INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Understanding the effects of practice, process and ability on abstract problem-solving performance: A study of group intelligence

Lasek, Rebecca, Ph.D.

Case Western Reserve University, 1994
UNDERSTANDING THE EFFECTS OF PRACTICE, PROCESS AND ABILITY ON ABSTRACT PROBLEM-SOLVING PERFORMANCE:
A STUDY OF GROUP INTELLIGENCE

by
Rebecca Lasek

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Dissertation Advisor: Dr. Douglas K. Detterman

Department of Psychology
CASE WESTERN RESERVE UNIVERSITY
January, 1994
We hereby approve the thesis of

[Signature]

candidate for the Ph.D.
degree.*

(signed)  (chair)

[Signature]

[Signature]

[Signature]

date 9/28/93

*We also certify that written approval has been obtained for any proprietary material contained therein.
I grant to Case Western Reserve University the right to use this work, irrespective of any copyright, for the University's own purposes without cost to the University or to its students, agents and employees. I further agree that the University may reproduce and provide single copies of the work, in any format other than in or from microforms, to the public for the cost of reproduction.

[Signature]
UNDERSTANDING THE EFFECTS OF PRACTICE, PROCESS AND ABILITY ON ABSTRACT PROBLEM-SOLVING PERFORMANCE

Abstract

by

REBECCA LASEK

Two experiments were designed to investigate the effects of individual and group practice, group process and cognitive ability on abstract problem-solving performance. Experiment I compared the scores of three member groups with their best, average and worst members' individual scores. The findings showed that (a) ad hoc groups outperformed their best members, (b) a unique group ability and not the best member's individual ability determined the group product and (c) there was evidence of a relationship between group process and group performance. Experiment II focused on the effects of the amount and complexity of group practice on future individual performance. The results showed that (a) two group practice sessions improved future performance, (b) two individual practice sessions hindered future performance and (c) one group practice session did not affect future scores, regardless of the level of test item complexity during the group session. Two factors that may be responsible for the beneficial effect of group practice are (a) learning during group sessions and (b) the interpersonal relationships among group members.
Specific Aims

The present research included the following specific aims:

1) Assess group intelligence using a computerized multiple-step task that measures cognitive ability.

2) Compare the performances of ad hoc groups with the performances of their best members and determine whether the groups can outperform their best members on the task.

3) Determine, using multiple regression and the comparison of practiced groups to unpracticed groups, whether group performance is a function of the group’s ability or the best member’s overriding individual ability.

4) Measure group process with experimenter ratings and measure group interaction with Bale’s (1979) System for the Multiple Level Observation of Groups (SYMLOG). Compare the results of the group process/interaction measures with group performance scores to uncover any relationship between group ability and group process/interaction.

5) Demonstrate whether (a) group practice is better than individual practice, (b) more group practice is better than less group practice and (c) the complexity of the test items solved during group practice affects future individual performance.
Acknowledgements

I would like to gratefully acknowledge the guidance, support and many thoughtful contributions of Dr. Doug Detterman, my advisor, to my work over the years. I would like to thank Dr. Joseph Fagan for his constructive criticism of my writing-style which proved to be very helpful. I am grateful to Dr. Lee Thompson for her thoughtful suggestions regarding the theoretical applications of this research. I would especially like to acknowledge the support and contributions of Dr. Bob Greene who gave much thought to the prospectus for this research project, made many useful suggestions and raised many important issues that helped make this paper a better one.

A special thank you to Chris Schatschner and Dennis Mitchell for talking with me about my research and for taking time to help me out when I ran into problems. I am very grateful to Kate Spry for her kindness and help over the last few years.

I would like to thank Dr. Michael Dwyer at Baldwin-Wallace College, Dr. Steven Slane at Cleveland State University and Dr. Janet Larson at John Carroll University for the use of their subject-pools. I am ever grateful to the 300 students who diligently participated as subjects in these multiple-session experiments.

Much thanks to Jon, Alicia and Melissa.
Table of Contents

Abstract ........................................ii
Specific Aims .....................................iii
Acknowledgements ...................................iv
Table of Contents ................................v
List of Figures .....................................vii
List of Tables .....................................xiii
Introduction ........................................1
  Practice ..........................................3
  Process ..........................................9
  Ability ..........................................12
Preliminary Study ................................19
Method: Experiment I ............................21
  Subjects .........................................21
  Materials and Apparatus .......................22
  Design ..........................................28
  Procedure ......................................29
Results ...........................................32
Discussion .........................................44
Method: Experiment II ..........................52
  Subjects .........................................52
  Materials and Apparatus .......................52
  Design ..........................................53
  Procedure ......................................55
Results ...........................................58
Discussion .........................................63

v
List of Figures

Figure 1: Sample Test Item -----------------------------------24

Figure 2: Graphic Depiction of Design: Experiment I ----28

Figure 3: SYMLOG Profiles: Best and Worst Performing Clusters -----------------------------------------------37

Figure 4: Graphic Depiction of Design: Experiment II --54
List of Tables

Table 1: Mean PM Scores and Standard Deviations Based on the Total Percent Correct: Experiment I --------------32

Table 2: Mean PM Scores, Standard Deviations Based on the Total Percent Correct: GI and IG Conditions --------34

Table 3: Mean Progressive Matrices Scores for Groups and Best Members and Total Number of Groups Represented in each Category: Group Process -------------------------------35

Table 4: Mean PM Scores and Standard Deviations Based on the Total Percent Correct: Practiced Groups versus Unpracticed Groups and Practiced Best Members --------39

Table 5: Mean Progressive Matrices Scores Based on the Total Percent Correct: GI and IG Conditions by Subject-type -----------------------------------------------40

Table 6: Mean PM Scores and Standard Deviations Based on the Total Percent Correct: Best, Average and Worst Individuals Not Participating in groups and the Best, Average and Worst Members in Real Groups --------------41

Table 7: Mean PM Scores and Standard Deviations Based on the Total Percent Correct: Experiment II --------------59
Table 8: Mean PM Scores and Standard Deviations Based on the Total Percent Correct For Males and Females in All Conditions: Experiment II

Table 9: Mean Change in Scores from Pretest to Posttest for Males and Females: Experiment II
Understanding the Effects of Practice, Process and Ability on Abstract Problem-Solving Performance

Effective group problem-solving may be the cornerstone of our cultural development. Archaeological (Leakey & Lewin, 1978) and anthropological (Tiger & Fox, 1971) accounts of hunter and gatherer lifestyles have stressed how important organized reciprocity within small structured groups was for everyday problem-solving and survival. The value of group problem-solving for future societal advancement has been questioned in cultural, educational and scientific discussions over the years. The current circumstances surrounding Outcome-Based Education (OBE) provide a good example.

OBE is an innovative teaching philosophy which emphasizes small group learning as a central part of its curriculum. Following attacks by some traditionalists, OBE has received local (Kingsley, 1993) and national (ABC Nightly News, 1993) media attention. Despite the emerging controversy, research has provided encouraging support for the approach. The small groups (also referred to as student support systems) used in OBE have been shown to be effective alternatives to low teacher-student ratios. The groups have been found to (a) increase participation, interaction and cognitive and
affective outcomes, (b) encourage self-paced active involvement in learning, (c) allow for individualized practice and (d) promote fast error correction (Soled & Bosma, 1992). In a six year study, OBE was found to boost basic-skills scores above the national norms (Abrams, 1985).

Studies of other educational approaches have also supported the contributions of small-group work to achievement. In the field of engineering, for instance, an educational technique called Structured Controversy (Smith, 1984), which emphasizes cooperative learning in groups, has been associated with increases in engineering students' expertise (Smith, 1988). Another study done in the field of psychology suggested that numerous group problem-solving sessions over the course of an academic semester greatly improved students' skills as measured by a cognitive abilities test (Smith, 1989).

The theory that small group work is an effective educational technique has also been at the core of studies emphasizing cultural issues in education. Even though North American society continues to emphasize individualism, researchers have been considering the contributions of group-centered ideologies and cultures to educational outcomes. One study of Asian-American refugees has shown that families emphasizing in-home,
highly structured group study sessions produce students who excel in school well beyond their North American counterparts (Caplan, Choy & Whitmore, 1992).

The applications of group research continue to evolve while many of the questions concerning the advantages and disadvantages of working in groups remain unanswered: Can ad hoc groups perform better than their best members or do groups just exploit their best members? How important is group process in determining group performance on problem-solving tasks? Does working in groups improve cognitive skills and problem-solving abilities? If the answer is yes, how much time must one spend in a problem-solving group to benefit from the experience? And is there any truth to the old adage that a team (group) is only as good as its weakest member? Attempts to answer these questions were made in the following study of the effects of practice, process and ability on abstract problem-solving performance.

**Practice (Experiment I):** Several studies have demonstrated that *individual practice* does not have any beneficial effect on subsequent group performance (Laughlin & Sweeny, 1975; Tuckman & Lorge, 1962). One study even found that individual practice inhibits group performance (Schoner, Rose & Hoyt, 1974). According to Campbell (1968), group inferiority following individual
practice is a function of (a) group member commitment to former individual decisions and (b) group member reluctance to change personal views or response strategies. In other words, group member stubbornness may result in faulty group process.

The possibility that individual practice hinders future group performance by impeding process was tested in Experiment I. The key interest was whether individual practice promotes the same kind of negative response pattern noted by Campbell (1968) when the problems have multiple steps and measure cognitive ability. Because multiple-step problems provide an opportunity for group members to build on each other's ideas, having individual practice was expected to be an asset.

The influence of individual practice on future individual performance was also explored. Because individual practice has been found to stagnate or disadvantage groups (Laughlin & Sweeny, 1975; Schoner, Rose, & Hoyt, 1974; Tuckman, & Lorge, 1962), and groups generally perform better than individuals overall (Williams & Sternberg, 1988), it was expected that individual performance following individual practice would be more likely to stay the same than improve.

Research on individual to individual practice effects on some other cognitively based performance tasks
have shown the opposite effect. Research on the WAIS-R (Matarazzo & Herman, 1984), for instance, has shown an improvement in individual scores over time on tasks measuring performance IQ. Factors responsible for the practice effect on the WAIS-R (e.g., motor manipulation and use of the same test items), however, are not factors involved in the solving of progressive matrices problems. In addition, test-retest reliability on the Raven's Progressive Matrices tests is high (r=.70-90) (Anastasi, 1988). The reliability of the tests corroborates the idea that the occurrence of individual to individual practice effects on a progressive matrices task is not likely.

Unlike individual practice, group practice was expected to have a beneficial influence on future individual scores. According to Dillon, Graham and Aidells' (1972) study of brainstorming and Schoner et als. (1974) study of the solving of an economic problem, working in a group promotes homogeneity of individuals' ideas. The finding that individuals become more similar to one another in relation to the task has important implications for group practice on a cognitive abilities test, especially in regard to learning. Individuals may be applying the strategies that they learned during the group session to other similar, but not identical,
problems at a later time. If group members are learning from group participation then individual performance should improve following group practice. Experiment II embellished upon the idea that learning may occur during group practice. A hypothesis central to Experiment II was that learning would be affected by (a) the amount of group practice and (b) the difficulty of group practice.

**Practice (Experiment II):** Smith (1989), in a study he conducted in a university classroom, found that subjects may have learned problem-solving strategies from group experiences. He had students work on various projects in groups or individually over the course of a semester. Smith demonstrated that students who had a semester of group experience with problem-solving tasks scored higher on a subsequent individually administered Raven's Progressive Matrices test than students who had not had any in-class group experiences. The finding implied that experience with group problem-solving cultivated the skills that are necessary to be successful on a progressive matrices test.

Although the implication of Smith's (1989) finding had merit, there were a number of problems with his methodology. Smith collected his data over 10 years from 18 classes increasing the possibility of generational/cohort effects. He also acknowledged that
the conditions varied from class to class. For example, the size of the groups was reduced after 2 years possibly biasing the outcome. Most discrediting, however, was that assignment to condition was not random. Students chose to work in groups or individually. Students who chose to work in groups had the option to (a) move to another group at any point during the semester or (b) drop out of the group and work alone.

Smith (1989) argued that self-selection to condition did not affect his results because (a) major and academic GPA's were the same across conditions and (b) there were no differences across conditions on two individually administered classtime exams. Unfortunately, Smith did not consider psycho-social factors that may have influenced the selection process. Extraversion, introversion, temperament and other personality factors may have dramatically affected subjects' choices to work alone or in groups.

Smith (1989) biased his study by having the flexible self-selection to condition. The bias, however, may have significantly impacted the quality of the group processes and for that reason Smith's study makes an important contribution to the literature. Because Smith used what he considered to be multi-stage problems (e.g., students were assigned to critique and improve upon a theory using
specific rules), he believed that group information sharing and the logical coordination of that information would increase. In fact, self-selected groups in combination with multiple group sessions and multi-stage written problems may have greatly improved group processes by impacting relationships. In Smith’s samples, improved group processes may have led to better interpersonal relationships, more individual learning about the task (multi-step problem-solving) and ultimately improved individual problem-solving ability (as evidenced by high Raven’s test scores at the end of the term).

A primary hypothesis of Experiment II was that two sessions of group practice would benefit individuals more than one session of group practice because of advantages gained from increased contact. In other words, group process would be better in groups whose members had practiced together before than it would be for groups solving for the first time. The theory is supported by differences between studies of people who have a history of working together and studies of ad hoc groups. A group with a history of working together is more likely to outscore its best group member (see Michaelson, Watson & Black, 1989).

Other researchers have suggested there is a
relationship between the time spent working in groups and group performance. According to Shaw and Ashton (1976), groups working for extended time-periods on difficult tasks may have more stimulating group processes which could produce 'assembly bonus effects.' Assembly bonus effects (Collins & Guetzkow, 1964) refer to the positive contributions of group processes to group members' experiences, and ultimately group performance, above what might be expected from the best member. The hypothesis that group member learning may be one assembly bonus effect was tested in Experiment II.

Also tested in Experiment II was the inferred relationship between task complexity and success on the task (Shaw & Ashton, 1976). The idea that there is a relationship between task complexity and group performance was supported by a preliminary study done by the present author. Groups in that study outperformed their best members when subjects (a) had individual practice on the task (which contradicts other research) and (b) the practice problems were complex. A hypothesis of Experiment II was that individual practice followed by one group practice session on a complex task would benefit future individual performance.

Process (Experiment I): Steiner (1972) recognized the importance of process as a factor affecting group
outputs. He theorized that group performance equals the potential of the group members minus losses resulting from poor group process. Steiner focused his study on disjunctive tasks which require a choice between two or more alternatives. The progressive matrices task used in the present study, although a multiple-step task, fits the definition of a disjunctive task as well in that, ultimately, a choice must be made among six alternatives. Steiner speculated that group performance would never exceed the performance of the best member on a disjunctive task, except by chance.

Shaw and Ashton (1976) elaborated on Steiner’s (1972) theory when they hypothesized that group performance equals the potential of the best member minus losses resulting from poor group process plus assembly bonus effects. Shaw and Ashton’s hypothesis was partially supported by their study, but their data was not conclusive. They believed that task difficulty helped produce positive processes likely to be associated with assembly bonus effects. Whether different processes actually occurred among the groups in Shaw and Ashton’s sample remains unknown. Process was not measured.

Shaw and Ashton’s (1976) idea that the beneficial effects of group process probably occur during the solving of difficult tasks was addressed in Experiment I.
Group process was measured both directly and indirectly and the progressive matrices tests were designed to be difficult. The following hypothesis was made: If groups outperform their best members on the highly complex task, then the processes occurring in the successful groups will be different from the processes occurring in the unsuccessful groups. Confirming the hypothesis would support Shaw and Ashton's theory and provide support for the idea that group synergy furthers group ability beyond the potential of the best group member.

Group process was examined in three ways. The effectiveness of process was examined indirectly by studying the group's ability to perform better than the best group member. If the group's performance exceeds what the best member could have accomplished alone, then one could conclude that something occurred within the group (synergy) that enhanced the overall group ability. Because 'process' is the basis for what transpires among group members (making groups potentially different from individuals), process can be considered directly responsible for group output when group performance is superior to the best member.

Group interaction was studied directly to uncover what elements of process might be occurring in exceptional groups. Bale's (1979) System for the
Multiple Level Observation of Groups (SYMLOG) was used to measure interaction processes. Because individual differences among group members create a vast uniqueness among groups, a variety of processes were expected. In addition to looking at the overall differences among groups, some of the successful and unsuccessful groups were examined more closely. The experimenter also used naturalistic observation to directly assess group members' behaviors and group characteristics. The ratings and notes were expected to provide insight into the processes occurring in exceptional groups.

**Cognitive Ability (Experiment I):** According to Williams and Sternberg (1988), group intelligence is the functional ability of group members working together as a unit. Williams and Sternberg believed that group aptitude could not be measured using traditional means. An assumption of the present study was that group intelligence can be measured by using the same criteria used to measure individual intelligence. A cognitive task that correlates with IQ ($r=.70-.90$) was used in the present study to measure group problem-solving.

Although few studies have compared groups with their best members (see Hill, 1982; Lorge, Fox, Davitz, & Brenner, 1958, for reviews), studies that have made the comparison have found that groups rarely perform better
than the groups' best members would if they were working alone (Williams & Sternberg, 1988). Furthermore, only one study using ad hoc groups (group members without a history of a working relationship) and multiple-step problems has shown that groups can outperform their best members (Faust, 1959). Research with ad hoc groups is important because a greater level of control over the interpersonal relationships of the group members can be established; the level and stage of the relationships are more similar across groups, enhancing reliability and validity.

In his study, Faust (1959) compared groups with their best members on an anagram task. Groups outperformed their best members in one sample, but not in another. Faust believed that the multiple-step task was responsible for the groups' superiority over the best members in the one sample. However, the same task was used in both samples. Because the results were not consistent, Faust's conclusion could only be regarded as preliminary. Experiment I replicates Faust's study using ad hoc groups and multi-step problems in an attempt to find whether groups can outperform their best members. It was hypothesized that the cognitive ability of a successful group would be a reflection of the ability of that group as an unique entity and not the best member.
Group's Ability versus Best Member's Ability (Experiment I):
All three factors under investigation, (a) practice, (b) process and (c) ability, are important for understanding whether the group's ability or the best member's ability is responsible for the group product. When one refers to 'group ability,' one refers to what group members can accomplish by working together in a unified fashion as a unique entity. In other words, group ability is synonymous with group intelligence (see Williams & Sternberg, 1988).

Theoretically there is a high probability that groups will outperform their most competent members on multiple-step problems. One can determine the probability of obtaining the correct answer to a multi-step problem from a group by using the logic underlying the Guttman Scale (see Guttman, 1947). Using the Guttman Scale the probability associated with solving each step of the problem can be used to estimate the probability of obtaining a correct answer from a group with N members. Even when nobody in the group is able to solve the problem alone, the group members may be able to reach the correct solution by working together, if at least one person can solve each step (Lorge & Solomon's Model B, 1955). The present study tested subjects on problems with only a few stages, but mostly on problems that were
more complex. If the group scores exceed what the best members could have accomplished alone, then group ability and not the best member’s ability would be responsible for the group success.

Tuckman and Lorge (1962), in a study of groups that had a history of working together, theorized that if groups solving problems for the first time performed the same as groups whose members had individual practice, then individual ability (the ability of the best member) would be the determining factor. The rationale behind the theory is: If individual ability determines group performance, then practiced groups will ignore the opportunity to share and build upon any knowledge about the task gained from working alone and perform as if there are no advantages to having previous experience with the task. If, on the other hand, group ability is responsible for group performance, practiced groups will use the group members’ additional knowledge about the task and perform better than unpracticed groups.

Tuckman and Lorge’s (1962) data supported an individual ability model of group performance. The results were muddled, however, in that success on the task was estimated by a Quality Point Score (Lorge, Fox, Davitz, & Brenner, 1955) which gives points for the number of significant elements included in the solution.
Although a quality solution had multiple elements, there were not necessarily multiple steps. Thus, the task was not a true multiple-step task and was only capable of measuring the best member’s ability.

Furthermore, Tuckman and Lorge (1962) did not control for process because they did not consider that faulty group process might influence their findings. Group process may have been poor, however, as a result of the task that was used. The task was the Mined Road Problem (Office of Strategic Services, 1948) which requires groups to formulate strategies to get five men across a land mine-infested road without neutralizing or removing the mines. The tools available for the men to use are ropes, beams and other objects. The demands of the task may have encouraged possessiveness of ideas in the practiced groups whose members had already formulated their own best strategies individually. Campbell (1968) theorized that group members often adhere to their formerly constructed individual beliefs after they have solved a problem previously by themselves. The closed mindedness of individuals in the resolving groups may have resulted in poor group process. The present study challenges Tuckman and Lorge’s results using (a) the same experimental design (which is embedded in the larger design of Experiment I), (b) a true multiple-step task
and (c) a measure of group process.

In addition to investigating Tuckman and Lorge's (1962) theory, the relationship of the groups' scores to the best, average and worst members' individual scores was analyzed to provide more information about how the groups operate. If, following group practice, a linear relationship exists between the groups' scores and the best members' scores, then group performance would be considered a reflection of best member ability and not group ability. If, on the other hand, no linear relationship exists between the groups' scores and the best members' scores, then group ability would be the factor most likely to be responsible for the groups' performances. The data from Experiment I were analyzed to uncover any linear relationships between group and individual scores. The findings were expected to be supportive of a group ability model of group performance.

Lastly, research on learning from group participation has established that group members who have the least to contribute to the group experience gain the most from the process (see Hill, 1982 for a review). A group ability model of group intelligence would be supported if the best members were not only contributing to the quality of other group member's experiences, but benefitting from the experience themselves. The
following hypothesis was derived to further test the group and best member models of group intelligence: If multiple-step problems stimulate the group process, affording the group an opportunity to break through the 'ceiling' imposed by the best member's ability, then the individual performances of all group members will show improvement over the scores of their unpracticed counterparts, regardless of ability.
Preliminary Study

A preliminary study was done during 1991 on the effects of practice, process and cognitive ability on future group and individual performance (N=120). There were two testing conditions. Subjects in each condition were tested twice. Subjects were tested either individually and then in a group (condition 1), or in a group and then individually (condition 2). Sessions were held 3 weeks apart. A multi-step task was used to measure the cognitive ability of groups and individual group members. The test was chosen because subjects must correctly solve a series of steps in order to solve the problem and because the test correlates with IQ tests (r=.70-.90). According to Lorge and Solomon’s (1955) Model B theory, a group may be able to outperform the group’s best member on a multi-step task if all members contribute and process information adequately. There were two parallel test forms (form A and form B) that were crossed within each condition. No subject was supposed to receive the same test form twice. There was a flaw in the test construction that caused some of the same items to appear on both test forms. One of the test forms was longer than the other. The flaw, however, did not bias all of the results.

Group scores on the progressive matrices test were
compared with the best and worst individual scores of the same subjects. Performance was determined by the percent correct. The results indicated that (a) group ability is more important than individual ability in determining group intelligence (practiced groups significantly outscored unpracticed groups by 7%), (b) group intelligence can exceed the best member's intelligence as measured by the same test, but this effect is limited to instances in which groups are working on complex problems and members have practiced on a long version of the test, (c) group practice affects subsequent individual performances differently depending on individual ability; the worst members benefit the most, scoring significantly higher (10% higher) than their unpracticed counterparts and (d) individual practice affects subsequent group performance. The results prompted further study. Experiment I of the proposed project replicates the study just described with the addition of two control conditions (individual-individual and group-group). The test items were designed to be more complex and the tests were made to be equal in length.
Method

Experiment I

Subjects

Subjects were 192 students enrolled in introductory psychology classes at four Cleveland area universities. Participants included 96 students from Case Western Reserve University, 60 students from Baldwin-Wallace College, 21 students from Cleveland State University and 15 students from John Carroll University. Using subjects from different universities increased the diversity of the sample. Specifically, a greater range of cognitive and social-cognitive characteristics were sampled.

There were 103 females and 89 males. Fifty-four percent and 46%, respectively. The unequal samples of females and males resulted from a largely female subject pool at Baldwin-Wallace College. Over the two-semester period of data collection, 14 students failed to return for the second session. The absentees caused the loss of four groups (12 subjects) and the loss of 9 individuals from a control group. Four new groups and 9 more individuals were tested in place of the lost subjects. The lost subjects are not reflected in the above numbers. All subjects consented to participate and were informed of their right to terminate their participation at any time.
Forty-eight subjects were assigned to each of the four conditions. Subjects signed up for the experiment in groups of three. Instructions on the sign-up sheet informed subjects that they were required to attend two sessions in order to receive credit. The instructions clearly stated that failure to attend both sessions would jeopardize others’ participation. Reminder calls were made the night before each session to encourage promptness.

**Materials and Apparatus**

*Computerized Cognitive Abilities Test (CAT):* CAT is a battery of 11 independent cognitive tests. One test in the battery is modelled after Raven’s Progressive Matrices (PM) standard intelligence tests and correlates with IQ, \( r = .70-.90 \). A more difficult PM test was developed by the experimenter by transferring the algorithms of the Raven’s advanced test items to new computerized stimuli. Eighty items selected from both the original PM test and the newly developed stimuli were split into two parallel 40 item tests. Internal consistency reliabilities were calculated for the two test forms. Alpha coefficients (Cronbach, 1970) were .80 and .84 based on the individual test scores from the first session (total \( N=96 \)). The items in each test were ranked in order of difficulty, each item was
progressively more difficult than the one before. Problem solving ability on the CAT was determined by the number correct converted into a percentage score. Testing time lasted an average of 45 minutes.

**Testing:** When the computer was turned on test instructions appeared on the screen. The test taker indicated to the computer that he/she had finished reading the instructions by touching the screen. A set of 2 x 2, 3 x 3, or 4 x 4 matrices appeared in the form of an incomplete square or rectangle. The bottom right hand box was empty in each set. A row of six boxes were presented at the bottom of the screen. All boxes on the screen were filled with a pattern of black and white squares. The test taker had to decide which of the alternative boxes at the bottom of the screen would logically complete the larger configuration above (Figure 1). The test taker selected an answer by touching the screen and received audio feedback about whether the answer was correct. A correct response caused the computer to emit a high pitched tone, while an incorrect response caused the computer to emit a low sounding tone. The correct response was always shown in the empty space in the matrix after a choice had been made. For a full description of CAT see Detterman (1990).

**Computers:** Tests were administered on three 286 IBM
Figure 1 Sample Test Item

The test item could be solved using several different strategies. One possibility is: 1) Combine blocks 7 and 8 which would lead to solution F. Check the strategy by combining blocks 1 and 2 or blocks 4 and 5 and you will see that the combinations do not add up to blocks 3 and 6. F can be eliminated as a possible solution. 2) Add blocks 3 and 6 which would lead to solution E. Check the strategy by adding blocks 1 and 4 or blocks 2 and 5 and you will notice that the combinations do not add up to blocks 7 and 8. E can be eliminated as a possible solution. 3) Next you might try adding the number of darkened squares in the 1st, 2nd, 3rd and 4th rows of blocks 1, 2 and 3. You will see that the darkened squares in row 1 add up to 3, the darkened squares in row 2 add up to 2 and the darkened squares in rows 3 and 4 both add up to 4; check the strategy by adding the darkened squares in the rows of blocks 4, 5 and 6, the squares add up to the same numbers as the darkened squares in blocks 1, 2 and 3 (the numbers are 3-2-4-4). Now you know that the darkened squares in the rows of blocks 7, 8 and 9 should also add up to 3, 2, 4 and 4. You can easily see that the solution will have 2 darkened squares in row 1, 1 darkened square in row 2, 1 darkened square in row three and 1 darkened square in row 4. C can be eliminated as a possible solution. 4) You may notice that there are two shapes within each block: an L-shape and a "dot." 5) In each column of blocks, the "dot" stays in the same place. For example, the "dots" in blocks 1, 4 and 7 remain in the lower right hand corner in each block. You now know that the "dot" in the correct answer will be in the last two columns of rows 3 and 4 just as the "dot" in blocks 3 and 6 are in that position. Answer A can be eliminated as a possible solution. 6) Notice the L-shapes move around to 3 possible positions in each row and each column of blocks. Careful study of the different positions of the L-shapes indicate that in the correct solution the L-shape should fit into the upper right hand corner. The answer is B.
computers. Each computer had an EGA Touch-Screen color monitor. The computers were stationed in the same room and were separated by dividers.

*Naturalistic Observation*: Naturalistic observation was used as a means for discovering what events occurred in successful groups that were different from events that occurred in unsuccessful groups. The experimenter observed groups during testing and took notes on individuals' behaviors and on group characteristics. Aside from note taking, the experimenter also examined the group climate by rating the behaviors of group members using three main criteria.

The first criterion observed was how freely the group members communicated. For instance, the experimenter noted whether group members' communicative behaviors were immobile and passive or easy going and open. The experimenter observed the frequency in which group members addressed each other and the direction of their discussions. The second criterion observed was the overall attitude of groups toward the task. For example, comments like "you want to just guess" suggested a disinterested attitude, while comments like "maybe we should go about this a whole different way" expressed a more positive attitude toward the task. The experimenter also relied on observations of the group members' body
positions to draw conclusions about attitude toward the task. Group members who leaned forward intently while solving problems, for example, were viewed differently than group members who leaned back in their chairs away from the computer. Third, the overall attitude of group members toward each other was evaluated. The experimenter observed whether group members actively engaged each other in conversation, tolerated each other, or ignored one, or both, of the other parties.

All ratings were made on a 5-point scale consisting of the following criteria: (a) very poor, (b) poor, (c) average, (d) good and (e) very good. The three ratings were averaged into one overall rating for the group, (i.e., average). Although highly subjective, the ratings and notes provided unique information about the communicative modes and problem-solving styles of the groups. The study’s design curbed the subjectivity of the ratings in that success on the task was measured by the amount that groups scored above the best members. Even though the experimenter was not blind to how well the groups were performing, she was blind to how well the groups were performing in relationship to the best member. Group scores were compared to their best members’ scores within each of the five rating categories.
System for the Multiple Level Observation of Groups (SYMLOG):

SYMLOG is a paper and pencil task that can be used to examine cooperation strategies of groups and characteristics of group members' behaviors. The task consists of an adjective rating form with 26 parameters describing three dimensions. The dimensions are: (a) friendly/unfriendly (b) dominant/submissive and (c) acceptance of authority/non-acceptance of authority. Groups were observed by an experimenter during testing. Following testing the observer rated each group member. The group members rated themselves and each other. Group SYMLOG scores were determined by compiling the group member, self and experimenter ratings for each group into one composite group profile. The profiles were analyzed in a hierarchical complete linkage (furthest neighbor) cluster analysis.

Bar charts were used to compare the average profile of the optimally performing groups with the average profile of the worst performing groups. The optimum and worst group profiles were determined by the cluster analysis. For the best and worst performing clusters, the profiles were broken down to reflect the three dimensions of the scale. For a detailed description of SYMLOG see Bales (1979). (See Figure 3 pp.37 for a visual representation of the test items and how the items
relate to the three dimensions of the scale.)

**Design**

*Independent Variables:* The 192 subjects were randomly assigned to four conditions of 48 subjects each. All subjects attended two sessions scheduled 3 weeks apart. Subjects participated either alone both times, in a group of three and then alone, alone and then in a

**Figure 2**

*Graphic Depiction of Design:* (I = individual; G = group)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N=48</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>N=48</td>
<td>G (SYMLOG)</td>
</tr>
<tr>
<td>3</td>
<td>N=48</td>
<td>I</td>
</tr>
<tr>
<td>4</td>
<td>N=48</td>
<td>G</td>
</tr>
</tbody>
</table>

Alternate test forms were crossed within each condition. The best, average and worst individual scores were matched with their respective groups for the analyses.

group of three, or in a group of three both times. There respective abbreviations for the four conditions are II, GI, IG, and GG. "I" signifies an individual test session and "G" signifies a group test session. There were two test forms (A and B) which were administered in
two different orders (AB and BA). Test sequence was
crossed within the four conditions in a 4 X 3 X 5 X 2
(Condition X Ability X Process X Order) design (Figure
2).

**Dependent Variables:** The main dependent variables were:
(a) group and individual test scores on the progressive
matrices tests, (b) experimenter ratings of group process
and (c) the group SYMLOG scores.

**Procedure**

**Session One:** Subjects arrived at the laboratory in
groups of three. Upon arrival subjects were told that
the purpose of the study was to compare group and
individual performance on a computer task. The
experimenter explained the importance of promptly
returning for the second session and gave each subject an
appointment card. The appointment cards listed the
experimenter’s phone number along with instructions for
canceling appointments if an emergency arose.
Appointments were to be canceled at least 24 hours in
advance. Subjects were told they would be called and
rescheduled if one group member had to cancel. Subjects
signed an informed consent form and were randomly
assigned to one of the four conditions.

Subjects who were tested individually were taken
into the testing room one at a time and seated within
arm's length of a computer. Subjects tested alone were told to read the test instructions on the computer screen and follow the directions. Subjects were asked to do their best and work as quickly as possible. Following testing, subjects were dismissed.

Subjects who were tested in groups were taken into the testing room three at a time. Group members were seated within arm's length of one computer. Group members were told to read the test instructions on the computer screen and follow the directions. The experimenter verbally instructed the group members to work together on the task, to do their best and to work as quickly as possible. Following testing, group members were taken into separate rooms. SYMLOG rating scales were given to each group member (in the GI and IG conditions) to complete privately. Instructions for SYMLOG were written on the front cover of the SYMLOG booklet. Subjects were dismissed after they completed the rating scales.

Session Two: Subjects were called and reminded of the second session the night before the subjects' scheduled appointments. The procedure for the second session was identical to the first session. The only difference was that, per their assigned condition, some subjects who were previously tested in groups were tested
individually. Conversely, some subjects who were previously tested individually were tested in groups. All subjects received the alternate test form during the second session. All subjects were debriefed.
Results

**Experiment I**

*Initial Data Analyses:* The first session scores of each experimental condition and each experimental condition’s respective control group were analyzed using 2 X 2 ANOVAs (order X condition). There were no statistically significant differences between the first session scores of (a) the II and IG conditions, or (b) the GG and GI conditions. Because the order of test administration did

**Table 1**

Mean Progressive Matrices Scores and Standard Deviations Based on the Total Percent Correct: Experiment I

<table>
<thead>
<tr>
<th>Condition</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>N=48</td>
<td>58.81 (14.78)</td>
</tr>
<tr>
<td>GI</td>
<td>N=48</td>
<td>75.38 (9.54)</td>
</tr>
<tr>
<td>IG</td>
<td>N=48</td>
<td>57.44 (14.33)</td>
</tr>
<tr>
<td>GG</td>
<td>N=48</td>
<td>72.06 (9.62)</td>
</tr>
</tbody>
</table>

A star (*) indicates statistical significance p<.05 for comparisons of session 1 and session 2 by condition.

not significantly affect scores, order was excluded from further analyses.
One-way ANOVAs were used to compare the first and second session scores within each condition. There were no statistically significant differences from session one to session two in either the II or GG conditions indicating that (a) one session of individual practice did not affect future individual scores and (b) one session of group practice did not affect future group scores. As expected, group scores exceeded overall individual scores in both the GI, F(1,95)=29.09, p=.00 and IG, F(1,95)=88.40, p=.00, conditions (Table 1).

A between-subjects one-way ANOVA was done on the session 1 scores from the four conditions to compare group scores with the best individual scores. Although there was a trend suggesting that groups were outperforming the best individuals by 5%, statistical significance was not reached, F(1,63)=3.15, p=.08. Thus, unpracticed groups did not significantly outperform the best scoring unpracticed individuals in a between-subjects design.

Influences of Best Members and Group Processes in the Experimental Conditions (GI and IG): A 2 X 5 X 2 ANOVA was done with two levels of subject-type (best member vs. group), five levels of group process (very poor vs. poor vs. average vs. good vs. very good) and two levels of condition (GI vs. IG). The best member scores were
determined by identifying the highest ranking individual score within each assigned group. Group process was measured by experimenter ratings of group member communication, attitudes expressed toward the task and attitudes expressed toward each other. There was no effect for condition. There was a main effect for subject-type, F(1,63)=10.1, p=.002, indicating that groups outscored the best group members by 8% (Table 2).

Table 2

Mean Progressive Matrices Scores and Standard Deviations Based on the Total Percent Correct: GI and IG Conditions

<table>
<thead>
<tr>
<th>Conditions</th>
<th>GI and IG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-type</td>
<td></td>
</tr>
<tr>
<td>Group Scores N=32</td>
<td>77.25 (8.66)</td>
</tr>
<tr>
<td>Best Scores N=32</td>
<td>69.50 (10.37)</td>
</tr>
<tr>
<td>Average Scores N=32</td>
<td>60.50 (11.50)</td>
</tr>
<tr>
<td>Worst Scores N=32</td>
<td>49.56 (13.24)</td>
</tr>
</tbody>
</table>

Paired t-tests showed that for the GI and IG conditions combined, groups outperformed (a) the best group members by 8%, t(1,31)=4.22, p=.00, (b) the average members by 18%, t(1,31)=9.11, p=.00 and (c) the worst members by 28%, t(1,31)=13.44, p=.00. Other comparisons showed that the best members scored 9% above the average members, t(1,31)=5.94, p=.00 and 20% above the worst members, t(1,31)=10.94, p=.00. The average scores were 11% above the worst scores, t(1,31)=8.26, p=.00.
There was also a main effect for group process, $F(4,63)=2.48$, $p=.054$, indicating that groups benefitted from process as measured by experimenter ratings.

The experimenter was blind to how well the groups were doing relative to the scores of the best members. Table 3 illustrates how much group scores exceeded best members' scores in each rating category. The differences show that the groups outscored the best members by larger

**Table 3**

Mean Progressive Matrices Scores for Groups and Best Members and Total Number of Groups Represented in each Category: Group Process

<table>
<thead>
<tr>
<th>Group Process</th>
<th>Very Poor</th>
<th>Poor</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Scores</td>
<td>83.00</td>
<td>64.20</td>
<td>73.14</td>
<td>70.92</td>
<td>66.71</td>
</tr>
<tr>
<td>No. of Groups</td>
<td>(1)</td>
<td>(5)</td>
<td>(7)</td>
<td>(12)</td>
<td>(7)</td>
</tr>
<tr>
<td>Group Scores</td>
<td>60.00</td>
<td>68.20</td>
<td>74.86</td>
<td>80.83</td>
<td>82.43</td>
</tr>
<tr>
<td>No. of Groups</td>
<td>(1)</td>
<td>(5)</td>
<td>(7)</td>
<td>(12)</td>
<td>(7)</td>
</tr>
<tr>
<td>Difference between Group Scores and Best Member Scores</td>
<td>-23</td>
<td>+4</td>
<td>+1</td>
<td>+10’</td>
<td>+16’</td>
</tr>
</tbody>
</table>

A star (*) indicates statistical significance $p<.05$ for the difference between group and best member scores in a category.
amounts when process was rated as good. One-way ANOVAs showed that: (a) for groups rated as average or below average there were no statistically significant differences between the scores of the groups and their best members, (b) groups rated as good outscored their best members by 10%, $F(1,23)=7.7, p=.01$ and (c) groups rated as very good outscored the best members by 16%, $F(1,13)=13.4, p=.003$ (Table 3).

The SYMLOG group interaction ratings were analyzed first in a hierarchical complete linkage (furthest neighbor) cluster analysis. The distance between clusters was calculated as the distance between the two furthest points in the clusters. Specifically, euclidian distances were used. For each variable, distance was calculated using the square root of the sum of the squared differences in the values (Norusis, 1990). The average ratings for each group on the 26 parameters of the scale were entered into the analysis. The scores came from groups in the GI and IG conditions. Although up to nine emerging clusters were identified, the five clusters that emerged on the fifth stage were chosen to represent the data to ensure that all of the clusters would be greater than one.

The five clusters were entered into a $2 \times 5 \times 2$ ANOVA with two levels of subject-type (best member vs.
Figure 3  SYMLOG Profiles: Best and Worst Performing Clusters

<table>
<thead>
<tr>
<th></th>
<th>NEVER</th>
<th>RARELY</th>
<th>SOMETIMES</th>
<th>OFTEN</th>
<th>ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average amount of</td>
<td>0</td>
<td>0.5</td>
<td>1.5</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>group scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>above best member</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scores and key:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Cluster = +13.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst Cluster = +2.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U  Active, dominant,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>talks a lot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UP Extroverted, outgoing,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UFF A purposeful democratic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>task leader</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UF An assertive business-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNF Authoritarian, controlling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disapproving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UN Dominating, tough-minded,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>powerful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNB Provocative, egocentric,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shows off</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UB Jokes around, expressive, dramatic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPB Entertaining, sociable,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smiling, warm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P  Friendly, equalitarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PF Works cooperatively with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F  Analytical, task-oriented, problem-solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF Legalistic, has to be right</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N  Unfriendly, negativistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB Irritable, cynical, won’t cooperate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B  Shows feelings and emotions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB Affectionate, likable, fun to be with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DP Looks up to others, appreciative, trustworthy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPF Gentle, willing to accept responsibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF Obedient, works submissively</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNF Self-punishing works too hard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DN Depressed, sad, resentful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNB Alienated, quits, withdraws</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB Afraid to try, doubts own ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPB Quietly happy just to be with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D  Passive, introverted, says little</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

U = Upward = Dominant Behaviors  D = Downward = Submissive Behaviors
P = Positive = Friendly Behaviors  N = Negative = Unfriendly Behaviors
F = Forward = Acceptance of Authority  B = Backward = Non-Acceptance of Authority
group), five levels of SYMLOG categories and two levels of condition (GI vs. IG). Again there was a main effect for subject-type, $F(1,63)=9.42$, $p=.004$ and no effect for condition. The SYMLOG categories only neared statistical significance, $F(4,63)=2.08$, $p=.10$.

The average level of each of the 26 parameters in the best and worst clusters are represented in profile form (Figure 3). The success of a cluster was determined by the average amount that groups in that cluster scored above the best members. Groups in the best cluster scored 13.5% above their best members ($N=4$), whereas groups in the worst cluster performed only 2.25% above their best members ($N=4$). The profiles show that the successful groups were characteristically more dominant, friendly and accepting of authority than the less successful groups.

**Group Ability versus Best Member’s Ability in Group Performance:** Several between-subjects analyses using one-way ANOVAs revealed a series of findings, all of which support a group ability model of performance. The results indicated that groups that had one session of individual practice outperformed (a) groups that did not have any kind of practice, $F(1,31)=5.35$, $p=.003$, (b) the best individual scores following individual practice,
F(1,31)=4.25, p=.048 and (c) the best individual scores following group practice, F(1,31)=5.01, p=.0327. The findings suggest that a single session of individual practice is beneficial to the future performance of groups on a similarly complex task (Table 4).

**Table 4**

Mean Progressive Matrices Scores and Standard Deviations

Based on the Total Percent Correct: Practiced Groups versus Unpracticed Groups and Practiced Individuals

<table>
<thead>
<tr>
<th></th>
<th>GG Groups</th>
<th>II Best Members</th>
<th>GI Best Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG Groups</td>
<td>79.13</td>
<td>72.06</td>
<td>71.81</td>
</tr>
<tr>
<td>Session 2</td>
<td>(7.24)</td>
<td>(9.62)</td>
<td>(12.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>71.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.88)</td>
</tr>
</tbody>
</table>

A star (*) indicates statistical significance p<.05 for comparisons of the indicated means with the average score of groups in the IG condition, session 2.

Further support for a group ability model of group performance came from stepwise multiple regression analyses performed on data from the GI and IG conditions. The best predictor of former group performance in the GI condition was the interaction between the best X average X worst individual scores, R^2=.46, F=11.86, p=.004,
whereas the best predictor of future group performance in the IG condition was the worst individual score, \( R^2 = .42, F = 10.16, p = .007 \). Thus, group members in the GI condition may have learned about the task from each other and then applied the new knowledge when working alone. Data from the IG condition showed that a group can expect to perform about 34\% better than the worst member did during individual practice, \( t(1,15) = 15.15, p = .00 \), (Table 5).

**Table 5**

**Mean Progressive Matrices Scores and Standard Deviations Based on the Total Percent Correct: GI and IG by Subject-type**

<table>
<thead>
<tr>
<th>Condition</th>
<th>GI</th>
<th>IG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject-type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Scores</td>
<td>N=16</td>
<td>75.38 (9.75)</td>
</tr>
<tr>
<td>Best Members</td>
<td>N=16</td>
<td>71.81 (10.88)</td>
</tr>
<tr>
<td>Average Members</td>
<td>N=16</td>
<td>62.13 (11.60)</td>
</tr>
<tr>
<td>Worst Members</td>
<td>N=16</td>
<td>52.88 (12.59)</td>
</tr>
</tbody>
</table>

**Evidence for Learning from Group Practice:** The implication of the above multiple regression statistics for the data from the GI condition was that group members may have
learned from each other during the group practice session. Comparisons of the first session scores of the II condition with the second session scores of the GI condition were made to test the hypothesis that group members may have benefitted from the group experience.

A between-subjects 3 X 2 ANOVA with three levels of subject-type (best member vs. average member vs. worst

**Table 6**

**Mean Progressive Matrices Scores and Standard Deviations Based on the Total Percent Correct: Best, Average and Worst Individuals Not Participating in Groups and the Best, Average and Worst Members in Real Groups**

<table>
<thead>
<tr>
<th>Condition</th>
<th>II Session 1</th>
<th>GI Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Member</td>
<td>N=16</td>
<td>70.63 (10.66)</td>
</tr>
<tr>
<td>Average Member</td>
<td>N=16</td>
<td>58.25 (11.36)</td>
</tr>
<tr>
<td>Worst Member</td>
<td>N=16</td>
<td>47.56 (12.54)</td>
</tr>
<tr>
<td>Grand Mean</td>
<td>N=48</td>
<td>58.81 (14.78)</td>
</tr>
</tbody>
</table>

A star (*) indicates statistical significance p<.05 for comparisons of session 1 and session 2 by condition. There were no statistically significant differences between the two conditions.
member) and two levels of condition (II session 1 vs. GI session 2) was done. There was a main effect for subject-type, $F(2, 95) = 26.11$, $p = .00$, but not for condition, $F(1, 95) = 2.12$, $p = .15$. Thus, there was no evidence that group members improved significantly following one session of group practice. The means, however, were suggestive of a trend toward improvement for individuals who had participated in a group practice session (Table 6).

**Summary:** In a within-subjects design groups outperformed their best members regardless of condition. In a between-subjects design groups outperformed their best members, but the difference was not statistically significant. Experimenter ratings of group process offered some insight into how process affects a group's ability to score better than the best member. The SYMLOG interaction scores, when they were converted into profiles, provided visual representations of the different characteristics of successful and unsuccessful groups. The data provided much support for a group ability model of group performance. There was some suggestion in the data that subjects in the GI condition were learning from participating in a group session. Further analyses using a between-subjects design showed that there was no statistically significant improvement
in scores following the group practice session. The study indicates a need for further research using more within-subjects designs and applied problems.
Discussion

The hypothesis that groups can outperform their best members on a complex multi-step task was confirmed. In a within-subjects design, groups scored 8% above their best members overall and 11% above their best members following individual practice. In a between-subjects design, unpracticed groups scored 5% above unpracticed best scoring individuals. The difference, however, was not statistically significant (p=.08). The findings suggest that individual practice may contribute to future group success, possibly by improving group process.

Experimenter ratings of group process were associated with the ability of groups to succeed on the task. Groups that were rated as good or very good outscored their best members, while groups rated as average or below average did not. Thus, verbal and non-verbal communication, as well as the attitudes of group members toward the task and toward each other may have influenced the group's ability to excel as a dynamic unit.

The finding that ad hoc groups have the ability to outperform the potential of the best members supports a group ability model of group intelligence, as does the comparison of practiced groups to unpracticed groups. In the present study, practiced groups outperformed
unpracticed groups by 7%. The finding seems to be reliable in that it replicates the results of a preliminary study which also found a 7% difference. The original hypothesis stated that process would determine the success of practiced groups. Practiced groups were successful even when process scores were eliminated from the analysis. The effect was likely a result of the level of task difficulty (Shaw & Ashton, 1976). The difficult task may have promoted intense group discussion which may have in turn cultivated the group processes. Based on the group ability model, one can infer that group members in the IG condition shared and built upon personally held information that was gained from the individual session, a process which helped the groups to excel at problem-solving.

The group ability model of group intelligence has practical implications in regard to cooperative learning. College students who have, for example, been given the opportunity to review, or work on, a problem-solving task by themselves and then work in a group on a similar task may develop a better understanding of how to successfully solve problems by sharing and building on the information of other group members. The experimenter's ratings and group SYMLOG scores have shown the importance of group process to successful group functioning.
The hypothesis that if groups outperform their best members, then the processes occurring in the successful groups will be unique was substantiated by the SYMLOG clusters. Although the SYMLOG clusters only neared statistical significance when entered into an analysis of variance with group and best member scores, the clusters provided useful information about group processes when studied individually. Figure 3 (pp.37) shows how the most successful groups were rated more highly on the factors associated with dominance, friendliness and acceptance of authority. More specifically, the successful groups were rated as more confident, extroverted, cooperative and task-oriented (among other characteristics) than the less successful groups. Successful groups appear, not surprisingly, to have better communication skills and higher levels of characteristics often associated with a strong work ethic.

Future research might try to diagram more specifically the actual communication processes transpiring in successful groups. The experimenter noticed two phenomena that were occurring in the most successful groups. Either (a) the solution was known by one (two) subject(s) who taught the other two (one) how they arrived at the solution or (b) all three subjects
were working on the problem together, literally building a solution bit by bit, trial by success, without giving up. The events just described were most likely the processes that helped group members to (a) learn from each other about strategies that can be used for success on the task and (b) go beyond the ability of the group's best member. Diagramming step by step how groups share, teach and build on individual contributions would provide maps for understanding effective and noneffective cooperative problem-solving strategies.

Other support for a group ability model of group performance came from tests of linearity. If group intelligence was a function of the best member's ability, re-solving group scores and their best members' individual scores would be linearly related. There was not a linear relationship in the IG condition between the scores of the best members and the groups, but there was a linear relationship between the scores of the worst members and the groups. Groups in the IG condition performed 34% better than the worst members. The difference disputes the old adage that a team is only as good as its weakest member. In the present case the 'team' actually performed 34% better than the weakest member.

For groups in the GI condition, there was a linear
relationship between the interaction of the best, average and worst individual scores and the group scores. The relationship suggests that group members may have learned from group practice and become more similar in relation to the task. The finding corroborates past research on other tasks that have shown group members become alike in relation to the task following group practice (Dillon, Graham, & Aidells, 1972; Schoner, et al., 1974).

The effect of group practice on individual scores was tested in a between-subjects analysis. If the group product was a function of group ability, then all members should have been gaining (improving, learning) from the process. The data showed a trend indicating better scores for individuals with group practice (regardless of ability). There was not, however, any statistically significant evidence to support the differences. The nonsignificant result is the only evidence in the present study not in support of the group ability model of group intelligence. Future research might find a different result by using a within-subjects design.

Also supported only by trends, and not statistics, was the hypothesis that group members who contribute the least to the group experience gain the most from the process while the best group members are not affected. Studies dating back as far as 1924 (see Bechterev &
Lange, 1924) generally have shown that the lowest ability group members benefit the most from group practice (see Hill, 1982). The trends of the present study are in accord with the other research in that the worst members benefitted more than the others (worst members improved 5%, average members improved 4% and the best members improved 1%). Although the data offered no proof that learning occurred during one session of group practice, there was an implication in the multiple regression statistics that it did (see pp.39-40). Whether learning occurs during more than one session of group practice was a hypothesis tested in Experiment II.

A purpose of the present study was to uncover how a group assembled randomly can reach the peak of its functional ability. The present experiment has shown that groups can break through the 'ceiling' imposed by the best member's ability. The benefits, however, may be confined to the group experience and not spill over into long term improvements (at least not for the best member, not by much and not statistically). The support offered for the group ability model of group intelligence should be regarded in terms of the fundamental aspects pertaining to the particular task being studied. Because successful group performance is known to be task specific, we would benefit from knowing exactly what
kinds of real problems can be better solved by groups. Future research might help address this question by testing the theories outlined in the present study using applied problems that have the following properties: (a) multiple-steps, (b) conceptually-driven, (c) cognitively based and (d) require good visual/spatial abilities. One could look to mathematics, engineering and computer science for problems with these qualities.

Experiment I has answered some of the questions about how practice, process and ability affect group performance. Experiment II was designed to further explore the effects of group and individual practice on future individual performance. Experiment I demonstrated that group members may have learned during group practice and then applied what they learned at a later time when they were working alone on the task. A between-subjects analysis showed a trend indicating better scores for individuals who had group practice (regardless of ability) in comparison to a control group, but the differences were not statistically significant. Experiment II has a pretest-posttest design so that within-subjects comparisons can be made.

Another purpose of Experiment II was to follow-up on (a) Smith's (1989) theory that the amount of group practice may have a significant influence on future
individual performance on a progressive matrices task and (b) Shaw and Ashton's (1976) theory that task difficulty affects performance. Experiment II was designed to demonstrate whether (a) group practice is better than individual practice, (b) more group practice is better than less group practice and (c) the complexity of the test items solved during group practice affects future individual performance.
Method

Experiment II

Subjects

Subjects were 108 undergraduates enrolled in psychology classes at Case Western Reserve University. Subjects participated for class credit. There were 53 males and 55 females, 49% and 51% respectively. Two subjects did not return for the second session causing the loss of 2 groups (6 subjects). To compensate for the loss and maintain equal sample sizes, 6 fewer subjects were run in the experiment. The lost subjects are not reflected in the above numbers. Data from 2 subjects who received scores below what they would have had by chance were thrown out and replaced by data from 2 new subjects. All subjects were informed of the right to terminate participation at any time. All subjects were debriefed.

Materials and Apparatus

Computerized Cognitive Abilities Test (CAT): For a general description of CAT see Experiment I. Four tests were developed for Experiment II using a total of 116 test items. There were four test sessions. Two parallel test forms were used in the pretest and posttest periods. The test forms were crossed within each condition. Internal consistency reliabilities were calculated for the two test forms. Alpha coefficients (Cronbach, 1970) were .71
and .80 based on the pretest scores of the entire sample (N=108). The pretest consisted of 32 items and the posttest consisted of 38 items. The six additional items in the posttest were added to increase difficulty and prevent ceiling effects. The different test lengths did not pose a problem for the analysis because all scores were converted into percentages. The pretests and posttests consisted of both easy and difficult items. The test used during the second session consisted of 24 easy items. The test used during the third session consisted of 22 difficult items. The difficult test was slightly shorter than the easy test because the items were more time consuming. The pretest and posttest sessions took an average of 45 minutes.

**Computers:** Tests were administered on three 286 IBM computers. Each computer had an EGA Touch-Screen color monitor. The computers were stationed in the same room and were separated by dividers.

**Design**

The experiment was designed to be analyzed by a 2 X 4 X 2 (session: pre/post X condition X order) MANOVA for repeated measures. Order was withdrawn from the analyses after initial statistics found that order had no affect on scores. Sex was added to the analysis post hoc as a covariate. See Figure 4 for a graphic depiction of the
design.

**Figure 4**

**Graphic Depiction of Design: Experiment II**

(I = individual; G = group)

<table>
<thead>
<tr>
<th>Session</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pretest</td>
<td>simple</td>
<td>complex</td>
<td>posttest</td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 N=27</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>2 N=27</td>
<td>I</td>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>3 N=27</td>
<td>I</td>
<td>I</td>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td>4 N=27</td>
<td>I</td>
<td>G</td>
<td>G</td>
<td>I</td>
</tr>
</tbody>
</table>

Alternate test forms were crossed within each condition.

_Independent Variables:_ The 108 subjects were randomly assigned to four conditions of 27 subjects each. Each subject participated in four sessions scheduled one per week for 4 weeks. Subjects in the control condition participated alone during all four sessions (IIII - no group practice). Subjects in condition two worked in groups during the second session (IGII - group practice on easy test). Subjects in condition three worked in groups during the third session (IIIG - group practice on
difficult test). Subjects in condition four worked alone during the pretest and posttest periods and in groups during sessions two and three (IGGI - group practice on easy and difficult tests).

There were two test forms and two sequences of test form administration (AB and BA) for the pretest and posttest periods. The sequence in which the test forms were administered was crossed within each condition. Every subject received different test forms during the pretest and posttest sessions.

The test items were the same for all subjects during the second and third sessions regardless of condition. See pages 22 and 23 for a detailed description of the tests.

**Dependent Variables:** The main dependent variables were pretest and posttest scores. Problem solving ability on the CAT was determined by the number correct converted into a percentage. Also of importance were the third session scores for the III and IGII conditions. Third session scores for the two conditions provided information regarding practice effects incurred from previous group work on simple items.

**Procedure**

Subjects arrived at the laboratory in groups of three. Subjects were randomly assigned to one of the
four conditions and told the general purpose of the study was "to compare group and individual performance on a cognitive task." All subjects were asked to sign a consent form. Treatment of subjects who were tested individually was the same across sessions. Similarly, treatment of subjects who were tested in groups was the same across sessions.

Subjects who were tested individually were taken to a back room and seated in front of a computer within arm's length of the screen. Subjects were reminded that the test is a cognitive abilities task and were asked to read the test instructions displayed on the computer screen. The experimenter told the subjects to "try to work as fast as you can and do the best that you can." There were no restrictions on testing time. Following test sessions 1, 2 and 3, subjects were given appointment cards and dismissed. Following session 4, all subjects were debriefed.

Subjects participating in groups were taken to the back room in a group of three and seated in front of one computer. All group members were within arm's length of the computer screen. Subjects were reminded that the test is a cognitive abilities task and were asked to read the test instructions displayed on the computer screen. Group members were told to "try to work as fast as you
can and do the best that you can." Again no time limit was set for test completion. At the end of the forth session all subjects were debriefed.
Results

Experiment II

Data Analyses: Subject scores were analyzed using a MANCOVA for repeated measures with sex as a covariate. The data were analyzed in a 2 X 4 factorial design (session: pre/post X condition). Order of test administration, which was previously found to have no significant affect on scores, was excluded from the analyses. Grouping of subjects by condition (IIII, IGII, IIGI and IGGI) was a between-subjects variable, while session (pretest vs. posttest) was a within-subjects variable. Separate MANCOVAs for repeated measures were done for each condition to pinpoint specific effects. A one-way ANOVA was done on the third session scores of subjects in the IIII and IGII conditions. A 2 X 2 ANOVA (session X sex) and t-tests were done to determine the extent of the affect of the covariate in the pretest and posttest sessions.

Repeated Measures (Condition x Session): Group practice did affect future individual performance on the progressive matrices test. A MANCOVA for repeated measures showed a significant interaction between condition and session, F(3,104) = 4.07, p = .009, (Table 7).
Table 7

Mean Progressive Matrices Scores and Standard Deviations Based on the Total Percent Correct: Experiment II

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIII</td>
<td>N=27 72.89 (10.32)</td>
<td>69.63 (11.81) -*</td>
</tr>
<tr>
<td>IGII</td>
<td>N=27 77.11 (10.12)</td>
<td>75.67 (10.33) -</td>
</tr>
<tr>
<td>IIGI</td>
<td>N=27 70.37 (13.92)</td>
<td>75.07 (10.04) +</td>
</tr>
<tr>
<td>IGGI</td>
<td>N=27 72.59 (13.10)</td>
<td>76.41 (8.52) +*</td>
</tr>
<tr>
<td>Grand Mean</td>
<td>N=108 73.24 (12.06)</td>
<td>74.19 (10.45)</td>
</tr>
</tbody>
</table>

A star (*) indicates statistical significance p<.05. A minus (-) indicates a decrease from pretest to posttest. A plus (+) indicates an increase from pretest to posttest.

Additional MANCOVAs analyzing pairs of conditions revealed that the IIII and IGGI conditions were responsible for the condition by session interaction, $F(1,52)=11.73$, $p=.001$. Separate MANCOVAs done for each condition indicated statistically significant differences between the pretest and posttest scores of subjects in the IGGI condition $F(1,26)=5.68$, $p=.025$ and in the control group $F(1,26)=6.24$, $p=.019$. Mean scores demonstrate that the performances of subjects in the IGGI
condition improved from pretest to posttest, whereas the performances of subjects in the IIII condition deteriorated.

The scores of subjects in the IGII condition did not change significantly from pretest to posttest, $F(1,26)=.55$, $p=.47$, but a one-way ANOVA comparing the third session individual scores of the IIII and IGII conditions showed that subjects who had group practice on the easy test did better on the complex test than subjects in the control group, $F(1,53)=6.27$, $p=.016$. The third session means for IIII and IGII were 44% and 52%, respectively. The difference between pretest and posttest scores in the IIGI condition was not significant, although there was a trend toward improvement, $F(1,26)=3.20$, $p=.085$.

**The Covariate:** Sex was not controlled for as an independent variable because sex was not expected to influence scores. Gender did influence scores in some way: Sex was a statistically significant covariate, $F(1,103)=5.33$, $p=.023$. A 2 X 2 X 4 ANOVA (session X sex X condition) indicated that there was a trend in the data. The interaction between sex and session approached statistical significance $F(3,215)=1.77$, $p=.154$. The trend implied that (a) males did slightly better than females during the posttest sessions (Table 8) and (b)
Table 8
Mean Scores and Standard Deviations Based on the Total Percent Correct for Males and Females in All Conditions

<table>
<thead>
<tr>
<th>Session</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males N=53</td>
<td>74.85 (11.86)</td>
<td>76.62 (10.85)</td>
</tr>
<tr>
<td>Females N=55</td>
<td>71.69 (12.16)</td>
<td>71.85 (09.57)</td>
</tr>
</tbody>
</table>

Table 9
Mean Change in Scores from Pretest to Posttest for Males and Females

<table>
<thead>
<tr>
<th>Sex</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIII</td>
<td>-4</td>
<td>-4</td>
</tr>
<tr>
<td>IGII</td>
<td>+1</td>
<td>-4</td>
</tr>
<tr>
<td>IIGI</td>
<td>+5</td>
<td>+3</td>
</tr>
<tr>
<td>IGGI</td>
<td>+6</td>
<td>+3</td>
</tr>
</tbody>
</table>

males improved more relative to females following group practice (Table 9). The trend may be reflecting a
difference in competitive reactions of males and females to group experiences. It is possible, for example, that competitive feelings ignited during the group sessions inspired males to try harder than females during the posttest session. The trend in the data warrants further investigation.

**Summary:** The findings indicate that group practice did benefit future individual performance and a lot of practice was better than a little practice. The scores of subjects who worked in groups on two occasions improved over time. Group practice on only one occasion did not significantly improve posttest scores regardless of whether the practice was on the simple problems or on the complex problems. Group practice on only the simple test did benefit subjects when they took the complex test, but the effect did not carry over to the posttest session. There was a trend in the data suggesting that posttest scores improved following group practice on only the complex problems, but the improvement was not statistically significant.
Discussion

The results support the implication of Smith's (1989) finding that working in groups benefits the future individual performances of group members on the progressive matrices task. The hypothesis that more group practice is better than less group practice was confirmed. Having no group practice (IIII) actually hindered performance. The findings are discussed in relation to theories about learning from group participation and a theory about how the developing relationships among group members may have influenced outcomes.

The finding that having only individual practice (IIII) impedes performance may have resulted from an inability to learn much about the task when working alone. Because performance during the pretest tends to reflect most accurately an individual's ability on a task, scores should have stayed the same, not decreased. It is possible that a sense of defeat or dullness led to the poor performance during the posttest session. Being unable to improve may have caused frustration and boredom.

The changes in standard deviations from pretest to posttest support the speculation that learning did not occur in individual sessions and suggest that learning
did occur in groups. Table 7 illustrates how the standard deviations increased from pretest to posttest in the IIII and IGII conditions and decreased over time in the IIGI and IGGI conditions. The smaller the standard deviation the more similar the scores. The amount of change appears to be a function of the amount of and the intensity of the group practice sessions (given the assumption that increased stimuli complexity increases intensity). Thus, two group practice sessions produced the most homogeneity in posttest scores and no group practice sessions the least. The implication is that when working alone on similar but not identical problems, former group members are applying the same strategies they learned during the group sessions.

In the present experiment, learning did occur when subjects participated in one group practice session, although much more learning occurred when subjects worked in two group sessions. Group practice on only the simple test (IGII), for example, benefitted subjects when they took the complex test. The advantage, however, did not carry over into the posttest session. There was a trend suggesting that group practice on only the complex problems (IIGI) improved posttest scores, but the improvement merely came close to statistical significance (p=.085). Only participants who had two group practice
sessions (IGGI) had significantly better posttest scores. If learning does occur in groups, then the phenomenon of IGGI>IIGI>IGII>III for posttest improvement is in accord with the learning theory suggesting that spaced individual learning is superior to massed individual learning (see Rea & Modigliane, 1985). Distributed group learning seems to benefit group members more than a single session of group learning.

The finding that two group practice sessions are better than one for improving individual scores can also be explained by research dealing with the impact of changes in group member relationships on performance. Traditioned groups (groups whose members have a history of working together) have been shown to be more likely to outperform their best members on certain types of tasks (see Michaelson, Watson & Black, 1989). It is feasible that when groups met for a second time (IGGI) they became traditioned groups and profited from what were possibly superior interpersonal relationships.

Determining whether groups benefit from repeated exposure to the same people or just multiple group sessions could be easily tested. An experiment could be done using the same IGGI condition, but varying whether group members are tested in newly formed groups or with the same people during the second group session.
Another study is needed to tease out whether the beneficial effect of a lot of group practice is a result of the presentation order of the simple and complex stimulus items. In the present experiment the session in which simple items were solved always preceded the session in which complex items were solved. A study controlling for the order of stimulus complexity would provide more information about the extent of the beneficial effects of group practice.

Sex was not controlled for in the experiment, but was included as a covariate after preliminary analyses indicated that the performance of men and women was slightly different over time. A trend in the data approaching statistical significance suggested that following two sessions of group practice, the scores of males improved more relative to the scores of females (Table 9). The trend may have resulted from sex differences in one or more of the following areas discussed below: (a) confidence, (b) competitiveness and (c) spatial ability.

Following group practice males may have felt more confident than females about their ability to do well on the task. Siry (1990), for instance, found that after completing a task for the first time in a two session experiment, males expressed greater feelings of
confidence than females did about their ability to perform in the second session. The males in Siry's study did perform significantly better relative to females during the second session. In the present experiment, some aspect of the group experience may have increased confidence in males. Possibly feelings of competition were ignited.

Differences in competitive feelings may explain the trend in the present study. Because the amount of group practice appeared to be a factor related to the differences in scores, working in a group may have brought about more competitive feelings in males than in females (Block, 1979). Block found that males are socialized to be competitive, whereas females are socialized to be cooperative. Furthermore, Garza and Borchert (1991) found that when males are given cooperative feedback in a problem-solving group they tend to behave more competitively. Groups in the present study were told to work together suggesting a cooperative atmosphere. Feasibly, working in the groups may have incited more aggressive ambitions in the men than the women.

Another possible explanation for the trend is that there were sex differences in subjects' abilities to perceive and interpret the spatial relations (see Gouchie
& Kimura, 1991; Stumpf & Klieme, 1989) and do the mental rotations (see Harris, 1978; Kerns & Berenbaum, 1991; Maccoby & Jacklin, 1974) that are necessary to succeed on the spatially oriented cognitive task. Group experience may be tapping into the potential of the males more than the potential of the females resulting in better posttest scores for the males. Further research is needed to find whether the differences are real.
General Conclusions

Two experiments investigated the effects of group and individual practice, group process and cognitive ability on the performance of ad hoc groups on a multi-step task. From these experiments, several meaningful findings have emerged. First, and most notable, was that in a within-subjects design groups outperformed their best members. The success of the groups occurred just as Faust’s (1959) data and Lorge and Solomon’s (1955) Model B theory predicted. The finding reinforces the "group as organism" metaphor used by Williams and Sternberg (1988). The metaphor, which has a Gestaltten quality, suggests that groups operate as unique and dynamic entities, not unlike living organisms, with a potential independent of the individual group members.

Also fostering the credibility of the "group as organism" metaphor is the relationship between group process and group success. The SYMLOG scale showed that the processes occurring in successful groups were different from the processes occurring in unsuccessful groups. The members of successful groups were rated as more dominant, friendly and accepting of authority. Naturalistic observation provided some insight into the different strategies used by the effective and ineffective groups. As others have found (see Hill,
1982), building on shared information was important to the performance outcome. Teaching other group members problem-solving strategies also surfaced as a factor that is beneficial to group success, as did communication style and attitude. The data supports the idea that positive synergic properties can emerge during group problem-solving and increase group potential.

In accord with the results just discussed, the data also provided much support for a group ability model of group performance. The data directly contradicted Tuckman and Lorge's (1962) study. Tuckman and Lorge's data supported an individual ability model. The finding of the present study points to group process and individual practice as factors responsible for group success.

Experiment II demonstrated that there are long-term benefits for individuals who have had two group practice sessions. The finding raised questions about two factors that may have affected individual performance (a) the familiarity of group members with each other and (b) the possibility that group members were learning from the group experiences. The experiment introduced another question regarding sex differences: Does group practice affect the future problem-solving performance of males and females differently?
The findings of the present study and the findings of a preliminary study have proved to be systematic in the direction and relative magnitude of their effects. The use of computers with Touch-Screens for test administration probably aided in reducing the confounds that may have occurred had equivalent pencil and paper measures been used. The findings appear to be reliable and to have practical significance.

In a time when educational systems are grasping for alternatives to large classroom instruction, there is a place for the small group approach to learning. The present study supports the philosophy that using the small group as a learning tool can benefit college students' problem-solving skills. Encouraging group study intermixed with individual study may have lasting benefits for students in classes emphasizing multiple-step problems, (i.e., mathematics, computer program design, theory, etc...). For best results, students should approach the group session after some independent experience with the problem.
Footnote: Using the task as a class exercise.

One application of the present research involves using the individual/group testing sessions as a tool for developing in-class work groups in college classrooms. A class instructor who wants to use small group problem-solving as part of his/her class curriculum might benefit the students by adopting the following exercise. Test students on the computerized progressive matrices task individually and then in groups. Provide feedback about how the group performed in relation to the individual scores and then have the group members reflect on how the group dynamics may have influenced the outcome. The exercise will promote self-analysis and possibly enlighten the students about effective and ineffective interaction styles, communication skills and attitudes. Examining the self in relation to one's experiences in the group may steer students toward productive and engaging group work in the future. Learning to work well in groups may, in turn, help to improve students' problem-solving skills.
References


