outcome or productivity and stage of development. Studies with strong experimental design, especially those on work or training groups and organizational task groups, show a substantial connection between group development and productivity. However, using a cross-sectional design does not permit an exploration of the transition of stages. It is possible that successful outcomes could occur in a group that achieves maturation with little time spent in the earlier stages. Process-outcome studies with psychotherapy groups generally have lacked the methodological rigor of studies on work groups. A major problem is the lack of a clear and consistently applied standard for what constitutes group development. These studies rely excessively on self-report measures and are frequently characterized by small sample sizes. Studies of process and training groups consistently support the relationship between stage development, allow for evaluation of individual and group outcome, but sometimes lack the statistical power (because of the number of groups) to explore relationships with moderate effects.  

CONCLUSION

Considerable evidence exists that many interpersonal domains change systematically over the life of the group. To date, the model that has been most carefully studied in relation to the psychotherapy group is the progressive stage model, which embraces a number of stages (ranging in number from three to eight). Evidence from group climate, therapeutic factor, and leadership studies supports a stage model. Furthermore, evidence suggests that without this progression, interpersonal and intrapersonal changes are attenuated for its members. Rivaling the progressive stage model is a punctuated equilibrium model, which embraces a two-stage process in which the initial period is characterized by little accomplishment followed by a transition about midway through the duration of the group after which unfolds new productive interactional exchanges. Currently, insufficient empirical

---

*If moderate or small effects are predicted, a greater N is required than if the expected effects are large.
work has been conducted directly pitting one model against the other. It is possible that the particular model of development is context and task specific. It is also possible that the level at which observation takes place affects whether the change is viewed as punctuated or of a more continuous nature. Many studies have chosen somewhat arbitrary points at which to measure the phase, and little work has actually attempted to capture the shift in phases. Finally, much more work is needed to determine the extent to which the development of a psychotherapy group is associated with favorable outcomes.