The Development of an Instrument to Assess Cohesion in Sport Teams: The Group Environment Questionnaire

A.V. Carron
University of Western Ontario

W.N. Widmeyer and L.R. Brawley
University of Waterloo

The purpose of this paper was fourfold. The first purpose was to demonstrate the need to develop an instrument to assess group cohesion while the second was to outline a conceptual model of group cohesion upon which such an instrument could be based. This model reflected four related constructs which were the a priori basis for developing a large item pool and initial versions of the Group Environment Questionnaire (GEQ). The third purpose was to outline the four projects conducted to obtain construct-related information and to develop an initial version of the GEQ. The final purpose was to outline the two reliability and validity studies conducted with two different sport team samples. The results of these studies revealed that an 18-item version of the GEQ was internally consistent, reliable across studies, and content valid. Factor analyses with oblique rotation revealed preliminary evidence for construct validity. The GEQ is comprised of four scales reflecting the constructs of group integration-task, group integration-social, individual attractions to group-task, and individual attractions to group-social.

The importance of developing an instrument to assess group cohesion stems from the importance of groups. Shaw (1981) notes that "throughout history people have joined together in groups to accomplish a wide range of purposes" (p. 1). Momentary reflection seems sufficient to make individuals aware of the numerous and varied groups to which they belong or encounter in their daily lives. In 1951, James estimated that there were 4 or 5 million groups in existence at any given time. The prevalence and prepotency of

Each investigator contributed significantly to the research for this article and considers its authorship as joint. This research was supported by the Social Sciences and Humanities Research Council of Canada and by a Foundation Western Inc. grant to the authors. Appreciation is extended to Dr. P. Chelladurai for his contributions during the earlier phases of this project and to R. Summers for his assistance in conducting the studies and associated data analyses.

Requests for reprints and/or copies of the Group Environment Questionnaire should be sent either to A.V. Carron, Faculty of Physical Education, Univ. of Western Ontario, London, Ont., N6A 3K7, Canada or W.N. Widmeyer and L.R. Brawley, Dept. of Kinesiology, Univ. of Waterloo, Waterloo, Ont., N2L 3G1, Canada.
groups have led many social scientists to recognize the importance of studying them (e.g., Davis, 1969; Mills, 1967; Shaw, 1981; Steiner, 1972; Zander, 1979).

The study of groups falls under the rubric of group dynamics—a term initially advanced by Kurt Lewin (1935) to represent two principal processes occurring within groups. One of these, cohesion, is concerned with the development and maintenance of the group. The second, locomotion, is the activity by which the group seeks to achieve its objectives. Cattell (1948) pointed out that these two major group functions are stochastic (i.e., without group maintenance there can be no group locomotion). The essence of cohesiveness is inherent in its definition. For example, Carron suggests it is “a dynamic process which is reflected in the tendency for a group to stick together and remain united in the pursuit of its goals and objectives” (Carron, 1982, p. 124). Thus, by its very definition cohesiveness contributes directly to group maintenance and therefore indirectly to group locomotion. Considering the importance of these functions, it is not surprising that some social scientists (Golembiewski, 1962; Lott & Lott, 1965) have considered cohesion to be the most important small group variable. Consequently, studies dealing with the antecedents and consequences of cohesion are prevalent in the group dynamics literature of psychology, industry, and sport.

The extensive research on cohesion in a wide variety of settings has yielded very equivocal results. These findings have usually been attributed to problems associated with measuring cohesion. The apparent futility of measuring, weighting, and combining all the factors that attract members to a group led most early researchers to treat one factor, namely attraction to the other members, as synonymous with the total factors that attract members to a group. Thus, cohesion was usually assessed by some measure of interpersonal attraction such as reciprocal sociometric choices (e.g., Deep, Bass, & Vaughan, 1967), social activity of the group (e.g., Horsfall & Arensberg, 1949), group congeniality (e.g., Faunce & Beegle, 1948), cooperativeness versus competitiveness of groups (e.g., Deutsch, 1949), presence or absence of cliques (e.g., Lenk, 1969), similarity of interpersonal perceptions (e.g., Fiedler, Hartman, & Rudin, 1952), and members’ perception of group closeness (e.g., Martens & Peterson, 1971). Although the problems associated with this operational definition of cohesion have been discussed by various authors over the years, possibly the strongest criticism of this procedure was advanced by Escovar and Sim (1974). They pointed out that operational measures of cohesion based upon interpersonal attraction (a) underrepresented the concept—there are other factors at work within the group in addition to attraction to other members which keep individuals in a group; (b) fail to account for cohesiveness in situations characterized by negative affect (i.e., dissatisfaction, dissen- sion, hostility); (c) have not been supported empirically—interpersonal attractiveness has not been shown to correlate with other attractiveness measures in groups (Eismay, 1959; Gross & Martin, 1952); and (d) do not totally account for the conditions necessary for group formation.

In investigations in which interpersonal attraction has not been treated as synonymous with group cohesion, researchers have measured such different aspects of group attraction as members’ expressed desire to remain in the group (e.g., Schachter, Ellerton, McBride, & Gregory, 1951), members’ identification with the group (e.g., Converse & Campbell, 1968), and the value members place on group membership (e.g., Arnold & Straub, 1973). Some investigators have used members’ perceptions of group cohesion as the measure of group closeness. These measures include members’ perception of group functioning (e.g., Torrance, 1955) and teamwork (e.g., Martens & Peterson, 1971). Some researchers have assessed the perception of others (e.g., observers’ ratings of group in-
Figure 1 — Source of problems in group cohesion research.

teintegration during group performance; Stogdill, 1963). Still others have examined the similarity of members' interest in the task (e.g., Smith, 1968). Finally, attrition of group members has been used as an indication of low cohesivenes (e.g., Vander Velden, 1971). Most of these measures were devised by the researcher and rarely if ever subjected to psychometric analyses to establish their reliability and validity.

In summary, previous studies of cohesion have been characterized by one or more of the following three major measurement problems: They have utilized some measure of interpersonal attraction, a concept that underrepresents the attractions of groups for their members. They have assessed cohesion by such a wide variety of other measures that findings cannot be compared. They have rarely determined the psychometric soundness of the measures employed.

There is little doubt that the use of different and sometimes inappropriate measures of group cohesion has made it impossible to determine the reliability and/or validity of relationships between cohesion and its suspected antecedents/consequences. Rather than employing patchwork methods to repair existing measures or devising new ones with old problems, it is necessary to go to the root of the measurement issue (cf. Figure 1). Although several reviewers (e.g., Carron, 1980, 1982; Cartwright, 1968; Escovar & Sim, 1974; Evans & Jarvis, 1980; Gill, 1977; Zander, 1979) have suggested that measurement problems stem from the lack of a clear conceptualization of cohesion, rarely have researchers attempted to clarify this construct before measuring it.

Conceptual Framework for the Development of a Group Cohesion Inventory

Cohesion has been defined in a number of ways. One of the most frequently cited is the now classic Festinger, Schachter, and Back (1950) proposal that views cohesion
as the total field of forces causing members to remain in the group. One significant aspect of this definition is that it focuses on the individual and forces that attract individuals and cause them to remain with the group. Another significant aspect is their suggestion that cohesion is the result of numerous factors. Subsequently, Gross and Martin (1952) criticized this definition on the grounds that it fails to consider the group as a totality. They preferred to view cohesion as the resistance of the group to disruptive forces.

The need to distinguish between the group and the individual has been one of two major issues dominating the group dynamics literature (e.g., Cattell, 1948; Zander, 1971). For example, Cattell observed that a group can be studied at three different levels—population, structure, and syntality. At the population level the focus is on the individual group members and their needs, aspirations, and motives. At the structure level the focus is on the patterns of interaction within the group. And finally, syntality involves focusing on the group as a whole.

A second major distinction takes into account the task oriented and socially oriented concerns of groups and their members. The study of leadership provides perhaps the best illustration. Certainly every leadership theory, whether it focuses on the traits of leaders or their behaviors, incorporates an assessment of both the task and social dimensions (e.g., Fiedler, 1967; Hersey & Blanchard, 1969).

This first issue, the need to distinguish between the individual and the group, has also been evident in discussions on the nature of group cohesion. For example, as a result of their dissatisfaction with the popular research strategy of summing individual attractions to group (ATG) scores to produce a cohesion index, Van Bergen and Koekebakker (1959) advocated that the concepts of ATG and cohesion be differentiated. They defined ATG as the interaction of motives working on the individual to stay in the group and cohesion as the degree of unification of the group field. Thus, they identified both an individual concept (ATG) and a group concept (cohesion). Although the latter was not well defined operationally, it recognized cohesion as a group property distinct from ATG, a property reflecting the individual group member.

More recently, Evans and Jarvis (1980) reiterated the ideas of Van Bergen and Koekebakker (1959). They noted that the Van Bergen and Koekebakker definition of cohesion implies "a closeness among members, a similarity in perception of events, and...a bonding together in response to the outside world" (p. 366). Thus, according to Evans and Jarvis, measurement of this closeness/similarity/bonding might rely more on variability rather than a group mean. Evans and Jarvis also noted that, by contrast, ATG is a composite of individual members' feelings about the group in terms of their desire to be accepted and identified as a group member, their personal role involvement with the group, and their involvement with other group members. Thus it might be best assessed by some measure of central tendency.

The second issue discussed above—the need to distinguish between the task and social concerns of groups and their members—has also been considered in discussions on group cohesion. For example, Festinger et al. (1951) considered factors that caused members to remain in the group to be of two types: attractiveness to the group and means control. Enoch and McLemore (1967) considered ATG as intrinsic attraction and instrumental attraction. Anderson (1975) found that ATG (cohesion) became greater in task-oriented groups when they were successful and greater in socially oriented groups when members had similar values. Perhaps Mikalachi (1969) made one of the clearest distinctions when he advocated that cohesiveness be subdivided into task and social components. According to Mikalachi, task cohesion exists when the group coheres around the task it was organized to perform while social cohesion exists when the group coheres around social (nontask)
functions. Mikalchki also suggested that these two components of cohesiveness should be considered separately in terms of their antecedents and consequences.

**Cohesion: A Conceptual Model**

These two important distinctions—the individual versus the group and task versus social concerns—had a major impact on the development of the conceptual model. The model is divided into two major categories: a member's perceptions of the group as a totality and a member's personal attractions to the group. The former category is labeled *group integration*, and the latter, *individual attractions to the group*. Both perceptions help to bind members to their group. Also, the member's perceptions of the group as a unit and their perceptions of the group's attractions for them can be focused on task or social aspects. Thus, four constructs can be identified—*group integration–task, group integration–social, individual attractions to group–task,* and *individual attractions to group–social*. Operationally, members could indicate the extent to which they agree that each of the constructs plays a role in the development and maintenance of group cohesion.

Conceptually, group integration is the category that represents the closeness, similarity, and bonding within the group as a whole—the degree of unification of the *group* field. Individual attractions to group, on the other hand, is the category that represents the interaction of the motives working on the individual to remain in the group—the composite of the *individual* members' feelings about the group, their personal role involvement, and involvement with other group members. Further, the social aspect can be seen as a general orientation toward developing and maintaining social relationships within the group. The task aspect can be seen as a general orientation toward achieving the group's goals and objectives. It is assumed that the four constructs of the conceptualization are correlated. They are related through the perceived interaction of various task and social orientations as viewed through the eyes of the individuals for themselves and their group. The conceptual model which formed the basis for the Group Environment Questionnaire is presented schematically in Figure 2.

![Figure 2 - Conceptual model of group cohesion.](image-url)
This conceptualization and the questionnaire which ultimately evolved from it must be viewed within the perspective of its delimitations. One of these is that a questionnaire approach is utilized rather than direct behavioral assessments. Another is that individual perceptions are assessed rather than a total group viewpoint. Also, although other aspects of cohesiveness have been discussed in the literature (e.g., normative cohesion), only the four selected constructs are included, that is, the ones thought to account for the greatest variance among groups. And finally, only intragroup perceptions of cohesiveness are tapped; extragroup perceptions are not considered.

These delimitations notwithstanding, the conceptualization and the questionnaire which evolved from it do offer some advantages. One of these is that the antecedents and consequences of cohesiveness are not incorporated into the measurement of cohesion. A second is that cohesion is treated as a multifaceted concept—more than one factor binding members to the group is considered. And finally, the individual perceptions of the group’s task and social cohesiveness are considered through two major perspectives: personal ("what am I doing?") and group ("what is the group doing?").

**Operationally Defining the Constructs:**
**Phase 1—Identification of Constructs and Representation of Content**

Although four related constructs had been suggested as aspects of cohesion in the initial conceptual framework, the possibility that group members might perceive cohesiveness as involving other concepts was recognized. Also, in the construction of instruments to assess any construct, the actual representation of the construct (i.e., the semantics and the descriptors used) might be more clearly expressed by the actual subjects than by the investigators. In past research, cohesion instruments have too often been based upon the investigators’ rather than the group’s ideas. Thus, the intention in the present investigation was to involve group members as active agents in expressing the meaning of the construct of cohesion (cf. Adair, 1973; Russell, 1982; Sherif & Sherif, 1969). To this end then, group members’ perceptions of cohesion formed part of the basis for identifying cohesion concepts and the items reflecting the expression of these concepts. In this first phase, four projects were conducted to obtain input, three of which used group members as contributors.

**Project 1: Methodology**

Each member of an undergraduate class in the social psychology of physical activity interviewed three members from each of two sport teams. The interview questions concerned (a) the personal meaning of cohesion to group members, (b) the behavioral manifestation they could cite to reflect cohesion, (c) the incidents that group members recalled which would denote a low level or absence of cohesion, and (d) factors that contributed to the development of cohesion on the respondents’ team. A total of 234 respondents from a variety of different sport teams \((N = 78)\) were subjects. University, municipal, and industrial teams from hockey, soccer, baseball, lacrosse, football, track, and swimming were examples of the sports in which athletes were interviewed. More than 1 team in each sport was examined and athletes of both sexes were interviewed. The most consistently appearing responses were tallied, taking care to use the subject’s own wording, then coded according to whether responses concerned one of the four constructs (i.e., individual attractions to group–task, individual attractions to group–social, group integration–task, and group integration–social) or concerned another construct altogether.
Project 2: Methodology

The subjects were 63 members of an undergraduate kinesiology class, all of whom had previously competed in team sports. Each subject responded to a series of open-ended questions about why people join groups, leave groups, or stay with groups. A third of the respondents provided self-focused reasons (i.e., "Why did you..."), another third provided reasons for their teammates, (i.e., "Why did your teammates...") and a final third gave responses for athletes in general ("Why do individuals..."). In this manner, subjects used their own words and their own response rate to describe both personal (i.e., self and group focus) and general factors contributing to individual attractions to group and group integration. The different-person focus provided a range of perceived reasons so that several alternative perceptions, and thus greater potential response variability, could be obtained. Once again, the most frequently appearing responses were tallied and coded as in the first project.

Project 3: Methodology

The subjects were 60 members of intact teams actively involved in ongoing competition. These groups differed in the amount of interaction required to perform their sport. All teams examined were at the university level and involved in the activities of competitive swimming, cross-country running, and cheerleading. The same open-ended questions used in Project 2 were administered. The three sets of questions were randomly distributed so that some members from each group responded to each form of question. Coding proceeded as with Projects 1 and 2.

Project 4: Methodology

A literature search of 29 different articles and studies on the topic of cohesion provided a complement of questions (i.e., operational definitions) that had been previously used to examine the phenomenon. The search covered group dynamics articles published between 1948 to 1982 in the areas of social psychology, sociology, industrial psychology, and sport psychology.1 These questions were scrutinized for similarity, applicability to sport teams, and frequency of appearance. Coding of the applicable and most frequently appearing questions proceeded as with other projects.

Summary of Phase 1 Projects

All of the responses from the four projects just described were collapsed to form a response pool representing information concerning the four constructs. The responses represented the subjects' own wording of the affective and behavioral manifestations of proposed cohesion-related constructs as well as previously developed items purported to measure the phenomenon. No other responses emerged from the series of projects often enough to require adding constructs to the proposed model.

The resultant pool of affective and behavioral manifestations from subjects and from the group dynamics literature formed a major portion of information used for item development in Phase 2. It is important to note that Phase 1 represents one of the few

---

1The literature searched in Project 4 was only to 1982 because this was when the initial four projects were completed and instrument development began.
instances in the development of a cohesion instrument where subjects have been actively involved in providing information for item construction. The use of subjects’ as well as investigators’ information encourages greater information variability, lessens the possibility of exclusive investigator bias, and encourages item development using language and concepts that subjects can understand (cf. Adair, 1973; Russell, 1982; Sherif & Sherif, 1969).

**Instrument Formation:**

**Phase 2—Item Development and Content Validation**

The purpose of this phase was twofold. The first purpose was to develop an item pool that would form the basis for the Group Environment Questionnaire (GEQ). The information gleaned from the four projects in Phase 1 were used to write the items. When this large item pool was reduced to a manageable size, the second purpose of establishing content (face) validity was undertaken.

**Methodology: Item Development Project**

Four investigators and a senior research assistant used the Phase 1 information and their knowledge of group dynamics to generate an initial large item pool of 354 statements (cf. Guilford, 1954; Nunnally, 1978). Each individual independently wrote items representing each of the four constructs (individual attractions to group–task, individual attractions to group–social, group integration–task, group integration–social). The pool of items concerning each construct was then scrutinized by the five item writers and statements were grouped according to similarity of content. Each of the similar content areas was termed a subarea for a given construct. Several criteria were then used to agree upon the list of items to be included on the first version of the questionnaire. They were (a) frequency of appearance, (b) clarity in writing, (c) amount of ambiguity, (d) use of group members’ as opposed to investigator’s terminology, and (e) duplication (cf. Guilford, 1954). An agreement of 80% among the investigators was required to retain an item.

In this item-trimming process, care was taken to ensure that no one subarea (e.g., friendship, role clarity, role acceptance) was overrepresented for a given conceptual construct. A bias in representation could have led to erroneous results in subsequent item analyses. As well, care was taken to ensure that each construct’s scale had approximately equal numbers of positively and negatively worded items (cf. Nunnally, 1978). Finally, statements were written to ensure an approximately equal representation of affective and behavioral manifestations of individual attractions and group integration.

**Methodology: Content Validity Project**

Obviously, the item development, refinement, and selection conducted by the five item writers represented a form of content validation. As an additional step toward content validity, the reduced item pool was also sent to five experts in the area of group dynamics from the fields of social psychology, industrial psychology, and sport psychology. Each expert was sent a summary of the conceptual model with instructions to comment on the model and delete or make suggestions about the items representing each construct (cf. Bohrnstedt, 1970). The responses from the five experts were examined and a decision was then made whether to modify, delete, or add items. Experts’ responses confirmed the deletion choices of the research team.
Phase 2: Results and Discussion

The GEQ: Version 1

The original 354-item pool contained the following distribution of items representing each construct: individual attractions to group–task, 100 items; individual attractions to group–social, 80 items; group integration–task, 97 items; and group integration–social, 77 items. The process of item deletion by the five researchers reduced this original item pool by 85%. The reduced pool of items \((n = 53)\) was then put in questionnaire form. In a manner consistent with the recommendations of Guilford (1954), a Likert-type scale was adopted which varied from strongly disagree \((value = 1)\) to strongly agree \((value = 9)\). In the 53-item questionnaire, the distribution of responses in each of the four construct scales was as follows: individual attractions to group–task, 13 items; individual attractions to group–social, 13 items; group integration–task, 16 items; and group integration–social, 11 items.

The content validity assessments by the five experts were used in conjunction with the preliminary item analysis (Phase 3 of the research) and therefore are discussed with those results. A degree of content validity was assured through the protocol followed by the research team which required 80% agreement across a range of item characteristics.

Social Desirability

Social desirability (SD) is an issue considered by many who develop instruments because of a concern that socially desirable responses can contaminate the validity of the construct of interest. Investigators and authoritative sources often assume that SD instruments or the inclusion of SD scales within the main instrument is a psychometrically sound methodology (e.g., APA, 1974, Standards for Educational and Psychological Tests). However, the majority of test developers who accept this view fail to differentiate whether SD is a property of the items/scales or an individual difference variable.

As ingenious as many of the SD devices may be, they are commonly self-reports which fail to differentiate the two aforementioned SD types. McCrae and Costa (1983) have noted that such instruments may lead researchers to identify the honest, cooperative, and conscientious subject as a “liar.” Eliminating such subjects may cause the loss of very valuable data. An example in the case of group cohesion might be the highly involved, honest group member who responds to group integration items based upon an invariant perception about his/her group’s “highly desirable” behavior. Elimination of his/her responses would not be eliminating an artifact, but rather, substantive data. On the basis of their own and others’ data, McCrae and Costa (1983) have noted that instrument scores “corrected” by use of values from popular lie and SD scales decreased, not increased, the main instrument’s validity coefficients. This result is opposite to that desired through the use of SD scales.

Based on this rationale, the decision not to use an SD scale in the present questionnaire was anything but arbitrary. As McCrae and Costa have stated, other investigators using known—or worse, their own hastily constructed—lie scales as a methodological benefit might well be doing themselves a disservice.

Reliability of the GEQ: Phase 3—Item Analyses

The purpose of this phase in the test development was to assess the reliability of the questionnaire with respect to a measure of equivalence (i.e., internal consistency) through various item analytic procedures. Two studies were undertaken to accomplish this purpose.
Internal Consistency Issues

Stability and Equivalence. In terms of the reliability of an instrument, consistency can be considered in two categories called measures of stability and of equivalence. Measures of stability or test-retest reliability are time-honored procedures often cited in the literature as one of the indices of an instrument's worth. Unfortunately, most attention is often paid to these estimates even though they have greater limitations than do measures of equivalence (i.e., the internal consistency of items within a scale). Although the test-retest approach is thought to reflect stability, the degree of stability can vary due to between-test time interval, reactivity, or increased probability of the individual's response change over time. Bohrnstedt (1970) has suggested that these problems have turned researchers toward a greater reliance on measures of equivalence.

The previous concerns and suggestions were taken into account in the statistical analyses of the GEQ. The notion of equivalence and its assessment focuses on the internal consistency of a measure. For example, in the GEQ, when items representing a particular scale are summed to yield a score, it is assumed that each item is a relatively unique (i.e., not identical) measure of that scale. Thus, items in the scale are internally consistent with the construct measured by that scale.

In attempting to achieve high internal consistency, both statistical procedures and practical factors were considered. They were (a) increasing reliability by adding/deleting items from a scale; (b) ensuring that each item represented only a single construct; and (c) ensuring that the scale was of a practically administrable length.

Intra- and Interscale Equivalence. The internal consistency approach to equivalence examines the covariance among all items of a scale simultaneously. The statistical indicant generally used to express this reliability is a generalization of the Kuder-Richardson 20 formula (1937) called Cronbach's alpha (1951). The latter statistic is used when items are scored on a continuous measurement scale. Coefficient alpha is calculated for the scale when each item's variance is statistically removed from the total scale. Thus, both an appreciation for an item's relation to a construct and a statistical criterion are gained and used in making a scale reliable and practically shorter.

However, internal consistency within a scale is not sufficient for an instrument having more than one scale. Items should also be correlated with their own scale to a greater extent than with any other scale in the instrument. Items correlating otherwise reduce reliability because of their overlap with the other assessed constructs. Eliminating such items ensures a greater probability of the constructs being relatively unique.

Item Analysis. The processes for determining which items to retain in the various scales of an instrument are called item analysis techniques. The two item analysis techniques described in the previous section were used in reducing the 53-item questionnaire (cf. Phase 2—Results).

The process of item analysis requires not only a pool of subjects large enough to examine item reliability but also a sample representing a broad range of subjects (cf. Nunnally, 1978). An inventory attempting detection of both different aspects and intensity levels of a multifaceted concept requires statements that elicit responses across any population's normal distribution. However, to determine if such statements have the sensitivity to reflect sample differences, one must examine them through the responses of a heterogeneous sample. Improper sampling can cloud the sensitivity of the instrument to detect differences. Unfortunately, heterogeneous sampling is the exception rather than the rule in the sport cohesion literature.

In the case of the present investigation, the unit of analysis was the individual
within the group. Inasmuch as groups can be conceived as varying in their stages of development, it seemed wise to sample groups across this continuum.

**Study 1 — Method: Item Analysis**

**Subjects**

In this study, athletes \(N = 212\) from a number of teams \(N = 20\) of heterogeneous characteristics were tested. The sample consisted of forming through to well-developed teams (i.e., group tenure ranged from 1 month to 9 years). The teams were interactive, requiring members to integrate their roles in order to accomplish the group task. A variety of sport types was used (i.e., ball hockey, baseball, basketball, ice hockey, rugby, soccer). Teams were from intercollegiate and adult municipal associations. All participants were adult Canadians representing both male \((n = 13)\) and female \((n = 7)\) teams. Athletes averaged 12 years of playing experience with their sport. The percentage of new members on a given team ranged from 10 to 50\% \((M = 20\%)\). At the time of assessment, the mean time members had been together that season was 5.8 months. In addition, the teams had participated in an average of 26 competitions during that season. In summary, these teams represented a heterogeneous sample.

**Procedure**

The GEQ—Version 1 was administered by individual research assistants, who informed teams that they were part of a larger investigation concerning how teams interact. Team members were assured of anonymity (i.e., that no player or coach would see their responses).

Each respondent completed the 53-item GEQ (cf. Phase 2—Results). The GEQ also contained a variety of demographic questions concerning group and member characteristics. These questions not only helped to describe the group but also served to check the heterogeneity of sample characteristics.

Research assistants were given a standard protocol for questionnaire administration. Teams were always assessed before a weekday practice during their regular season. Assessments immediately before or after competitions were avoided so that responses would not be competition-specific.

**Results and Discussion**

With respect to internal consistency, the two procedures mentioned above provided the statistical bases for item elimination. The first criterion examined was whether an item written to represent one of the four constructs correlated well with its own scale’s total score (i.e., intrascale equivalence). Thus, for example, if an item written to measure individual attractions to group–task correlated poorly with the total score for that scale, its internal consistency was questionable and it was considered for elimination from the scale.

The second criterion was whether an item was more related to its own scale than to other scales (i.e., interscale equivalence). For instance, an individual attractions to group–task item that correlated \(r = .50\) with its own scale but also correlated \(r = .55\) with the
scale of individual attractions to group-social was not considered to be internally consistent. This example reflects a case of construct overlap and the item would be considered for elimination from the questionnaire.

While these statistical criteria determined retention or deletion of items, it was necessary to seek a maximum Cronbach’s alpha for each of the four scales and simultaneously consider the overall length of the questionnaire. As items were deleted from each scale, the item analysis procedure was reiterated, Cronbach’s alpha was recalculated on the basis of the deletion, and item correlations with other scale total scores were recomputed. Four iterations of this process were conducted, reducing the questionnaire from 53 items to 24 while still retaining good internal consistency. It should be noted that any items questioned by the external reviewers in the content validity phase were among those deleted by the item analysis process.

The scales in the reduced version of this GEQ were represented by the following numbers of items: individual attractions to group–task, 7 items; individual attractions to group–social, 5 items; group integration–task, 7 items; group integration–social, 5 items. The respective values for Cronbach’s alpha for each of the aforementioned scales were .74, .58, .78, and .61. These results are presented in Table 1, which also illustrates a matrix of interscale correlations.

The interscale correlations indicate that the scales are moderately related. However, since these relations do not exceed .80, there is no cause for concern about multicollinearity. In fact, they reflect relations suggested by the conceptual model. That is, the scales are assessing constructs that are related but are sufficiently unique not to be considered redundant.

### Table 1

**Internal Consistency of the GEQ: Study 1**

<table>
<thead>
<tr>
<th>Scales</th>
<th>ATGT</th>
<th>ATGS</th>
<th>GIT</th>
<th>GIs</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual attractions to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group–task (ATGT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG–social (ATGS)</td>
<td>.43</td>
<td></td>
<td></td>
<td></td>
<td>.58</td>
</tr>
<tr>
<td>Group integration–task</td>
<td></td>
<td>.54</td>
<td>.32</td>
<td></td>
<td>.78</td>
</tr>
<tr>
<td>(GIT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIs–social (GIS)</td>
<td>.35</td>
<td>.32</td>
<td>.39</td>
<td></td>
<td>.61</td>
</tr>
</tbody>
</table>

Reliability values are based upon the 24 items selected from the 53-item GEQ (N = 212). Interscale correlations are calculated by correlating a scale total score (sum of all items in a given scale) with its counterpart for each of the other scales (e.g., the ATGT scale score correlates $r = .54$ with the scale score for GIT).
Although the internal consistency values suggested good reliability in a questionnaire of practical length, it was felt that greater confidence could be placed in these findings if they could be replicated with a different sample. Also, the reliability coefficients just presented for the 24 items reflect relations among items and scales physically embedded in the original 53-item format. Thus, it was concluded that if the scales were reliable, a similar pattern of relations would be evident in another study when the 24-item questionnaire was used.

**Study 2 — Method: Item Analysis**

The basic purpose of Study 2, then, was to determine the reliability of Version 2 of the GEQ (n items = 24). A second purpose was to determine if the internal consistency values found in Study 1 could be replicated with a different athlete sample.

**Subjects**

The athletes (N = 247) examined were from 26 different teams, heterogeneous in sport type (i.e., basketball, cross-country running, cross-country skiing, curling, figure skating, gymnastics, ice hockey, precision skating, ringette, speed skating, swimming, wrestling). Of these teams, 14 were female and 12 were male. Teams were both interactive, n = 10 (i.e., requiring members of different roles to interact to accomplish the group task), and individual, n = 16 (i.e., members act individually and their performances sum for a common team outcome) in nature. Teams also varied in their tenure together and in the number of competitions in which they had participated. They were sampled from various intercollegiate and municipal adult leagues in the province of Ontario.

**Procedure**

Test administration procedures identical to those in Study 1 were followed using the 24-item GEQ. Again, care was taken to ensure that the questionnaire was not administered immediately before or after a competition in order to avoid situation-specific responses.

**Results and Discussion: Studies 1 and 2**

A comparison of the reliability values between studies is illustrated in Tables 2 and 3. The values for Cronbach’s alpha are quite similar between studies as the scale-by-scale results in Table 2 indicate. Cronbach’s alpha for individual attractions to group–task and social and group integration–task and social are $r = .65, .64, .71, \text{and} .72$, respectively.

The interscale correlations of Studies 1 and 2 are compared in the matrix presented in Table 3. Correlations from Study 2 can be compared with those of Study 1, the latter in parentheses immediately below the former. The interscale correlations in Study 2 were slightly lower than those from Study 1. They still reflect the relatedness of the assessed constructs but their size indicates that the measures are relatively unique.
Table 2

Internal Consistency Comparison: Studies 1 and 2

<table>
<thead>
<tr>
<th>Scales</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study 1</td>
</tr>
<tr>
<td>Individual attractions to group - task</td>
<td></td>
</tr>
<tr>
<td>(ATGT)</td>
<td>.74</td>
</tr>
<tr>
<td>ATG - social</td>
<td></td>
</tr>
<tr>
<td>(ATGS)</td>
<td>.58</td>
</tr>
<tr>
<td>Group integration - task</td>
<td></td>
</tr>
<tr>
<td>(GIT)</td>
<td>.78</td>
</tr>
<tr>
<td>GI - social</td>
<td></td>
</tr>
<tr>
<td>(GIS)</td>
<td>.61</td>
</tr>
</tbody>
</table>

Reliability values in Study 2 are based upon the responses to the 24-item GEQ while those from Study 1 are based upon responses to the same items within the 53-item GEQ (Study 1, N = 212 vs. Study 2, N = 247).

Table 3

Scale Intercorrelation Comparison: Studies 1 and 2

<table>
<thead>
<tr>
<th>Scales</th>
<th>ATGT</th>
<th>ATGS</th>
<th>GIT</th>
<th>GIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual attractions to group - task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ATGT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG - social</td>
<td>.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ATGS)</td>
<td>(.43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group integration - task</td>
<td>.41</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GIT)</td>
<td>(.54)</td>
<td>(.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI - social</td>
<td>.30</td>
<td>.40</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>(GIS)</td>
<td>(.35)</td>
<td>(.32)</td>
<td>(.39)</td>
<td></td>
</tr>
</tbody>
</table>

Parenthetical values are those of Study 1.
Thus, good internal consistency for the GEQ—Version 2 (hereafter referred to as GEQ) was evident both within and between scales. This consistency was stable across two different samples of sport groups, heterogeneous in their characteristics. It was concluded that the GEQ was reliable from an equivalence perspective and that this equivalence was replicable (cf. Bohrnstedt, 1970). It is important to realize, however, that the reliabilities of any scale are only estimates dependent upon sampling and estimation technique. A greater degree of confidence may be accorded the GEQ than its predecessors because its estimates are reliable within and between studies.

While it is important to demonstrate the reliability of a measuring instrument, it is equally important to examine its validity. This was the purpose of the fourth phase in this instrument’s development.

**Validity of the GEQ: Phase 4—Factor Analysis**

**Content Validity**

The GEQ had been shown to have good face or content validity (i.e., the degree to which the scale represents the constructs of interest) as mentioned in the Phase 2 discussion. Briefly summarized, the following procedures initially undertaken also contribute to this content validity: (a) broad based literature search, (b) subjects used as active agents in concept definitions, (c) reliance on a conceptual model which provided the rationale for the various scales, (d) assessments of item content made by unbiased experts and the research team, (e) intercorrelations of each item with own and other scale total scores. All of these steps represent logical and statistical procedures that help to ensure content validity (cf. Bohrnstedt, 1970; Kerlinger, 1973).

**Construct Validity**

Two other types of validity, criterion-related and construct, also require estimation in developing the GEQ. The former mainly concerns the predictive ability of an instrument while the latter concerns the meaning of the instrument—the extent the concepts upon which it is based account for subject responses. A variety of procedures and investigations are required to determine the degree of confidence an investigator has in the “validities” of an instrument (cf. Bohrnstedt, 1970; Campbell & Fiske, 1959; Kerlinger, 1973).

One group of procedures that can be used as a preliminary indicator of construct validity is factor analysis. These data-reduction procedures examine the intercorrelations among items with the notion of collapsing them into a smaller set of variables or constructs. Validity is present when the factors reflect the constructs in the proposed conceptual model. A practical benefit of this procedure in the early stages of test development is that the sample used to achieve the instrument’s reduced form (e.g., the 24-item GEQ) can be factor-analyzed. Thus, the investigator can obtain a preliminary indicator as to whether the items on the instrument do represent the proposed conceptual structure. This allows for decision making about further steps toward achieving validity. A factor structure not representing the conceptual model may mean that the instrument, the model, or both need further work. By contrast, preliminary verification of the model may indicate it is worthwhile spending the considerable effort required to conduct predictive studies. The factor analysis in this case was conducted using the 212 team sport athletes sampled in Study 1. The data analyzed were the responses to the 24 items selected for Version 2 of the GEQ.
It should be noted that the factor analytic approach taken in the present investigation represents a theory-driven as opposed to data-driven approach to instrument development. The latter approach proceeds without a conceptual model but with the expectation that a model will emerge. Nunnally (1978) suggests that this avenue of test construction can be viewed as a "shotgun" approach. As emphasized earlier, this approach has been the root of the problems associated with empirical cohesion research. Thus, the opposite approach was taken where preliminary verification of an a priori proposed model was attempted.

**Factor Analysis of the GEQ: Method**

A variety of factor analytic procedures can be used to determine the structure of a set of variables. The procedure selected is partly dependent upon what the investigator is trying to accomplish. With theory-driven test development, there is an a priori notion of what the underlying factors in the data might be. As well, the theoretical model may also suggest a degree of relatedness among the proposed variables.

**Oblique Rotation**

In the present investigation, the factor analysis chosen was principal factoring with oblique rotation. The goal of rotation is to achieve the most theoretically meaningful, and if possible the simplest, factor structure (Harris, 1975). The advantage of oblique rotation is that the factor axes need not be orthogonal to each other. The absence of an orthogonality assumption in a factoring approach is probably more realistic when the theoretical constructs under examination are considered related. Gorsuch (1974) noted that if prior evidence suggests these relations are present, oblique rotation is a realistic approach. As previously stated, the interscale correlations in both Studies 1 and 2 (cf. Table 3) illustrated moderate relations between scales. Given this evidence, the choice of oblique rotation seemed logical and appropriate. It also seemed logical to constrain the solution to four factors as proposed in the conceptual model. This is consistent with the choice of procedures which was governed by an a priori set of criteria for determining simple structure (cf. Harris, 1975).

**Interpretation Procedures**

Although a factor analytic procedure is chosen logically, factor structure may be interpreted in a variety of ways. When the factors are correlated, the contribution of variables (i.e., items in the present case) to the interpretation of factors differs according to the factor matrix examined. The two matrices of chief concern in the oblique solution are the pattern (i.e., pattern coefficients define the grouping of variables) and the structure (i.e., correlations of the variables with the factors). Gorsuch (1974) recommended initial interpretation on the basis of the factor structure matrix. The advantages of this interpretation are that (a) the relation of the variable (item) to the full factor is indicated, and (b) correlation coefficients have more interpretive meaning to the majority of investigators in the sense that their range and the practical importance of that range is known. The relationship of the items to the factors should remain the same regardless of the context of another study and other factors in that study. By contrast, the pattern matrix coefficients can only
be interpreted within the context of a given study. Thus, the structure matrix from the oblique rotation was the primary choice for interpretation in the present study.

The salient loadings within a factor help with its interpretation. Once again, however, the investigator is faced with various approaches in assessing what constitutes a sufficiently high loading. There is no single rule to aid in deciding what is a sufficient loading (Gorsuch, 1974). The general approach is to conduct the analysis with a sample of sufficient size so that any variable of interpretive interest is significant. In using this approach, Gorsuch suggested doubling the minimum significant correlation coefficient in the structure matrix to ensure that the loading is high enough to have interpretive meaning. For example, the minimum significant correlation with an $n$ of 100 is $r = .20$. Thus, the doubling guide would suggest that a variable with a minimum loading of .40 is interpretable. However, with a large enough sample, even some doubled correlations would be questionable from an interpretive standpoint. Consequently, a loading of .30 is the criterion generally accepted as a minimum for samples of $n > 175$.

Of course, the size of a structure matrix coefficient is only clearly interpretable if the variables load primarily on one factor. The clearer the strength of the relationship of an item with a single factor, the easier the interpretation. This criterion encourages interpretation in accord with Thurstone's (1947) notion of simple structure.

It should be noted that the item analytic procedures previously discussed are also a form of implicit factor analysis. Thus, elimination of test items by those procedures is not unlike dropping items that do not meet minimum loading criteria in factor analysis. However, the reason for using item as opposed to factor analysis for item deletion is that item analysis has been shown to be a somewhat superior procedure for reliability and validity estimates (Larsen & LeRoux, 1983). It was expected that the majority of the 24 items of the GEQ would meet the loading criterion suggested. Whether the items would relate to the factors for which they were written remained for the factor analysis to determine.

**Results and Discussion**

The structure matrix of the four-factor, obliquely rotated model is illustrated in Table 4. The four factors each had an eigenvalue greater than 1.0. The majority of items met the minimum loading criterion of .30. No interpretable factor had less than three items. This solution indicated that in this sample, the GI scales (group integration–task, group integration–social) and the ATG scales (individual attractions to group–task, and individual attractions to group–social) were ranked most to least important, respectively (i.e., as assessed by amount of variance accounted for by factor eigenvalues exceeding 1.0).

What has to be remembered when interpreting a correlated factor model is the purpose of the procedure and what the various correlation loadings represent. Correlations among the factors are based upon the correlation between an item and a factor as well as the factor's relation to the item through other correlated factors. Thus, estimates of an item's contribution to the factor are not exclusive to the factor. It is therefore questionable how the variance accounted for should be interpreted. In this preliminary analysis, what is perhaps more important is the grouping of items within factors. Recall that the primary intent for utilizing factor analysis was to obtain a preliminary indicant of construct validity. Thus, interpretation in terms of the conceptual model was of chief interest.

In considering the structure matrix, four groupings of variables were evident that can be named according to the factors proposed in the conceptual model. For this sample,
Table 4

Structure Matrix: Oblique Rotation Factor Analysis of the GEQ

<table>
<thead>
<tr>
<th>Scale</th>
<th>Items</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Group integration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– task</td>
<td></td>
<td>.53</td>
</tr>
<tr>
<td>(GIT)</td>
<td>2</td>
<td>.61</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.62</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.66</td>
</tr>
<tr>
<td>GI – social (GIS)</td>
<td>1</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.47</td>
</tr>
<tr>
<td>Individual attractions to</td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>group – social (ATGS)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.48</td>
</tr>
<tr>
<td>ATG – task (ATGT)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

The values in the matrix are correlations of scale items with the factors. Only items loading highest on their own factor are reported for brevity. Exceptions within a factor are noted.

aItems lower than the minimum .30 criterion. bItems that overlapped with other factors.

some of the items in two of the scales (ATG scales) either did not meet the minimum loading criteria or overlapped too much with other factors. Thus, they were not useful in interpretation. However, the majority (n = 19) of the 24 items from the GEQ were satisfactory.

It is also noteworthy that the correlation matrix among factors shown in Table 5 reflected a degree of relatedness similar to that of the interscale correlations between scales (cf. Table 1). Much higher or much lower correlations among factors would have been one indication that the rotation used was inappropriate for interpretation.

In summary, results of the exploratory factor analysis revealed a factor structure that would be representative of the constructs proposed in the conceptual model. However, these results also suggested that a third version of the GEQ might be more internally con-
Table 5
Factor Intercorrelation Matrix

<table>
<thead>
<tr>
<th>Factor</th>
<th>GIT</th>
<th>GIS</th>
<th>ATGS</th>
<th>ATGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group integration task (GIT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI - social (GIS)</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual attractions to group - social (ATGS)</td>
<td>.28</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG - task (ATGT)</td>
<td>.37</td>
<td>.27</td>
<td>.28</td>
<td></td>
</tr>
</tbody>
</table>

sistent than the second as some of the 24 items did not meet the minimum factor loading criterion. Items from the two individual attractions to group scales were viewed as possible candidates for elimination (cf. Table 5). Therefore, this same data set \((n = 247)\) was further item-analyzed.

The GEQ: Version 3

Version 2 of the GEQ was examined for one criterion of internal consistency, (i.e., the relation of each item with its own scale; cf. results of Studies 1 and 2). However, the second criterion for inclusion of items in any given scale had not been applied to the 24-item GEQ. Briefly reiterated, each item must be assessed for its relation to its own and other scale total scores. Items relating as well or greater with scales other than their own are not considered internally consistent to the parent scale. Removal of such items from their scale changes the item-to-scale correlations observed in the first iteration of item analysis.

Subsequent iterations are usually conducted. Obviously, when an already refined instrument has been reduced in size, the elimination of various items after two or three iterations must be done cautiously. As well, when items are removed in later steps of the process, an item eliminated at an earlier step may subsequently become more internally consistent and thus "reclaimed" in the scale. The reason the reclaimed item may now correlate best with its own scale is that undesirable items, which were later removed, had originally suppressed the item's contribution to its own scale. The reclaimed item's contribution to internal consistency could only be evaluated given the final list of items retained for all scales.

To summarize, several iterations of item analysis, item elimination, and reclamation occurred. The overall objective was improvement of internal consistency. The final result was Version 3 of the GEQ, an 18-item questionnaire. The number of items retained for each scale was as follows: individual attractions to group–task, 4 items; individual
Table 6

GEQ Internal Consistency Comparison: 24 Items Versus 18 Items

<table>
<thead>
<tr>
<th>Scales</th>
<th>Cronbach's Alpha</th>
<th>ATGT</th>
<th>Interscale correlations</th>
<th>ATGS</th>
<th>GIT</th>
<th>GIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual attractions to group – task (ATGT)</td>
<td>.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(ATGS)</td>
<td>(.65)</td>
<td>(.28)</td>
<td>(.29)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Group integration – task (GIT)</td>
<td>.70</td>
<td>.41</td>
<td>.30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(GIS)</td>
<td>(.71)</td>
<td>(.41)</td>
<td>(.30)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GI – social (GIS)</td>
<td>.76</td>
<td>.30</td>
<td>.40</td>
<td>.34</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(.72)</td>
<td>(.30)</td>
<td>(.40)</td>
<td>(.34)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Reliability values from the 18-item GEQ are compared to the 24-item GEQ. Values for the latter version of the inventory are in parentheses. The sample is N = 247.

attractions to group–social, 5 items; group integration–task, 5 items; group integration–social, 4 items. As can be seen in Table 6, the net effect of this process was to increase the internal consistency of the individual attractions to group–task (ATGT) and group integration–social (GIS) scales with no substantive change in individual attractions to group–social (ATGS) or group integration–task (GIT) values. For ATGT, ATGS, GIT, and GIS, the Cronbach’s alphas were $r = .75, .64, .70$, and .76, respectively. Interscale correlations were similar to those of the 24-item GEQ (cf. Table 6). Thus, Version 3 of the GEQ continued to have good internal consistency after elimination of problem items observed in the previous 24-item version.

General Discussion

The present research project evolved from the view that there is a need within group dynamics in sport to develop a psychometrically sound instrument to assess group cohesion. Based upon this need, it was also felt that the route to such an instrument required a conceptual model of group cohesion. From this, an instrument could evolve in a systematic fashion.

The perceptions of active group members and the group dynamics literature provided the information base for item development. Through a variety of item analytic procedures, an 18-item, four-scale instrument emerged. This instrument, the Group Environment Questionnaire, is (a) practical, (b) reflects good internal consistency and is stable across two independent samples, (c) assesses a wide variety of sport groups having
heterogeneous characteristics, (d) is content-valid, and (e) has preliminary psychometric support for its construct validity.

The suggestion is made that the construct validity is preliminary because evidence for that form of validity must come from a variety of investigations and methods (cf. Cook & Campbell, 1979). One factor analytic study alone does not suggest sufficient validation. The factor structure must be replicated across different samples having different characteristics. Gorsuch (1974) provided an even stronger statement regarding the general use of factor analysis in determining representations of constructs:

It must be stressed that interpretations of factors are post hoc unless a hypothesis testing procedure is used. In any post hoc situation, no interpretation is regarded as final but only as a lead for further research....The widely followed practice regarding interpretation of a factor as confirmed solely because it "makes sense" is to be deplored. (p. 188)

Further validation research on the GEQ is necessary and is currently under way.

At this stage of instrument development, the GEQ is both similar to and different from other cohesion instruments reported in the sport group literature. The similarities lie in the attempt to measure group members' perceptions of cohesion by self-report. The differences can be grouped into conceptual and psychometric categories. From the conceptual perspective, this instrument is based upon a model concerning group members' perceptions of cohesion (admittedly, other situational factors also contribute to cohesion). By contrast, preceeding instruments neither specify a model nor use it for development.

From a psychometric viewpoint, the form of the GEQ developed to date represents a more concerted attempt to (a) operationalize constructs by independent methods, (b) separate previously confounded perceptions, (c) be generalizable to a large cross-section of sports, (d) have a form of reliability consistent across samples, and (e) satisfy more than one form of validity.

As yet, the GEQ is still in its infancy as part of a continuing program of research. Its validity and future worth as an assessment device require further analyses by ourselves and other investigators.

References


Manuscript submitted: November 5, 1984
Revision received: March 3, 1985