

Blending Individual and Group Assessment: A Model for Measuring Student Performance

Elana Reiser
St. Joseph's College
ereiser@sjcny.edu

Abstract: Two sections of a college discrete mathematics class were taught using cooperative learning techniques throughout the semester. The 33 students attending these sections were randomly assigned into groups of three. Their final examination consisted of an individual and group blended examination where students worked in their groups and discussed questions, but each student had a different version of the examination and their own questions to answer. Afterwards, students were given a questionnaire and one group from each section was interviewed to see how they rated blended assessment. The majority of students preferred the blended examination over an individual examination, however not every student felt this way. One benefit of the blended examination was being able to verbalize test questions, which led to a better understanding of them. Concerns about the blended examination such as time, fairness, and difficulty concentrating were raised.

Keywords: alternative assessment; cooperative learning; cooperative assessment; groups

Cooperative learning is a popular teaching method. Roger Johnson, David Johnson, and Edythe Holubec, prominent figures in the field of cooperative learning, wrote, "one of social psychology's great success stories is the widespread use of cooperative learning" (1998, p. 9). Organizations such as The National Council of Teachers of Mathematics (NCTM) recommend that teachers use cooperative learning techniques for mathematics instruction. In fact, the 1989 Curriculum and Evaluation Standards written by NCTM suggest using a variety of instructional methods in classrooms, including small-group work and peer instruction.

There are many reasons why a teacher uses cooperative learning. Group work has been shown to be a very effective method of learning. Kagan & Kagan (1994) suggest that the three most important benefits of cooperative learning are: (1) academic gains, (2) improved race relations, and (3) improved social and affective development. Cooperative learning also allows students to monitor their progress through group processing. Authors of the NCTM *Principles and Standards for School Mathematics* (2000) assert that "students learn more and learn better when they take control of their learning by defining their goals and monitoring their progress" (p. 15).

The method of assessment in a classroom should match the manner of instruction (Boe, 1996). Webb et al. (1995) suggest that a major reason for using group work in assessment is to link assessment more closely to group work used in classroom instruction. Boud, Cohen, and Sampson (1999) argue that "if assessment gives students the message that only individual

achievement is valued, and that collaborative effort is akin to cheating, then the potential of peer learning will not be realized" (p. 413). Not using cooperative assessment when it is appropriate (i.e. when using cooperative learning) may have detrimental effects on student learning (Boud et al., 1999). Boud et al. (1999) also state that "inappropriate forms of assessment appear to encourage students to take a surface approach to learning, that is, they emphasize rote learning, conforming to the narrowest interpretations of assessment tasks, and working to beat the system rather than engage in meaningful learning" (p. 413). Therefore, it is crucial to match the assessment to the teaching methodology. When students are taught individually, individual testing makes sense, but when students are taught in groups they should be tested using group assessment.

Group assessment is an assessment style that teachers can use with cooperative learning (Klecker, 2002; Griffin, 1994). Group assessment is a testing framework in which group members work together on a test or project and the group as a whole is given one grade. Students work together and combine their knowledge to complete the assessed task. This method of assessment does have the benefit of enhancing students' social skills and teaching them to work cooperatively. In one study chiropractic students working on a group assessment said that it helped improve their critical thinking and boost their confidence (Meseke, Bovee, & Gran, 2009). Group assessment also provides the benefit of serving as a learning experience when students work collaboratively to debate and discuss the test questions (Zipp, 2007). However, group work and assessment have their own problems. The principal failure of group assessment is that it sometimes fails to measure individual achievement. Webb (1993) found that it was unclear if the group score was representative of the score that each individual in that group would have received if they took the assessment individually. In her study, Webb looked at the scores of students in a group test and then in an individual test and found that "students working in cooperative groups performed at uniformly high levels, so it was not possible to predict their performance in the individual setting from their performance in the group setting" (Webb, 1993, p. 145). Ewald (2005) agrees that a group assessment is not a valid or reliable predictor of individual scores. Meseke, Nafziger, and Meseke (2010) gave group quizzes in a neuroanatomy course and found that while students taking the quizzes collaboratively performed better than the control group, there was no significant difference between the two groups on the individually administered final examination. This is further proof that group assessments fail to measure individual achievement.

Neither individual nor group assessment is adequate to assess fully a student's ability, particularly if cooperative learning is used in the classroom. An alternate form of assessment is needed.

Purpose of the study

The purpose of this study is twofold. The first goal is to develop an assessment alternative for teachers who use cooperative learning. This means of assessment should accurately measure students' individual mathematical achievement. The second goal of this study is to see what students think of this new form of assessment.

One way to determine whether a student has learned a subject is to combine group assessment with individual assessment. The term "blending" is used to describe this mixture of the individual and group assessment techniques in which students work in groups during the

assessment, but each student is given a different, parallel version of the examination. In blended assessment students can discuss test questions, but each student is responsible for writing and handing in his or her own answer.

Procedures of the Study

The target population for this study consisted of college-level students taking a discrete mathematics course. There were two sections of this course with each section consisting of approximately 16 students.

The course's final examination was used to test the stated hypothesis. Students were placed into appropriate groups, based on previous research, which they stayed in throughout the semester. One aspect of this research was regarding heterogeneous grouping. The composition of groups should be heterogeneous in terms of gender, race, and ability (Mulryan, 1995; Webb, 1997). When groups are maximally heterogeneous, students "tend to become tolerant of diverse viewpoints, to consider others' thoughts and feelings in depth, and seek more support and clarification of others' positions" (Stahl, 1994, p. 3). One way that heterogeneous grouping tends to occur is when students are randomly assigned into groups (Johnson et al., 1994).

Another area the research focused on was group size. In particular, groups should be small since cooperative groups are most effective when there are two to four students per group (Gatto & Daniels, 2005). Having too many group members does not give each student enough opportunity to interact with other students while having too few students does not sufficiently allow for diverse opinions within a group (Johnson et al., 1994). A group of three seemed best for this study as it allowed diversity as well as time for students to actively participate. The students were randomly put into groups of three with at least one female and at least one male. The group members were allotted time to practice working together within their group.

There were various ways that this course integrated cooperative learning techniques with cooperative assessment. The discrete math class met three times a week. After new material was presented, the students met in their groups. The students received questions based on the new material. Only one set of questions was given to each group in order to encourage task interdependence. In this way, students could not work by themselves to answer the questions; the members of the group would have to depend on each other at least to see the questions. After groups were given enough time to complete the questions, students were randomly selected to put up the solution on the board. One student from each group was chosen to share their solution. This was done to ensure individual accountability. If the solution was correct then the student's team got one bonus point. Each student was given the same amount of turns for presenting a solution.

A similar method was used for the quizzes. Students took each quiz individually. This also was done to ensure individual accountability. The first quiz set each student's quiz average. On each subsequent quiz, if every group member got his or her average score or better, the group got a bonus point. For example, three students in a group got a 73, 86, and 89 on their first quiz. If they then got a 74, 87, and 91 on the next quiz then this group got a bonus point. This was done to ensure positive interdependence. It made students care about and feel responsible for other group member's learning. The quizzes were still taken individually, but each group member would have a vested interest in helping the other group members achieve high quiz scores. Any bonus points that the group accumulated were then added to their examination scores.

Other cooperative learning structures were also used at various points in the class. These were adapted from Kagan's book *Cooperative Learning* (1985). For example, on the first day of school the students participated in a classbuilding activity called "Roam the Room." Kagan (1994) argues that although taking time for class and teambuilding activities may seem to be off-task, such activities actually create a social context that will then make the classroom more conducive to learning. A ditto with ten mathematics trivia questions was given to each student. The students completed what they could by themselves in the first ten minutes. The students then roamed the room looking to get all of the trivia questions answered by talking with other students. After fifteen minutes one student was randomly called on to answer the trivia questions. If the chosen student got eight out of the ten correct then each student in the class was given a small prize. Randomly calling on a student was done to ensure individual accountability. This way, each student would feel that he or she should have all of the questions answered. Roaming the room was done to promote interdependence. If one student did not know the answer to a question he or she could find another student who did. This activity was useful for several reasons. It was mainly used because it was the first day of school and the students used this opportunity to talk to their classmates and get to know each other. It was also used to get the students thinking about mathematics since they had just returned from summer vacation.

During the final examination each of the three group members received a different, parallel version of the test. In each group the students were tested on the same concepts but each student had different questions to answer. Students were able to discuss the test questions and receive all of the benefits of working in a group, while still being individually accountable for the material.

The best way to ensure that the three parallel tests closely resembled each other was to make a list of all of the topics that would be on the test. The two main topics were logic and proof writing. Each logic section of the textbook used in this class (Epp, 2004) was looked through and a list of the types of questions to put on the test was compiled. These included logical arguments, truth values, conditional statements, negations, and proving that two statements are logically equivalent. The same method was used to get questions for the sections on proof techniques. The topics chosen were proof by contradiction, proof by contraposition, induction, proofs with the ceiling or floor function, parity proofs, rational proofs, and divisibility proofs. Some miscellaneous topics, such as the mod and div functions and converting from decimal to binary notation, were also included. Once the list of the topics to be included on the test was completed, three equivalent questions for each topic were constructed. Some of these questions were taken from an online test bank provided by the publisher of the course's textbook. Other questions were taken from additional discrete mathematics textbooks. The rest of the questions were either modified from one of these sources or written by the researcher. A Java program was created in order to randomly assign one of each of the three questions in each topic to one of the three versions of the test. This program listed the numbers one through three in a random order. The first question written down was assigned to the test number that the program gave first, and so on.

Results

The descriptive statistics for the individual and blended scores are shown in Table 1.

Table 1. Descriptive statistics for blended and individual scores

	Blended Score	Individual Score
mean	84.00	81.27
median	84.00	83.00
sd	9.37	10.42
r	0.63	

An Analysis of Variance (ANOVA) was performed to determine if there were any statistical differences in the score that a student received on the blended part versus the score that same student received on the subsequent individual part. The observed F value (3.58) was less than the critical F value (4.12), therefore the results are not statistically significant ($\alpha = .05$). This means that there is no statistical difference between the scores on the blended examination and the scores on the individual examination. After a student has taken this blended examination, that student was found to be individually accountable for knowing the material that the examination covered. This provides evidence for the thesis that when classes are taught using cooperative learning, blended assessment provides students with all of the benefits of working in a group while still being held accountable as individuals. It was found that the score that a student received on the blended examination was similar to the score that the student received on an individual examination that took place after the learning experience of the blended examination. This was further reinforced by looking at the correlation ($r = .63$) between the two parts of the examination. This is a fairly strong correlation.

Since one of the goals of this research was to determine if blended assessment accurately predicted individual assessment. This study was structured with the blended assessment occurring prior to the individual assessment. Since the individual scores were statistically equivalent to the blended scores, the conclusion that blended assessment accurately tests learning is supported.

Students' Feelings about Blended Assessment

A questionnaire was given to all students after they had taken the examination. The questions were developed using a five point Likert scale. The students were asked to rate each of the following statements as strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree. The last question was open-ended and asked if students preferred the individual or blended format better and why they felt as they did. Table 2 summarizes the statistics for each question.

Table 2. Descriptive statistics for student questionnaire

	mean	median	sd	n
Q1: In my opinion, each group member was prepared for the blended examination	3.79	4	.35	33
Q2: I discussed my test problems with other group members	4	4	.35	33

	mean	median	sd	n
Q3: I studied for the blended examination the same amount I would have for an individual examination	3.71	4	.33	33
Q4: I got a higher score on the blended examination than I would have on an individual examination	2.97	3	.31	33
Q5: Taking a blended examination helped me understand the material better	3.38	4	.33	33
Q6: Taking a blended examination helped the other members of my group understand the material better	3.85	4	0.35	33
Q7: I would prefer to take an examination blended or individual? Why do you have this preference?	19 B	10 I	4 Mixed	

In general, the results showed an agreement with each statement, other than question four, but there weren't any that had a strong agreement. This result was somewhat surprising at first. It would seem that students should feel that working in groups would help them get a better grade (question 4). It may be possible to better understand these results by considering the order of assessment. When answering this survey, students had already taken a blended examination. After taking the blended examination, students felt they would score the same on a subsequent individual examination. This is consistent with the statistical analysis, which showed that there was no difference between the scores on these two parts. This is precisely because the blended portion served as a learning experience.

Question four was related to the last question, which asked if students preferred blended or individual exams better. The following table shows the answers to question four grouped by preference:

Table 3. Perceived learning by exam style preferences

Preference	1	2	3	4	5
Blended N=19	15.8%	5.3%	10.5%	31.6%	36.8%
Individual N=10	60%	0%	20%	10%	10%
Mixed N=4	0%	50%	50%	0%	0%

It is interesting to note that the students who preferred blended exams tended to think that they performed better on the blended part of their midterm examination. This is supported by the fact that the majority of the question four answers for students who preferred blended examinations were a 4 or 5. The majority of the question four answers for the students who preferred individual examinations were a 1. This shows that students who preferred individual

examinations thought they performed better on the individual portion of their midterm. The students who had mixed feelings all answered question four with a 2 or 3 score. This means that they felt they scored about the same on the two parts of the midterm examination.

While, overall, students felt that they scored the same on the blended and individual examinations (question 4 mean = 2.97), it is notable that students felt that they helped their teammates understand the material better (question 6 mean = 3.85). Students may have had a self-serving bias, thinking that they helped their groupmates more than they received help.

Question two, "I discussed my test problems with other group members," was put on the questionnaire to determine if group discussions were actually taking place during the blended part. It is possible that students would choose to work alone, even though the group members were allowed to discuss the test questions. One way to discourage isolated work was for students to take the examination in the same groups that they had been in all semester. In this way, students would feel comfortable with their groupmates and would be more likely to engage in discussion. The median on this question was 4, which indicates that group discussions did in fact take place. Additionally, while the blended part was occurring, the instructor noticed that the groups were discussing all or most of the test questions.

Students felt that they took responsibility for knowing the material on their own and that other group members also took on that same responsibility. This was shown in the students' answers to question three, "I studied for the blended examination the same amount I would have for an individual examination" and their answers to question one, "in my opinion, each group member was prepared for the blended examination." The median score on both of these questions was 4.

Qualitative Survey Findings

There are various reasons why students might have preferred an individual or blended examination. The results of the open-ended question gave rise to several common themes. The students who preferred a blended examination liked the interaction with other students. For example, students said that the blended exam offered them a second opinion as well as help with examination questions. The fact that group members were there to answer any questions boosted their confidence, but, at the same time, the students still felt responsible for their own learning and did not rely entirely on their group members to explain all of the concepts.

Students also thought that discussing the questions with other people allowed them to understand the material better. By explaining answers to others, they felt it helped reinforce the material and organize and articulate their thoughts. Students stated that this reinforcement led to a deeper understanding. It is interesting that the students noted this because previous research has concluded this same notion (Webb et al., 1995).

Other students liked that they got to socialize and be around other people while they were taking the blended examination. Most testing in schools is done individually, so when students get to socialize while being evaluated, it may help them to relax and this could help some students with test anxiety (Meseke, Bovee, & Gran, 2009).

Students who preferred to take an examination individually stated that learning should be a personal responsibility and that they did not feel comfortable placing that responsibility on the shoulders of others in their group. Additionally, they did not think it was fair for them to be forced to take on the responsibility of other people's learning. This idea of feeling responsible for your group member's learning is an integral component of cooperative learning that is hard to

instill. The students had been participating in cooperative learning for only one semester. They may not have preferred cooperative learning because it is hard to undo thirteen or fourteen years of traditional schooling where the emphasis has been placed on individual responsibility. Additionally, there are some people who feel less comfortable working in a group setting. It therefore makes sense that some students felt responsible for their own learning and uncomfortable in helping or being helped during an examination.

Students also felt that the noise made by other students during the blended part interfered with their ability to concentrate. Lastly, they felt that the blended part took too long. To that effect, the results of studies have shown that students working in groups take longer to answer questions because of the discussions that take place (Beebe & Masterson, 1994; Stearns, 1996). There was no time limit on the examination. However, even though students were given unlimited time, the longer they stayed, the more mentally exhausting the test became. This exhaustion could have affected their feelings towards the examination and their ability to perform.

Student Interviews

One group from each of the two sections of discrete mathematics was chosen to be interviewed to further explore how they felt about cooperative learning and the blended examination they completed. These particular groups were selected by the instructor because they worked well together. The interviews were conducted with the whole group together. The interviews were recorded on tape and transcribed.

The first set of questions was about cooperative learning. The students were familiar with both the term “cooperative learning” and the elements defining cooperative learning because class time had been spent on going over these concepts. All six students that were interviewed liked using cooperative learning in the classroom. Students felt that it sometimes helped to have peers explaining the material because other students could sometimes explain it in simpler terms. As one student explained, “sometimes it’s easier when you have one of us explaining it. Not that the teachers doing it wrong, it’s just easier sometimes when someone on your level can explain it easier in simpler terms.” They also thought that working in groups was helpful because if they had trouble getting started on a question, they had other people to discuss the question with. This is the same concept as peer assistance in cooperative learning terms, the feeling that the members of the group should help each other learn. The students also felt that cooperative learning better prepared them for quizzes and tests because they did not want to let the group down if they were randomly called up to the board to earn bonus points. It “encourages you to try harder to understand the problem and try to understand it enough to go up to the board because you don’t want to let your group down.” This captures the idea of positive interdependence - in order for one student to succeed all group members have to succeed. Another student felt that the randomization of calling students up to the board made all of the group members contribute. It “encourages you to try harder to understand the problem and try to understand it enough to go up to the board because you don’t want to let your group down.” This supports the notion of individual accountability because students felt responsible for their own learning. Since all group members did the work and discussions took place within the group, it also helped to ensure equal opportunity for students to participate in the group work. Additionally, students liked that once the solutions were put on the board, the answers were then checked with a whole class discussion. This allowed all students to review the answer and offered the possibility of seeing a

new way of solving the problem. It also allowed them to see common mistakes, which might help them to avoid the same mistakes in the future. Checking answers and any group discussion that followed are part of the group processing component of cooperative learning.

It was interesting that the students in these interviews brought up all of the elements of cooperative learning (positive interdependence, individual accountability, peer assistance, equal opportunity, and group processing). Through their answers, it was verified that the class was, in fact, using cooperative learning rather than just having students placed in groups.

The second set of questions in the interview was about the blended examination that the students took. The students felt that the blended examination had some strengths and some weaknesses. One strength was that the students had the opportunity to discuss the test questions with their group members, and this led to a better understanding of the questions. Other students felt that it helped them to understand the material better by explaining the answers to others and by talking out loud. Students noted that “if you understand something and you can explain it to somebody else you're only learning it better.” Students felt that the real reason that helping others was useful was because it was a form of repetition and it gave them a chance to reflect on their own knowledge. These ideas were also expressed in the student questionnaires and are held by researchers (Webb et al., 1995; Slavin, 1984). Another student thought that the blended part served as a review for the individual part. This notion of assessment serving as an episode of learning is also held by many researchers (Shavelson & Baxter, 1992; Shepard, 2001; Wiggins, 1989).

The students felt that the blended examination had some limitations. They felt pressured for time. Even though the students had unlimited time for the examination, they may have had outside time pressures or other commitments after the scheduled class time was over. One student suggested that “you definitely need more time when you're in a group session or less problems.” Some students also had trouble concentrating during the blended part. However, this would depend on the specific students in each class. One student commented that the noise in the background actually helped him to concentrate better because working in a completely silent environment made him more likely to daydream and lose his train of thought.

Based on the student questionnaires and the student interviews, it seems that, overall, students enjoyed taking the blended examination. One student summed it up nicely by saying, “we're helping each other out to learn better.” They did, however, see some limitations to it.

Conclusion

Using blended assessment may be a way for teachers who use cooperative learning to align their assessment method with their teaching method while maintaining individual accountability. This study has provided evidence for the fact that the individual student is held accountable in a blended assessment. Blended assessment overcomes the principal weakness of group assessment, that the individual student is not held accountable, and, therefore, provides a genuine alternative to individual assessment. Teachers who were once reluctant to use group assessment because they thought that individuals were not held accountable for the material can now consider blended assessment as an alternative to individual assessment.

This study pointed out some of the strengths and weaknesses of blended assessment. One of the strengths of blended assessment is that students are able to talk to other group members and get help on the test questions, which often leads to discussion. Regardless of whether the student is the speaker or the listener, these discussions can lead to a deeper understanding of the

material. Because these discussions are taking place, the assessment is acting as an episode of learning. “An ideal assessment would be a good teaching activity and, indeed, might even serve as a teaching activity when not used for assessment” (Shavelson & Baxter, 1992, p.20). Through the student surveys it was seen that students agreed that they learned from each other, even though they seemed to think that they helped others more than they received help. To help students realize this self-serving bias, instructors can give students an individual examination followed by a blended examination. This test sequence can lead to students scoring significantly better on the blended portion. By seeing their grade increase on the blended portion, students should realize that it was actually helpful for them (and not just their groupmates) to discuss the questions as they took the assessment.

Using blended assessment also allows the teacher to write tests at a higher level. Vygotsky's (1978) experiments with children's zone of proximal development showed that children can achieve more by working with others than by working alone.

One of the weaknesses of blended assessment found in this study is that students had trouble concentrating while taking the blended examination. The blended examination also took longer than an individual examination and this led to time issues such as fatigue.

Recommendations

A sample size of thirty-three students was used in this study. This is an insufficient number to generalize to a larger population. A larger sample of students needs to be tested to see if the same results would occur. Additionally, different populations need to be tested. The students used in this study were all college students at a private school in the suburbs of a large city. Similar studies should be done with different grade levels, in different areas of the country, and in public colleges and universities to see if the individual is held accountable in these different situations. Studies could be done to see if blending has different effects for different populations. Variations on this study could test blended assessment with different genders, racial groups, grade levels, and school subjects.

A similar study could be done where groups are further spread apart in the classroom or groups take tests in different rooms to help with concentration problems. Another follow-up study could explore the issue of time constraints. The issue of time could have been resolved by shortening the three blended tests.

References

Beebe, S.A., & Masterson, J.T. (1994). *Communicating in small groups: Principles and practices*. New York: HarperCollins College Publishers.

Boe, B. (1996, August). A democratic assessment strategy. Paper presented at the Annual Meeting of the Association of Teacher Educators, Tarpon Springs, FL.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning assessment. *Assessment and Evaluation in Higher Education*, 24(4), 413-426. <http://dx.doi.org/10.1080/0260293990240405>

Epp, S.S. (2004). *Discrete mathematics with applications*. Belmont, CA: Brooks/Cole.

Ewald, J.D. (2005). Language-related episodes in an assessment context: A small-group quiz. *The Canadian Modern Language Review*, 61(4), 565-586.

<http://dx.doi.org/10.3138/cmlr.61.4.565>

Gatto, M., & Daniels, E. (2005). *Cooperative learning in your classroom: A handbook for teachers*. New York: SUNY Stony Brook.

Griffin, M.M. (1994, April). Learning through testing: An investigation of cooperative assessment. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Johnson, R.T., & Johnson, D.W. (1994). An overview of cooperative learning. In J. Thousand, A. Villa & A. Nevin (Eds.), *Creativity and collaborative learning* (pp. 31-44). Baltimore, MD: Brookes Press.

Johnson, D., Johnson, R., & Holubec, E. (1998). *Cooperation in the classroom*. Boston: Allyn and Bacon.

Kagan, S. & Kagan, M. (1985). *Cooperative learning*. San Clemente, CA: Kagan Publishing.

Kagan, S. & Kagan, M. (1994). The structural approach: Six keys to cooperative learning. In S. Sharan (Ed.), *Handbook of cooperative learning* (pp. 115-136). Westport, CT: Greenwood Press.

Klecker, B. (2002, October). Formative classroom assessment using cooperative groups: Vygotsky and random assignment. Paper presented at the annual meeting of Educational Psychology, Oxford, OH.

Meseke, C.A., Bovee, M.L., & Gran, D.F. (2009). Impact of collaborative testing on student performance and satisfaction in a chiropractic science course. *Journal of Manipulative and Physiological Therapeutics*, 32(4), 309-14.

Meseke, C.A., Nafziger, R., & Meseke, J.K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *The Journal of Chiropractic Education*, 24(1), 19- 29.

Mulryan, C.M. (1995). Fifth and sixth graders' involvement and participation in cooperative small groups in mathematics. *The Elementary School Journal*, 95(4), 297-310.

<https://doi.org/10.1086/461804>

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: The Council.

National Council of Teachers of Mathematics. (2000). *Principals and standards for school mathematics*. Reston, VA: The Council.

Shavelson, R.J., & Baxter, G.P. (1992). What we've learned about assessing hands-on science. *Educational Leadership*, 49(8), 20-25.

Shepard, L.A. (2001). The role of classroom assessment in teaching and learning, in V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). Washington DC: American Educational Research Association.

Slavin, R.E. (1984). Students motivating students to excel: Cooperative incentives, cooperative tasks, and student achievement. *The Elementary School Journal*, 85(1), 53-63.
<https://doi.org/10.1086/461391>

Stahl, R.J. (1994). *The essential elements of cooperative learning in the classroom*. (ERIC Document Reproduction Service No. ED 370881).

Stearns, S.A. (1996). Collaborative exams as teaching tools. *College Teaching*, 44(3), 111-112.
<http://dx.doi.org/10.1080/87567555.1996.9925564>

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Webb, N.M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment*, 1(2), 131-152.
http://dx.doi.org/10.1207/s15326977ea0102_3

Webb, N.M. (1997). Assessing students in small collaborative groups. *Theory into Practice*, 36(4), 205-213. <http://dx.doi.org/10.1080/00405849709543770>

Webb, N.M., Nemer, K., Chizhik, A., & Sugrue, B. (1995, April). Using group collaboration as a window into students' cognitive processes. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 7033-7713. Retrieved from <http://www.jstor.org/stable/20404004>.

Zipp, J. F. (2007). Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology*, 35(1), 62-76.