5-1987

Verbal Interactions Among Elementary Students with the Jigsaw II Learning Method

Katherine Gilbert
Utah State University

VERBAL INTERACTIONS AMONG ELEMENTARY STUDENTS WITH THE JIGSAW II LEARNING METHOD

by

Katherine Gilbert

A thesis submitted in partial fulfillment of the requirements for the degree

of

MASTER OF ARTS

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1987

ACKNOWLEDGMENTS

I would like to express sincere appreciation to my major professor, Dr. Richard Powers, for help in designing the study and in particular for his encouragement and sense of humor. I would also like to thank Dr. Hyrum Henderson and Dr. Elwin Nielsen for their helpful comments. I am grateful to Maria Norton of the Colleges of Family Life and Education Computer Center for her help with the statistical analysis.

I am indebted to Sandy Wallentine, the teacher, who so willingly gave of her time in designing the curriculum materials and accommodating her classroom to the requirements for videotaping. She was so gracious in meeting the procedural demands of the project.

I would also like to thank the observers Mike Moulton, Steve Pemberton and Jennifer Bauer who gave so willingly of their time for the analysis of the videotapes.

I am grateful to Cache County School District for permission to do the study and particularly to North Park Elementary which welcomed the opportunity to be part of a research project.

Katherine Gilbert

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Verbal Interactions Among Elementary Students with the Jigsaw II Learning Method

by

Katherine Gilbert, Master of Arts

Utah State University, 1987

Major Professor: Dr. Richard B. Powers
Department: Psychology

The cooperative learning method, Jigsaw II, was implemented in a grade four social studies class for the purpose of examining the verbal interactions among students as they learned from each other. Jigsaw II is structured to enhance cooperation because each student has exclusive information that is needed by other group members to do well on a test. It was hypothesized that the more capable students in a heterogeneous learning group would help the less capable ones learn the material. As the lower ability students gained proficiency in teaching their information, the variance in the rates of speaking would be less at the end of the implementation of Jigsaw II than at the beginning. This did not happen. There was homogeneity of variance between the rates of speaking at the beginning and the end. The rate of positive verbalizations (learning the information and group functioning) was over 80% at the

beginning and increased slightly during the implementation of Jigsaw II, but was not statistically significant. There was large variability in the rates of verbalizations among students, as well as large variability in rates of speaking for individuals across different learning group sessions. Any trends in changes of rates of speaking were obscured by the high variability. The verbalization rate of the high ability students doubled, the rate of the middle ability increased 32% and the rate of low ability students remained unchanged. On five quizzes administered over the learning unit, the high ability students attained the highest quiz scores, but the low ability students performed as well as the middle ability students.

(83 pages)

INTRODUCTION

Cooperative learning is a technique in which students come together in small groups to help each other learn; it is different from teacher-directed learning in small groups because the students work together without direct supervision. The group members, four to six in number and of varying abilities, become responsible for each other's learning. The role of the teacher changes from the prime dispenser of information and supervisor to a consultant that helps facilitate group functioning either in the interpersonal or academic domain. Cooperative group learning has been used successfully across all academic subjects and grade levels (Slavin, 1980b).

There are several ways to foster cooperation between students for learning. The developers of group learning techinques at Johns Hopkins University (Slavin, 1980a) use a group reward structure. When a member of the team does well when competing against others of similiar ability, points are won for the team. The highest scoring team for that week is given special recognition.

Another method, known as the "Jigsaw Classroom" (Aronson, 1978), fosters helping behavior between students by giving each group member exclusive access to information that everybody in their group must know for

an individually administered quiz. More capable students must help the poorer performing student if they want to have this information for the test.

Still another technique is that used by the Johnsons (1975) who believe that working together is natural for children. They advocate the formation of a cooperative goal structure by the teacher to foster group learning. The teacher encourages group work by presenting problems in a manner that allows the group to solve the problem as a unit. They share materials and ideas, divide up the labor, and when the task is completed they are rewarded as a group. Through interactions in the group they begin to view each other as prime sources of information instead of the teacher.

Cooperative group learning has produced increases in academic achievement and self-esteem as well as liking for school and peers. Although many studies have examined the effects of cooperative group learning, very few studies have examined what the students say to each other when they are working together to produce these positive outcomes. The purpose of this investigation was to examine the verbal interactions of students working in a cooperative learning environment.

Types of Cooperative Learning

Cooperative learning groups can be broadly characterized into two kinds. One kind uses a group reward structure, promotes high task interdependence, and provides for individual accountability. The second variety, the group investigative approach, leaves the decision of the learning task and how it is to be accomplished up to the students. The product is a group one and its evaluation is made by both the teacher and other groups.

Group Reward and High Task Interdependence Methods

The group reward methods developed at the Center for Social Organization of Schools, The Johns Hopkins University, are Student Teams-Achievement Division (STAD), Teams-Games-Tournament (TGT) and Jigsaw II (Slavin, 1980a). With STAD, students are assigned to teams which are heterogeneous with regard to academic ability, gender, and ethnic origin. The students are responsible for mastering material presented by the teacher. This can be accomplished through group discussions, questioning, or work sheets. The incentive for making sure that everybody masters the material is a quiz taken individually at the end of the topic. The greater the improvement in a student's score over past performance the more points he contributes to the team.

This system allows students of all abilities to contribute the maximum number of points to the team. The team that earns the highest number of points for the week is given special recognition in a class newsletter and if necessary another favor, such as free time.

TGT uses the same team formation procedure and rehearsal of materials as STAD but academic game tournaments are used instead of quizzes to determine improvement in performance. Team members compete with students from other teams who are of the same academic ability. The top scorer in the tournament receives the greatest number of points for his team while the middle and low scorer receive proportionally reduced scores. The team with the highest score receives recognition as described for STAD. Jigsaw II teams are formed in a similiar fashion but in addition each team member is designated as an "expert" in a particular subtopic of the material being studied. Experts from different teams meet to discuss their topics and then go back to their teams to teach the material to their groupmates. There is motivation to present the "expert" material well and for team members to encourage a good presentation because a quiz is taken individually over all the material. Team scores and rewards are the same as for STAD.

Aronson's (1978) original jigsaw classroom uses high task interdependence but no group reward. Heterogeneous groups are formed on academic, gender and race variables. Members of teams consisting of five or six students are given exclusive access to task related information. After meeting in groups with students assigned to the same subtopic, they return to their own team and teach it to the other members. Learning from a groupmate is the only way to get the information which will be on the quiz at the end of the unit. Poorer performing students are soon helped by more able ones as higher ability ones learn their is no advantage to ignoring or ridiculing the less capable. Poorer students start to become valued contributing group members.

The Group Investigative Method

Johnson and Johnson (1975) advocate the group approach to learning through the development of a cooperative goal structure as opposed to an individualistic or competitve one. A cooperative goal structure exists when students perceive that they can only obtain their goal when others also obtain it. This contrasts with an individualistic goal structure where achievement by one student is unrelated to others, or a competitive one where the success of one student means another fails. A cooperative goal structure is

established in the classroom by the teacher presenting the learning task to the group as a whole, encouraging the sharing of ideas and material, and rewarding the group as a unit. The Johnsons use the example of determining how long a candle will burn in a glass jar. Instead of producing a single answer, the students speculate on what factors determine the length of burning (eg. the shape of the jar, whether or not the jar has been used for a previous burn) and record their findings on a common worksheet. They are invited to compare their results with other groups and speculate on why everybody's results are not the same. The students are encouraged to decide on their own how to divide up the task so as to produce the best solution. Cooperative learning contains the essential elements of positive interdependence, face-to-face interaction among students, individual accountability for mastering material, and appropriate use of interpersonal and small group skills (Johnson & Johnson, 1983).

The group investigative method is based on group inquiry in a social context (Sharan, Kussell, Hertz-Lazarowitz, Bejarano, Raviv, Sharan, 1985). According to Sharan (1980), the Johnsons' cooperative learning methods promote social interaction but not problem solving and interpretation. The academic material to be learned is primarily facts and

information. Sharan, Ackerman, and Hertz-Lazarowitz (1979) have developed a cooperative learning method for promoting higher level cognitive processing such as analysis and judgement. It is more complex than the Johnsons' because more planning is left up to the students. Although the general problem area is still delineated by the teacher, the students decide on the subtopics. With each group working on a different subtopic, group members decide how to divide up the task and select means for accomplishing the task. They evaluate and summarize the information for presentation to the class. Both teacher and students evaluate the final product (Sharan, 1980).

Research Results of Cooperative Learning Methods

The effects of cooperative learning techniques have been measured on such variables as academic achievement, liking for school and peers, self esteem, race relations and mutual concern. Slavin (1980b) points out the difficulty of comparing studies due to different methodologies. Some studies have used curriculum specific tests and other standardized tests. The primary comparisions have been made between classes where cooperative learning has been introduced and control classes where regular classroom instruction, usually whole-class, teacher-directed, instruction, has continued.

Student Team Learning
Developed at Johns Hopkins

Slavin's (1980b) summary of Teams-Games-Tournament (TGT) research (12 studies) and Student Teams-Achievement Division (6 studies) demonstrated statistically significant positive results in the experimental group for academic achievement, race relations and mutual concern across a wide variety of settings and subjects. For TGT all twelve studies measured academic achievement with either a curriculum specific or standardized test and some used both. Ten studies showed a statistically significant effect for achievement in favor of the experimental group. Race relations were improved in three out of four studies and mutual concern in five out of seven. Only two studies, using social studies as the subject area, found no difference between the experimental and control classrooms on achievement. Neither of these two studies measured changes in race relationships, but one found a statisticlly significant improvement in mutual concern. Examination of five studies in language arts, and one in mathematics for Student Teams-Achievement Division (STAD) showed a statistically significant postitive effect for academic achievement in four studies for the experimental group. These results are impressive because the control groups in these studies used exactly the same academic materials rather than just being held

to the same objectives. Three of the studies which measured changes in race relationships showed a positive effect, and five studies which measured mutual concern demonstrated improvement in three.

Two studies done by Hulten and Devries (1976) and Slavin (1979) attempted to determine whether it is the opportunity for peer tutoring or the group reward structure that accounted for increases in achievement with TGT and STAD. Both studies concluded that it was the group reward. The Slavin (1979) study also demonstrated that the focused schedule of instruction contributed to the success of STAD.

Slavin and Karweit (1981) demonstrated that cooperative learning techniques could be used as a serious alternative to traditional teaching methods, that their postitive effects were not due to novelty, and that they could be used for a major part of the school day. They implemented TGT for mathematics, STAD for language skills and Jigsaw II for social studies in ten experimental classrooms. Using the Comprehensive Test of Basic Skills (CTBS), they demonstrated a statistically significant effect for reading vocabulary, language mechanics, and language expression. On the affective measures, the experimental groups had statistically significant increases in self-esteem, liking for school, the number of friends in school, and

a decrease in the number of non-friends in school.

The Jigsaw Classroom

Aronson's (1978) jigsaw classroom developed out of a need for better race relations in the desegregated Austin, Texas city schools. In a series of field experiments the effect of the jigsaw technique on academic achievement, liking for school and peers, self-esteem and egocentrism were examined. Blaney, Stephan, Rosenfield, Aronson and Sikes (1977), using social studies in fifth grade, found that students in classrooms using the jigsaw technique liked their groupmates better at the end of the experiment than at the beginning. This increase in liking did not come at the expense of disliking others outside the group. On the liking-for-school measure, Anglos in the experimental group compared to the controls increased their liking for school; blacks and Mexican-Americans decreased slightly. However, in the control group, blacks decreased markedly in their liking for school, while Mexican-Americans increased. The authors speculate that the Mexican-American children were uncomfortable with the verbal demands made upon them in the jigsaw classroom compared to the controls who could avoid verbal interactions. On the self-esteem measure experimental subjects increased significantly while controls decreased. A followup study done by Geffner

(cited in Aronson & Osherow, 1980) found that Mexican-Americans increased their liking for school when they learned social studies with the jigsaw technique. This study lasted two weeks longer than Blaney et al.'s (1977) and the school had an almost 50% ratio of Mexican-Americans who had been in the school for several years.

Lucker, Rosenfield, Sikes and Aronson (1976) examined the effects on achievement for social studies with fifth and sixth graders using jigsaw and found that students in the experimental class performed significantly better than the controls. However, the improvement in performance was due primarily to a substantial improvement in minority performance in the jigsaw classroom. Since, this experiment lasted only two weeks, it seems likely that an improvement in Anglo performance in the experimental classroom compared to the control would be detected with a longer exposure to the technique. Bridgeman (cited in Aronson and Osherow, 1980) found that jigsaw students exhibited less egocentrism than those taught by traditional methods or in small-group classrooms. Students in this study were shown a series of pictures depicting an event with the final scene only explicable in terms of preceeding ones. For example, a boy was shown crying over the departure of his father on an airplane and then later

was upset when a mailman delivered a package containg a toy airplane. The mailman was not privy to the information that the father had left on an airplane. Children in the jigsaw group were more likely to say that the mailman would be confused by the crying than the control group which would say that the mailman knew his father had left on the airplane.

Group Investigative Methods

Johnson, Johnson and Maruyama's (1983) meta-analysis of cooperative versus competitive and individualistic learning experiences concluded that cooperative learning promoted greater interpersonal attraction among homogeneous students, students from different ethnic groups, and handicapped and non-handicaped students. Johnson and Johnson (1983) concluded that cooperative learning is just as effective as competitive or individualistic learning on achievement. For some tasks, such as concept attainment, verbal problem solving, categorization, spatial problem solving, retention and memory, it is more effective.

Cooperative group learning promotes higher cognitive processing than in an individualistic condition where students work on their own (Johnson, Johnson, Roy & Zaidman, 1985). They assessed three levels of cognitive processing through analysis of

verbal interactions. Using a grade four social studies unit, they defined low cognitive processing as repetition of information, an intermediate level as giving new information and a high level as elaborations, explanations or relating to previously learned information. With their verbal interaction instruments which measured factors other than cognitive processing, they determined that of the total talking time 1.8% was devoted to low level processing, 46.8% to medium level and 11.5% to high level.

Sharan et al. (1979) measured cognitive processing with achievement tests based on Bloom's taxonomy. An example of low level cognitive processing was recall of information and sequencing of events, while higher cognitive processing was represented by questions requiring analysis of problems, judgement and evaluation. Classes taught by the whole-class traditional method were compared with ones divided into small groups. There was no difference between the classes for low level items, but the group learning method produced superior achievement on higher order thinking. Sharan (1980) collected essays written by the students in the experimental classrooms asking them to express their feelings about small group learning compared to their experience in traditional whole-class instruction. They were scored by outside examiners and

not seen by the teachers. The children expressed a greater sense of independence and responsiblity as well as being better accepted by other children because they were listened to.

The Johnson et al. (1985) study examined the verbal interactions of small cooperative learning groups in an attempt to link types of verbal interactions with achievement. Using a grade four social studies unit, they demonstrated that 80% of oral interactions dealt with the exchange and elabortion of task related information, 10% with encouraging learning and disagreeing, and 10% with non-task related material. Whether a student is speaking or listening in the group appears to affect achievement. Although there is only a moderate correlation (.35, $p<.05$) between achievement and speaking about task-related information for students of all abilities, there is a marked negative correlation when low ability students are listening to new facts and information (-.56, $p<.10$) or to questions (-.56, $p<.10$). Therefore, low ability students benefited by talking about the information to be learned. Yager, Johnson and Johnson (1985) found greater achievement when a group leader was designated daily to lead the discussion.

Johnson and Johnson (1985) have developed several instruments for recording verbal interactions. Their most recent instrument provides data about the speaker,

the receiver and the type of verbal interactions. Verbal statements are characterized into four modes: (a) the task mode which is a statement about the academic material, (b) the management mode which is a statement about classroom structure, rules, directions or plans, (c) the social mode describes a person's feelings and emotions and interactions about non-task and non-management subjects, (d) the process mode which are statements about group participation, communication skills, and group process skills. The Johnsons (1985) found that 77% of the verbal interaction in cooperative groups was directed toward the task, 17% to management, 5% to social and 1% to process.

The results of the Johnson et al. (1985) and Yager et al. (1985) studies may shed some light on why the jigsaw classroom is successful in fostering better achievement. Does the structure of the jigsaw learning group promote talking about the subject matter which is correlated with achievement? Does a group leader emerge who facilitates learning by everybody in the group? In the Lucker et al. (1976) study, it was the improvement in the minority students' performance that accounted for the increase in achievement in the experimental group. This improvement could be accounted for by a larger percentage of time spent on task (ie. exchanging task related information) or it could be that the lower

achieving students were given the opportunity to talk and thus practice the material. Everyone must impart his/her "expert" information to the group, providing everybody with the chance to talk. The "expert" in the jigsaw classroom closely parallels the leader designated in Yager et al.'s (1985) groups. Exchange of task related information in the jigsaw classroom is done in a positive social context because other group members need the information.

PURPOSE AND OBJECTIVES

Many studies have compared cooperative learning techniques with other methods of instruction. Achievement and social-emotional variables have been measured before and after the intervention, and the findings are in favor of cooperative learning methods. Examination of all the variations in cooperative group learning reveals several factors that are associated with increases in achievement, mutal concern and self-esteem. They are: (a) a group reward structure, (b) very specific learning objectives, (c) the opportunity to talk, plan and interact and (d) the opportunity for higher cognitive processing. Very few studies have examined in detail what the students are saying to each other as they work together in groups. Does the small group finally provide the lower achieving student the opportunity to talk which increases self-esteem and eventually the motivation to engage in the intellectual task? Do groupmates provide encouragement and support giving confidence to express oneself no matter how feebly? Does the nature of the verbal interactions change over the implementation of a cooperative learning technique? Are there perhaps more support and encouragement statements with time? Do task-related discussions increase as members gain

confidence and grow to know each other's strengths and weaknesses?

Recent verbal interaction studies (Johnson & Johnson, 1985; Johnson, et al., 1985) have been done in classrooms in which cooperative group learning is a regular part of the school curriculum. They have characterized the verbal interactions that occur in a classroom with a well established cooperative goal structure. No studies have documented changes in the frequencies and types of verbal interactions that occur during the implementation of a cooperative learning method. Observation of verbal interactions during the implementation of a cooperative learning technique has the potential for discovering what type of verbal interactions contribute to the positive effects for achievement and social-emotional variables.

Since Jigsaw II has all the factors associated with positive outcomes for group learning, its use would be appropriate for examining the types and changes in verbal interactions that occur when a cooperative learning method is introduced. If students are indeed helping each other learn, changes should occur in the type and frequency of verbalizations among the students. It was predicted that in the beginning, the more capable students would speak more because of their verbal fluency, but over the implementation of Jigsaw II the

less capable should talk more as they became more proficient at teaching their "expert" material. They should also score higher on tests compared to scores on tests before the implementation of Jigsaw II. Since everybody cannot talk at once, it was hypothesized that the higher ability students would be speaking less at the end.

Therefore, the objectives of this study were to:

1. determine if the homogeneity of variance changed in the rates of verbalizations from the beginning to the end of the implementation of Jigsaw II.
2. determine if the types of verbalizations changed during the implementation of Jigsaw II.
3. determine whether the type or frequency of verbalization was associated with a student's ability (as rated by the teacher) in social studies.

METHOD

Sample

Subjects

The subjects were 23 students in fourth grade in an elementary school of Cache County School District. The students were assigned to three, five-member and two, four-member learning groups according to the following procedure. They were first ranked from 1 to 23 by the teacher according to their ability in social studies. Students ranked 1 through 5 were assigned to groups 1 through 5 (eg. Rank 1--Group 1, Rank 2--Group 2,) and then ranks 6 through 10 were assigned to groups, reversing the group order (eg. Rank 6--Group 5, Rank 7--Group 4). The assignment of Ranks 11--15 was the same as for Ranks 1--5 (eg. Rank 11--Group 1, Rank 12--Group 2). This procedure produced heterogeneous learning groups with regard to ability in social studies and gender. Some minor adjustments in group composition were made to ensure that there was a girl in each group as only 7 of the 23 students were female.

In order to determine improvement over past performance on social studies quizzes using Jigsaw II, the teacher assigned a base score to each student, which was approximately five points below what he/she would be expected to score on a quiz. With 30 points possible on

the quiz, the two highest ranked students were assigned a base score of 20, 19 for the third and fourth ranked students to 10 for the two lowest ranked students. Students were not made aware of either their rank or base score, but were told that they could make points for their team if they scored higher than their usual score on social studies quizzes. For example, if a student had a base score of 17 and scored 27 on a quiz out a possible 30, he received 10 improvement points, the maximum points possible. A student with a base score of 20 and a quiz score of 27 received 7 points. A perfect quiz score automatically received 10 points. There was no penalty for a quiz score less than the base score; a 0 was assigned. After the third quiz, each student's base score was adjusted so it was five points below the average of his/her three quiz scores.

Curriculum

The learning unit was about Indians of Utah, part of the social studies curriculum. It was divided into 5 topics and each of those 5 topics was further divided into 5 "expert" subtopics. The expert information was printed on a sheet which was given to the students in their learning groups. The schedule for each of the five topics was as follows: (a) general introduction of the topic by the teacher while students were seated in their learning groups (5 - 10 minutes), (b) dissemination of

the expert handouts (2 minutes), (c) meeting of the "expert" groups (10 - 15 minutes), (d) meeting of the "learning groups" 15 - 30 minutes), (e) review sessions in learning groups and administration of the quiz (20 - 30 minutes). On the first day of the cycle the topic was introduced by the teacher and the expert groups met for ten to fifteen minutes. On the second day the learning groups met. Some topics required a second meeting on a succeeding day for the learning groups. There was always a review session on the day of the quiz. While the expert or learning groups met, the teacher circulated to give assistance with the academic material or facilitated appropriate communication among team members. On the day after the quiz, each team was given a summary sheet which listed each member's improvement score (how much the student scored over his/her base score), the total team score (sum of individual improvement points) and the team's standing in relation to the other teams' performance for that quiz. At the same time everybody received a newsletter announcing the winning team and individuals who had contributed the most points to their teams. The individual quizzes were returned after giving the teams several minutes to study the team summary sheet and the newsletter.

Implementation of Jigsaw II

Teacher Training

The teacher was trained in the use of Jigsaw II through a pilot study conducted in the spring of 1986. The Utah Indian topic was not used in the pilot study, because it had already been taught that year. It was selected for this study because of student interest and enthusiasm in previous years. While this study was in progress, the teacher taught another fourth grade class the same topic using Jigsaw II at an earlier time. This allowed the teacher to detect any difficulties with the written material or with the time alloted for a particular subtopic. The class used in the study was chosen because of a more flexible schedule which permitted an extension of the lesson if neccessary.

Introduction of Jigsaw II to Students

When the teacher introduced Jigsaw II to the students she followed the procedure outlined in the teacher's manual Using Student Team Learning (Slavin, 1980a). The students were told that they would be learning about Utah Indians in a new way. They would be in learning groups where each member would need to teach his expert topic to teammates. They would first meet with other students who had the same expert topic so

they could help each other learn the material. The students were told that they must learn the material well enough to be able to teach the material in their own words and were specifically told not to read the expert information to their team mates. A quiz would be taken individually over all the information. Students who improved the most over previous social studies scores would contribute more points to their team than those who did not improve. Everybody had an equal opportunity to contribute points so they should work hard to help each other because team scores were figured by summing individual improvement scores. After each quiz, a newsletter would be published with the winning teams and the names of students who had contributed the most points to their team.

Pre-Implementation Procedures

Before the actual implementation of Jigsaw, the students met as a group to decide on a name. The teacher lead a discussion on ways to help each other learn. They reviewed the importance of eye contact when listening, how to ask questions and how to handle a situation when a group member was not listening or disturbing the group. On the follwing day, these skills were briefly reviewed and the teams played the "broken squares" game, a group cooperative exercise described by Aronson (1978). During these two days, the students

were videotaped to acclimatize them to the presence of the cameras.

Before Jigsaw II was implemented and during the week after the study was finished, the teacher completed a Behavioral Academic Self-Esteem Rating Scale (Coopersmith, 1982) on each student. This scale consists of five factors, of which the following are abbreviated definitions taken from the test manual:

> (a) Student Initiative - how often students participate in classroom activities, (b) Social Attention - how well the student "fits into" the classroom envrionment, (c) Success/Failure - how successfully students cope with failure, criticism, correction, admonitions and other responses that could be perceived as negative, (d) Social Attraction - how compatible youngsters are with peers, (e) Self-Confidence - a youngster's level of verbal expression about school accomplishments. (p. III-1)

Adjustments to Jigsaw II Procedure

After the first week of Jigsaw II, the teacher became concerned that the students were still reading their expert material to group members and that students were not helping each other learn the information. Between the fifth and sixth session, three brainstorming activities, which according to Kagan (1986) promote

group cooperation, were done. For example, within one minute, they were to think of how many combinations of 2 numbers equal 25. They also watched a five minute videotape exerpt from the pilot study of a student teaching his expert material without reading it. After the viewing, the teacher invited the students to try teaching in their own words and led a brief discussion with them on how they could accomplish it.

Data and Instrumentation

A schedule was developed to videotape on a rotating basis three of the five groups each day during a learning session. There were a total of 14 learning group sessions (15 - 30 minutes duration), resulting in one group being videotaped 7 times, three groups 8 times and one group 9 times. The study lasted 4 weeks. The verbalizations occurring during the learning sessions were coded by trained observers into categories summarized below. These categories were developed from the Johnson and Johnson (1985) verbal interaction measure which coded the speaker, the receiver and the type of verbalization. In their study, the observer recorded in the classroom a two-minute continuous segment of all group members' verbalizations. The observer would then move to another group. The Johnsons categorized verbal statements into a "task mode",

"management mode", "social mode" and a "process mode". The four verbalization categories used in this study were as follows (see Appendix A for the complete definitions):

1a. On-task - Information (OT-I): a statement or a question that is an exchange of task related information. eg. a pueblo is an Indian village.

1b. On-task - Elaboration (OT-E): any statement that elaborates on the material or a question about the material that cannot be answered directly from the expert sheet. eg. Debbie's mother brought a craddelboard last year.

2. Not-on-Task (NT): a statement that is not related to the material being studied eg. let's play after school.

3a. Management-Positive (MP): a postive statement relating to the functioning of the group eg. it's Greg's turn.

3b. Management-Negative (MN): a statement that is critical of the group's functioning eg. you took too long.

4. Miscellaneous (MS): a statement that does not fit into the above categories or cannot be deciphered.

Coding

To code one learning session, the observer watched the tape without stopping and recorded the elapsed time. During this viewing, the observer selected the student whose verbalizations were the clearest and recorded that person's speaking on the second viewing. Only one person's verbalizations were coded with each viewing. Viewing the tape as many as six times increased the chances of deciphering complex interactions or hard to hear passages. Each incidence of speaking was recorded in one of the four categories, (see Appendix B for coding form). To be scored, the verbalization had to contain a fact or a comment, but not necessarily be grammatically correct. If the student paused and then made another statement, it was counted separately as long as it made sense on its own. For example, if a student said "they stole horses," paused and then said "they traded them," two verbalizations were scored. If both facts were said without pausing it was counted as one. If a phrase such as "it seems that" was said, no verbalization was scored. When a student was reading the expert material, the stopwatch was started and was stopped when he/she paused or said "um," "and" or some other connecting comment. For every five seconds of elapsed time, one verbalization was scored in the appropriate category. If the student, while reading the

expert information in the on task-information category, made a statement in a different category, that verbalization was recorded separately.

Observer Training

Three observers were trained using videotapes that were made in the pilot study. Each observer was provided with the definitions of the categories, the steps to follow for using the verbal interaction form (Appendix A) and copies of the expert handout. Copies of the expert material familiarized the observer with the subject matter. After the observer had learned the definition of the four categories, the investigator and the trainee observed one individual on the training tape. Each time the student spoke, the tape was stopped and the trainee coded the verbalization. An incorrect categorization was discussed with the investigator until agreement was reached. Coding of another student's verbalizations on the same tape was done alone by the trainee. Agreement to within 10% of the investigator's score constituted readiness to proceed. Each observer, before being assigned a group, coded the verbalizations of all individuals in Group 5 for the first learning session. The same 10% agreement criteria was required before coding of the assigned group began.

The investigator (KG) coded the verbalizations of

Group 1 and observer JB performed the reliability check. The investigator coded this group because one of the trained observers (MM) was unable to continue with the study. Although it would have been preferable to have a trained observer do the coding for Group 1, it is the investigator's opinion that her coding did not bias the study. The coding of Group 1 by the investigator was done concurrently with reliability checks for the two remaining observers. Since these reliability checks were satisfactory and the investigator's coding of Group 1 tallied with the trained observer JB, it appeared to be the best solution to the loss of one observer. The observer SP coded the verbalizations for Groups 2 and 3 with the investigator (KG) as the reliability check. JB and MM coded Groups 4 and 5 respectively with the investigator (KG) as the reliability check. To determine agreement between observers, JB coded four sessions of Group 3 scored by SP and four sessions of Group 5 coded by MM. She coded the verbalizations for two individuals not selected on the first check, performed by the investigator.

Reliability Procedures

For purposes of reliability, it was decided initially that three learning groups would be randomly selected for each week of the study and then two

students within each group randomly chosen. Interobserver agreement was not achieved using this method. Consultation between the observer and the investigator indicated problems in the following areas:

1. Disagreement as to who was talking. This difficulty was caused by several students talking at once, poor sound quality due to background noise, excessive movement of students as they jostled for positions around the table, tapping on the table and talking in silly voices so as to be unrecognizable. This difficulty was solved by analyzing the later learning sessions first. In the later sessions the tapes were technically better and the students were working better together. There was less physical movement around the table and the quiet ones were speaking louder. As the observers gained experience they were better able to code the tapes of inferior quality.

2. Students talking so softly that the microphone did not pick up their verbalizations. The soft talking was a problem for several students who were self conscious about the camera in the beginning. Soft talking disappeared after the fifth session. Although nothing could be done to enhance the voice level, the observers were able to pick up more cues through the experience of analyzing the later tapes.

3. Coding of the verbalizations when the students were

teaching their expert material. Students who read their material fluently presented no difficulties, but the ones who read haltingly and slowly did. The discrepancy in timings was solved by running the stopwatch only when the student was speaking. When the student paused or said "and" or "um" the stopwatch was stopped.

4. Redefining the unit of verbalization. The students tended to speak in short phrases and grammatically incorrect sentences. The expression of one idea or piece of information could include several pauses or one student would be interrupted by another. This problem was solved by redefining a unit of verbalization. It had to be a unit of information that made sense on its own (eg. they got horses by trading). If another piece of information was added after a momentary pause (eg. and they stole them), it was counted as a second verbalization. If the student said only "they got horses by," it was not coded.

5. Discriminating between an actual statement about the information and a statement relating to the information. Students responded to information with such statements as "okay", "I don't know" or "what". These verbalizations were to have been coded under on-task in the subcategory of "closure". A closure statement indicated the end of a verbal interaction. However, closure statements were the greatest source of

disagreement in the on-task category. Rather than code these closure statements as "on-task-closure", they were coded as management-positive statements.

6. Redefining the criteria for agreement. The frequencies within a category obtained by the observer and the investigator were averaged. If adding or subtracting 10% of the mean to either coder's score yielded the mean, then the score for that category was considered reliable. For purposes of reliability, the two subcategories of on-task-information and on-task-elaboration were combined and considered as one category. It was decided to combine the not-on-task and miscellaneous categories because they were infrequent and difficult to distinguish. It will be recalled that the miscellaneous category included verbalizations that could be heard but not deciphered. Some observers deciphered more than others. Since the frequencies of management-negative and not-on-task/miscellaneous statements were so low and a difference of only one incidence would not meet the 10% criteria, only the on-task-information and the management-positive statements were considered for determining observer agreement. If agreement was not obtained for these two categories, part of the learning session was reviewed with the investigator to determine the source of the discrepancy and the session was recoded by the observer

for all group members. The second reliability check was done for two different group members. If the 10% criteria was not obtained on the second coding, the verbalization rates coded by the observer and the investigator were averaged.

Refining the definition for the unit of verbalization and more specificity about the timing procedure when the expert material was being read improved agreement in the categories with higher frequencies. When these revisions were made, the procedure for reliability checks was also changed. It was determined that within a group there were students who did the majority of the talking and others who did little. The decision was made that for each session, the reliability check would include the coding of a frequent speaker and a less frequent speaker. If agreement were attained for two persons, then it was assumed that the other members' scores were accurate. A schedule was drawn up so that each member within a group was checked an equal number of times over the duration of the study. The following formula was used to calculate interobserver agreement:

$$\frac{\text{No. of agreements}}{\text{No. of agreements + No. disagreements}} \times 100$$

It is readily apparent that categories with higher frequencies will yield better interobserver agreement. A difference of only one or two verbalizations in a low

frequency category produces wide fluctuations. For example, if an observer recorded 3 incidents and the checker scored only 1, then agreement would be only 33%. On the other hand, a frequency of 3 and 4 would yield a 75% agreement rate. Therefore, interobserver agreement was not calculated for a category when the frequency for that learning session fell below 5 incidents. However, the reliability calculations do reflect the agreements between observers for both frequent and less frequent speakers, because one of each was chosen for each check.

The adjustments made in the coding system produced satisfactory agreement between the investigator and the observer MM assigned to Group 5. Unfortunately, he was unable to continue with the study. For the other two observers, this procedure did not result in satisfactory agreement for several of their assigned learning sessions. Part of the problem was due to poor tape quality and a high level of physical activity in the learning groups assigned to them, but it was also due to the need for more training. Since the observers could only code three to four hours per week and it was undesireable to extend the coding over many weeks, the investigator coded five of the thirteen sessions assigned observer SP and five of seven assigned observer JB. For the sessions coded by the investigator, the

observers did the reliability checks on 2 persons per session.

Results of Reliability Checks

The results of the rates of agreement between the observers and the investigator across all sessions are listed in Table 1. They are listed for the four verbalization categories (on-task, management-positive, management-negative, not-on-task/miscellaneous) and for "positive verbalizations" and "negative verbalizations". The positive verbalizations are the combined verbalization rates of the on-task and management-positive statements; the negative verbalizations are the combination of the management-negative and the not-on-task/miscellaneous statements. A dash (-) means that no reliability figure was calculated because the frequency was less than 5 incidents per session.

The most frequently occuring verbalization categories produced the best agreement rates: 89% for on-task and 84 % for management-positive. The least frequent category, management-negative, yielded the poorest agreement rate (59%). When the management-negative statements were combined with the not-on-task/miscellaneous statements to yield negative verbalizations the rate of agreement was 78%. The

positive verbalizations (on-task and management-positive) produced an agreement rate of 91%.

Table 1.

Interobserver Agreement (%) By Groups Across All Sessions

	OT	MP	MN	NT/MS	Positive Verb.	Negative Verb.
Group						
1	86	87	-	78	90	76
2	92	84	66	94	92	82
3	90	86	-	89	92	78
4	92	84	50	58	90	70
5	86	78	-	82	90	85
Mean:	89	84	59	80	91	78

Note. OT = on-task; MP = management-positive; MN = management-negative; NT/MS = not-on-task/miscellaneous; dash (-) = reliability not calculated.

The agreement rate for positive verbalizations between trained observers was lower than that between an observer and the investigator. The agreement rate between observer JB with observer SP for positive verbalizations was 80% and with MM was 78%. The agreement rate of 78% for negative verbalizations of JB with SP was similar to that between an observer and the

investigator. No negative verbalization agreement rate was computed for MM and JB because of the low rate in the students selected for the check. It will be recalled that the reliability checks between two trained observers were performed on two individuals not selected the first time. Thus the slightly lower reliability figures between trained observers may be partially due to differences in the individuals selected.

Data Analysis

The categories of verbalizations are expressed as rates (verbalizations per minute) because the learning sessions varied in length. The rates of verbalizations for each category were computed for each person for a particular learning session. The on-task (OT) category was combined with the management-positive (MP) category since both types of statements related to learning within the group. This combination was labelled positive verbalizations. The management-negative (MN), not-on-task (NT) and miscellaneous (MS) were combined and labelled as negative statements. They were statements that detracted from learning.

For purposes of analysis, the students were divided into three ability groups: High = Ranks 1-8, Middle = Ranks 9-15, Low = Ranks 16-23. The rankings were the teacher's rating of the students' achievement level in

social studies before the implementation of Jigsaw II.

If in fact, during the implementation of Jigsaw, students learned how to help each other learn, the rates of verbalizations should change with more homogeneity of variance in the rates of talking at the end of the study than in the beginning. Since the prediction was that the rates of verbalizations for the higher ability students would decrease and that the rate for the less capable would increase, an "early" and a "later" session were compared. The early session was either session 1, 2, 4 or 7 and the later session was either session 12 or 13. The wide spread in the choice for an early session was due to the videotaping schedule and the need to choose a session in which no students were absent. When it became evident that the rates of verbalizations for an individual were highly variable from session to session and that there were no trends toward an increase or a decrease in speaking for individuals, the rates of speaking were averaged for session 1 through 7 to yield verbalization rates for the "first half". Session 8 through 13 were combined to form the "second half". Session 14 was not used because it differed slightly from the traditional learning group. That lesson was about Indian legends and although they were instructed to help each other learn the legends, the students had not met first in expert groups.

RESULTS

The rates of verbalizations were organized into two sets of data. The verbalization data for an early session (Session 1, 2, 4 or 7) was compared with that for a later session (Session 12 or 13). The second comparision was made between the mean verbalization rate for the first half (Sessions 1-7) and the second half (Sessions 8-13). The statistical analysis for both data sets yielded similar results. The only exception was an increase in the rate of negative verbalizations when an early and a later session were compared. When it became evident that the rates of verbalization for an individual were highly variable from session to session, the decision was made to report only the analysis for the first and second half. However, for the reader that is interested, the rates of verbalizations for the early and later session are listed in Tables 6 and 7 in Appendix C. They are listed according to the four verbalization categories for the three ability groups (Table 6) and for the five learning groups (Table 7). It will be recalled that the positive verbalizations are the total of the on-task and management-positive statements and the negative are the total of the management-negative, not-on-task and miscellaneous statements.

Objective 1: Changes in the Homogeneity of Variance

It was predicted that initially the higher ability students would speak more than the lower ability, but that the rates of speaking would become more similar over the implementation of Jigsaw II. Examination of Figures 1 through 4, in which the rates of positive and negative verbalizations are plotted for the learning sessions 1 through 14, indicates that there was no trend toward an increase or a decrease in the rate of verbalizations for individuals. Figure 1 exhibits the rates of speaking for individuals in a five member group. The rates of speaking are highly variable between group members, and the rate of speaking for an individual fluctuates widely from session to session. For example, student #8 spoke at a rate of 2.1 in Session 1 while in Session 3 his rate was 8.0. The rate of "positive" verbalizations in a four member group (Figure 3) demonstrate the same degree of variability.

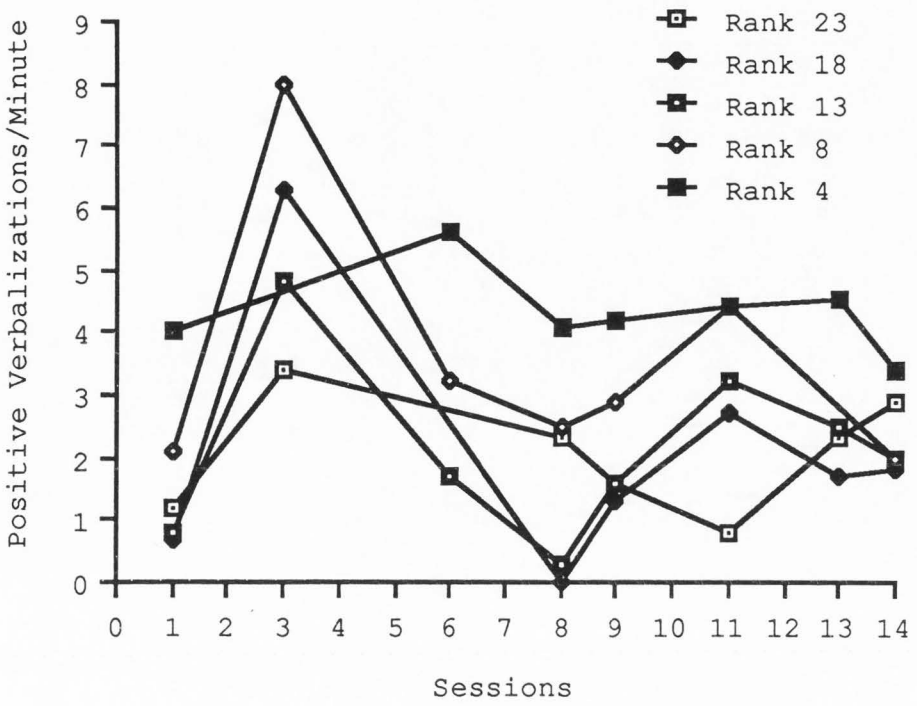

Figure 1. Rates of positive verbalizations per minute for group 4.

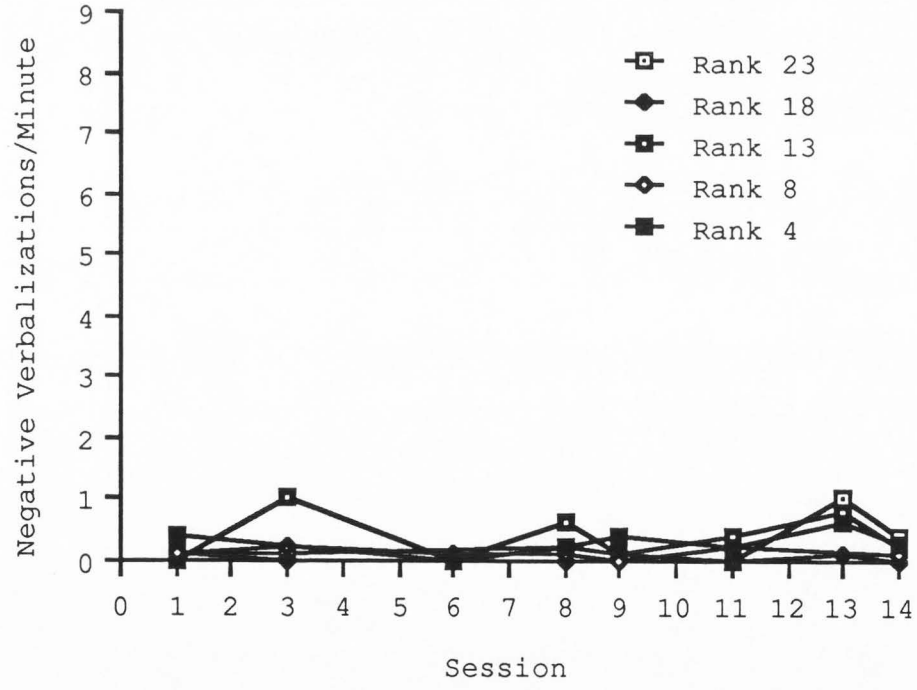

Figure 2. Rates of negative verbalizations per minute for group 4.

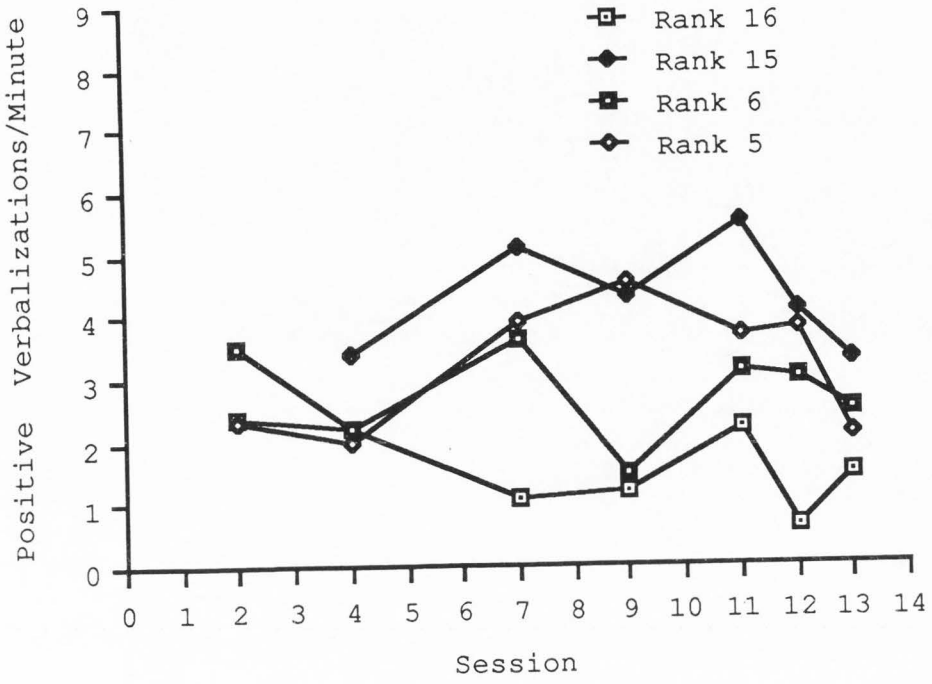

Figure 3. Rates of positive verbalizations per minute for group 1.

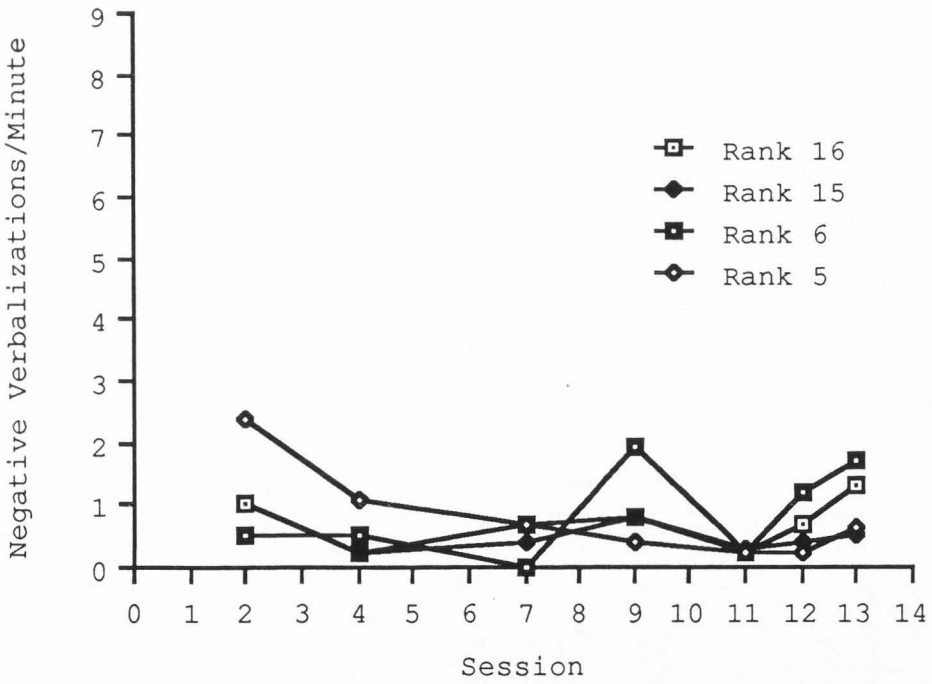

Figure 4. Rates of negative verbalizations per minute for group 1.

The rates of negative verbalizations (Figures 2 & 4) did not vary as much as the positive verbalizations for a five and four member group. However, there were incidents of wide variability between individuals in a particular session as well as fluctuations in the level of negative verbalizations between the sessions. For example, in Figure 2, the range in negative verbalization rates was from 0.0 to 1.0 for Session 3, while in Session 13 there was a higher rate of negative verbalizations for everybody ranging from 0.6 to 1.0.

The rates of positive and negative verbalizations for the first and second half according to the ability group of the students are listed in Table 2. A paired T-test for the homogeneity of variance (Ferguson, 1981) between the first half (Sessions 1-7) and the second half (Sessions 8-13) indicated that the variance in rate of talking did not differ significantly for either the positive verbalizations, $t(21) = .91$, $p > .05$ or for the negative verbalizations, $t(21) = .69$, $p > .05$.

Table 2.

Rates of Verbalizations for First and Second Half by Ability Group.

Category	Ability Group	First half Verb/Min	First half Std. deviation	First half % of total verb	Second half Verb/Min	Second half Std. deviation	Second half % of total verb
POSITIVE	High (8)	3.33	.86	88.0	3.75	.95	88.6
	Middle (7)	2.63	.83	80.4	3.29	1.35	87.0
	Low (8)	2.09	.62	78.8	2.00	.59	83.6
	Mean	2.68	.91	82.9	3.00	1.22	87.0
NEGATIVE	High (8)	0.45	.43	12.0	0.48	.51	11.4
	Middle (7)	0.64	.45	19.6	0.49	.33	13.0
	Low (8)	0.56	.35	21.2	0.39	.22	16.4
	Mean	0.55	.40	17.1	0.45	.36	13.0

One of the reasons for lack of significance in the homogeneity of variance between the first and second half of the study was due to the large standard deviations in the rates of speaking. For example, from Table 2, the rate of positive verbalizations in the first half for high ability students ranged from 2.47 to 4.19. For the middle level students the range was from 1.80 to 3.46 indicating that some middle ranked students spoke as much as some high ranked students. A similar overlap occurred between the middle and low ranked students whose range of positive verbalizations were from 1.37 to 2.71. Indeed, there were some low level students who spoke as much as some high ability students

when the low rate (2.47) for the high ability group is compared with the high rate (2.71) for the low ability group.

Objective 2: Changes in Types of Verbalizations

The mean rate of positive verbalizations for all students increased from 2.68 to 3.00 from the first to second half (Table 2). A paired T-test indicated that the increase was not statistically significant, $t(21) = -1.1.66$, $p > .05$. However, the rate of positive verbalizations was initially high with 83% of the total verbalizations being positive in the first half and 87% in the second half.

The decrease in the rate of negative verbalizations from 0.55 to 0.45 was not statistically significant, $t(21) = .69$, $p > .05$. It should be noted that there was statistically significant increase in the rate of negative verbalizations if an early and later session are compared, $t(21) = 2.98$, $p < .01$. However, the sessions that were used as the later one, sessions 12 or 13, for some reason, had much higher negative verbalization rates. The mean for sessions 12 and 13 was 0.67 compared 0.20 for sessions 8 through 11.

In summary, the rate of positive verbalizations increased and the negative decreased, but the change was not statistically significant.

Objective 3: Changes in Verbalizations Associated with Ability Groups

In the first half of the study the high ability students made positive statements at a rate one third greater than the low ability students (3.33 compared to 2.09, Table 3). The rate of negative statements was approximately the same for all groups (0.45 to 0.64). The predicted decrease in the rate of positive verbalizations for the high ability students and an increase for the low ability students did not occur. What happened was that the rate of verbalization for high ability students increased 12%, the rate for middle ability students increased 32% (2.63 to 3.29) and the rate for the low level students remained the same.

During the second half of the study, the high ability students made positive statements at almost twice the rate of the low ability students. The middle ranked students spoke at a rate almost equal to the high ability.

The rate of negative statements remained approximately the same for the high ability students (.45 to .48) and decreased for the middle (.64 to .49)

and lowest ranked students (.56 to .39).

An analysis of covariance for the mean positive verbalizations in the second half (Table 3) with student ability (high, middle and low) as the independent variable and positive verbalizations as the dependent variable indicated no statistical difference among the positive verbalization rates for the three ability groups, $F(2,19) = 2.68$, $p > .05$. An analysis of covariance for the mean negative verbalizations in the second half (Table 4) was also not statistically significant $F(2,19) = 0.26$, $p > .05$.

Table 3.

Analysis of Covariance for Mean Positive Verbalization Rate in Second Half (Sessions 8 - 13).

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F	Significance of F
Covariate:					
First half positive verbalizations	10.70	1	10.70	11.82	.003
Main Effect:					
Ability groups	4.84	2	2.42	2.68	.095
Error	17.16	19	0.90		

Table 4.

Analysis of Covariance for Mean Negative Verbalizations in Second Half (Sessions 8-13).

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F	Significance of F
Covariate:					
First half negative verbalizations	0.42	1	0.42	3.35	.083
Main Effect:					
Ability groups	0.07	2	0.03	0.26	.770
Error	2.39	19	0.13		

Although the rate of verbalization for the low ability students did not change, the percentage of positive verbalizations increased from 78.8% to 83.6% (Table 2). The percentage of positive verbalizations also increased for the middle ability students (80.4% to 87%).

Analysis of Quiz Scores

The mean scores (maximum possible = 30) of the high, middle and low ability students for five quizzes are plotted in Figure 5. On the first two quizzes the high ability students obtained the highest scores of 24.1 and 22.3. The middle level students obtained mean scores of 21.4 and 17.4 followed by the low ranked students whose scores were 19.4 and 16.7.

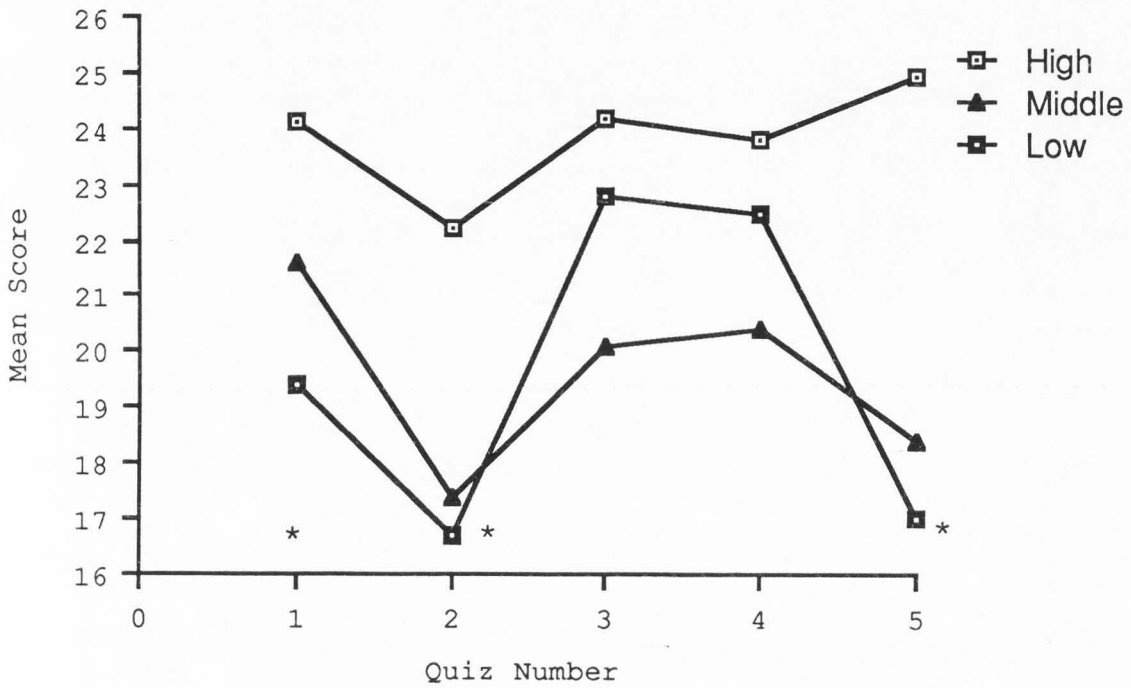

Figure 5. Mean quiz scores for high-,middle- and low-ability groups. [* Only difference in scores between high- and low-ability groups is statistically significant ($p < .05$)].

On quizzes 3 and 4 the high ranked students continued to do the best but the low ranked students scored higher than the middle ranked (22.9 vs 20.1 and 22.5 vs 20.4). On the fifth quiz, the high ability students did the best with a score 24.9 followed by the middle and low ability with scores of 18.4 and 17.0 respectively.

An analysis of variance for repeated measures (Table 5) indicated a statistically significant difference among high, middle and low ability students on quiz scores $F(2,20) = 4.08$, $p < .05$. The mean scores on all five quizzes for the high, middle and low ranked students were 23.9, 19.6 and 19.7 respectively. A posteriori comparisions of quiz scores for the three ability groups indicated that only on quizzes 1, 2 and 5 was the difference between the scores of the high and low ranked students statistically significant ($p<.05$). On quizzes 3 and 4 where the low ranked students scored higher than the middle ranked, the difference was not statistically significant.

In summary, the high ability students maintained the highest quiz scores throughout, but the low ability students improved to the level of the middle ability students.

Table 5.

Analysis of Variance for Quiz Scores.

Sources of variation	Sum of squares	Degrees of freedom	Mean squares	F	Significance of F
Ability group	415.6	2	207.8	4.08	0.033
Subjects	1019.9	20	50.99	2.68	0.001
Time	204.9	4	51.23	2.69	0.037
Ability group x time	150.9	8	18.87	0.99	0.449
Error	140.3	74	19.02		

Behavioral Academic Self-Esteem Measure

The mean of the Behavioral Academic Self-Esteem (BASE) rating before the implementation of Jigsaw II was 100.5 and upon completion was 103.1. These scores were standard scores with a mean of 100 and a standard deviation of 15. The difference between the pre and post scores was not statisitically significant when a paired t-test was computed ($t(21) = .79$, $p > .05$).

DISCUSSION

The main hypothesis was that the rate of speaking for the higher ability students would decrease and that the rate for the less capable would increase. This change should occur over the implementation of Jigsaw as the less capable students learned how to teach their expert material better and were encouraged by their groupmates. This did not happen. The higher functioning students spoke more than the lower ability students at the beginning of Jigsaw II and were speaking at a slightly higher rate at the end. The rate of speaking for the lower ability students remained the same throughout the study, although their percentage of positive verbalizations increased. The type of verbalizations changed little over the implementation of Jigsaw II. More than 80% of the verbalizations were directed toward the task at the beginning and the rate had increased only slightly by the end. One explanation for such little change appeared to lie in the rather rigid functioning of the learning group.

Each group member took his turn teaching his expert material from a handout provided by the teacher. They seldom varied from the following procedure in their teaching of the expert material. They read directly from the expert sheet, occasionally mispronouncing a

word or missing a line altogether. Illogical statements or gaps in the story elicited no questioning. Seldom did a group member ask for clarification or make a correction.

Students in this study made very few statements that added to the information provided to them. In the Johnson et al., (1985) study, 10% of the statements about the material being learned were high-level oral rehearsal, which were explanations, elaboration, providing rationale, and relating information to previously learned information. Their study also examined the verbal interactions in a grade 4 social studies class, but the data were collected in a classroom which had an ongoing cooperative goal structure and highly trained teachers.

In this study, it was as though each student was waiting his/her turn to speak. At other times they appeared to be competing within the group. For instance, there were no incidents of social reinforcement on any of the videotapes for a good presentation or for improvements in teaching. In the last learning session, two team mates had no better solution to the correct pronunciation of a word than "you say it your way and I will say it mine." When a student finished reading, the study guide questions were asked and group members answered spontaneously. There was no effort to make

sure that everybody knew the answer and there were no discussions about different answers. If everybody had taken their turn teaching and the session was not yet finished, they started over, reading from the expert sheet in a similar fashion. Other examples with concern for having an equal turn were in the two lessons requiring the production of a group product. For instance, in one lesson, the students were required to draw a time line and mark on it a summary statement about different Indian periods. Seventy percent of the verbalizations were devoted to whose turn it was to write and what colored pen should be used. Although the instructions for the lesson were to summarize important points about each time period, very little information about Indians was exchanged. A similar phenomena occurred in the lesson for writing a story with Indian symbols, based on their knowledge of Indians acquired so far. The students were very concerned with the mechanics of the project, making sure they had their turn writing and that the writing was neat.

Students of all abilities had great difficulty presenting the material in their own words or extracting the main ideas, but the more capable ones taught more of their material because they were better readers. The less capable students did not read all of their expert

material outloud and left out some of the main points. Nobody questioned the paucity of material presented. In other words, the better students did not help the less able teach their topic. During the learning sessions, there was no reference to the quiz unless students were alerted by the teacher that it would be administered in five minutes. They did not appear to see the relevancy between their team score and its effect on their individual grades in social studies.

This concern with taking turns also helps to explain the large fluctuations in in the rate of speaking for individuals across sessions. Although large variability in speaking rates among indivduals was expected at the beginning, it was not expected that the rate of speaking for an individual would vary so dramatically from session to session. The fluctuation in the rate of speaking for individuals across time was primarily due to whether a student had an opportunity to teach his/her expert material. In sessions which were at least 20 minutes in duration, each group member had an opportunity to teach. On days when the learning group met only briefly for a review of the material before taking the quiz, not all members had an opportunity to present their material. This was primarily due to difficulties with summarizing and extracting the main ideas as previously mentioned. If a

student presented his material, he/she read all of it from the expert sheet. As a result, in a review session, there was not enough time for everybody to present. The second factor that contributed to the variability in individual speaking rates was the lack of accountability from session to session. The students tended to treat each learning session as a fresh start, failing to remember who had taught the day before. Sometimes the more assertive students would reteach their expert material before permitting a groupmate to have a turn. The result was that not everybody each day had an equal opportunity to talk.

Teammates did not come to view each other as prime sources of information, one of the positive outcomes of cooperative learning (Johnson et al., 1985). In this study, one group decided to exchange expert sheets as the most effective way of mastering the material. Fortunately, this was discovered early in the study and curtailed by the teacher.

To summarize what happened in the learning group, a high proportion of the verbalizations was directed toward exchanging information about the task and statements relating to the management of the group. Students did not help or encourage each other to learn, nor did they check to see whether groupmates had adequately learned the information to do well on the

test. The large variability in rates of talking between individuals within the same ability groups and for individuals between sessions confounded observing any trends in the rates of speaking during the implementation of Jigsaw.

Although the learning groups did not function in the manner hypothesized in the literature, the students performed satisfactorily on the quizzes. If the mean scores for all five quizes are considered, the low ability students did as well as the middle ability students and on two of the five quizzes they scored higher than the middle ranked students. The high ability students maintained their lead throughout. The higher quiz scores of the low ability students is similar to Lucker et al.'s (1976) finding in which the low ability minority students improved significantly, while the high ability students maintained their good scores. The improved performance on the quizzes by the lower ability students was not linked to increases in verbalization. The Johnson et al. (1985) study reported a significant negative correlation between low ability students listening to information and achievement. In this study, the verbalization rates of the low ability students remained the same throughout, but their quiz scores were as high as the middle level students whose verbalizations approached that of the high abilty

students by the end of the study. It would appear in this study that the learning environment provided by the group was sufficient to improve test scores for the low ability students.

Approximately 85% of the verbalizations were devoted to learning the material, similar to what the Johnson et al. study (1985) reported. It would seem logical that the more time devoted to engagement with the academic material, the greater likelihood of better test performance. Small group learning in which the students are responsible for the teaching probably results in more active processing of the material than if the teacher were lecturing to the whole class. The large percentage of verbalizations devoted to the subject matter indicated that there was active engagement with the material. Student team teaching requires the development of a very specific curriculum and the stipulation of a time frame for accomplishing it. Slavin (1979) determined that such a focused schedule of instruction was related to improved test performance. Perhaps then, it was the focused schedule of instruction and the high proportion of academic engaged time rather than peer tutoring that accounted for the good quiz scores of the lower ability students.

Although Jigsaw II was implemented according to the teacher's manual (Slavin, 1980a), it appeared that the

students did not understand the concept of a group contingency and a group reward. The fact that the team's score affected individual grades and that only the whole team's performance would merit recogniton did not seem relevant to the students. Slavin (1979) states that the two most important components of classroom organization are the task structure and the reward structure. Slavin (1979) observed more tutoring in classes with a team reward structure than with an individual reward structure. The lack of tutoring among students in this study would indicate that they did not understand the team reward. Failure to comprehend the connection between the learning session and the group reward may have been partly due to a time factor. Poor performance on a quiz by oneself or a team member could not be related to what happened in the learning group two days previously. The concept of a group reward may have been foreign to the students. Small group learning where the teacher does not lead the group was a new experience for these students. They had had no previous experience with this system and there were no other activities in their curriculum that would require this degree of cooperation. The teacher did attempt to foster cooperation and stressed the importance of everybody doing well on the test because the group score affected individual grades. She did this by instructing

the class as a whole and by interacting with the individual groups during the learning sessions. When the group scores were disseminated along with the newsletter, she checked with each group individually to make sure they knew their group score and where they ranked in the class. She frequently praised a group for working well together and encouraged the students not only to ask each other questions about the material, but modelled how they could do it. They were asked to think of ways to bring members back into the group who were not paying attention and how this problem could be avoided in the first place. All this was to little avail, apparently.

In this study, behavioral academic self-esteem ratings did not exhibit significant gains. This may be partially explained in the general demeanour of the class. The teacher communicated to the investigator that the class used for the study was more compliant, homogeneous and less enthusiastic about classroom activities either in the academic or social realm compared to her other fourth grade. The use of Jigsaw II for social studies did not arouse nearly the level of excitement that it did in her other class. This lack of enthusiasm would contribute to the teacher observing no change in academic self-esteem. In completing the rating scale, the teacher may have also been influenced

by the students' behavior in classes not related to Jigsaw II.

This study points out the importance of fostering group cooperative skills. Their importance is not new (Johnson & Johnson 1975; Aronson, 1978). The Johnsons (1975) contend that establishing a cooperative goal structure fosters cooperation and to this end they provide extensive training for their teachers. Aronson (1978) in the Jigsaw classroom provides time at the end of the lesson for processing how well the group worked together. Graves and Graves (1985) in their work with cooperative education in New Zealand and California found that the introduction of cooperative learning methods for one or two subjects per day was meaningless. The students would cooperate temporarily to please the teacher for an external reward, but there was no transfer to other school activities. It is the Graves' opinion that cooperative learning requires a holistic approach. Tasks need to be restructured to promote cooperation and students need feedback on their patterns of communication. For example, the drawing of a mural is more conducive to cooperation than individual art projects. Interactions between students should be analyzed to facilitate insight and understanding rather than blame, so that the likelihood of the new skill being used in future interactions is enhanced. Kagan

(1986) has broken down the skills for cooperative learning into task skills and maintenance skills. At the elementary level, task skills include giving ideas, talking about work, checking other's understanding of work, asking questions, following directions and staying in their seats. The maintenance skills include encouraging, using names, inviting others to talk, responding to ideas, looking at others, saying "thank you", sharing feelings, disagreeing in a nice way and keeping things calm. Through processes such as task restructuring, games, communication training, role playing and feedback from students and teachers, cooperative skills are developed. The activities used to promote cooperation are an integral part of learning. For example, when task restructuring is used cooperation is essential for completion of the task. Games can be tacked on to regular learning tasks. Kagan (1986) likes the "talking chip game" where speaking is limited to one minute and a chip must be placed in the center of the table when a person is finished speaking. Everyone must take a turn talking before all the chips are retrieved and the process starts over. Another approach is the assignment of roles to individals to keep the group on task or to summarize ideas, the results of which can be processed at the end of the learning session.

The results of the analysis of the verbal interactions in this study support the importance of facilitating cooperative skills. The implementation of a group reward was not enough to foster the development of cooperative skills. Jigsaw II did transfer the responsibility for learning from the teacher to the students and yielded satisfactory achievement levels in social studies. However, it did not develop helping skills and concern for each other's learning.

CONCLUSIONS

Although this study took place in only one classroom with one teacher, it is very likely that what occurred is typical of what happens when Jigsaw II is implemented in a classroom with no previous cooperative learning experience. Although the teacher implemented the method according to procedures that have been successful in other classrooms and the students were compliant, concern for each other's learning did not develop. Skills in how to cooperate need to be incorporated into the learning process. The fact that this is a long term process need not be of too much concern because it appeared that the students mastered the material, as measured by the quiz scores, inspite of the deficits in cooperating skills.

Along with cooperating skills, students at this grade level appear to need specific instruction in summarizing information to promote a freer flow of information and to advance beyond just reading facts to each other. Although the acquisition of this skill would take time, its ready acquistion by some students would provide models for others.

It is imperative that the students understand the reward structure right from the beginning. This could be accomplished by requiring that each group member

teach a very small unit (eg. meaning of pueblo, mesa), administering the quiz and giving the scores in the same day. The students would have immediate feedback on the consequences of failing to teach or learn the material.

The effort for cooperative team teaching is certainly more work initially for the teacher than regular teacher-directed classroom instruction. However, once materials are prepared, the responsibility for learning is transfered to the students. The teacher is freed from dispensing factual information to interacting with the group on a level that promotes broader cognitive processing and the development of interpersonal skills. Responsibility for one's own learning and the acquisition of good interpersonal skills are worthy goals for future academic endeavors and the workplace, where few people work alone.

REFERENCES

Aronson, E. (1978). The jigsaw classroom. Beverly Hills, CA: Sage Publications.

Aronson, E. & Osherow, N. (1980). Cooperation, prosocial behavior, and academic performance: Experiments in the desegregated classroom. In L. Bickman (Ed.), Applied Social Psychology Annual (Vol. 1, pp. 163-196). Beverly Hills, CA: Sage Publications.

Blaney, N. T., Stephan, C., Rosenfield, D., Aronson, E., & Sikes, J. (1977). Interdependence in the classroom: A field study. Journal of Educational Psychology, 69, 121-128.

Coopersmith, S. (1982). BASE, Behavioral academic self-esteem, A rating scale Palo Alto, CA: Consulting Psychologists Press Inc.

Ferguson, G. A. (1981). Statistical analysis in psychology and education. New York, N.Y: McGraw Hill Book Company.

Graves, N. B. & Graves, T. D. (1985). Creating a cooperative learning environment. An ecological approach. In R. Slavin, S. Sharan, S. Kagan, R. Hertz Lazarowitz, C. Webb & R. Schmuck (Eds.), Learning to cooperate, cooperating to learn. (pp. 403-436). New York: Plenum Press.

Hulten, B. H. & Devries, D. L. (1976). Team competition and group practice: Effects on student achievement and attitudes. (Rep. No. 212). Baltimore, MD: The Johns Hopkins University, Center for Social Organization of Schools.

Johnson, D. W. & Johnson, R. T. (1975). Learning together and alone. Englewood Cliffs, NJ: Prentice-Hall Inc.

Johnson, D. W. & Johnson, R. T. (1983). The socialization and achievement crises: Are cooperative learning experiences the solution? In L. Bickman (Ed.), Applied social psychology annual (Vol. 4, pp. 119-164). Beverly Hills, CA: Sage Publications.

Johnson, D. W. & Johnson, R. T. (1985). The verbal interaction measure: A preliminary report. Unpublished manuscript. University of Minnesota, Cooperative Learning Center, Minneapolis.

Johnson, D. W. & Johnson, R. T. & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. Review of Educational Research, 53, 5-54.

Johnson, D. W., Johnson, R. T. Roy, P, & Zaidman, B. (1985). Oral interaction in cooperative learning groups: Speaking, listening, and the nature of statements made by high-, medium-, and low-achieving students. The Journal of Psychology, 119, 303-321.

Kagan, S. (1986). Cooperative learning resources for teachers. Riverside, CA: University of California, Printing and Repographics.

Lucker, G. W. Rosenfield, D., Sikes, J., & Aronson, E. (1976). Performance in the interdependent classroom: A field study. American Educational Research Journal 13, 115-123.

Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes, and ethnic relations. Review of Educational Research, 50, 241-271.

Sharan, S., Ackerman, Z. & Hertz-Lazarowitz, R. (1979). Academic achievement of elementary school children in small-group versus whole-class instruction. Journal of Experimental Education, 48, 125-129.

Sharan, S., Kussell, P, Hertz-Lazarowitz, R, Bejarano, Y., Raviv, S., Sharan, Y. (1985). Cooperative learning effects on ethnic relations and achievement in Israeli junior-high school classrooms. In R. Slavin, S. Sharan, S. Kagan, R. Hertz Lazarowitz, C. Webb & R. Schmuck (Eds.), Learning to cooperate, cooperating to learn (pp. 313-344). New York: Plenum Press.

Slavin, R. E. (1979). Effects of student teams and peer tutoring on academic achievement and time on-task. Journal of Experimental Education, 48, 252-257.

Slavin, R. E. (1980a). Using student team learning Baltimore, MD: The Johns Hopkins University, Center for Social Organization of Schools.

Slavin, R. E. (1980b). Cooperative learning. Review of Educational Research, 50, 315-342.

Slavin, R. E. & Karweit, N. L. (1981). Cognitive and affective outcomes of an intensive student team learning experience. Journal of Experimental Education, 50, 29-35.

Yager, S. Johnson, D. W. & Johnson, R. T. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. Journal of Educational Psychology, 77, 60-66.

APPENDICES

Appendix A

Use of the Verbal Interaction Form:

1. Fill in the information at the top of the data sheet.

2. Select Start point by picking a very clear verbalization by a group member. Enter the number from the VCR and start all analysis from this point. Write the verbalization in the space provided.

3. The Stop point is when the teacher turns off the lights.

4. Watch the session once without stopping. Record the elapsed time for that session using the stopwatch.

5. Select the student(s) to observe. Record each incidence of speaking in the appropriate row. A new incidence of speaking is recorded if there is a pause. eg. How did the Navho Indians get horses? (1 slash). Answer: They steal them (pause) and they trade them (2nd slash). For a verbalization to be scored it must contain a fact although it may not be a gramatically correct sentence. eg. if the student says "and they" or "it seems that", pauses and then says "stole them", it is counted as one verbalization. When a student is teaching his expert material to the group and is essentially reading it uninterrupted, run the stopwatch only while he is speaking. Do not include the asking of the study questions in the timing. Record the time in minutes and seconds and circle the figure. Each five second interval is coded as one incidence of verbalization.

Each verbalization is scored according to whether it is:

1. <u>On Task (OT)</u> : A statement that is an exchange of task related information. It is about the task - it usually is correct information but incorrect information is recorded as OT as long as it relates to the the task. These statements are further categorized into:
   a. Information (I) - a statement of fact. An information statement can also be in a question form. eg. if one student asks questions of the other students to determine if the material has been learned.

b. Elaboration (E) - any statement that elaborates on the material or a question about the material that cannot be directly answered from the text. eg. why are there so many kinds of birds?

2. Not on Task (NT): a statement that is not related to the material being studied. eg. "I like your shirt", "let's play after school." If 2 students are talking to each other while a third student is talking to the group and their verbalization is not picked up by the microphone, it is automatically categorized as NT.

3. Management (M): a statement that relates to group processes. They can be either positive or negative.

Examples of management-positive statements: "it is Greg's turn", let's hurry up and finish".
When words such as "okay", "whatever", "yes", "no", "what" are used alone they are counted as M-P. When used with other statements, that statement defines the category. eg. "okay, they could hunt better on horses" is OT.
Examples of management-negative statements: "you took too long", "you didn't tell us that". The criteria for M-N statements is complaining. When one student tells another student that they are not supposed to read their material or any other reference to reading, it is M-N because the teacher has given specific instructions not to read the material from their paper.

4. Miscellaneous (MS): a statement that does not fit into the above categories. If a verbalization is too faint to hear or the statement cannot be understood it is coded as MS.

Appendix B

Data Form

Group 4

Date of Lesson:_____ Date Coded:_____

Coder:_____ Elapsed Time:_____

Start:_____
Stop:_____

List Absent Students_____

	S 1	S 2	S 3	S 4	S 5	TOTAL
OT-I						
OT-E						
M-P						
M-N						
NT						
MS						
TOTAL						

COMMENTS:

Appendix C

Table 6 .

Rates of Verbalizations for Early and Later Sessions by Ability Groups.

Verbalization category	Rank	Early session			Later session		
		Verb/Min	Std. deviation	% of total verb	Verb/Min	Std. deviation	% of total verb
OT (On task)	High (8)	1.63	.83	49.5	1.87	.61	48.1
	Middle (7)	1.03	.50	38.0	2.06	.69	57.4
	Low (8)	1.08	.50	57.5	1.50	.44	56.8
N = 23	Mean	1.26	.67	47.7	1.80	.61	53.9
MP (Management-positive)	High (8)	1.22	.50	37.0	1.16	.58	29.8
	Middle (7)	1.16	1.01	42.4	0.83	.37	23.2
	Low (8)	0.54	.44	28.5	0.48	.34	18.1
N = 23	Mean	0.97	.72	36.7	0.80	.51	23.9
POSITIVE	High (8)	2.85		86.6	3.03		77.8
	Middle (7)	2.19		80.4	2.89		80.6
	Low (8)	1.62		86.0	1.98		75.0
	Mean	2.22	.95	84.3	2.63	.85	77.8
MN (Management-negative)	High (8)	0.10	.08	3.0	0.13	.08	3.3
	Middle (7)	0.19	.32	7.0	0.06	.05	1.4
	Low (8)	0.05	.08	2.7	0.03	.05	0.7
N = 23	Mean	0.11	.19	4.2	0.07	.07	2.1
NT/MS (Not on tasks miscellaneous)	High (8)	0.34	.28	10.0	0.73	.64	18.7
	Middle (7)	0.34	.39	12.5	0.64	.55	17.9
	Low (8)	0.21	.22	11.2	0.64	.51	24.2
N = 23	Mean	0.30		11.3	0.67	.54	20.2
NEGATIVE	High (8)	0.44		13.4	0.86		22.2
	Middle (7)	0.53		19.6	0.70		19.4
	Low (8)	0.26		14.0	0.69		25.0
N = 23	Mean	0.41	.45	15.6	0.74	.54	22.2

Table 7.

Rates of Verbalizations for Early and Later Sessions by Student Groups.

Verbalization category	Group	Early session Verb/Min	Early session Std. deviation	Early session % of total verb	Later session Verb/Min	Later session Std. deviation	Later session % of total verb
OT (On-task)	1	0.45	.17	15.1	1.60	.53	47.3
	2	1.70	.36	48.9	1.60	.36	39.4
	3	1.30	.47	45.7	1.92	.58	58.8
	4	1.20	.80	64.1	1.68	.51	49.5
	5	1.56	.72	71.5	2.08	.94	74.8
	Mean	1.26	.67	48.0	1.80	.61	
MP (Management-positive)	1	2.0	.74	67.1	0.75	.51	22.1
	2	1.02	.56	29.4	1.10	.36	27.1
	3	0.94	.42	33.0	0.72	.44	22.1
	4	0.56	.64	29.9	1.08	.82	31.8
	5	0.52	.32	23.8	0.48	.26	17.2
	Mean	0.96	.72	36.6	0.80	.51	
MN (Management-negative)	1	0.13	.05	4.3	0.08	.10	2.3
	2	0.27	.43	7.7	0.13	.05	3.2
	3	0.14	.06	4.9	0.04	.05	1.2
	4	0.04	.05	2.1	0.08	.09	2.3
	5	0.00	.00	0	0.04	.05	1.4
	Mean	0.10	.18		0.07	.07	
NT/MS (Not-on-task) miscellaneous	1	0.40	.36	13.4	0.95	.66	28.1
	2	0.48	.31	13.8	1.23	.52	30.2
	3	0.46	.33	16.2	0.58	.42	17.8
	4	0.10	.12	5.3	0.55	.39	16.2
	5	0.10	.07	4.5	0.18	.11	6.4
	Mean	0.30	.29	3.8	0.67	.54	
POSITIVE	1	2.45		82.2	2.35		69.5
	2	2.72		76.1	2.70		88.2
	3	2.24		78.8	2.64		80.9
	4	1.72		91.9	2.76		81.4
	5	2.48		95.7	2.56		92.1
	Mean	2.22	(2.3)	84.7	2.60		77.8
NEGATIVE	1	0.53		17.7	1.03		30.5
	2	0.75		21.6	1.36		11.8
	3	0.60		21.1	0.62		19.1
	4	0.15		8.0	0.63		18.6
	5	0.11		5.0	0.22		7.9
	Mean	0.40		15.3	0.74		22.2