ABSTRACT

DEVELOPMENT OF THE INTERPERSONAL SKILLS INTERACTION ANALYSIS: AN INTERACTION ANALYSIS TECHNIQUE TO MEASURE INTERPERSONAL COMMUNICATION SKILLS IN SMALL GROUP SETTINGS

By

Randall M. Isaacson

An examination of the field of affective and socio-emotional education shows an increasing interest in these fields, particularly in relation to the classroom teachers facilitation of human relations and interpersonal communication skills. The present study reviewed this research with an emphasis on those pre-service and in-service teacher education programs which instruct and evaluate interpersonal communication skills. The review of the research in the field pointed out the almost total lack of objective instruments available to researchers to measure group members communication skills.

The emphasis of the present study was the development of an interaction analysis observation instrument which would be capable of measuring interpersonal communication skills in small group settings. Three areas were examined: the instruments reliability, validity and the interpretation of matrices and flow charts.

The reliability of the Interpersonal Skills Interaction Analysis (ISIA) was measured by Scott's π in three areas. The

inter-rater reliability (coefficient of observer agreement) was estimated using Scott's π for four observers with inter-correlations ranging from .72-.88. A live versus taped reliability coefficient was calculated to estimate the loss in reliability due to audio-tape recordings. These coefficients were .72 and .79 demonstrating the acceptability of the audio-tapes. A stability coefficient was also calculated, which demonstrated a greater within group stability than between group stability.

The validity of the ISIA was demonstrated by using participant and expert opinion's ratings of each group. The ISIA distinguished between those groups judged effective and ineffective (Wilcoxon Matched pairs = .002) and further discriminated the differences between the effective and ineffective groups to be due to the communication skills under study. Using a non-parametric statistic (Friedman Anova) the effective and ineffective groups were found to differ on self-disclosure (.002), active listening (.035), feedback (.077) and affective interactions (.031). The validity of the ISIA was further supported by a rank order correlation which showed the individual group opinionnaire data to correlate with the ISIA category data. The findings were further discussed and illustrated through an examination of matrix and flow chart interpretation.

DEVELOPMENT OF THE INTERPERSONAL SKILLS INTERACTION ANALYSIS: AN INTERACTION ANALYSIS TECHNIQUE TO MEASURE INTERPERSONAL COMMUNICATION SKILLS IN SMALL GROUP SETTINGS

By

Randall M. Isaacson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1976

Dedicated to
Elmer P. Isaacson,
my grandfather, who
waited with more patience
for a longer time than
anyone else to call me
Dr. Isaacson
Thank you, Grandpa

Durin

disagreement

not least" in

has been more

below. She he

and other impo

me keep my san

in me. Thank

Others

committee memb

Joe By

forgetful, but

good chairman

Judy H

but who gave m

many years, in

encouraging my

Howard

late innings.

modify my stat

a great deal o

you'll always

ACKNOWLEDGMENTS

During this ordeal I read one author who stated his disagreement with the frequent mentioning of wives as "last but not least" in acknowledgments. I agree. My wife, Christie Lee, has been more helpful to me than all the helpful people listed below. She helped me with ideas, with coding, with proofreading and other important odds and ends. But, most important, she helped me keep my sanity and she believed in me when I didn't believe in me. Thank you Chris.

Others who helped me in more ways than just being committee members were my committee:

Joe Byers, who was often times hard to find and sometimes forgetful, but really right with me when the going got tough. A good chairman and friend.

Judy Henderson, who is now next to impossible to find, but who gave me ideas and guidance in my professional growth for many years, in Ed Zoo and other important areas. Thanks for encouraging my exploratory behavior.

Howard "Nolan" Teitelbaum, a friend who pinch hits in the late innings. A late addition to my committee who helped me modify my statistics after I had collected my data and tried, with a great deal of patience, to understand Ed Zoo. Thanks Nolan, you'll always be my ace in the bullpen.

John Lo
mittee and also
hundreds of thi
ally. I can be
every day.

Abbey S
my writing thar
finish, but I'
loss.

There
acquaintances
I am. A few s
assistants who
I thank you Ro
To two good fr
the ISIA, than
new friend who
getting into,

And la
my attempts a
patient, but I

John Lopis, a close friend who happened to be on my committee and also happened to be my expert judge. Thanks for the hundreds of things you've shown me both professionally and personally. I can best thank you by passing them on to others, and I do, every day.

Abbey Shur, the person who probably helped me more with my writing than anyone else. I'm sorry you weren't around at the finish, but I'm sure the students in Israel have gained from our loss.

There are, of course, hundreds of students, friends, and acquaintances to thank. Each helped me to become the professional I am. A few specifics immediately come to mind. To those teaching assistants who allowed me to invade the privacy of their IPL group; I thank you Ron Jones, Mike Radke, Mary Samuelson, John Graves. To two good friends who dealt with the frustration of learning the ISIA, thank you Felice Shulman and Laurie Moffot. And to a new friend who drew my flow charts and didn't know what he was getting into, thanks Dave Robinson.

And last, but not least, thanks to my family who supported my attempts all through school. I can't really say you were patient, but I always knew you were right with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Since the first groups in Bethel, Maine, in the late 1940's, a great deal of change has come about in the field of group dynamics and interpersonal communication skills. From a few individuals whose primary concern was personal and social change in the area of education, group work has now expanded to many diverse areas of our society. In fact, it would be unusual for an individual, born in the last 20 years not be faced with the decision of participating or not participating in some sort of intensive group experience sometime in his life. The general public has been bombarded by facts and fictions about the intensive group encounter through the media and from individuals who have experienced such groups in business, medicine, religion, weekend workshops, or the formal educational institutions. An enormous amount of energy has been expended by a great many people to call attention to the importance of communication skills in man's everyday life. Some of the proponents of this movement have used their expertise in the field to work with marital problems, business (e.g., National Training Lab), parent-child communications (Gordon, 1970), psychological therapy sessions (Berne, 1961, 1966) and individual growth and self-actualization (e.g., Esalen).

Perhaps the fastest growing, and perhaps the largest sector of the intensive group experience movement is attempting to bring about changes in our educational institutions (Christ, 1972). These changes encompass diverse experiences derived from a variety of educational viewpoints at all levels of our schools, from early elementary school through college. While the changes very often involve changes in educators' viewpoints and behavior toward students they sometimes involve changes in curriculum. Interpersonal communication skills seem to have become a curricular area in and of themselves, not just a side issue to be dealt with when communication becomes a problem. Communication skills have become a subject matter to be taught in schools just as science or math.

In spite of the variety of settings in which group work is being used and the diversity of viewpoints which underlie the various approaches of the group leaders, one finding of most group work is the lack of systematic evaluation and research. In the preface to one of the most in-depth studies of sensitivity groups, it is stated that, "The explosive expansion of the use of groups for personal change has not been matched by corresponding concern for information about what such groups do and how well they do it. Innovation has exceeded evaluation" (Lieberman, et al., 1973, p. vii).

The lack of evaluation and research of group work leaves many with questions as to what group participation involves and what possible benefits or harm might accrue to those involved in such intense group work. This presents a real concern to many

individuals who are faced with the choice of joining such groups. At this time, their only recourse is to seek out subjective reports from those who are, or have been, involved in such experiences. Moreover, there exists greater problems for those individuals who are exposed to group work in a less than voluntary situation. With the increasing emphasis on communication skills and affective education, a substantial portion of the school population is now exposed to some sort of group work in their schools. Programs such as DUSO (Dinkmeyer, 1970) and FOCUS (Anderson and Miner, 1971) at the elementary school level and value clarification programs in high schools put a great deal of emphasis on teaching communication skills, and to the layman those programs may seem similar, if not identical, to sensitivity groups for children. This has caused concern for parents and educators, as much of the publicity and popularity associated with the sensitivity movement has centered around what might be called a Bob-and-Carol-and-Ted-and-Alice fantasy, the "touchy-feely" aspect of the intensive group experience. In light of the paucity of facts and the almost complete lack of research and evaluation, this fantasy has become a reality for a great many people.

For that portion of the society which chooses to voluntarily participate in intensive group work, research on the possible benefits or harm of such experiences would be helpful but perhaps not essential. If the choice is left to the individual and even if the experience had no long-term benefit, many individuals might elect to participate in such a group as recreational adventure.

In education, however, the value of an activity which lacks a specific goal is now being questioned. The public wants to know what their children are learning in schools and more importantly many parents want to know why their children are learning those facts and skills. The public wants schools to be held accountable for the "what and why" of learning.

Educators are looking more closely at what teachers need to know and do to be more successful in the classroom and have developed curricula which explicitly stipulate what skills have to be mastered (e.g., competency-based education). Educators are concerned with teacher's in-class performance; what skills teachers need to exhibit rather than simply what knowledge they possess about math or any other subject matter. One area in which teachers need new knowledge and skills is interpersonal communications.

Colleges of Teacher Education are beginning to change also. Standard methods courses do not fully prepare teachers to use programs such as DUSO (Dinkmeyer, 1970) and FOCUS (Anderson and Miner, 1971) which has created a growing need for per-service teacher education course work in interpersonal communications, human relations, and value clarification training. Programs and course work in these areas are moving slowly but the interest and need will certainly bring about more changes at a faster rate in the coming years in both schools and colleges of teacher education.

In all areas of education, the need for evaluation and research on intensive group experiences is essential. Accountability

on a national as well as a local level calls for evaluation of specific objectives. Performance-based teacher education requires basic competencies for teachers which demands an assessment of the criteria by which we shall judge teachers. New Teacher Education programs which emphasize interpersonal communication and human relation skills need to implement both formative and summative evaluation procedures (Scrivens, 1967). The sensitivity movement has had an influence on the recognition of the need for socio-emotional and affective education in school; but it has not helped develop the criteria for the competencies needed by teachers to deal with these domains, nor has it aided in developing a method of evaluating such skills.

A few colleges of Teacher education, e.g., University of Georgia (Gazda, et al., 1973a), University of Massachusetts (Allen & Cooper, 1967), and Michigan State University (Lopis, 1973), have begun to develop programs which focus on the need for teachers to be trained in interpersonal communication and human relation skills. Recently, two state legislatures, Minnesota and Wisconsin, have made it mandatory that schools of teacher education instruct their prospective teachers in human relation skills. Programs dealing with human relations, effective education, and interpersonal communication skills are increasing in number, but methods of evaluating such programs are still somewhat primitive. Evaluation is not keeping pace with development. The present research will begin to look at the evaluation of such intensive group

experiences and propose one method, an observation instrument for interpersonal communication skills, of doing such evaluation and research.

Purpose of the Study

It is time for those involved in groups whose goal is change, to specify the desired change and assess that change. Thus far, the emphasis has been on process, but in a society that is increasingly calling for accountability we must not neglect the product. To understand the group process as it unfolds is an important part of the group functioning; but to understand the product of group participation in terms of skill acquisition and changes in groups members' behavior is also important. Added significance for greater understanding results from societal pressure for such information.

This research will attempt to begin to look at groups--Interpersonal Process Laboratories (IPL) in particular--to examine the behavior of both the facilitator and group members while participating in the group, and to develop an observation system to measure those behaviors.

Rationale for the Research

The increasing interest in intensive small group involvement has created the need to evaluate such group in terms of what participation entails and also of what long term consequences participants can expect. In the preface to his book Beyond Words: the Story of Sensitivity Training and the Encounter Movement, Back

(1972) states, "Renewed interest in formal evaluations and studies underway may soon relieve the gloomy picture of the state of the research shown here" (p. xi). But the gloomy picture painted by Bach in 1972 is still with us and, as noted below, may be getting worse.

Recently the intensive group movement and the interest in communication skills has gone beyond the social movement phase and has become the interest of business and professional people. Professions which have as one of their tasks relating with people in the socio-emotional domain, that is helping others relate effectively in the social world with others and with their own emotions, are particularly interested in communication skills and human relation training. Nurses, doctors, clergy, social workers, counselors, and teachers all need the skill of communicating with people in order to optimally deliver their services.

With this in mind, many professional schools are implementing programs which instruct the pre-service professional in communication skills. Nurses training (Aiken, 1973), doctor-patient relationship courses, group counseling course work, and social work training all involve experiences in groups or direct communication skill training. These programs are being developed very rapidly, but they have put little if any emphasis on the evaluation of such programs. Granted, most of these programs are too new to begin a summative evaluation program, but there is a place for formative evaluation in these programs. Presently the most popular, if not the only, method of evaluation is the subjective

opinion of an experienced facilitator or communications "expert." But the reliability of such evaluations can certainly be questioned. Given the vested interest such experts have in these new programs, it opens the posibility for them to spuriously find success in their programs. This is pointed out by Lieberman et al. (1973), who note that encounter group theorists often allow their perceptions to become self-fulfilling prophecies. Objective formative evaluation of such programs is imperative if the programs are to improve. Objective formative evaluation will also aid later summative evaluation which must demonstrate the programs' success if such programs are to continue.

Group Work in Education

An evaluation technique for communication skills is urgently needed in education just as in the other professions. A teacher must communicate effectively with the students to discern their needs. Moreover, teachers have the added responsibility of facilitating effective communication between students, a responsibility that requires the teacher to have effective communication skills himself and also be able to teach these skills to others. This points out the importance of evaluation in the educational profession: first, the teachers of teachers must be evaluated, (i.e., are the instructors at the college level skilled in communicating with others?); second, the pre-service teachers should use effective communication skills in their college environment, (i.e., are preservice teachers demonstrating the appropriate

skills in their communication classes?); third, a teacher who has experienced a communication skills program in college should use those skills in the public schools, (i.e., does he or she model and instruct the pupils in those skills?)

With the need for communication skills programs in the public schools and colleges of teacher education, and the emphasis on accountability in education, one wonders what has caused the delay in initiating evaluation procedures in this area. One reasons for the delay is that the movement in education is still in the development stage. Socio-emotional education is only a few years old as an academic discipline, and the leaders in the field spent their energies on development, they want to finalize the product before they evaluate it for the public. This delay may also be influenced by the anti-research bias which Back (1972) notes, "In fact, an investigator's concern with assessment techniques is frequently taken as an expression of hostility" (p. 15).

In most school subjects the evaluation is done by standardized test or some other sort of paper and pencil test. The objective of the instruction is to teach knowledge or in some cases a particular skill which can be demonstrated by performing some task that is measured objectively through a paper and pencil test. While knowledge of group dynamics, which could be evaluated by means of a paper and pencil test, is important, the demonstration of the skill in a group situation is the true test of the instruction. Here lies the roadblock to the evaluation of communication

skills programs; no objective evaluation procedures exist to examine communication skills in group settings.

Presently group evaluation is done by an experienced observer or even by the facilitator himself by means of subjective report. Since traditional testing cannot measure group functioning, subjective evaluation was (is) seen as the only alternative by most people concerned with evaluation and research. But there is an alternative which has been used to examine classroom interaction. Since the first Handbook of Research on Teaching (Gage, 1963) literally hundreds of observation instruments have been developed, a large portion of which have been developed for teaching (Simon and Boyer, 1970). Some of these instruments have been developed for a particular subject matter such as math or foreign language but most simply examine general classroom interaction. These observation systems would seem to be a solution to objective evaluation of intensive small group experiences, but they have not been the complete cure all.

Recently, William Childers (1973) used one of these observation systems (Flanders, 1970) to examine the effects of the Georgia program (Gazda, 1973) on student teacher behavior. He found few significant results and in his recommendations suggests that a new instrument be developed, "A more sensitive instrument should be developed that will more directly reflect differences in communication styles" (Childers, 1973, p. 72). These results are not surprising in light of the fact that Flander's system examines pupil-teacher interaction at a very

general level (e.g., lecturing, giving directions, or asking questions). These general behaviors may not change as a result of more effective communication skills. The teacher may ask different types of questions after a human relations program, but the number of questions may be the same; therefore, the Flanders system would not discriminate those differences. Observation systems are a promising answer to the question of objective evaluation; but now that interpersonal communications is a subject matter in and of itself, it seems necessary to develop an observation system that is directed at examining specific types of communication skills. Flanders (1970) noted that in developing an observation system, the first requirement is that you know what you're looking for: an observation system is like walking in the park; if you're looking for birds, you won't notice the rabbits. Of all the hundreds of observation instruments (Simon and Boyer, 1970), none list interpersonal communication skills as its focus of interest. Observation systems may well be one answer to objective evaluation and research, but to look for rabbits we cannot expect to find results using binoculars that see only birds. There is a specific need for an observation instrument whose objective is the examination of specific interpersonal communication skills.

Definitions, Deliminations, and Terms

Before moving into an examination of observation instruments and interaction analysis, definitions of terms that delimit

the area of interest that will be examined in this research will be presented. Three general areas in which there are terms which have been defined in a multitude of ways (or undefined) will be examined. These are: affective education, group work, and observation instruments.

I. Affective Education

The term affective education is in some ways a misleading term. For the proposes of this study affective education will refer to that position which encourages students to talk about how they feel (affective awareness) and understand their own feelings and the feelings of those around them (affective and cognitive awareness). Although named affective education, this field also deals with the cognitive domain in the students must have a cognitive understanding of the process used in dealing with those feelings effectively. In the area of affective education, there are four terms or concepts which are commonly used but which may be so general (and misused) as to be confusing to the reader.

A. Cognitive. Cognitive refers to knowledge or facts. It is usually used in regard to what a person knows about a situation, person, incident, or body of information. Most teacher preparation is aimed at the cognitive domain, and prepares teachers to instruct students in a body of knowledge (facts). These facts cover a wide spectrum of information from historical dates to math skills. Knowledge of the specific criteria which must be included

in particular communication skills would also constitute a cognitive area, i.e., knowledge about feelings (e.g., understanding group dynamics) is cognitive even though the content is feelings (affective).

B. Affective. Affective refers to feelings or emotions. It is usually used in regard to how a person feels about a situation, person, incident, or body of information. Very little time is spent in teacher education programs to prepare teachers to deal with the affective concerns found in public school classroom.

C. Socio-Emotional Domain. Man's needs, as well as the subject matter taught in school, can be broken down into three interrelated areas or domains: intellectual domain, physical (or psycho-motor) domain, and socio-emotional domain. The first two domains are relatively well defined and researched and have received most of the attention of the educational community (e.g., the three "R's", hot lunch programs, school nurses, physical coordination, etc.). The socio-emotional domain deals with the social interaction of people. This involves the cognitive knowledge of how to communicate effectively and get along with others in social situations and also the affective component of how one feels about himself or herself and those around him/her. A major aspect of the socio-emotional domain involves handling emotions constructively in social settings.

E. Interpersonal Communication Skills. Interpersonal communication skills constitute a subset of the socio-emotional domain and would probably be classified as affective education by

many people. With the variety of programs being started, many terms are used to label a few general skills. Three skills that are included in nearly all the programs are active listening, self-disclosure, and feedback. These skills encompass nearly all the specific skills generated for any program (Allen, 1968; Becvar, 1974; Gazda, 1973; Gordon, 1970; and Lopis, 1974).

II. Group Work

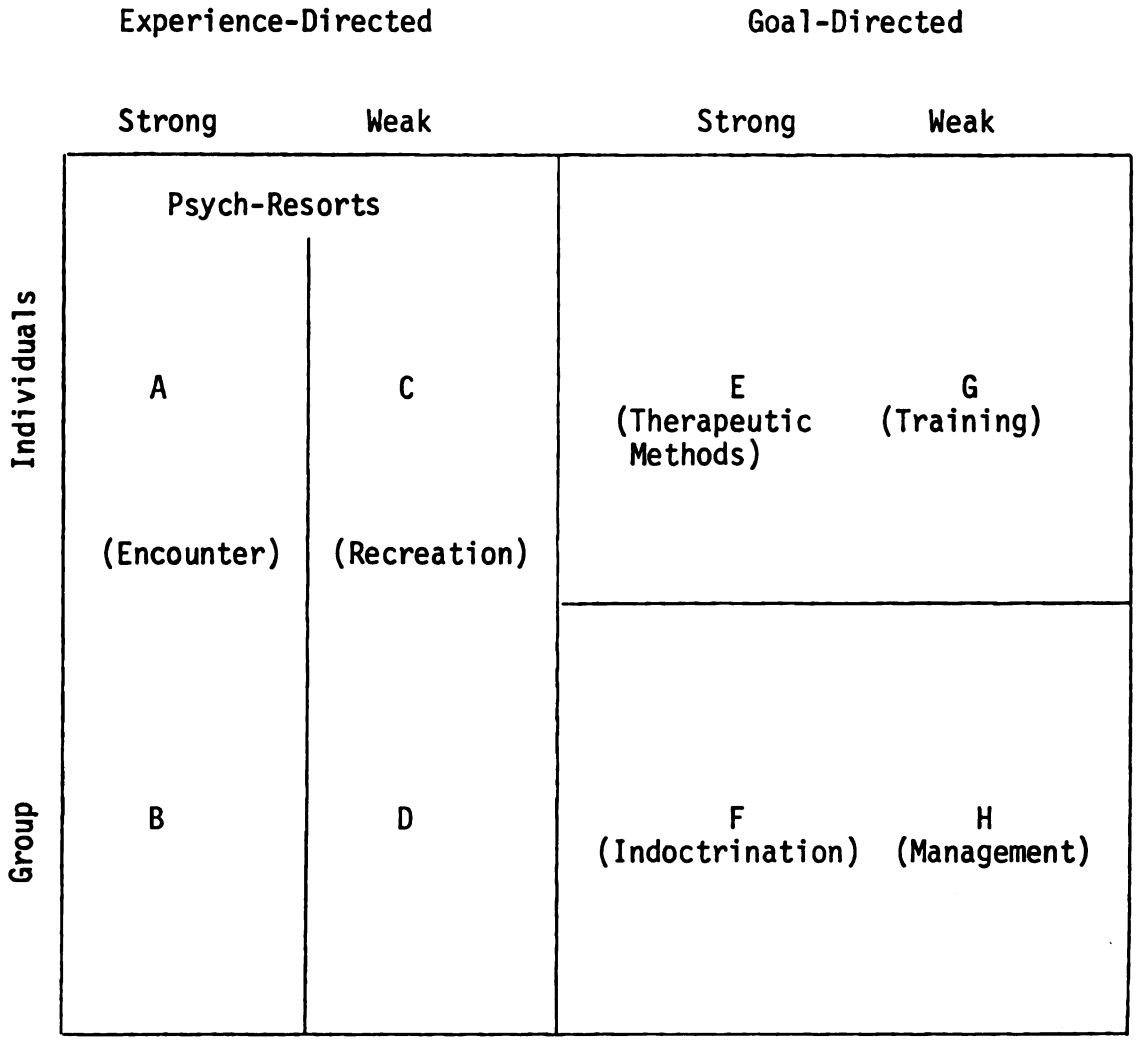
The area of group work, or the intensive group experience, is very difficult to define because the terms do not have precise meanings that are accepted by even the experts in the field. One of the major difficulties in understanding the group work field is the distinction between the various sectors of the movement. Since the present study is concerned almost exclusively with group work in education, it is important to delineate exactly where group work and education fit together. Back (1972) addresses this problem in Beyond Words:

> In the early period of sensitivity training, however, the idea of making changes through a group experience multiplied in education as well as in the medical and social work fields where education was needed. This philosophy has permeated the whole group-work field to the extent that sensitivity training has become confused with all of group work. . . . The basis of sensitivity training still remains the strong experience, the subjective feeling of change, while group work is generally much more goal oriented and wary of strong emotions (p. 176).

To understand the group movement, Back (1972) has constructed a map to locate and distinguish the various types of groups using three dimensions: (1) experience for itself or goal directed (2) strong or weak emotional impact on members, (3) the

Tentative Assignments of Experiences to the Scheme

(Back, 1972)



The accompanying chart shows the eight possibilities of sensitivity training according to this scheme. They are labeled A through H. . . . The boundary lines of the field of sensitivity training are not very definite and are continually shifting, especially as long as sensitivity training is expanding, for instance in the field of therapy (p. 122).

individual versus group orientation. Using these three dimensions, Back constructs a matrix (cube, 2 x 2 x 2) with eight areas which correspond to the various approaches of the group movenent. The first dimension (experience for itself or goal directed) examines the purpose for group membership, is the experience itself the goal (i.e., self-expression) or does the experience have a goal beyond participation specifically, is the purpose of group membership change? The second dimension considers the emotional impact on the group members. The distinction is not clear cut but the extremes point out the difference: a weekend marathon at Esalen aimed at sensory awareness arouses entirely different emotions than an afternoon workshop for businessmen on how to get along better with employees. The third dimension examines whether the group emphasis is on individual growth or group development, is the purpose of the group to better the individuals involved or the group as a whole?

The group movement involves many approaches and styles which all fit into Back's matrix someplace depending upon how they meet the three criteria, but two types of groups in particular need further clarification.

A. Sensitivity, Encounter, T-Groups. Although the proponents of each of these movements would no doubt object to grouping them together, these three particular types of groups all occupy the same cell in Back's matrix, experience-directed groups with a strong emotional impact that can have either a group or individual orientation (cells A & B). These groups have no definite aim beyond

encouraging people to understand themselves and encourage strong emotions and feelings. This is the area of the group movement where the Bob-and-Carol-and-Ted-and-Alice fantasy comes closest to becoming a reality. In the present study, this group will be referred to as the encounter group or movenent.

B. Educational Process Group. This term is used to label those groups which have a definite aim or goal (i.e., there are objectives that extend beyond the time and space of the group) and also have individual development as a priority over group development. Their emotional impact would depend upon the particular group although the experience would rarely have the emotional impact of an encounter group. The most critical difference between these groups and others in the movement is their strong emphasis on specific objectives and change within individuals. The most frequently stated objectives for these groups is the acquisition of particular communication skills. These groups generally fall in cell G, although some may be close to cell E.

C. Interpersonal Process Laboratory. This study will explore one particular type of group, the interpersonal process laboratory (IPL), which exemplifies the educational process group. This group consists of approximately 15 students and one instructor (facilitator) meeting two hours twice a week for ten weeks. This group is part of a course (Education 200: The Individual and the School) at Michigan State University which is a required course for education majors. The objectives for this course appear in Appendix D.

III. Observation Techniques

In Medley and Mitzel's (1963) discussion of systematic observation, they refer to the term observational technique as ". . . procedures which use systematic observations of classroom behavior to obtain reliable and valid measurements of differences in the typical behaviors which occur in different classrooms, or in different situations in the same classroom" (p. 250). This covers a broad range of observational techniques (or systems) of both the category and sign type. A category system includes a number of specifically defined categories into which all observable behavior which is of interest fall and also includes the number classifiable in each category. Therefore, if verbal behavior is the area of interest, all verbal statements theoretically fall into one and only one category and the instrument specifies how often they occur in a given segment of time. A sign system on the other hand specifies what behaviors an observer is to watch for and records only those behaviors. The two systems differ in that the category system is theoretically exhaustive of behaviors of the type recorded.

A. Interaction Analysis. Interaction analysis is a specific type of category system devised by Flanders (1960) which studies the chain of classroom events in such a way as to take into account each recorded event in sequence with every other recorded event. This is done by recording, in sequence, each event according to a specific category definition and then transferring the list of code symbols, one symbol to one event, onto a matrix which shows graphically the

relationship between the events, i.e., what precedes and follows the individual events (see Flanders, pages 54-75, 1970). Although this system does not specify what precedes and follows each specific individual event once the codes are tabulated on the matrix they do portray the probability of each category being followed or preceded by every other category. This increases the amount of information retrievable from the data by going beyond a simple frequency count by adding the dimension of time.

B. Interpersonal Skills Interaction Analysis (ISIA). The ISIA is the title of an interaction analysis technique derived from Flanders' Interaction Analysis Categories (FIAC), (Flanders, 1970) and Ober's Reciprocal Category System (RCS), (Ober, et al., 1970) and from programs in communication skills (e.g., Lopis, 1975). It is a multiple cluster category system that examines interpersonal communication skills in groups whose goals are specifically related to those skills. The clusters and categories are listed in Appendix A.

Procedure

The development of an observation instrument involves a number of steps. The first task in developing an observation instrument is to set broad limits on the types of behaviors to be investigated. In the present study this broad area of interest is interpersonal communication skills. The investigator must then review the literature to discover what instruments have already been developed in the field. This is followed by the specification

of the type of observation procedure to be used (e.g., sign system, category system) and the categories to be included in the observation instrument. After the type of system and categories are decided upon, operational definitions must be written for each of the categories. This is perhaps the most important step as reliability and validity rest on the extent to which the categories are behaviorally defined and mutually exclusive. After the categories are defined the appropriate population must be delimited and pilot testing of the instrument must be carried out. Following pilot testing any needed modifications of the instrument must be included in the preparation of a training manual for observers. The training manual must include category definitions, ground rules for using the instrument and some type of practice exercises for observers.

Once the instrument training manual is prepared the instrument is ready for field testing. Field testing involves specifying the procedures for the data collection, verifying the reliability of the instrument (including the training of observers) and showing that the instrument is valid for the stated population. These steps will be used to develop the ISIA.

Population

The populations to which the ISIA may be applied encompasses a great variety of environments but is restricted by the goals of the groups. The ISIA may be used to examine the communication skills of people (both adult and children) involved in groups whose goal is the development of more effective communication

skills. Many programs have been developed recently which focus on these skills and the ISIA may be an appropriate tool for evaluating and researching these groups. Public school programs, (e.g., DUSO and FOCUS), parental training programs (e.g., Parent Effectiveness Training), teacher education programs, and any other academic programs whose goal is more effective communication are all possible populations in which the ISIA may be used. The population to be examined in the present study will be those individuals enrolled in an introductory course in education, The Individual and The School (ED200), at Michigan State University during the Summer, and Spring terms, 1974-75. A major segment of this course consists of participation in interpersonal process laboratories (IPL) whose objectives are: self-disclosure, active listening, questioning, observation, and feedback skills (see IPL above in definitions, delimitations, and terms).

Data Collection

The actual data used in developing an interaction analysis technique is the sequence of codes recorded by trained observers. This data may be obtained in a number of ways each of which has advantages and disadvantages related to how removed the coding is from the actual group interaction. The least removed method of collecting the data would involve actual in-class coding by trained observers. The advantage to this method of data collection is that the observer is exposed to all the verbal and non-verbal stimulae to aid in the coding. Due to the subtle nature of

interpersonal communication all these cues may be important and must be explored. But because of the possible effects the observer may have on the group (i.e., changes in behavior due to the presence of an observer) and time limitations of the observers, in-class observation of all groups would be inefficient. For the present study, in-class observation would be examined only as a check for the reliability of other data collection procedures. That is to say that some in-class observation would be carried out and then compared to other methods of data collection with the aid of exploring the possible loss of data by other observation methods. A second method of data collection would involve the use of videotape equipment to collect the group interactions. The advantage of using this method would be the recording of both verbal and non-verbal interactions, but the cost and possible interruption due to the recording equipment prohibit the collection of data by video-recordings. The use of tape recording is the third data collection method. Flanders' (1971) work and pilot testing by the author indicate this method to be both efficient and reliable. The majority of the data collected for the present study will be done by audio-tape recordings, as this procedure involves a minimal amount of group disruption while still retaining all the verbal interaction.

The data for the present study was collected by tape recording during the Summer Term, 1974, and Spring Term, 1975. To examine the validity of the instrument, student and instructor opinionnaire data were collected during Summer Term, 1974.

Reliability

The establishment of the reliability of an observation schedule is perhaps the most crucial element in developing such an instrument. Two types of reliability will be analyzed in the present study: the coefficient of observer agreement and the reliability of in-class observations versus tape recorded observations.

The coefficient of observer agreement (Medley and Metzel, 1963) is the amount of inter-rater agreement and is defined as the correlation between scores based on observations made by different observers at the same time. This is the most common form of reliability when examining an observation instrument. For the present study, Scott's π (Scott, 1955) will be used to estimate inter-rater agreement. This method of estimating reliability can be interpreted as the extent to which the coding reliability exceeds chance. Research on interaction analysis has used this method of estimating reliability (Amidon and Hough, 1967).

In-class observation versus tape recorded observation will examine the possible decrease in reliability due to the loss of non-verbal cues through the use of tape recordings. This will be done by having an observer code a live group session at the same time as the group is being tape recorded. At a later date the same observer will recode the tape recording and the reliability of the in-class versus tape recorded data will be estimated using Scott's π. Because of the verbal nature of the ISIA, it is felt that the loss will be minimal.

Validity

Validity measures of observation instruments are difficult to define and are seldom addressed in the literature. Medley and Metzel (1963) refer to the validity of an observation as the extent to which the observation data reflect actual differences in behavior as opposed to different impressions by different observers. Most developers of observation instruments seem to assume that a high degree of observer agreement demonstrates the similarity of impressions by observers which, in turn indicates actual differences in behavior. This may be evidence of direct or primary validity (Ebel, 1972), but it would seem important to examine the derived or secondary validity including the correlations of the observed behavior to "actual" behavior or secondary measures of that behavior.

In the study of the intensive group experience, no previous objective measures of the participants behavior exists; that is the purpose of developing the ISIA. This being the case, the researcher must look to a less reliable but useful subjective evaluation, the opinions of the participants. The IPL Evaluation (Appendix B) is an instrument developed to systematically collect the opinions of the group members as they relate to the skills measured by the ISIA. During the Summer Term, 1974, this instrument was administered to all group members immediately following the taped group meetings. The opinions of group members rating their own interactions is subjective and can be very biased. Because of this an expert in interpersonal communications

listened to the sessions and rated each session on the same opinionnaire (Appendix B). This data was used to validate the group member's opinionnaire data.

The opinionnaire data are used to choose extreme groups (effective versus ineffective; see questions #8 and #20, Appendix B) which are compared to examine the primary validity of the instrument. If the instrument is, in fact, sensitive to the communication skills it purports to measure, it should be able to distinguish between those groups seen as effective by the members as opposed to those groups viewed as ineffective. The opinionnaire data will also be used to verify the quantity of particular skills used in the groups. For example, questions #7 and #19 refer to the amount of active listening demonstrated in the group: do groups who differ significantly on those questions show a difference in the ISIA categories which represent active listening?

Questions to be Addressed

Intensive group work in education is expanding at a rapid rate. A great deal of research and evaluation is needed in this area in the near future to point out the strengths and weaknesses of group work in all areas of formal and informal education. Before any research or evaluation can be started, the tools of evaluation must be developed. This study will focus on the development of one tool which, if shown to be reliable and valid, will be valuable in examining programs whose primary concern is communication skills. The answers to the following questions may offer more systematic

objective evaluation and research in the field of interpersonal communication skills programs.

Are observation techniques, interaction analysis in particular, suitable tools for examining communication skills in intense group experiences?

Can observers be trained to code interpersonal communication skills reliably? Do audio-tape recordings disclose enough of the cues of communication to reliably code interpersonal interaction or must observation be done live, in-class?

Can an instrument, such as the ISIA, reflect the subjective judgments of the participants? Can such an instrument discriminate between groups judged effective and those judged ineffective? If so, what particular skills are evidenced in those effective groups? What behaviors occur less frequently? What patterns of behavior occur in such groups?

CHAPTER II

REVIEW OF RELEVANT AND RELATED RESEARCH

During the past quarter century, man has made great strides towards understanding the behavior of individuals in groups. During the same period, education has undergone great changes and innovation have become standard procedure. This chapter will explore the interaction of these two phenomena, specifically the impact of the sensitivity movement on the institution of education. An examination of the literature reveals numerous attempts at research and program evaluation in the area of sensitivity training, human relations training, affective education and other related fields. But one is astounded at the number of researchers and reviewers in the field who cry out the same old song: we need research on human relations training, we need instruments to measure the outcomes of groups, we need better methodology to study the treatment effects (what happens in those "black boxes" called training groups?), or group goals must be behaviorally stated, to mention a few of the verses. But, like the proverbial weather problem, it seems everyone is talking about it, but no one ever does anything about it; the lack of adequate tools to study small group interactions still impedes research and evaluation in education as well as in other fields. For the sake of brevity, this review will only

examine this problem as it relates to educational research and evaluation. But the problem is just as pressing in other disciplines.

Overview

By way of orientation, this chapter will begin with a general discussion of the role of group dynamics and interpersonal communication skills in education, relating the felt need for such programs, their intergration into the schools, and a brief overview of the types of programs being instituted at various levels of schooling. This will be followed by a discussion of the in-service human relations programs for teachers, and then a review of the research on the pre-service teacher training programs. The difficulties involved in researching the attitudes and behaviors of small group members will be reviewed including a discussion of the methods of observing the outcomes of group participation. The last section of this chapter will focus on one method of quantifying group behavior; interaction analysis as developed by Flanders will be explained particularly as it relates to the development of the Interpersonal Skills Interaction Analysis.

Introduction

> All of my professional life I have heard quotes of surveys which showed that 75-80% of human beings failed in the work-a-day world because they cannot relate effectively with other people; yet the major part of our educational effort is directed toward improving instruction--how to teach students more math earlier, interesting ways to present new and old facts . . . we have bigger and better reading programs, and we are producing so many non-readers that we are creating jobs in school after school for remedial reading teachers. (And I

> am happy to say that many of the remedial reading teachers that I know are finding that their best results occur when they set aside the textbook for awhile and relate to the child as a human being.) (Tatum, 1969)

With all the time and energy put into educational research on curriculum, learning, teacher education, etc., and all the changes that have been brought about in the classroom, a large proportion of our society is still disenchanted with the institution of education. The above quote by Tatum (1969) echoes a position that is a growing concern to many educators--the need for affective education, human relations, and communication skills in the classroom.

Recently there has been a great deal of controversy among the public, students, and professional educators concerning the direction education ought to be taking. For example, Ebel (1972) states ". . . it seems clear that the principal task of the school is to facilitate cognitive learning" (p. 33). But those educators of a more humanistic persuasion claim that the school's function is to bring about what Rogers (1969) called "significant" or "experiential learning." He defines this type of learning as having ". . . a quality of personal involvement--the whole person in both his feeling and cognitive aspects being in the learning event" (p. 5). As with any philosophical argument, no single fact can be brought to bear that will settle this issue. There is a great deal more to be said on both sides of this issue, but it seems clear that education is expanding beyond merely the facilitation of cognitive learning and all signs seem to indicate it will continue to do so despite a great deal of resistance.

Despite the resistance, the fields of affective education and communication skills are being incorporated into an ever increasing number of programs. As Reece and Passmore (1971) point out, education has emphasized knowing and doing for the past four decades, but feelings may be the primary focus of the seventies. In the years to come, society may mandate a more humanizing educational experience and part of that experience must include some instruction on relating to those around you. More than likely these experiences will include a human relations model similar to sensitivity training. Educators must be prepared to show the usefulness of such a program (what do the participants gain?) and also be aware of the skills teachers will need to facilitate such programs.

Although sensitivity training has always been connected with education (NTL is and always has been formally related to the National Education Association), the impact of the group movement was negligible up until the middle to late sixties. Even today much of what the public knows (teachers included) is based on rumor, subjective report, or sensationalism from the media. In an opinion poll in Nation's Schools (1970), half of superintendents interviewed seemed to be saying they'd suspend judgment on sensitivity training until they received more information. But twice as many felt such experiences had a positive effect than negative. Their uncertainty and concern related to the proficiency of the group leaders and the conflicting information on the effects of the groups. These concerns are legitimate whether a person is deciding on attending a

weekend marathon himself or deciding on implementing a human relations program in his school district. The competency of the trainer or facilitator is extremely important, particularly in a situation in which the participants are not volunteers such as the public schools. This will be explored in depth in the later discussion of pre-service and in-service training, but one must note the risk a superintendent or principal takes when s/he implements a human relations program in his/her district or school if his/her teachers have no training in group work. The issue of conflicting information is also a great concern even today as so little fact exists on the effects of groups other than the "fact" that most participants have a positive opinion of their group experience.

Although there is a shortage of objective knowledge concerning training groups, teachers have been exposed to group work through professional journal articles and various workshops run specifically for teachers. In 1970, Educational Leadership devoted an entire issue to "Sensitivity Education: Problems and Promise." In that article teachers, principals and other educators from various parts of the country shared information about their programs including the opinions of their students and staff relating to sensitivity training. The enthusiasm generated by the programs comes through in one typical statement, "It is hard to imagine anything more important at the present time than the improvement of human relations, and that is what successful sensitivity education furthers. Our material wealth is unbelievable, but we often seem to be in the Dark Ages in our human

relations" (Corey, 1970, p. 238). Other teacher education publications such as Childhood Education (Lippit, 1970, and Trubowitz, 1975) and Scholastic Teacher (Harrison, 1971) show this same enthusiasm.

Summer workshops for teachers are offered at colleges and NTL (at Bethel) as well as programs which may be contracted by school districts. One contract program called Talent Awareness Training (Sponberg, 1969) which holds workshops mostly in the Rocky Mountain states had already reached 20,000 elementary teachers as of 1969. These programs, designed to introduce teachers to sensitivity education, are numerous and the numbers are growing. This increase is also true in the types of programs which are being developed and used in the public schools.

Group Work in Education

The growing use of sensitivity training in business, industry, religion, and as recreation has been phenomenal in the past fifteen years. The areas of education in which groups are being used are equally diverse. From nursery school through graduate school, from nursing homes (Diekman, 1972) to campus police (Abramson, 1973), people who meet in groups are finding uses for sensitivity training and communication skills training.

Public School Programs

The most wide spread reported use of group work in education is in pre-service and in-service teacher training programs, but the implementation of group work is by no means limited to teacher training. Many programs have been developed at all levels of

education both public and private. A number of these programs have been developed to simply open up the communication or discussion aspect of a particular course, others have been used to develop related skills (e.g., speech and theatre), while others have been directed at particular problems of the school or students (e.g., racial tension or drug problems). The applications of the sensitivity training experiences have also varied; a number of the programs consist of one teacher reporting his personal application in his classrooms while other programs are formal decisions at a school district to implement a city-wide program. One similarity among all these programs is the use of a subjective evaluation technique, if, in fact, any evaluation is performed.

Elementary School Programs

> There is no lower or upper age limit for some form of sensitivity education. Children in nursery schools have been helped to become more aware of the effects of their interpersonal behavior on other children and on themselves. They can be helpful to keep in closer touch with the way they feel about and perceive what other people do to them and what they do to other people (Corey, 1970, p. 240).

An intensive group experience can be very involved and sometimes upsetting to the participants, and for this reason many people might feel that young children should not be exposed to it. But applications of sensitivity training have been used in nursery school (Human Development Program, HDP, Bessell, 1968) and elementary schools (MacDougal, 1973) and curricula have been developed in human relations for elementary school children (Van Camp, 1973). In fact, Dinkmeyer (1970, 1972) has developed a packaged program which is used in elementary schools across the country. Developing

Understanding of Self and Others (DUSO) (Dinkmeyer, 1970) is a human relations program which includes tapes, pictures, teachers' guide, and other materials for kindergarten through sixth grade. The objectives of the program are similar to those for groups of older participants and include listening skills, self-disclosure, and value clarification.

DUSO, HDP, FOCUS and programs that include techniques such as Glasser's classroom meetings (Glasser, 1969 and O'Donnell and Maxwell, 1971) all derive a portion of their practices from sensitivity training. These programs are being implemented in an ever increasing number of schools without any systematic evaluation or research of the outcomes. This causes concern for educators and the public because the answers to the questions; what are my children being exposed to? and what can I expect my child to learn from these groups? are still being answered very subjectively, if any attempt is made to answer them at all.

High School Programs. The programs in high school have been used to augment the normal classrooms as they have in elementary schools, but group work has taken on the added dimension of facilitating particular problems of schools or districts such as drug problems or racial problems. In regular classrooms, group work has been used in speech (Heiman, 1974; Galvin, 1974) and English courses (Harrison, 1971; Simon and Sarkotich, 1967). There are also examples of schools for drop-outs which use human relations training and one integral part of their program (Caine and Lindenaver, 1973).

There are classes aimed more specifically at human relations and communications skills (Wells, 1970) and some governmental support for high school programs which focus on socio-emotional growth (Springport High School, 1967). Specific problem areas in schools are a new curricular development with interracial relations (Curkhuff and Banks, 1970; Price, 1969) and drug education (Deardon and Jekel, 1971; Southern Regional Education Board, 1974) receiving the most attention.

Human relations training and sensitivity education has been used in a number of settings in high schools across the country. Many personal accounts appear in professional teacher journals which indicate the variety of uses for group work. But other than questionnaire data from the participants or subjective observations of the teacher, very little has been done in the way of evaluation. High school programs exist but at this time no definitive statement can be made in relation to their effectiveness.

College Programs

Outside of teacher education, a number of college related studies have been reported which deal with sensitivity training and human relations programs. The first course to implement sensitivity training occurred at Harvard in the early 1950's (Mann, 1967). These first studies generated a number of research reports (Bales, 1950; Hore, 1973) that laid the groundwork for later research. Today the vast majority of college and universities have some type of intensive group experience available, some of them similar to

the program at Antioch College (Solomon, et al., 1970) which involves a cross disciplinary approach including social psychology, drama, and speech. At Antioch, as with many college programs, the evaluation was somewhat systematic but the first paragraph of summary of research findings forewarns the reader of the problems to be expected:

> Our research measures and findings have been limited in their applicability and usefulness. This is partly the fault of our practice and partly due to the lack of valid and reliable testing instruments (p. 59).

Other areas in which sensitivity training have been used in college environments include training programs for counselors (Schroeder, et al., 1973; Perkins and Atkinson, 1973; Dendy, 1971) and counseling of patients (Arbes and Hubbel, 1973). Studies such as those have used global ratings of empathy and understanding as their dependant measure (e.g., Empathetic Understanding [EU], Carkhuff, 1969a) or they have used self-report or attitudinal changes as their measure. The choice of these subjective measurements resulted from the lack of established objective instruments. These difficulties will also be noted in carefully planned in-service and pre-service teacher education programs. The lack of adequate measurement tools has impeded the needed evaluation and research on the outcomes of group participation.

In-Service Teacher Training. In-service teacher education programs which involve sensitivity training techniques come in all shapes and sizes. From an uncontrolled study of three small Manitoba (Canada) High Schools (Benmen and Capelle, 1971) to an in-depth controlled study of the teaching-learning process, done

in conjunction with a major university (Bowers and Sour, 1961), many school districts are using group training techniques to improve their staff relations and their teachers in-class communication skills. Perhaps the most interesting aspect to examine in reviewing these studies is the range of subjectivity in the evaluation techniques of the programs.

The Buffalo (New York) Board of Education's final evaluation of their Human Relations Education Project (1970) will give the reader some insight into the unspecific nature of many of the evaluation reports (and perhaps the programs). To begin with, the report offers no definition of human relations as implemented in their program (one is unsure if the program involves racial relations, communication skills or some other objective), and the reader is further confused by "the specifically stated objectives" which are, "Teachers will assert increased awareness of the importance of human relations in their own lives and the lives of their students." These types of program objectives are not uncommon, as the programs of both West Virginia (Forman, 1968) and Tennessee (Khana, 1969) have similar objectives which focus on the "awareness of the need for human relations." The unspecific nature of many of the programs may be a result of the infancy of the field. But, if sensitivity education and human relations programs are to improve, formative evaluation must be undertaken and that must be attempted based on the goals of the programs (i.e., specific objectives).

Sensitivity training has been used in a variety of educational settings (e.g., junior college staff, Keile and Gallessieh,

1971), but the majority of the reported programs have involved principals and/or teachers. A number of these programs report no evaluation (Hendrickson, 1968 and Kimple, 1968, 1969, 1970) or a limited questionnaire evaluation (McElvaney, et al., 1967), but an interesting number of programs are including formal albeit subjective evaluation programs.

Before examining the evaluation programs, it may be important to make a more definitive statement concerning the subjectiveness of evaluation. The author uses three criteria in assessing the subjective nature of a study. The first criterion to apply involves the subjective nature of the data: is the reported data personal opinion or fact based on systematic behavior observation? A large portion of the studies already cited include opinion data (e.g., the teacher noticed that the students got along better) and are questionable because of the probable biased perception of the reporter (i.e., s/he sees what s/he expects and wants to see). The data must also be considered subjective whenever the data are of a self-reported nature. This is particularly important with volunteer participants in sensitivity training, as they may have expected to gain from the experience and, therefore, perceived the gains they expected.

The second criterion in assessing studies is the subjective nature of the methodology. Campbell and Stanley (1963) refer to the experiment ". . . as the only means for settling disputes regarding educational practice . . ." (p. 2) and list various factors jeopardizing internal and external validity. Although it may be too

critical at this time to judge sensitivity training research by all the standards proposed by Campbell and Stanley, some of the criteria are directly relevant to a great deal of the research being conducted in this area. Many studies of sensitivity training do not employ a control group (Harrison, 1971 and Diamond and Shapiro, 1973), therefore eliminating the possibility of any comparison (would they have changed if they had received no treatment or a placebo treatment?). A second area of concern is the differential selection of subjects or lack of random assignment to control and treatment groups. As noted previously, this is very important when using volunteer participants. The third methodological criterion is related to the measurement of the data; is the instrument being used reliable and valid? This is of particular concern when the instrument is made expressly for the study and no data on reliability or validity are reported.

The third criterion in assessing the subjectiveness of the evaluation relates to the type of research or evaluation being performed. Dunkin and Biddle (1974) point out the four possible variables in educational research; pressage variables (formative experiences, training experiences and personality characteristics), context variables (conditions to which the teacher must adjust, i.e. environment), process variables (the actual activities of classroom teaching, what teachers and pupils do), and product variables (the outcomes of teaching, the changes that come about in pupils). These four variables can be combined to examine a number of cause-effect relationships. In relation to pressage, process, and product

variables sensitivity training has been lacking because researchers have not examined the behaviors being demonstrated in groups nor the behavioral outcomes of the groups. In training teachers in communication skills these three variables are closely related and can all be measured in the same way. The pressage variable involves the teacher's training experiences and should be measured by examining the teacher's behavior in training groups. The process variable includes teacher's and student's in-class behavior and must be evaluated by measuring the communications being used in the classroom. The product variables are the changes (or lack of changes) in student's communication skills and must be behavioral measures of student's behavior. To effectively compare these three variables a researcher must be able to measure in-class behavior, whether that class is in a college, in-service teacher training, or elementary classroom. This is where most research on sensitivity groups is lacking; they do not examine in-class behaviors. Dunkin and Biddle (1974) point out the problem as related to teacher effectiveness research:

> Perhaps the most significant shortcoming of these early studies is that they assiduously avoided looking at the actual processes of teaching in the classroom . . . if teachers do vary in their effectiveness, then it must be because they vary in the behaviors they exhibit in the classroom. To shed light on this point, one must study classrooms--where the action actually is (p. 13).

This problem is as prevalent in sensitivity research as it was (is) in teacher effectiveness research. To use a pressage variable (sensitivity training) as an independent variable (in many cases an undefined variable) and then expect a significant

change in a product variable (student opinion or behavior) requires an extremely powerful treatment and an equally sensitive measurement instrument. When one considers that most sensitivity research is basically exploratory and most of the measurements are crude by almost any psychometric standard, non-significant results should be expected. The cause-effect relationship is strained by pressage-product research and without the experimental controls called for by Campbell and Stanley (1963) the research results begin to look quite subjective.

In teacher effectiveness studies the results of pressage-process research has not revealed training experiences to have as great an impact on teacher in-class behavior as might be expected and process-product research, when it is undertaken, is equally discouraging (Dunkin and Biddle, 1974). But to make the jump from pressage variables to product variable with very little experimental control is mostly a subjective leap of faith. This will be explored more fully in the section on research, for now the studies will reveal many of the shortcomings.

A doctoral dissertation by Bailey (1967) clearly points out the pressage-product difficulty. He studied the effects of sensitivity training upon a high school faculty using student perceptions as measured by the Student-Opinion Questionnaire as the dependent variable. The design of the study controlled for most sources of invalidity, as it followed Campbell and Stanley's (1963) "Non-equivalent Control Group Design" and included two post-tests, one approximately one month following the sensitivity training and a

second post-test three months after training. The main thesis of the study was, "If change in teachers is observable by and has an effect upon the students, then change should be recorded by students. The students are the product of the educational process and should be one of the eventual criteria for evaluating in-service programs" (p. 9). The Student-Opinion Questionnaire had ten objectively scored items and the four hypotheses were based on the data from these items. All hypotheses were found not to be significant. All items on both administrations of the instrument were investigated for differences between the experimental group and the control group. Of the twenty comparisons, one significant difference was found, "ability of teacher to explain clearly," on the first post-test. In explaining the non-significance, the author stated four possible reasons: (1) there was no change as a result of sensitivity training, (2) the laboratory was not long enough to bring about change, (3) the students were unable to perceive change if it did occur, or (4) the instrument was not sensitive enough to the change if it was perceived by the students. As a final word the author said, "If one accepts other research that has demonstrated positive changes as a result of sensitivity training and the positive reactions by the teachers following this laboratory, it may be suggested that the explanation for the lack of significant differences may be related to the instrument" (p. 108). The questions raised by this study do not relate to the subjectivity of the data (the reliability and validity of the instrument are substandiated) or the methodology, but rather the type of research

involved. Bailey offered four explanations for the non-significance of the results but because of the type of research he chose it is not possible to say which explanation is the most plausible. The study is about behavior change in a faculty as a result of human relations training, but there is no measure of any behavioral changes. The in-class process is missing. We have no measure of the behavior of the individuals while in the sensitivity training group, no measure of those individuals' subsequent behavior in the classroom, and yet researchers want to know if those hypothesized changes effect students who are supposed to observe those hypothesized changes. Sensitivity training research is in its infancy and must be measured one step at a time. More sensitive tools are important, but an equally important question is--sensitive to measure what?

Pressage-product research is used frequently with in-service training programs. Some programs using sensitivity training report significant findings; others seem to rationalize their non-significant findings away. Schmuck (1967, 1968) found significant and positive changes in students' perceptions of classroom groups, their own status and influence, attitudes, and friendship patterns. Nelson (1969) found no significant results and points out the distance between training and the student product, "The tests of student anxiety, alienation and opinion surveys are perhaps not germain to an assessment of the kinds of changes human relations training can effect in a short term project" (p. 31).

In various types of sensitivity training from business to education, a common dependent measure is ratings by peers, co-workers or supervisors of perceived changes in behavior following a sensitivity training experience. In two studies at NTL, Miles (1960, 1965) found perceived change as reported by participants and job associates in relation to "listens more," "communicates better," and "shares decisions," for elementary school principals. Krafft (1967) studied the changes in behavior, due to a human relations laboratory, of secondary school seminar instructors. He found no instrument to measure their behavior and chose instead to measure their behavior by the perceptions of the participant himself, a co-worker, and the principal of each subject. He collected the data by interview, but had difficulty because the subjects, co-workers and principals knew the identity of the experimental and control group. The experimental subjects knew what behavioral changes the interviewer was interested in and the principals talked almost exclusively about the experimental subjects. This points out the difficulty in perceptual data; the subjects, and those they frequently come in contact with are sensitized to the desired changes. They expect changes and their perceptions may simply be revealing those expectations. Data which are based on behaviors and not perceptions will necessarily be more objective and valid.

Values and attitudes is another frequently used dependent variable to measure changes in teachers due to a human relations experience. Benmen and Capelle (1971) found high school teachers to improve their self-actualization, attitudes toward educational

process, and values of inclusion and affection as measured by the Personal Orientation Inventory (POI) (Shostrom, 1964), Educational Process Opinionnaire (EPO) (Wehling and Charters, 1969), and Fundamental Interpersonal Relations Orientation (FIRO-B) (Schutz, 1958) respectively. Lee (1967), in a study of the effectiveness of a human relations training program for in-service teacher training, found that teachers' attitudes (towards pupils in interpersonal relations and teaching as a vocation) increased as measured by the Minnesota Teacher Attitude Inventory (MTAI) as a result of sensitivity training. In another in-service training program undertaken to examine teacher attitude change as a result of sensitivity training, Sweeney (1969) found elementary and secondary school teachers to score significantly higher on the MTAI after training than did a control group. Two statements by Sweeney (1969) point out the possible misapplication of research efforts in the area of sensitivity education:

> Teacher-pupil attitudes are simply indicators of the teacher's classroom behavior and the mere introduction of better attitudes by instruction may not produce any change in behavior (p. 4).

But then a few pages later he seems to contradict himself:

> The study focus was on teacher attitudes. What is needed, among many other possible approaches, is a focus on pupil perception of the teacher prior to T-group sensitivity training and then after the experience. It may be that the learning which the teacher experiences, the insight, the awareness, etc., may not always be brought out from the group experience to the classroom (p. 7).

The application of learning, which occurs in a sensitivity experience, to the classroom is the goal of in-service group experiences. If teachers cannot apply what they learn in an in-service workshop,

the workshop has failed them. The learning should involve changes in behavior and, if the researcher wishes to examine the changes in teachers' in-class behavior, he would do well to measure that behavior, not the student, co-worker, or principal's perception of that behavior. Contradictory statements such as those by Sweeney (1969) are disturbing: perhaps the absence of adequate instruments to measure behavior encourages researchers to examine non-process variables.

Pre-Service Teacher Training

> The new trends in education have implications for present and future teachers, as well as teacher educators. Teachers must be trained or assisted to assume their new roles comfortably and effectively. They need to be oriented toward working more with smaller groups and individuals; they must be trained in the skills needed to function within this orientation. . . ." (Crist, 1972, p. 73).

Since the first college program at Harvard (Mann, 1967) in the early sixties, the use of sensitivity training on the college campus has expanded enormously, particularly in colleges of teacher education. In this review alone the programs of research of approximately twenty institutions will be cited and one would suspect that for every reported university or college program, numerous programs exist which have no published results. The number of programs is substantial and multiplying every year because of the intensified interest in their use for the personal demands of teachers and for the facilitation of elementary and secondary school programs which include human relations and communication skills. The new trends in education that Crist (1972) speaks of exist in every type of school environment and at every

age level, but teachers do not have all the necessary skills to implement these programs. A school district can purchase a DUSO kit for every classroom, but, unless the teachers are prepared to use those materials, they will be wasted, either because the teacher passively resists using them because she has never been introduced to them (a quite common practice) or because she attempts to teach a subject matter (human relations) she is unfamiliar with and fails. "New math" was (is) not as successful as it could have been because teachers were not prepared to use it in their classrooms despite the fact that almost every teacher had been instructed in some form of math education course. Human relations is not a new version of what teachers are already teaching; it is a new curriculum that some teachers have never heard of, much less taught. In-service training is one way of introducing the in-class teacher to human relations, but, if these programs are to be implemented successfully, future teachers must be instructed in the knowledge and techniques needed so they will be comfortable and effective in their own classrooms. This section will begin with an overview of the programs in pre-service teacher training, noting some of the program evaluation being conducted. That will be followed by a more in-depth look at four particular programs, Minnesota, University of Massachusetts, Carkhuff's human relation training, and the program at Michigan State University.

In an article entitled "Sensitivity Training: Solution or Conspiracy?" Wiggins (1970) examines the benefits of sensitivity training and some of the deficiencies in school programs. He notes

that NTL reports at least eight graduate programs which include sensitivity training and that programs at every educational level are investing money and time in programs and training. However, many of these programs have encountered troubles: unclear or non-existent objectives, poorly trained personnel, the lack of research and evaluation to establish the programs as beneficial to the participants. All these difficulties have brought sensitivity training to the point where Wiggins (1970) feels that the role of training in education must be reevaluated. He suggests that the status of sensitivity training in schools would improve if: (1) the term sensitivity training were replaced with human relations training, (2) standards for trainers were developed and enforced, (3) "Human relations training were used only when clearly defined goals and behaviorally defined objectives are established," (4) "Research could be conducted to provide empirical evidence as guide posts to direct application of human relations training," and (5) "Evaluation models to assess the results of training programs could be developed" (p. 257). These suggestions can be used to examine some of the human relations programs that schools of teacher education offer.

Human relations training is essentially a subjective experience. Researchers of sensitivity training have consistently encountered difficulty in describing or having others describe such an experience (Lieberman, et al., 1973), and one seldom finds a group experience which will describe its goals any more clearly than the goals at NTL: (1) self insight, (2) better

understanding of other persons and awareness of one's impact on them, and (3) better understanding of group process (Sweeney, 1969). Considering this type of subjectivity, it is not surprising to find that most human relations programs have a subjective goal such as increased awareness or no stated goal at all. The Syracuse University Model Elementary Teachers' Education Program (Benjamin, et al., 1968) had as its goal for a teacher the increased awareness of and sensitivity to him/herself as a: (1) person, (2) teacher of children, and (3) member of the educational system. This is at least as specific as a number of other programs in teacher education (Central College, Roelofs and Sears, 1971; Carnegie-Mellon University, Borke and Burstyn, 1970; Lehman College, O'Hare, 1968; and University of Maryland, Baltimore County, Calliotte, 1971).

The contrasting approach to a subjectively defined experience is proposed by Egan (1970) who suggests contract groups as a structured approach to encounter groups. Contract groups define for the members the expected outcomes of the group and a broad boundry for their behavior. He asks participants to engage in the following kinds of activities; support, self-disclosure, express feelings, confront others, and respond to confrontation (all of which are defined). His thesis revolves around the contract which he sees as having research potential because it points out the behaviors of interest. He feels the contract defines categories that can be used in a scoring system which could be used for research and evaluation. In examining research, Egan

(1970) feels that many of the non-significant findings in the sensitivity movement would better be labeled irrelevant because the measures have so little relationship to the expected outcomes of the groups. He further points out the relationship between clearly defined goals and research and evaluation:

> Perhaps it is time to review the criteria we use to judge the success or failure of sensitivity-training experiences. If measurement is to have any meaning at all, it is necessary to delineate clearly the specific goals of any laboratory experience, to determine what means are associated with achieving these goals, and to devise measures to determine whether these goals have been reached or not. Perhaps the criteria we have used to measure success or failure have been too gross or have not reflected the real goals of the experience (p. 366).

The specification of the goals of an experience such as human relations training is difficult due to the complexity of the behaviors involved and the variance of the experience itself. Movement has been made toward specifying human relations goals in behavioral terms in programs such as Northwest Regional Educational Lab (Wallen, 1968), Indiana University at South Bend (Peterson, et al., 1973), University of Illinois (Gross, et al., 1971) and others but it should be noted that in the majority of the published reports of teacher education programs no mention is made of specific behaviors as outcomes of the programs. It should also be noted that the research and evaluation of these programs does not aid the reader to any great extent in evaluating the effectiveness of human relations programs, particularly in relation to teachers' in-class behavior. Although no correlation is necessarily established because of this trend, it might be

said that some credence is lent to the previous quote from Egan (1970), evaluation may depend on specifically defined outcome behaviors.

In the section to follow, four programs will be explained, a state program, two university programs and a number of university programs related to the human relations training model developed by Carkhuff.

Minnesota Human Relations Requirement

In 1971 the State Board of Education of Minnesota adopted EDUC 521, a human relations component in all programs leading to certification in education. As is the case with many university programs, the goals of this state program are open ended (Hatfield, 1972), including: knowledge and understanding of racial and cultural differences, the ability to recognize one's own attitudes and feelings, ability to create learning environments conducive to successful experiences, ability to communicate effectively with all pupils, and ability to express and encourage others to express honest emotions and understand the effect of one's behavior on others.

Since the adoption of EDUC 521 very little reported research has been conducted. Carl and Jones (1972) reported on a study to determine the effects of the program on teachers but the extent of their evaluation was a questionnaire at the conclusion of the workshop. The questionnaire discovered that the participants felt the experience was helpful in understanding other people's feelings, and Jones (1972) report on a study to determine the

effects of human relations training on teachers. The extent of their evaluation was the administration of a questionnaire at the conclusion of the workshop. The questionnaire discovered that the participants felt the experience was helpful for understanding other peoples' feelings.

A study at the University of Minnesota (Thorman, 1971) examined the effectiveness of four methods of training pre-service teachers in interpersonal skills. The study involved a hundred education students randomly divided into four treatments: (1) control, (2) academic study of interpersonal relations, (3) T-group, and (4) work with school children. The dependent measures of the study were (1) MTAI, (2) Behavioral Inventory of Interpersonal Skills (part I, student rates him/herself; part II, a friend rates the student's interpersonal skills), (3) FIRO-B. The results were not significant although self-report questionnaires showed students' attitudes toward T-group and child experiences, direct experiences (face-face) with people to be more valuable than academic experiences with the same objectives. The findings of the study led to the clear recommendations by Thorman (1971). "(1) present programs for training prospective teachers in interpersonal relations should be subject to close scrutiny, and (2) efforts to construct instruments which are increasingly sensitive to the objectives of interpersonal skills training should continue . . . the results of the study confirmed the need for instruments specifically related to the situation being evaluated" (p. 22).

The Educ 521 Human Relations Component is a requirement that may become a standard part of teacher education certification (Wisconsin has a similar plan). The more these programs are exposed to the public light, the more important close scrutiny will become. This will make Thornman's recommendation for instrument development all the more important as teacher and educational accountability must include evaluation of all programs.

University of Massachusetts: A Behavioral Objective Curriculum in Human Relations

The Model Elementary Teacher Education Program (METEP) at Massachusetts (Allen and Cooper, 1967; Ivey and Rollins, 1970, 1972; Ivey, et al., 1970) is one of nine proposals for elementary teacher funded by the department of Health, Education and Welfare, Office of Education, in 1968. At least six of the nine proposals contain a component or module that is directed towards human relations, sensitivity training or communication skills (Fattu, 1968). Of the nine funded proposals, the University of Massachusetts is perhaps the most visible and includes more focus upon human relations training. The program attempts to teach pre-service teachers the possible options in three areas: content knowledge, behavioral skills, and human relations skills. This review will focus on the human relations skills.

The program is committed to teaching specific behaviors the teacher should be able to engage in using specific behavioral objectives and performance criterion in evaluation.

The human relations portion of the METEP curriculum is called Human Interaction (HI). It is written from a behavioral frame of reference building from traditional human relations, sensitivity training, and behavioral psychology. The teacher trainees participate in a "Do-Use-Teach" program in which they show they can demonstrate (do) the skill, then practice (use) it in their lives, and finally they must teach the skill in the university laboratory school. The program defines, with behavioral objectives, the skills of relaxation, listening (attending behaviors) and non-verbal communication.

The reported evidence of evaluation of the Human Interaction program consists of one experimental study (Iver and Rollins, 1970). The design included random assignment to treatment and control with pre- and post-testing of both groups. The treatment consisted of the "Do-Use-Teach" program including four hierarchies: relaxation, non-verbal awareness, attending behaviors, and decision-making. Two instruments were selected for each hierarchy: one to measure attitudinal changes (a semantic differential), the second to measure changes in skill level as a result of the training. An additional instrument was used to examine the subject's discrepancy between his/her self-concept and his/her goal-self-concept. Each instrument used (all of which had been developed prior to the present study except one) had a reliability of better than .80. Each instrument was used as a pre-test and post-test for each of the hierarchies. The data were collected in settings other than the Human Interaction

groups, and it is not clear if the testing environments relate to groups or interpersonal interaction. The results indicated no change in self-concept discrepancy; but for the two measures of each of the hierarchies, significant changes in the treatment group's attitudes (all except attending behaviors) and skills (all except relaxation) were demonstrated. In the conclusion, the authors make the following recommendation, "The study ought to be seen as an observational study of a human relations program that was performance based. What needs to be done is a replication of this study in which more precise instrumentation is used. . . ." (p. 65).

This study begins to show the effectiveness of a human relations training program, particularly in relation to attitude changes. But the measurement of skill acquisition requires closer scrutiny. The measure of relaxation was based on reading errors due to delayed auditory feedback (a secondary measure of anxiety) and the dependent measure of decision-making was a paper and pencil test. Both non-verbal awareness and attending behaviors were measured by an observation system designed specifically for measuring those skills, but the stimulus and environment in which those skills were demonstrated is not defined. More precise instrumentation is needed, but it is also imperative that the environment in which the data are collected be more precisely defined. It is also important that that environment closely approximate the environment in which the student is expected to display the acquired skill. This may

mean that data should be collected on the pre-service teacher in groups and/or in the classroom.

Carkhuff's Systematic Human Relations Training Model

The most systematically designed and thoroughly researched teacher education program in human relations is the Systematic Human Relations Training Model (SHRT) at the University of Georgia (Gazda, et al., 1973). Based on the model developed by Carkhuff (1969) for lay and professional helpers, this program has at its foundation Rogers' (1957) therapeutic concepts: accurate empathy, non-possessive warmth, and genuineness. The human relations training classes are small groups of approximately ten students who meet with a facilitator for two hours, once a week for ten weeks. The course is quite structured and is theoretically devided into three phases (Gazda, et al., 1973a, 1973b) which introduces and requires mastery on the following skills: phase 1--empathy, respect, and warmth, phase 2--concreteness, genuineness, and self-disclosure, phase 3--confrontation and immediacy.

The entry level of the students is assessed by a modified version of Carkhuff's (1969) communication and discrimination indexes. The global rating of responses (Gazda, et al., 1973a, p. 96) is used to analyze and assign a rating to any helper response. Each of the eight dimensions (empathy, etc.) also has an individual rating scale, similar to the global scale, which is used in instruction to aid students in discriminating facilitative

responses and also a communication scale to rate their own responses as a helper in helper-helpee interaction.

The program is systematically designed to train teachers in the counseling skills which Carkhuff (1969) and others have shown in research to be effective in the helping relationship. The emphasis is on training. While many human relations programs focus on here-and-now feelings and personal awareness, this is not the goal of the Systematic Human Relations Training Model (SHRT). Rather the goal is to have pre-service teachers leave the experience with a set of counseling type skills which they may use in one-to-one teacher-student situations.

There is a great deal of research related to SHRT that has examined many aspects of education and related fields. The results have generally been very supportive of the program. Research by a number of authors in the 1960's showed a high correlation of empathy, warmth, and respect with various measures of teacher behavior and product outcomes. Those students become the impetus for developing a program such as the SHRT. Dixon and Morse (1961) found teachers identified by pupils as "more open" to be significantly more empathetic, warm and respectful. A number of authors (Cogan, 1958; Christianson, 1960; Solomon, et al., 1964) found teacher warmth related to general pupil achievement. In a number of related studies by Aspy (Aspy, 1965; Aspy, 1969; Aspy and Hadback, 1967), reading achievement in elementary students was found to be related to high levels of the facilitative dimension. Other studies have shown the facilitative

dimension related to students' in-class behavior in pre-school adjustment (Truax and Tatum, 1966) and for children with behavior and academic problems (Staffer, 1970). Other studies which show similar results for other student populations include Hefele (1971) with deaf children and Pierce and Schaubel (1970) with graduate student counselors.

Since 1970 a number of studies have evaluated the effects the SHRT had on pre-service and in-service teacher behavior. Berenson (1971) studied the effects of SHRT on student teachers' behavior using a number of dependent measures (Carkhuff's index of responding, a classroom supervisor rating form, the Teacher Situation Reaction Test (TSRT) and Amidon and Flanders Interaction Analysis). The experimental design included an experimental group which received SHRT, a training control group which received didactic instruction in human relations training, a "Hawthorne" effect control group, and the control group proper. The SHRT experimental group showed significant results in: (1) higher levels of helping as measured by the written index of responding, (2) the assessment by classroom and college supervisors, (3) solving problems as measured by the TSRT and (4) differing from the control group in classroom behavior as measured by an interaction analysis (more positive reinforcement, less criticism, less emphasis on subject matter). Other studies using the Index of Responding (Global Scale) have shown significant gains in discrimination and communication of the facilitative dimensions for pre-service teachers (Bixler, 1972; Balzer, 1973; Hornsby,

1973), in-service workshops (Taylor and Barnes, 1970) and at other universities (University of Maryland, Baltimore County, Calliotte, 1971, and Boston University, Marshall, 1970, and Hartzell, et al., 1973).

The SHRT model has been researched by educators for a number of populations examining the effects of group composition (Hornsby, 1973) and other training variables. A majority of these research studies use the Global Scale (Gazda, 1973) as at least one of their dependent measures. This measure, with modification, has been used with success since the late 1960's, but it has some shortcomings when used in an educational setting. The SHRT model attempts to train teachers in specific skills, but the ultimate goal is that they use these skills in the classroom. But the Global Scale cannot be used in a natural environment. It is designed to measure single responses to a helpee stimulus, and the classroom environment is more complex than that. Classroom interaction includes statements which are uncodable when using the Global Scale. As was noted in reviewing the research on the Massachusetts program, it is important to examine the product outcomes of a program in terms of the teacher's in-class behavior. The Global Scale seems to be incapable of categorizing classroom behavior. It is designed for testing and perhaps with modification could be used in one-to-one counseling-type interactions, but classrooms and group interactions involve more complexity than that.

Another difficulty with the Global Scale is that it is a high-inference scale. High inference scales are composed of codes which are not denotable or countable behaviors (Rosenshine and Furst, 1973). An examination of the scale (Appendix C) reveals the inferential nature of coding the categories and the footnote points out the possible subjectivity involved in coding (i.e., how is a coder to interpret "the rater must be guided by the level(s) of the condition(s) that are offered or withheld in the helper's response?"). The high inference nature of the categories is shown in another light by the results of studies by a number of authors (e.g., Muehlberg, et al., 1969; Kiesler, et al., 1967). In examining empathy, positive regard, and congruence, the studies challenged the independence of these scales. A global therapist quality or "good guy factor" was found which accounted for nearly 90% of the variance among empathy, regard, genuineness, concreteness, and self-disclosure. Two explanations could account for these high correlations: therapists high on one dimension are high on all dimensions or the dimensions are not separate. The second explanation could relate to the high inference nature of the scales. An examination of the scales for the eight skills (see Gazda, 1973) reveals a striking similarity. In a study by Childers (1973) of the effects of the SHRT model on student teachers' in-class behavior, the need for a low-inference observation system for group environments is pointed out. Childers (1973) found practically no significant results and, in his recommendations for further research, states: "A more sensitive instrument

should be developed that will more directly reflect differences in communication style" (p. 72).

Michigan State University--Interpersonal Process Laboratories

The human relations program at Michigan State University is part of an introductory educational psychology course, The Individual and the School (Educ 200) which focuses on socio-emotional education. The course is divided into three inter-related parts: the carrel portion which involves the cognitive tasks of teaching concepts (e.g., assessment techniques, respondent learning, etc.), the large group presentation which is a lecture presentation of relevant issues in education, and the Interpersonal Process Laboratory (IPL) which involves the presentation, demonstration, and practice of interpersonal communication skills. The IPL sections of the course consist of approximately fifteen students and one instructor. These sections meet for two hours, twice a week for the entire term (ten weeks). In these sections the instructor presents and explains the seven objectives of the IPL (see Appendix D) to the students and discusses their value and implications for personal relationships in general and for classroom teaching. The major purpose of the IPL section is the practice and demonstration of the seven objectives. That is, the instructor's responsibility is to facilitate and evaluate the students' mastery of the interpersonal communication skills. This is done through the use of strategies similar to those used in sensitivity groups (Lopis, 1973). Each instructor is free to use whatever

strategy s/he wishes (or no structured strategy) to facilitate his/her students mastering the IPL objectives. The course is based on a mastery model and is graded on a pass/no credit basis. To receive a pass, each student is required to "master" each of the IPL objectives. The evaluation of the students rests with the IPL facilitator who is required to prepare a "feedback sheet" for each student twice during the term. The feedback sheet is composed of various behavioral indicators for each of the objectives on which the facilitator rates the students' communication skills. The course is behaviorally oriented, emphasizing the pre-service teachers' understanding and demonstration of specific interpersonal communication skills to aid the in-class teachers to communicate with those around them in both cognitive and affective domains.

A number of research reports have been written concerning the entire Ed 200 course, but very little of the research focuses solely on the IPL phase of the course. In a study of attitude changes as a result of the Ed 200 course, Stiggins (1972) found significant attitude changes using a semantic differential pre-test, post-test design. Using the evaluation, potency, and leniency dimensions, Stiggins (1972) found the carrel concepts (e.g., shaping behavioral objectives) to change meaning more significantly than the IPL concepts (e.g., questioning and listening skills), although most concepts became more valuable, potent and lenient. A student questionnaire study by Schulman and Byers (1974) examines the entire Ed 200 course, but focuses

on the laboratory experience. The questionnaire form was used to gather data on the IPL because of "the lack of adequate alternative means of gathering this data" (p. 1). Results showed over ninety percent of the students felt the course increased their ability to teach; over sixty percent felt the course increased their desire to teach; close to seventy-five percent responded that they would participate in an IPL even if it were not required; and seventy percent said they would like to participate in an advanced IPL.

Using a questionnaire sent by mail one year after their completion of Ed 200, Radke (1975) studied the possible benefits or harm to IPL participants. Using a random sample of twenty-eight respondents, he found two possible casualties (perceived harm, present and past) and fourteen students who perceived growth present and past as a result of participating in an IPL experience.

A study by Schulman (1974) examined facilitator grading and decision-making. She found facilitator grading decisions to vary widely which confirmed a theory that a student's chances of passing vary depending upon the instructor that student was assigned. She theorized that this was a function of either (1) the instructional skills of the facilitator or (2) the varying criteria used by different facilitators. This presented a problem which could not be solved because (1) "there are no objective criteria for determining TA (facilitator) competency levels" (p. 12) and (2) "there are no objective measures of student performance against which the accuracy of TA criteria can be compared" (p. 12).

The human relations program at Michigan State University encounters the same evaluation difficulties as many other similar programs; one must use a questionnaire form (and accept biased perceptions of students), develop an observation system (and accept questionable reliability and validity), or engage in no evaluation and rely on positive comments by enthusiastic students to show the value of the program. These problems lead us to an evaluation of the research that may begin to assist the person charged with the evaluation of a human relations program.

Human Relations Research and Observation: Problems and Solutions

> And the infrequency with which change in teacher and pupil behavior has been the criterion in educational research seems notable, when change in behavior is the goal of education. Much of the available research has suffered from the lack of a planned and coherent design. Faith in laboratory training has sometimes depended on questionable data; measures of known validity and reliability have often been lacking; and reliance, sometimes of necessity, has been placed in questionable self-ratings, loosely and hurriedly constructed self-report inventories: ratings completed by individuals who have little or no opportunity to observe behavior adequately and hard-to-interpret unquantifiable projective devices (Bowers and Sears, 1961, p. 154).

This chapter has pointed out a number of human relations, sensitivity, and/or encounter group programs in education and some related fields. Most of these programs have reported serious obstacles in evaluating their effectiveness. This section will review some of the difficulties encountered in researching group work in education, looking particularly at two general barriers that the field must grapple with if it is to show the potential of these programs in the schools. The quote by Bowers and Sears

(1961) points out the two hurdles which must be cleared if evaluation and research in human relations is to progress: what should researchers measure and how shall they measure it? This section will begin with an examination of the need for goals and objectives in human relations programs. That will be followed by a brief review of the measurement problems associated with subjective data and secondary data. Types of behavioral measures will then be examined including an introduction to the observational devices used to measure group participants' behavior.

As has been noted previously in this review, sensitivity training has been viewed primarily as a subjective experience. The majority of the research in the field has accepted that premise as a given, and this may have been the basic problem that undermined many studies. But this premise is no longer viewed as tenable in education or other fields. Campbell and Dunnette (1968), in a report on industrial T-group experiences, point out three major problems facing T-group research; (1) lack of theory which relates to change, "Presently, it is unclear what kinds of outcomes to expect from any specific T-group effort," (p. 79); (2) the difficulty in relating learning in training groups to organizational settings--what is transfer and how do you measure it? and, (3) the measurement problem is compounded by the slippery notion of "interpersonal awareness." In summary, Campbell and Dunnette (1968) state, "Research must devote more effort to specifying the behavioral outcomes they expect to observe as a result of T-group training" (p. 68).

The ambiguity in goals, training methods and evaluation causes confusion in the consumer since there are conflicting interpretations of the same research data. Proponents of sensitivity training will find that evidence is supportive of a hypothesis that training leads to behavioral change. Critics will review the same results and find no indication of change (Barber, 1969). Many changes are needed to alleviate this problem, but two of the most basic are adequate specification of the independent and dependent variables (Diamond and Shapiro, 1973). "In light of the multitude of critical parameters then, the use of generic terms like 'sensitivity,' 'encounter,' and 'T-group' are inadequate as defining operations. At this stage, it becomes most important for researchers and theoreticians to isolate and specify exactly what goes on in their groups" (p. 2). In relation to dependent variables, it is equally important to employ dependent measures specifically consistent with the group goals.

Once the goals have been behaviorally defined and the nature of the training has been revealed, the next issue which must be examined is the measurement of those goals. Since the goals of the training will involve the behavior of the participants, one method of examining the appropriateness of the measurement tool will be to judge how removed the actual data is from the participants' behavior.

Previously discussed studies involving self-reported perceived change (Bunker, 1965; Miles, 1960 and 1965; and Sperber,

1972) are examples of subject bias contaminating the data. Peer-reported changes (Miles, 1960, 1965, and Kraft, 1967) or student perceptions (Bailey, 1967) suffer from the same problem of perceptual distortion. These approaches to the measurement of behavioral changes seem sound and, since the results are usually encouraging, their use will probably continue. But the effects of perceptual bias are so strong that no conclusions can be drawn from these studies except perhaps the verification of the participants' enthusiasm.

Personality tests and attitude and value inventories are often used to collect secondary data on group participants. Some of the inventories most often used include: Minnesota Teacher Attitude Inventory (Lee, 1967; Sweeney, 1969; Thurmon, 1971), Fundamental Interpersonal Relations Orientation Behavior (Thorman, 1971; Bonmen and Capelle, 1971; Solomon, 1970), Edwards'--"Personality Preference Schedule" (Solomon, 1970), and the Personal Orientation Inveotory (Banment and Capelle, 1971). The results of these studies have generally been disappointing, although some significant results have been found. Nevertheless, the trouble in interpreting even significant data still exists; as Sweeney (1969) pointed out, teacher attitudes are only indicators of the teachers' classroom behavior and even the most drastic change in attitude may not produce any change in behavior.

Studies that examine changes in participants' behavior encompass a wide variety of dependent measures. A study by Schmuck (1967) of in-service teachers' innovative behavior used

the number of innovative practices tried out by the teacher in his/her classroom (self-reported) as a dependent measure. He found significant change as a result of the group experience. Studies by Heck (1971) and Hunt, et al. (1969) use a task developed by Hunt (1965) to measure teachers' interpersonal sensitivity and flexibility. The task involved teaching a lesson in which a student acted as though s/he obviously did not understand the concept. The criteria used to evaluate the lesson was a measure of the teacher's ability to understand another person's perspective, to approach the teaching task from the child's understanding. No significant changes were observed as a result of participation in a sensitivity training experience. A comparison of the Schmuck (1967) dependent measure with the task developed by Hunt (1965) points out an important aspect of behavioral measures. While the Schmuck measure may seem trite, it has an important attribute. He was using sensitivity training to bring about changes in teachers' use of innovative experiences in the classroom. He found significant results and part of that significance must be attributed to the fact that he was measuring the behavior he was attempting to teach. A statement by Heck (1971) points out the dilemma of his study, "Another feature of this particular training project was that the sensitivity training program had one primary objective: that being a behaviorally defined skill labeled communication effectiveness, it was important to measure that skill by using a behavioral method" (p. 505). The non-significant results of this study may have

been caused by the fact that although the dependent measure was a behavioral measure, it did not approximate the group goal. Measuring a skill using a behavioral method is important, but that dependent measure must be consistent with and a direct measure of the behavioral goal.

Direct measures of group outcomes vary in the specificity of their criteria. Meador (1971) reports on the Process Scale developed by Rogers and Rablen (1958) which tends to be very general. The scale measures self-disclosure and the definition of the highest stage will give the reader a feeling for the inferential nature of the scale, "Seventh stage. The individual lives comfortably in the flowing process of his experiencing. New feelings are experienced with richness and immediacy and this inner experiencing is a clear referent for behavior" (Meador, 1971, p. 72).

As was pointed out earlier in this chapter, the work of Carkhuff (1969b) and Gazda (1973a) suffers from some of the same subjective scale definitions as Rogers and Rablen's Process Scale. An article by Gormally and Hill (1974) examines the strengths and weaknesses of research on Carkhuff's training model. One of most notable weaknesses of the Carkhuff rating scales is the difficulty in systematizing judgments of helper responses. Carkhuff's scales have been a valuable contribution to measuring group effectiveness, but they also present problems, "For example, the scale points lack of operational specificity which makes it difficult to maintain objectivity and standardization of scale

use in ratings," and this prompts the authors to suggest ". . . other measures should be related to rating scale changes. Conclusions based entirely on trainee movement on rating scales should be regarded tentatively" (Gormally and Hill, 1974, p. 542).

Another difficulty with the Carkhuff scale is the generalizability of data on skill acquisition to real life situational responding (Gormally and Hill, 1974). A training group's growth is normally reported through data collected by the discrimination and communication indexes. As stated previously, these scales involve multiple choice tests and written responses to client stimulae. But numerous difficulties have been discovered using written responses as a dependent measure. Researchers (Carkhoff, 1969c, and Greenburg, 1968) found evidence that only highly functioning therapists have high correlations among written, oral and live interview situations. Other studies have demonstrated that trainees can write stylistically correct responses but are unable to respond empathically in interviews (Butler and Hansen, 1973). Gormally and Hill (1974) point out the problem, "Learning to communicate empathically requires a different and more difficult level of skill than writing a response. . . . Although written responses are easy and economical to use, they lack generalization to real helping situations: this limits their utility in research" (p. 541). A suggestion by Gormally and Hill (1974) anticipates the need for a category system based on frequency data and independent categories to be used with group recordings:

> An alternative to the use of rating scales is to listen to the entire interview and record frequency data, for example, number of responses that identify a feeling, number of nonverbal referents, etc. The responses can be identified for simple presence or absence, and the categories are fairly independent. Use of frequency data reduces the subjectivity involved in rating scale measures (p. 544).

Another measurement technique which suffers from the lack of generalizability is Kagen and Krathwohl's Affective Sensitivity Scale (A.S.S., 1967). This scale measures sensitivity by having the subjects identify (by multiple choice testing) the feelings of possible clients who are shown on video-taped vignettes. Several studies have used the A.S.S. to examine changes in individuals' sensitivity following a group experience (Danish, 1971 and Dendy, 1971) and a study by Danish and Kagen (1971) points out one of the difficulties. They found significant positive change in some groups but not all groups and the results left them unsure of the reason for the variance. This leaves two questions; what occurred in the various groups to account for the variance? (a process question) and what do significant results mean in terms of the subjects' "real world" behavior? (does success in identifying feelings on a multiple choice test relate to interpersonal empathy or communications?)

The studies reviewed so far suggest two important criteria in the evaluation of behavioral change in group participants. First, the measurement must relate to the goals of the experience. There are two reasons for this; if the researcher is interested in significant results, specific measurement is more likely to produce them (you wouldn't use a test of multiplication following instruction

in addition) and also specific measures will be more generalizable to real life setting (e.g., the classroom) if the goals were appropriate. The second criteria is that the technique should include or be capable of measuring the process variables of the group experience. We must open up the black box called group training so that the process variables can be related to the product variables. If participants change as a result of group experiences, what is it that happens in those groups that bring about those behavioral changes?

Most studies reviewed so far have used product outcomes as the dependent measure although some have been used to measure process variables. The discussion will now turn to two observational systems that can be used to measure process variables.

The Group Assessment of Interpersonal Traits (GAIT) is a report schedule to measure interpersonal skills (Goodman, 1969). The schedule is used in a structured small group situation in which the measurement technique resembles group training. The group is composed of about eight participants and three observers. The participants are asked to write on a card an interpersonal concern which they will voluntarily share during the group meeting. One person volunteers to start (s/he is the discloser) and proceeds to read his/her concern. Another participant may volunteer to engage the person in a five-minute dialogue (s/he is the understander). This continues until every participant has engaged in both roles. The participants and observers are then required to rate all the participants on a six-point Likert-like scale in relation to

statements which reflect the following dimensions: empathic understanding, emotional honesty-openness, warmth-acceptance. A study by D'Augelli (1973) reported difficulty in establishing high reliability with the GAIT. For empathic understanding, emotional honesty-openness, and warmth-acceptance, he noted the following reliabilities: observers - .78, .69, .64 and for participants - .61, .48, .35 respectively. It should be noted that the subjective nature of the three rating categories may have lead to the low reliabilities. These reliabilities along with the structured environment needed to use this technique are definite drawbacks to using the GAIT for group evaluation.

A more specific category observation system was developed by Whalen (1969) for measuring group verbal behavior. Her system has raters score all verbal responses into the following categories: (1) personal discussion, (a) personal self-disclosure, (b) immediate feelings, (c) personal questions; (2) feedback, (a) positive feedback, (b) negative feedback, (c) neutral feedback, (d) accepts feedback, (e) rejects feedback, (f) requests feedback; (3) impersonal discussion, (a) impersonal self-disclosure, (b) extra group process, (c) impersonal questions; (4) group process; (5) descriptive aspects of communicative speech; (6) unscoreable utterances. The reliability estimates were computed for each of the categories individually and the majority of the categories had reliabilities of about .90. The continuous coding by the raters was facilitated by an event recorder which collected frequency and duration data for each of the categories. The

frequency and duration data were analyzed but no analysis was reported concerning the order of the events. Whalen (1969) notes the criteria for the selection of the dependent measure: "The classes (categories) were selected so as to include the behaviors modeled in the film as well as those which typically occur in newly formed groups" (p. 511). Recently, the Whalen categories have been combined with the GAIT technique in research on counseling skills (Rappaport, et al., 1973 and D'Augelli and Chinsky, 1974).

Because Whalen's categories are appropriate for newly formed groups (strangers), they may not all be appropriate for evaluating the outcomes of groups composed of individuals who have met for longer periods of time. But the promise of a category procedure which uses the goals of the human relations programs as the categories needs to be explored. Such a system would be directed toward both specific measurement and process measurement. The development of such a technique will be examined in the next section.

Observation: Measurement of Communication Skills

> As with other fields of study having to do with interpersonal interaction, curiosity about issues outstrips methodological resources. Often the researcher is confronted with a choice between a well-established, tested instrument which has doubtful or tangential relevance to the laboratory situation, or a tailor-made but untested new instrument. There has been a tendency to utilize established, validated measures rather than to rely on homemade devices whose deficiencies may become apparent only after all the data have been collected. Yet . . . instruments must be developed specifically for the social context under study (Stock, 1964, p. 437).

In examining the research on Carkhuff's human relation model, Gormally and Hill (1974) noted the many drawbacks to using rating scales in measuring communication skills. They suggested using frequency data based on specifically defined categories to reduce the subjectivity involved in rating scale measurements. One study has been cited (Whalen, 1969) which collected frequency data using specific, well defined behaviors to delimit a set of categories. It was noted that the system developed by Whalen (1969) was used for research on the behavior of groups which were composed of strangers. Because human relations training participants are not strangers (at least not strangers following the first few meetings), and because the goals for human relations training groups differ from the goals of Whalen's research on modeling and counseling, it would be appropriate to develop a category system similar to Whalen's but directed toward categories encompassing more of the goals and objectives of human relations training.

The categorization procedures used with Whalen's system allow for the analysis of frequency and duration. Another important consideration in evaluating small group behavior may be an analysis of patterns of interaction. A system which will allow the researcher to examine recurring patterns of interaction and what precipitates those patterns would be valuable in assisting the examination of a macro-view of groups. Frequency counts look at groups from a micro-view and often times pick out otherwise unnoticed differences. Patterns of group interaction look at

groups from a macro-view and could help the investigator understand the larger picture of the group. This section will review two important considerations in developing a category system for human relations training. A procedure which simplifies the collection of frequency data and patterns of interaction will be viewed to be followed by a survey of the goals of human relations programs to be used as categories in a new category system.

Interaction Analysis

Classroom and group observation have been a topic of research interest since the early 1940's and many systems of observation have been developed to study classroom climate (Anderson and Brewer, 1945; Lewin, et al., 1939; Withall, 1949; Bales, 1950, and Medley and Mitzel, 1958). In looking at all the observation systems, one stands out as being the most influencial. Flanders' System of Interaction Analysis (FIAC or FSIA) (Flanders, 1960) has been used more often for classroom observation and has stimulated a wider variety of studies concerned with the classroom than any other observation system (Dunkin and Biddle, 1974). But perhaps more important in relation to the present review is the fact that the FSIA has spawned a number of other observation systems based on modification of the FSIA (e.g., Amidon and Hunter, 1967; Hough, 1967; and Ober, 1966) which can be used in a variety of settings to measure a variety of behaviors.

There has been an enormous amount of research derived from the FSIA. Many hypothesis related to teacher effectiveness have been studied using Flanders' system and some of those studies have given education and teachers a better way of looking at themselves. These studies are enlightening but in most instances do not pertain to the present study. But recently the FSIA and other observational systems have received some criticism (Dunkin and Biddle, 1974 and Rosenshine and Furst, 1973). This section will review the advances made in the field of interaction analysis particularly as they relate to the development of an interaction analysis system for measuring communication skills in small group settings. This will include response to some of the recent criticism of observation systems.

Flanders' System of Interaction Analysis (FSIA) or Flanders' Interaction Analysis Categories (FIAC) was developed in the late 1950's to estimate the amount of interdependence between successively coded statements. When developed, the categories themselves were viewed as secondary in importance to the interaction analysis procedure which involved sequential time unit categorization and matrix display (Amidon and Hough, 1967). The impact of the FSIA during the last twenty years has had more to do with the procedures involved in coding with the system than with the categories themselves.

The FSIA is an observation system consisting of ten categories (see Appendix E), seven which refer to teacher behavior, two which refer to student behavior, and a category for silence or

confusion. To use the FSIA (or any other interaction analysis system based on Flanders' procedure) an observer, who has been trained in discriminating the various categories, listens to a classroom interaction, (either live, video-taped, or audio-taped) decides which category best represents each event and writes down the code symbol of that category (Flanders, 1970). The observer codes at a steady coding tempo (from twenty-five to twelve symbols per minute depending on the observer and the system used) and produces a long series of code symbols, one symbol to one event. This list of symbols can be analyzed for category frequencies or a matrix can be generated that will allow the investigator to examine the sequences of events by comparing every event to the event immediately before and after it. The time unit, sequencial coding, and matrix generation have been the prime contribution of the FSIA as they have allowed the investigator to examine the data in many ways. Simple frequency counts may at times be important as the FSIA'S time unit procedure allows the investigator to analyze not only the frequency of certain categories but also their duration. But more important, the matrix generation allows the investigator to examine recurring patterns of behavior within the matrix.

Flanders emphasizes that the ten categories of his original system may only be a starting point for many researchers interested in questions beyond the scope of the FIAC. This encouragement to other researchers to develop new categories for interaction analysis has not fallen on deaf ears. Simon and Boyer (1970), in a

publication devoted solely to observation instruments, note that at least fourteen of the seventy-nine observation instruments they report on are derived from the FSIA. Rosenshine and Furst (1973) report similar findings, of the twenty-five systems they report on that are expansions or modifications of other systems, twenty are at least partially derived from the FIAC.

In Flanders' most complete treatise on interaction analysis (Flanders, 1970) he devotes two chapters to the development of alternative interaction analysis systems. One method of modifying interaction analysis is the use of multiple coding with category clusters. Multiple coding with category clusters is based on the coders ability to code more than one symbol for each event. A code would include more than one diget, each diget representing a category from a different cluster. Therefore if an investigator were interested in who was talking and what kind of communication skill the speaker was using (as is the case in the present study) he would use two clusters, one to specify the speaker, the second to specify the communication skill. The number of possible categories would be the product of the number of categories in the first cluster times the number of categories in the second cluster.

Ober (Ober, et al., 1971) developed a multiple category system, based on the FSIA, in an attempt to overcome what he felt was an overemphasis on teacher talk. Ober felt that although research emphasized the teacher's behavior in the classroom a great deal could be learned from the students' verbal behavior. His system, called the Reciprocal Category System (RCS) (see

Appendix F), devotes equal attention to student talk and teacher talk. The RCS is a multiple category system which has nine categories (plus a category for silence or confusion). Each category can be coded as a single diget (e.g., 1, 5, 9) to represent teacher talk or a two diget number (e.g., 11, 15, 19) to represent student talk. Ober's equal emphasis on student and teacher talk is relevant to the present study because the interactions in human relations training groups are frequently dominated as much by student talk as teacher talk. In developing an interaction analysis technique for human relations training groups it will be important to devote equal attention to teacher and student talk.

The cognitive-affective distinction has been encorporated in many observation systems (80% of the systems in Simon and Boyer (1970) report some emphasis on the affective domain) but analysis of the effects of the affective-cognitive interaction have received very little attention. This may be a function of classrooms and teachers as emphasized by the fact that Flanders and others (e.g. Amidon and Hough, 1967) report an average of less than one percent of teacher talk being categorized as accepting feelings. This seems to have caused most observation systems to focus less on the affective dimension although some systems do emphasize students' feelings (e.g., McRel Interaction Analysis System (Simon and Boyer, 1970, #58) and Hough System (Simon and Boyer, 1970, #9). The affective domain cannot be overlooked in developing an observation system for human relations training as a great deal of emphasis is placed on people's feelings

in such groups. In fact a great deal of data would be overlooked if the affective domain was not equally represented, as often times groups may devote more time to the members' affect than cognition. Just as student talk was noted to be just as important as teacher talk, so too should affective messages be considered just as important to categorize as cognitive messages.

These considerations suggest a cluster for a human relations observation schedule. Since all group interaction must be either student or teacher talk, and either cognitive or affective, the combination of these two dimensions (along with a code for silence) would make up a cluster to denote who is speaking. This cluster would be made up of five categories: (1) silence, (2) teacher-cognitive talk, (3) teacher-affective talk, (4) student-cognitive talk, and (5) student-affective talk. Before turning to an examination of various communication skills which could make up the second cluster an examination of some of the criticisms of observation systems would be appropriate.

One of the most pressing problem with any observation system is the categories, their definition, and the extent to which they are mutually exclusive. For systems to be reliable and valid they must not include more than one category to code a single event. That is there must be a one to one relationship between observed behavior and one possible code. For a system to be reliable an observed behavior should be coded with only one category by any trained observer. Dunkin and Biddle (1974) are critical of many category systems which do not have mutually

exclusive categories. They feel that systems whose categories are not mutually exclusive suffer in reliability, may show weak and contradictory findings, and make interpretation difficult because the researcher cannot be sure what the reported data means. This is certainly an important question to examine in developing a new instrument and not a simple problem to solve. Simon and Boyer (1970) point out that optimally, observation systems should represent a set of mutually-exclusive, all-inclusive behavior but that in reality this is only a goal to reach for. For most systems many behaviors seem to fall into two or more categories which requires observer training which still does not result in 100 percent reliability. Because observation systems are used to measure complex behaviors they cannot be easily defined. One could use a system which had only two categories, "someone talking" and "no one talking." This would result in high observer agreement and would also satisfy the other requirements set down by Dunkin and Biddle (1974) but it would yield results which, in most cases, would be useless. In developing an observation system categories must be chosen with as little overlap as possible. These categories must also be defined as specifically as possible to avoid the confusion of having many categories applicable to one behavior.

Closely related to this problem is the question of the inference the observer must make in coding categories. The previous discussion would seem to indicate that observers should make no inferences at all in coding. But low-inference measures have shown

less success in predicting student success when compared with high-inference measures such as enthusiasm and clarity. High-inference measures are many times less reliable and leave the investigator somewhat in the dark about what the data means in terms of the behaviors the teacher demonstrated. Low-inference measures overcome these problems but have not proven successful as significant predictive measures. Because of this problem Dunkin and Biddle (1974) suggest pursuing the low-inference components of high-inference concepts and Rosenshine and Furst (1973) state, "One way to combine the two observational procedures would be to use student questionnaires as the source of high-inference measures and tape recordings of the instructional period as the source of low-inference measures" (p. 166). The present study uses student questionnaires and tape-recordings in the collection of data.

Dunkin and Biddle's (1974) most general concern for observation is that to some extent the instrument has become the focus of attention rather than the concepts they measure. This is the foundation of their concern that researchers are turning out more and more "new" category systems without being able to state why the categories are chosen or how the research categories are related to other research. In developing new category systems it is important to state why certain categories are chosen and relate those categories to the work that has already been done in the area. The purpose of the next section will be to show how the categories chosen for the present instrument relate to the work done in the field of human relations.

Human Relation Goals: Categories for Observation

In the last section a note was made of the concern raised by Biddle and Dunkin (1974) about researchers specifying why they choose to use various categories in their observation systems. In the field of human relation training and sensitivity education there are numerous terms to denote a small group of behaviors which are viewed as helpful in interpersonal interactions. For an observation system to have the widest possible usefulness it should incorporate categories to measure each of these helpful behaviors. This section is devoted to specifying why the proposed interaction analysis system incorporates the categories it has by showing generally how those categories relate to the communication skills training that is being performed in education.

As has been noted time and again during the review of educational programs, the majority of programs do not have specific behavioral objectives. Of those programs which do have specific goals it is possible to categorize their objectives into three broad skills, self-disclosure, listening, and feedback. These three areas of skills will be examined as they have been defined in communication programs and as they relate to an observation system.

One of the most basic attributes to sensitivity training is self-disclosure. The merits of disclosing one's self to others has been expounded in almost every type of sensitivity group under a wide variety of names and rationale. The term self-disclosure has been used by a number of programs (Barbour and Goldberg, 1974;

Wallen, 1968; Johnson, 1972; Egan, 1970; and Lopis, 1975) and is probably the most well known term deriving much of its exposure from the work of Jourard (1964, 1971). It includes such related terms as openness, leveling or authenticity (Springport High School, 1967) and also the general area of expressing feelings (Egan, 1970 and Belvar, 1974). Johnson (1972) defines it as follows. "Self-disclosure may be defined as revealing how you are reacting to the present situation and giving any information about the past that is relevant to understanding how you are reacting to the present" (p. 10). It is generally what a layman might call a "personal discussion" although it need not relate intimate details of the speakers past life. Whalen (1969) distinguishes this "personal discussion" from "impersonal discussion" in her observation system. She uses the term impersonal discussion to mean the offering of biographical information or other generally accessible information about the speaker. In developing an observational system it is important to distinguish self-disclosure from the offering of other information which does not reveal relevant (relatively non-accessible) data about the speaker. If sensitivity training helps people to level with others and be more open and authentic this should be revealed in an increase in self-disclosure behavior. An observation system for evaluating a human relations program must include a category for measuring self-disclosure and another category to record the offering of other information, similar to Whalen's (1969) impersonal discussion.

Self-disclosure involves the speaker revealing himself to those around him, to others who are listening. The second skill incorporated into most sensitivity education experiences involves the listener. For people to feel comfortable revealing their ideas or feelings they must feel that others are listening to them and accepting what they say and feel. Numerous behaviors are involved in helping others feeling comfortable in self-disclosure, such as attending behaviors (eye contact, posture, etc., Ivey and Rollins, 1970), understanding and exploratory responses (Dendy, 1971), and paraphrasing and behaviors which reflect empathy (Gazda, et al., 1973). Generally these behavioral indicators, and other non-behaviorally defined skills, involve letting the speaker know that the receiver is listening with understanding in a nonjudgmental way. This type of listening is called active listening by Gordon (1970) and is described as listening with the purpose of understanding the complete message of another person (both the content and feelings of the message) and communicating to the speaker, through your behavior, that you are understanding. This involves paraphrasing, asking clarifying and exploratory questions, and showing through your behavior that you understand and empathize with the speaker. In human relations programs this has been referred to as: reflecting and summarizing feelings (Perkins, 1973; Barbour and Goldberg, 1974), empathetic understanding or attention (Bervar, 1974; Barbour and Goldberg, 1974), empathetic understanding or attention (Bervar, 1974; Barbour and Goldberg, 1974) or checking for understanding

and summarizing (Springport High School, 1967). Behaviors such as these are important in any human relations program and for an observation system to evaluate such programs it must have a category which allows for the measurement of active listening behaviors.

The third general skill area of interest to human relations programs is feedback. The following definition by Nylen, et al. (1967) approximates the definitions offered by other programs, whether they call the skill feedback, confrontation, or conflict resolution. Feedback is "communication which gives back to another individual information about how he has affected us and how he stands with us in relation to his goal or intentions" (p. 75). Feedback can be either positive (showing support for the person's behavior) or negative (non-support for his behavior) and generally this skill is further defined as to the feedback's usefulness or constructiveness. Numerous criteria are used in various programs for evaluating how constructive or useful a feedback statement is but generally the following criteria are included: the feedback must describe the specific behavior, the statement should be presented as tentative knowledge not as fact, and the sender of feedback should include his own feelings about the behavior or alternative behaviors the receivers could follow (Bervar, 1974; Johnson, 1972; and Springport High School, 1967). As an integral part of communication training it is important to include in an observation schedule, categories which measure positive feedback,

negative feedback and also some measure of the constructiveness of that feedback.

Conclusion

A portion of the literature related to sensitivity training and its evaluation has been presented here. The increased use of sensitivity training or some variation of human relations training in education combined with the continued lack of reliable measurement instruments in the field points out the need for the present research. Although there is an immediate need for the evaluation of many human relations programs, researchers such as Bowers and Sears (1961) and Stock (1964) have pointed up the more pressing need to develop reliable valid instruments to use in evaluation. The need of the sixties has become the need of the seventies, very little has been done to solve the problem. It seems investigators are more concerned with showing the value of their program than with developing instruments capable of reliable, valid measures. The present research is an attempt to begin to show ways to collect valid and reliable data about the communication skills of group participants in small group settings.

CHAPTER III

INSTRUMENT DESCRIPTION AND PILOT TEST

Introduction

The need for a method of evaluating intensive small group experiences is clear: many groups exist and a recognition of the need for such groups is creating many new programs that incorporate these experiences. However, research and evaluation in this area are lacking and instruments for measuring communication skills are practically nonexistent. The present study is an examination of this problem with an emphasis on one possible solution, an interaction analysis observation system. This chapter will introduce the reader to an interaction analysis system designed specifically for measuring communication skills in small group settings and will be referred to as the Interpersonal Skills Interaction Analysis (ISIA).

The initial section of this chapter will describe the procedures for using an interaction analysis system, specifically the coding and matrix generation, and also briefly describe the categories of the ISIA. The last section in this chapter will describe the ISIA Pilot Test and instrument modifications.

Interpersonal Skills Interaction Analysis

The ISIA follows a multiple coding category system (Flanders, 1970) derived in part from the observation schedules of Flanders

(FIAC) (Flanders, 1960) and Ober (RCS) (Ober, et al., 1971). Being a classroom interaction analysis system, the ISIA seeks to abstract communication by ignoring the content characteristics (i.e., what is being talked about) and focusing on the process characteristics (i.e., the types of communication skills being used in the interactions). Interaction analysis systems are a process of encoding and decoding. Encoding is defined as assigning, by coded symbol, statements to previously defined categories. This assignment is done by trained observers and the recording of the data is done chronologically. Decoding is the reverse process. A trained analyst interprets the coded data from which inferences can be made about the original interactions even though the decoder may not have been present when the original data was collected (Flanders, 1970). In this way the communication process can be examined and compared with other interactions apart from the specific content of the interaction. The purpose of observation schedules is descriptive rather than prescriptive although the understanding to be gained from the data may facilitate future modification of the communication process.

In a multiple coding category system a single code consists of two or more numbers or letters which symbolize a single event. The ISIA which uses two category clusters requires a two place symbol for each event. One place designates the category within the first cluster and a second place which designates the category in the second cluster. For example, a single event might be coded "46" to indicate "4 ," the fourth category in the first

cluster, and " 6," the sixth category in the second cluster. The first category cluster in the ISIA (category A) designates who is speaking (teacher, student, or silence) and in what domain (cognitive or affective) the speaker is talking. The second cluster (cluster B) indicates what communication skill the speaker is using. The following are the codes for the ISIA:

Cluster A	Cluster B
0 - Silence or confusion	1. Positive Feedback
1 - Teacher - cognitive	a. responsible
2 - Teacher - affective	b. irresponsible
3 - Student - cognitive	2. Active Listening
4 - Student - affective	3. Elicits Information
	4. Directs or Suggests
	5. Offers Information
	6. Self-Disclosure
	7. Negative Feedback
	a. responsible
	b. irresponsible

The ISIA has three exceptions to the use of two numbers per code. The code for silence or confusion (A:0)[1] is necessarily a single digit, 0, since it includes no particular communication skill.[2]

The other exceptions are the codes for positive feedback (B:1) and negative feedback (B:7) which include a letter in addition

---

[1]The symbol "A:0" indicates the 0 code in category A, "B:2" indicates the 2 code in category B, "32" indicates the third code in category A and the second code in category B.

[2]Although there is an appropriate and inappropriate time for silence, laughter, etc., this will not be examined by the ISIA.

to the two number codes. The letter is used to indicate the type of positive or negative feedback that is used, either responsible (a) or irresponsible (b). For example, the code B:1a indicates positive responsible feedback and B:1b indicates positive irresponsible feedback. Letters are used with negative feedback as well.

Flanders (1970) describes interaction analysis as a label which "refers to any technique for studying the chain of classroom events in such a fashion that each event is taken into consideration," (p. 5). To extract the optimal amount of information from such a technique, two important conditions must be met. First, the events must be recorded in sequence, allowing the interpretation of the order of the events. Second, the observer must have a coding tempo which will allow the interpretation of the duration of specific events. When these two conditions are met, the coding data can be decoded and interpreted for total time spent in specific types of interactions, what types of interactions precede and follow specific communication skills, what patterns of communication skills exist in the classroom or group, and other questions of interest to the observer, teacher, or researcher.

The ISIA category system is a totally inclusive system (Flanders, 1970) which exhausts all the possibilities of any potential event. That is, the five categories in the A cluster combined with the nine categories in the B cluster allow for the coding of any verbal statement. This makes possible the continuous

coding at a constant rate throughout the observation. The continuous coding at a constant rate (coding tempo) is an important consideration as it allows the interpretation of the sequence of events in relation to their duration. This is essential whenever the observer (or researcher) wishes to investigate the proportion of time spent in any specific category and also in examining the sequencing of particular categories.

Redundancies and contradictions should be avoided in a multiple category system to make the system workable and complete. A code symbol that contains two digits is redundant to the extent that any other two digit symbol can be used to code the same behavior. A code symbol is contradictory when any two serial digits represent categories that are mutually exclusive and therefore produce a meaningless symbol (Flanders, 1970). The ISIA has no redundancies, i.e., one and only one code symbol can be used for any observed event. The ISIA has some contradictory categories which are meaningless and cannot be used. However, all possible interactions can be coded in the ISIA scheme. Hence the categories of the ISIA can be said to be jointly exhaustive.

Table 1 (and Appendix G) shows a listing of the thirty-three possible categories in the ISIA, it does not include the contradictory categories that are by definition meaningless, e.g., the code 45 would signify student talk, affective, offering information, but by definition the offering of information must be cognitive, therefore, the code 45 is meaningless.

Table 1

List of all ISIA Categories

			Cluster B								
			1a	1b	2	3	4	5	6	7a	7b
CLUSTER	A	0									
		1	11a	11b	12	13	14	15	16	17a	17b
		2	21a	21b	22	23			26	27a	27b
		3	31a	31b	32	33	34	35	36	37a	37b
		4	41a	41b	42	43			46	47a	47b

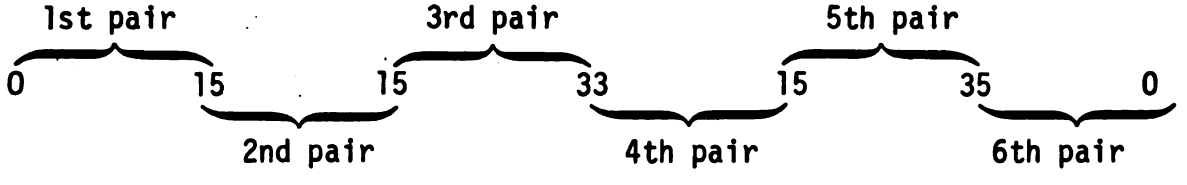
Procedure

The procedure for recording events in sequence involves having a trained observer sit in on a group or listen to a tape recording of a group and decide which category best represents the communication events just completed. This categorizing and recording of the codes in sequence is done as often as possible at a constant tempo. Pilot testing revealed that a reasonable coding rate for the ISIA is approximately one code every five seconds, although some variance between observers is expected. Variance between observers is not as critical a variable as a variant tempo for each observer.

Having a regular tempo is important, because most conclusions depend on rate consistency, for example, the comparison of two categories during the same observation can only be done if a code in one category represents an equal amount of time in another category: twenty coded units of one category should

equal approximately the same elapsed time as twenty coded units of another category.

Once the data is recorded in sequence on an observation sheet (see Appendix H), it can be transferred to a matrix which facilitates the interpretation of patterns within the data. One of the most significant contributions to the field of observation techniques has been the interaction matrix which Flanders (1960) introduced with the FIAS. Given any category system designed for classifying events at a constant rate, in sequence, the information obtained by the data will be increased (in fact, more than squared) by considering pairs of events as the unit to be tabulated rather than single events (Flanders, 1970). This can be done efficiently by generating a matrix with n rows and n columns (n being the number of meaningful categories in the interaction analysis system) and using this matrix to show the sequential order of the observed events. This is done by using pairs of events whereby the first number of any given pair designates the row and the second number designates the column. The following example will help to illustrate the relationship:



Note that in the above example each code symbol is used twice to form a pair (with the exception of the first and last codes which are used once). Each code symbol is first used as the

second number in a pair (except the initial number) and then used as the first number in the following pair. In this way, the n x n matrix facilitates the observer's investigation of the patterns of interaction, i.e., what precedes and follows certain communication skills. For example, the above illustration shows the third pair to consist of the code symbol 15 (teacher - cognitive - offers information) followed by the code symbol 33 (student - cognitive - elicits information). This pair has the address 15 - 33, it is located at the intersection of row 15 and column 33. By using this pairing system on the hundreds of coding symbols recorded in a thirty-minute observation, an observer can generate a matrix that yields a great deal more information than the individual codes themselves. Whenever observation code symbols are recorded in a fashion which preserves the original sequence at a stable coding tempo, a matrix can be tabulated which yields the added information needed to examine an interaction beyond a simple frequency count. This type of matrix combines individual code symbols and short interaction patterns into one matrix which may delineate large more frequent patterns of communication which may go unnoticed if one were examining individual interaction codes or the overall perception of the entire interaction. Primary and secondary communication patterns can be examined for frequency and duration. Individual rows and columns can be inspected to answer questions such as: "What response most frequently follows negative feedback?" or "What most frequently precedes student - affective self-description?"

ISIA Category Description

The ISIA uses two category clusters to discriminate types of interactions in small group settings. This section will briefly describe each category in those clusters. A more complete description of the categories including examples is presented in the ISIA Training Manual (Appendix A).

Cluster A

A:0--Silence or Confusion - This category includes pauses, short periods of silence, or periods of confusion in which the observer cannot understand the interaction clearly enough to code it (e.g., laughter).

A:1 and A:2--Teacher - Both of these categories refer to verbal statements of the teacher (classroom) or group leader (small group, process lab, etc.).

A:3 and A:4--Student - Both of these categories refer to verbal statements of the student (classroom) or group members (small group, process lab, etc.).

A:1 and A:3--Cognitive - Cognitive statements refer to verbal comments which have a factual or content input. Cognitive statements are related to knowledge, the process of knowing. Statements which are coded cognitive are the presentation of how the person thinks about something as opposed to how they feel about it.

A:2 and A:4--Affective - Affective descriptions are those which refer directly to feelings. These statements may refer to either the speaker's feelings or the feelings of other group members.

They often include words which refer to affective states such as love, hate, anger, frustration, shy, etc. (see also, Appendix E in Gazda, 1973). For statements to be coded affective they must label and/or refer directly to a feeling. This is, if someone is obviously angry but does not label that anger ("I'm really angry") the message is coded cognitive. The coding is based on verbal communication not on inferences to feeling states in the group members. This is a very conservative approach but it eliminates the problem of false positives.

Cluster B

Cluster B of the ISIA is used to denote what communication skill is being used by the speaker. These skills are closely related to helper-helpee skills used in counseling and other communication skills programs (e.g., Carkhuff, 1969) although the words used to label the particular skill may be different. The nine categories are not all "communication skills" but rather are particular communication skills and other categories which make the ISIA a totally inclusive system.

B:1--Positive Feedback - Feedback is the response or reaction a person gets from or gives to others regarding one's personal being or actions. It is a verbal response of a sender (the person giving the feedback) to a receiver (person to whom feedback is directed) which is focused on the receiver's being or actions (stimulus behavior). In the case of positive feedback, the sender's message (the positive feedback) shows support for

the stimulus behavior of the receiver. It is, in effect, positive reinforcement for the stimulus behavior. Feedback can be either cognitive or affective. If it is directed toward the receiver's affective behavior (e.g., sharing of feelings) and/or includes the sender's affect associated with the stimulus behavior (e.g., sender stating how the stimulus behavior makes him feel) it is coded affective (21 or 41). Cognitive positive feedback (11 or 31) would be coded for any positive feedback that refers to a cognitive stimulus behavior and does not include the sender's affective reaction to that stimulus behavior.

B1:a--Responsible - For positive feedback to be responsible (B:1a), it must meet two criteria, it must be specific to the stimulus behavior and it must be potentially helpful. For feedback to be specific, it must describe to the receiver the stimulus behavior in specific rather than general terms. That is, the receiver must be aware of exactly what he is getting feedback about. The helpful quality refers to the nature of the stimulus behavior itself. That is, what the sender is approving of must be something that should be continued or increased. To meet the criteria of helpful, the sender must be giving positive feedback about a stimulus behavior that is potentially growth-producing to the receiver. Both of these conditions must be met for positive feedback to be responsible.

B1:b--Irresponsible - Positive feedback which does not meet the criteria of both helpful and specific is coded positive irresponsible feedback (B:1b).

B:2--Active Listening - Active listening is a sentence, word, or phrase which puts the focus of an interaction on the person who has previously been talking and encourages that person to elaborate further in the interaction. This may be accomplished by paraphrasing, reflection of feelings, or the asking of a clarifying or exploratory question. The important ingredient in active listening is that the listener communicate to the speaker that he has understood what the speaker said (or that he does not understand and wishes clarification) and also communicates to the speaker the listener's desire to hear and understand more of the speaker's ideas or feeling.

B:3--Elicits Information - This type of talk asks a question or requests information about the content, subject, or process of the group with the intent that another should answer (respond). The purpose of behaviors in this category is to elicit or secure information. It differs from active listening in that it is the initiation of an interaction and not the encouragement of an ongoing interaction. Eliciting behaviors may be cognitive or affective.

B:4--Directs or Suggests Solutions - This type of talk gives directions, instructions, orders, or assignments with which another is expected to comply. It differs from B:3 statements in that directions are given and compliance is indicated. Statements which are part of an interaction made up of active listening and self-disclosure but which direct the person to a specific solution are also coded B:4.

B:5--Offers Information - The code B:5 is used when a statement is the offering of facts or information concerning the content, subject, or procedures being considered. It is also to code responses to questions or requested information by others. This code is used for statements which are the presentation of facts outside of one's own experience, i.e., it relates what the speaker knows rather than what he has done. It is the presentation of cognitive information and can never be affective.

B:6--Self-Disclosure - Self-disclosure is the offering of information of a personal nature and includes the sharing of values, opinions, personal experiences, and feelings. The B:5 versus B:6 distinction depends on whether the information presented is fact outside of one's experience (B:5) or facts or feelings within one's experience (B:6).

B:7--Negative Feedback - Negative feedback differs from positive feedback in that the sender is stating non-support for the stimulus behavior. The predictable effect of the negative feedback is that it weakens the stimulus behavior. The criteria for responsible feedback (either negative-B:7a or positive - B:1a) is that the feedback must be specific and potentially helpful. If the feedback lacks either specificity or helpfulness, it is coded as irresponsible feedback.

ISIA Pilot Test and Instrument Modification

During the winter and spring term, 1974, a pilot test was run in a sample of IPLs of ED 200. The pilot test was important

for two reasons. Because of the personal nature of intensive group experiences, it was felt that the affective reaction of the members, both facilitators and students, was an important variable to examine. Secondly, because the ISIA was a new instrument, it was important to collect some data on its feasibility for use in groups in relation to useability, reliability, and specificity.

During the winter and spring terms approximately ten IPL sessions were tape recorded. The nature of the research was explained to the members of the groups by the group facilitator and permission was requested to record the group session for that day. In all cases, the groups consented to the taping, although some initial hesitancy by some members was evidenced. In some cases, this hesitancy was discussed after the taping had begun and it is the author's judgment that this reluctance was quickly overcome. Although some facilitators stated an initial apprehension concerning the taping, the stated reactions following the tapings were all positive. The tapes were made available to the facilitators and this was seen by many of them to be an asset in their working with the groups. There were no stated negative reactions by either facilitators or students.

During the spring term, the opinionnaire (Appendix B) was given to two groups following the taping of their sessions. Brief instructions were given orally to the students. They were then instructed to read the directions, fill out the opinionnaire, in class, and hand it in to the instructor. The opinionnaire required less than five minutes to complete for the majority of the students.

Although there was no negative affect associated with filling out the opinionnaires, some difficulties were discovered. These difficulties were all associated with the directions: some students were unsure of the difference between self-description and offering information, some students were confused over how to make the judgments they were asked to make (i.e., they were unsure of what to compare the session with), and some students did not take into account the phrase "first half of the session," (i.e., they used the entire class as the unit of analysis on the first twelve questions). Each of these difficulties was remedied by more specific directions both oral and written.

From the results of the pilot test, the author concluded that the affective concerns were not the problem they initially were felt to be. With the new wave of privacy invasion, it was felt many people would resent tapings, but this was not found to be the case. This is not to say that taping groups is unquestioningly accepted, rather the author found that by explaining the nature of the research and the use to be made of the recordings, the members of the group were quite willing to allow the taping. The key to this success seems to be the honest communication of the objectives and procedures involved in the use of the recorded material.

A majority of the ten IPL sessions that were taped were coded by the author using the ISIA. These tapes were used to examine the feasibility of the instrument in terms of the following questions: How much time is involved in the training of an observer

to use the ISIA? Is it possible to code group sessions using categories such as self-description, active listening, and feedback, or are these too vague? How reliable is the observer and instrument? What modifications need to be made in the instrument, manual, or procedures to conduct future research?

Although no observer was actually trained to use the ISIA (except the author himself) there is some indication as to the length of time it would take to train an observer. After choosing the categories, writing and refining the manual, and listening to some tapes, it took the author less than ten hours of training to reach a level where he had a stable coding tempo and a subjective feeling that he was coding with reasonable reliability. Taking into account the author's familiarity with the instrument and his experience in teaching the IPLs (three years), it seems reasonable to assume the following: (1) the training of an observer who has experience in facilitating IPLs (and therefore the IPL objectives) would require approximately ten to fifteen hours to reach an inter-rater reliability of .80 (using the author as the criterion). (2) It is difficult to judge whether knowledge of group dynamics or knowledge of the IPL objectives was helpful in learning to use the ISIA. It is possible that group facilitators (e.g., sensitivity groups, encounter groups) would be able to use the ISIA in the same period of time as an IPL facilitator. (3) It is felt that individuals with no such experience would take from fifteen to twenty hours of training to become proficient in the use of the

ISIA, depending upon their understanding of interpersonal communication skills.

Because the ISIA was developed partially from the objectives of the IPLs, it would be assumed that the behaviors exhibited in the IPLs would reflect these objectives. The coding of the tapes from winter and spring terms reflected this. Although the communication skills were not equally distributed across all the possible categories, there was evidence that most of the categories were represented and that the category definitions and ground rules are specific enough to make the instrument useable and reliable. For example: One of the sessions from winter term was coded twice (one week between the two codings) to examine the intra-rater reliability. In this particular twenty-five minute segment of one session, nineteen of the thirty-three categories were used. Of the nineteen categories used, some categories were used much more frequently than others (four categories accounted for 66% of the coded data; 15-20%, 36-19%, 35-14%, 0-13%), while the remaining coded data were more evenly divided between the remaining fifteen categories.

The reliability of an observation instrument is a difficult question to address as there is little agreement as to exactly what such a reliability should measure (Medley and Mitzel, 1963; Mitchell, 1970; and Rosenshine and Furst, 1973). This question will be addressed in depth in the section on Reliability, but for the pilot test, it was felt that an intra-rater reliability measure would give some indication of the "agreement coefficient,"

(Rosenshine and Furst, 1973), potential of the instrument. Scott's "pi" (Scott, 1955) was chosen to estimate the reliability as it is unaffected by low frequencies, can be adapted to percent figures, and takes into account the number of categories. The results showed an intra-rater reliability of no less than .70. Although only one reliability check was done (that being on one-half of one session), it is felt that this indicates sufficient specificity to warrant further investigation of the ISIA without major modifications of the instrument itself.

As with any observational instrument, it is important to be able to use the ISIA in more than one situation (e.g., IPLs). During the spring, 1974, the investigator had access to a fourth grade classroom in a local public school (Southridge Elementary School, Charlotte, Michigan) which was using the DUSO program (Developing Understanding of Self and Others, Dinkmeyer, 1970). Three classes were tape recorded while the class was involved in DUSO. These were coded using the ISIA. The ISIA was found to be appropriate for this environment as the categories covered all the interactions and the distribution of codes covered a majority of the possible code categories.

Modifications

From the results of the pilot, a number of modifications were made and implemented in the present study. These modifications fall into three areas, the taping of IPL sessions, the opinionnaire, and the ISIA.

Recording of IPL sessions--The results of the pilot test showed that apprehensions to the tape recording by the group members could best be avoided or alleviated by the honest communication of the objectives and procedures involved in the use of the recorded material. Therefore, it was crucial that these objectives and procedures were made as clear as possible to the students and facilitators who were involved in the study. During the summer term, when a number of groups were taped for the final four weeks, the author personally described the study to each group, answered any questions the group members had and allowed them to decide whether they would participate in the study. This seemed to cause the least interference in group functioning and also avoided the possible negative affect associated with required participation.

Opinionnaire

Several difficulties were discovered in the pilot test in relation to the opinionnaire. The students' difficulty with some of the terms indicated the need for a brief description of terms, particularly those terms which were not found in the objectives for the course (e.g., offering information). A glossary of the terms was included with each opinionnaire.

The directions needed to be spelled out in more detail as students seemed to just glance over them. In light of the students rushing through, the directions (and the glossary) were put on a separate page, as the first page of the opinionnaire. It was particularly important to emphasize two parts of the directions:

(a) it must be clearly pointed out that the opinionnaire was asking the students to look at the class in two parts, the first twelve questions relating to the first half of the class, and the second twelve questions relating to the second half of the class. That is, question #1 is identical to #13 except for the words first in #1 (referring to the first half of the class) and second in #13 (referring to the second half of the class). (b) It must be clearly pointed out that students were to make judgments for the twenty-four questions in relation to other IPL sessions (i.e., individual class periods) they have experienced during the term. Students seemed to confuse "session" and "section," stating that they hadn't attended any other section. This was more clearly explained on the introductory page.

Generally, the observation instrument was acceptable in the form used during the winter and spring terms, 1974. Two minor modifications were seen as beneficial to the present study. The investigator's concern over fine discriminations and a personal communication with Ned Flanders brought about the collapsing of what were reflective statements, clarifying statements, and exploratory statements into one code, active listening. Convergent and divergent questions were also collapsed into what is now Elicits Information. The possible loss of reliability and Flanders' work, which has shown such fine discriminations to contribute little to the data, led to the elimination of those categories. The second modification of the ISIA involved the rewriting of the observation manual. The training of observers and the possible use of the

ISIA in other settings by other investigators requires that the ISIA have an in-depth manual to instruct observers in coding procedures.

CHAPTER IV

METHODOLOGY

Introduction

Chapter III described the procedures, categories and the pilot test of one observation technique for measuring communication skills in small group settings, the Interpersonal Skills Interaction Analysis. This chapter will review the procedures involved in testing the suitability of the ISIA in one particular type of communication skills group, the Interpersonal Process Laboratory (IPL). The answers to the following questions, first posed in Chapter I, are the focus of the development of the ISIA:

1. Can the ISIA be shown to be a reliable observation instrument? What conditions influence that reliability?

2. Can the ISIA be shown to be a valid observation instrument for recording interpersonal communication skills? Using the available subjective criterion, is there any correlation between the ISIA and those criteria?

3. Is the ISIA capable of delineating particular types of communicative behavior in small group settings? Do the subjective reports of group effectiveness relate to the behaviors demonstrated in the groups as recorded by the ISIA?

The answers to these questions will be sought through the analyses described in this chapter. The first area of discussion will be the specific methods of data collection including the population, tapings, observations and observers. The next area of discussion will be the reliability of the instrument, of various types of

reliability, of the specific methods to be used in the present study, and a review of the procedures involved in training observers. A discussion of the validity of the instrument will include consideration of the types of validity measures, the specific instrument used, and the procedures involved in estimating the validity of the observation instrument.

Data Collection

The actual data in the present study were the recorded codes of the ISIA which represent the communication skills being used by group members in a small group setting. Except for the data collected for the in-class versus taped reliability estimate, all data in the present study were taken from audio-tape cassettes. This section will include an examination of the population and sample, raw data, and observers.

Population and Sample

The population for this study will be those individuals involved in an introductory course in education, The Individual and the School, at Michigan State University during the summer term, 1974. The course is divided into three parts: the carrel portion which involves the cognitive tasks of teaching concepts (e.g., assessment techniques, respondent learning, etc.), the large group presentation which is a lecture presentation of relevant issues in education, and the Interpersonal Process Laboratory (IPL) which involves the presentation, demonstration, and practice of interpersonal communication skills. The latter portion of the course, the IPL, will be the area of interest for this study.

Raw Data

During the summer term, 1974, five IPL classes were tape recorded for the last three weeks of the term. During that term, fifteen IPL sections were offered to students. Because of a limited number of tape recorders, only eight sections could be taped. The researcher explained the study to the eight sections, outlining the objectives of the research, explaining how the tapes were to be used, and answering any questions the students had. They were then told to make their decision with their instructor regarding participation in the study after the researcher had left. Six of the eight groups decided to participate in the study. One of these groups was dropped from the sample because of missing data. The five remaining groups make up the sample. Each of the five IPL sections was recorded for the last three weeks of the term, two groups had six recorded sessions and three groups had five recorded sessions. Each recorded session included a two-side cassette tape (forty-five minutes per side) and the opinionnaire data from students and instructor.

Observers

Four observers (those trained in coding the ISIA) were used to estimate the reliability of the ISIA: the researcher, an instructor in ED 200, a school teacher, and a student. It was felt to be important to estimate the reliability of the ISIA using a group of observers with a variety of experiences in interpersonal communications and educational background. The researcher was

experienced in group dynamics, had worked with the objectives under study and had facilitated more than forty IPL groups. The instructor in ED 200 was experienced in group dynamics, had worked with the objectives under study for more than two years and had led approximately five IPL groups. The school teacher was the wife of the researcher who had an understanding of group dynamics but no formal instruction in the objectives of the ED 200 course. She had never participated in an IPL group but was experienced in group work through the facilitation of DUSO exercises in her classroom. The student was a pre-service teacher who had experienced two IPL sections. Except for the researcher, all the observers received the same training with the ISIA (see training manual, Appendix A).

Reliability

The definition of the reliability of an observation instrument involves a number of variables and it would seem to vary according to the environment in which the observation instrument was being used. Medley and Mitzel (1963) define reliability as follows: "A measure is reliable to the extent that the average difference between two measurements independently obtained in the same classroom is smaller than the average difference between two measurements obtained in different classrooms," (p. 250). This definition takes into account three variables, the amount of inter-rater agreement (what Medley and Mitzel call "coefficient of observer agreement"), the amount of within-class variability, and the amount of between-class variability.

The coefficient of observer agreement is defined as the correlation between scores based on observations made by different observers at the same time. This is the most common form of reliability when examining an observation instrument. This type of reliability can be estimated by a variety of reliability indexes; the most common being the percentage of judgments on which the coders agree. Unfortunately, a measure which only takes into account the percentage of agreement is biased in favor of systems with small numbers of categories. For example, a random assignment to a two-category system would yield a much higher reliability estimate than random assignment to a ten-category system. Therefore, a reliability estimate must take into account the number of possible categories and also the number of categories used. Otherwise, one would only need add categories that could not possibly be used to increase the reliability.

When the data to be analyzed is on a nominal scale, as is the case with most observational instruments, one method of determining the reliability is by Scott's π (Scott, 1955). This method can be interpreted as the extent to which the coding reliability exceeds chance. It is calculated by the following formula:

$$\frac{P_o - P_e}{1 - P_e},$$

where P_o (observed percent agreement) represents the percentage of judgments on which the two observers agree when coding the same

data independently; and P_e is the percent agreement to be expected on the basis of chance. This formula takes into account the number of categories used, the number of codes recorded, and the percentage of the agreement between the observers.

Scott's π has been used extensively by those involved in observation research (Amidon and Hough, 1967, Ober, et al., 1971) but it has received some criticism. Mitchell (1969) notes that methods of reliability such as Scott's take into account total events in each category rather than the reliability of individual codes by the observers. This could be a serious problem if the researcher were interested in using individual codes in his analysis. This is not the case in the present research. The units to be analyzed in the present research involve frequency count totals (column totals of the individual class matrices) and patterns within a matrix, not individual codes. A difficulty may also arise in relation to observer's coding tempo. Since it is unlikely that observers' coding tempos will be exactly the same, the only possible way to examine the reliability of individual codes would be to use transcripts of the tapes. This would be extremely inefficient for the present study and is probably not necessary in any case due to the use of a frequency count in the analysis of the data. Therefore, Scott's π is the preferred reliability index in estimating the coefficient of observer agreement. The coefficient of observer agreement is the most basic and most essential step in showing the reliability of an observation technique. In the present study, Scott's π was used to estimate

this coefficient, the reliability of inter-rater agreement or the correlation of trained observers coding the same group sessions.

The coefficient of observer agreement is a necessary condition for reliability but not always a sufficient condition. Unreliability can also arise from within-class variability and between-class variability. If the interactions that are observed and coded do not differ sufficiently between group sessions, even perfect inter-observer agreement will not result in acceptable reliability. For example, if an instrument were developed to measure a trait which everyone demonstrated in exactly the same way, the observer agreement could be perfect, but the between-group variability would be zero and the instrument would be worthless. On the other hand, if the within-class variability was as great as the between-class variability, the trait or behavior being measured would be very unstable and even perfect observer agreement would result in a limited reliability because of what Cronbach (1972) labels the lack of generalizability of the results. If the within-class variability is as great as the between-class variability, whether that variability be very high or very low, the instrument cannot discriminate one class from another. This would diminish the usefulness of the instrument to the point where it could only be used in a descriptive manner.

In examining the reliability of an observation instrument, two separate factors come into play. The inter-rater reliability (or coefficient of observer agreement) relates to the instrument itself, but the stability of the trait or behavior being measured

also influences the reliability. The within-class variability is therefore an important consideration in examining an observation schedule's reliability. Medley and Mitzel (1963) refer to a stability coefficient as the correlation between scores based on observations made by the same observer at different times. They contend that any instability across occasions is due to random error in the environment or the persons. McGaw, et al., (1972) contends that this is not necessarily the case as it does not allow for lawful changes in behavior. In the present study this is an important consideration as there is little expected stability from one IPL session to the next due to the differing objectives of each session. That is, one session may have as its objective listening skills while the next session may have as its objective, feedback skills. This would constitute lawful changes in behavior from one session to the next and would naturally lower any stability estimates between sessions.

McGaw, et al., (1972) compared an observer agreement coefficient to the reliability coefficients associated with alternate forms of a test. If the psychometric analogy were extended, the stability coefficient could be examined by means of the split-half reliability coefficient. Because many small group experiences have lawful changes in the behavior of the group members due to alteration of the objectives of the group session, a stability coefficient comparing different sessions would be predictably low. If it were found that different sessions which had different behaviors were coded in a way that indicated stable behavior across

sessions, this would indicate unreliability rather than reliability. An alternative way of estimating the stability coefficient is to apportion the group sessions into two- or three-minute divisions and then use an odd-even correlation to estimate the stability coefficient.

Using the split-half reliability coefficient, it is possible to begin to examine the ISIA in relation to the definition stated earlier by Medley and Mitzel (1963, p. 250). One judgment of the within-class versus between-class stability would be the extent to which a split-half correlation of one session is greater than the correlation of split-halves of two different sessions. This examines whether within-class variability is less than between-class variability but it may leave some questions unanswered. Because of the flexible approach of instructors and the changing objectives between sessions stability coefficients would be predictable low. Although low correlations are to be expected, within-class stability should show higher correlations than between-class stability.

One additional question of reliability is raised in relation to the influence of non-verbal behavior. In interpersonal communication, part of any message is disclosed through non-verbal cues. Since this study uses data from audio-tape recordings, it is important to investigate the possible loss of information due to using audio-tape recordings as opposed to in-class observations. During the spring term, 1975, the researcher investigated this question by comparing the coded data collected during an in-class

observation to data collected on the same session by means of a tape recording. The researcher observed and coded a group session while it was going on while tape recording the same session. Later, the tape recording was coded and then the in-class observation data was compared with the tape-recorded data using Scott's π to check the reliability of the two observations. A coefficient of observer agreement of .70 or better supports the researcher's contention that the information lost due to the use of audio-tape recordings is not significant enough to justify in-class observations.

Critical to the reliability of an observation instrument is the training of observers to use the instrument. In this case, the training of observers to use the ISIA was accomplished by the use of a training manual (Appendix A). The training manual is essentially a self-teaching guide which briefly explains the procedures of interaction analysis, defines the categories with exemplars and non-exemplars for each category, and finally leads the observer through some exercises which introduce him/her to first the basic distinctions and then gradually incorporate more of the categories until s/he uses all of the categories to code a short transcript. When the observer had mastery of all the categories on typed transcript examples, s/he was introduced to audio-tape observations. The observer was trained on audio-tapes until s/he reached a level where s/he felt comfortable in coding a forty-five minute tape.

For a more detailed description of the training procedures, see Appendix A, ISIA training Manual.

Validity

With observation instruments, a great deal of confusion and debate centers on the question of reliability. The issues of validity for observation instruments seem to receive less debate in comparison. Herbert and Attridge (1975) point up the problem in their article "A Guide for Developers and Users of Observation Systems and Manuals:"

> Though much time and space has been devoted to discussion of the reliability problem in observational research, precious little has been assigned to that of validity. Most of the instruments developed in the observation field have yet to prove the validity of their measures (Rosenshire and Furst, 1973, pp. 125-126). System designers and users frequently do not go far enough in the development of their tools to establish validity against measures of student growth or other pertinent criteria. Still the current progress towards proving the validity of the systems measures must be reported (p. 15).

The validity problems encountered in previous observational studies are also experienced in the present research. A major part of the difficulty in estimating validity arises from the vagueness in the various definitions. In discussing observation instruments, Medley and Mitzel (1963) state: "A measure is valid to the extent that differences in scores yielded by it reflect actual differences in behavior--not differences in impressions made on different observers," (p. 250). They go on to say that a valid observational scale provides a record of the behaviors that actually occurred in such a way that the scores are reliable. Herbert and Attridge (1975) point out the lack of data on validity for observation instruments but one cannot appreciate the neglect (either conscious or unconscious) of the topic until one goes through the literature.

Not only is the validity data absent from the literature, the present researcher found the topic of validity for observational instruments mentioned in only three articles, and two of those were commenting on how rarely the topic was examined. The lack of validity measures for observational instruments can be more easily understood when one considers the accepted measurement definition of validity--validity is ". . . the degree to which it measures what is purports to measure, . . ." (Ebel, 1972, p. 567). Using this definition, an author of an observation instrument would seem to have reason to claim validity for his instrument if the instrument were shown to have a high degree of observer agreement. Meeting the requirement of observer agreement does, in fact, show evidence for both construct validity and face validity. Using these definitions of validity, the ISIA can be shown to be a valid instrument if it evidences a high degree of observer agreement.

Construct and face validity are both what psychometricians would call direct or primary validity (Ebel, 1972 and Thorndike and Hagan, 1955). But it is also important for an instrument to show derived or secondary validity. Derived validity depends on the extent to which a measurement correlates with a criterion score. There are two types of criterion-related validity, predictive validity and concurrent. Because the ISIA is a new instrument in a field which has had very little research (affective education), it is impractical to attempt to show predictive validity. But because the instrument is being developed to add objectivity to previously subjective reports of group effectiveness,

the issue of the concurrent criterion-related validity is an important consideration. The most effective way of demonstrating the criterion validity of the ISIA is by showing the relationship of the ISIA to the most objective standard now being used. As was pointed out in Chapter II, very few evaluations use an instrument even approximating an objective measure but it would seem that the subjective reports by observers and participants comes closest to being an acceptable instrument. Herbert and Attridge (1975) refer to this procedure, "An appropriate, though somewhat primitive, procedure to determine concurrent criterion-related validity might be the comparison of the instrument's findings with the opinion of one or more observers assessing the same behavior" (p. 15).

Rosenshine and Furst (1973) in their discussion of the selection of variables for future observational studies recommend the use of both high-inference and low-inference variables together. They advise using student questionnaires as the source of high-inference measures and tape recordings as the source of low-inference measures. In examining the secondary validity of the ISIA, high-inference measures (questionnaires or opinionnaires) are used as a criterion measure to compare with the ISIA data (low-inference measure).

The procedure for gathering criterion measures involved the collecting of opinionnaire data relevant to particular taped IPL class sessions. The collection of data involved taping IPL class sessions and at the conclusion of the class session having

the participants immediately fill out an opinionnaire (Appendix B) on the group interaction. During the summer term, 1974, both students and instructors were asked to fill out the opinionnaires. Five classes were tape recorded the last three weeks of the term.[1] The twenty-seven recordings are, in effect, fifty-four observations since each tape is a two-sided cassette, forty-five minutes per side. The opinionnaire was designed so that the forty-five minute halves of the tapes could be examined independently: the first twelve questions of the opinionnaire refer to "the first half of the session: and the second twelve questions refer to "the second half of the session."

In the present study, the primary reason for collecting the criterion data is to contribute to the examination of the discrimination ability of the ISIA. The purpose of the ISIA is to add objectivity to the subjective reports of the participants or observers and for this objectivity to be valid, it must relate to the subjective reports. One aspect of a discrimination index is the ability of a scale or test to distinguish accurately between extremes (Ebel, 1972). With a test, the discrimination index examines good and poor testees according to some standard, for observational instruments the judgment rests with the instrument's ability to distinguish extreme examples of the interactions or behaviors under investigation according to a standard. In the present study, the standard is the subjective reports of the group participants.

---

[1]A sixth class was recorded for two weeks but because of missing tapes and opinionnaires, it was excluded from the sample.

The opinionnaire data were used to choose extreme groups using question #8 ("In comparison to other IPL sessions to date, the first half of the session was: (1) One of the best, (2) Above average, (3) Average, (4) Below average, (5) One of the worst") and question #20 (identical to #8 except reference is to "the second half of the session") as the criteria, the mean scores of the responses to questions #8 and #20 were rank ordered to choose the upper and lower 10% as the extreme groups. At first, the choice of the extreme groups was to be simply the top five and bottom five in the ranking. An examination of the ranking showed very small mean differences among the extremes, so an additional criterion was used in choosing the extreme groups. The use of a subjective opinionnaire leaves open the possibility of a certain halo effect in the ratings: One group might see every session as "one of the best," while another group might see every session as "one of the worst." For this reason, it was decided to use IPL sections to choose the extreme groups, selecting the group sessions rated highest and lowest among each of the five IPL sections. This resulted in extreme groups which were very similar to the original upper and lower 10% rankings.

The data for the extreme groups consisted of one audio-tape recording of approximately forty-five minutes per session (with five sessions in each extreme group) and the opinionnaire data for each of the sessions (the number of opinionnaires collected per session varied from ten to sixteen). The effective group consisted of five sessions, two of the recordings were of the first half of

the session, and three were of the second half. The ineffective group consisted of five sessions, all five ineffective recordings were of the first half of the session.[2] This difference is probably a function of the feeling that most group members have that it takes a certain period of time for "things to get warmed up."

A t-test was used to examine the opinionnaire data in reference to mean differences between the effective and ineffective groups. The results supported the hypothesized differences originally stated in the research proposal; the extreme groups were shown to be significantly different in relation to questions #8 and #20,[3] the results being significant at the .001 level. This statistically verified the choices of the extreme groups. The group members did, in fact, perceive the groups to be different.

The questions dealing with time spent in the affective domain (question #1 and #13) showed a significant difference for extreme groups at the .01 level. Group members perceived the effective sessions to have spent more time in the affective domain than the ineffective sessions. Although the questions related to active listening did not show significant statistical differences (the differences being at about the .1 level), they did point out a

---

[2]One of the sessions originally chosen as part of the ineffective sample had to be replaced because of taping difficulty.

[3]For data analysis parallel question data on the opinionnaire (e.g., 8 and 20, 1 and 13, etc.) were combined in the analysis. The questions are identical except for the reference to the first or second half of the session.

difference worthy of examination when comparing the ISIA data. Group members tended to perceive more active listening in the effective sessions than in the ineffective sessions. Group members also perceived the effective sessions to be more genuine (question #9 and #21), relaxed (#10 and #22), constructive (#11 and #23), and involved (#12 and #24) than the ineffective sessions. These questions cannot be compared to the ISIA data for validity but they do add descriptive data, lending support to the contention that the groups are, in fact, different. The original hypothesis related to self-disclosure (#2 and #14) and feedback (#3, #6 and #15, #18) were not supported.

These results were used as the criterion in validating the ISIA. Group members perceived more time being spent in the affective domain in the effective groups when compared to the ineffective groups. They also perceived more active listening in the effective groups compared to the ineffective groups. For the ISIA to be a valid instrument for evaluating communication skills, it must reflect these same differences.

Group members' opinionnaires are very subjective reports and this problem is compounded by the fact that students are not experts in communication skills. For this reason, an additional criteria was used to validate the opinionnaire data. The opinion of an expert in the field of communication skills was used to validate the group members' opinionnaire data. The ten tapes of the extreme groups (five tapes of effective sessions, five tapes of ineffective sessions) were randomly ordered to be

listened to by the expert. The expert was told that the ten tapes included five effective and five ineffective sessions as judged by the group members. No indications were given as to which tapes were effective or ineffective and the expert was instructed to listen to the ten tapes in the order they were numbered. Following listening to a tape, he was instructed to fill out the opinionnaire sheet for each of the tapes and after listening to all the tapes, he was to rank order the tapes from most effective to least effective. The expert opinion data lend support to the validity of the opinionnaire data. Four of the five effective tapes were ranked effective by the expert and four of the five ineffective tapes were ranked ineffective by the expert. The twp tapes which were judged differently by the group members when compared with the expert may be explained in light of the expert's written comment on one of those tapes, "This (tape) is hard to rate because there was some excellent data collected--I feel as though the potential for an exceptional group was present, but confrontation was needed." This particular tape was of the first half of the session and because the group members rated the session after experiencing the second half of the session (an experience the expert was not exposed to) this could have influenced their ratings, the "potential" may have been realized in the second half. This tape was rated effective by the group members and ineffective by the expert. This may be explainable in light of the group members rating the session following the entire taping. The second tape which was rated differently by the expert was rated

ineffective by the group members and fifth in effectiveness by the expert, very close to being ineffective.

The expert opinion data supported the group member opinionnaire data and also shed light on another important area of interest. Each small group session is different and although it is possible and important to examine what behaviors constitute an effective group, it is also valuable to examine the behaviors demonstrated in each individual session within those samples labeled effective and ineffective. This is pointed out in the opinionnaire data where all but one of the effective sessions may rank high on one particular question. That one group is effective but for different reasons than the other four groups. The analysis must examine and describe that difference.

Data Analysis

Reliability

All reliability estimates will be analyzed using Scott's π (Scott, 1955). For each reliability estimate a sum of the number of codes recorded in each category is the unit of analysis. Using Scott's π these totals are used in comparing pairs of observations (observer to observer or live-class coding to taped coding) to estimate to what degree the two observations exceed chance.

To estimate the coefficient of observer agreement each of the four observers (the researcher, ED 200 instructor, public school teacher, and student) were required to code one forty-five minute tape. The reliability tape was randomly selected from the tapes

not previously selected as effective of ineffective. Each observer was required to code the entire reliability tape, using earphones, in its entirety without stopping. That is, each observer was instructed to code the tape from start to finish without stopping or going back even if they missed a statement or section. This insured similar conditions for all observers. In addition to the four observers listed above, it was decided that an observation with corrected codes should be generated. Given unlimited time and the opportunity to change his/her codes an observer could generate what could be called a corrected code observation. This was done in the present study by the researcher who coded the reliability tape a second time. The second coding was corrected by listening to the same tape a number of additional times, stopping when he felt it was necessary, and using a stop-watch to insure a coding tempo of one code every five seconds. This coding represents the ideal coding, or the coding the researcher felt was perfect given unlimited time and the option to make any and all changes. Using Scott's π (Scott, 1955) the corrected observation and the four observer observations were compared to produce a correlation matrix of reliabilities.

To estimate the loss of information due to using tape recordings as opposed to in-class observations two observations were made. During the Spring term, 1975 the researcher coded an IPL section while tape recording the same group: three months later the tape recording was used to code the group. During the summer Term, 1975 the researcher coded a different IPL section

while tape recording the group: seven months later the tape recording was used to code the second group. The conditions for the taped coding were the same as stated previously, the observer was not allowed to stop the tape or to go back and correct any codes. Scott's π was used to estimate the in-class versus taped reliability of the observer in relation to these two groups.

To estimate the within-class stability and compare that to the between-class stability the codings of ten groups (5 pairs, validity tapes) were divided into two minute segments. To estimate the within-group stability the two minute segments within each group were combined to form the sum for each category of the odd two minute segments and the sum of the even two minute segments. These split-half (odd-even) category sums were used to estimate the within-group stability using Scott's π. To estimate the between-group stability the category sums of the odd two minute segment for groups one to five were compared with the category sums of the odd two minute segments for groups six to ten (by pair, 1-6, 2-7, etc.). The results of these computations were an odd-even stability coefficient for ten groups and five between-class stability coefficients. These were examined and the results explained in terms of within-class versus between-class stability.

Validity

Before beginning a discussion of the validity analysis it is important to examine the types of data to be analyzed. Unlike the reliability estimates, the ISIA data used in the validity

analysis were not simply a frequency count of the categories used. The ISIA validity data was based on the ISIA codes recorded by the researcher on the five effective groups and the five ineffective groups. These codes were then used to generate a matrix, on each of the ten groups, which reflects the chronological nature of the data. The frequencies within the cells of the 25 x 25 matrix (see Appendix I) shall be referred to as category pairs. These pairs allow an investigator to state what categories proceed and follow any other category. For example; in group 1 (see Appendix I) a 7 occurs in column 0, row 0, this indicates category 0 is followed by category 0, seven times. Also in group 1 an 8 occurs in row 36, column 0, this indicates that category 36 is followed by category 0 eight times. Of a possible 625 category pairs (25 x 25 matrix), 81 category pairs had a frequency of at least three in at least one group. The frequencies of these 81 pairs, in each of the ten groups, were the unit of analysis for all the validity statistics.

Since the comparisons to be made were between the effective and ineffective groups and because it could be assumed that groups are effective because they include the objectives to be demonstrated in the group, it could be stated that some categories (e.g., active listening, self-disclosure, feedback) were more preferred than others. Siegal (1956) notes this type of relationship in defining an ordinal scale, "It may happen that the objects in one category of a scale are not just different from objects in other categories of that scale, but that they stand in some kind of relation to them. Typical relations among classes are: higher,

more preferred . . ." (p. 24). Therefore the data to be examined in the validity study are on an ordinal scale and non-parametric statistics, particularly rank-order statistics, would be most appropriate.

The establishment of the validity of an observation schedule is a difficult and complex task. This is particularly true when one considers that a given system cannot be said to be valid or invalid, rather only degrees of validity can be supported. Assuming that the ISIA showed a high degree of observer agreement, a high correlation between in-class versus taped observations and a stability coefficient which supports greater within class stability than between class stability, there would be support for face validity and construct validity. To further support the construct validity of the system and to demonstrate the concurrent criterion-related validity of the ISIA a number of questions will be posed. The extent to which the researcher can answer and explain the following questions is the degree to which the system can be said to be valid.

> Question 1: Is the ISIA capable of even the most basic distinctions? That is, can the ISIA distinguish differences between the effective and ineffective groups, irregardless of what those differences are?

As previously noted the participant opinionnaires and the expert opinionnaire showed a significant difference between the effective and ineffective groups. The most basic distinction the ISIA must be capable of making is to show a significant difference between the groups judged to be effective in comparison to the

groups judged ineffective. This question was answered by using the non-parametric Wilcoxon matched-pairs signed-ranks test (Siegal, 1956). The Wilcoxon test was chosen to compare the effective versus ineffective because the comparison is between two related samples. To perform the analysis the sum of the five frequencies (one from each group) within the effective group for each of the eighty-one category pairs, became the effective group data. The same procedure was done for the ineffective groups. This resulted in two related samples, 81 category-pair sums for the effective group and 81 category-pair sums for the ineffective group. The Wilcoxon test was used to test the null hypothesis that there was no significant difference for the frequencies of various categories between the effective and ineffective group. A two-tailed test of significance was appropriate as no inference could be made as to which group is "better." A significant result would give no indication as to what the differences were, only that there were significant differences in relation to the categories used.[4]

> Question 2: If there were significant differences between the effective groups and ineffective groups are those differences related to the category-pairs which represent the objectives under study in those groups? That is, were the differences between the effective groups and ineffective groups due to category-pairs which represent self-disclosure, active listening, feedback, and the affective domain?

To show that an instrument is reliable gives support to the instrument's primary validity. But if it could be shown that the

---

[4]All validity statistics were performed using Indiana University's Statistical Package for the Social Sciences (SPSS), particularly the nonparametric statistical package (Tuccy, 1974).

difference between groups were a result of the demonstration of the group's objectives in one group while in another group those objectives were not demonstrated, then that would accord greater support for the construct validity of the instrument. This is particularly true in the present study, if it could be shown that the effective groups demonstrated the previously mentioned skills more frequently.

This question actually asks a number of questions that will be explored in this section. There are four areas that were examined. These four areas relate directly to the objectives for the group and can be phrased as questions.

1. Is there a significant difference between the groups in relation to self-disclosure?
2. Is there a significant difference between the groups in relation to active listening?
3. Is there a significant difference between the groups in relation to feedback?
4. Is there a significant difference between the groups in relation to the amount of interaction in the affective domain?

These four questions were examined in three ways. First it was important to examine whether there was a significant difference between all ten groups, secondly whether there was a difference between the five effective groups, and thirdly whether there was a difference between the five ineffective groups. These questions were examined by the non-parametric Friedman

two-way analysis of variance (Siegel, 1956). In reference to the Friedman ANOVA, Siegel (1956) states, "When the data from k matched samples are in at least an ordinal scale, the Friedman two-way analysis of variance by ranks is useful for testing the null hypothesis that the k samples have been drawn from the same population" (p. 166). In the present study the Friedman ANOVA was used to test whether each of the 81 matched samples (81 categories) are randomly ranked within the ten groups (null hypothesis) or whether some groups consistently rank higher (in comparison to other groups) on a chosen set of categories. Unlike the Wilcoxon analysis (which used all 81 categories) the analyses using the Friedman ANOVA used only those category pairs which related to the objective being examined. For example, in examining self-disclosure the category pairs 36-36, 0-36, 32-36, etc. were used.

The Friedman ANOVA was used twelve times to examine the four objectives under three conditions in the following way: all groups for self-disclosure, active listening, feedback, and affective interaction; effective groups for self disclosure, active listening, feedback, and affective interaction; ineffective groups for self-disclosure, active listening, feedback, and affective interaction. The sums of the category pairs for positive and negative feedback were combined to form feedback because the categories for feedback were recorded so seldom.

If differences were found in the direction of more of these skills being used in effective groups this would support the construct validity of the ISIA. If differences were found between

the effective and ineffective groups the next question must compare those differences to the opinionnaire data.

> Question 3: If differences are found between the effective and ineffective groups on the objectives, do those differences correspond to the opinionnaire data?

To demonstrate the concurrent criterion-related validity of the ISIA the category-pair frequencies should relate to the ratings of the group participants and the expert. This caused something of a problem in choosing a statistic to correlate the ISIA codings to the subjective judgments of participants and expert opinion. The most obvious problem centered around the fact that the participants ratings were based on only the groups they participated in, that is, different participants were judging each group and they may not have been using the same criteria in their judgments. This problem was not the case in the expert's judgments but his opinionnaire contained a large number of tied scores which also created problems. A simple (although perhaps not entirely statistically sound) solution to this problem was to add the participants ratings of each group on each skill (active listening, etc.) with the experts opinionnaire ratings, creating a score for each group on each skill. These scores reflected a higher weighting of the expert opinion (he had observed all the groups) but allevaited the problem of ties in the experts judgments (there were no ties in the participants judgments). The opinionnaire data scores for each of the ten groups on active listening, self-disclosure, affective interaction, and feedback (positive plus negative) were then rank ordered. Data from the ISIA matrices were summed for each group

on the category-pairs for active listening, self-disclosure, affective interaction, and feedback and these sums were rank ordered. The opinionnaire data rank orders and the ISIA category-pair rank orders were compared using the Spearman rank-order correlation (Bruning and Kintz, 1968) to examine the relationship between the opinionnaire data and the ISIA data.

Answers to the three questions posed above will indicate the degree to which the ISIA may be said to be valid. Another important addition to analyzing the communication patterns within groups may be matrix interpretation (Flanders, 1970). As previously stated the categories of the ISIA may be used to generate a matrix (25 x 25) which can be examined for major and minor patterns of communication. To demonstrate the usefulness of these matrices in analyzing groups, a matrix for each of the ten groups was generated and a flow chart (Flanders, 1970) drawn to graphically illustrate the major and minor patterns within each group.

Conclusion

The question raised concerning the reliability and validity of the instrument were subjected to the tests described in this chapter. They represent all the questions posed earlier in this study. The answers to these questions will be discussed in relation to the usefulness of an observation schedule for evaluating and interpreting the communication skills used in small group settings. Before presenting this discussion, the results of the reliability and validity estimates will be presented.

CHAPTER V

RESULTS

Introduction

The preceding chapter made note of three questions, first discussed in the initial chapter of this study, which make reference to the two essential ingredients in developing an observation instrument, reliability and validity. The methodology chapter also outlined the specific procedures that were used to address the reliability and validity issue relevant to the development of the ISIA. This chapter will be addressed to those questions and more specifically will present the results of the procedures used to answer those questions. The chapter will be organized as was the preceding chapter; beginning with the results relevant to the reliability of the ISIA, then proceding to the issue of validity, and concluding with a discussion of matrix and flow chart interpretation.

Reliability

The most basic question to be answered in the development of an observation schedule is the reliability of the instrument. Although many approaches can be taken in establishing the reliability of an observation instrument three issues will be examined in the present study, the coefficient of observer agreement or inter-rater

reliability, the live (in-class) versus taped reliability, and the stability coefficient.

Coefficient of Observer Agreement

In the present study the coefficient of observer agreement was examined by having four observers, with a wide range of skills and experience in group work and education, code a single forty-five minute tape under identical conditions. The four observers were the researcher (hereafter referred to as R), an ED 200 instructor (F), a public school teacher (C), and an undergraduate student (L). In addition to the four observers a corrected code (K) was generated by the researcher. Scott's π (Scott, 1955) was the statistic used to estimate the reliabilities. The results are presented in the form of a reliability matrix (Table 2).

Table 2

Reliability Matrix - Intercorrelations of the Five Observations Using Scott's π

	K	R	L	F	C
K	--				
R	.88	--			
L	.79	.80	--		
F	.72	.75	.77	--	
C	.78	.77	.72	.72	--

Table 1 presents inter-rater reliabilities for the five observations which range from approximately .70 to .90. The question now arises as to how reliable an observation schedule should be to be considered acceptable. That is not an easy question to respond to because the answer may well depend on how the system will be used and what form of evaluation it will be replacing. For perhaps just this reason very little mention is made in the literature in reference to acceptable levels of reliability. Flanders (1967), in referring to the training of observers in use of his schedule (FIAC), notes that a Scott' coefficient of .85 is a "reasonable level of performance." This gives some indication of what to compare the reliabilities in Table 2 with, but the .85 level of reliability cannot be used as the sole standard of comparison. Flanders' FIAC is a well tested schedule with very low inference categories. This is not true of the ISIA. The ISIA is a new schedule that requires a higher level of inference by the observer in some categories and the ISIA would be replacing a subjective evaluation which has no known reliability. Taking these points into consideration the researcher feels that the ISIA has demonstrated a reasonable level of reliability and although some modifications of the system will be explored (see Chapter VI) generally it may be stated that the ISIA has been shown to be a reliable measure of interpersonal communication skills.

It may also be said that although there are differences in the degrees of reliability between observers, the ISIA may be used reliably by a variety of observers. It is obvious that the researcher

has a higher degree of reliability with the correlated code (in a sense, intra-observer reliability) than any other observer. This may be a result of the intra-observer nature of that reliability but the greater reliability may also be explained by the familiarity of the researcher with the system. Any other explanation for the differences in the levels of reliability would be speculative at best and will not be explored.

Live Versus Taped Reliability

The second reliability estimate that was examined related to the possible loss of information due to tape recordings. It was felt that because interpersonal communications relied so heavily upon non-verbal cues that tape-recorded observations might be much less reliable than live, in-class observations. Although it is impossible to decifer which is the more accurate, an in-class observation or a tape-recorded observation, Table 3 shows the coded observations to be very similar.

Table 3

In-Class Versus Taped Observations

Observation I - .72
Observation II - .79

Both reliability estimates are calculated by Scott's π and are based on an in-class observation followed much later by a taped

observation. In Observation I there was a three-month interval between the in-class and taped observation, for Observation II there was a seven-month interval. The lengthy interval between in-class and taped observations insured against the observer recalling the categories used to code particular interactions. One additional point should be made in reference to Observation II. For six months prior to the taped observation (the last six of the seven months) the observer did not code a single group and did not refresh his memory concerning actual coding. This certainly gives support to the notion that although the ISIA may be difficult to learn it is not something that is easily forgotten (at least not for the researcher).

The reliabilities presented in Table 3, although not extremely high, do alleviate a great deal of the concern over tape-recorded observations. The two reliabilities presented do not differ significantly from the inter-rater reliabilities (Table 2). One question which does arise out of the design of the in-class versus taped reliability is: What if the in-class observer and taped observer had been different people? This question was considered in planning this design but because of the inaccessibility of observers this consideration had to be excluded from the design. If future research were conducted this should be an important consideration in designing the research.

Stability Coefficient

The two previous sections have shown the ISIA to be a reasonably reliable instrument. But what about the behaviors

being categorized: are there differences between groups? and are the behaviors stable within a group? To answer these questions a split-half reliability coefficient was proposed in Chapter IV. Although the split-half reliability coefficient proposed is not an established procedure it was one way of addressing Medley and Mitzel's (1963) concern for a stability coefficient in estimating reliability, while still taking into consideration the lawful changes in behavior due to differing objectives between group sessions. Table 4 presents the stability coefficients using Scott's π to estimate the reliabilities.

Because the proposed stability coefficient is not an established procedure there are no statistical methods to verify that there is greater within-group stability than between-group stability. But an examination of Table 4 certainly lends support to this hypothesis. In every pair of groups except one, the within-group stability is at least twice as reliable as the between-group stability. The one exception (Groups 2 and 7) may be explained by the similarity of the groups (the expert opinion rated groups 2 and 7 the same on six of nine questions and the flow chart of ISIA category-pairs, Appendix I, shows the groups to be very similar).

The stability coefficient is an important yet difficult estimate to judge empirically. The split-half method in the present study lends support to the stability of the behaviors the ISIA is observing but it has shortcomings. In future research the stability coefficient could be more effectively examined by using group objectives as one dependent variable. That is, a more important

Table 4

Within-Group and Between-Group Stability Coefficient Estimates Using Scott's π

Group 1, odd-even, r_{xx} = .74
Group 6, odd-even, r_{xx} = .75
Group 1 odd-Group 6 odd, r_{xy} = .38
Group 2, odd-even, r_{xx} = .62
Group 7, odd-even, r_{xx} = .69
Group 2 odd-Group 7 odd, r_{xy} = .66
Group 3, odd-even, r_{xx} = .75
Group 8, odd-even, r_{xx} = .84
Group 3 odd-Group 8 odd, r_{xy} = .34
Group 4, odd-even, r_{xx} = .48
Group 9, odd-even, r_{xx} = .80
Group 4 odd-Group 9 odd, r_{xy} = -.21
Group 5, odd-even, r_{xx} = .67
Group 10, odd-even, r_{xx} = .73
Group 5 odd-Group 10 odd, r_{xy} = .12

issue to examine would be the stability between groups who have the same objective (e.g., positive feedback) in comparison to the stability between groups who have different objectives. This is not to take away from the information gained by using the split-half procedure, the ISIA has demonstrated within-group stability, but additional methods should be used in future research to explore this issue, particularly in light of the changing objectives between groups.

Validity

The second issue to be addressed in developing an observation instrument is the validity of the instrument. As was pointed out in Chapter IV, the face validity and construct validity of an observation instrument can be demonstrated by the reliability of the instrument. Now that the reliability of the ISIA has been shown it is important to consider other methods of establishing the construct validity of the ISIA and to examine the concurrent criterion related validity of the instrument. The approach to those issues was established in Chapter IV in the form of three questions. This section will answer those questions by presenting the results of the procedures recommended in the previous chapter and interpreting those results.

> Question 1: Is the ISIA capable of even the most basic distinctions? That is, can the ISIA distinguish differences between the effective and ineffective groups, irregardless of what those differences are?

Table 5 presents the results of the Wilcoxon matched-pairs test which answers the most fundamental of the validity questions.

Table 5

Wilsoxon Matched-Pairs Signed-Ranks Test - Effective Versus Ineffective Groups

Cases = 81	Two-tailed probability = .0020

It is clear from the results of the Wilcoxon matched-pairs test that there is a significant difference between the summed category-pair totals for the effective versus ineffective groups. This result indicates that the distribution of the frequencies of the codes within the 81 category-pairs for the effective and ineffective groups is significantly different. This confirms that there are differences but sheds no light on what those differences are.

> Question 2: If there were significant differences between the effective and ineffective groups are those differences related to the category-pairs which represent the objectives under study in those groups? That is, were the differences between the effective groups and ineffective groups due to category-pairs which represent self-disclosure, active listening, feedback, and the affective domain?

Question 2 was first posed in Chapter IV along with the procedures directed toward answering the question. In this section the results of the Friedman two-way analysis of variance by ranks will be presented along with an interpretation of those results. The results will be presented in a fashion which allows an examination of; first all ten groups, next only the effective groups, and finally the ineffective groups.

The reliability data already presented supports to some degree the construct validity of the ISIA. The Wilcoxon test has shown the ISIA to be sensitive to differences between effective and ineffective groups. But the most important issue in supporting the construct validity of the ISIA is the system's ability to discriminate the objectives of the groups. To verify the ISIA's ability to discriminate these skills a Friedman ANOVA was performed on all groups in relation to the category-pairs associated with self-disclosure, active listening, feedback, and affective interaction. The results are presented in Table 6.

The analysis involves the ranking of the ten groups on each of the category-pairs associated with the objective under study. For example, the analysis of self-disclosure involves 38 cases (category-pairs) and the ten groups are rank ordered (from 1-10) on each of the 38 cases. The sum of the ranks for each group indicates the degree to which each group (in comparison to the other groups) demonstrates self-disclosure. Significant results are evidence of non-random rankings between the groups and indicate significant differences between the groups in relation to the amount and type of self-disclosure (different category-pairs being different types of self-disclosure, i.e., different in relation to what precedes or follows self-disclosure).

The results from table 6 show a significant difference between the ten groups at least the .05 level for self-disclosure, active listening, and affective interaction and a difference at the .08 level for feedback. This confirms the differences between the groups

Table 6

Friedman Two-Way Analysis of Variance

by Ranks--All Groups

All groups, all self-disclosure

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	229.0	215.0	195.0	246.0	249.5
	Group 6	Group 7	Group 8	Group 9	Group 10
	197.5	237.5	188.5	151.5	180.5

Cases	Chi-Square	D.F.	Significance
38	26.1947	9	.0019

All groups, all active listening

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	91.5	94.0	121.0	115.5	145.5
	Group 6	Group 7	Group 8	Group 9	Group 10
	98.0	101.5	85.0	86.0	107.0

Cases	Chi-Square	D.F.	Significance
19	17.9914	9	.0353

All groups, all feedback

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	58.0	76.0	55.0	94.0	82.5
	Group 6	Group 7	Group 8	Group 9	Group 10
	55.0	65.0	55.0	55.0	64.5

Cases	Chi-Square	D.F.	Significance
12	15.5227	9	.0775

All groups, all affective interactions

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	111.5	91.0	78.5	129.5	126.5
	Group 6	Group 7	Group 8	Group 9	Group 10
	102.5	97.0	86.0	78.5	89.0

Cases	Chi-Square	D.F.	Significance
18	18.3788	9	.0310

on all the objectives.[1] An examination of the rank sums shows these differences not to be random among the ten groups but rather the effective groups consistantly have rank sums which exceed the ineffective group's rank sums. The Friedman ANOVA ranks every category-pair by assigning a rank of one to the smallest number and therefore the smallest rank sum indicates the group with the least amount of self-disclosure, active listening, etc. The significance level of the four analyses clearly points out the differences between the ten groups and an examination of the rank sums shows the effective groups to be displaying each of the objectives more frequently than the ineffective groups. This distinctly confirms the construct validity of the ISIA.

To take the analysis one step further sheds even more light on the differences between the groups. Table 6 reports the results of the Friedman ANOVA on all ten groups, Table 7 reports the results of the same statistic performed on only the effective groups and Table 8 reports the results for the ineffective groups. The results reported in these two tables afford the opportunity to examine more closely the differences between the effective and ineffective groups.

The results of Table 7 indicates significant differences between the effective groups on active listening, feedback, and affective interactions but non-significant differences for self-disclosure. The results of Table 8 indicates significant differences

---

[1]Although the analysis of feedback does not show a significance at the .05 level it will be assumed to be significant. This was not a formal hypothesis testing but rather an exploratory study in which the researcher defines .08 as significant.

Table 7

Friedman Two-Way Analysis of Variance by Ranks--Effective Groups

Effective groups, all self-disclosure

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	111.5	108.5	99.5	124.5	126.0

Cases	Chi-Square	D.F.	Significance
38	5.2737	4	.2604

Effective groups, all active listening

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	45.0	47.0	61.0	58.0	74.0

Cases	Chi-Square	D.F.	Significance
19	11.5789	4	.0208

Effective groups, all feedback

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	28.5	37.0	27.0	47.0	40.5

Cases	Chi-Square	D.F.	Significance
12	9.3167	4	.0537

Effective groups, all affective interactions

	Group 1	Group 2	Group 3	Group 4	Group 5
Rank sums:	55.0	46.0	40.0	65.5	63.5

Cases	Chi-Square	D.F.	Significance
18	10.7444	4	.0296

Table 8

Friedman Two-Way Analysis of Variance by Ranks--Ineffective Groups

Ineffective groups, all self-disclosure

	Group 6	Group 7	Group 8	Group 9	Group 10
Rank sums:	117.0	139.0	110.5	94	109.5

Cases	Chi-Square	D.F.	Significance
38	11.2263	4	.0241

Ineffective groups, all active listening

	Group 6	Group 7	Group 8	Group 9	Group 10
Rank sums:	58.5	60.0	51.5	52.0	63.0

Cases	Chi-Square	D.F.	Significance
19	2.1579	4	.7067

Ineffective groups, all feedback

	Group 6	Group 7	Group 8	Group 9	Group 10
Rank sums:	34.0	39.5	34.0	34.0	38.5

Cases	Chi-Square	D.F.	Significance
12	1.0167	4	.9073

Ineffective groups, all affective interaction

	Group 6	Group 7	Group 8	Group 9	Group 10
Rank sums:	59.5	57.5	51.5	47.5	54.0

Cases	Chi-Square	D.F.	Significance
18	2.0222	4	.7317

for ineffective groups on self-disclosure but non-significant results for active listening, feedback and affective interaction. These results point out a possible difference between self-disclosure and the other skills.

Self-disclosure is a skill which is demonstrated to some degree in all groups. An examination of the rank sums in Table 6 would indicate that self-disclosure occurs more frequently in effective than ineffective groups. To verify this difference a Wilcoxon Matched-Pairs test was performed comparing the effective and ineffective groups on self-disclosure. Table 9 shows a significant difference between the effective and ineffective groups at the .0001 level verifying that the effective groups do demonstrate self-disclosure more frequently than ineffective groups.

Table 9

Wilcoxon Matched-Pairs Signed-Ranks Test--Effective Versus Ineffective Groups, all Self-Disclosure

Cases = 38	Two-tailed Probability = .0001

This result along with the fact that effective groups have less variance on self-disclosure (more homogeneity, non-significant difference, Table 7) while ineffective groups have more variance on self-disclosure (more heterogeneity, significant difference, Table 8) tends to support the theory that there is a minimal amount (or mastery level) of self-disclosure that must occur in a group for it to be considered effective.

Beyond this minimal level there is not a great deal of difference between effective groups on self-disclosure, self-disclosure is a necessary but not sufficient condition for a group to be considered effective. But in ineffective groups the only significant difference between the groups is self-disclosure. It may be that the ineffective groups with the most self-disclosure are the least ineffective (this is supported by the fact that group 7 has the most self-disclosure of the five ineffective groups and was rated highest by the expert in effectiveness, of the ineffective groups).

While self-disclosure is the skill which distinguishes between the ineffective groups (mostly because it is the only skill demonstrated to any extent), in the effective groups each of the other skills shows a significant difference between the effective groups. This supports a hypothesis that effective groups are not all effective for the same reason. Some have a great deal of active listening, other a lot of feedback, while others deal more in the affective domain. These findings are significant and certainly support the construct validity of the ISIA but they are to some extent speculative. These interpretations are construed from statistical analysis and the subjective observations of a group leader (the researcher) and seem to make a lot of practical sense. But to state as a fact that the difference between effective and ineffective groups is self-disclosure or feedback is not possible until further research is conducted.

This analysis supports the validity of an instrument to be used in such research. There are differences between effective and ineffective groups and there is strong indication here that the difference involves self-disclosure, active listening, feedback, and affective interaction. Further research is needed. The ISIA is one tool to aid conducting such research.

Question 3: If differences are found between the effective and ineffective groups on the objectives, do those differences correspond to the opinionnaire data?

In Chapter IV a method to establish the concurrent criterion-related validity was proposed which summed the expert opinionnaire ratings with the average participant ratings and correlated those sums with the ISIA matrix data. A Spearman rank order correlation coefficient established the relationship between opinionnaire data and ISIA data for self-disclosure, active listening, feedback, and affective interaction. The results are reported in Table 10.

Table 10

Spearman Rank Order Correlation Coefficient: Opinionnaire Data--ISIA Category Data

Self-Disclosure	, rho = .70
Active-Listening	, rho = .32
Feedback	, rho = .85
Affective	, rho = .63

With the exception of the active listening correlation, the rank order correlations show a relatively high relationship between

the opinionnaire data and the ISIA. These seem particularly impressive when one considers the subjective nature of the opinionnaire data. The opinionnaire data were not originally ranked by the expert or participants. Rather the participants rankings were generated from the opinionnaire data which is based on raters (participants) who had not observed all the groups. The expert observed all the groups but his ratings had many tied scores (a number of cases where four of the ten groups were rated the same). These two problems were alleviated to some extent by summing the expert and participant ratings but this probably weakened the power of the rank order statistic.

The exception to the high correlation (active listening) may reflect a difference in definition between the expert and the researcher. In the rank ordering of the groups on active listening the expert's ratings ranked group 3 in tenth position while the ISIA data rank group 3 second. This one difference in rankings lowered the correlation from .71 to .32. The difference in the rankings of group 3 may be a function of the types of interactions in group 3. Group 3 was a very cognitive group with a high frequency of offering information and numerous questions related to the information offered. In the ISIA category system, continued questioning of information is coded active listening. In questioning the expert it was found that he did not consider these questions active listening. This further supports the sensitivity of the ISIA. An expert who uses a different definition of even a part of one category will dramatically effect the results.

Generally it may be said that the ISIA has demonstrated a high level or reliability and validity. In Chapter IV a helpful aid in analyzing and interpreting the interactions in groups was mentioned. To close out this results chapter a brief explanation of matrix and flow chart interpretation will be presented.

ISIA Matrix and Flow Chart Interpretation

Up to this point the ISIA has been used as a source of data to be used in statistical analyses to verify the reliability and validity of the instrument. But the ISIA can be used in other ways to examine the processes of small groups. One of Flanders' (1970) greatest contributions to the field of observational studies was the interaction matrix. As described earlier, the matrix allows an investigator (group member, facilitator, evaluator, researcher, etc.) to abstract the patterns of communication within a group from the content of the group. This is a valuable tool in comparing groups because an observer can be biased by the content of a group (i.e., what is talked about), and there are times when the process is more important than the content.

In lieu of an abstract discussion of the value of matrix and flow chart interpretation this section will use a concrete example to demonstrate the kinds of interpretations that can be drawn from the comparison of two flow charts. The reader should be aware that every group is different and the discussion to follow is only one interpretation of two very dissimilar groups. Other groups would raise different questions. Other investigators might

even arrive at different interpretations of the same groups. But the generation of matrices and flow charts simplifies what can be an overwhelming amount of data into a manageable picture which points out significant patterns of interaction. This is not to say that the role of the expert observer (albeit, subjective observer) will not be needed in the future. The ISIA matrix or flow chart quantifies the interactions but cannot interpret the quality of the skills used.

The interpretation made in this section will be based on the flow charts of an effective group (group 4, Figure 1) and an ineffective group (group 9, Figure 2). These groups are the same IPL section (same group of participants) and represent the most effective (group 4) and ineffective (group 9) session tape recorded during that term for that group. The flow charts were generated from the matrices for each group.

Flow charts are best used to interpret the amount of time devoted to particular categories and the communication patterns (or directional flow) within a group. This interpretation is facilitated by three aspects of the flow chart: Each box within the flow chart depicts a stable cell of the interaction matrix. A stable cell being a category-pair (e.g., 35-35, 22-22) which represents a category code followed by the same code. The size of the boxes in the flow chart depict the frequencies of the category-pair. The second aspect of interest in the flow chart is the arrows connecting the boxes. The arrows represent the transition cells within the matrix. A transition cell indicates the frequency of

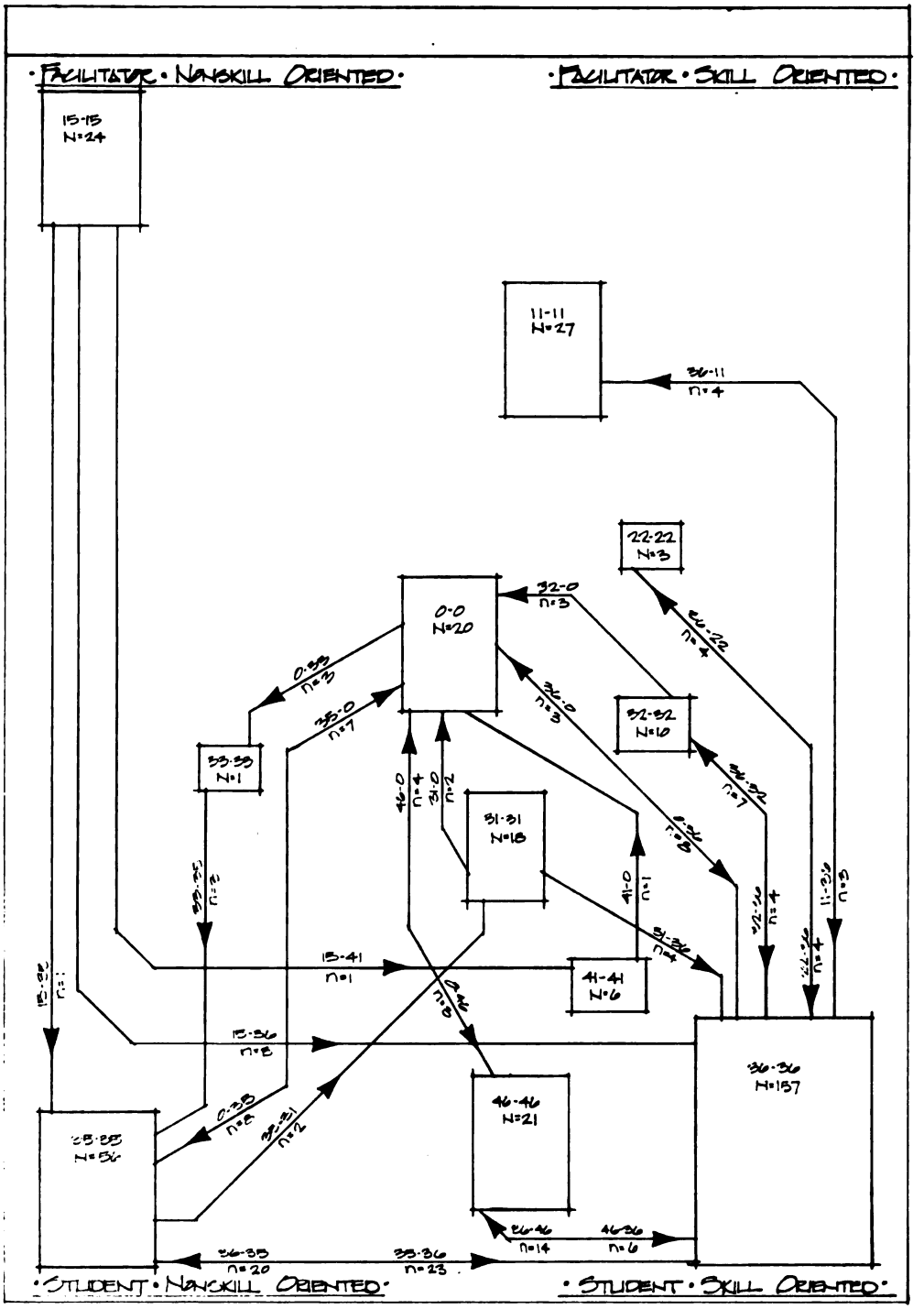


Figure 1.--Communication Flow in an Effective IPL Group (group 4) as Recorded by the ISIA.

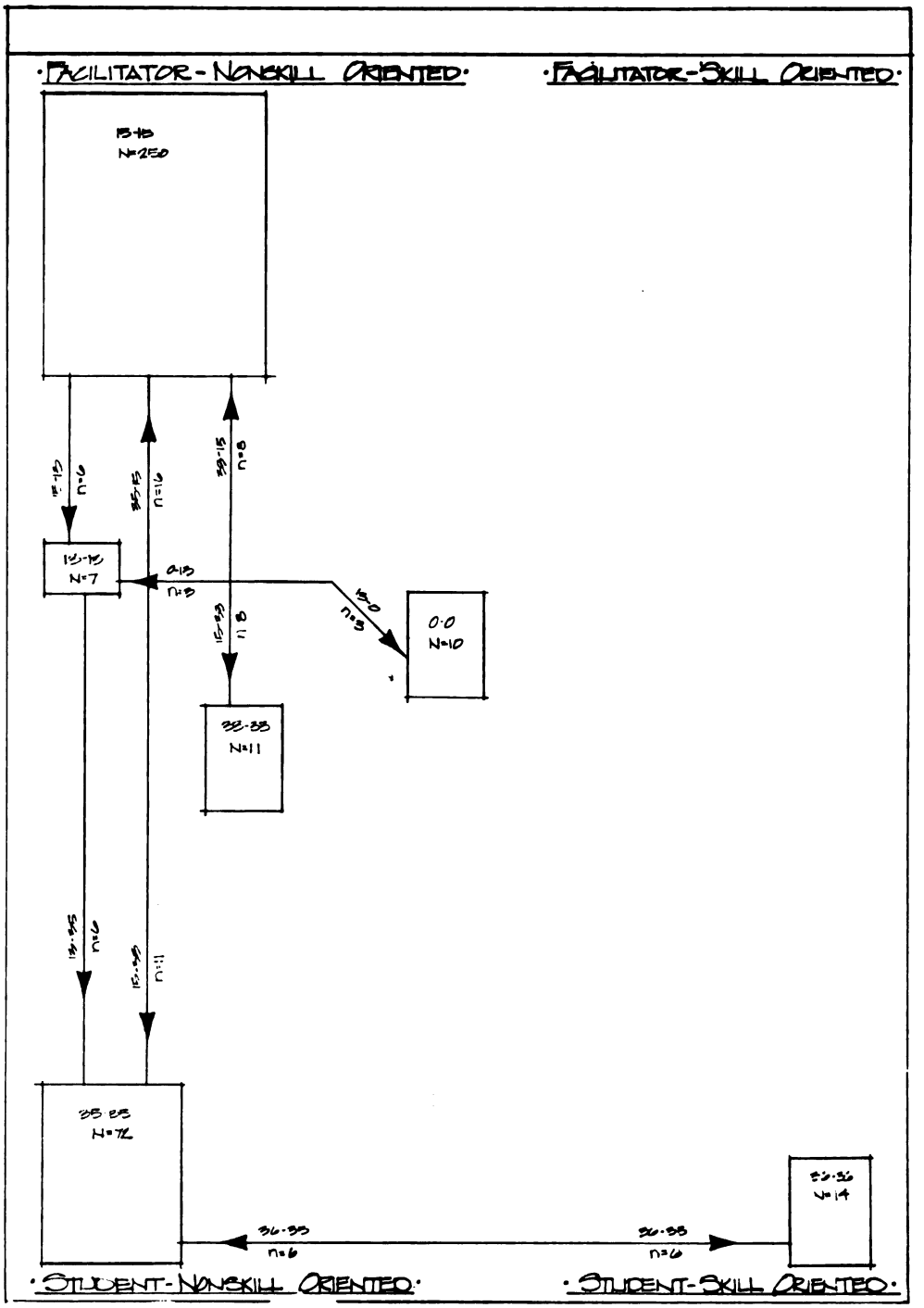


Figure 2.--Communication Flow in an Ineffective IPL Group (group 9) as Recorded by the ISIA.

category-pairs which show movement from one stable cell to another (e.g., the 15-35 category-pair represents facilitator-offering cognitive information followed by student-offering cognitive information, which occurred eleven times in group 9, Figure 2). The exact frequencies of stable cells are written within the box, the frequencies of transition cells are written along the arrows. The third aspect of interest encorporated in the flow chart is the placement of the stable cell boxes. The flow chart is divided into four areas. An imaginary horizontal line from mid-left to mid-right divides the flow chart into facilitator categories (above the line) and student categories (below the line); a vertical line from mid-top to mid-bottom divides the flow chart into skill oriented categories (self-disclosure, active listening, feedback and affective categories, right of the line) and other categories (offering information, elicits information, etc., left of the line). Location of the boxes within quadrants is based upon convenience of graphical presentation and has no interpretive meaning.

There is no one best way to interpret a flow chart and perhaps there are not even any ground rules that one must follow. But there are some questions to ask that may help an investigator interested in differences between groups. There are some very obvious differences between Figure 1 and Figure 2. Perhaps the most obvious difference involves the quadrant in which most of the "action" is taking place. In Figure 1 a large portion of the category-pairs occur in the lower half (student skill categories) while in Figure 2 a large portion of the category-pairs occur in

the upper half (facilitator, non-skill categories). This points out the first difference between the groups, group 4 (Figure 1) spends a great deal of time with students demonstrating the group objectives while group 9 (Figure 2) spends time having the facilitator offer information. This leads to a second criteria in examining the flow chart, the frequencies of particular categories of importance to the investigator. In examining Figures 1 and 2 the most striking difference relates to the 15-15 and 36-36 category-pairs. Group 4 spends a great deal of time in student self-disclosure and very little in facilitator offering information. The opposite is true for Group 9. In examining particular categories (or objectives) the flow chart is particularly helpful. For example, if one were interested in comparing active listening and questioning, group 4 has quite a bit of active listening (32-32, 22-22, 36-22, 32-36) but very little eliciting of information. The opposite is true of Group 9 (e.g., 13-13, 33-33, 15-13, 33-15).

These are obvious differences that jump off the page, but the inquisitive investigator may be more interested in the subtleties of a group. Let us examine a few of the subtleties in the two groups. Group 4 has quite a bit of feedback, what precedes and follows that feedback? The flow chart indicates that student feedback is preceded by the offering of information (35-31 and 15-41) while the facilitator's feedback is preceded by student self-disclosure (36-11). The categories which follow feedback indicate that feedback brings about either silence (31-0, 41-0) or student self-disclosure (31-36, 11-36). These findings by themselves may not

be great insights into the world of group dynamics but in conjunction with similar findings from other groups flow charts these findings could have theoretical, practical, and evaluative applications.

Another important aspect of flow charts is the total picture they give. The flow chart picture in Figure 2 creates an entirely different impression than Figure 1. The frequency of the 15-15 category-pair alone gives the impression that the interactions in group 4 centered around the facilitator's offering information. This is substantiated by the transition cell frequencies (arrows). Only two transitions (35-36, 36-35) do not involve the facilitator. All other transitions either begin or end with the facilitator's information or question.

These are only a few of the interpretations that could be made from these two group sessions. Given a hundred flow charts from groups categorized by effectiveness, objective of each group, type of facilitator, size of the group, etc., an investigator could make innumerable comparisons and perhaps generate hypotheses and information not possible with subjective reports. The possibilities are numerous, now that an instrument is available that quantifies the subtleties that subjective observers always knew were there.

Conclusions

The major purpose of this study was to develop an instrument which would be capable of evaluating interpersonal communication skills in small group settings. The two important criteria in evaluating the development of such a system are the instrument's

reliability and validity. This chapter has presented the results of an examination of the ISIA which demonstrate both the reliability and validity of the instrument. These results are encouraging and invite further research into this area. The possibilities for future research are exciting and the researcher urges others to use the instrument to examine some of the questions concerning group work. But flaws were found. The next chapter will discuss some of the possible modifications of the system, modifications of the training, and possible research questions that could be addressed.

CHAPTER VI

DISCUSSION

Introduction

The focus of the present study has been to bring to light the pressing need for evaluation techniques in the area of group work and interpersonal communications and to develop an observation system that will meet some of those evaluation needs. The primary objective in the development of any evaluative technique is to demonstrate the reliability and validity of the instrument: this has been the fundamental priority of this research. The results are in, the data has been presented and in reflecting upon the outcomes, three areas are of major importance. This chapter will first review the results with an emphasis on the outcomes in relation to some of the major issues posed in prior studies. This will include a critique and recommendation for the procedures. This will be followed by an examination of the observation system, including possible modifications of the ISIA and the training of observers. The study will conclude with suggestions for future research using the ISIA.

Review of Results: Solutions to Old Problems

The review of the literature presented in Chapter II continually pointed out the need for instruments capable of measuring

interpersonal communication skills. Suggestions and recommendations abound with no clear cut answers to the problems. A number of the difficulties discussed in the review of the literature can now be reflected upon in light of the results of the ISIA development. These will be considered here as part of the summary of this study.

Literally dozens of studies referred to in Chapter II call for more objective means of evaluating the changes in participants behavior. Many problems such as the lack of reliability and validity estimates, the global nature of many scales, the inconsistancy of the dependant measure in comparison to the group goals were mentioned. The reliability and validity of the ISIA have been examined and discussed in previous chapters. These results stand alone and are supportive of the future usefulness of the ISIA. But an examination of past difficulties and suggestions will further support the ISIA system.

One of the first and most frequent criticisms of the research and evaluation done on group work has been the subjective nature of the data collected. The review of the literature cites numerous examples of opinionnaires, participant self-report, and other uses of instruments which do not directly examine changes in participants behavior. These subjective reports are highly suspect, but simply using behavioral ratings does not always put the evaluation above reproach. This difficulty is best exemplified by the concern over global rating scales. Studies such as Muehlberg, et al. (1969) point out the limitations of global instruments such as Carkhuff's scale (1969b). The extremely high inference nature of such scales

leaves questions as to what the scale is actually measuring. This is an important area to examine in reviewing the ISIA. Although the ISIA requires some inference on the part of the observer (e.g., the difficulty in coding some affective-cognitive or offering information-self-disclosure transitions), the nature and frequency of those inferences makes the ISIA much less subjective than a global rating scale. This is supported by an examination of the ISIA categories in comparison to the rating points on a scale such as Gazda's (Appendix C). But the real indication of what the ISIA is measuring comes from the validity data. The validation of the ISIA clearly points out what the instrument is measuring; communication skills, the same communication skills that were observed (rated) by facilitators, students, and an expert.

One article which points out the global scale difficulties associated with Carkhuff or Gazda's scales also suggests the usefulness of an instrument such as the ISIA. A quote from Gormally and Hill (1974) seems worthy of repeating as it anticipates so clearly the instrument which has been developed in the present study.

> An alternative to the use of rating scales is to listen to the entire interview and record frequency data, for example, number of responses that identify a feeling, number of non-verbal referents, etc. The response can be identified for simple presence or absence, and the categories are fairly independant. Use of frequency data reduces the subjectivity involved in rating scale measures (p. 544).

This, in many ways, described the ISIA: an instrument which reduces the vast amounts of data in a group to relatively small pieces (by categories) which can be combined (by matrices) to paint the larger picture (through flow charts).

The ISIA along with the IPL and IPL objectives addresses another problem pointed out by a number of authors. Gormally and Hill (1974) and Muehlberg (1969) point out the need for behavioral measures, but behavioral measures by themselves are not enough. In evaluating or researching groups it is important that the dependant measures are consistent with, and a direct measure of, the group goals. Egan (1970) recommends that the answer to this dilemma is to have groups clearly delineate their goals and use these objectives to devise a measure to determine whether the objectives have been met. This is important whether the group is in education (Bowers and Sears, 1961) or in industry or other related fields (Campbell and Dunnette, 1968), the expected behavioral outcomes must be specified and measured. This concern was taken into consideration in the development of the ISIA. The objectives of the IPL (which are similar if not identical to many other educational process groups) were the foundation of the ISIA category development. The validity data presented in Chapter V point out this relationship. The objectives specify for the students and facilitator the types of behaviors to be expected in an effective group and the ISIA confirms that those behaviors are evidenced in groups judged to be effective.

Studies such as those done at the University of Massachusetts (Iver and Rollins, 1970) point out another difficulty in measuring group changes. Precise instruments are important, as are instruments which measure the specific group goals. But unless the instrument measures the changes in participants behavior in the environment

which the behavior would normally be demonstrated, there will be problems in the generalizability of the results. That is, there would be some question as to where an individual would be capable of exhibiting that behavior. There would also be difficulties in doing longitudinal studies on process and product variables. For example, measuring a communication group's effect (pressage variable) on a public school teacher's in-class behavior (process variable) one year later, and the modeling effect her behavior has on her children's behavior (product variable), requires an instrument that can be used in a variety of environments. The nature of the ISIA may substantially aleviate this problem. It can be used in any small group situation where the goals of the group are communication skills. As was pointed out in the instrument description and pilot test chapter, the ISIA (after initial pilot testing) appeared appropriate for use in public school classrooms. This study does not fully explore the appropriateness of the ISIA outside of a college population but initial indications show the potential for research in a variety of settings, particularly public school classrooms.

Procedural Recommendations

These potential answers the ISIA presents, along with the strong reliability and validity data presented in the previous chapter, certainly promise potential for the use of the ISIA in future research and evaluation endeavors. But what else has been learned from the development of this system? A number of procedural

recommendations can be made to strengthen the investigation of reliability and validity, whether the investigation further examines the ISIA or other observation instruments.

Validity

It is difficult to critique the validity procedures used in the present study without tripping over one's own criticism. Chapter II was very critical of the subjective nature of participant's self-reports and questionnaire data. This is a bonafide criticism which makes validating a system difficult because subjective reports are the only criterion available in validating an objective observation system. At the risk of tripping, it must be stated that not only did the validity findings indicate the validity of the ISIA, the findings also gave some support to the subjective opinionnaires, particularly the opinionnaire data of an expert judge.

But some on the opinionnaire data is in contradiction with the ISIA data. Specifically, the opinionnaire data showed no significant difference between effective and ineffective groups in relation to self-disclosure and feedback. In contradiction to those t-test findings, the ISIA data (see Friedman ANOVA, all groups, Table 5) showed a significant difference between all groups on self-disclosure and feedback. This may be a function of the statistics used, but if further analyses were conducted with parallel statistics and differences were found between the opinionnaire and ISIA data, the researcher would conclude that there were differences and that either the participants were unaware of those

differences or the opinionnaire was not sensitive to them. With the data collected no significant differences were found between expert opinion and the ISIA data except for the rating of active listening for one group (see Question #3, Chapter V).

With the exception of this minor inconsistancy the validity procedures used in the present study are viewed as acceptable and somewhat innovative considering the lack of validity estimates reported in the literature. One procedural suggestion is appropriate at this time. The ISIA has demonstrated concurrent criterion-related validity and perhaps could be tested for predictive validity. One way to examine this validity would be to tape a number of groups, having one observer listen and code these groups (decode). These codings would be used to generate matrices and flow charts. The matrices and flow charts could be interpreted and summarized (encoded) by a second person (who is familiar with group dynamics and the ISIA but who has not listened to the groups). These interpretations and tapes could be given to a third person (an expert in group dynamics but not necessarily familiar with the ISIA, since he will not see the matrix or flow charts) who could attempt to pair the tape to the interpretations reported by the second person. Successful pairings would further support the validity to the ISIA.

Reliability

The procedures used in establishing the inter-rater reliability of the ISIA were traditional and acceptable. The training of a larger number of observers would be beneficial but that must depend on

volunteers and/or funds. In regards to the taped versus in-class observations and the stability coefficient two procedural recommendations are appropriate.

In examining the taped versus in-class reliability, future studies should include a number of live codings by a number of observers. This would allow the taped versus in-class observations to be done by different observers. Since the reliability coefficient can be estimated for two observers (on taped groups) an estimate could be made of any loss of reliability. In the present study a high degree of relationship was demonstrated between in-class and taped codings but no loss of reliability could be estimated.

In the present study an estimate was made of the stability coefficient by using a split-half reliability estimate. This was a creative approach to a previously ignored issue in reliability, and although the approach may be questioned it did demonstrate within-class stability. In reexamining this issue a new solution to the problem seems more appropriate. Originally the researcher defined stability in terms of within-group behavior. But taking into account the lawful changes in participants behavior due to the objectives of the group during a particular strategy, a more appropriate approach would be to examine stability by objectives or strategies rather than groups. That is, two different groups which use the same strategy (same objectives) that are both considered effective should be more similar (stable) than two groups using different strategies. This could easily be explored by taping different groups which participate in the same strategy and then

tape the same groups later both using a different strategy. This would demonstrate the stability of the objectives rather than the stability of the group. It would also lend validity to the ISIA system.

This study has examined a number of questions posed by previous research in the area of interpersonal communication skills. The development of the ISIA has opened the door for a more systematic approach to the answering of those questions. The development of the ISIA has also afforded the opportunity to explore the procedures that can be used in developing an observation system. Some recommendations have been made for refining these procedures and some of these recommendations could become areas of research in and of themselves (e.g., Is there stability between different groups using the same strategies?). But before we move on to a discussion of possible research using the ISIA it would be appropriate to discuss the suggested modifications of the ISIA and the ISIA training manual.

Proposed Modification of ISIA and Training

In reviewing the results presented in Chapter V all areas of investigation show the ISIA to be a valuable tool for the research and evaluation of groups. But if one were to choose an area in which to invest more time and energy to improve the ISIA that area would be the inter-rater reliability: Table 1 in Chapter V presents a matrix of inter-rater reliabilities which are acceptable but by no means perfect. This section will discuss proposed changes for the ISIA in two areas that should benefit the reliability, the training of observers and modification to the instrument itself.

An examination of Table 1 reveals that the inter-rater reliabilities are not as highly correlated as was expected. Interviews with each of the observers also revealed that the training took somewhat longer than was anticipated and created frustration in some of the observers. In discussing this problem with the observers it was found that these difficulties centered around one problem, the transition from coding written transcripts to audio tapes. All the observers found that they had no difficulty in mastering the categories and could demonstrate almost perfect reliability in coding written transcripts. But in attempting to code the practice tapes (see practice tapes in Appendix A) they were frequently overwhelmed initially by the need for a coding tempo, the problems of not being able to go back over an interaction, the fidelity of the tape, etc. The research had anticipated this problem and had produced a number of practice tapes by dividing a group recording into shorter five to twenty minute segments. These tape segments were coded by the researcher and the tapes were duplicated so that each observer would have a copy of the tape and the researcher's codes. The observers said these tape segments and codes were helpful, in fact essential (they felt the task would have been impossible without them) but that the transition from written transcripts to real groups was still too difficult.

The discussion of this problem brought out the need for a more gradual shaping of the skill of observation. It is proposed that any future training of observers should incorporate an intermediate step between written transcripts and groups coding. The

major problems in coding an actual group is that group participants talk very fast with no pauses (causing a difficulty in observer's coding tempo) and the communication skills they use are not pure examples of any one skill (causing difficulty in choosing categories, e.g., jumping from affective to cognitive and back, or mixing self-disclosure with feedback and questioning). This could be remedied by producing short (one to two minute) practice tape segments to demonstrate each of the skills. These practice tapes should be recorded in an environment conducive to taping (e.g., a sound studio) to a climate the observer to listening without the interference of extraneous noise. The practice tapes should also be recorded by role playing written scripts so that the observer can be shaped from simple obvious category codings to the more difficult subtle category distinctions. Practice tapes such as these would aleviate a number of the problems voiced by the observers: e.g., pauses would help observers reach a stable coding tempo, typical coding sequences could be pointed out, difficult discriminations (e.g., affective-cognitive) could be pointed out.

This is not to say that coding would become cut and dry or that every possible difficulty could be pointed out. Interpersonal communication is an extremely complex process, everyone communicates in their own unique way, and coding with an instrument such as the ISIA will always be difficult. Reliabilities will never be perfect but that is not to say we should never explore the complex. We must strive to be more specific about these skills

and categories with the aim of more understanding and reliability. But we must also keep in mind our limitations.

Before moving to an examination of possible modifications, of the ISIA system two additional points should be made concerning observer training. A difficulty emphasized by one observer related to the fidelity of the tape. To reliably code group interactions it is important to be able to concentrate on exactly what is said. In the present study no special equipment or rooms were used in the recording of the groups (to minimize group interruptions) but in future studies such accommodations might be helpful. For three of the four observers the fidelity was considered acceptable but all observers noted that the task would be much easier with "perfect" tapes. The last recommendation concerning the training of observer relates to who can be trained. Table 1 shows little or no relationship between experience in groups and reliability. Further research needs to be done before any definitive statement can be made about who can be trained to use the ISIA. But from the present results it seems that anyone familiar with group dynamics has the potential to learn to code group interactions using the ISIA. The ISIA was developed solely for research and evaluation but one side benefit the researcher noted in his own group facilitation was an increased awareness of skills being used in a group and an unexpected refining of processing skills in groups.

The pilot testing of the ISIA showed no need for major modifications of the system: a final examination of the system generally supports that conclusion. The results of the statistical

analysis, interviews with each of the observers, and the researcher's coding of numerous groups generally support the present ISIA categories. Two minor modifications of the system are seen as potentially helpful in increasing the reliability of the observers (by reducing conflicts between categories) and in increasing the discrimination power of the ISIA when comparing groups.

Perhaps the most obvious question to be raised in examining the ISIA data is the absence of irresponsible feedback. This was not discovered or considered until all the data had been collected. The question is difficult to answer but some hypotheses can be examined. The first hypothesis that comes to mind is that there was no irresponsible feedback given in any of the groups. This can quickly be dismissed as that is so unlikely, particularly in view of the participant and expert opinions (they felt there was irresponsible feedback given in the groups. The only other explanation is that the irresponsible feedback that was given was coded using another category. This seems very likely, but the more important question is which other category. By going back over validity tapes the researcher found that what might be viewed as irresponsible feedback was not coded as responsible feedback, rather that feedback was coded offering information (B:5) if it was non-specific, or directs or suggest (B:4) if it was ordering rather than tentative. This creates a dilemna of sorts; is it important that those interactions be coded irresponsible feedback or is it acceptable to leave the distinctions as they stand? Both sides of the issue can be defended but perhaps the solution to the problem is to explain

these fine discriminations (e.g., B:7b rather than B:4 or B:5) in the manual and produce practice tapes which would demonstrate the fine discrimination of such codings.

A related problem in coding feedback is that feedback is defined as showing support or non-support. This is a black and white distinction on paper but unfortunately people in a group interaction frequently give what seems like feedback with no indication whether they are stating support for the stimulus behavior or stating non-support for the stimulus behavior. This problem can best be resolved by creating a new category, Processing or Stating Observations. The feedback which states neither support nor non-support is frequently just an observation with no value put on the observation. Statements such as, "What I see happening is . . ." or "When you got angry last week I saw Bob . . ." are frequently made by facilitators and students. The problem for the observer is that the statement obviously feeds-back to the group what has happened, but to judge the statement as positive or negative requires an inference that may not be appropriate. A category for processing or making observations would aleviate this problem and also more clearly refine the Offering Information (B:5) category, as these observation statements are frequently coded B:5.

The problem with the category for offering Information is that it is a "catch-all" category. It is important to have such a category in an observation system (see Manual Appendix A) but if important information is being lost within the category, revisions of the system should be made. Two problems are seen as

important in examining the Offering Information category. The first problem relates to group processing and observations: these interactions were coded Offering Information (B:5) but with the addition of a new category, Processing or Stating Observations, some of the observer decision making (e.g., B:1 or B:5, B:7 or B:5) would be simplified and inter-rater reliabilities may increase.

The second problem in examining the Offering Information category is the distinction between Offering Information (B:5) and Self-Disclosure (B:6). In categorizing the B:5 and B:6 codes the extreme examples of each category are easy to code, e.g., a lecture is obviously B:5, a group member talking about a tramatic moment in his life and how it effects his present behavior is obviously B:6. But unfortunately a large percentage of the B:5 and B:6 statements are examples of the grey area between the extremes. To increase reliability and to discriminate finer distinctions between groups the researcher proposes that the Self-Disclosure category be expanded somewhat to include some of the "grey areas" between B:5 and B:6. Part of the information gained from observers and coding was that observers were capable of using more categories than the ISIA contains. This ability along with the expansion of the B:6 category leads to the creation of three sub-categories within Self-Disclosure. The proposed categories are: B:6a - Self-Disclosure; past history or incidents ("story-telling"), B:6b - Self-Disclosure; values, opinions or attitudes, and B:6c - Self-Disclosure; relating past experience to the present, feelings, sharing of why I feel or behave as I do. By expanding

the Self-Disclosure category and at the same time breaking it down into sub-categories, the discrimination power of the instrument would increase (being able to distinguish different levels of self-disclosure) and the reliabilities of observers may increase because the B:5 - B:6 distinction would become much clearer.

Suggestions for Future Research

This study has described the development and validation of an observation instrument capable of objectively measuring interpersonal communication skills. Instrument development is a tedious and in some ways unrewarding experience but as Stock (1964) notes, when a researcher is interested in exploring an issue such as interpersonal communications he is frequently faced with the problem of using well established instruments with doubtful or tangential relevance or a tailor-made but untested new instrument. This study has been a beginning toward solving this dilemma.

Research that can come to insightful conclusions tends to be exciting and intrinsically rewarding. Studies that develop the instruments to be used in future research sometimes seem like they will never end and the investigator frequently bathes in the frustration of "I bet that . . ." only to return to the task at hand. For the investigator involved in the present study delay of gratification was sustained by the fantasies of future research to be conducted using the ISIA. As a conclusion to this study a few of the more important questions to be addressed in future research will be explored.

The most important single issue to be investigated in the intensive group movement relates to the change or lack of change in people's behavior due to participation in a group experience. In relation to education groups whose goals are interpersonal communications there is now an instrument capable of measuring those changes. For studies to be carried out in the future the ISIA is a reliable, valid instrument (assuming the group goals are consistent with the ISIA categories) but care and judgment must be used in constructing the design of future research. Pre and post measurements must be encorporated along with control groups and a special emphasis on the objectives and strategies that are demonstrated in the groups being measured. For example, a study might be designed which tapes and codes a number of groups on two strategies, some groups using strategy A early in the term and strategy B late in the term. Other groups would use strategy B early and strategy A late. Control groups could be exposed to both strategies. This design controls for objectives, strategies and group membership (control group) in its measure of change during the term. Designs such as this (c.f., Campbell and Stanley, 1963) would demonstrate change or lack of change in group participants.

Other research is needed which examines the differences between effective and ineffective groups. In the present study the investigator was tempted to draw conclusions about the differences between effective and ineffective groups. But the nature, goals, and design of the present study are such that any conclusions stated as facts would be presumptious. What are needed now are

studies which explore the effects of the objectives of the group, the facilitators role, and the time during the term. One example of such a study would examine whether the facilitator's role changes as a group develops. It might be hypothesized that early in the term the facilitator must take a more active role in a group for it to be considered effective (i.e., model the skills) while later in the term his role would become more of a processor or observer.

The ISIA is capable of discriminating very small, sometimes inconspicuous differences in groups. Matrix and flow charts could illustrate these differences given many groups categorized on a number of variables. Some of the questions might include: (1) What precedes and follows silence? (2) What follows feedback? (3) Is there any difference between what follows student as opposed to facilitator feedback? (4) Is there a most dominant "cycle" in effective groups? (5) Is there any category that is most prevalent preceding affective categories such as a 22-46-22-46 pattern?

These research questions are only a few of the many possibilities that are now open to researchers who are willing to learn an observation system and use it to investigate the many questions that have been "unanswerable." To those educators who have stated that affective education has no place in schools because it was not ameniable to measurement; and to those affective educators who said that affective education was too subtle and important to be put to measurement; you can't hide behind those words anymore. The time has come to bring affective education out in the open, expose it to the light, and examine the impact it has on others.

Conclusion

The original purpose of this study was to review the research concerned with the effects of participation in an intensive group experience and to develop an instrument which would be capable of systematic objective evaluation of educational process groups. An instrument, the Interpersonal Skills Interaction Analysis, was developed and the reliability and validity of the ISIA were examined by field testing the instrument in an educational process laboratory. A series of statistical tests pointed out the extent to which the ISIA may be said to be reliable and valid. These statistical results show the ISIA to have a relatively high degree of both reliability and validity. This study concludes with recommendations for future observation instrument development, possible modification of the ISIA system, and recommendations for future research using the ISIA.

The major conclusion to be drawn from this study is that observation instruments are a valuable source of data in the evaluation of educational process groups. Interaction Analysis techniques such as the ISIA do have limitations (the lack of perfect reliability and exacting training) but their benefits do outweigh the risks and the time involved. It must be concluded that there is a future in the research and evaluation of small group experiences. The time has come for educators involved in affective education and interpersonal communication skills to demonstrate the value of their disciplines in terms of specific behavioral outcomes.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramson, S. A., & Nielson, R. C. Human relations training for campus police. College Management, 1973, 8(9), 21-22.

Allen, D. W., & Cooper, J. M. Model Elementary Teacher Education Program. Final report (HEW, Office of Education #OEC-0-8-089023-3312 (010)). Washington, D.C.: U.S. Government Printing Office, 1967.

Amidon, E. J., & Hough, J. B. (Eds.). Interaction analysis: Theory , research and application. Reading, Mass.: Addison-Wesley, 1967.

Amidon, E. J. & Hunter, E. Improving teaching. New York: Holt Rinehart & Winston, 1966.

Anderson, H. H., & Brewer, H. M. Studies of teachers' classroom personality. I. Dominative and socially integrative behavior of kindergarten teachers. Applied Psychology Monograph, No. 6, 1945.

Anderson, Judith L., & Miner, Patricia. Focus on self development guide. Chicago: Science Research Associates, 1971.

Arbes, B. H., & Hubbell, R. N. A structured communication skills workshop. Journal of Counseling Psychology, 1973, 20, 332-337.

Archer, J., & Kagan, N. Teaching interpersonal relationship skills on campus: A pyramid approach. Journal of Counseling Psychology, 1973, 20, 535-540.

Aspy, D. N. A study of three facilitative conditions and their relationship to the achievement of third grade students. Unpublished doctoral dissertation, University of Kentucky, 1965.

Aspy, D., & Hadlock, W. The effects of high and low functioning teachers upon student performance. In R. R. Carkhuff & B. G. Berenson, Beyond counseling and therapy. New York: Holt, Rinehart and Winston, 1967.

Aspy, D. The effect of teacher-offered conditions of empathy, positive regard, and congruence upon student achievement. Florida Journal of Educational Research, 1969, 11, 39-48.

Bailey, W. J. Student perceived behavioral changes occurring in a secondary school faculty as a result of a human relations inservice workshop. (Doctoral dissertation, Michigan State University.) Ann Arbor, Michigan: University Microfilms, 1967.

Bales, R. F. Interaction process analysis: A method for the study of small groups. Cambridge, Mass.: Addison-Wesley, 1950.

Benjamin, W., et al. Specifications for a comprehensive undergraduate and inservice teacher education program for elementary teachers. Syracuse University, Final Report Project No. 8-9018 OEC-0-8-0918-3313 (010), 1968. (ED 018 677)

Banmen, J., & Capelle, R. Human relations training in three rural Manitoba high schools. Paper presented at the Annual Convention of the National Council on Human Relations, Winnipeg, April, 1971. (ED 052 478)

Barbour, A., & Goldberg, A. Interpersonal communication: Teaching strategies and resources. ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill., 1974. (ED 085 798)

Barker, W. H. Human relations training and the innovative consultant. Paper presented at conference on the Affective Domain of Learning: Implications for Instructional Technology, March, 1969. (ED 034 317)

Becuar, R. J. Skills for effective communication: A guide to building relationships. New York: John Wiley, 1974.

Berenson, David H. The effects of systematic human relations training upon classroom performance of elementary school teachers. Journal of Research and Development in Education, 1971, 4(2), 70-85.

Bessell, H. The content is the medium: The confidence is the message. Psychology Today, 1968, 1(8), 32-35.

Bixler, J. E. Influence of trainer-trainee cognitive similarity on the outcome of systematic human relations training. Unpublished doctoral dissertation, University of Georgia, 1972.

Blakeman, J. D., & Emener, W. G. Training in interpersonal communication: Suggested techniques and approaches. Journal of Research and Development in Education, 1971, 4(2), 36-46.

Borke, H., & Burstyn, J. W. The new teacher and interpersonal relations in the classroom. Journal of Teacher Education, 1970, 21, 378-381.

Bowers, N. D., & Soar, R. S. Studies of human relations in the teaching-learning process. Final report: Evaluation of laboratory human relations training for classroom teachers. Cooperative research project #469, Chapel Hill, North Carolina, 1961.

Bruning, J. L., & Kintz, B. L. Computational handbook of statistics. Glenview, Ill.: Scott, Foresman, 1968.

Buffalo Board of Education, New York. Human relations education project. Final evaluation report. August, 1970. (ED 083 240)

Bunker, D. R. Individual applications of laboratory training. Journal of Applied Behavioral Sciences, 1965, 1(2), 131-148.

Caine, E., & Lindenauer, G. G. Human relations training in teacher education. Journal of Emotional Education, 1973, 13, 27-37.

Calliotte, J. A. Initial attempts at developing appropriate human relations experiences for potential teachers. Paper presented at American Personnel and Guidance Association Convention, March, 1971. (ED 056 004)

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Campbell, J. P., & Dunnette, M. D. Effectiveness of T-group experiences in managerial training and development. Psychological Bulletin, 1968, 70, 73-104.

Carkhuff, R. R. Helping and human relations: A primer for lay and professional helpers. Vol. 1. New York: Holt, Rinehart and Winston, 1969(a).

Carkhuff, R. R. The prediction of the effects of teacher-counselor training: The development of communication and discrimination selection indexes. Counselor Education and Supervision, 1969(b), 8, 265-272.

Carkhuff, R. R., & Bonks, G. Training as a preferred mode of facilitating relations between races and generations. Journal of Counseling Psychology, 1970, 17, 413-418.

Carkhuff, Robert R. Training as a necessary pre-condition of education: The development and generalization of a systematic resource training model. Journal of Research and Development in Education, 1971, 4(2), 3-16.

Carl, M. E., & Jones, L. W. Some practical observations regarding the Minnesota State Human Relations Requirement and its effects on in-service teachers. 1972. (ED 083 202)

Chaney, R., & Passmore, J. L. Affective education: Implications for group process. Contemporary Education, 1971, 42 213-216.

Childers, William C., Jr. An evaluation of the effectiveness of human relations training model using in-class student teacher observation and interaction analysis. (Doctoral dissertation, University of Georgia). Ann Arbor, Michigan: University Microfilms, 1973, No. 74-4768.

Christensen, C. M. Relationship between pupil achievement, pupil affect-need, teacher warmth and teacher permissiveness. Journal of Educational Psychology, 1960, 51, 169-174.

Cogan, M. L. Theory and design of teacher-pupil interaction. Harvard Educational Review, 1956, 26, 315-342.

Corey, S. M., & Corey, E. K. Sensitivity education. Educational Leadership, 1970, 28, 238-240.

Crist, J. Group dynamics and the teacher-student relationship: A review of recent innovations. Research and Development Memorandum #81, Stanford University, School of Education, 1972. (ED 062 292)

Danish, S. J. Factors influencing changes in empathy following a group experience. Journal of Counseling Psychology, 1971, 18, 262-267.

Danish, S. J., & Kagan, N. Measurement of affective sensitivity: Toward a valid measure of interpersonal perception. Journal of Counseling Psychology, 1971, 18, 51-54.

D'Augelli, A. R. Group composition using interpersonal skills: An analogue study on the effects of members' interpersonal skills on peer ratings and group cohesiveness. Journal of Counseling Psychology, 1973, 20, 531-534.

D'Augelli, A. R., & Chinsky, J. M. Interpersonal skills and pre-training: Implications for the use of group procedures for interpersonal learning and for the selection of nonprofessional mental health workers. Journal of Consulting and Clinical Psychology, 1974, 42, 65-72.

Deardon, M. H., & Jekel, J. F. A pilot program in high school drug education utilizing nondirective techniques and sensitivity training. Journal of School Health, 1971, 41, 118-124.

Dendy, R. F. A model for the training of undergraduate residence hall assistants as paraprofessional counselors using videotape playback techniques and Interpersonal Process Recall (IPR). (Doctoral dissertation, Michigan State University.) Ann Arbor, Michigan: University Microfilms, 1971.

Diamond, M. J., & Shapiro, J. L. Toward the long-term scientific study of encounter group phenomena. I. Methodological considerations. Paper presented at the 53rd annual meeting of the Western Psychological Association, April, 1973. (ED 084 465)

Dickman, I. R. Getting it all together (GIAT): Teaching human relationships and communication skills in nursing homes. (ED 098 287)

Dies, R. R., & Sadowsky, R. A brief encounter group experience and social relations in a dormitory. Journal of Counseling Psychology, 1974, 21, 112-115.

Dinkmeyer, D. Top priority: Understanding self and others. Elementary School Journal, 1972, 72, 62-71.

Dixon, W. R., & Morse, W. C. The prediction of teaching performance: Empathetic potential. Journal of Teacher Education, 1961, 12, 323-329.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, and Winston, 1974.

Egan, G. Encounter: Group process for interpersonal growth. Belmont, Calif.: Brooks/Cole Pub., 1970.

Fattu, N. A. Nine proposals for elementary teacher education. A description of plans to design exemplary training programs, 1968. (ED 018 677)

Flanders, N. A. The problems of observer training and reliability. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, research and application. Reading, Mass.: Addison-Wesley, 1967.

Flanders, N. A. Teacher influence: Pupil attitudes and achievement. Final Report, Cooperative Research Project No. 397, U.S. Office of Education. The University of Minnesota, 1960.

Flanders, N. A. Analyzing teaching behavior. Reading, Mass.: Addison-Wesley, 1970.

Forman, J. R. (Ed.). Human relations handbook. Suggestions for improving interpersonal relationships in West Virginia Public Schools. A guide prepared for West Virginia educators. (West Virginia State Department of Education, Charleston, West Virginia, 1968.) (ED 099 439)

Galvin, K. M. Communication approaches for the intergroup educator. Massachusetts Communication Journal, 1974, 6(1), 31-46. (ED 095 599)

Gazda, G., Asbury, F. R., Balzer, F. I., Childers, W. C., Desselle, R. E., & Walters, R. P. Human relations development: A manual for educators. Boston: Allyn and Bacon, 1973(a).

Gazda, G., Asbury, F. R., Balzer, F. J., Childers, W. C., Desselle, R. E., & Walters, R. P. Instructor's manual to accompany: Human relations development: A manual for educators. Boston: Allyn and Bacon, 1973(b).

Glasser, W. Schools without failure. New York: Harper and Row, 1969.

Goodman, G. An experiment with companionship therapy: College students and troubled boys--assumptions, selections, and design. In B. G. Gurney (Ed.), Psychotherapeutic agents: New roles for nonprofessionals, parents and teachers. New York: Holt, Rinehart and Winston, 1969.

Gordon, T. Parent effectiveness training: The "no-lose" program for raising responsible children. New York: Peter W. Wyden, Inc., 1970.

Gormally, J., & Hill, C. E. Guidelines for research on Carkhuff's training model. Journal of Counseling Psychology, 1974, 21, 539-547.

Gross, T., Lepper, C., & Rappaport, J. Effects of sensitivity training, video-taped modeling and instruction on group verbal behavior. Paper presented at the annual meeting of the Midwestern Psychological Association, May, 1971. (ED 052 479)

Harrison, C. T. The teacher and the T-group. Scholastic Teacher, 1971, 24, 7-8.

Harrison, R. Problems in the design and interpretation of research on human relations training. Explorations in Human Relations Training and Research, 1967, 1, 1-9.

Harrison, R. Research on human relations training: Design and interpretation. Journal of Applied Behavioral Science, 1971, 7, 71-85.

Hartzell, R. E., Anthony, W. A., & Wain, H. J. Comparative effectiveness on human relations training for elementary student teachers. The Journal of Educational Research, 1973, 66, 457-461.

Hatfield, D. Teacher education and functional humanity. A position paper of the Task Force for Human Relations in Teacher Training-Minnesota. Journal of Intergroup Relations, 1972, 1(4), 38-45.

Heck, E. J. A training and research model for investigating the effects of sensitivity training for teachers. Journal of Teacher Education, 1971, 22, 501-507.

Hefele, T. L. The role of teacher interpersonal communication factors on the academic achievement of deaf students. Unpublished doctoral dissertation, State University of New York at Buffalo, 1969.

Hendrickson, S. Group interaction for inner-city teachers. Elementary School Journal, 1968, 68, 400-405.

Heiman, H. Teaching interpersonal communications. North Dakota Speech and Theatre Association Bulletin, 1974, 2(2), 7-29. (ED 097 745)

Herbert, J., & Attridge, C. A guide for developers and users of observation systems and manuals. American Educational Research Journal, 1975, 12, 1-20.

Hornsby, J. L. The effects of group composition on systematic human relations training. Unpublished doctoral dissertation, University of Georgia, 1973.

Hough, J. B. An observation system for the analysis of classroom instruction. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, reserach and application. Reading, Mass.: Addison-Wesley, 1967.

Hunt, D. E. The communication task. 1965. Cited by Hunt, D., Dirr, P., & Joyce, B. Sensitivity training for teachers: An experiment. The Journal of Teacher Education, 1969, 20, 75-83.

Hunt, D., Dirr, D., & Joyce, B. Sensitivity training for teachers: An experiment. The Journal of Teacher Education, 1969, 20, 75-83.

Ivey, A. E., & Rollins, S. A. The human relations performance curriculum: A commitment to intentionality. University of Massachusetts, 1970. (ED 043 060)

Ivey, Allen E., Rollins, S. A., Cooper, J. M., Schleiderer, A., & Gluckstern, N. Human interaction: A behavioral objective curriculum in human relations. University of Massachusetts, 1970. (ED 051 113)

Ivey, A. E., & Rollins, S. A. A behavioral objectives curriculum in human relations: A commitment to intentionality. Journal of Teacher Education, 1972, 23, 161-165.

Johnson, D. W. Reading out: Interpersonal effectiveness and self-actualization. Englewood Cliffs, N.J. Prentice-Hall, 1972.

Jourard, S. M. The transparent self: Self-disclosure and well-being. Princeton, N.J.: Van Nostrand Co., 1964.

Jourard, S. M. Self-disclosure: An experimental analysis of the transparent self. New York: John Wiley, 1971.

Kagan, N., & Krathwohl, D. R. Studies in human interaction (Final report, Grant OE 7-32-0410-270). East Lansing, Mich.: Educational Publication Services, 1967.

Khana, J. L. An evaluation of the human relations program. Project Upper Cumberland, Final report, Title III ESEA. Contract No. 67-03525, Livingston, Tennessee, 1969.

Kimple, J. A. Innovative summer school. National Educational Association Journal, 1968, 57(3), 28-29.

Kimple, J. A. How South Brunswick schools developed an in-service training program. Nation's Schools, 1969, 83(3), 85-86.

Kimple, J. A. Sensitivity: A superintendent's view. Educational Leadership, 1970, 23(3), 266-269.

Koile, E. A., & Gallessich, J. The Dallas human relations (abs. Junior College Journal, 1971, 41(6), 31-37.

Kraft, L. J. The influence of human relations laboratory training upon the perceived behavioral changes of secondary school seminar instructors. (Doctoral dissertation, Michigan State University.) Ann Arbor, Michigan: University Microfilms, 1967.

Lee, W. S. A study of the effectiveness of sensitivity training in an in-service teacher-training program in human relations. Dissertation Abstracts, 1967, 28(5), 1680A.

Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created "social climates." Journal of Social Psychology, 1939, 10, 271-299.

Lippitt, R. Sensitivity training: What is it? How can it help students, teachers, and administrators? Childhood Education, 1970, 46, 311-313.

MacDougal, D. B. The classroom council: A method for improvement of interpersonal classroom relationships. Elementary School Guidance and Counseling, 1973, 7, 244-246.

Mann, R. D. Interpersonal styles and group development: An analysis of the member-leader relationship. New York: Wiley, 1967.

Marshall, S. A. Sensitivity training: A report. Educational Leadership, 1970, 28, 250-253.

Mason, J., & Blumberg, A. Perceived educational value of the classroom and teacher-pupil relationships. Journal of Secondary Education, 1969, 44, 135-139.

McElvaney, C., Brozgal, J., Miller, N., & Watson, G. Preliminary report of session I: Human Relations Laboratory for School Administrators. August, 1967. (ED 017 073)

Meador, B. D. Individual process in a basic encounter group. Journal of Counseling Psychology, 1971, 18, 70-76.

Medley, D. M., & Mitzel, H. E. A technique for measuring classroom behavior. Journal of Educational Psychology, 1958, 49, 86-92.

Miles, M. B. Human relations training: Process and outcomes. Journal of Counseling Psychology, 1960, 7, 301-306.

Miles, M. B. The T-group and the classroom. In L. D. Bradford, J. R. Gibb, & K. D. Beene (Eds.). T-group theory and laboratory method. Innovation in re-education. New York: Wiley, 1964.

Miles, M. B. Changes during and following lab training: A clinical experimental study. Journal of Applied Behavioral Sciences, 1965, 1, 215-242.

Muzingo, D. A mini-course in interpersonal communication in a junior high trimester program: Teacher's potpourri. Speech Teacher, 1974, 23, 144-149.

Nelson, H. Impact of change in attitudes under three variations of sensitivity training. Final report, Colonie Central School District, Albany, New York, July, 1969. (ED 039 193)

Nicholas, W. R. Group exploration of educational beliefs, values, and interpersonal relationships, a descriptive study. (Doctoral dissertation, Michigan State University.) Ann Arbor, Michigan: University Microfilms, 1972.

Nylen, D., Mitchell, J., & Stout, A. Handbook of staff development and human relations training: Materials developed for use in Africa. Washington, D.C.: National Training Laboratories Institute for Applied Behavioral Sciences, 1967.

Ober, R. L. Predicting student teacher verbal behavior. Unpublished doctoral dissertation, Ohio State University, 1966.

Ober, R. L., Bentley, E. L., & Miller, E. Systematic observation of teaching: An interaction analysis-instructional strategy approach. Englewood Cliffs, N.J.: Prentice Hall, 1971.

O'Donnell, D. J., & Maxwell, K. Reality therapy works here. Instructor, 1971, 80(7), 70-73.

O'Hare, M. R. Sensitivity training in teacher education, 1968. (ED 028 959)

Opinion Poll, Schoolmen on the fence about sensitivity training. Nation's Schools, 1970, 85, 43.

Palomares, U. H. Communication begins with attitude. The National Elementary Principal, 1970, 50(2), 47-49.

Perkins, S. R., & Atkinson, D. R. Effects of selected techniques for training resident assistants in human relations skills. Journal of Counseling Psychology, 1973, 20, 84-90.

Peterson, V., Calvin, R., & Yutzy, J. Human relations training, contracting and field experiences: An integrative approach to the teaching of an educational foundations course. Teacher Education Forum, 1973, 2(3), 1-12.

Pierce, R., & Schauble, P. Graduate training of facilitative counselors: The effects of individual supervision. Journal of Counseling Psychology, 1970.

Price, R. A. Final report: First year human relations project. (A cooperative project sponsored jointly by the Syracuse Public Schools and the Social Studies Curriculum Center, Syracuse University.) Syracuse, N.Y.: Syracuse University, Social Studies Curriculum Center, 1969. (ED 032 595)

Radke, M. W. A follow up survey of education 200, the individual and the school, examining harm and growth in students and how useful they perceive the course concepts. Research report #17. East Lansing, Mich.: College of Education, Michigan State University, 1975.

Rappaport, J., Gross, T., & Lepper, C. Modeling, sensitivity training, and instruction: Implications for the training of college student volunteers and for outcome research. Journal of Consulting and Clinical Psychology, 1973, 40 99-107.

Ratliffe, S. Non-print resources for teaching interpersonal communications in the middle school: Reviews of teaching/ learning resources. Speech Teacher, 1974, 23, 173-174.

Roelofs, B., & Sears, C. Awareness groups for prospective teachers. Journal of the Student Personnel Association for Teacher Education, 1971, 9(3), 67-70.

Rogers, C.R. The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology, 1957, 21, 95-103.

Rogers, C. R., & Rablen, R. A. A scale of process in psychotherapy. 1958. Cited by B. D. Meador. Individual process in a basic encounter group. Journal of Counseling Psychology, 1971, 18, 70-76.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In R. M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Schmuck, R. A. Helping teachers improve classroom group process. Journal of Applied Behavioral Science, 1968, 4, 401-435.

Schroeder, K., et al. Systematic human relations training for resident assistants. Journal of College Student Personnel, 1973, 14, 313-316.

Schulman, F. TA decision making styles. Research report #16. East Lansing, Michigan: College of Education, Michigan State University, 1974.

Schulman, F., & Byers, J. L. Development and analysis of a student report evaluation form for an introductory education course. Research report #14. East Lansing, Michigan: College of Education, Michigan State University, 1974.

Schutz, W. C. FIRO: A three dimensional theory of interpersonal behavior. New York: Holt, Rinehart & Winston, 1958.

Shostrom, E. L. An inventory for the measurement of self-actualization. Educational and Psychological Measurement, 1964, 24, 207-218.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Simon, A., & Boyer, E. G. (Eds.). Mirrors for behavior: An anthology of classroom observation instruments. Supplementary Vols. A and B. Philadelphia: Research for Better Schools, 1970. (ED 042 937)

Simon, D., & Sarkotich, D. Sensitivity training in the classroom. Today's Education, 1967, 56(1), 12-13.

Solomon, A., Perry, S., & Devine, R. Interpersonal communication: A cross-disciplinary approach. Springfield, Ill.: Charles C. Thomas Publisher, 1970.

Solomon, D., Rosenberg, L., & Bezdek, W. Teacher behavior and student learning. Journal of Educational Psychology, 1964, 55(1), 23-30.

Southern Regional Education Board, Atlanta, Georgia. Beyond the three R's. Training teachers for affective education. Sponsored by the National Institute of Drug Abuse. (ED 096 581)

Sperber, S. M. Development of a training program for teachers using interpersonal process, communications, and problem solving. (Doctoral dissertation, University of Pittsburgh.) Ann Arbor, Michigan: University Microfilms, 1972, No. 72-22-310.

Sponberg, R. A. Talent awareness training: A program that considers and changes teacher performance in terms of behavioral goals. Instructor, 1969, 78(9), 61.

Springport High School, Michigan. Human relations laboratory training student handbook. 1967. (ED 018 834)

Stiggins, R. J. Stability of semantic meaning space and change in concept meaning during teacher training. (Doctoral dissertation, Michigan State University.) Ann Arbor, Michigan: University Microfilms, 1972.

Stock, D. A survey of research on T-groups. In L. D. Bradford, J. R. Gibb, & K. D. Benne (Eds.), T-group theory and laboratory method: Innovation in re-education. New York: Wiley, 1964.

Stoffer, D. L. Investigation of positive behavior change as a function of genuineness, warmth, and empathetic understanding. Journal of Educational Research, 1970, 63, 225-228.

Sweeney, D. B. T-group method of sensitivity training: Its effects on teacher attitudes as indicated on the Minnesota Teacher Attitude Inventory. Report of a study of the Three Village Central School District #1. Setauket, New York, June, 1969. (ED 038 372)

Taylor, M., & Barnes, J. Developing professional leadership to create a climate for individualizing and vitalizing instruction. Final report for Cycle I of the project for developing professional leadership to create a climate for individualizing and vitalizing instruction. Atlanta, Georgia, 1970.

Thorman, J. H. Relative effectiveness of four methods of training prospective teachers in interpersonal skills. The Journal of Educational Research, 1971, 65, 19-21.

Truax, C. B. The use of supportive personnel in rehabilitation counseling: Process and outcome in supportive personnel in rehabilitation centers: Current practices and future needs. Pittsburgh, Pa., and Colorado Springs, Co., 1967.

Truax, C. B., & Tatum, C. R. An extension from the effective psychotherapeutic model to constructive personality change in preschool children. Childhood Education, 1966, 42, 456-462.

Trubowitz, S. The listening teacher. Childhood Education, 1975, 51, 319-322.

Tuccy, J. A. SPSS 5.8 manual supplement: subprogram npar tests (non-parametric statistical tests.) Wrubel computing center, report #273. Bloomington, Indiana, 1974.

VanCamp, S. S. A human relations curriculum. Childhood Education, 1973, 50, 73-76.

Wallen, J. L. Appendix R: A basic communication skill for improving interpersonal relationships. Northwest Regional Educational Laboratory. Portland, Oregon, 1968. (ED 026 323)

Weaver, R. L. Sensitivity training and effective group discussion. Speech Teacher, 1971, 20, 203-207.

Wehling, L. J., & Charters, W. W., Jr. Dimensions of teacher beliefs about the teaching process. American Educational Research Journal, 1969, 6, 7-30.

Wells, H. G. To get beyond the words. Educational Leadership, 1970, 28, 241-244.

Whalen, C. Effects of a model and instructions on group verbal behavior. Journal of Consulting and Clinical Psychology, 1969, 33, 509-521.

Wiggins, T. W. Sensitivity training: Salvation or conspiracy? Educational Leadership, 1970, 28, 254-257.

Withall, J. The development of a technique for the measurement of social-emotional climate in classrooms. Journal of Experimental Education, 1949, 17, 347-361.

APPENDICES

APPENDIX A

ISIA TRAINING MANUAL

ISIA TRAINING MANUAL

(Research on IPL Skills)

The Interpersonal Skills Interaction Analysis (ISIA) is an observation tool to examine the communication skills of people in small group setting. It is primarily designed to be used to investigate the communication skills of groups whose objective or goal is to teach or practice those communications skills. It is, in effect, an evaluation tool or test of the group members interpersonal communication skills just as a math test evaluates students math skills in a classroom. Unlike a math test, which is usually given to individuals, the ISIA evaluates the entire groups skills as a whole and not the skills of particular individuals.

The ISIA is one of a number of techniques for measuring verbal interactions based on the notion that verbal interaction can be perceived as a series of events which occur one after another.[1] Each event occupies a small segment of time, and the chain of events can be spaced along a time dimension. The analysis of this chain of events is called an interaction analysis. The individual segments of the chair are called events, an event being defined as the shortest possible act that a trained observer can identify and record. Interaction analysis is defined as any technique for studying this chain

---

[1]A great deal of the introductory information contained in this manual is similar to that of the FIAC, much of this is paraphrased from Flanders, (1970). Credit must be given to Ned Flander for his trail blazing in the area of interaction analysis.

or interaction events in such a fashion that each event is taken into consideration.

This manual will attempt to prepare you to use one interaction analysis technique, to become a trained observer of communication skills using the Interpersonal Skills Interaction Analysis. You will be introduced to the 33 possible ISIA categories. Each of these categories will be defined and examples will be presented. You will be given guidelines and rules of thumb to help you distinguish each of the categories. After this initial exposure to the categories you will be given practice examples to help you distinguish the categories. The practice examples will be done by groups of categories so that you will not be overpowered by all 33 categories at once. After doing these module-exercises you will be given a transcript exercise that will let you use all 33 categories on a short written interaction. The last step will be to listen to and code a tape of a group interaction. All these exercises are part of this manual and are self checking.

At first glance this may seem like an impossible task, particularly when you realize that an experienced observer (coder) should identify events and code at approximately one code every five seconds. If that is discouraging to you take heart and have patience, I felt the same way and I developed the system. To begin with it may be somewhat frustrating but taken bit by bit it is not nearly as difficult as it first sounds, you will probably become a skilled observer using the ISIA in less than 20 hours depending upon your present understanding of communication skills and your own persistance.

In learning to use the ISIA the most important skill is being able to distinguish and label categories accurately but first I'd like to explain something of the product of interaction analysis to give you a total picture of the system you'll be using. Using the ISIA, a trained observer sits in the classroom or group, or views a video-sound playback, or just listens to an audio-recording and keeps a record of the flow of events on an observation form. The observer may punch the keys of some mechanical devise but in most situations s/he will simply keep track on a sheet similar to the one on page 40. S/he is trained to use a set of categories. S/he decides which category best represents each event and then writes down the code symbol of that category. The recorder's speed of coding will vary but with the ISIA it has been found that a coding tempo of about one code every 5 seconds is a comfortable speed for most observers. The product of the observation is a long list of code symbols, one symbol to one event in chronological order. The list of codes can then be analyzed for the frequency of certain codes and also for patterns within the list by use of a interaction matrix which will be discussed much later. One of the features of this type of system is that it ignores the subject matter people are discussing and focuses of the communication skills they are using.

ISIA Categories

The ISIA is a multiple coding category system. It is called multiple coding because it requires more than one code per event.

The codes (in most instances two codes per event) come from two category clusters; the first category cluster (cluster A) identifies who is speaking (either teacher, student, or silence) and in what domain the speaker is talking (either cognitive or affective). The second category cluster (cluster B) identifies what particular communication skill the speaker is using. Every event is coded with one code from cluster A and one code from cluster B (there are two exceptions to this that will be explained later). The categories are listed below according to clusters.

Cluster A	Cluster B
0-Silence or confusion	1. Positive Feedback
1-Teacher - cognitive	a. responsible
2-Teacher - affective	b. irresponsible
3-Student - cognitive	2. Active Listening
4-Student - affective	3. Elicits Information
	4. Directs or Suggests
	5. Offers Information
	6. Self-Description
	7. Negative Feedback
	a. responsible
	b. irresponsible

A few examples would probably be helpful. The following chain of events might occur early in a class: 0, 15, 15, 32, 15, 14. Notice that right away we have one of the exceptions to the rule about one code from each cluster. The code for silence or confusion consists of a code from cluster A (0) but no code from cluster B, because silence or confusion indicates no particular communication skill (at least no skill that the ISIA measures). Silence or confusion is the only event which is coded with a single diget.

Following the 0 code is the code 15. This indicates the offering of cognitive information by the teacher; A:1[2] indicates teacher talk cognitive, B:5 indicates the offering of information.

The third coded event is another 15. This shows that the cognitive information offered by the teacher lasted longer than five seconds. Once an observer is trained s/he will have a coding tempo of about one code every five seconds and if the teacher were to go on offering cognitive information for a whole minute the observer would record twelve consecutive 15's (one every five seconds for sixty seconds equals twelve). This indicates the duration of the interaction.

The forth coded event is a 32 indicating cognitive active listening by a student; the A:3 indicates student talk, cognitive, the B:2 indicates active listening.

The fifth coded event is another 15, cognitive information offered by the teacher. And the sixth coded event is a 14 indicating the teacher directed or suggested something cognitive. This is a ypical chain of events and might be interpreted to mean the teacher was lecturing about something, a student might have asked a question to which the teacher replied with more information and then the teacher directed or suggested the class (or a student) to do something. This is rather typical interaction pattern that indicates the type of communication skills that were used in a classroom. As is typical of interaction analysis it ignores the topic of conversation

---

[2]The symbol "A:1 indicates the 1 code in category A," "B:5" indicates the 5 code in category B, "15" indicates the first code in category A and the fifth code in category B.

and focuses in on the types of interactions that occurred. That is, in this example we have no idea what the people were talking about but we do know what kind of interaction was taking place. We also know another bit of information. If the person who was coding this interaction, coded at about one code every five seconds then this interaction took about thirty seconds. That is important information that can only be obtained if the coder has what is called a coding tempo, that is if s/he codes at a constant rate.

You have now been presented an overview of the system and an outline of what the categories are and how they are recorded. Before we move on to specific definitions of the categories I must explain one last point about interaction analysis. Systematic coding of complex behavior such as verbal interaction is a skill that takes practice. Do not read this manual casually and expect to be an expert in an hour, it won't happen. The single most important factor in you becoming an expert in using the ISIA is practice, your persistance in going over examples and transcripts and tapes. You can expect to devote about four hours or more to practicing before the various skills involved begin to come under control. These initial efforts may be frustrating and are not always rewarding but practice will pay off and you should feel comfortable using the ISIA by the tenth to fifteenth hour.

It is now time to define each of the categories, present some examples of each of the categories, give some ground rules for discriminating between the categories, and give you a chance to show yourself what you learned (or didn't learn) in what will be

called Training Modules. You'll start with Cluster A and learn the distinction between affective and cognitive.

TRAINING MODULE I

Cluster A

At first glance Cluster A seems like the easier of the two clusters to master when learning to code the ISIA and in many ways this is true. The A:0 category is very straight forward and the teacher-student distinction is obvious once you know the voice of the teacher. But the cognitive-affective distinction can be very tricky. Of all the categories in the ISIA the cognitive-affective distinction is the most subjective. Because of this most of the exercises in this module will be presented to help you make that distinction. Below are the definitions of each of the categories in Cluster A.

0-Silence or Confusion.--This category includes pauses, short periods of silence, or periods of confusion or laughter in which the observer cannot understand the interaction clearly enough to code it. In listening to tapes or sitting in on live group meetings this code is used for those silent periods or times when confusion (e.g., laughter, everyone talking at once, etc.) makes coding impossible. When coding transcripts or doing exercises in this manual the symbols ##### will stand for silence or confusion, each # representing a second of silence. Therefore ##### would be coded 0, ##### ##### would be coded 0, 0 and so forth, a single 0 representing five seconds

of silence. Remember coding is done in five second intervals (at least you're shooting for a tempo of approximately one code every five seconds) so the examples and transcripts will indicate that interval and ask you to code an event that represents approximately five seconds.

A:1 & A:2-Teacher.--Categories 1 & 2 from Cluster A both represent verbal statements of the teacher or group leader, category 1 is used for teacher-cognitive and category 2 for teacher-effective. In coding audio-tapes it is the responsibility of the person doing the taping to identify for the coders which individual in the group (which voice) is the teacher.[3] In transcripts and examples the teacher statements will be identified by a T.

A:3 & A:4-Student.--Categories 3 & 4 both represent verbal statement of the student or group member, category 3 is used for student-cognitive and category 4 for student-affective. In transcripts and examples the student statements will be identified by an S.

A:1 & A:3-Cognitive.--Categories 1 & 3 from Cluster A both represent verbal statements which have a factual or cognitive input, Category 1 is used for teacher-cognitive and category 3 for student-cognitive. Cognitive statements are related to knowledge, the process of knowing. Statements which are coded cognitive are the presentation of how the person thinks about something as opposed to

---

[3]Teacher will represent teacher and/or group leader, student will represent student and/or group member unless otherwise specified.

how they feel about it. The coder must be sensitive to the fact that people often use the phrase "I feel . . ." when the meaning of the statement refers to thoughts, e.g., S: "I feel the major decisions should be made by the teacher." This is a cognitive statement even though the student says "I feel . . ." because the statement refers to what the student thinks and could also have been stated "I think . . ." This example would be coded A:3, student-cognitive. For the example to be affective it would have to include his feelings about the teacher making the decisions such as S: "I feel comfortable with the teacher making the major decisions." Many statements that people make that begin with "I feel . . ." could just as well be stated "I think . . ." but the reverse is not true. When you are stating how you feel you cannot begin the statement with "I think. . . ." If you are frustrated about reading this manual it wouldn't make sense to say "I think frustrated. . . ." Statements which do not include how someone is feeling are coded cognitive. All statements are either cognitive or affective and it would seem that the easiest way to define cognitive (and easiest way to remember it) is to say that cognitive statements are those statements which are not affective. This may seem useless until one realized that affective statements have a much clearer criteria.

A:2 & A:4-Affective.--Categories 2 & 4 both represent verbal statements which refer to feelings; category 2 is used for teacher-affective, and category 4 is used for student-affective. Affective statements may refer to either the speaker's feelings or the feelings of someone else. They often begin with "I feel . . ." or contain

words that refer to affective states such as love, hate, upset, shy, excited, foolish, frustrated, etc. Whenever a statement refers to how the speaker is feeling or when the speaker refers to how someone else is feeling the event is coded A:2 or A:4. One of the keys to coding affective statements is to look (or listen) for feeling words.

Module I, Example-Exercise

OK, you've been exposed to the definitions and a few examples. Now you probably need some more examples so you can decide if you need to re-read this section. Below are a number of examples on the left hand side of the page. On the right hand side of the page are a number of blanks and the answers. You may choose to fold the edge of the page under to use these examples as a quiz. In this way you may use the examples as a way of evaluating what you've learned up to now. There is an explanation on following each paragraph to explain the answers.

Notice that the codes are written A:1 or A:5 because these aren't complete codes, to be complete codes they would need a number from cluster B (except the code for silence or confusion but you knew that was the single code 0 anyhow, didn't you?).

T:	"I'm really glad you all showed up on time today." (1)	(1) _____	A:2
S:	"I'm afraid to go back to my history class." (2)	(2) _____	A:4
	##### (3)	(3) _____	0
S:	"I think the class meets at 10:00." (4)	(4) _____	A:3
T:	"Tomorrow we'll talk about feelings."(5)	(5) _____	A:1

In the first example the teacher talks about his feelings of being glad which is affective (A:2). The second example states how the student is feeling about going to history class (A:4). The next example is five seconds of silence (0). The forth example is a statement of fact, when the class will meet (A:3). The last example is also a statement of fact by the teacher (A:1). This example points out how feelings can be talked about cognitively, that is the teacher is not referring to any individual's feelings but rather points out the fact that the class will talk about feelings. This is important to remember, to talk about the concept of feelings is congitive. When I wrote about categories A:2 and A:4 I was talking about facts and not how I felt about those facts. If I had written that I was frustrated writing about them that would be coded affective, it explains my feelings.

T: "Bob, how did you feel when Jane said that? (6)	(6)	_____	A:2
S: "Well, I felt she shouldn't have said it." (7)	(7)	_____	A:3
T: "But how did you feel?" (8)	(8)	_____	A:2
S: "Oh, pretty threatened. I was afraid."(9)	(9)	_____	A:4
S: ## "I'm glad you said that Bob." (10)	(10)	_____	A:4

(6) is an example of someone asking for feelings from some one else (A:2). The reply to the question (7) includes the word "felt" but it is a factual reply, a cognitive answer. This implies feelings but does not state how Bob felt and is therefore cognitive rather than affective. The teachers next question (8) refers again to feelings (A:2) and Bob's answer is affective this time (A:4).

In the last example (10) there is two seconds of silence (##) before the student's affective response (A:4). Unless the silence is an entire coding event (approximately five seconds) do not code the brief silence.

Before going on to the next section there are some helpful hints that may benefit you in your coding. It is important to remember to code statements as the receiver hears them rather than as you might expect the speaker to mean them. That is, stay away from inferring what is not said. Going along with that remember that statements are coded cognitive unless the statement actually refers to the speaker's feelings or the feelings of someone else. There is one exception to this rule. In situations where there has been a flow of conversation that has been all affective do not change to coding the statements cognitive unless the subject is changed. That is, in a situation where the group knows the topic of conversation is about someone's feelings the speaker does not need to label that feeling all the time for the statement to be coded affective. This leads us into the first ground rule for the ISIA.

Ground Rule 1

When using any observation scheme there will be times when the coder faces a certain amount of reasonable doubt in categorizing statements no matter how specific the category definitions are. The first ground rule that when doubt exists the coder will classify doubtful statements into categories which are consistant with the prevailing balance of the interaction. That is to say if you have

just coded ten 26 in a row and now you are unsure whether the present statement is a 25 or a 26 go with the flow of the conversation which would mean code it 26. It is important to define reasonable doubt to mean that you can't decide between two equally possible categories.

TRAINING MODULE II

Cluster B

You should now have a pretty good grasp of cluster A, you should be able to identify who is speaking and in what domain. We now move on to an examination of what particular communication skill the speaker is using. There are seven categories (nine including responsible and irresponsible feedback) to examine in this module and at the risk of confusing you we will not go through them in order. Rather we will leave B:1 - Positive Feedback til the last to be explained with B:7 Negative Feedback. The other five categories will be done in two groups: Active Listening (B:2), Elicits Information (B:3) and Directs or Suggests (B:4) will be explained, examples given, and an example-exercise on the three categories will be presented, then Offering Information (B:5) and Self-Disclosure (B:6) will be presented in the same way. Finally Feedback will be presented.

B:2-Active Listening.--For a statement to be coded B:2 it must meet two criteria: First the statement should keep the focus of the interaction the person who has previously been talking and

on the subject that person was talking about. Second, the statement must encourage the previous speaker to go further in the interaction, to elaborate on his/her ideas or feelings. Active listening serves as a reinforcement to the speaker in that it is communicating to the speaker that he has been heard and that someone is interested in what he said and would like to hear more. There are generally three types of statements which are coded as active listening; paraphrasing, clarifying statements, and exploratory statements.

Paraphrasing involves reflecting back to the speaker his/her ideas or feelings. Paraphrasing communicates an understanding which in effect reinforces the person to continue with his feedback, self-disclosure, or his offering of information. An example of this might be statements such as S: "It really makes you angry when a number of people are absent." (B:2) or T: "You wish I'd leave you alone." (B:2) These statements reflect back to the person what he has said and allow him to elaborate on his statement if he wishes to do so.

Active listening often times takes the form of a question. Clarifying questions are questions from a listener to the speaker which seek out further information, e.g., T: "I'm not sure I understood what you mean when you say you were upset about?" (B:2) Clarifying questions in effect tell the speaker that the listener is interested in what the speaker is talking about but that the listener does not completely understand the speakers message.

Active listening can also take the form of an exploratory question. Exploratory statements or questions encourage the speaker

to expand on his meaning or feelings at a deeper level. They often acknowledge the stated meaning or feelings and then ask the speaker to elaborate more fully the facts and the relationship of those facts, or to examine deeper feelings. Examples T: "So it is important for a teacher to model self-disclosure. Is there any other reason for a teacher to be self-disclosing?" (B:2) or S: "So they just forgot all about you and you were mad. Were you feeling anything else?" (B:2) Active listening is often times a question in an on-going interaction, a question that encourages the speaker to continue sharing the facts or feelings he was presenting. This is important as it distinguishes active listening from the next category.

B:3-Elicits Information.--This type of talk asks a question or requests information about the content, process, or subject of the group with the intent that another should answer. This category differs from active listening in two ways. First, the intent of active listening is to reinforce or encourage a previous speaker to talk more, while the intent of eliciting information is to secure information for the person asking the question. This is not presented to say that the intent of the question is the major criteria in making the distinction between B:2 and B:3 statements but it is something to keep in mind. The major criteria for making the B:2 - B:3 distinction has to do with when the question is asked, what is the context in which the question is asked?

If a question is part of an interaction where a person is talking about themselves and the question encourages them to talk

more about that subject it is active listening. If a question starts a conversation, or changes the focus of the conversation to another subject or person it is coded B:3, elicits information. Example: T: "Bob, how would you define responsible negative feedback." (B:3) But if Bob had been talking about responsible negative feedback and the teacher asked him what that meant to him or said something like T: "Bob, I'm not sure what you are meaning by that, could you define responsible negative feedback for me." (B:2) Then, it would be coded active listening because it is then a clarifying question. The context is very important.

B:4-Directs or Suggests Solutions.--This type of talk gives directions, instructions, orders, or assignments with which another is expected to comply. Statements which are parts of an interaction made up of active listening and self-disclosure but which direct the person to a specific solution are also coded B:4. For a suggestion to be coded B:4 it must imply compliance. B:4 suggestions give the solution and often close off further communication, they in effect say--here is what you ought to do, now go do it. Example: S: "You're having difficulty getting into that course? Well, go see the prof, he might let you in." (B:4) This is different than suggesting alternate ways of looking at a problem which encourages this person to explore the problem and come to his/her own decision, such as: S: "You're having difficulty getting into that course? Have you tried seeing the prof yet?" (A:2) This second example would be coded active listening since it encourages the person to explore talk about the problem some more. It tends not to cut off

communication. B:4 statements are always coded cognitive, either 14 or 34, never affective.

Module IIa, Example-Exercise

The definitions and a few examples of categories B:2, B:3, and B:4 have been presented and now it's time for some examples. Because the context is so important in making these distinctions. The examples below are presented in paragraph form as in exercise 1. Some of the events in the paragraph are already coded because they are categories you have not been introduced to yet. Following each paragraph will be an explanation of the codes.

T:	"OK, let's be quiet and get started. (1)	(1)	_____	B:4
	Fran, we left off with you why don't			
	you begin. (2) How did you feel about	(2)	_____	B:4
	where people put you on their sheets?"(3)	(3)	_____	B:4
S:	"Well, some of them I already knew but			
	others sort of bothered me." (4) ##	(4)	B:6	B:6
S:	"Which ones bothered you, Fran?" (5)	(5)	_____	B:2

(1) is a direction telling the class what to do and (2) is a suggestion to one particular student. In (3) the teacher is asking a question of a student to elicit information. (4) is self-disclosure which will be discussed in Module III. (5) is an example of active listening where a student acknowledges what Fran said (that she was bothered) and encourages her to explain more about what she has said.

S:	"A couple of people said I showed the			
	least concern for the group." (6)	(6)	B:6	B:6
S:	"How do you feel about that?" (7)	(7)	_____	B:2

S:	"Kind of confused, I don't know why they said that." (8)	(8)	B:6	B:6
T:	"Would you feel better if they told you." (9)	(9)		B:2
S:	"Sure, I'd like to know why." (10)	(10)	B:6	B:6
	##### (11)	(11)	0	0
S:	"You'd better ask them, it doesn't seem like they're going to volunteer it." (12)	(12)		B:4

In (6) Fran states which things bother her or her sheet, which is self-disclosure. Another student follows up with active listening (7) to explore how she feels about what they said. (8) is more self-disclosure by Fran to which the teacher asks her a question (9) to help her explore that feeling more, an example of active listening. Fran responds with a yes (10) which is followed by something I hope you remember from Module I, silence (11). In (12) a student suggests to Fran that she ask others why they put her down.

S:	"Jack, why did you put me down for that?" (13)	(13)		B:3
S:	"Because you just sit back an say so little. It seems like you're bored."(14)	(14)	B:7a	B:7a
S:	"Do I do any thing else?" (15)	(15)		B:2
S:	"Well, you're always the first one to leave." (16)	(16)	B:7a	B:7a
T:	"Fran, do you understand what Jack is saying?" (17)	(17)		B:3

S: "Yea, but I didn't know anyone took it to mean I didn't care." (18) (18) B:6 B:6

S: "You didn't see your behavior as indicating not caring?" (19) (19) _____ B:2

In (13) Fran asks Jack a question to elicit information from him. Jack states what behavior made him feel the way he does toward Fran which is negative feedback (14) and she actively listens to him (15) so he will explain his feedback more, which he does (16). The teacher then redirects the interaction back to Fran with a question (17) that elicits information to which she responds (18). The last statement (19) is a paraphrasing of what Fran said. This last example leads us into the next ground rule.

Ground Rule 2

During an interaction which consists of active listening and self-disclosure or offering information any time the focus of the interaction changes to another, individual or another topic the question which changes the focus is coded B:3. In the example above a question is asked (13) which brings about some self-disclosure and active listening. Then another question (17) changes the focus of the interaction to another individual. The question which changes the focus is coded B:3. The B:2 - B:6 interaction will occur frequently in small groups. Be aware of any change in focus be that change in relation to individuals or the topic of the conversation.

How did you do? If you are having trouble why not go back and re-read the section and then try the examples again. If you are

not folding the answers under I hope you're not fooling yourself, when listening to a tape you can't look over at the answers to give you a clue!!!

This section will explain the differences between the B:5 category (Offers Information) and the B:6 category (Self-Disclosure). As with the other sections definitions will be presented, examples will be given, and an example-exercise is the last step.

B:5-Offers Information.--This type of talk presents responses to questions or self-initiated facts or information concerning the content, subject, or procedures being considered. It consists of facts outside of one's own experience, i.e., it relates what the speaker knows rather than what he has done or experienced. For example if a class were discussing negative feedback and a student said that for negative feedback to be responsible it must be specific, that statement would be coded B:5. It is the presentation of a fact the student knows. But if the student said that when people give him feedback that is unspecific he thinks it is irresponsible because he doesn't know what to change, that would be coded B:6 because it explains personal experience. Unless the statement is actually descriptive of the person's personal experiences or feelings it is coded B:5. A comparative example might be helpful to show the difference: S: "When I was in high school we couldn't even leave the classroom without some kind of a pass. (B:6) But now kids are free to go anywhere they please in some schools." (B:5) The first sentence explains the students experience. The second sentence explains a fact the student is aware of. A hint to help you make

the B:5-B:6 discrimination is to ask yourself the question--Is the information of a personal nature or is it the recitation of acquired knowledge?

Very often during a group the facilitator or a group member will comment on the ongoing interaction. These types of statements are explanations or observations of the groups behavior and are called processing or critiquing in the field of group dynamics. Example: T: "The group has spent most of the time today talking about low-risk topics." (B:5) As long as these statements are presented as observations and do not show support or non-support for the observed behavior they are coded B:5. If the statement showed support for the behavior it would be coded positive feedback (B:1), and non-support is coded negative feedback (B:7), this will be discussed in the next section. This may be a difficult discrimination because people often present observations and infer non-support (or support) but unless the support or non-support is clear code the statement B:5.

Ground Rule 3

Two categories in cluster B can never be teamed with either of the affective categories in cluster B (A:2 or A:4). Categories B:4 and B:5 are never affective.

It is not possible to be affective in directing or suggesting anything. You could suggest to a person that they share more of their feelings but that is still a cognitive process, that is you would be saying that the fact is that they would be better off sharing their feelings.

B:5 statements are never affective because statements which refer to feelings but are not self-disclosures of personal feelings are actually offering information of a cognitive nature. For example, if a teacher were to talk about how it has been shown that talking about one's feelings is healthy, it would be coded B:5 since it is actually a cognitive presentation, i.e., facts about feelings.

B:6-Self-Disclosure.--Self-Disclosure is the sharing of your feelings or experiences about yourself. It may occur as a response to a question or it may be self-initiated. The essential ingredient in B:6 statements is that the information be of a personal nature, information that tells others more about you as an individual. Information which tells others about you, the individual is broadly defined but does not include the "I know someone who did that, they . . ." type statements (those statements are coded B:5).

There are some particular events which may make it difficult to distinguish B:6 from other categories. The following hints should help you. Whenever a person relates how they are feeling or how they were feeling, and the statement is not a reaction to something someone else has done (if the feelings were because of someone else's behavior it would be coded feedback) the statement is coded B:6.

Example: T: "I really feel great today." (B:6) or S: "I was really nervous going into that meeting." (B:6)

Opinions and values are very difficult to code. There is some over lap between categories B:5 and B:6 in relation to opinions

or values because people do not always clearly distinguish between facts and opinions, in fact many people are unaware of what is fact and what is opinion. Whenever an opinion is stated as an opinion code it B:6. The Clues to this are phrases like "I think . . . ," "I feel . . ." or "My opinion is . . . ," for example S: "I feel children are more honest than adults." (B:6) This is a clear example of an opinion stated as an opinion, but what if the phrase was just "Children are more honest than adults," is that presented as opinion or fact? One key to help you decide will be the tone of voice of the speaker. Very often the presentation of opinion as fact will include an authoritative tone. Another key that may help you is to remember that quite often when people present opinions as facts they over generalize and are prone to argue that their opinions are facts. For example, "Bussing is bad for all people involved" (B:5) is an opinion (although here it is presented as fact), there may be some facts to support that opinion but it is still an opinion. In coding groups watch for people arguing opinion as fact. Clear examples of opinions being presented as facts should be coded B:5. If there is reasonable doubt code the statement B:6.

Ground Rule 4

In group interaction the B:6-B:2-B:6 interaction will occur frequently, there will be some self-disclosure followed by active listening and then more self-disclosure followed by active listening and then more self-disclosure. It is sometimes difficult to

discriminate between when a listener is displaying active listening and when a listener changes the focus to himself. The key to the B:2-B:6 discrimination in an ongoing interaction is, where is the focus of attention? If the focus remains on the speaker the statement is active listening (e.g., S: "Sometimes complicated instructions confuse me. If that how you are feeling?", (B:2), but if the statement brings the focus of attention to the listener it is coded self-disclosure (e.g., S: "That happens to me also. In a history test the instructions were harder than the test. I spent . . ." B:6). In an ongoing interaction the focus is often changed when a response becomes self-disclosure rather than a brief clarifying statement. If the self-disclosure is short (a few seconds) and is used to show the other person that the listener understands, code it active listening. If it extends more than five seconds code it self-disclosure.

Module IIb, Example-Exercise

Things are probably getting a little more complicated now as the number of categories you learn increases. Only two major categories are left after these examples and then lots of practice in using all the categories at once. In case you're wondering, using all the possible combinations of cluster A and B there are thirty-three possible categories.

All the events which should be coded with the categories you have already been introduced to in cluster B are left blank (with all the answer on the far right). B:5 and B:6 categories are explained as are other events which may be difficult to categorize.

T: "Today we're going to continue with the shields we did on Tuesday. (1) Is there anyone who would like to start? ## (2) Gail why don't you begin. (3) #### (4)

(1) _____ B:5
(2) _____ B:3
(3) _____ B:4
(4) _____ 0

S: "For the value I wouldn't budge on I put down abortion. (5) I know that some people get upset talking about that but I feel that abortion is wrong. (6) Everything I've ever learned from my parents and church and everywhere else tells me that it is a sin. (7) The people who are for abortion are wrong. I know it!" (8)

(5) _____ B:6
(6) _____ B:6
(7) _____ B:6
(8) _____ B:5

S: "Why are you so sure of that Gail."(9) (9) _____ B:2

S: "I just am, I know they're wrong."(10) (10) _____ B:5

(1) is coded offering information because it gives the students information about the class. It tells them nothing about the teacher as a person. (5) is a piece of information that tells you something about Gail, it is descriptive of Gail the individual. (6) and (7) are also coded self-disclosure. (8) is a statement of opinion presented as fact, therefore coded offering information. (10) is coded B:5 for the same reason.

T: "Could you explain to us more about why you feel that way?" (11)

(11) _____ B:2

S: "You see, people are getting upset, let's talk about something else." (12) (12) _____ B:4

S: "It doesn't seem like everyone is upset, but it does seem like you are upset. (13) Gail, would you feel better if we talked about something else?" (14) (13) _____ B:5 (14) _____ B:2

S: "Yea, I guess it's mostly me. I'm real sensitive about the topic. (15) I could talk about it some more but I'm real nervous now. (16) Why doesn't someone else do their shield, we can get back to me later." (17) (15) _____ B:6 (16) _____ B:6 (17) _____ B:4

T: "OK, I feel that you getting upset like that is something you should try to work at." (18) (18) _____ B:7a

S: "Yea, I notice that people stay away from me sometimes when I'm like that." (19) (19) ##### (20) (19) _____ B:6 (20) _____ 0

(11) is active listening in response to (10). (12) is a suggestion to talk about another topic. (13) is an example of critiquing by a student. The student is making an observation of the class which is coded B:5. The student then asks a clarifying question (14) which is active listening. (15) is self-disclosure as is (16). (18) is an example of responsible negative feedback which will be explained in the next section. (19) is self-disclosure of her experience with other people.

	##### (21)	(21)	_____	0
T:	"Who would like to go next?" (22)	(22)	_____	B:3
	##### (23)	(23)	_____	0
T:	"Sam, Why don't you go next." (24)	(24)	_____	B:4
S:	"The value I wouldn't budge on is justice. I really feel strongly about that. (25) Justice is treating everyone equally under the law and is the cornerstone of democracy. (26) In the courts today many people are not treated equally and are deprived of their rights." (27)	(25)	_____	B:6
		(26)	_____	B:5
		(27)	_____	B:5
S:	"Sam, I agree, many people are deprived of their rights, only last week I got this ticket and the judge and the cops. . . ." (28)	(28)	_____	B:6
S:	"It seems to me that we are getting off the subject. Sam was sharing his shield and now we're on to a different subject. (29) Why don't we get back to Sam." (30)	(29)	_____	B:5
		(30)	_____	B:4

(22) is a question eliciting information. Silence (23) follows so the teacher directs (24) Sam to go next. Sam self-discloses (25) his strongest value and then goes on to offer information about justice (26) and (27). This information tells us nothing about Sam and is a good example of how a person can begin

to self-disclose himself and then change the topic so that he is only offering factual information.

(28) starts out with what seems like a paraphrasing of Sam but becomes a self-disclosure of an incident by someone else. (29) is an instance of someone sharing their observation of the class (critiquing) which is coded B:5 and then a suggestion (30) is made.

This is the last section in Module II and will explain four categories: B:1a-Positive Responsible Feedback, B:1b-Positive Irresponsible Feedback, B:7a-Negative Responsible Feedback, and B:7b Negative Irresponsible Feedback. This section will be organized differently from the preceding sections because of the similarities of these categories. First a definition of feedback will be presented which will be followed by an explanation of the difference between responsible and irresponsible feedback, next the four categories will be defined and that will be followed by the examples and exercises.

Feedback is the response or reaction a person gets from or give to others regarding one's personal being or actions. It is a verbal response of a sender (the person giving the feedback) to a receiver (person to whom feedback is directed) which is focused on the receiver's being or actions (stimulus behavior). That may sound very complicated so let me translate. When someone acts in such a way that you feel compelled to tell them how good or bad their behavior made you feel, you are being compelled to give that person feedback. The person giving the feedback is called the <u>sender</u>, the person to whom the feedback is directed is the <u>receiver</u>, and the

behavior of the receiver that caused the reaction in the sender, the behavior that compelled the sender to say some thing, is called the stimulus behavior. Now as you might have guessed there is positive and negative feedback. Positive feedback shows support for the stimulus behavior, it encourages the person to exhibit the stimulus behavior more often. Negative feedback shows non-support for the stimulus behavior, it discourages the person from exhibiting the stimulus behavior. These will be explained more fully in the definitions of the categories below. In addition to the two kinds of feedback mentioned so far (positive and negative) feedback can also be broken down into responsible and irresponsible. There are a variety of ways of defining responsible feedback but for the ISIA feedback will be defined as responsible if it is specific. That is to say that feedback which states the stimulus behavior is coded as responsible. Positive and Negative Feedback can be either responsible or irresponsible which gives us four categories of feedback. Each will be defined below.

B:1a-Positive Responsible Feedback.--Positive feedback shows support for the behavior of the receiver. It is, in effect, positive reinforcement for the stimulus behavior. Any time a group member (either teacher or student) shows support through his/her verbal statements for the behavior of the group as a whole or any individual, that verbal statement is coded Positive Feedback. For the feedback to be responsible it must be specific rather than general. For feedback to be specific it must describe to the receiver the stimulus behavior in specific terms, that is, the

receiver must be aware of exactly what he is getting feedback about. Sometimes the sender will not state the stimulus behavior in his feedback but it is obvious what the sender is referring to (e.g., T: "I'm glad you said that, that made me feel much better") in these cases the feedback is still considered specific (and responsible) because the sender is aware of what behavior of the feedback refers to. When feedback shows support and is specific it is coded Positive Responsible Feedback (B:1a).

B:1b-Positive Irresponsible Feedback.--Feedback which shows support but which is unspecific is coded Positive Irresponsible Feedback (B:1b). For positive feedback to be unspecific the receiver must be aware of what behavior the feedback refers to. Because most people feel comfortable giving positive feedback they usually give the feedback at the time the stimulus behavior is exhibited which usually makes the feedback specific. For this reason B:2b is normally in infrequently used category.

B:7a-Negative Responsible Feedback.--Negative Feedback differs from Positive feedback in that it shows non-support for the receiver's stimulus behavior. The sender implies through his response that he disagrees with and/or disapproved of the receivers behavior. The predictable effect of the negative feedback is that it weakens the stimulus behavior. As with Positive Feedback the feedback must be specific to be responsible. This is more important with Negative Feedback because very often the sender will not give the feedback at the time the stimulus behavior is exhibited. Instead the sender will wait, perhaps waiting for the behavior to be exhibited again,

and when the feedback is finally given it is imperative that the feedback is very specific in stating the stimulus behavior that compelled the sender to give the feedback. If the feedback shows non-support and is specific it is coded B:7a.

B:7b-Negative Irresponsible Feedback.--Feedback which shows non-support and which is unspecific is coded Negative Irresponsible Feedback (B:7b).

Here are some examples of the four types of feedback.

S: "I'm really glad you let us know how you felt." (B:1a). This shows support and is specific (it states that the receiver's stating how he felt was the behavior that made the sender glad).

S: "Your great Bill, you always do nice things." (B:1b). There might be a situation where this could be responsible but as it stands, out of context, it looks very vague and unspecific. What nice things does Bill do? Unless that if obvious from the context code it B:1b.

T: "Judy you've interrupted a number of people today and it seems like your not listening, that upsets me" (B:7a). The last phrase shows non-support and the statement is very specific to the stimulus behavior. If the phrase "that upsets me" was not included the statement would still be coded B:7a because interrupting and not listening show a strong indication of non-support by the sender.

S: "This is really stupid." (B:7b) This shows non-support but is very unspecific. What is studied and why is it stupid?

Before going on to examples and exercises there are a couple of ground rules concerning feedback that should be helpful to you.

Ground Rule 4

Feedback can be either cognitive or affective. If the feedback is directed toward the receivers affective behavior (e.g., sharing of feelings) and/or indicates the sender's feelings associated with the stimulus behavior (e.g., sender stating how the stimulus behavior makes him feel) it is coded affective. By coded affective I mean that the category from Cluster B (either B:1a, B1b, B:7a, or B:7b) is paired with either A:2 or A:4. If the feedback is not directed toward affective behavior and does not include the senders feelings it is coded cognitive and paired with either A:1 or A:3. (For example: 21a-teacher-cognitive responsible positive feedback, 47a-student-affective-responsible negative feedback, etc.).

Ground Rule 5

To be coded as feedback (either positive or negative) the statement must show support or non-support of behavior. Statements which simply state how the person behaves with no hint as to whether that behavior is "good" or "bad" (support or non-support) are coded B:5. This sounds a lot easier than it actually is since support or non-support is often times inferred from the behavior. For example, in an earlier example (Module IIa) Fran asks Jack why he put her down for "shows least concern for the group," he responds with "Because you just sit back and say so little. It seems like you're bored." If Jack had brought that up in class as an observation of Fran it would be coded B:5, but because it is an explanation of why he feels Fran shows the least concern for the group it is coded B:7a. He is

showing disapproval of that behavior. If the statement is made as an observation with no value judgment code it B:5. If the statement includes support or non-support or if the context implies support or non-support code it as feedback.

Module IIc Example-Exercise

All the categories have now been explained. This module will contain some examples of feedback and an explanation of those examples.

S: "I'd like to be the next person to get feedback. Does anybody have anything they'd like to share with me?" (1)	(1) _____	B:3
S: "I really like you John, you've really helped me." (2)	(2) _____	B:1b
T: "Jean, that isn't specific enough to be helpful to John. Could you tell him why he's helped you." (3)	(3) _____	B:7a
S: "OK, that day I was talking about my mom, you asked a lot of questions that helped me out. (4) And then after class I really felt good that you stuck around to talk to me. I felt you understood." (5)	(4) _____	B:1a
	(5) _____	B:1a
S: "Thanks Jean, that's helpful. (6) Does anyone have anything else?" (7)	(6) _____	B:1a
	(7) _____	B:3
##### (8)	(8) _____	0

In (1) John elicits information. Jean give him some feedback (2) that is irresponsible because it isn't specific. The teacher points out that the feedback is irresponsible and asks her to make it specific. He is giving her negative feedback (3) about her feedback. She then states a specific instance of why she feels good about John (4) and (5). John then supports (6) Jeans feedback, supporting her behavior. His statement is considered specific because she knows what he is referring to. He then asks a question (7) to elicit more information which is followed by five seconds of silence (8).

S: "Well, this is hard to say because I don't really know why I feel this way. (9) Sometimes I feel sort of uncomfortable around you cuz you know so much about me and I know so little about. (10) ## I guess it's because when we talk we mostly talk about me. # Like you're there to help me but I never get to help you. (11) Do you know what I mean?" (12)

(9) _____ B:6

(10) _____ B:7a

(11) _____ B:7a

(12) _____ B:3

S: "Yea, people have told me that before, like I'm the doctor and your're the patient." (13)

(13) _____ B:2

S: "That's just how I feel sometimes. I wish you'd talk about yourself more." (14)

(14) _____ B:7a

The first three events can be difficult to categorize because they seem to depend so much on one another. The first event in this section (9) is pretty clearly self-disclosure, the person is talking about themself. The next event (10) starts out like self-disclosure but if you look at (10) and (11) together it is clearly B:7a. Sometimes in coding tapes I find myself slowing down my tempo somewhat because statements such as (10) depend on what follows. It is then possible to write down a few quick codes such as B:7a, B:7a to catch up. After some taping practice I think you'll see what I mean. In (10) and (11) the student is pointing out behavior which makes him uncomfortable, showing non-support for that behavior. In (11) he seeks information regarding his feedback. (13) is a good example of Ground Rule 4. John states that people have told him that before to lead into his paraphrasing of the feedback. It is coded active listening because the focus is put on the other student (14) is the completion of his negative feedback and is coded B:7a partially because it follows the prevailing balance of the interaction (see Ground Rule 1).

Now you are finished with the last exercise in Module II. You've gone over all the categories and are ready for Module III, which presents exercises to code using complete codes (using both categories at once). Are you ready? Maybe you should go over a category or two again? It's up to you the exercises in Module III await you.

TRAINING MODULE III

Complete Coding

You've now done some coding using the categories from Cluster A and some coding using the categories from Cluster B. But when listening to tapes or observing a classroom you must use complex codes which involves the coding of every event (except silence) with a category from Cluster A and a category from Cluster B. This module will be primarily composed of example-exercises for you to code using both clusters at once.

There is no great mystery to complete coding. If you know all the categories well it just involves putting the two clusters together and thinking of events in terms of who's talking in what domain, with what communication skill? Just remember that B:4 and B:5 can never be affective and B:0 (silence) stands alone. Other than those exceptions all Cluster A categories go with all Cluster B categories to make up 33 possible ISIA categories. Before you go on to the example-exercise there are a couple things you might do: (1) look at page 41, it lists the 33 categories, (2) write down on a piece of scratch paper the categories as given on page 5. That will help you from fumbling around to remember. What number Active Listen is or whatever category your mind goes blank on. Most of the events should be relatively straight forward. I have picked out some which may be questionable and those are identified by having the event number underlined. They will be explained in a paragraph following the transcript.

	##### (1)	(1) _____	0
T:	"Has everyone finished drawing their shields?" (2)	(2) _____	13
	##### (3)	(3) _____	0
	##### (4)	(4) _____	0
T:	"OK, there are a number of ways we can do this. We could do it in small groups or we could share the shields in large groups. (5) Does anyone have a particular preference? (6) #####	(5) _____	15
		(6) _____	13
	(7) Well then I'd like to do it in a large group because then I get to hear everyone's shield. (8) Why doesn't someone just start and people can just volunteer. (9)	(7) _____	0
		(8) _____	26
		(9) _____	14
	##### (10)	(10) _____	0
S:	"I might as well start with the first category. The thing I hate about myself is that I'm always procrastinating. (11) I don't know why but I always put things off till the last minute, (12) I can be given an assignment for a twenty page paper and wait til the day before to begin. (13) ## I really don't like doing things this way but no matter what	(11) _____	46
		(12) _____	36
		(13) _____	36

	I tell myself I always put things off." (14)	(14) _____	46
S:	"How much does it really bother you? Do you think about it very often?"(15)	(15) _____	42
S:	"It doesn't bother me very much til the ninth week of the term." (16)	(16) _____	46
	##### (17)	(17) _____	0
S:	"That may sound funny but it really upsets me. (18) Everyone else is studying their notes and I'm beginning to worry about how I haven't read a single thing." (19)	(18) _____ (19) _____	46 46
S:	"It seems like you don't even worry til the last minute. (20) Does the experience you have during finals week have any impact on you early the next term?" (21)	(20) _____ (21) _____	42 42
S:	"Sometimes. Oh, I make a lot of resolutions but I always break them (22) I've written up lists of things to do and I've even checked off some I've done but it doesn't last. (23)	(22) _____ (23) _____	36 36
S:	"What do you do with your time since you're not studying?" (24)	(24) _____	32
S:	"Nothing really exciting. Sit around the dorm and talk with people, watch TV, the usual." (25)	(25) _____	36

S: "You ought to make some kind of plan so that you can't do those things til you've studied an hour or so a night." (26) (26) _____ 34

S: "Well, I haven't tried that yet. Maybe I'll give that a try." (27) (27) _____ 36

T: "You started this by saying it was something you hated about yourself. What exactly do you mean by hate/"(28) (28) _____ 22

S: "Well, it just kinda haunts me. I get mad at myself." (29) (29) _____ 46

T: "How mad? Mad enough to get you off you butt to do something about it, if you knew what to do?" (30) (30) _____ 22

S: "I'd like to think so, but I don't know what to do." (31) (31) _____ 36

S: "Do you think anyone can answer that ror you? What to do?" (32) (32) _____ 32

S: "I guess there's no easy answer, I feel like I've really tried (33) I really have tried but it's so frustrating, it's so easy to forget about when there's nothing due tomorrow." (34) (33) _____ 36 (34) _____ 46

T: "So you need some pressure. Can you put that pressure on yourself?" (35) (35) _____ 12

S: "I'm not sure what you mean." (36)	(36)	_____	33
T: "Well the pressure that has motivated you to do some work has been from outside yourself. (37) I'm thinking that if you could begin to put pressure on yourself you'd have more control of your own life. (38) When people let demands outside of themselves control their lives they are bound to become frustrated (39) Do you think that applies to you? (40)	(37)	_____	15
	(38)	_____	15
	(39)	_____	15
	(40)	_____	13
S: "Yea, when I'm frustrated its usually someone else who's telling me what to do (41)	(41)	_____	46
##### (42)	(42)	_____	0

This will be a short explanation of some of the more difficult or tricky events to code. As stated before the transcript these were identified by having the unit number underlined. If you have questions about any other events please let me know as others may be having the same difficulty and I may be able to write a ground rule or hint that could help you and others.

The very first event (1) is underlined not because it is difficult but to inform you that when listening to a tape or observing a live class you always start and end with a 0. (8) is affective-self-disclosure because the teacher is disclosing a

personal preference, a feeling toward something and although the feeling isn't terribly strong it is still an affective disposition (21) probably we could argue a long time on whether this is 32 or 42. Is the question cognitive or affective? Points could be made both ways as this is one of the few really subjective judgments. I coded it 42 because the paraphrasing statement before it is affective (26) is a solution to the problem and is not presented as a question, "have you tried this," with the other person to make the decision. Rather it is the solution with the connotation of, do this it will solve your problem.

(27) the reply is short and does not explore the problem solution. Although this isn't a description of an experience it is coded self-disclosure because it is a self-disclosure of what the person may choose to do, choose to experience, it is not a statement of fact. (33) is coded 36 because it is a statement which describes experience (self-disclosure) and it is cognitive because it does not state any of his feelings toward what he has done (i.e., he could have said "I think I've really tried"). The statements leading up to and including (39) are all statements of theory, as if to be read out of a book. The question (40) if put in other contexts could be active listening but because the long offering of information took the focus off the other person it is coded elicits information (42) remember to end tapes with a 0.

(43) (43) _____ __0__

S: "Well, I'm glad I told you that. I
can't say I've got any solutions to

	my problem but at least things to think about."(44)	(44)	_____	46
S:	"Jim, before you go on to something else I'd like to tell you that I'm really glad you shared all that with us. (45) Sometimes you hold back but today you told us alot about how you felt, and you even volunteered." (46)	(45)	_____	41a
		(46)	_____	41a
S:	"Thanks, I felt better saying those things." (47)	(47)	_____	46
T:	"Can anybody make any observations, or process what has happened so far." (48)	(48)	_____	13
S:	"I noticed that people were very attentive when Jim was talking, a lot of active listening and good eye contact." (49)	(49)	_____	35
	##### (50)	(50)	_____	0
T:	"Anything else?" (51) ##	(51)	_____	13
S:	"I think the feedback Kathy gave Jim was positive and responsible." (52)	(52)	_____	35
T:	"Jim, do you want to go on with the rest of your shield or would you like someone else to continue?" (53)	(53)	_____	13
S:	"Why doesn't someone else go ahead.(54)	(54)	_____	34
	## Any volunteers? ##(55)	(55)	_____	33

T: "Why don't you pick someone." (56) (56) _____ 14

S: "George, your shield looked interesting. Why don't you go next." (57) (57) _____ 34

S: "OK, but I'm not going to start with what I hate about myself. I'll start with the value I wouldn't budge on. (58) Some of you might think I'm a (58) _____ 35
Jesus-freak or something for saying this but the value I hold most important is the love of God. (59) (59) _____ 36
I feel very strongly that the love of God is a strong motivation in my life and could be the same for other people. (60) I don't really go out preaching it (60) _____ 46
but I enjoy talking to people about it if they show me they're interested. (61) (61) _____ 46
I don't particularly enjoy talking about it in a group, large groups that is, because there is usually someone who'll get bored or it'll turn someone off (62) I'd rather not turn someone away (62) _____ 46
from something I consider so valuable." (63) (63) _____ 46

S: "It seems like you'd rather not talk about it here but I'd like to hear it sometime." (64) (64) _____ 32

S: "I'd be happy to talk to you about it sometime. (65) (65) _____ 46

S: "George, are you embarrassed or shy or what? I'm not sure how you feel about talking about religion?" (66) (66) _____ 42

S: "I used to feel embarrassed but that's not so much it anymore ## (67) Now I'm not real sure how I feel about it except maybe I'm afraid I'll turn someone off to Jesus and all that goes with that." (68) (67) _____ 46 (68) _____ 46

T: "Are you afraid you'll turn somebody off to Jesus or are you afraid to turn someone off to yourself?" (69) (69) _____ 22

S: "Oh, I'm sure it's a little of both. It used to be that I was afraid people would dislike me but that doesn't bother me so much now. (70) I just see the reaction people have to door-to-door Jesus salesman, boy they really turn me off too, and I want to tell people that it isn't all like that. (71) Religion is sort of a personal thing, not sort of, it is a personal thing and it bothers me to see it sold on neon signs or door-to-door." (72) (70) _____ 46 (71) _____ 46 (72) _____ 46

S: "That's really right on George, I feel the same way. Some of those creeps that come to the door try to push religion like it was dope." (73)	(73)	_____	35
S: "Yea, I had this guy, just last week, come knocking at my door, and . . ." (74)	(74)	_____	36
T: "Wait a minute, it seems the conversation is moving away from George to everyone else's experiences (75)	(75)	_____	15
Let's get back to keeping the focus on George." (76)	(76)	_____	14
S: "George, how do you feel about the door-to-door people? What do you say to them?" (77)	(77)	_____	43
S: "Oh, I usually tell them I'm a christian and let them lay some of their rap on me." (78)	(78)	_____	36
S: "George! That's really bogue!" (79)	(79)	_____	37b
T: "Ron, could you be more specific." (80)	(80)	_____	14
S: "Well, you tell us that it's not good for people to lay that kind of trip on others and then you let them do it to you. I don't think that's helpful to them." (81)	(81)	_____	37a
S: "Yea, I see what you mean." (82)	(82)	_____	36

(47) includes the word "thanks" which if it were elaborated would possibly be considered positive feedback but as it stands the event is self-disclosing how he feels. (52) is an example of processing or critiquing. As with most processing the line between observations and feedback are very thin. Since this is clearly labeled processing it is coded 35. (55) includes quite a bit of silence but when there is a verbal statement use that for coding. Only code silence for five continual seconds or more. The sequence (59) through (63) is a good example of continual self-disclosure where you should use Ground Rule 1. (59) is coded cognitive because he is not talking about how he feels about religion. Starting with (60) the self-disclosure is clearly affective and that continues through (62) but (63) is questionable. Because there is reasonable doubt (cognitive or affective) code it with the flow of the interaction (affective). (73) is also questionable but with no flow of interaction to help us code this one we have to think in other terms. This is coded 35 because it is actually closer to offering information than to self-disclosure. It tells us nothing about the speaker really unless we look at it as offer opinion as fact and that is also coded B:5. Comparing (73) and (74) points out something else, the speaker in (74) is about to tell the class a personal incident which must be coded B:6. (80) is difficult because you could make a case for that being feedback but it is more clearly a suggestion or direction for Ron to follow. Short statements such as (82) should be coded as self-disclosure if they

refer to the person himself (as in this case) and as information if they refer to ideas.

I hope this last section hasn't demoralized you too much. I can just hear you saying "But there are so many exceptions and I just can't . . ." and I wish I could tell you it just ain't so but I must admit there will be some statement that coders may not agree on. If you find some glaring omissions please let me know, the ISIA is still in the developmental stage. But for now it is time to move on to listening and coding rather than reading and coding.

These transcripts have asked you to code specified events. Tape recordings do not specify the length of the events. The events on the transcripts would probably be longer than five seconds if they were spoken but that was done so that you wouldn't be asked to code in the middle of sentences. In listening to tapes and coding first try to code events as fast as it is comfortable to you. Try not to stop the tape recorder. Your first reaction will be to stop the tape every 30 seconds or so to think about what just been said. My experience has been that that hurts your coding more than it helps it keeps you from really becoming involved in what is happening in the group. Try to continue coding for ten to twenty minutes and then maybe to go back and code the same section of tape again. Practice is what is going to be most beneficial and you should not be overly concerned with speed (coding tempo) or accuracy at first that will come later.

Once you are comfortable with coding tapes the next step is to work on setting a coding tempo for yourself. To examine this, time yourself for ten minutes and see how many events you are able to code. You should aim at getting a tempo of twelve codes per minutes (one code every five seconds) and once you can do approximately that pace try to force yourself to continue it, it will probably become like a metronome in your head.

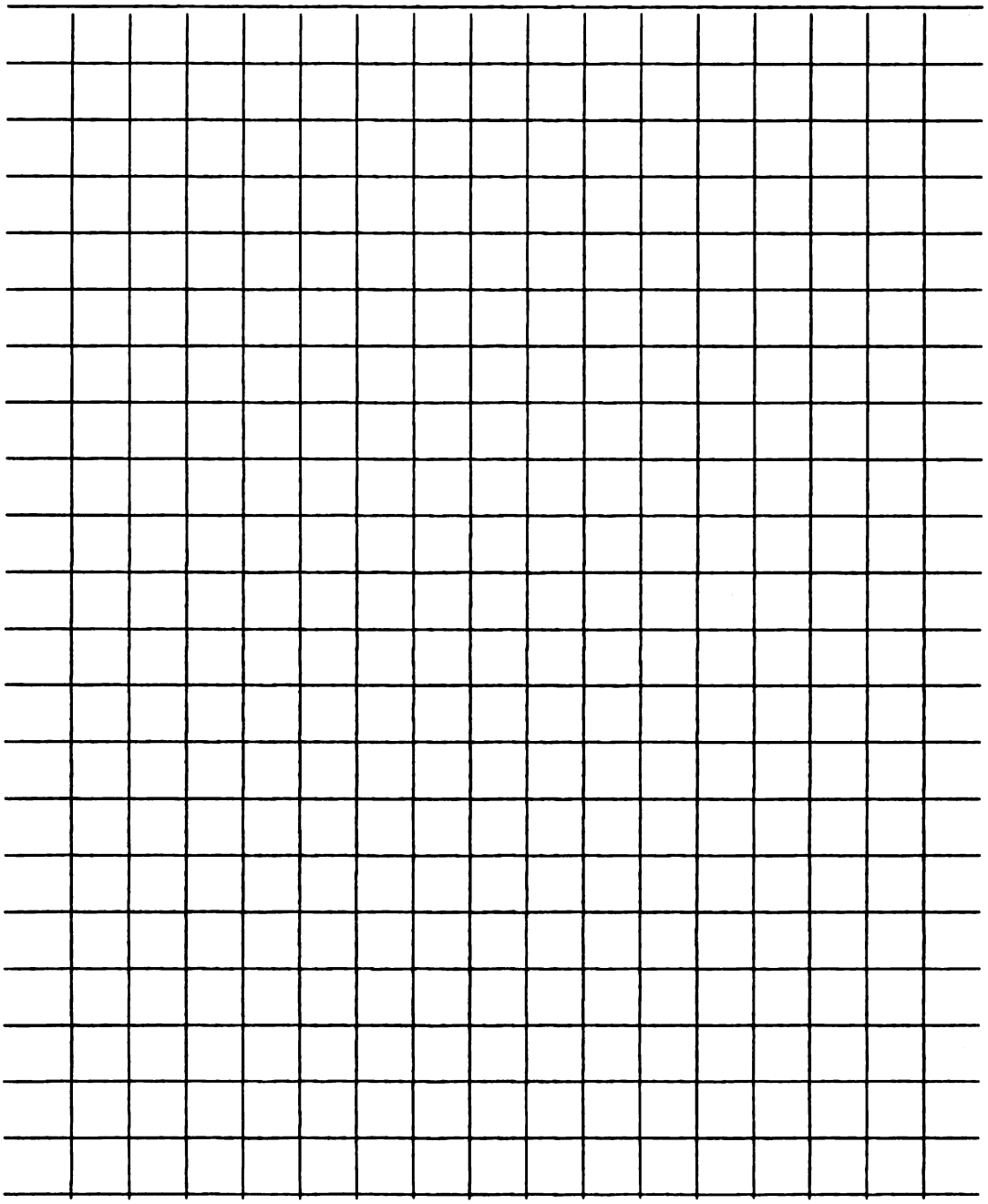
The last step is accuracy and this is most easily examined by comparing you to yourself. (Later you'll be compared to others but if you are consistant with yourself that is a big step in the right direction.) You can do this by coding the same segment of tape a number of times. Then total up each of the categories, i.e., how many 32's were there in that segment the first time you coded it?, the second time? Do not try to compare individual codes because your tempo may have been different. Comparing the totals for each category will give you some indications of consistancy and accuracy.

Please see me for any questions you have. Thank you.

Section number _____ Date _____

Instructor _____ Observer _____

Date of coding _____ Time _____ Page # ____



33 Possible ISIA Categories

0	31a
11a	31b
11b	32
12	33
13	34
14	35
15	36
16	37a
17a	37b
17b	41a
21a	41b
21b	42
22	43
23	46
26	47a
27a	47b
27b	47b

Revisions, Clarifications and Helpful Hints in Coding Tapes

This appendix is to help you in coding tapes and also to clear up some confusions and make clearer some distinctions between various categories. This appendix should be read before you attempt to code any audio-tapes.

One of the basic difficulties in coding interpersonal communications is that very few statements obviously fit into one code. Theoretically a statement is either one category or another but in real life settings people do not talk according to the category rules of the ISIA. Often times people begin a sentence with the intention of giving someone feedback and then midway through the sentence change their minds and start talking about themselves. This makes for variety in social situations but frustrates the observer. Below are some helpful hints to aid you in listening to audio-tapes. Some of these may modify slightly what you read in the ISIA manual, others are presented to help you deal with audio-tapes.

Distinguishing the B:5 category from many other categories. The B:5 category may cause you to want to give me a great deal of B:7a and perhaps B:7b. It seems to overlap with many other categories and so here are some hints to help you distinguish it in a few of the usual sticky situations.

1. In almost any tape you listen to you will encounter the difficult B:5 - B:6 distinction. When is a statement self-disclosure and when is it offering information? This may be the

most difficult distinction because there is no single rule to tell you how to code the statement when the distinction is B:5 - B:6 and also because this distinction has to be made so often. One very general subjective hint that should help you is to keep in mind this definition: Self-disclosure is the sharing of feelings, experiences, or values that allow others to know the speaker's inner-self. It is the act of revealing personal information about one's self and has been labeled "real self-communication." This may sound straight forward but when applied to real life spontaneous statements you will find some statements obviously B:5 or obviously B:6 but all too many fall short of being obvious. In situations where you are unsure, listen to see if the statement discloses feelings, experiences or values about the speaker that would help others to know the speaker. For example: S: "When we were in the small groups Monday I didn't say very much." Code that statement 35, it is an observation of the speaker's own behavior. But if the speaker added "because I didn't think anyone was listening to what I had said." Code that 36 because it adds something that was happening inside the person and helps others to know why the speaker behaves as he does.

2. B:5 statements can also be difficult to discriminate from B:1 and B:7 statements because often times people offer what might be considered feedback without stating support or non-support. When using the ISIA to code interactions only statements which show support or non-support are coded as feedback. Statements which "feed-back" information to others in the form of "this is what

you are doing" but do not show support or non-support for that behavior are coded B:5 (offer information). That sounds pretty easy but because people often infer support or non-support it can often times be difficult to code. If you are sure the speaker and the listeners (receivers) are aware of the support or non-support code the statement feedback. If the statement is reasonably vague to leave some doubt code the statement B:5.

3. In some situations the B:5 category may even be difficult to discriminate from the B:4 category. In those situations remember that the B:4 category is a direction or suggestion encouraging someone to do something (or not to do something). For example: S: I've got a book on problem solving." (35) is different from S: "I've got a book on problem solving, why don't you read it." (34) because in the second statement the speaker is encouraging the person to take some action. By now you're probably thinking the B:5 category is just a garbage category that is part of the ISIA to confuse you. Well it isn't there to confuse you but in some ways it is a catch-all category to be used to code statements which fit nowhere else. Jokes, side remarks, and many social amenities should be coded B:5. It is a category that is not really a "communication skill" and encompasses a large portion of everyday speech. In fact there are probably many everyday conversations which, if coded, would include almost exclusively the B:5 category. This brings me to the last hint which I hesitate to write, so use it with caution. When you encounter a statement which you are sure fits <u>none</u> of the other cluster B categories code that statement B:5. This is <u>not</u> an

invitation to code everything B:5 but rather a rule to help you with those statements which just don't seem to fit anyplace.

The following are some ground rules to help you in coding audio-tape recordings.

1. The first thing you will notice is that the transcripts you've been coding are quite a bit different from the way people converse in a group. One of the greatest differences being that people often draw out a short sentence or thought to twenty or thirty seconds. This causes two problems: you may be waiting 10-15 seconds to figure out how to code that short sentence and then once you decide on a code should you go back and record silence (0) or the code you've decided upon? Because tempo is so important there will be times when you must go back and "catch-up" when you felt it important to wait to decide on a code. Waiting 10-15 seconds to decide on what code is appropriate is perfectly ok in many situations because people often times talk so slowly that coding at a constant rate is impossible. But in situations where you have been waiting, remember to then "catch-up" by recording more than one of those codes (for example, if you waited 15 seconds and then decided that a statement should be coded 26, you should quickly code 26, 26, 26). This may sound confusing and may be confusing when you first start listening to tapes, as you will probably be uncomfortable making quick decisions. But as you become more used to listening and coding you will spend less time waiting to decide on codes for particular events. The key to this is to remember that you want to establish a coding tempo in

your mind which will hopefully be about twelve codes per minute. Any time you have to wait to decide on a code you will break up that tempo. No matter how skilled an observer you become there will always be times you will have to wait to make a decision on a code but you will want to keep these pauses to a minimum.

2. Another difficulty you will run into right away is that people in groups do not talk one at a time and often interrupt one another. When there is background noise (laughing, wispering, etc.) if the person who has been talking is still talking and seems to still be the "center of attention" (that is, hasn't been drowned out) keep coding that person. Also if a person has been talking (e.g., giving suggestions) and is interrupted only to return to his statement, continue to code it as you were coding it before the interruption. If the background noise or interruptions become chaos and you can't code the speaker, code that time as silence or confusion (0).

3. If a person is talking affectively and someone actively listens to that person keeping the subject the same, code the active listening as affective, even if the person doing the active listening does not refer directly to the speaker's feelings.

4. If during a five second event two different categories occur, code both categories. This often times occurs with active listening when a person asks a very short question. This will mean that sometimes you will be coding at a rate that is faster than 12 codes per minute.

5. Before you begin to code tapes I'd like to remind you of Ground Rule 1: When doubt exists the coder will classify doubtful statements into categories which are consistant with the prevailing balance of the interaction. This is perhaps the most important ground rule in listening to tapes. You must always be listening and coding every five seconds but keep in mind that a five second event may be part of a long sentence or interaction that is one thought, emotion, or question. Don't be disturbed if you write twenty 15's in a row. Be aware that the appropriate code can change at any time but that extended periods of the same code are not unusual. This is particularly important in distinguishing cognitive from affective. If a person begins by talking about a certain feeling they had (e.g., 46) and talk for five minutes about that feeling remember that the flow of the conversation has been affective self-disclosure and unless the speaker changes the focus away from that feeling (or a question changes the focus) it is still coded affective self-disclosure.

Practice Tapes

This is a practice tape for you to practice coding. It is divided into five segments with short pauses between each segment so that you can stop the recorder and check yourself or recode the same segment. The segments are 5 minutes, 6 minutes, 6 minutes, 12 minutes, and the last segment is 16 minutes.

This practice tape should help you in two areas. First it can help you develop a stable coding tempo. By checking the number

of codes you recorded for the first five minute segment (and other segments) you can estimate your rate of coding. Try to aim for 12 codes per minute (that would give you 60 codes for the five minute segment, 72 codes for the six minute segment, etc.). Second, this practice tape should be able to help you become more consistant and reliable in your coding. By coding the same segment a number of times you can compare yourself with yourself. More important you should compare your coding with an experienced observer. Included with this tape is an observation sheet with the codes I have recorded for these five segments. Compare your codes with the codes on that sheet looking at the differences. It may be helpful to listen to the tape segment to examine the differences. Please note any questions you have and see me if you would like to discuss those differences. The segments increase in difficulty as you move from segment I to V. The most difficult segment is V particularly the first minute or so. Examine the codes we differ on and follow the codings I have written as you listen to the tape if you have difficulty (this is after you have coded the segment yourself). Recoding the same segment can be very helpful. You may also want to put a watch or clock in front of you to help you get a stable coding tempo.

Segment I

Section Number 5 Date Aug. 8, 1974

Instructor Randy Observer Randy

Date of coding _____ Time 5 minutes Page #___

0	35	14														
35	0	35														
35	0	35														
35	36	35														
13	12	35														
14	12	35														
0	36	0														
35	33	0														
35	36	0														
35	32	33														
0	36	0														
35	34	35														
35	34	0														
35	35															
13	35															
35	35															
35	35															
35	35															
35	35															
35	15															
0	15															
35	15															
0	15															
0	15															
13	14															

Segment II

Section number 5 Date Aug. 8, 1974

Instructor Randy Observer Randy

Date of coding _____ Time 6 minutes Page # ____

0	46	46														
13	46	46														
36	46	46														
46	46	46														
46	46	46														
35	35	46														
35	15	35														
35	36	46														
13	36	32														
35	14	36														
35	14	46														
35	36	46														
35	15	46														
35	36	46														
35	36	46														
35	34	46														
15	34	36														
35	34	36														
35	35	36														
33	35	36														
35	36	35														
35	36	35														
35	46	33														
15	22	0														
46	22	36														

Segment III

Section number 5 Date Aug. 8, 1974

Instructor Randy Observer Randy

Date of coding _____ Time 6 minutes Page #___

0	0	46	34												
13	14	0	35												
11a	0	34	0												
11a	0	33													
11a	33	46													
36	0	46													
33	15	46													
35	15	46													
33	35	46													
36	35	46													
35	35	46													
15	11a	46													
15	11a	46													
35	33	46													
35	36	46													
35	33	46													
35	46	32													
35	36	36													
0	36	36													
35	0	36													
35	31a	33													
33	31a	46													
15	46	46													
15	46	46													
0	0	46													

Segment IV

Section number 5 Date Aug. 8, 1974

Instructor Randy Observer Randy

Date of coding _____ Time 12 minutes Page #____

0	46	32	34	36	16	35							
35	46	0	34	35	16	36							
33	42	0	33	15	15	35							
36	42	13	34	15	15	0							
31a	46	14	34	34	15								
0	46	35	36	34	14								
0	46	35	34	35	14								
35	46	35	35	33	36								
46	46	35	35	36	36								
46	46	35	35	36	36								
46	46	35	35	36	36								
46	34	46	35	0	36								
15	35	46	35	36	36								
46	35	35	35	32	15								
46	35	35	35	36	15								
46	35	35	33	36	35								
46	35	36	33	35	15								
46	35	36	36	35	15								
46	35	36	36	35	15								
46	35	36	35	35	34								
42	35	36	36	15	46								
46	35	34	0	16	46								
46	32	34	34	16	33								
46	36	34	35	16	35								
46	32	34	35	16	33								

Segment V

Section number 5 Date Aug. 8, 1974

Instructor Randy Observer Randy

Date of coding Time 16 minutes Page #

0	35	46	35	36	15	36	22						
35	35	46	35	36	15	36	22						
35	35	46	35	35	32	36	22						
35	31a	46	46	35	15	36	46						
35	31a	15	46	46	15	36	46						
35	31a	15	46	46	15	12	46						
46	31a	15	46	46	15	36	46						
46	31a	15	42	46	35	36	46						
46	31a	15	42	0	35	36	46						
31a	31a	15	46	15	33	32	46						
35	31a	15	46	15	36	36	42						
42	46	15	46	15	36	36	46						
36	46	15	22	33	35	36	46						
36	46	35	46	15	35	36	46						
36	46	31a	36	15	36	36	0						
36	46	32	36	15	35	36							
36	42	0	36	15	15	32							
35	42	0	36	15	33	36							
35	46	33	36	15	0	36							
35	46	0	36	15	34	36							
35	46	35	36	15	0	36							
35	46	33	46	15	0	0							
35	46	36	22	15	33	22							
35	46	35	46	15	36	46							
35	46	35	36	15	36	22							

APPENDIX B

INTERPERSONAL PROCESS LAB OPINIONNAIRE

Appendix B

IPL Evaluation

This is a questionnaire attempting to discover how people perceive IPL sessions in relation to specific IPL skills. Your IPL has been tape recorded and will be examined by means of an observation technique, in reference to what IPL skills have been used during this session. This questionnaire will be compared with the findings of the observation technique to verify those results. This will not be used to evaluate you or your instructor. Please answer the following 24 questions taking into account the following:

1. It is sometimes the case that the parts of an IPL vary a great deal. For example, part of an IPL may deal exclusively with self-description while later the class will focus on feedback. This opinionnaire begins to take that into consideration by looking at halves of the IPL. The first 12 questions of this opinionnaire refer to the first half (initial hour) of the class, the second 12 questions (#13 - #24) refer to the second half (final hour) of the class.

2. Some question has arisen as to what to compare the class with in answering questions such as #8 and #20. Please compare this IPL session (session referring to a single 2 hour class period) with other IPL sessions you have experienced this term. You are not being asked to compare other IPL sections to your section because you have not experienced them. Please make your judgments in relation to sessions you experienced earlier this term.

3. Most of the words presented in this opinionnaire you have probably come across previously in the IPL or carrel work. The following words are defined to help you with concepts you may not have encountered.

Cognitive vs. Affective - Cognitive refers to facts, knowledge, and the process of thinking. Affective refers to feelings, attitudes, values, and the process of feelings. Question 1 and 13 ask you to judge whether the class was talking about how they think (cognitive) or how they feel (affective).

Personal Self-Description vs. Factual Information - Personal self-description refers to the sharing of feelings or experiences. Factual information refers to facts outside of one's own experiences, to what people know rather than what they feel or have done.

Please enter your instructor's name, section number, and date on this sheet (or IBM sheet) along with your responses. Also check the box marked "student." Thank you, this will help us keep our records in order.

IPL Evaluation

Student ___ Instructor's Name _____

Instructor ___ Date _____

Observer ___ IPL Section # _____

1. During the first half of the IPL session, the ratio of time spent in the affective domain in comparison to the time spent in the cognitive domain was:

1.	100% (all affective)	:	0% (no cognitive)
2.	75% (affective)	:	25% (cognitive)
3.	50% (affective)	:	50% (cognitive)
4.	25% (affective)	:	75% (cognitive)
5.	0% (no affective)	:	100% (all cognitive)

2. The ratio of time in the first half of the IPL session which was devoted to personal self-description in comparison to the time devoted to the presentation of factual information was:

1.	100% (all personal self-description)	:	0% (no facutal infor.)
2.	75% (personal self-description)	:	25% (factual infor.)
3.	50% (personal self-description)	:	50% (factual infor.)
4.	25% (personal self-description)	:	75% (factual infor.)
5.	0% (personal self-description)	:	100% (factual infor.)

3. The amount of positive feedback given during the first half of the session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

4. If positive feedback was given during the first half of the IPL session, how much of that positive feedback was responsible?

None				All of it	No Positive Feedback was Given	Unsure of the Concept
1	2	3	4	5	6	7

5. The amount of negative feedback given during the first half of the session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

6. If negative feedback was given during the first half of the IPL session, how much of that negative feedback was responsible?

None				All of it	No Negative Feedback was Given	Unsure of the Concept
1	2	3	4	5	6	7

7. The amount of active listening which occurred during the first half of the IPL session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

8. In comparison to other IPL sessions to date, the first half of the session was:

1. One of the best
2. Above average
3. Average
4. Below average
5. One of the worst

9-12. Please circle the number which indicates how you were feeling about the first half of the IPL session in relation to the following four pairs of words.

9.	Genuine				Artificial
	1	2	3	4	5
10.	Relaxed				Tense
	1	2	3	4	5
11.	Construc-tive				Destruc-tive
	1	2	3	4	5
12.	Involved				Uninvolved
	1	2	3	4	5

13. During the second half of the IPL session, the ratio of time spent in the affective domain in comparison to the time spent in the cognitive domain was:

1. 100% (all affective) : 0% (no cognitive)
2. 75% (affective) : 25% (cognitive)
3. 50% (affective) : 50% (cognitive)
4. 25% (affective) : 75% (cognitive)
5. 0% (no affective) : 100% (all cognitive)
6. Unsure of the concepts

14. The ratio of time in the second half of the IPL session which was devoted to personal self-description in comparison to the time devoted to the presentation of factual information was:

1. 100% (all personal self-description) : 0% (no factual infor.)
2. 75% (personal self-description) : 25% (factual infor.)
3. 50% (personal self-description) : 50% (factual infor.)
4. 25% (personal self-description) : 75% (factual infor.)
5. 0% (personal self-description) : 100 % (factual infor.)

15. The amount of positive feedback given during the second half of the session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

16. If positive feedback was given during the second half of the IPL session, how much of that positive feedback was responsible?

None				All of it	No Positive Feedback was Given	Unsure of the Concept
1	2	3	4	5	6	7

17. The amount of negative feedback given during the second half of the session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

18. If negative feedback was given during the second half of the IPL session, how much of that negative feedback was responsible?

None				All of it	No Positive Feedback was Given	Unsure of the Concept
1	2	3	4	5	6	7

19. The amount of active listening which occurred during the second half of the IPL session was:

None				A Great Deal	Unsure of the Concept
1	2	3	4	5	6

20. In comparison to other IPL sessions to date, the second half of the session was:

    1. One of the best
    2. Above average
    3. Average
    4. Below average
    5. One of the worst

21-24. Please circle the number which indicates how you were feeling about the second half of the IPL session in relation to the following four pairs of words.

21.	Genuine				Artificial
	1	2	3	4	5
22.	Relaxed				Tense
	1	2	3	4	5
23.	Construc-tive				Destruc-tive
	1	2	3	4	5
24.	Involved				Uninvolved
	1	2	3	4	5

Please specify on the reverse side of this sheet your reasons for responding to #8 and #20 in the manner you chose.

APPENDIX C

GAZDA'S GLOBAL SCALE

GLOBAL SCALE [1]

1.0	1.5	2.0	2.5	3.0	3.5	4.0
A response in which the helper attends to neither the content nor the surface feelings of the helpee; discredits, devalues, ridicules, or scolds the helpee; shows a lack of caring for, or belief in the helpee; is vague or deals with the helpee in general terms; tries to hide his feelings or uses them to punish the helpee; reveals nothing about himself or discloses himself exclusively to meet his own needs; passively accepts or ignores discrepancies in the helpee's behavior that are self-defeating; ignores all cues from the helpee regarding their immediate relationship.		A response in which the helper only partially attends to the surface feelings of the helpee or distorts what the helpee communicated; withholds himself from involvement with the helpee by declining to help, ignoring the helpee, responding in a casual way, or giving cheap advice before really understanding the situation; behaves in a manner congruent with some preconceived role he is taking, but is incongruent with his true feelings; is neutral in his nonverbal expressions and gestures; is specific in his verbal expressions (e.g., gives advice or own opinion) or solicits specificity from the helpee (e.g., asks questions) but does so prematurely; does not voluntarily reveal, but may briefly answer questions regarding his own feelings, thoughts, or experiences relevant to the helpee's concerns; does not accept discrepancies in the helpee's behavior but does not draw attention to them either; comments superficially on communications from the helpee regarding their relationship.		A response in which the helper reflects the surface feelings of the helpee and does not distort the content; communicates his openness to entering a helping relationship; recognizes the helpee as a person of worth, capable of thinking and expressing himself and acting constructively; communicates his attention and interest through his nonverbal expressions or gestures; shows that he is open to caring for and believing in the helpee; is specific in communicating his understanding but does not point out the directionality emerging for helpee action; shows no signs of phoniness but controls his expression of feeling so as to facilitate the development of the relationship; in a general manner, reveals his own feelings, thoughts, or experiences relevant to the helpee's concerns; makes tentative expressions of discrepancies in the helpee's behavior but does not point out the directions in which these lead; discusses his relationship with the helpee but in a general rather than a personal way.		A response in which the helper goes beyond reflection of the essence of the helpee's communication by identifying underlying feelings and meanings; is committed to the helpee's welfare; is intensely attentive; models and actively solicits specificity from the helpee; shows a genuine congruence between his feelings (whether they are positive or negative) and his overt behavior and communicates these feelings in a way that strengthens the relationship; freely volunteers specific feelings, thoughts, or experiences relevant to the helpee's concerns (these may involve a degree of risk taking for the helper); clearly points out discrepancies in the helpee's behavior and the specific directions in which these discrepancies lead; explicitly discusses their relationship in the immediate moment.

[1] Rarely if ever would all of the conditions described in each level be represented in a single response; therefore when the scale is used to rate responses, the rater must be guided by the level(s) of the condition(s) that are offered or withheld in the helper's response.

APPENDIX D

INTERPERSONAL PROCESS LAB OBJECTIVES

Appendix D

Goals and Objectives for the IPL

A. Goal: The student will become aware of the interpersonal skills of communication necessary for constructive social-emotional growth, assess the effectiveness of these skills, and describe the transfer of these skills and processes to teaching experiences.

Objectives:

1. To learn and exhibit self-description skills. The student will be able to share his own ideas, opinions, and feelings about himself as it relates to his perceived willingness and readiness to (1) teach and (2) explore, respect and be responsible to himself and others.
2. To learn and exhibit listening skills. The student will be able to not only re-state what has been said, but also to relate the feelings and intended meaning of the speaker to the speaker's satisfaction.
3. To learn and exhibit questioning skills. The student will be able to seek further information or clarification for self and others without cueing a particular response.
4. To learn and exhibit observation skills. The student will be able to recognize and interpret, through description and explanation, devise modes of non-verbal expression, i.e., hands, face, arms, etc.

5. <u>To learn and exhibit responsible feedback skills</u>. The student will be able to relate honestly his feelings about another person (to that person) in an effective manner.
   a. Discriminate between feedback that is responsible in intent from feedback that is irresponsible.
   b. To decide a course of action based on the feedback given and the evaluation of that feedback, i.e., to change self, to accept differences, etc.

APPENDIX E

FLANDER'S SYSTEM INTERACTION ANALYSIS

Appendix E

Flanders' System of Interaction Analysis (FSIA)

1. Accepts Feelings
2. Praises or Encourages
3. Accepts or Uses Student Ideas
4. Asks Questions

5. Explaining or Informing
6. Gives Directions
7. Scolding/Reprimanding or Defending Authority

8. Student Talk - Expected or Predictable Response
9. Student Talk - Initiated Response

10. No Talk/All Talk

APPENDIX F

OBER'S RECIPROCAL CATEGORY SYSTEM (RCS)

Appendix F

The Reciprocal Category System (RCS)

Category Number Assigned to Teacher Talk		Category Number Assigned to Student Talk
1	Warms (informalizes) the climate	11
2	Accepts	12
3	Amplifies the Contribution of Another	13
4	Elicits	14
5	Responds	15
6	Initiates	16
7	Directs	17
8	Corrects	18
9	Cools (formalizes) the climate	19
10	Silence or Confusion	10

APPENDIX G

ISIA CATEGORIES

Appendix G

33 Possible ISIA Categories

0	31a
11a	31b
11b	32
12	33
13	34
14	35
15	36
16	37a
17a	37b
17b	41a
21a	41b
21b	42
22	43
23	46
26	47a
27a	47b
27b	

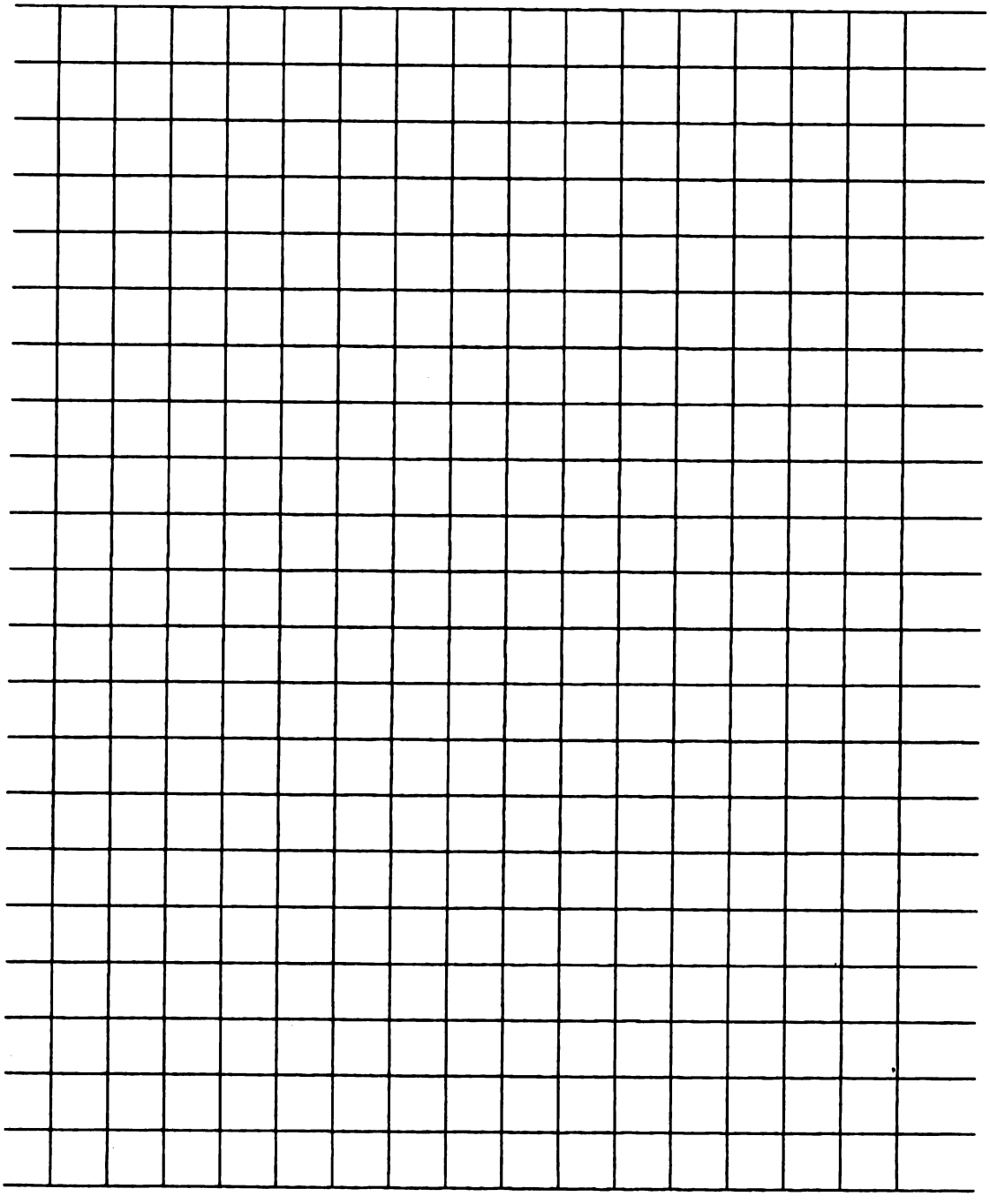
APPENDIX H

ISIA OBSERVATION SHEET

Section number _____ Date _____

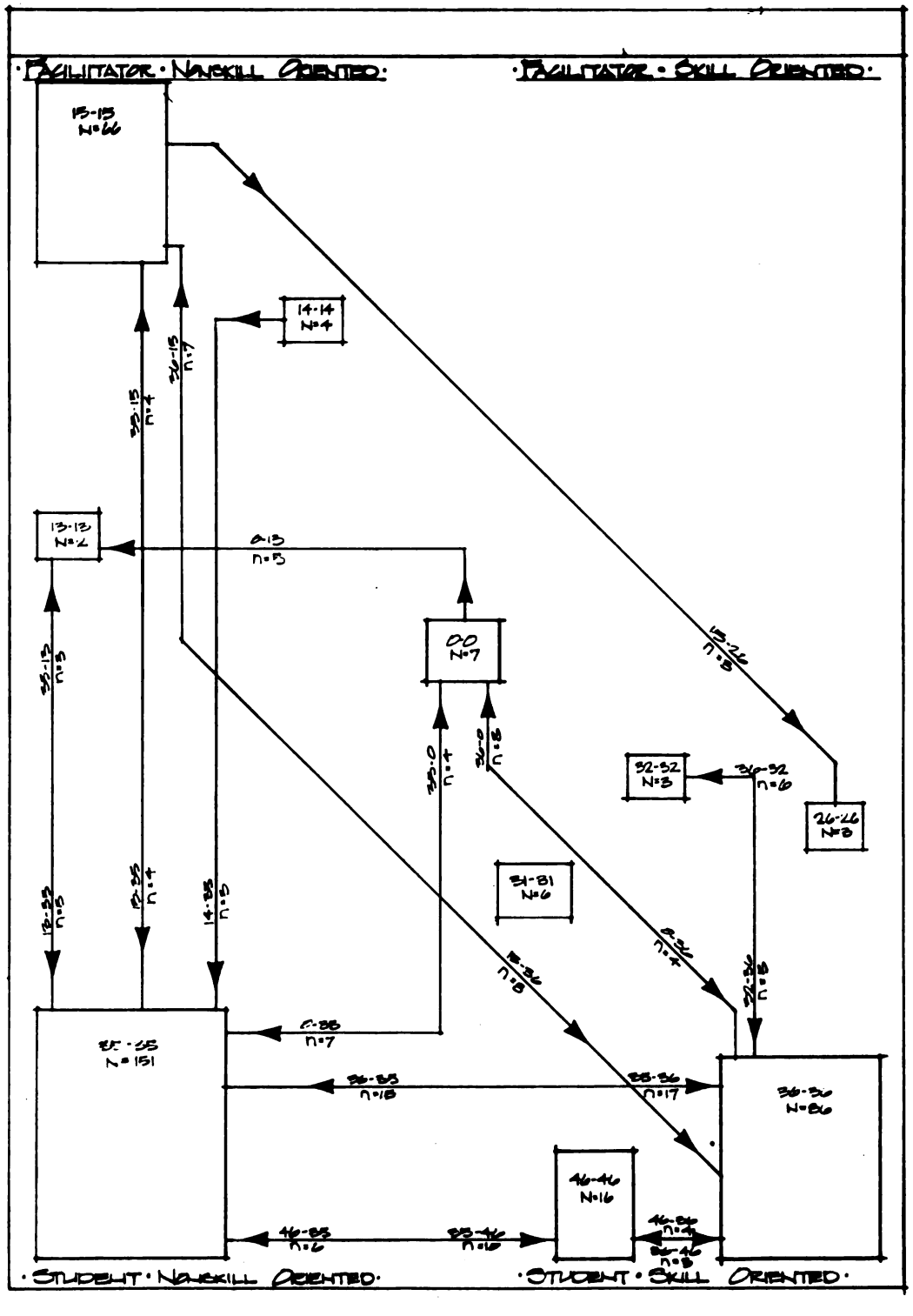
Instructor _____ Observer _____

Date of coding _____ Page # ___

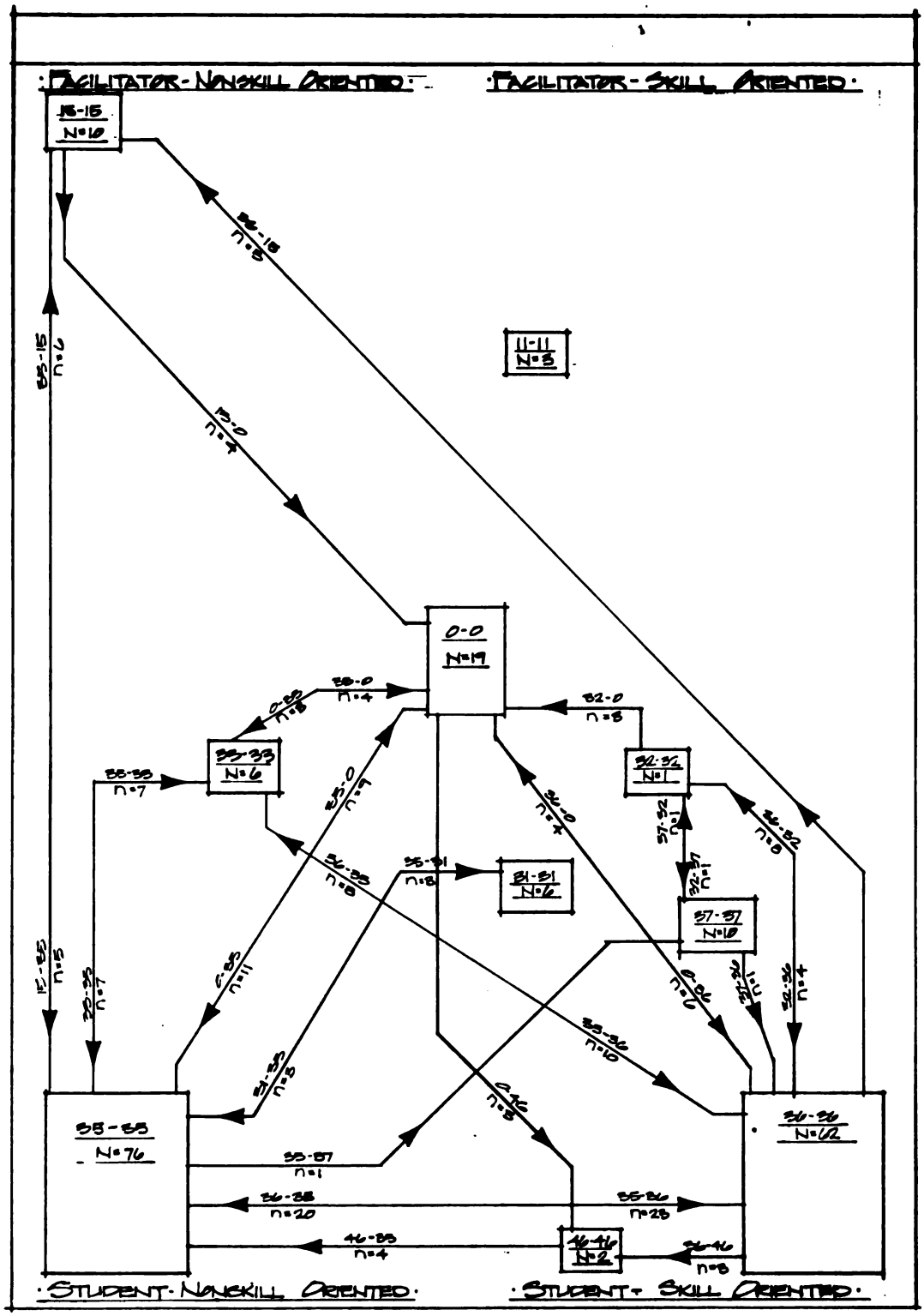


APPENDIX I

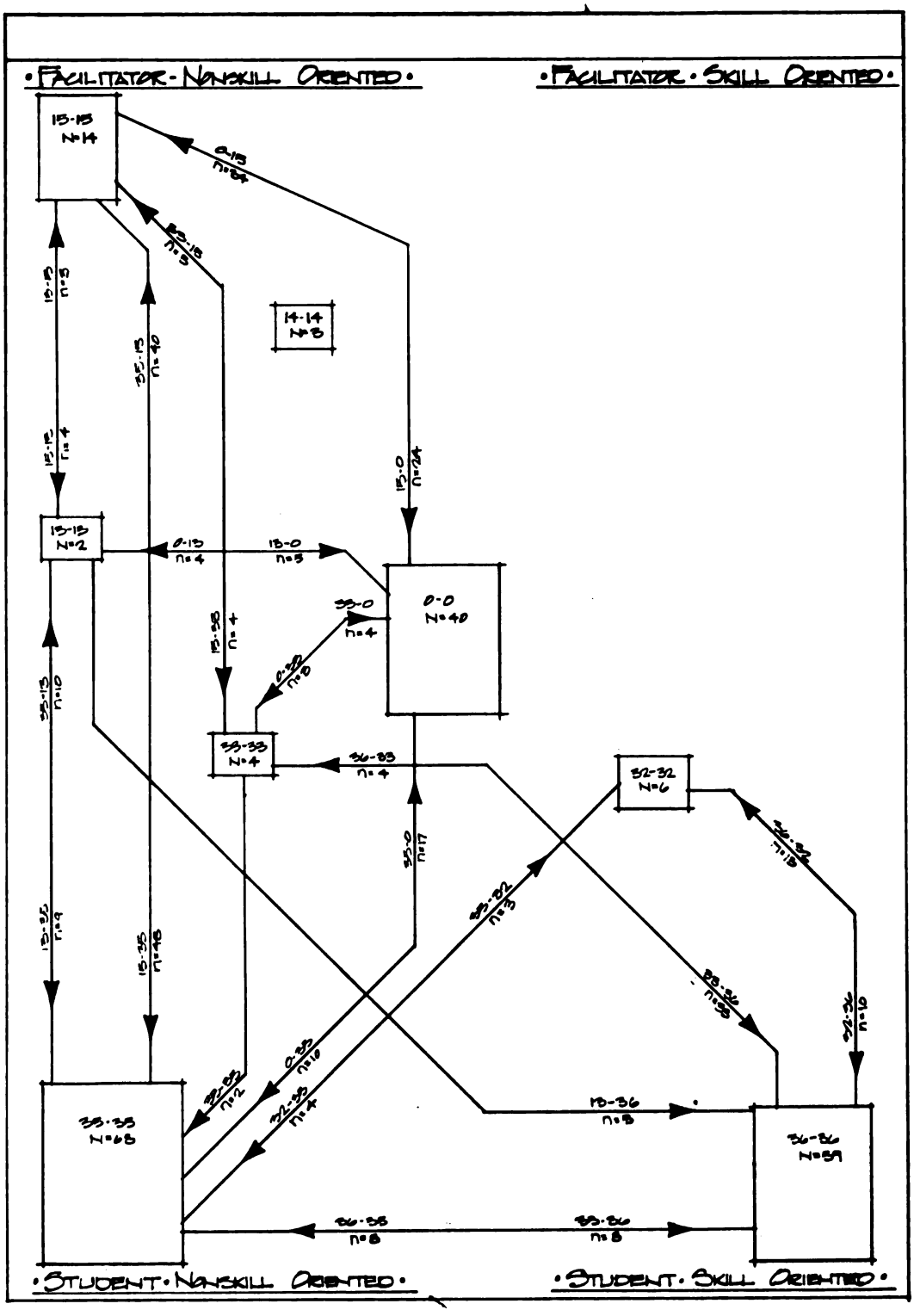
VALIDITY GROUP MATRICES AND FLOW CHARTS



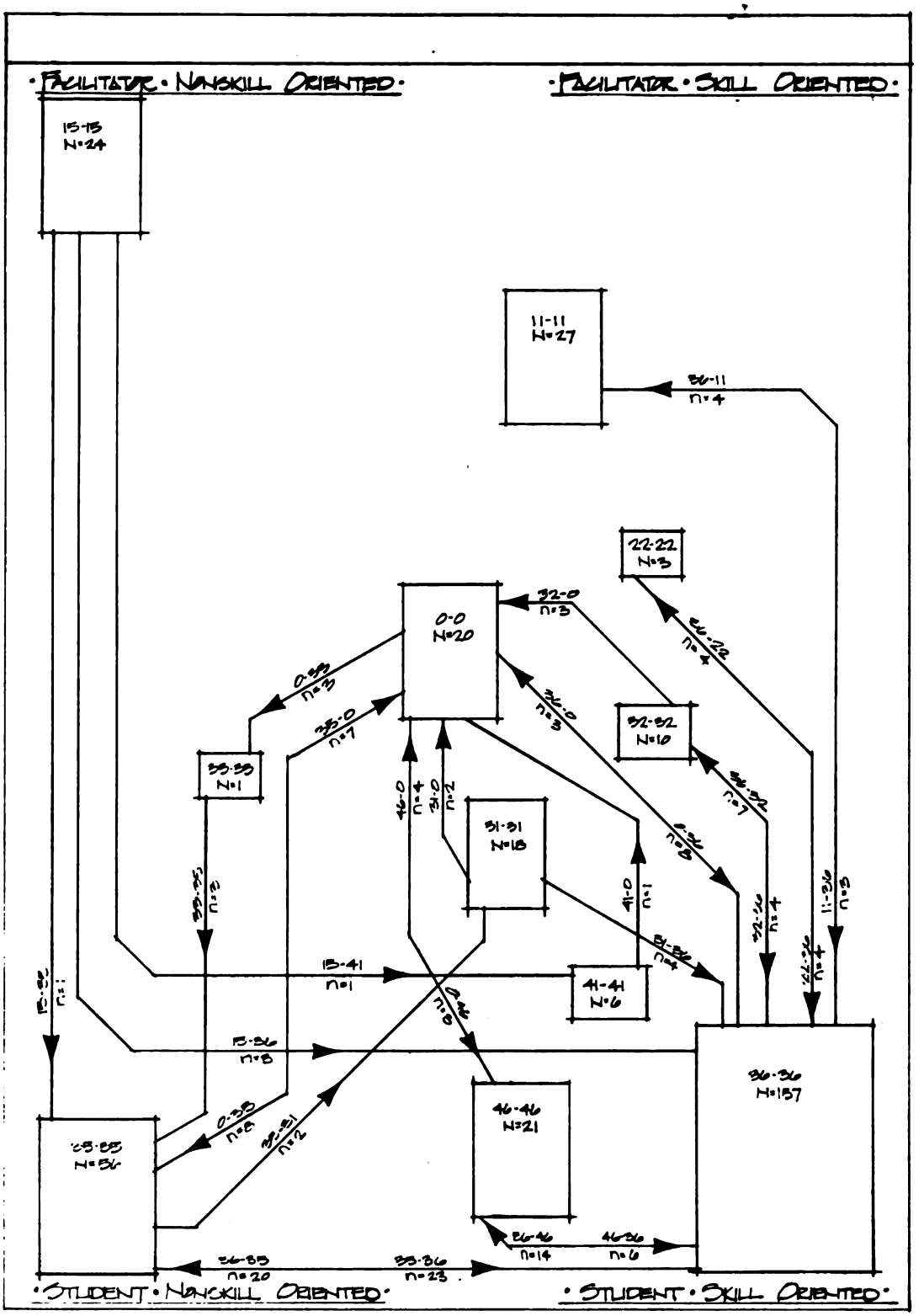
Appendix I: 1b.--Group 1 (effective) Flow Chart.



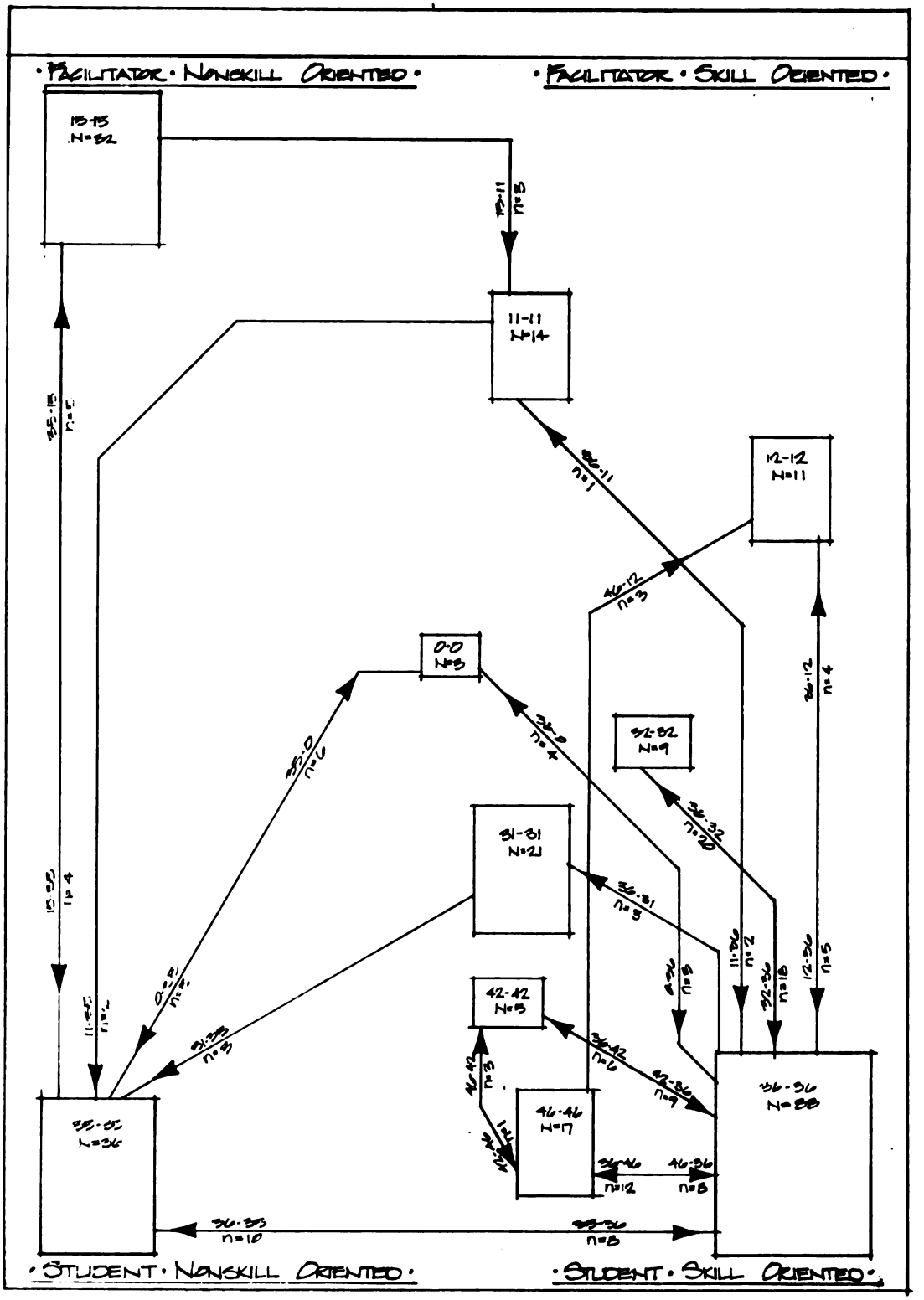
Appendix I: 2b.--Group 2 (effective) Flow Chart.



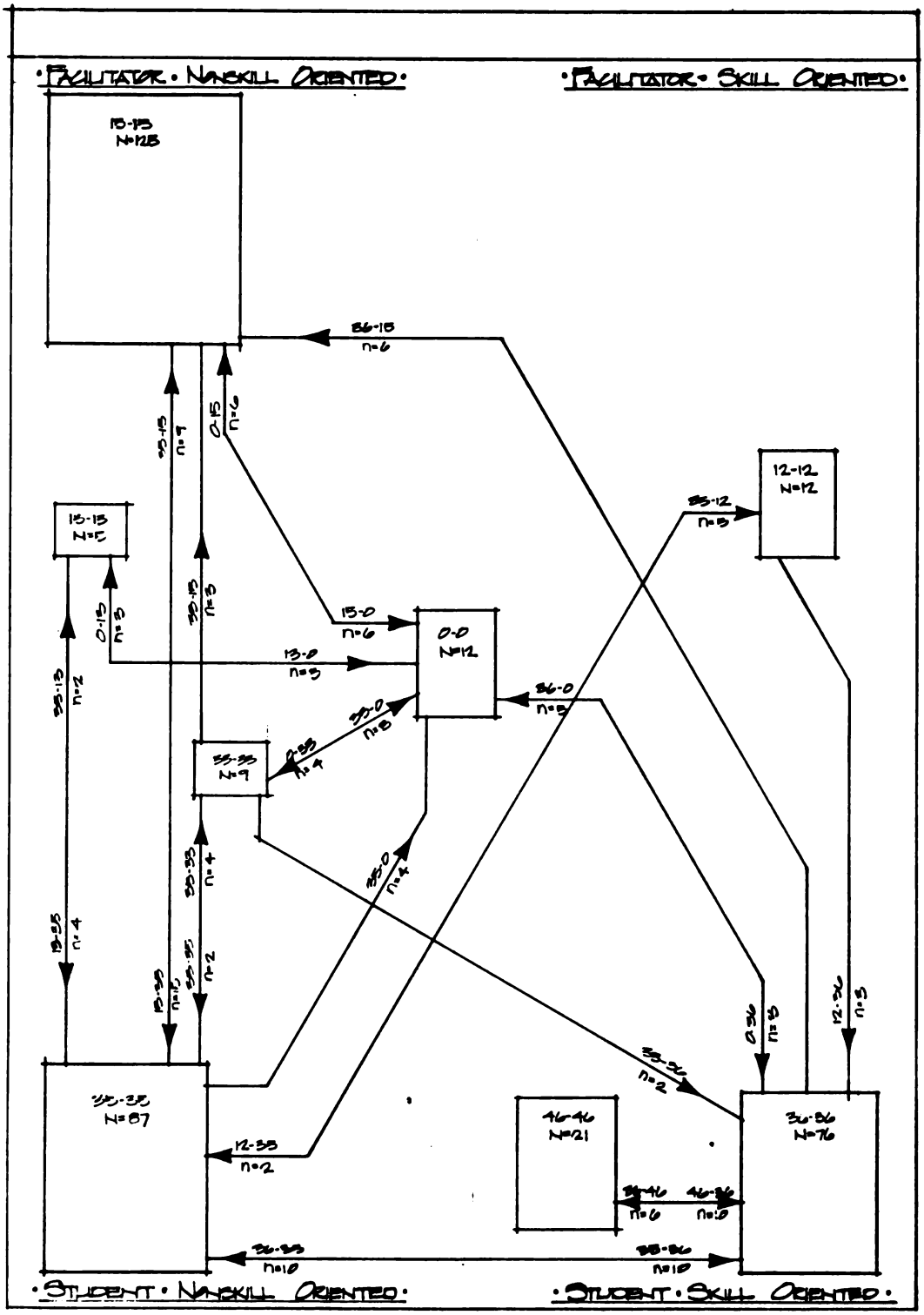
Appendix I: 3b.--Group 3 (effective) Flow Chart.



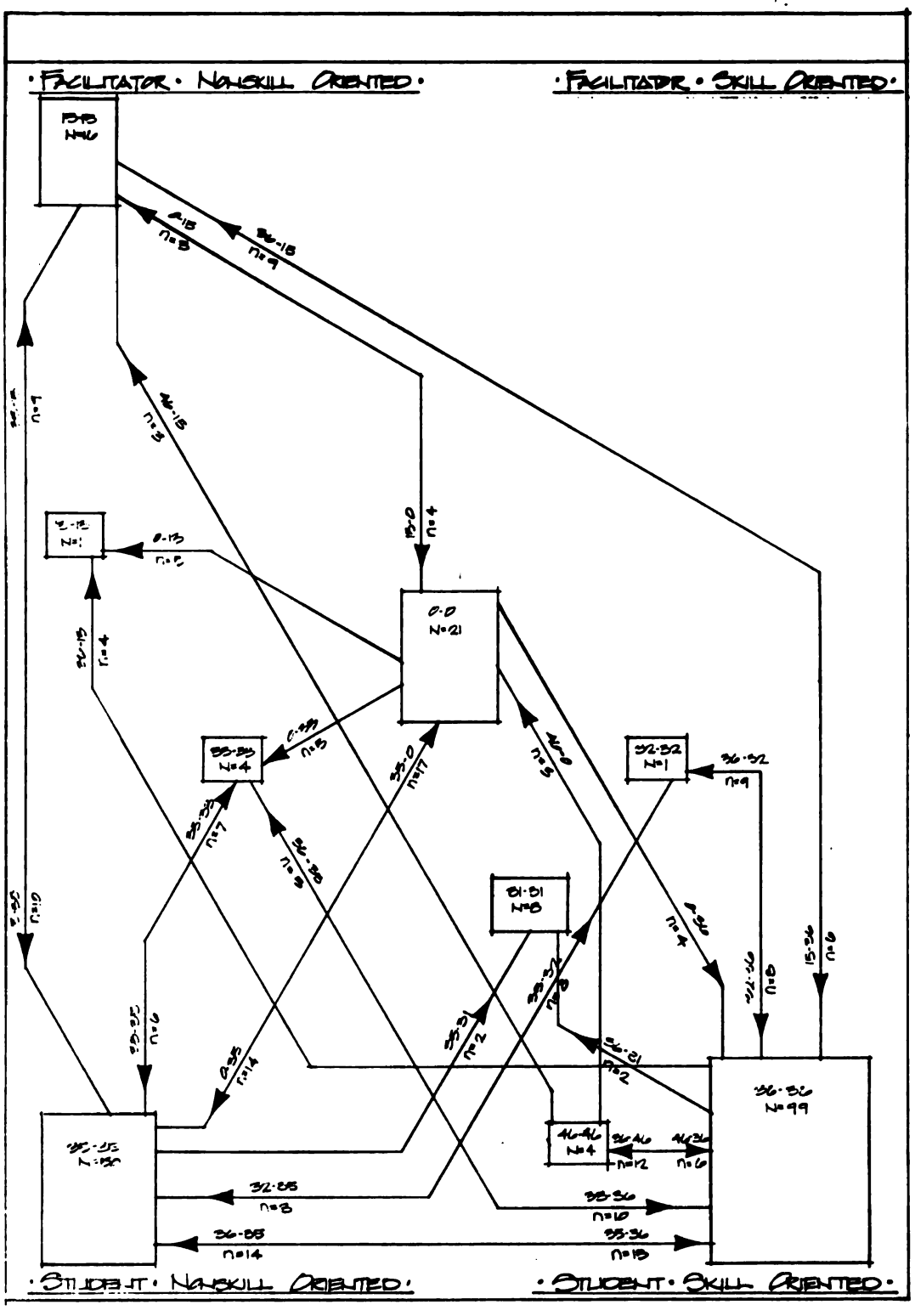
Appendix I: 4b.--Group 4 (effective) Flow Chart.



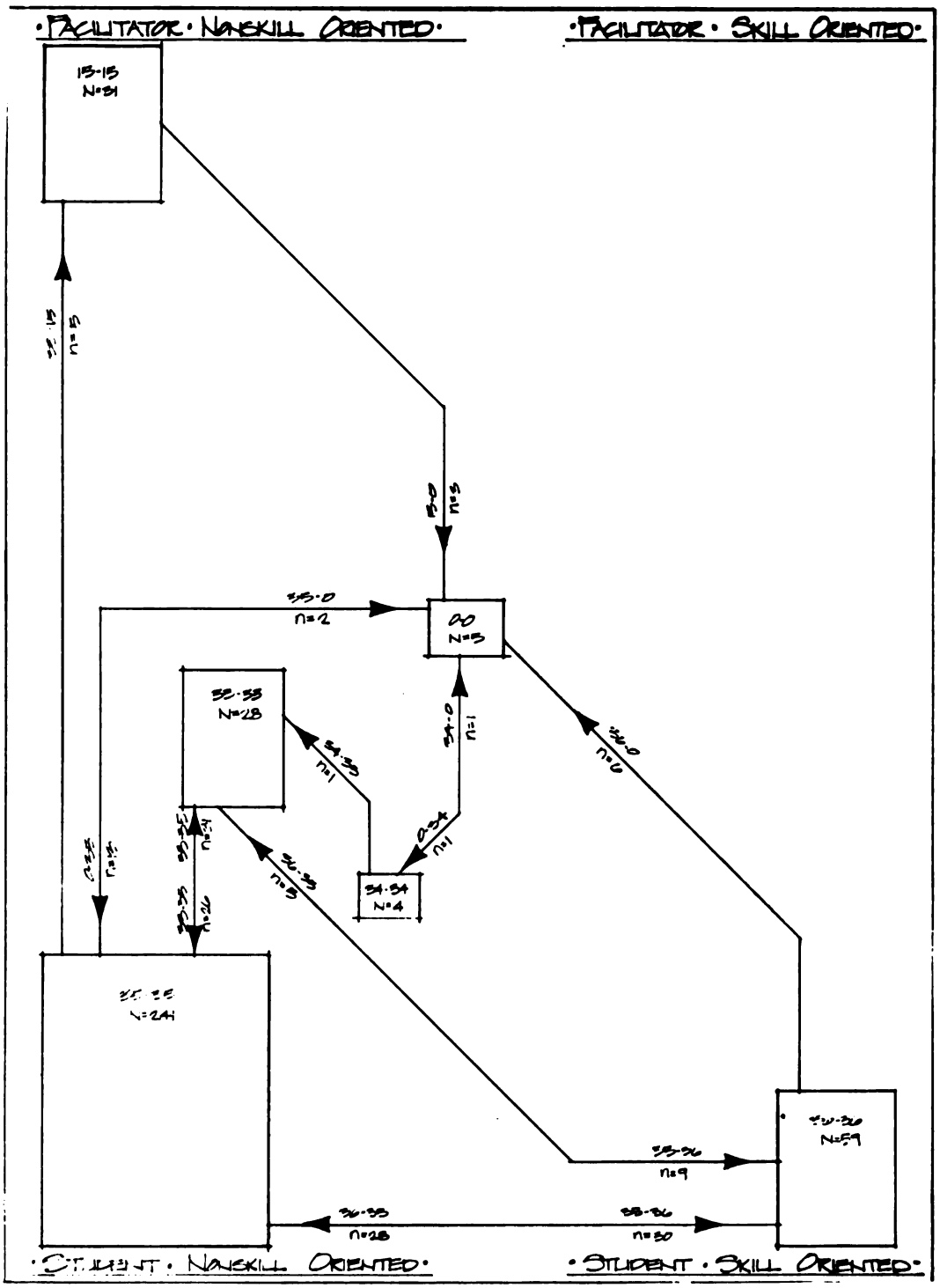
Appendix I: 5b.--Group 5 (effective) Flow Chart.



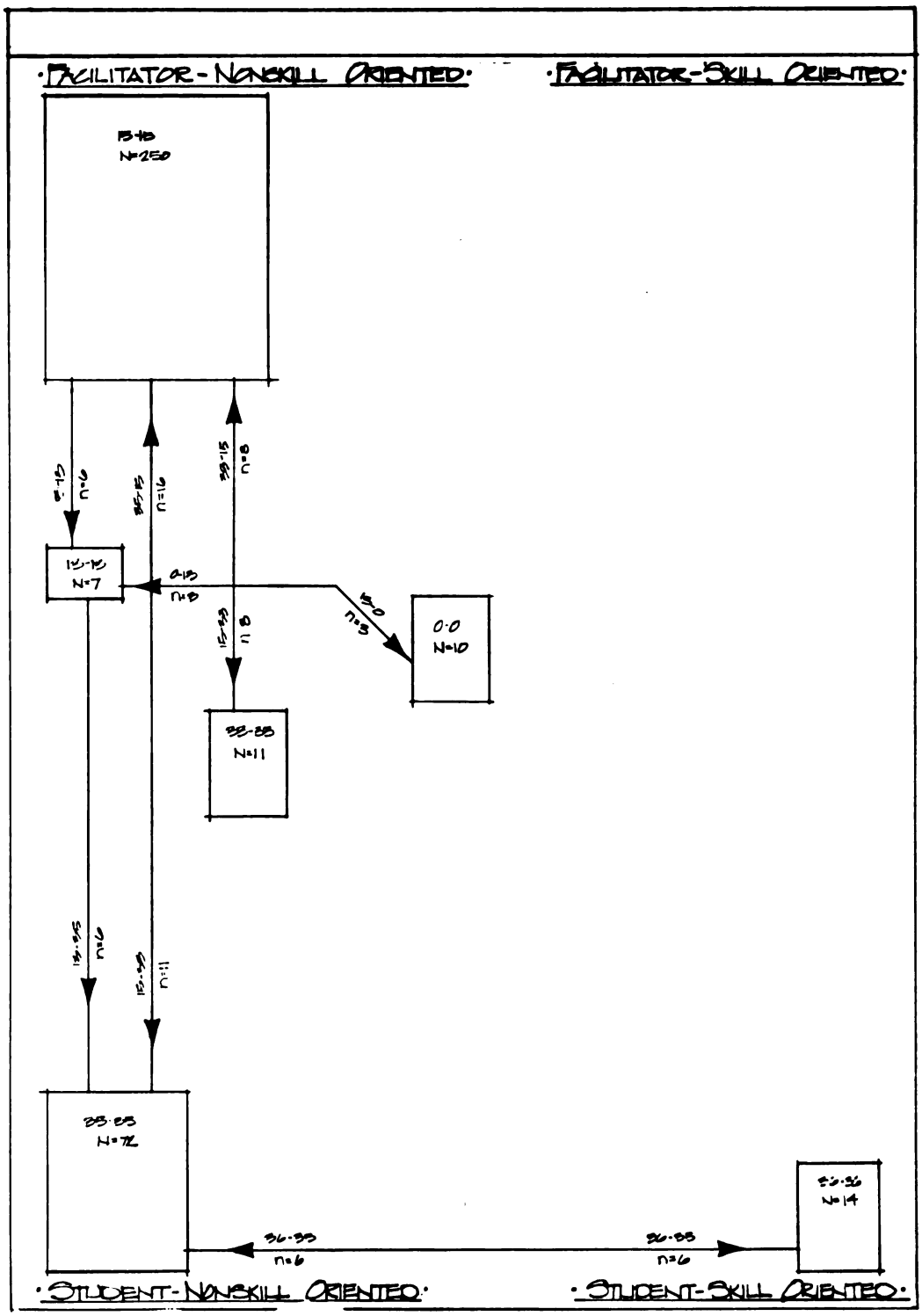
Appendix I: 6b.--Group 6 (ineffective) Flow Chart.



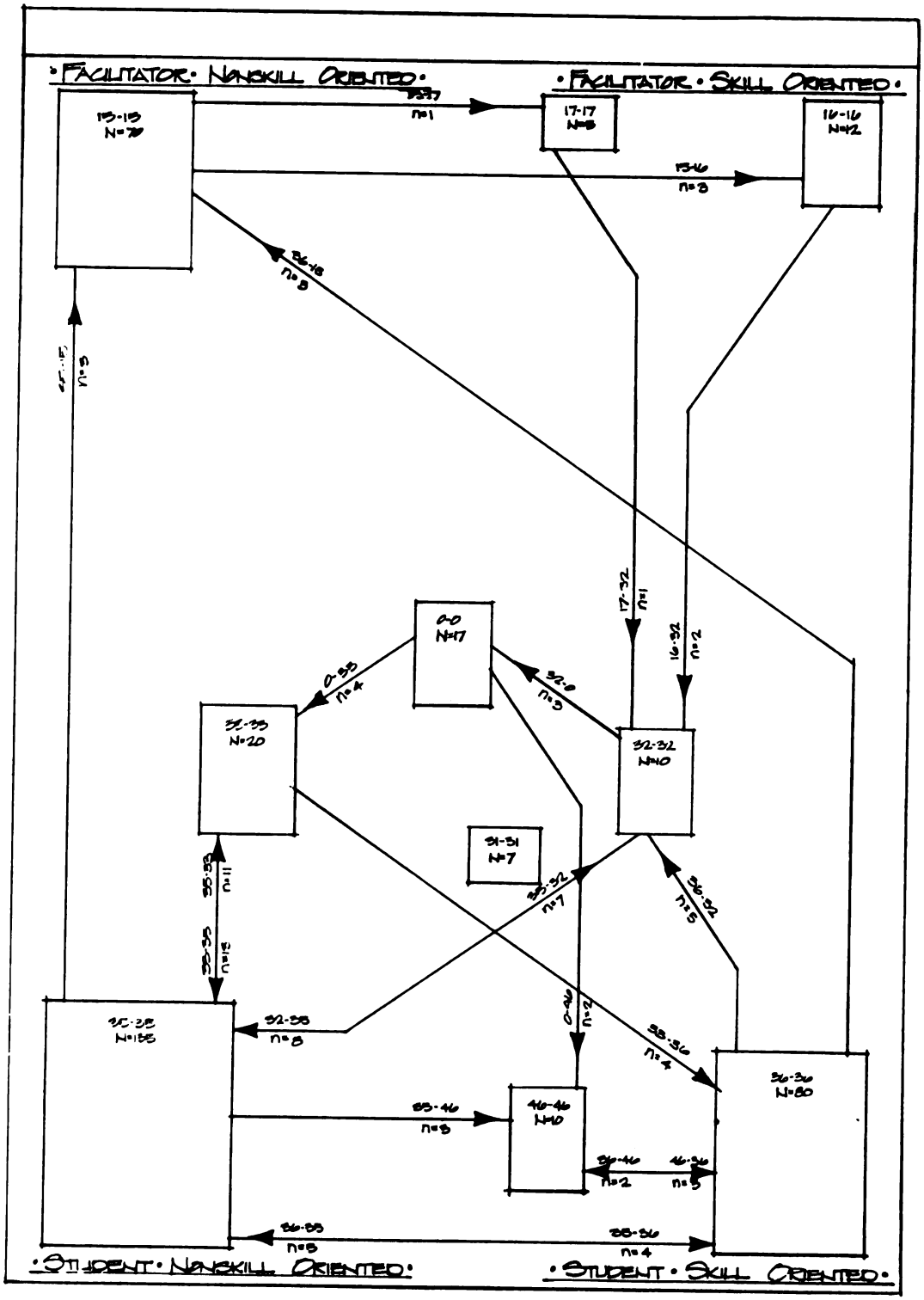
Appendix I: 7b.--Group 7 (ineffective) Flow Chart.



Appendix I: 8b.--Group 8 (ineffective) Flow Chart.



Appendix I: 9b.--Group 9 (ineffective) Flow Chart.



Appendix I: 10b.--Group 10 (ineffective) Flow Chart.