DOCUMENT RESUME

ED 392 812 TM 024 449

AUTHOR Webb, Noreen M.; And Others
TITLE Using Group Collaboration as a Window into Students' Cognitive Processes.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 95
CONTRACT R117G10027
NOTE 30p.; Based on a 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/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Cognitive Processes; *Comprehension; *Cooperation; Educational Assessment; Evaluation Methods; Grade 7; Grade 8; *Group Discussion; Interaction; Junior High Schools; *Junior High School Students; Performance Factors; Research Methodology; Student Evaluation; Test Results; *Verbal Communication; Videotape Recordings

ABSTRACT

This study was part of a larger study investigating the influence of group collaboration during a group assessment on performance of the group and subsequent widespread performances. This study compared the verbal interactions of students while they completed a science assessment in small collaborative groups with their performance on the same test administered individually. What their verbalizations revealed about their cognitive processes and how verbalizations corroborated their performance scores and helped clarify their responses were analyzed. From a sample of 662 seventh and eighth graders, a subsample of 6 students from each class (mainly eighth graders), for a total of 126 students, was selected. Group discussions during the assessment were videotaped and later coded for analysis. Results indicated that analyzing verbal interactions might be a useful source of information about students' thinking to supplement individual test results. More than half of the students gave evidence of their degree of comprehension during group discussion, and a substantial portion gave different information in their verbal interaction than they gave on the individual test. The less information they gave on the test, the more useful the verbal interaction was in revealing their understanding. An appendix gives sample questions from the hands-on test. (Contains 5 tables and 21 references.) (SLD)

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – April 1995

Project 2.1 Designs for Assessing Individual and Group Problem Solving

Review of Factors Influencing Group Assessment Outcomes

Using Group Collaboration as a Window Into Students' Cognitive Processes

Noreen Webb, Project Director

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program, cooperative agreement number R117G10027 and CFDA catalog number 84.117G, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Using Group Collaboration as a Window Into Students' Cognitive Processes[1]

Noreen M. Webb, Kariane Nemer and Alex Chizhik
Center for Research on Evaluation, Standards, and Student Testing
University of California, Los Angeles

Brenda Sugrue
University of Northern Colorado

Introduction

One of the goals of alternative forms of assessment is to accurately measure students' ability to reason and solve problems. An important validity issue is the extent to which the scores based on students' performance reflect their higher order thinking. A number of approaches to examining the degree to which performance assessments measure higher order thinking skills have been proposed or are being developed (Baker, O'Neil, & Linn, 1993).

One approach is that used by Baxter and Glaser in which they interview students and observe them while they conduct scientific investigations. The results of several of these studies (Baxter, Elder, & Glaser, 1994; Baxter, Glaser, & Raghavan, 1993) show that the scores on their science assessments do differentiate between different levels of cognitive processes concerning planning, monitoring, the quality of the solution strategy, and students' explanations of their solutions: "High scorers provided a more complete plan for approaching the task, were more strategic in their problem-solving approach, engaged more frequently in self-monitoring activity, and generated better explanations of content related concepts than low scorers" (Baxter et al., 1994, p. 2).

Using another approach, Sugrue and Valdes (1995; see also Sugrue, in press) investigated whether alternative formats and approaches for measuring the same science construct give rise to the same interpretations of students' conceptual

[1] This report is based on a paper presented at the 1995 annual meeting of the American Educational Research Association, San Francisco, April 19.

understanding. The same concept, such as voltage in electric circuits, was tested in a variety of ways, including asking students to select one of several electric circuits with the highest voltage, selecting the correct prediction about what would happen to voltage if an electric circuit were altered, drawing diagrams of circuits with certain characteristics, manipulating real objects (batteries, bulbs, wires) to construct circuits with certain characteristics, and writing explanations for why voltage was higher in one of their drawn or constructed circuits than in another. Sugrue and Valdes reported only moderate correlations among students' scores on concepts tested in different ways, suggesting that the different approaches were measuring different aspects of students' thinking in science.

Still another approach is that used by Hamilton, Nussbaum, and Snow (1995), who carried out concurrent interviews to identify reasoning processes used by students taking various kinds of tests. Hamilton et al. asked students to think aloud while they solved multiple-choice items, constructed response items, or completed hands-on performance tasks in science. The interviews provided important information about the reasoning students used to solve problems, the relationship between their knowledge and reasoning used, and the role of other skills, such as reading comprehension ability, in students' performance on science tests. The verbal protocols revealed that students sometimes used quite different processes to solve problems than were intended by the test developers (e.g., analysis of spatial models in memory instead of declarative knowledge) and showed which kinds of previous experiences inside or outside the classroom helped students solve problems on the tests. Furthermore, the verbal protocols showed that some tasks did not necessarily evoke scientific reasoning and that correct answers sometimes could be obtained or constructed by using everyday experiences instead of scientific reasoning or by using flawed scientific reasoning.

The present study used group collaboration as a window into students' thinking processes. Previous research on collaborative group work in the classroom shows that students verbalize their thinking in the process of helping one another, working together to solve a problem or complete a task, resolving disagreements, and justifying their actions, strategies, and decisions (Webb, 1991, 1993; Webb & Farivar, 1994; Webb & Palincsar, in press). The few studies of group collaboration in assessment contexts suggest that students also verbalize their thinking while solving problems on group assessments (Fall, Webb, & Wise, 1995; Saner, McCaffrey, Stecher, Klein, & Bell, 1994; Wise & Behuniak, 1993).

The present study, then, compared students' verbal interaction while they completed a science assessment in collaborative small groups with their performance on the same test administered individually. The purposes of this study were (a) to determine whether students' verbalizations during group discussion revealed information about their cognitive processes, (b) to determine the degree to which their verbalizations during group discussion corroborated their performance scores on the individual test, and (c) to determine how students' verbalizations during group discussion helped to clarify their responses on the individual test.

The test used here was itself designed to elicit a variety of information about students' understanding of the science concepts being assessed. The assessment was a hands-on test of electric circuits in which students were asked to assemble materials to produce two circuits, one brighter than the other, draw a diagram of their completed circuits, answer multiple-choice questions about voltage, resistance, and current in their circuits, and generate explanations to explain their multiple-choice responses. This assessment, then, used several different formats that may tap different dimensions of students' knowledge and understanding: constructing solutions to a problem, selecting a response from a set of alternatives, and generating explanations of certain concepts or principles (see Sugrue, in press).

The study reported here (comparing information provided by students' performance on individual tests and their verbal interaction during group discussion) was part of a larger project designed to investigate the influence of group collaboration during a group assessment on performance of the group and subsequent performance of students on individually-administered tests. The issues being investigated in the larger project are the influence of group composition on (a) the processes that take place during the group assessment, (b) the group's performance on the group assessment, (c) how much students learn from the group collaboration, and (d) students' performance on subsequent individual assessments. Those analyses are forthcoming.

Method

Sample

The sample for the study consisted of 662 seventh-grade and eighth-grade students from five schools in Los Angeles County. Six teachers volunteered for the study; the number of classes for each teacher ranged from 2 to 5, for a total of 21 classes.

The schools came from different locations in the county and represented a wide mix of student characteristics. The total sample of students was 86% eighth-grade, 14% seventh-grade; 50% female, 50% male; 45% White, 40% Hispanic, 10% Black, and 5% Asian.

The subsample analyzed here consisted of six students from each class, for a total of 126 students. The subsample of students was 86% eighth-grade, 14% seventh-grade; 52% female, 48% male; 40% White, 40% Hispanic, 14% Black, and 6% Asian. Students in the subsample, then, had similar characteristics to those in the whole sample.

Design

The design of the study was carried out in two stages, called Phase 1 and Phase 2. At the beginning of Phase 1, all students were administered a variety of pretests measuring verbal and nonverbal reasoning skills: the New Jersey Test of Reasoning Skills, a vocabulary test, and Raven's Progressive Matrices. Then teachers conducted a three-week unit on electricity and electric circuits in their classrooms. Teachers used their own curriculum materials and were given minimal instructions on the content and procedures of the instructional unit. At the end of the instructional unit, students took two tests: (a) a paper-and-pencil test on concepts in electricity and electric circuits, and (b) a hands-on test on the same concepts with similar format and questions (with batteries, bulbs, wires, and resistors for students to connect). The order of the tests was counterbalanced in each classroom: half of the students completed the written test on the first day and the hands-on test on the second day; the other half of the students completed the hands-on test on the first day and the written test on the second day. All students completed these tests individually.

After a one-month interval without any instruction or review of electricity or electric circuits, students were administered the written and hands-on tests again

(Phase 2). On the first day, all students completed the hands-on test either in three-person collaborative small groups (81% of the students in each class) or individually (19% of the students in each class). On the second day, all students completed the written test individually.

For the group test, students worked on the hands-on test in three-person collaborative small groups. Although each student in the group was given a test paper to complete and turn in, students were encouraged to work together on the test questions and to help one another understand how to solve the problems. Students in all groups were videotaped as they completed the hands-on test. Students in each class were assigned to groups of three kinds: (a) one group with three students whose performance on the Phase 1 tests showed that they had little understanding of electricity and electric circuits (called "low achievers" in this paper), (b) one group with one knowledgeable student (called a "high achiever") and two students with little knowledge or understanding, and (c) groups with a mixture of knowledge and understanding in the group. All groups were heterogeneous on gender and ethnic background to reflect the mix of characteristics of students in the class.

The first two group compositions (uniformly low-achieving groups and mixed groups) were formed to determine the impact of a knowledgeable student on the performance and group processes that occur in group tests and the subsequent performance of students on an individual test (results of these analyses are forthcoming). Each student in the those types of groups had his or her own high-quality clip-on microphone to improve the audio quality of the videotape. The other groups were videotaped using the built-in microphone of the camcorder. On the basis of the high audio quality for the uniformly-low and mixed groups and the planned differences in group composition, those two groups in each class were selected for the analyses presented here.

After data collection had been completed, it was discovered that one high-achieving student has been mistakenly classified as a low achiever when groups were formed. Instead of the 21 groups of each type that were initially planned, then, there were 22 mixed groups (n = 66) and 20 uniform groups (n = 60).

Tests

Because the hands-on test was the only test that students completed individually (Phase 1) and in collaborative groups (Phase 2), only those tests were analyzed here. The hands-on test had two tasks. Each task consisted of a collection of materials for students to assemble into two electric circuits, one brighter (or dimmer) than the other. For Task 1, students were given two 1.5-volt batteries, two 9-volt batteries, 3 light bulbs in bulb holders (labeled A, B, and C), and wires. For Task 2, students were given two 9-volt batteries, two light bulbs (labeled A and B), three pieces of graphite (resistors), and wires. After assembling the circuits and drawing diagrams of their circuits, students answered three questions on voltage, resistance, and current. Each question had two parts: a multiple-choice item followed by an explanation. The questions were the same for both tasks. The instructions and questions appear in the Appendix.

Scoring of Individual Tests (Phase 1)

Students assembled a great variety of electric circuits. Several alternative circuits fulfilled the stated requirements of the task. Other students' circuits did not (for example, some students did not use all items when assembling their circuits). The scoring of the questions on voltage, resistance, and current was based on the circuits that students assembled, regardless of whether the circuits fulfilled the stated requirements. To avoid penalizing students for their drawing ability or their ability to draw a diagram that corresponded exactly to the circuits that they assembled, their circuits were videotaped. In most cases, students' multiple-choice items were scored based on the circuits that they assembled rather than on their drawings. In a few cases, students clearly answered the multiple-choice items using their diagrams rather than the actual circuits, so their diagrams were used to score their multiple-choice responses.

Because a number of students chose the correct multiple-choice alternative but gave an explanation that indicated misunderstanding of the concept, students' responses to the multiple-choice items were scored separately from their explanations for why they chose a particular response (1 for correct, 0 for incorrect). (Procedures are underway to develop a more comprehensive system for coding students' explanations that will distinguish between different levels of knowledge and understanding and will clarify specific misconceptions that students hold.)

Scores on the hands-on tests in Phase 2 were not analyzed here because they could not be assumed to reflect the understanding of individual students. Students' responses on the Phase 2 hands-on tests may have been a reflection of the group discussion and other students' responses, not their own understanding or misunderstanding.

Coding of Videotapes of Verbal Interaction in Group Discussion (Phase 2)

The videotapes of group discussion were used to code instances where students demonstrated understanding or misunderstanding of each concept. The only instances coded were those that demonstrated the thinking of a student that was not influenced by what other students said or did. Any utterance that could have been influenced by previous discussion in the group was not coded. For example, if a student said something that was similar to what another student had already said, the utterance was not coded. Students' verbal interaction was coded for each concept (voltage, resistance, current) as 1 (shows understanding), 0 (shows misconception), or as "gives no information."

Results

Overall Performance Level of Subsample in Phase 1

As described previously, six students in two groups were analyzed from each class. One group had all low-achieving students; one group had one high-achieving and two low-achieving students. Because five out of six students in each class analyzed here were low-achieving students compared to the rest of the class, the subsample analyzed here was assumed to have a lower level of performance than the whole sample of the study. Table 1 compares the performance of the subsample analyzed here with the performance of the whole sample of the study on the multiple-choice items. The results given in Table 1 confirm the assumption that the subsample was a low-achieving subset of the whole sample. On every concept on the Phase 1 test, the mean scores of the subsample were lower than the mean scores of the whole sample. This finding suggests that the findings of the remaining analyses may not generalize to the remainder of the whole sample.

Comparison of Information on Individual Test and From Verbal Interaction in Group Discussion

Table 2 compares the information from the Phase 1 individual test with the information from verbal interaction in group discussion in Phase 2. For each concept, Table 2 shows the percent of students for whom (a) the same information was shown by their Phase 1 individual test scores and their verbal interaction

Table 1

Student Achievement in the Subsample With Verbal Interaction Data and in the Whole Sample (Phase 1)

	Concept in electric circuits					
	Voltage		Resistance		Current	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
	M[a] (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*S)D*
Subsample analyzed here (n = 126)	.40 (.49)	.35 (.48)	.41 (.49)	.45 (.50)	.47 (.50)	.54 (.50)
Whole sample (n = 439)	.60 (.49)	.49 (.50)	.48 (.50)	.67 (.47)	.59 (.49)	.67 (.47)

Note. All students completed the Phase 1 hands-on test individually.
[a] Proportion of students giving correct response.

during group discussion, (b) there were discrepancies between students' understanding as shown by their Phase 1 individual test scores and their verbal interaction during group discussion during Phase 2, and (c) group discussion showed no unambiguous information on student understanding or misunderstanding that could be coded. Of the 126 students in the subsample, 18 students did not perform any work on Task 1 and 19 students did not perform any work on Task 2. Because there was no way of knowing what their performance would have been if they had taken the test, their scores were not included in the analyses, resulting in 108 and 107 students with usable data on Tasks 1 and 2, respectively. Students who did perform work on a task but skipped one or more of the multiple-choice items were included in Table 2.

The overall results in Table 2 show that about half of the students revealed information about their level of understanding of the science concepts in their verbal interactions with other students. For some of these students, the information revealed during group discussion was similar to the information they provided on the individual test. For others, the information revealed during group discussion differed from the responses they had given on the individual test.

Same information from individual test and group discussion. As can be seen in Table 2, about a third of the students (from 24% to 40% of the students across the three concepts for the two tasks) demonstrated the same level of

Table 2

Comparison of Students' Individual Test Performance and Verbal Interaction in Group Discussion (Percent of Students)

	Concept in electric circuits					
	Voltage		Resistance		Current	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Same information from verbal interaction and test	38	33	40	34	24	28
Test item correct	23	16	21	22	15	19
Test item incorrect	15	17	19	12	9	9
Discrepancies	20	21	23	22	20	16
Verbal interaction shows student understanding	14	12	16	15	13	6
Test item incorrect	13	12	15	13	10	5
Test item skipped	1	0	1	2	3	1
Verbal interaction shows student misconception	6	9	7	7	7	10
Test item correct	3	5	3	3	5	8
Test item skipped	3	4	4	4	2	2
No information from group discussion	43	47	37	43	57	57
Test item correct	13	8	11	11	18	17
Test item incorrect	24	31	19	24	29	28
Test item skipped	6	8	7	8	10	12

Note: Percentages do not always sum exactly due to rounding. n = 108 for Task 1; n = 107 for Task 2.

understanding of the concepts on the individual test and in their group discussion. They either gave the same multiple-choice response during group discussion that they had selected on the individual test, gave a similar explanation, or both.

Discrepancies between individual test and group discussion. As can be seen in Table 2, the percent of students for whom the individual test and verbal interaction in group discussion provided discrepant information about students' understanding ranged from 16% to 23% across the three concepts for the two tasks. As the detailed breakdown in the middle part of Table 2 shows, the majority of the discrepancies were in the direction of students showing greater understanding of the concepts in their verbal interaction in group discussion than on their individual tests. Most of these students showed understanding during their verbal interaction but had selected the wrong response on the individual test. A few students had skipped the item on the individual test.

Less than half of the discrepancies were in the direction of students showing a lower level of understanding during group discussion than they had demonstrated on the individual test. Some of these students had answered correctly on the individual test but showed in their verbal interaction that they did not understand the concept. The remainder had skipped the item on the individual test but showed a misconception in their verbal interaction.

It is instructive to look in detail at the information that students demonstrated during the group discussion. Most students for whom the individual test and the group discussion provided conflicting information about their understanding simply gave a different multiple-choice response than they selected on their individual test. The remainder, however, not only gave a different response during group discussion than they had given on the individual test, but also gave a rationale that showed a different level of understanding (or misunderstanding) than they had demonstrated on the test.

In many cases, students verbalized more information or more accurate information about their understanding than they had indicated on their individual test. The following excerpts illustrate five different ways in which students demonstrated more understanding during group discussion.

Sometimes students left an entire item blank on the individual test, but showed some understanding in their verbal interaction. Student 2 in the following excerpt had skipped the item on voltage on the individual test but showed in verbal interaction that he understood the role of the voltage of individual batteries:

1 The highest voltage is [circuit] A, I think.
2 Highest voltage . . . [circuit] B.
1 B? But it's different.
2 Yeah. Nine plus 1.5 [volts].
1 Oh, that's right.

Students often wrote no explanation on their test to accompany their multiple-choice selection but gave an indication in their verbal interaction that they had some knowledge about the concept. In the following excerpt, Student 1 recognizes a connection between voltage and the number of volts in the battery, but seems confused by the decimal numbers involved:

1 Which of the two circuits you made has the highest voltage? A has higher voltage? Huh?
2 Yeah.
1 Why?
2 I don't know.
1 Because it has . . . What's more voltage: a 9-volt or this one [points to a 1.5-volt battery]?
2 A 9-volt. This is only one and a half. That's nine.
1 I thought this was 1.5 or something like that.

The confusion about decimal points was a common problem. In the following excerpt, students interpreted 1.5-volt batteries as having 15 volts:

1 Why is [circuit] B brighter? 9 volt plus 15 volt . . .
2 I am waiting to see what you write.
1 Just put 9 volt plus 15 volt equals high voltage.
3 Is used in circuit B.

Sometimes students had written an illogical explanation on their test but verbalized a more sensible explanation during group discussion. In the following case, in which Circuit A has more resistors, and higher resistance, than Circuit B, Student 2 had stated on his individual test that Circuit B had more resistance because "it looks like its resistance will last longer," an explanation that makes little sense. In his verbal interaction, he selected the correct response and indicated a partial, if poorly stated, understanding of the role of resistors:

1 Why is resistance higher in Circuit A?
2 Because it will slow down the energy to the graphite.

Some students gave a more specific or precise explanation in their verbal interaction than they had given on the individual test. In the following example, Student 1 had written on the individual test that circuit B had more resistance than circuit A because "circuit B doesn't light up as bright as circuit A". This response gives little information about the student's understanding of resistance but instead shows only that the student can correctly interpret evidence of higher resistance (in this case, the voltage in the two circuits was the same, so the evidence was correct). This student's verbal interaction clarified his understanding of resistance:

2 Why is bulb A in circuit A brighter than bulb B in circuit B?
1 Because [there are] two light bulbs [he points to circuit B].
2 Because there are more batteries . . .
1 No, there is the same amount [of batteries]. There is more resistance [pointing at circuit B] because there is an extra thing [bulb].
2 Resistance. OK.

Finally, some students corrected an obviously wrong explanation that they had given on the individual test. In one example, a student indicated a conception on the individual test that resistance was related to the *kinds* of wires and *kinds* of bulbs in a circuit. This student wrote on the individual test that a circuit with two bulbs and a circuit with one bulb had the same resistance because "they have the same kind of wires and same kind of bulbs." In the group discussion, he gave a more accurate response that related resistance to the *number* of bulbs: "Which got the higher resistance? They both have the same resistance. No, B does [correct]. Want to know why? Because more bulbs [pointing to circuit B] takes more resistance."

In other cases, students' clarification during verbal interaction indicated a misconception that was not evident on the individual test. In the following example, the student had written a similar explanation on the written test to that of a student described above: Circuit B has higher resistance than circuit A "because it's not as bright as circuit A." This student's verbal interaction, however, revealed that he associated resistance with the number of wires, a partial misconception: "Higher resistance is still [circuit] B because there are more wires."

Some students demonstrated in their verbal interaction that they either did not know the vocabulary or did not understand the underlying concept, even though they had selected the correct response on the individual test. In the following excerpt, circuit A has less resistance than circuit B, but the group has decided that circuit A has more resistance:

1 Why does A have more resistance?
2 More resistance? What's resistance?
3 More light?
2 Yeah? Are you sure?
3 I don't know. I'm just saying.
2 Just put A does not have any resistance. I don't even know what resistance is.

Sometimes students demonstrated in their verbal interaction some confusion between terms, such as the confusion between current and voltage shown by Student 1 in the following excerpt:

1 [reading from the test] Which of the two circuits you made has the highest voltage?
2 They have equal, but . . . see they are both the same because it does not matter the current, because they both have two, wait . . . [looking closely at the batteries] 1.5 and . . . oh, it's the same because they are both [a 1.5-volt battery and a 9-volt battery] here and they are both there [pointing to the batteries in the two circuits].
1 But, doesn't voltage become less when there is more resistance?
2 No. That was the current.
1 Oh, that's right. OK, so they have the same voltage.
2 Why?
1 Because we used the same batteries in each.

The following excerpt for another group shows a similar confusion between current and voltage, but without any resolution:

1 I don't remember what current is.
2 Neither do I.
1 Current is the push of this [pointing to one of the batteries].
2 I think it is the force that it comes out of the

anode [pointing to one of the bulbs].
1 So it means how much voltage there is?
2 I think [circuit] A [has more current] obviously.
1 Yeah. They have the same [current] because, how couldn't they? They have the same batteries. Unless one is older than the other.

Sometimes students indicated a lack of confidence in their correct answers, as shown by Student 2 in the following excerpt:

1 Circuit A has the highest resistance?
2 I don't know.
1 Why did you put that then? [Student 2 reads from student 1's paper:] "It has more graphite"?
2 I don't know.

Finally, some students made it clear in their verbal interaction that their answer on the test was a chance result:

1 [reading from test] Which of the circuits you made has the highest resistance?
2 Eenie, meenie, minie, mo, catch a tiger by his toe: [circuit] A! [as she points back and forth between the circuits]
1 [circuit] A because it has more wires.
3 Wait. Which one is brighter?
2 [circuit] A! Now just put it!
3 No, [circuit] A is dimmer. [circuit] B because it's brighter.

The excerpts given above show the great variety of ways that students demonstrated their understanding or misconceptions during verbal interaction that was not evident from their responses on the individual test.

No information from group discussion. Roughly half of the students (from 37% to 57% of students across the tasks) either did not say anything about the concept during group discussion or were not coded because it was unclear whether their verbalization reflected their own prior understanding or was influenced by the verbalizations of other students in the group.

Role of Group Composition on Information From Group Discussion

The final analyses investigated whether the composition of the group influenced the foregoing results. First, based on previous research showing that students with high academic status tend to participate more actively, give more suggestions, direction, and information, and have more influence than students with low academic status in mixed groups (Berger, Rosenholtz, & Zelditch, 1980; Cohen, 1982, 1994; Cohen, Lotan, & Catanzarite, 1990), it was hypothesized that high achievers in this study would be more active in group discussion than low achievers. Consequently, it was expected that high achievers would reveal more evidence of their level of understanding than would low achievers.

If high achievers were particularly active, this could have different effects on the behavior of low achievers. The active verbalizations of high achievers could stimulate the discussion of low achievers by motivating them to participate in group discussion and giving them a model to follow. In this case, low-achieving students may verbalize their thinking. On the other hand, low achievers could feel intimidated by the much greater competence of the high achievers, could be too discouraged to participate actively in group discussion, and may believe that they have little to contribute to the group. Consequently, they may not verbalize much of their thinking.

In uniformly low-achieving groups, several outcomes were also possible. If students perceive that they all have the same level of competence, students may be willing to express their thinking without fear of being compared negatively with others. On the other hand, students may believe that their group has too little competence to solve the problems and may be too discouraged to work actively on the problems or verbalize their thinking.

Given the wide variety of possible influences of group composition on students' verbalizations of their thinking, no hypotheses were posed here about the influence of group composition on the behavior of low-achieving students.

Tables 3, 4, and 5 compare individual test performance and verbal interaction in group discussion for high-achieving students in mixed groups, low achievers in uniformly low-achieving groups, and low achievers in mixed groups. The results confirm the hypothesis that high-achieving students would be the most active and would show the most evidence of their thinking. As can be seen in Table 3, the percentage of high achievers who verbalized their thinking was fairly

Table 3

Comparison of Students' Individual Test Performance and Verbal Interaction in Group Discussion: High Achievers in Mixed Groups (Percent of Students)

	Concept in electric circuits					
	Voltage		Resistance		Current	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Same information from verbal interaction and test	64	69	77	64	51	73
Test item correct	64	64	68	59	46	68
Test item incorrect	0	5	9	5	5	5
Discrepancies	10	9	9	5	9	9
Verbal interaction shows student understanding	5	0	0	0	0	0
Test item incorrect	5	0	0	0	0	0
Test item skipped	0	0	0	0	0	0
Verbal interaction shows student misconception	5	9	9	5	9	9
Test item correct	5	0	9	5	9	9
Test item skipped	0	0	0	0	0	0
No information from group discussion	28	24	14	32	41	19
Test item correct	23	14	14	27	32	14
Test item incorrect	5	5	0	0	9	0
Test item skipped	0	5	0	5	0	5

Note: Percentages do not always sum exactly due to rounding. $n = 22$ for Task 1; $n = 22$ for Task 2.

high, ranging from 59% to 86% across tasks. The verbalizations of high achievers generally confirmed their test performance. The majority of students had answered the items on the individual test correctly and verbalized their correct thinking during group collaboration. Few high achievers showed discrepancies between their performance on the individual test and their verbalizations in group discussion.

The results in Tables 3 and 4 show very little difference between low achievers in uniform and mixed groups. Low achievers in the two kinds of groups showed similar patterns of revealing the same information as in their group discussion, showing discrepant information between the test and group discussion,

Table 4

Comparison of Students' Individual Test Performance and Verbal Interaction in Group Discussion: Low Achievers in Uniformly Low Groups (Percent of Students)

	Concept in electric circuits					
	Voltage		Resistance		Current	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Same information from verbal interaction and test	30	21	35	22	16	14
Test item correct	10	0	8	14	4	6
Test item incorrect	20	21	27	8	12	8
Discrepancies	28	16	26	27	20	18
Verbal interaction shows student understanding	20	6	18	19	10	8
Test item incorrect	18	6	16	15	8	8
Test item skipped	2	0	2	4	2	0
Verbal interaction shows student misconception	8	10	8	8	10	10
Test item correct	2	4	2	4	6	6
Test item skipped	6	6	6	4	4	4
No information from group discussion	41	64	38	53	63	70
Test item correct	12	10	10	8	20	21
Test item incorrect	25	44	22	37	33	37
Test item skipped	4	10	6	8	10	12

Note: Percentages do not always sum exactly due to rounding. n = 49 for Task 1; n = 52 for Task 2.

and providing no information about their level of understanding in their verbal interaction in group discussion. Therefore, the presence or absence of a high-achieving student seemed to have little effect on the extent to which low achievers verbalized their thinking.

The lack of any effect for group composition removes a potentially complicating factor in using group discussions to elicit information about student understanding. What remains to be seen in future analyses is whether the lack of effect of group composition found here will generalize to other group compositions (for example, groups with students of medium competence relative to other students in the class).

Table 5

Comparison of Students' Individual Test Performance and Verbal Interaction in Group Discussion: Low Achievers in Mixed Groups (Percent of Students)

	Concept in electric circuits					
	Voltage		Resistance		Current	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Same information from verbal interaction and test	32	27	27	36	19	21
Test item correct	16	9	11	12	11	6
Test item incorrect	16	18	16	24	8	15
Discrepancies	14	36	25	24	24	15
Verbal interaction shows student understanding	11	30	22	18	24	6
Test item incorrect	11	30	22	18	19	3
Test item skipped	0	0	0	0	5	3
Verbal interaction shows student misconception	3	6	3	6	0	9
Test item correct	3	3	0	0	0	9
Test item skipped	0	3	3	6	0	0
No information from group discussion	54	36	49	39	56	63
Test item correct	8	3	11	6	5	12
Test item incorrect	35	27	27	21	35	33
Test item skipped	11	6	11	12	16	18

Note: Percentages do not always sum exactly due to rounding. $n = 37$ for Task 1; $n = 33$ for Task 2.

Discussion

The results of this study show that analyzing verbal interaction during group collaboration may be a useful source of information about students' thinking to supplement their individual test performance. More than half of the students in this study gave evidence of their understanding or misunderstanding during group discussion. Moreover, a substantial portion of students gave different information in their verbal interaction than they gave on the individual test, revealing either more understanding or less understanding of the scientific concepts than they had shown on the individual test.

Verbal interaction during group collaboration provided the most useful information for students who had given little or no information on their individual test. The less information students provided on their tests, the more useful was the verbal interaction from the group discussion. The verbal interaction during group collaboration was especially informative when (a) students did not complete an item, either failing to indicate their selection on the multiple-choice part of an item or failing to provide a written explanation for their choice, or (b) their written explanations were ambiguous or difficult to understand or interpret. Students who gave little information on their individual test were more likely to talk about their thinking during collaboration than they were to write about their thinking on the test.

Verbal interaction also provided useful information for students who gave incorrect answers on the test. Among students who verbalized their thinking during group collaboration and who had given incorrect answers on the individual test, more than half demonstrated greater understanding of the concepts than they had shown on the test. The remaining students demonstrated the same level of misunderstanding that they had shown on the test.

Verbal interaction was least useful for students who demonstrated a fairly high level of understanding on the individual test. When students gave elaborate, detailed, or clear written explanations on the individual test, their verbal interaction during the group discussion usually confirmed their level of understanding.

Several unanswered questions from this study remain, with implications for future research. This study showed that analyzing verbal interaction during group collaboration can yield important information about students' understanding that tests administered individually may not. This was the case in the present study with an individual test that was designed to elicit a variety of information about student understanding. The information revealed by group collaboration may be even more significant for tests that reveal less about the reasoning that students used to solve problems. Further research is needed to examine the relative payoff of analyzing group collaboration for tests with different formats, tasks, and topics.

Another set of interesting issues for future research concerns the relative advantages and disadvantages of using individual think-aloud protocols (concurrent or retrospective) and collaborative group discussions to elicit information about students' thinking. One advantage of group discussions is the relative ease of eliciting information. When students work together on a problem, they reveal information about their knowledge, problem-solving strategies, and hypotheses through the normal course of communicating with their teammates. Although think-aloud protocols have been used successfully to reveal students' thinking while they solve problems in non-assessment contexts (Ericsson & Simon, 1984; Greeno & Simon, 1988) and on assessments (Norris, 1989), they may not elicit information about students' thinking processes as readily as in group discussions. A disadvantage of collaborative interaction, on the other hand, is the potential ambiguity of students' verbalization. A student's verbalization may be influenced by what others have said, rather than reflecting only his or her own understanding. Although students verbalized a great deal in this study, only a small portion of their verbalizations were clearly identified as reflecting only their own understanding. Future research should compare the relative information gained from these methods.

Finally, this study investigated two specific group compositions: groups with or without an "expert" in the material. Analyses of other kinds of groups should be conducted to determine whether other mixes of achievement levels may influence how much students reveal about their thinking during group discussion. In addition, there are a wide variety of other student characteristics that need to be considered as group composition variables, including gender, ethnic background, and peer status. Research on group collaboration in classroom settings shows that group composition on these variables can influence the level of activity of students in the group and the nature of their participation (Webb, in press; Webb & Palincsar, in press). These group composition variables may also influence the level and nature of participation of students working on assessments in collaborative groups.

References

Baker, E. L., O'Neil, Jr., H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48,* 1210-1218.

Baxter, G. P., Elder, A. D., & Glaser, R. (1994, April). *Cognitive analysis of a science performance assessment.* Report prepared for Center for Research on Evaluation Standards, and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Baxter, G. P., Glaser, R., & Raghavan, K. (1993, October). *Analysis of cognitive demand in selected alternative science assessments.* Report prepared for Center for Research on Evaluation Standards, and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Berger, J., Rosenholtz, S. J., & Zelditch, M. (1980). Status organizing processes. *Annual Review of Sociology, 6,* 479-508.

Cohen, E. G. (1982). Expectation states and interracial interaction in school settings. *American Review of Sociology, 8,* 209-235.

Cohen, E. G. (1994) Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-36.

Cohen, E. G., Lotan, R., & Catanzarite, L. (1990). Treating status problems in the cooperative classroom. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 203-230). New York: Praeger.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fall, R., Webb, N. M., & Wise, N. (1995, April). *Group discussion and large-scale language arts assessment: Effects on students' comprehension.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Greeno, J. G., & Simon, H. A. (1988). Problem solving and reasoning. In R. C. Atkinson (Ed.), *Steven's handbook of experimental psychology* (pp. 589-672). New York: Wiley.

Hamilton, L. S., Nussbaum, M., & Snow, R. E. (1995, April). *Alternative interview procedures for validating science assessments.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher, 18*, 21-26.

Saner, H., McCaffrey, D., Stecher, B., Klein, S., & Bell, R. *(1994) The effects of working in pairs in science performance assessments.* Santa Monica, CA: The RAND Corporation. Manuscript submitted for publication.

Sugrue, B. (in press). A theory-based framework for assessing domain-specific problem solving ability. *Educational Measurement: Issues and Practice.*

Sugrue, B., & Valdes, R. (1995, April). *Interpreting cognitive components of performance from converging sources of evidence.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22,* 366-389.

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

Webb, N. M. (in press). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis.*

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31*(2), 369-396.

Webb, N. M., & Palincsar, A. S. (in press). Group processes in the classroom. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology.*

Wise, N., & Behuniak, P. (1993, April). *Collaboration in student assessment.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Appendix

Hands-on Tests

CRESST/UCLA ELECTRIC CIRCUITS HANDS-ON TEST, 1994

BAG 1

Name: ______________________________

Date: ______________________

1. Use the items in Bag 1 to make **two circuits** on the white paper mat on your desk.

Follow these rules:

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be brighter than Bulb B.**
- **There should be one 9-volt battery in each circuit.**
- **Use all of the items in the bag but do not use any item more than once.** For example, if you put Bulb C in Circuit A, you cannot also put it in Circuit B.

2. In the boxes below, draw diagrams of the circuits you made.

Circuit A (brighter)	**Circuit B (dimmer)**

3. Why is Bulb A in Circuit A brighter than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

____

____

____

4. Which of the two circuits you made has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

5. Which of the two circuits you made has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

6. Which of the two circuits you made has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

____

____

____

CRESST/UCLA ELECTRIC CIRCUITS HANDS-ON TEST, 1994

BAG 2

Name: ______________________________

Date: ______________________

1. Use the items in Bag 2 to make **two circuits** on the white paper mat on your desk.

Follow these rules:

- **Bulb A should be in Circuit A. Bulb B should be in Circuit B.**
- **Bulb A should be dimmer than Bulb B.**
- **Use all of the items in the bag but do not use any item more than once.** For example, if you put one piece of graphite in circuit A, then you could not put three pieces in Circuit B (there would be only two pieces of graphite left to use).

2. In the boxes below, draw diagrams of the circuits you made.

Circuit A (dimmer)	**Circuit B (brighter)**

3. Why is Bulb A in Circuit A dimmer than Bulb B in Circuit B? (Try to use scientific terms in your answer.)

____

____

____

4. Which of the two circuits you made has the **highest voltage**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME VOLTAGE**

Why? (Try to use scientific terms in your answer.)

____

____

____

5. Which of the two circuits you made has the **highest resistance**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME RESISTANCE**

Why? (Try to use scientific terms in your answer.)

____

____

____

6. Which of the two circuits you made has the **highest current**?

Circle one: **CIRCUIT A** **CIRCUIT B** **BOTH CIRCUITS HAVE THE SAME CURRENT**

Why? (Try to use scientific terms in your answer.)

____

____

____