DOCUMENT RESUME

ED 353 299 TM 019 363

AUTHOR Webb, Noreen

TITLE Collaborative Group versus Individual Assessment in Mathematics: Group Processes and Outcomes. Review of Group Assessment Issues. Project 2.3: Enhancing the Utility of Performance Assessments: Domain-Independent R&D.

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 Nov 92

CONTRACT R117G10027

NOTE 28p.

PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Classroom Techniques; Comparative Analysis; Competence; *Cooperation; Evaluation Methods; Grade 7; *Group Dynamics; Junior High Schools; *Junior High School Students; Mathematics Skills; Middle Schools; *Student Evaluation

IDENTIFIERS *Individualized Evaluation; Middle School Students; Performance Based Evaluation

ABSTRACT

Large-scale assessment programs are beginning to design group assessment tasks in which small groups of students collaborate to solve problems or complete projects. However, little is known about the validity of data from group assessment for making inferences about the competence of individual students. The present study compared students' performance in small-group and individual assessment contexts to determine how well achievement scores from group work represented the skills of individual students, and to determine what additional information about students' skills was provided by data on group dynamics and group problem-solving processes. Two seventh-grade general mathematics classes taught by the same teacher at an urban middle school participated in the study. The sample included 53 students (45 percent males and 55 percent females). Sixty-six percent of the students were Hispanic American, 21 percent were Anglo American, 11 percent were African American, and 2 percent were Asian American. During a curriculum unit on operations with decimal numbers, students worked in collaborative small groups (three or four students in a group) for one class period to calculate the costs of long distance telephone calls. Two weeks later, after a review session, students worked on a similar problem individually without collaborating with others. The results show that performance in the group setting was much greater than performance in the individual setting, and that data on group processes gave important insights into students' mathematics skills and their behavior in collaborative groups. Eight data tables are included. (Author/RLC)

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

Final Deliverable – November 1992

Project 2.3: Enhancing the Utility of Performance Assessments: Domain-Independent R&D

Review of Group Assessment Issues

Collaborative Group Versus Individual Assessment in Mathematics: Group Processes and Outcomes

Noreen Webb, Project Director

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

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part under the Educational Research and Development Center Program cooperative agreement 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.

All opinions in this report are those of the author.

Acknowledgement

I would like to thank Sydney Farivar, Jack Hinds, Chris Iwasa, Colleen Nakamura, and Sandy Nonhof for participating in the study, and Maika Watanabe, Jonathan Troper, and Randy Fall for their help in preliminary coding of the data. I would also like to thank Rich Shavelson and Joan Herman for their helpful comments on an earlier draft of the paper.

Abstract

Large-scale assessment programs are beginning to design group assessment tasks in which small groups of students collaborate to solve problems or complete projects. Little is known, however, about the validity of data from group assessment for making inferences about the competence of individual students. The present study compared performance in small-group and individual assessment contexts to determine how well achievement scores from group work represented the skills of individual students, and to determine what additional information about students' skills was provided by data on group dynamics and group problem-solving processes. The results showed that performance in the group setting was much greater than performance in the individual setting, and that data on group processes gave important insights into students' mathematics skills and their behavior in collaborative groups.

Collaborative Group Versus Individual Assessment in Mathematics: Group Processes and Outcomes

Noreen Webb

CRESST/University of California, Los Angeles

Large-scale assessment programs are increasingly turning to group assessment in which small groups of students collaborate to solve problems or complete projects instead of, or in addition to, students working on tasks individually (e.g., Connecticut's Common Core of Learning Assessment: Lomask, Baron, Greigh, & Harrison, 1992; California Assessment Program: Pandey, 1991; Shavelson & Baxter, 1992). One reason for using group assessment is to reflect the growing importance being placed on group collaboration and group problem solving in instruction. Because group work can facilitate learning (Slavin, 1990), school districts and state departments of education have started to mandate the use of cooperative and collaborative learning methods on a large scale (e.g., California Department of Education, 1985, 1992). To the extent that assessment practices influence the curriculum, group testing affirms the importance of group collaboration in instruction.

Second, what students can accomplish in teams is important to potential employers who are increasingly using work teams to respond to global competition (Hackman, 1990). Assessing students in groups provides information about group productivity and group effectiveness that individual assessment of student skills does not.

Third, group assessment makes it possible to measure students' abilities to collaborate with others. Team effectiveness involves many dynamic processes including, for example, coordination, communication, conflict resolution, decision making, problem solving, and negotiation (Salas, Dickinson, Converse, & Tannenbaum, 1992). Observing students collaborating with others makes it possible to evaluate their ability to work with others and their ability to monitor and shape their own behavior (Redding, 1992).

Fourth, group assessment can be used to measure students' problem-solving processes. When students work with others to solve problems, they freely verbalize their knowledge, understanding, problem-solving strategies and misconceptions (see Shavelson, Webb, Stasz, & McArthur, 1988). They may reveal much more about their understanding than can be inferred from responses to questions on an individual test.

Fifth, the drive toward authentic assessment calls for complex problems in realistic contexts (Meyer, 1992). Complex problems may be less intimidating to students if they can work with others.

Finally, group testing is sometimes used for logistical reasons, such as making more efficient use of limited test materials. Some performance assessments use special equipment that would be very expensive to duplicate for every student to be tested, and so are used with groups of students to save costs (e.g., electric circuits, Shavelson & Baxter, 1992).

Many testing programs stress individual accountability and obtain achievement scores for individual students from group assessment. But it is unclear whether the performance of students in collaborative group contexts accurately represents their individual competence. Part of the uncertainty hinges on the definition of a valid measure of individual competence. From one perspective, individual competence is best measured by individuals working alone without assistance (the traditional individual testing context). Group assessment contexts that give students opportunities to collaborate may overestimate individual competence when students use resources in the group to solve problems that they would not be able to solve individually. This is especially a concern when students are allowed to collaborate on all aspects of the task, including the work that they will submit for evaluation (e.g., Shavelson & Baxter, 1992).

Differences between performance in group and individual settings have long been documented in out-of-school contexts (Hare, 1992; Kahan, Webb, Shavelson, & Stolzenberg, 1985), and occasionally in educational contexts (e.g., Johnson, Johnson, & Skon, 1979), but rarely have been studied in educational assessment contexts. In non-assessment contexts, students often perform better when collaborating with others, due to cognitive factors (e.g., greater intellectual resources available) and social variables (e.g., increased task motivation; Knight & Bohlmeyer, 1990). But negatively functioning groups can sometimes produce worse performance than individuals working alone (Hackman, 1990). So scores from group assessment contexts may overestimate or underestimate students' performance in an individual setting.

A social constructivist perspective presents a somewhat different view of individual competence. While individual competence can be measured by individuals working alone, it can also be demonstrated when individuals collaborate with others to learn how to solve problems that they could not previously solve by themselves (Vygotsky, 1978). In a truly collaborative context, all individuals are actively engaged in working toward a solution to the problem (Damon & Phelps, 1989; Tudge & Rogoff, 1989). From this perspective, the performance of students working collaboratively with others would be a valid measure of individual competence when students are actively involved in learning how to solve the problem. On the other hand, when students use the group's resources to obtain a solution or an answer without trying to learn how to solve the problem (e.g., copying other students' work without trying to understand it, carrying out the arithmetic operations after another student has set up the solution to the problem), scores from the group assessment context will overestimate their individual competence.

From both perspectives on what constitutes individual competence, then, scores from a group assessment context may not be valid indicators of students' individual competence. Furthermore, achievement scores from group assessment contexts provide little information about group functioning. Studies of group dynamics in instructional settings show that data on group processes are necessary for understanding how groups operate and the experiences of students in them (Webb, 1989, 1991). Group process data can reveal the extent and nature of individual student participation as well as the nature of the group's collaboration (e.g., conflict and controversy, joint construction of ideas and solutions, helping relationships, beneficial and debilitating social processes, see Webb & Palincsar, in preparation).

The present study, then, compared performance in small-group and individual assessment contexts to examine the following questions: (a) How closely do achievement scores from a group collaboration context correspond to scores of students working individually? (b) What additional information about students' skills is provided by data on group dynamics and group problem-solving processes? The group context used in this study allowed students to collaborate on all aspects of the task. The individual testing context allowed no collaboration.

Method

Sample

Two seventh-grade general mathematics classes taught by the same teacher at an urban middle school participated in the study (n = 53). The gender breakdown was 55% female, 45% male. The ethnic backgrounds of the students were Hispanic (66% of the sample), Anglo (21%), African-American (11%) and Asian-American (2%). Because the number of African-American students was too small to analyze separately, and because African-American and Hispanic students showed similar scores on all measured variables (statistical comparisons on all tests and behavior variables were nonsignificant, $p < .55$ or greater), these two groups were combined into one group for further analysis. Similarly, the one Asian-American student was combined with the Anglo students.

Design

During a curriculum unit on operations with decimal numbers, students worked in collaborative small groups (3 or 4 students in a group) for one class period to calculate the costs of long distance telephone calls. Two weeks later, after a review session, students worked on a similar problem individually without collaborating with others.

Procedures

Group problem-solving. Prior to the study, students participated in activities designed to help them work effectively in groups. They carried out activities designed to make them feel more comfortable in the classroom (e.g., learning their classmates' names and interests), and practiced basic communication and social skills (e.g., attentive listening, no put downs, moderate voice level, checking for understanding, sharing ideas and information, encouraging, checking for agreement; see Farivar & Webb, 1991).

Students were then assigned to groups that were heterogeneous on mathematics achievement, gender, and ethnic background to work on mathematical material (multiplication of decimals). Students were encouraged to collaborate when solving the problems. They were also instructed to make sure that everyone in the group understood how to solve the problems and to help students having difficulty. At the beginning of the class period, the teacher modeled the solutions to several problems. The teacher then assigned problems for the groups to solve. Her role was to monitor group functioning but not to give assistance. Several students asked her for assistance; in response, she directed them to work with their group ("You have to ask your group" and "Check with other people").

Students were each required to submit papers showing all of their work in solving each problem. Students had been working in small groups for about ten days at the time of data collection for this study.

To obtain records of group discussions, all groups were tape recorded for the entire class period. Using stereo tape recorders, a clip-on microphone for each student, and one observer per group, it was possible to identify the speaker of each utterance on the tapes. The classes were tape recorded on prior occasions to familiarize them with the procedures and the presence of the observers.

Individual test. The day before the individual test, students practiced solving problems similar to those that would appear on the test. The teacher modeled solutions to some problems; students practiced solving others. For individual testing, students worked individually without assistance from other students or from the teacher. As in the group context, the problems had a free-response format and students were instructed to show all work on their test papers.

Group and Individual Tests

Problems. In the group session, students were given a table of telephone rates for various prefixes (with three columns for the prefix, cost for the first minute, and the cost of each additional minute) and were asked to calculate the cost of three long distance calls. Because not all groups finished the third problem, only the first two were analyzed here: (a) Find the cost of a 30-minute call to the 771 prefix, and (b) Find the cost of a 11-minute call to the 781 prefix.

For both problems, the cost of the first minute was $0.22 and the cost of each additional minute was $0.13.

On the individual test, students were asked to solve the following problem: "A long distance call to San Francisco costs $0.30 for the first minute plus $0.08 for each additional minute. What is the cost of a 10-minute call?"

The problems on the group and individual tests were designed to be as comparable as possible. The major difference between them was the presentation of the costs in a table versus a sentence. Because some students in the group initially had some difficulty interpreting the table, it could be argued that the problems presented in the group were slightly more difficult than the problem on the individual test. Because computational accuracy (numerical accuracy of multiplication and addition, and placement of the decimal point) was not scored in this study (see below), the slight difference in numerical values between the two tests should not have influenced the results.

Scoring. Students' written work on each problem from group and individual testing was scored as correct or incorrect on nine components, reflecting all of the errors that students made: recognizing that the call involved multiple minutes, creating a subgroup of additional minutes that was less than the total number of minutes in the call, determining the correct number of additional minutes, applying a single cost to each additional minute, using the correct cost for each additional minute, using the correct arithmetic operation to calculate the cost of the additional minutes, creating a one-minute subset for the first minute, using the correct cost for the first minute, and using the correct arithmetic operation to combine the cost of the first minute and the cost of the additional minutes. The scoring emphasized conceptual understanding; numerical computational errors (e.g., multiplying two numbers incorrectly) were not scored. An overall score for each student was obtained by averaging over all components and problems on a test.

Two coders scored all written work; interrater agreement exceeded 99%. The consistency of scores across components of a problem was very high (internal consistency alpha ranged from .91 to .97 across group work and individual problems). Due to ceiling effects and restriction of range on the group work problems, it was not possible to obtain a reasonable measure of the consistency of total scores across the two group work problems. Due to the very high mean scores on the two group work problems (.93 and .96 respectively; the percent of the sample obtaining perfect scores on the two problems was 79% and 86%, respectively), the correlation between the total scores on the two group work problems was low (r = .32) and, consequently, internal consistency alpha was low (.34).

Coding of Group Processes

Transcripts of the tape recordings of group work were used to identify and categorize student behavior on each component of every problem. Five categories of behavior accounted for the majority of students' experiences in group work: students (a) solved the problems correctly with little or no assistance from others, (b) made errors, were corrected, and were told the correct procedures for solving the problem, (c) indicated that they did not understand or were confused, and were told the correct procedures for solving the problem, (d) copied other students' work without doing it themselves, or (e) did not contribute verbally to group discussion. Because most students had the same experience for many or all components of a problem, the category of behavior that best represented a student's experience was the one used for further analysis. Two raters independently categorized student behavior and agreed on 93% of the codes.

Coding of ability level. To determine the effect of ability level on performance and behavior, ability level was based on scores from a 13-item mathematics pretest administered at the beginning of the study. The pretest included numerical exercises and word problems using whole numbers and decimals (internal consistency alpha = .74). The sample was split into thirds corresponding to high ability (36%), medium ability (34%), and low ability (30%).

Results

Achievement

Table 1 gives the mean proportion correct for each component of the problem in the group and individual settings. On every component of the problem, as well as for the total problem, mean performance was significantly

Table 1

Performance in Group and Individual Settings

Component	Group *M*	Group *SD*	Individual *M*	Individual *SD*	Paired *t*
Recognizes that call has multiple minutes	.98	.10	.75	.43	3.62*
Creates subgroup of additional minutes that is less than the total length of call	.95	.18	.60	.49	5.23**
Creates correct size subgroup of additional minutes	.90	.25	.49	.51	5.90**
Applies single cost to each additional minute	.93	.20	.60	.49	4.91**
Uses correct cost for each additional minute	.98	.10	.74	.44	3.83**
Uses correct operation to calculate cost of additional minutes	.97	.12	.58	.50	5.67**
Creates a separate subgroup for first minute	.95	.15	.72	.45	3.68*
Uses correct cost for first minute	.93	.20	.58	.50	5.34**
Uses correct operation to combine costs of first and additional minutes	.92	.21	.58	.50	5.01**
All components	.95	.14	.63	.43	5.44**

* $p < .01$. ** $p < .001$.

higher in the group setting than in the individual setting ($p < .01$). Performance in group work was close to the ceiling of 1.00.

While there was a significant drop in mean performance from group to individual settings, the pattern of individual students' performance across the two settings varied considerably. Table 2 shows the percent of the sample whose scores increased, decreased, or did not change across the two settings. As can be seen in Table 2, about half of the sample (51%) obtained the same

Table 2

Patterns of Changes in Performance From Group to Individual Settings

		Score			
		Group		Individual	
Category of Change	Number of Students	*M*	*sd*	*M*	*sd*
Increase or no change	27 (51%)[a]	.94	.18	.96	.15
Increase	3 (6%)	.78	.24	.96	.06
No change	24 (45%)	.96	.16	.96	.16
Decrease	26 (49%)	.95	.09	.28	.34
Range:					
0 to -.25	5 (9%)	.99	.02	.89	.00
-.26 to -.50	2 (4%)	.83	.24	.44	.16
-.51 to -.75	7 (13%)	.92	.11	.29	.09
-.76 to -1.00	12 (23%)	.97	.06	.00	.00

[a] Percent of sample ($n = 53$).

score or increased slightly across testing settings. Most of these students performed well in both settings. The other half of the sample (49%) showed a decrease from the group to individual settings. Most of these decreases were quite large (see Table 2), showing that students performed well when they worked in groups but performed poorly when tested individually.

To determine whether background characteristics of the students influenced performance, patterns of performance were examined separately for students from each ethnic background (Anglo vs. Hispanic/African-American), gender (female vs. male), and ability level (high vs. medium vs. low). Because females and males did not differ on any test or behavior variable ($p < .29$ or greater) those results are not presented here. Because ethnic background was not significantly related to ability (difference between ethnic groups on the pretest was not statistically significant, $t = 1.54$, $p < .14$; distribution of each ethnic background across high, medium, low ability levels was not statistically significant, chi-square(2) = 1.81, $p < .41$), ethnic background and ability level were analyzed as distinct variables.

Table 3 shows the distribution of performance for students from each ethnic background and ability level. The pattern of performance was significantly different between Anglo students and Hispanic/African-American students. Most of the Anglo students (83%) performed as well on the individual test as on the group test (no change or an increase in scores from the group test to the individual test; see Table 3), whereas less than half of the Hispanic/African-American students did so. Conversely, very few Anglo students (17%) obtained lower scores on the individual test than on the group test, whereas over half of the Hispanic/African-American students (59%) did so (the difference between these patterns was statistically significant: chi-square(1) = 4.94, $p < .03$).

Table 3

Distribution of Performance of Students From Different Ethnic Backgrounds and Ability Levels

Change From Group to Individual Setting	Ethnic Background: Anglo ($n = 12$)	Ethnic Background: Hispanic/African-American ($n = 41$)	Ability Level: High ($n = 19$)	Ability Level: Medium ($n = 18$)	Ability Level: Low ($n = 16$)
No increase or change	10[a] (83%)[b]	17 (42%)	13 (68%)	10 (56%)	4 (25%)
Increase	1 (8%)	2 (5%)	2 (11%)	1 (6%)	0 (0%)
No change	9 (75%)	15 (37%)	11 (58%)	9 (50%)	4 (25%)
Decrease	2 (17%)	24 (59%)	6 (32%)	8 (44%)	12 (75%)
Range:					
0 to -.25	1 (8%)	4 (10%)	3 (16%)	1 (6%)	1 (6%)
-.25 to -.50	0 (0%)	2 (5%)	2 (11%)	0 (0%)	0 (0%)
-.51 to -.75	0 (0%)	7 (17%)	1 (5%)	2 (11%)	4 (25%)
-.76 to -1.00	1 (8%)	11 (27%)	0 (0%)	5 (28%)	7 (44%)

[a] Number of students. [b] Percent of ethnic background or ability subgroup.

A similar pattern emerged for the three ability levels. A majority of the high ability students (68%) performed as well on the individual test as on the group test (increase or no change, see Table 3); about half of the medium ability students (56%) did so; but only a small subset of the low-ability students did so (25%). The difference between the patterns of performance across the three ability levels was statistically significant (chi-square(2) = 6.78, $p < .04$). In summary, Hispanic/African-American and lower-ability students were overrepresented among those who scored lower on the individual test than on the group test.

Group Processes

Students' behavior in group work was analyzed to understand why their performance was often worse on the individual test than in the group setting. Table 4 presents the number of students in each behavior category and the corresponding mean scores in the group and individual settings. The results are presented separately for students who obtained the same scores on the group and individual tests and students who obtained lower scores on the individual test than on the group test.

Solved problems correctly without assistance. As can be seen in Table 4, nearly half of the sample (n = 25, 47%) solved the problems correctly during group work without assistance. These students also tended to do well on the individual test. Presumably they knew how to solve the problems and their score in group work was an accurate indication of their understanding.

Used group resources to obtain solutions. A large portion of the sample (n = 21, 40%) used resources of the group to obtain solutions to the problems. Over a quarter of the sample (n = 15, 28%) showed, by making errors or asking questions, that they were having difficulty with the problems and received assistance to solve them (see Table 4). Eleven of these students asked questions or made statements indicating confusion or lack of understanding. Whereas all of these students received enough assistance to show correct work on their paper in group work, only four of these students obtained high scores on the individual test. The remaining seven students obtained very low scores on the individual test.

Table 4

Student Behavior and Performance in Group Work

Behavior Category	Number of Students	Mean Score Group	Mean Score Individual	Mean Score Change
Increase or no change from group to individual settings	27 (51%)[a]	.94	.96	.02
Solved problems correctly	20 (38%)	.96	.99	.03
Made errors, was corrected and told procedures	1 (2%)	1.00	1.00	.00
Did not understand, was told procedures	4 (8%)	.99	1.00	.01
Did not contribute to group discussion	2 (4%)	.61	.61	.00
Decrease from group to individual settings	26 (49%)	.95	.28	-.67
Solved problem correctly	5 (9%)	.99	.89	-.10
Made errors, was corrected and told procedures	3 (6%)	.89	.30	-.59
Did not understand, was told procedures	7 (13%)	.96	.10	-.86
Copied others' work	6 (11%)	.94	.13	-.81
Did not contribute to group discussion	5 (9%)	.94	.11	-.83

[a] Percent of sample.

One reason for the difference between the two groups of students may be the effort they expended to try to understand the assistance they received. Table 5 gives representative excerpts from group work that contrast the experiences of students who obtained high scores on the individual test with the experiences of students who obtained low scores on the individual test. In the first excerpt, the student clearly tried to understand the procedures that the group gave him. This excerpt represented the efforts of all of the students who did equally well on the group and individual test. In the second excerpt, the student used the procedures she was given to obtain the correct solution but did not try to understand them. This was true of all students who performed

Table 5

Excerpts from Group Work for Confused Students

Problem: 30-minute call, $0.22 first minute, $0.13 each additional minute.

EXAMPLE 1: STUDENTS WITH HIGH SCORES ON INDIVIDUAL TEST (STUDENT INSISTS ON UNDERSTANDING PROCEDURES)

A I don't know how to do it.

D OK...the first minute is 22 cents.

B So each additional is 13.

D Yeah. So add 13 cents times...times 29...because for the first minute it's 22 cents, so there is a minute. And then 13 cents times 29. You didn't understand.

A Nope! OK, let me get this straight. It's a 30 minute call to 711...There is 30 minutes. So why do you [put 29]?

D There is the first minute, 22 cents. Now multiply 13 cents times 29. Because 29 minutes are left after the first minute.

A Well, it's 30 minutes. But you are saying, do what?

D Multiply 29 times 13 cents.

A Why 29? This is 30.

D Because they already got a minute. That's the first minute.

A Oh, Ok. Thank you.

EXAMPLE 2: STUDENTS WITH LOW SCORES ON INDIVIDUAL TEST (STUDENT USES PROCEDURES BUT DOES NOT TRY TO UNDERSTAND THEM)

C How come you got 29?

A First, you have to say 29 times 13. And then plus 22.

C 29 times 13?

A Yeah, cause you already have another minute right there. That's a minute right there.

C Are you sure it's 13 times 29?

A Yeah.

worse on the individual test than on the group test, with two exceptions. The two exceptions were students who did try to understand the explanations they received but nevertheless could not solve the problem on the individual test.

A similar picture emerged among the four students who made errors. All four students were given the correct procedures, but only one of them obtained a high score on the individual test. This result can be explained by whether students understood the assistance they received. Table 6 gives excerpts from group work that contrast the experiences of the student who obtained a high score on the individual test with those of the students who scored poorly on the individual test. The first excerpt in Table 6 shows the student making an error, being corrected, and receiving an explanation of the correct procedures. The second excerpt, in contrast, shows a student making the same kind of error, and using the procedures to solve the problem, but clearly not understanding them. This excerpt represents the experience of all three students who made errors, received assistance, but obtained low scores on the individual test. Interestingly, the tendency of the group to be unconcerned about whether students understood the procedures was typical.

A few students (n = 6, 11% of the sample) clearly copied other students' work without trying to understand it (see Table 7). Their low scores on the individual test show that they did not use the work they copied to learn how to solve the problems for themselves.

Of the students who used the group's resources to obtain correct solutions to the problems in group work, either by receiving assistance and explanations or by copying others' work, then, only a third of these students (n = 7) were actively engaged in understanding the procedures that the group used to solve the problems. The remaining students (n = 14) did not try to understand the procedures but instead used the group's resources to "get the right answer." Whether students tried to understand the procedures strongly predicted whether they would solve the problem correctly on the individual test. Among students who used the group's resources to try to understand the procedures, the probability of obtaining a high score on the individual test was 0.71 (5 out of 7 students). Among students who used the group's resources only to obtain the correct answer to the problems, the probability of obtaining a high score on the individual test was zero (0 out of 14 students).

Table 6

Excerpts from Group Work for Students Who Made Errors

Problem: 30-minute call, $0.22 first minute, $0.13 each additional minute.

EXAMPLE 1: STUDENT WITH HIGH SCORE ON INDIVIDUAL TEST (STUDENT IS TOLD PROCEDURES AND SEEMS TO UNDERSTAND THEM)

Problem: 30-minute call, $0.22 first minute, $0.13 each additional minute.

A She spoke for 30 minutes...We are going to [multiply] 13 [and] 30.

B You always minus one minute from the phone call.

A Look, 'cause you have to times 30 to this. 'Cause she spoke [for 30 minutes]...so we are going to put 30 right here.

C 29. Because you to take away a minute.

A So, it's 13 [for each additional minute].

EXAMPLE 2: STUDENTS WITH LOW SCORES ON INDIVIDUAL TEST (STUDENT IS TOLD PROCEDURES BUT DOES NOT UNDERSTAND THEM)

C You times it [13 cents] by 30. [wrong]

B 29.

C No, 30!

B [In a previous problem], there was 8 minutes so you subtract one and put 7. Here, it was 3 so you subtract and put 2. This is 30, so you minus 1 and use 29. See?

C If we get it wrong, it's your fault...I don't understand this.

B So? It's too bad!

C $3.77 plus 22 cents is $3.99.

Table 7

Excerpt From Group Work Showing Copying

D I don't know the homework thing, so you do your homework, OK? I don't want to do it.

C Why not?

D I don't know how.

C It's the same thing [as the problems we were doing today].

D I don't even know what we were doing right here. I was copying you guys.

Did not contribute to group discussion. The remaining students ($n = 7$, 13% of the sample) did not contribute to group discussion. Although the observers noted that some of these students seemed to be working independently, it was unclear from the audiotapes whether other students were listening to the group's discussion or were merely copying other students' work. Videotapes would help clarify these students' behavior, as would more detailed commentary from the observers. The low individual test scores of most of the "quiet" students, however, suggests that they did not understand how to solve the problems and did not, or could not, use the group's discussion to learn how to solve them by themselves.

Student characteristics predicting behavior. Behavior in the group was related to ethnic background and ability level but not gender. As can be seen in Table 8, Anglo students and higher-ability students were heavily represented among those who solved the problems correctly without assistance: 75% of Anglo students solved the problems correctly without assistance compared to only 39% of Hispanic/African-American students; 74% of high-ability students did so compared to 50% of medium-ability students and 13% of low-ability students.

Conversely, Hispanic/African-American and lower-ability students dominated the behavior categories showing a need for assistance. Combining the three categories of making errors, indicating confusion, and copying as showing a need for help, only 17% of Anglo students fell in this category compared to 46% of Hispanic/African-American students; only 21% of high-ability students fell in this category compared to 44% of medium-ability students and 56% of low-ability students.

To test the significance of the relationships between needing help from the group and ethnic background and ability level, two categories of need were compared: did not need help (solved problem correctly without assistance) and needed help (making errors, indicating confusion, copying). Ethnic background was marginally related to need for help (chi-square(1) = 3.06, $p < .09$). Ability level was significantly related to need for help (chi-square(2) = 9.80, $p < .01$; difference between mean ability scores of the two categories was also statistically significant, $t = 3.41$, $p < .01$).

Table 8

Distribution of Behavior of Students from Different Ethnic Backgrounds and Ability Levels

	Ethnic Background		Ability Level		
Behavior Category	Anglo (n= 12)	Hispanic/ African-American (n = 41)	High (n = 19)	Medium (n = 18)	Low (n = 16)
Increase or no change	10[a] (83%)[b]	17 (42%)	13 (68%)	10 (56%)	4 (25%)
Solved problems correctly	8 (67%)	12 (29%)	11 (58%)	8 (44%)	1 (6%)
Made errors, was corrected and told procedures	0 (0%)	1 (2%)	0 (0%)	0 (0%)	1 (6%)
Did not understand, was told procedures	1 (8%)	3 (7%)	2 (11%)	1 (6%)	1 (6%)
Did not contribute to group discussion	1 (8%)	1 (2%)	0 (0%)	1 (6%)	1 (6%)
Decrease	2 (17%)	24 (59%)	6 (32%)	8 (44%)	12 (75%)
Solved problems correctly	1 (8%)	4 (10%)	3 (16%)	1 (6%)	1 (6%)
Made errors, was corrected and told procedures	0 (0%)	3 (7%)	2 (11%)	1 (6%)	0 (0%)
Did not understand, was told procedures	1 (8%)	6 (15%)	0 (0%)	3 (17%)	4 (25%)
Copied others' work	0 (0%)	6 (15%)	0 (0%)	3 (17%)	3 (19%)
Did not contribute to group discussion	0 (0%)	5 (12%)	1 (5%)	0 (0%)	4 (25%)

[a] Number of students. [b] Percent of ethnic background or ability subgroup.

Whether students used the group's resources to try to understand the procedures, compared to merely using them to obtain the correct answer, was not significantly related to ethnic background or ability level. Nearly all of the students who used the group's resources were Hispanic or African-American. There was no significant difference between the proportion of each ethnic group who tried to learn the procedures and who did not try to learn. Similarly, all ability levels were equally represented in both categories.

Discussion

The results of this study showed that students' performance in group collaboration overestimated the ability of many of them to solve the problems individually. Furthermore, the data on group processes showed why scores in the group setting often exceeded those in the individual setting. Many students used the resources of the group to get the right answer but not to learn the procedures for solving the problems. They copied other students' work or used the procedures for solving the problems that other students provided. They were not actively engaged in constructing solutions to problems, but were merely using the work that other students had done. When faced with the problem on the individual test, they could not solve it. Achievement scores from the group setting, then, were not a valid indicator of these students' individual competence.

The group collaboration studied here was in the context of classroom instruction, not formal group testing. So an important question is how well the results found here would generalize to a formal group assessment context. Specifically, does the large discrepancy between group and individual performance found in this study overestimate what would be found in other assessment contexts? Several features of the group context in this study may lead toward that conclusion. First, students had some practice in working in small groups (for approximately two weeks) prior to the study. Second, they had received instruction in basic communication skills to help prepare them for working in groups. Both of these factors should have facilitated group functioning. Groups without previous experience in collaboration may spend more time negotiating how to work with others and less time on the academic task at hand, and may obtain lower scores as a result. This hypothesis remains to be tested, however. Third, the use of heterogeneous groups may have maximized group performance. All groups had at least one student who was able to solve the problems, providing resources for students who could not. If, in contrast, some groups were formed without any student who could solve the problems, students' performance in those groups may have been poor, lowering the mean score over all groups.

On the other hand, a number of features of the present study may have caused it to underestimate the discrepancy in group and individual performance that may occur in other settings. First, the teacher weighted the grade for group work less than the grade on the individual test, so students working in groups may have been less motivated to work hard than they were when tested individually. If the grade for group work had received the same weight as the individual test, scores in the group setting may have been even higher, and the resulting disparity between group and individual settings also greater. Second, by the time of the individual test, students had received an additional two weeks of instruction plus a review session, which should have helped boost their individual test scores compared to their scores in the group. Third, the emphasis in the current study placed on using the group to learn how to solve the problems rather than only to supply solutions should also have boosted individual scores. This was not a major factor, however, as few students used the resources of the group to learn how to solve the problems. In this respect, the group collaboration in the present study probably represented the mindset of students in typical group assessment situations quite well: to work jointly toward solutions to problems rather than toward greater understanding of how to solve problems. On balance, the conditions of the present study may have produced a reasonable estimate of the discrepancy between student performance in group and individual assessment that would occur in formal assessment settings.

Other questions emerging from the results of this study concern the possibility and desirability of modifying the group assessment context to produce achievement scores that are more valid indicators of individual student competence. If one believes that individual competence is best measured by students working individually, then a question to be explored is whether limiting the nature or extent of collaboration in group assessment would produce achievement scores that more accurately represent students' individual competence. The group context used in this study allowed students to collaborate when discussing how to solve problems and when writing their solutions. This represents the practice in a number of testing programs that give students opportunities for continuous collaboration (e.g., Shavelson & Baxter, 1992). Other testing programs, however, allow collaboration on some aspects of the task (e.g., discussion of a piece of literature) but not others (e.g., writing responses to questions based on that discussion, Connecticut Board of Education, 1992). One hypothesis to be tested is that the more extensive the opportunities for collaboration, the less valid will be the achievement scores for drawing inferences about individual students. In the extreme case, it is possible that any amount of collaboration will preclude informative scores about individual competence when interest lies in what students can accomplish individually.

From the perspective of social construction of knowledge and understanding, however, individual competence can be demonstrated when students work together to construct solutions to problems. Through collaboration with others, students can learn how to solve problems that they could not initially solve on their own. The issue here is not limiting the collaboration in groups, but instead is ensuring that students are actively engaged in learning how to solve problems rather than copying others' work or being told the procedures to use.

Limiting collaboration among students would also be counterproductive for other purposes of group assessment described at the outset of this paper, such as modeling assessment on collaborative instructional practices, measuring the productivity of students when working in groups, and measuring students' collaboration skills and problem-solving processes. Fulfilling these goals probably requires more, not less, collaboration.

One solution to the problem of obtaining valid information about group and individual performance lies in examining group processes. This study showed that, in contrast to the scores from group assessment, information about the *processes* taking place during group collaboration can provide important and accurate information about individual students' competence and behavior. The analyses of group processes in the present study shed considerable light on the understanding of individual students, as well as on their behavior in group collaboration. Collecting data on the dynamics of collaborative groups and students' behavior, then, may help group assessment efforts to meet multiple goals: measuring group productivity and students' collaboration skills and drawing inferences about the capabilities of individual students.

The implication of the results of this study for assessment practice is that scores from group assessment may not be valid indicators of many students' individual competence. Without data on group processes, it will be difficult or impossible to distinguish among students who solve the problems with little or no assistance from the group, students who learn how to solve the problems by working collaboratively with others, and students who use resources in the group to obtain the correct solution (by copying others' work or being told what to do) without learning how to solve the problem. All of these students will obtain high scores in group assessment, but not all of them will be competent.

In conclusion, group collaboration may have an important place in future assessment practices, but scores on work submitted from group assessment should not be used to make inferences about the competence of individual students. Without data on group processes, scores from group assessment are better interpreted as what students can produce when working with others.

References

California Department of Education. (1985). *Mathematics framework for California public schools, kindergarten through grade twelve.* Sacramento: California Department of Education.

California Department of Education. (1992). *Mathematics framework for California public schools, kindergarten through grade twelve.* Sacramento: California Department of Education.

Connecticut Board of Education. (1992). *Connecticut Academic Performance Test.* Hartford, CT: Connecticut Board of Education.

Damon, W., & Phelps, E. (1989). Critical distinctions among three methods of peer education. *International Journal of Educational Research, 13*, 9-19.

Farivar, S., & Webb, N. M. (1991). *Helping behavior activities handbook.* Los Angeles: University of California, Graduate School of Education.

Hackman, J. R. (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork.* San Francisco: Jossey-Bass.

Hare, A. P. (1992). *Groups, teams, and social interaction: Theories and applications.* New York: Praeger.

Johnson, D. W., Johnson, R. T., & Skon, L. (1979). Student achievement on different types of tasks under cooperative, competitive, and individualistic conditions. *Contemporary Educational Psychology, 4,* 99-106.

Kahan, J. P., Webb, N., Shavelson, R. J., & Stolzenberg, R. M. (1985, February). *Individual characteristics and unit performance* (Report R-3194-MIL). Santa Monica, CA: The RAND Corporation.

Knight, G. P., & Bohlmeyer, E. M. (1990). Cooperative learning and achievement: Methods for assessing causal mechanisms. In S. Sharan (Ed.), *Cooperative learning: Theory and research.* New York: Praeger.

Lomask, M., Baron, J., Greigh, J., & Harrison, C. (1992, March). *ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science.* A symposium presented at the annual meeting of the National Association of Research in Science Teaching, Cambridge, MA.

Meyer, C. A. (1992). What's the difference between *authentic* and *performance* assessment? *Educational Leadership, 49*, 39-40.

Pandey, T. (1991). *A sampler of mathematics assessment.* Sacramento: California Department of Education.

Redding, N. (1992). Assessing the big outcomes. *Educational Leadership, 49*, 49-53.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, . I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: ABLEX.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49*, 20-25.

Shavelson, R. J., Webb, N. M., Stas, C., & McArthur, D. (1988). Teaching mathematical problem solving: Insights from teachers and tutors. In R. I. Charles & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 203-231). Hillsdale, NJ: Erlbaum.

Slavin, R. E. (1990). *Cooperative learning: Theory, research, and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Tudge, J., & Rogoff, B. (1989). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In M. H. Bornstein & J. S. Bruner (Eds.), *Interaction in human development* (pp. 17-40). Hillsdale, NJ: Lawrence Erlbaum.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13*, 21-40.

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