DOCUMENT RESUME

ED 383 758 TM 023 288

AUTHOR Warkentin, Robert W.; And Others
TITLE An Exploration of the Effects of Cooperative Assessment on Student Knowledge Structure.
PUB DATE Apr 95
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Analysis of Covariance; *Cognitive Structures; Cooperative Learning; *Educational Assessment; Educational Psychology; Higher Education; Multiple Choice Tests; Pretests Posttests; *Student Attitudes; *Test Construction; Testing; *Undergraduate Students
IDENTIFIERS *Cooperative Assessment

ABSTRACT

Cooperative assessment was investigated in a classroom setting, examining achievement outcomes as measured by a multiple choice posttest of course content, a posttest of knowledge structure representation, and student perceptions of the cooperative assessment procedure. Eighty-three undergraduate psychology students participated in this nonequivalent control group study design. It was hypothesized that students taking tests using a cooperative assessment procedure would perform significantly better on a posttest of educational psychology course concepts than would students completing tests in a traditional format. Analysis of covariance indicated that there were no significant differences between the groups on the posttest and that the hypothesis was not supported. There were also no differences between groups on similarity or coherence measures of student knowledge structure. Student reactions to the cooperative assessment procedure were overwhelmingly positive. Students enjoyed taking tests in groups and felt that they learned more through this process as they discussed and debated the responses to the test items. One figure illustrates the discussion, and two appendixes provide supplemental information. (Contains 28 references.) (Author/SLD)

An Exploration of the Effects of Cooperative Assessment on Student Knowledge Structure

Robert W. Warkentin, Georgia Southern University
Marlynn M. Griffin, Georgia Southern University
Gwendolyn P. Quinn, Florida State University
Bryan W. Griffin, Georgia Southern University

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 18 - 22, 1995.

An Exploration of the Effects of Cooperative Assessment on Student Knowledge Structure

Robert W. Warkentin, Georgia Southern University
Marlynn M. Griffin, Georgia Southern University
Gwendolyn P. Quinn, Florida State University
Bryan W. Griffin, Georgia Southern University

Abstract

Amongst the plethora of cooperative learning studies, several investigations of collaborative, cooperative, or group assessment have appeared. These studies have investigated cooperative assessment in laboratory conditions (Lambiotte, et al., 1987), and in classroom settings (Bilsky-Torna, 1993; Webb, 1993; R. R. McCown, personal communication, April 13, 1993), and have examined the effects of cooperative assessment on learning (Lambiotte, et al., 1987; Bilsky-Torna, 1993; Webb, 1993; R. R. McCown, personal communication, April 13, 1993) and group process on assessment outcomes (Webb, 1993). In this study, we refined McCown's methodology and investigated cooperative assessment in a classroom setting, examining achievement outcomes as measured by a multiple choice posttest of course content and a posttest of knowledge structure representation, and student perceptions of the cooperative assessment process.

Eighty-three undergraduate educational psychology students participated in this non-equivalent control group design study. It was hypothesized that students taking tests using a cooperative assessment procedure would perform significantly better on a posttest of educational psychology course concepts than would students completing tests in a traditional format. In addition, the effect of the treatment on student knowledge structure representations was examined. The cooperative assessment group completed exams individually and then in groups; student exam grades were a combination of individual and group exam scores. The traditional assessment group took exams individually.

Analysis of covariance indicated that there were no significant differences between the groups on a posttest of educational psychology concepts, thus the hypothesis was not supported. There were also no differences between groups on either similarity or coherence measures of student knowledge structure. Student reactions to the cooperative assessment process, however, were overwhelmingly positive. These data indicated that although there were no statistically significant differences in achievement between the treatment groups, students enjoyed taking tests in groups and felt that they learned more through this process as they discussed and debated the responses to test items.

Cooperative learning has been thoroughly documented as an effective learning tool and teaching strategy (Johnson & Johnson, 1989; Johnson, Johnson, & Smith, 1990; Slavin, 1991). Cooperative learning, students learning from and with each other, helps to develop social skills, establish professional working skills, builds a sense of community within a classroom, and enhances student achievement, self-esteem, and attitudes toward school at all grade levels, including in college classrooms (Johnson, Johnson, & Smith, 1990; Slavin, 1991; Wynne, 1983). One explanation for these effects is that cooperative learning techniques augment the extent to which content is actively processed by students and offers participants the opportunity for discussion and negotiation which may lead to higher-level reasoning and the development of thinking strategies (Gabbert, Johnson, & Johnson, 1987; Johnson et al., 1990; Nystrand, 1986). The social support which often arises from a cooperative task has been shown to aid students in persisting on a challenging task, reducing frustration, increasing autonomy, and contributing to academic and career aspirations (Gabbert, et al., 1987; Sarason, Sarason, & Linder, 1983).

Cooperative techniques require students "to explain what they are learning to each other, learn each other's point of view, give and receive help from classmates, and help each other dig below the superficial level of understanding the material they are learning" (Johnson, et al., 1990, p. 11). While such "digging" and sharing of viewpoints can and does occur during cooperative in-class practice activities which take place as part of instruction, it is also likely that cooperative learning groups will negotiate under test conditions. Many believe that this negotiation of understanding is essential for knowledge construction (Duffy & Bednar, 1991; Kember & Murphy, 1990; Vygotsky, 1978). Brown (1989) argued that learning is "about the making of meaning, not just the receiving of it. Thoughtfulness is a constructive, not a passive, undertaking" (p. 32).

Proponents of authentic assessment have called for assessment to be more than an end product. Rather, assessment should be part of the learning process (c. f., Shavelson, Baxter, & Pine, 1992; Shepard, 1989; Wiggins, 1989). Many educators have had the experience of returning a test to the class and attempting to help students learn from their mistakes, only to find that many of the students are far less interested in learning from their mistakes than they are in trying to rationalize their incorrect responses in the hopes of earning extra points. Cooperative assessment places the ownership for learning in the hands of the students and offers opportunities for the negotiation of understanding. This negotiation can be a valuable learning tool and can encourage students to think about what they have learned and are learning as they discuss the assessment with their peers.

Several studies have begun to look at the benefits of an extension of the cooperative learning process - cooperative assessment. Lambiotte et al. (1987) found, under laboratory conditions with college students, that cooperative test-taking led to increased quantity of recall in reading and comprehension. Singer (1991) looked at the efficacy of cooperative testing with junior high students in a pre-algebra class. His study revealed an increase in test scores using cooperative test-taking, as well as an expressed preference by the students for this form of test taking. R. R. McCown (personal communication, April 13, 1993) reported similar findings in an unpublished pilot study of cooperative assessment.

Other studies have investigated the effects of group or collaborative assessment on achievement and social processes. Bilsky-Torna (1993) found that assigning grades based on group responses to an English quiz increased student motivation and the quantity of communication between students and teacher during the test. This study also reported disadvantages to cooperative testing such as group instability, noisy classrooms, "stifling" of academically stronger students, and weaker students riding on the coattails of stronger students during the group quizzes.

Webb (1993) analyzed the relationship between achievement scores obtained during small-group assessment tasks and individual assessment tasks. In Webb's study, students solved mathematics operations on decimal numbers in collaborative small groups for a 50-minute class period. Two weeks later, following a review session, students examined a similar problem without collaborating with other students. All students performed better in the group assessment situation regardless of prior ability. Student performance on the individual assessment was accurately predicted by both ability and behavior within the group assessment setting.

Griffin, Quinn, McCown, and Driscoll (1994) found that students do, in fact, discuss and negotiate responses to test items under cooperative assessment conditions. Despite the degree of interaction among group members on the cooperative assessment task, statistical differences were not found on the postmeasure of achievement between the treatment and control groups in their study. Possible explanations for the lack of significant findings in this study include (a) the achievement posttest may not have been sensitive to the kinds of understandings and conceptual knowledge fostered by the cooperative assessment technique and (b) the discourse of group members, while animated and on task, was not of a quality that improved understanding of the course concepts. The present study is intended to address the first possibility; that is, examination of student knowledge structures is expected to uncover differences in student cognitive structure, such as better organization and integration of course concepts, which cannot be detected with the current unit examinations.

Chi, Glaser, and Rees (1982) found the ability to perceive the underlying relatedness of concepts to be a measure of competence in that domain. Further research has shown that a high degree of correspondence between student and teacher network representations is correlated with achievement and degree of learning (Goldsmith & Johnson, 1990). Student knowledge structures seem to change as a result of instruction and students' subsequent increased understanding of the content renders them more "expert-like." The internal consistency of students' knowledge structure has also been found to predict classroom achievement (Warkentin, Griffin, & Bates, 1994). That is, students who rated the similarity between pairs of concepts more consistently tended to respond correctly to more test items.

This study, then, investigated the effects of cooperative assessment on achievement and knowledge structure in an educational psychology class for preservice teachers. The knowledge structure measures were included to provide a different, and perhaps more sensitive, measure of achievement on selected course concepts than that provided by the multiple choice posttest. The cooperative assessment design differs from that utilized by Lambiotte (1987) in that it occurs in a classroom context rather than a laboratory context, and is thus more ecologically valid. It also differs from those utilized by Webb (1993) and Bilsky-Toma (1993) in that students take tests first individually and then in groups to ensure individual accountability as well as group rewards. We have employed the design utilized by Griffin, et al., (1994) which was developed by R. R. McCown (personal communication, April 13, 1993) who found positive effects on student grades and student attitudes using this design.

Method

Subjects and Design.

Eighty-three students enrolled in four sections of educational psychology, taught by the second author, participated in the study. All students were education majors with a variety of areas of emphasis (e.g., early childhood, middle grades, physical education, etc.). There were 16 males and 30 females in the treatment group, and 9 males and 24 females in the control group. Nine students in the experimental and 4 students in the control group were Black, while the remainder were White.

A non-equivalent control group design was utilized, with treatments randomly assigned to groups. Scores from a pretest (discussed below) and self-reported GPAs were obtained and analyzed to determine if groups differed significantly in these areas. The groups were not statistically different on GPA F (1, 81) = 0.98; p = .33, though initial differences did exist on pretest performance, control M = 13.17, SD = 2.50, treatment M = 11.91, SD = 3.14, $F(1, 81) = 4.49$; p =.037. Analysis of covariance was utilized in later analyses in an effort to provide some statistical control for these initial differences in pretest scores.

The independent variable was testing condition in which participants either took tests first individually and then as part of a group (cooperative assessment) or completed tests on an individual basis only (individual assessment). Students in the treatment group could add up to ten points to their individual grade if performance on the group assessment exceeded their individual performance, but would not lose any points if individual performance surpassed group performance. Cooperative scores (i.e., the student's combined individual and group assessment scores) could not exceed group scores. For example, suppose Group A obtained a 96 on their group exam. Student A1 obtained an 82 on her individual exam, and is thus eligible for all 10 group points, bringing her cooperative assessment score to 92. Student A2 scored

88 on the individual exam, and is thus eligible for only 8 group points so that she does not exceed the group score, giving her a 96 on her exam. A score of 97 was obtained by Student A3 on the individual assessment, so she does not receive any group points and she keeps her grade of 97. Using this scoring procedure, we attempted to provide for both individual accountability and group incentives. Since students could add only up to 10 points to their individual score, they had to do some advance preparation for the exam to obtain a passing grade. By allowing students to earn extra points to add to their individual grades, incentive to work together as part of a group was also established as it was in everyone's best interest to obtain the highest group score possible to maximize the likelihood that students would earn group points and thus increase their exam grade.

The dependent variables were achievement, as measured on a posttest of educational psychology course content and similarity and coherence ratings of student knowledge structure representations as measured by KNOT Interlink (Schvaneveldt, 1990). The similarity index measures the effects of the instruction and testing condition on the similarity of the students' knowledge structures to the instructor's knowledge structure. The coherence measure provides an index of the internal consistency of each student's knowledge structure. Survey data from all students in the treatment group and follow up interviews with representative students in the treatment condition were also conducted in an attempt to learn more about the group process and attitudes toward cooperative assessment.

Materials

Pretest. The pretest consisted of 30 multiple choice items, 10 from each of the first three units of instruction for the introductory educational psychology course. The topics addressed on the pretest were operant conditioning and information processing (unit 1), observational learning, motivation, and outcome decisions (unit 2), and instructional models, instructional tactics, and classroom management (unit 3). The items on the pretest were primarily application type multiple choice items, and all were matched to unit objectives to ensure content validity.

Posttest. The posttest was comprised of the same 30 items which appeared on the pretest and addressed the first three units of instruction. The posttest was administered in three sections, 10 questions on each of the last three unit exams. For example, the first unit exam consisted of 40 items from chapters 5 and 6, and the second unit exam consisted of 30 items from chapters 7 through 9 along with 10 items from chapters 5 and 6. Thus, the last three exams were partially cumulative. The 10 cumulative items from each of the last three exams were added together to comprise the posttest, and student posttest score was defined as the number correct of these 30 items. This partially cumulative approach to testing was designed to provide a measure of the testing condition effects on individual learning. The posttest scores were taken from the first administration of the unit 2, 3, and 4 exams (for the experimental group) to measure the impact of treatment on individual performance.

KNOT Interlink. Students completed identical pretests and posttests of knowledge structures. For these tasks, students rated the similarity of pairs of concepts on a microcomputer using the program KNOT Interlink (Schvaneveldt, 1990). A total of 22 core concepts from the behavioral view of learning (positive reinforcement, negative reinforcement, punishment, shaping, timeout, operant conditioning, Premack principle, extinction, discriminative stimulus, token economy, response cost) and the information processing perspective (short term memory, long term memory, attention, encoding, retrieval, automaticity, prior knowledge, mnemonics, schema, metacognition, sensory register) were combined to form 231 unique concept pairs. Students rated all 231 pairs of concepts on relatedness (described in more detail below).

Group Assessment Survey. A 12 item survey was designed to tap student opinions and perceptions of the cooperative assessment process. These 12 items examined group process and student perceptions of the testing situation (See Appendix A).

Procedures

On the first day of the quarter students in all class sections which participated in the study completed the 30 item pretest of educational psychology concepts. Students also completed a questionnaire requesting demographic information which was used to assign students to heterogeneous groups.

During the first class meeting, the pretest of knowledge structure representation was also administered. Each student and the instructor of the courses rated each concept pair on the basis of semantic relatedness. Each concept pair was presented via microcomputer accompanied by a Likert scale with anchors "related" and "unrelated." Students indicated their judgment of relatedness by placing the cursor in the region of the scale which they thought best reflected the degree of relatedness of the two concepts. Using the Pathfinder algorithm (Schvaneveldt, 1990), KNOT then provided a graphic representation of the semantic network implied by the student's ratings of interrelatedness (see Figure 1), as well as a measure of the similarity between each student's network and the teacher's network. The similarity index is based on the proportion of links common to the student's and professor's network.

Students in all classes were assigned to cooperative groups of 4-5 members, which were mixed on the basis of gender, race, ability, and major. In terms of ability, each group consisted of a student with a high GPA (3.5-4.0), two students with average GPAs (2.5-3.49), and one student with a low GPA (below 2.49). Males and Black students were divided among the groups, and, to the extent possible after mixing the groups on all other variables, each group contained at least one student who was not an early childhood education major (the predominant major in all classes). Students in both conditions completed other types of group activities as well as cooperative assessment. Students worked in their groups to complete in-class, non-graded activities designed to facilitate acquisition of course concepts. Each group also worked to cooperatively complete a series of article critique papers concerning issues in educational psychology. These papers were graded and all group members received the same grade on each of the papers (4-5 papers, depending on the number of people in the group). In addition, the students in the cooperative assessment group completed the second administration of each unit exam working cooperatively with their group members.

With the exception of the difference in testing procedures, all sections of the course were taught in the same manner by the second author of the study. The two control group sections took place, one section each, during the summer and fall quarters of 1994. The experimental group sections were taught during the winter quarter of 1995. Care was taken by the instructor to use similar examples, to complete the same activities, and to cover the same amount of material per period in all classes. There were occasions, however, on which one group discussed something to a different depth or from a different angle than another group. Although this does not allow for strict control of the teaching conditions, more effort to deter these discussions was not made because the instructor was unwilling to interfere with the quality of the instruction in order to implement strict experimental controls. Class sessions were typically a mixture of lecture, discussion, generation of examples, examination of practical applications of course content, and group activities.

At the end of each of the four units students completed a unit examination. Each exam was comprised of 40 multiple choice questions (primarily higher-level items) and 2 essays items. Students in the experimental group were allowed to leave the room after completing the individual portion of the exam and were instructed to return at a time designated by the instructor. Students in the experimental group worked together to complete the test a second time after returning to class. Each group was directed to submit only one set of responses and come to consensus on the responses to the test items.

Exams were returned to students within two class periods for all sections of the course. During the period in which exams were returned, the instructor placed an answer key on the overhead projector so students could check the accuracy of the machine scoring. Copies of the exam were not distributed to students, but copies were available in the instructor's office for students' perusal. Very few students in either treatment condition chose to pursue this option.

Students in all groups completed the posttest knowledge structure task during the class meeting immediately after the first unit examination. The posttest of knowledge structure was identical to the pretest of knowledge structure.

The Group Assessment Survey was distributed on the last day of class. Students completed it at home and returned it on the day of the final examination. Following the final examination, all students were informed that they had been participating in an experiment throughout the quarter. They were briefly

told of the purpose of the study and assured that individual data would be kept confidential. At this point, students were also told that they could request that their data be withdrawn from the study if they chose not to participate. Students could exercise this option immediately or after grades were reported, in case they feared retribution for withdrawing their data from the study. No student chose to withdraw his or her results from the data pool.

Toward the end of the quarter in which data were collected for the treatment group, eight students from the cooperative assessment condition were asked to participate in individual interviews with the first investigator. The first investigator conducted the interviews, rather than the instructor of the course, in case students had any concerns to voice about the procedure. We anticipated that students would be less likely to offer criticisms of cooperative assessment to the instructor of the class in which they were completing the cooperative assessment procedure. Stratified random sampling was used to select the students; stratifications were based on self-reported GPA using the same categories used to assign students to cooperative groups. Initially, nine students agreed to complete interviews, but one student from the higher GPA group later was unable to attend the interview. The interviews were audiotape recorded (with the permission of the student). Each student was asked the same series of questions, but responses were explored in more depth as needed.

Results

Achievement and Knowledge Structure Posttests

All data were analyzed at the alpha = .05 level. Posttest data were analyzed using ANCOVA, with pretest scores and GPA entered as covariates. No significant differences were found between the groups on the posttest, control group $M = 21.03$, $SD = 3.29$, treatment group $M = 21.49$, $SD = 3.66$, $F(1, 77) = 3.28$; $p = .074$. The test for an interaction between pretest and treatment also yielded statistically insignificant differences, $F(1, 77) = 1.18$; $p = .28$, as did the interaction between GPA and treatment $F(1, 77) = 2.02$: $p = .159$. The effect of pretest scores was not significant $F(1, 77) = 3.85$; $p = .053$, but GPA was significant, $F(1, 77) = 19.65$; $p = .000$. Thus, students with higher GPAs scored higher on the posttest. The knowledge structure measures also indicated that no statistical differences existed between the groups on either the measure of similarity $F(1, 77) = 1.64$; $p = .20$, or the measure of coherence $F(1, 77) = .77$; $p = .38$. The effect of GPA was significant on both similarity $F(1, 77) = 5.03$; $p = .03$ and coherence $F(1, 77) = 7.71$; $p = .007$, but the effect of pretest was not significant in either analysis. Analyses of the interactions of GPA and pretest score were conducted with both the measures of coherence and similarity, but were not statisitically significant at any conventional levels.

Survey

Analysis of survey data indicated that student perceptions toward cooperative assessment were quite positive. Mean scores, modes, and standard deviations for the Group Assessment Survey are presented in Appendix A (note that items 3, 9 and 12 are scored in reverse). These results indicate that students tended to discuss test items with group members, prepared about the same amount for exams in this course (despite the fact they would receive extra points from group collaboration) as they did for exams in other courses, and felt that taking tests in groups was somewhat beneficial to their grades. Students also believed that they were learning more about the course content by discussing the exams with their groups.

Interview

The last source of data to be presented, interview data, was compatible with the outcomes of the survey. The questions asked during the interview fell into two major categories, affective and cognitive/metacognitive outcomes. Students indicated positive responses to most questions in both categories. That is, they reported that they were interested in learning, they valued the task, and they were determined to find the best answers and best supporting rationales for these answers. The responses to the metacognitive questions will be analyzed in more detail in the discussion section of this paper.

Discussion

Overview and Posttest Outcomes

This study examined the effect of cooperative assessment on the acquisition of educational psychology course content. It was predicted that students in the cooperative assessment group would outperform students in the individual assessment group on a posttest of educational psychology and on a measure of knowledge structure similarity and coherence. These hypotheses were not supported. Certainly the first explanation to be considered regarding this finding is that cooperative assessment does not improve achievement scores compared to an individual assessment format. Student perceptions of the procedure, however, appear to contradict this finding. As will be discussed in greater detail, students reacted positively to the treatment and felt that they learned more from this method of testing than from traditional methods. Still, it must be noted, student perceptions are not always accurate indicators of achievement gains.

Survey

The results of the survey administered at the end of the quarter indicated a generally positive regard by students for the cooperative assessment procedure. Survey items used a 7-point Likert scale with higher numbers indicating more positive responses except for the reverse scored items, 3, 9, and 12 (see Appendix A). Three major areas were emphasized in the survey questions: group testing process, study habits, and benefits of cooperative assessment.

Items 2, 4, 9, and 10 addressed the processes through which the groups went as they completed tests together. Note that item 9 is reverse scored, so that the lower mean corresponds with positive responses on items 2, 4, and 10. The survey indicates that students did tend to discuss the responses to test items more than they tended to rely on one person to provide a correct response and that all group members offered input to the discussion of individual items. Thus, it appears that students negotiated the responses to test items and attempted to reach a mutual understanding about the content. This finding is consistent with the results of the interview which indicate that for some items there was group discussion of the correct response, but for others items group consensus, rather than discussion, was the norm. Observations of the group testing phase by the instructor confirm that students did not discuss all items, but primarily discussed those items for which group members had selected a variety of responses. For these items, however, lively debates often ensued with different members trying to convince others of the accuracy of their individual response.

Preparation for unit exams was addressed in items 1, 5, and 6. It seems quite reasonable that students might study less for an exam on which they have the opportunity to improve their grades through group testing than in a traditional testing situation in which the grade they earn is solely the result of their own efforts. This possibility was, in fact, what prompted us to allow the group exam to count only a maximum of 10 points toward a student's exam grade. Students indicated that they and other members of their group were well-prepared for the exams (item 1). Survey results (item 5) indicate that students studied about the same amount as they usually do, despite knowing they had a group exam to complete. Of slightly more consequence to the students' preparation was the difficulty level of the exams (item 6); students indicated that they prepared somewhat more than usual due to the difficulty of the exams.

Last, the survey examined the potential benefits of the cooperative assessment method. As in interviews, students focused on both the benefits of cooperative assessment to their course grades as well as to their understanding of course concepts. Note that items 8 and 11 (intrinsic, learning goal) have higher means and lower standard deviations than item 7 (extrinsic goal). Question 12 also looked at students' goal orientation, and seems to indicate that students leaned more toward a learning goal orientation. Note that the lower mean on this reverse scored item indicates that students were concerned with receiving feedback from the group assessment, but were somewhat more concerned with the benefits of learning from their group. Overall, then, students perceive that the cooperative assessment procedure did benefit their understanding of course material. This finding is another plus for using cooperative assessment - students believe they are learning through this method as they work in groups to complete

testing tasks. Group discussion of exams places the responsibility of learning into the hands and minds of students, encouraging them to be responsible for their own acquisition and construction of knowledge.

Interview

The interviews focused primarily on issues related to affective outcomes (goals, social interactions of group members, self-efficacy) and cognitive/metacognitive outcomes (planning, monitoring and regulating group testing, evaluating, form of discussion) (see Appendix B). The primary focus of discussion for this paper is student responses to metacognition questions.

Interview data indicate that the procedure used by some of the groups to decide which items to discuss seemed to preclude many (perhaps too many) items from being debated and discussed. In some groups, if there was no disagreement about the correct response to an item, the item was not even read aloud. This appeared to limit the number of items, and therefore the amount of content, discussed. Perhaps, then, problems did not lie in the level or quality of the discourse, but the breadth of the discourse. As for the items which were discussed, we know little about them. Students were not asked to indicate which items they discussed nor to what extent they were discussed. Items may have been discussed because they were vague or ambiguous, not because they were particularly important. Perhaps, then, these items constrained students' attention to details or unimportant information. If students found it necessary to debate all test items maybe the strength of the cooperative assessment procedure would be increased.

Data also indicate that students perceived the cooperative assessment activity as a problem solving task. They focused on explaining why their answers were correct or incorrect and tried to clear up confusing issues. They also evaluated each other's reasoning about information and tried to determine the best answers and rationales for the answers. Furthermore, they discussed different perspectives on particular items and valued comparing different viewpoints. Students indicated that they were actively monitoring their peers' discussion when there was disagreement, especially when there was an "even split" with equal numbers of group members arguing for two different responses. Many diverse examples were generated during group discussion, indicating that students were clarifying and re-examining their knowledge of the information through interaction. However, all of this discourse activity was limited to the few items which were actually discussed.

Cognitive structures are modified when students begin to consider the relationships among concepts. The group testing phase of the cooperative assessment procedure allows students to wrestle with these relationships in groups as well as alone. The interview responses indicated that students did engage in metacognitive activities centered around the relationships among course concepts as they shared perspectives on and developed examples for the various concepts. These study strategies, however, affect the relationships of those concepts actually discussed. Therefore, if students did not discuss those concepts measured by the knowledge structure task or those measured on the content posttest, it is not surprising that no differences between treatment conditions would appear on these measures. Although interview data indicate that students did not discuss all of the content on the exams, students were unable to pinpoint exactly which content they did discuss during the group exam. The challenge becomes, then, to (a) determine which items students are actually discussing and to what extent, or, better still, to (b) somehow encourage students to discuss all content presented on the unit assessments.

One possible, and obvious, solution to the latter problem is to have students work on the tests as a group from the outset - that is, eliminate the individual portion of the assessment. While this might encourage more discussion, it also eliminates the provision made for individual accountability and may encourage more "free-riders." Another possible solution is to generate test items of a different type and/or difficult enough across all items to encourage more discussion. A third potential solution, and one which will be investigated in future studies, is to generate parallel items for the individual and group assessments. By taking this measure, individual accountability could be retained while students would be more likely to discuss all of the items because they had not seen any of them previously. Future studies will also examine the former problem, determining which items students are discussing and to what extent, through observation of group exams and discourse analysis of the discussions which take place during testing.

References

Bilsky-Torna, Z. (1993). "Son of S.T.A.D." - A comparison of group quiz and individual test scores. English Teachers Journal, 46, 35-43.

Brown, A. L., Bransford, J. D., Campione, J. C., & Ferrara, R. A. (1983). Learning, remembering, and understanding. In J. Flavell & E. Markman (Eds.), Handbook of child psychology: Vol. 3. Cognitive development (pp. 77 - 166). New York: Wiley.

Brown, R. (1989). Testing and thoughtfulness. Educational Leadership, 46, 31-33.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, 1. Hillsdale, NJ: Lawrence Erlbaum Associates.

Corno, L. (1986). The metacognitive control components of self-regulated learning. Contemporary Educational Psychology, 11, 333-346.

Corno, L., & Mandinach, E. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18, 88-100.

Duffy, T. M. & Bednar, A. K. (1991). Attempting to come to grips with alternative perspectives. Educational Technology, 31, (September), 12-15.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. Journal of Personality and Social Psychology, 54(1), 5-12.

Gabbert, B., Johnson, D. W., & Johnson, R. T. (1987). Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. The Journal of Psychology, 120(3), 265-278.

Goldsmith, T. E. & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt, (Ed.), Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex Publishing Corporation.

Greene, B. A., & Miller, R. B. (1993, April). Influences on course performance: Goals, perceived ability, and self-regulation. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Griffin, M. M., Quinn, G. P., McCown, R. R., & Driscoll, M. P. (1994, April). Learning through testing: An investigation of cooperative assessment. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Johnson, R. T., Johnson, D. W., & Smith, K. A. (1990). Cooperative learning: An active learning strategy for the college classroom. Baylor Education, 15(2), 11-16.

Kember, D., & Murphy, D. (1990). Alternative new directions for instructional design. Educational Technology, 30(8), 42-47.

Lambiotte, J. G., Dansereau, D. F., Rocklin, T. R., Fletcher, B., Hythecker, V. I., Larson, C. O., & O'Donnell, A. M. (1987). Cooperative learning and test taking: Transfer of skills. Contemporary Educational Psychology, 12, 52-61.

Nystrand, N. (1986). The structure of written discourse: Studies of reciprocity between readers and writers. New York: Academic Press.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33-40.

Sarason, I., Sarason, B., & Linder, K. (1983). Assessment and experimentally provided social support. Seattle: University of Washington Press.

Schvaneveldt, R. W. (Ed.). (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex Publishing Corporation.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21(4), 22-27.

Singer, R. S. (1990). The effects of collaborative testing on the test scores and the classroom attitudes of junior high school pre-algebra students. ERIC Document TM016072.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46, April, 4-9.

Slavin, R. E. (1991). Synthesis of research on cooperative learning. Educational Leadership, 48, (February), 71-82.

Vygotsky, L. (1978).

Warkentin, R. W., Griffin, B. W., & Bates, J. A. (1994, April). The relationship between students' study activities, content knowledge structure, academic self-efficacy, and classroom achievement. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Webb, N.M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. Educational Assessment, 1(2), 131-152.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership, 46, (April), 41-47.

Wynne, E. A. (1983). Teaching about cooperation. Phi Delta Kappan, 64, 497-499.

Appendix A

Means, Modes, and Standard Deviations for the Group Assessment Survey

Item		Mean	Mode	SD
1.	To what extent was each group member prepared for the exams? (unprepared to very well prepared)	6.04	6	0.94
2.	To what extent did each group member participate and offer input as you completed the group exams? (did not participate or offer input to participated and offered extensive input)	6.47	7	0.94
3.	My group members seemed to study less and rely on the group's success to raise their exam grades.	1.70*	1	1.26
4.	In general, our group discussed and debated the answers to most of the test items.	5.78*	7	1.68
5.	What effect did taking the group exam have on your study habits? (I studied much less than usual to I studied much more than usual)	4.77	4	0.97
6.	What effect did the difficulty level of the exams have on your study habits? (I studied much less than usual to I studied much more than usual)	5.30	6	1.08
7.	Taking the tests as part of a group was beneficial to my grade.	6.10*	7	1.19
8.	Taking the tests as part of a group was beneficial to my understanding of required course concepts.	6.42*	7	0.77
9.	In general, our group did not debate the responses to most items, but relied upon one or two members to provide the answers.	2.33*	1	1.93
10.	The group as a whole worked together to complete all of the essay items.	4.94*	6, 7	1.91
11.	As I completed the group exam, I gained a better understanding of content I missed when I took the test individually.	6.50*	7	0.83

12.	As I completed the group exam, I was more concerned with determining which items I answered correctly on my own than with learning from the group.	3.00*	4	1.41

*All items used a 7-point Likert scale, with higher numbers indicating more positive responses with the exception of items 3, 9, and 12 in which case the lower numbers indicate a more positive response. End points of the Likert scale for all items (unless otherwise indicated following the item) were "Strongly disagree" (1) to "Strongly agree" (7).

Appendix B

Interview Questions and Response Summaries

I. Affective/Process Outcomes

A. Goal

Q: What was the purpose of the peer-group test?

1. Problem solving
   - to explain why answers are correct or incorrect (clear up confusing points or wrong answers)
   - discuss reasoning behind answers
2. Multiple views are highlighted
   - to see others' perspectives, others' thinking
3. Benefits of group discussion
   - discussion should improve learning, memory, understanding

B. Social/Behavioral Interactions of Group Members

Q: Did everybody in the group do about the same amount of talking or did one particular person dominate the discussion? Was it a pleasant experience?

1. Amount of talking
   - students who knew more or had more experience talked more, but most felt that all students had equal power
2. Pleasant experience
   - pleasant experience to discuss the test with peers

Q: How did you respond to each other's comments during the group testing session? Can you give me an example of your response to someone in the group?

Students typically
- engaged in positive exchanges (equity, respect, fairness, acceptance, no hostility)
- tried to understand the reason behind peers' responses to items
- tried to evaluate the best answer

C. Self-efficacy

Q: Was the group test a difficult challenge for you and for your group? Why or why not?

- students apparently did not feel challenged except when a lot of disagreement occurs, then they feel challenged to resolve the disagreement

II. Cognitive/Metacognitive Outcomes

A. Planning

Q: Did you do anything special to prepare for the peer-group testing activity outside of class? If so, what?

1. Study strategies
   - Group assessment did not change students' typical methods (strategies) for test preparation. Rather, they tried to apply their usual study techniques to "fit" the group discussion activity.
2. Level of preparation
   - some felt more responsible to be more prepared than usual because they were concerned about peer pressure

Q: What was your goal during the activity? What did you hope to accomplish?

1. External goal
   - All students mentioned increasing grades, number of points, or score on test. This seemed to be their primary goal.
2. Internal goal
   - 7 of 8 students also mentioned that they wanted to learn, to understand, to get other perspectives on the material, or to clarify misunderstandings

B. Monitoring and Regulating

Q: What was the most difficult part of the group test?

1. Disparity of answers
   - difficulty occurred especially when there was great disparity in group answers
2. Consensus-seeking
   - trying to get everyone to agree which was difficult because either
     1) no one is sure of the content (lack of knowledge)
     2) each person is more likely to believe their answer is correct
     3) more persuasion had to be performed, which requires more effort

Q: Did you spend you time on some particular items more than others? If so, why?

1. Multiple choice items
   - particular items created more disagreement, taking more time to work through discussion and reach consensus
2. Essay
   - required students to first decide who responded to which item, to discuss individual answers, and to choose or construct best group response

Q: Did you ever go back to items you already answered? Why?

-students paid attention to and utilized test taking strategies to regulate their thinking, such as noting difficulty of item, elapsed time, illumination of previous response by later discussion

Q: What did you do when you didn't know an answer or didn't know if your answer was correct?

Various strategies used
- went with best explanation or example, deciding this person must know answer
- temporarily skip the item and come back to it
- went with majority opinion
- went with "expert" opinion (rarely done)
- guessed (rarely done)

Q: When you were taking the group test, how did you decide on the best answers to the multiple choice items? Give an example.

•First, groups would ask for each person's response to an item. Procedures varied based on consensus or lack of consensus.

1. Consensus
   - If all agreed, then go on to next item with little or no discussion
2. Lack of consensus
   - some disagreement, but majority agree then groups discussed a little bit but often opted for majority view
   - major disagreement, with equal representation of two responses lead to much discussion and a search for the best explanation or justification, followed by a decision on correct response

Q: When you were taking the group test, how did you decide on the answers to the essay? Give an example.

•First students determined which essay question was answered most frequently.

1. All or majority answered same question
   - each reported and explained their answer, then group discussed the answers and selected the best response to use as the group response
2. If different questions were answered
   - group decided on one to respond to, but some people were left out of the response generation

C. Evaluating Activity

Q: What did you learn in the group testing process?

Process helped to clarify information

- generated more and diverse examples of concepts from different perspectives
- remembered teacher's or textbook's examples (review)
- clarified relationships among concepts
- applied concepts rather than merely memorizing definitions

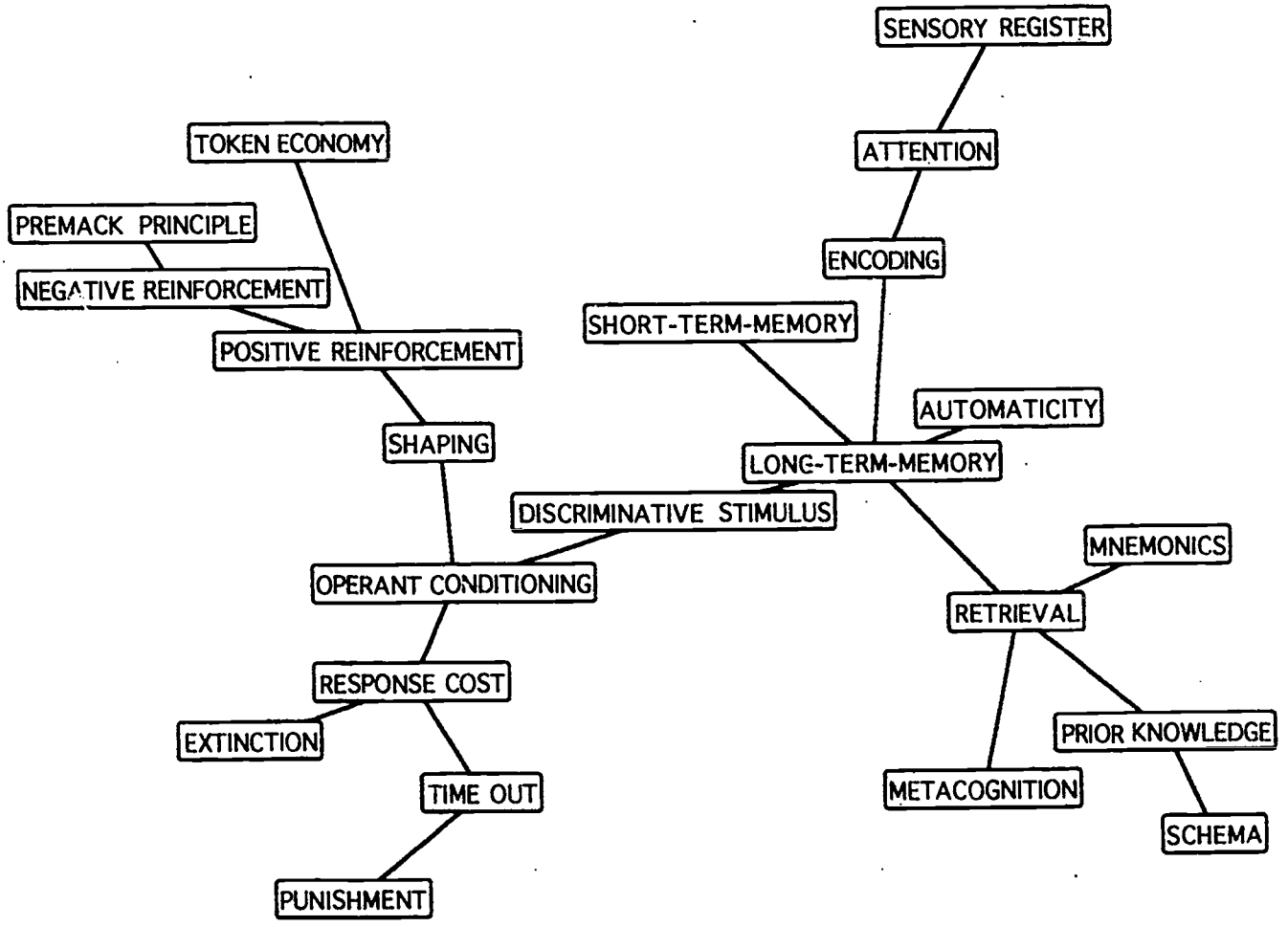

Figure 1. Average student's Pathfinder network of 22 educational psychology concepts.