International Journal of Teaching and Learning in Higher Education 2007, Volume 19, Number 3, 289-296
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Student Assessment: A Comparison of Solitary, Cooperative, and Competitive Testing

Marcie N. Desrochers, Herbert C. Fink, Andrea Thomas, Joe Kimmerling, and Wendy Tung
State University of New York, College at Brockport

Alternatives to a solitary testing format can involve students working in teams to arrive at the correct answer. We compared two group assessment methods, cooperative and competitive, to a solitary testing approach. In most comparisons examining the undergraduate respondents' (N=77) performance, the two group-testing methods were equivalent. Both group methods were superior to the solitary testing format in determining students' knowledge of the course material and confidence in their answers. Moreover, rather than one person generating the team's answer, most team members actively participated in the decision-making. The cooperative group surpassed the solitary testing group in terms of preference for their format of assessment.

Assessment empirically determines whether student learning or other educational outcomes, such as positive affect or change in values, have been attained. One of the most widely used assessment approaches in North American higher education is the objective test, most frequently a multiple choice exam, where students answer questions individually. Administration of this test format is a relatively quick and easy way to measure students' knowledge of a subject matter. Faculty, administrators, and students believe it to be a legitimate assessment device. As grade level increases, teachers are more likely to choose objective measures of student assessment and take steps to improve the quality of the assessment tool (Zhang & Burry-Stock, 2003).

Selection of multiple choice tests over other assessment methods like essay tests, are often based on reliability and validity criteria. Machine scored multiple choice tests have perfect scoring reliability while essay tests have problems with interrater reliability (Hogan, 2007; Johnson, Penny, & Gordon, 2001; Longford, 1994). Bridgeman and Lewis (1994) found that multiple choice tests used in College Board Advanced Placement exams in the United States were always equal or superior to essay tests in validity, predicting freshman grade point averages. Presumably freshman grades are a composite of different methods assessing knowledge and higher order thinking, such as objective items, essay tests, papers, and class participation.

Despite the extensive use of multiple choice tests in some parts of the world, their validity has been questioned. A student's test score when taking the test alone may be adversely affected by low student motivation, test anxiety (Mendl, 1999), cheating (Norton, Tilley, Newstead, & Franklyn-Stokes, 2001) and differences in question interpretation (Ingram & Nelson, 2006). Moreover, content knowledge may not reflect behavioral skills (McGimsey, Greene, & Lutzker, 1995). The learner's cognitive style (Lu & Suen, 1995) and preferences for different instructional styles (Birenbaum, 2007) may be negatively correlated with outcomes using this solitary testing approach.

One method that may overcome some of the shortcomings of assessing the individual student is cooperative testing. This approach involves a small group of students working together to arrive at a common solution to a problem. This situation is comparable to what often occurs in a workplace setting or on a sports team. Group members work together cooperatively to achieve a goal. Acquisition and generalization of important interpersonal skills (e.g., explanation, negotiation) are potential benefits of the cooperative assessment approach (Zimbardo, Butler, & Wolfe, 2003). Moreover students seem to enjoy working in teams more than solving a task alone (Hinsz & Nickell, 2004). This positive emotion associated with cooperative assessment may impact the task in a beneficial way. Additional interest in the topic being examined as well as an increase in time spent examining the issue may result from use of this approach.

Another benefit of cooperative testing is that there are consistently higher achievement scores when students work in groups compared to solitary testing (Cortright, Collins, Rodenbaugh, & Dicarlo, 2003; Rao, Collins, & DiCarlo, 2002). Jensen (1996) compared individual to cooperative testing and found that, on average, students improved their exam scores in the cooperative testing condition by seven percentage points. Jensen, Moore, and Hatch (2002) found that biology students performed significantly better on a cooperative portion of a class quiz compared to the individual portion.

In addition to improving student scores, students in the Mitchell and Melton (2003) study listed more advantages than disadvantages with cooperative testing. These advantages included immediate feedback, testing as a learning experience, better retention because of the discussion and interaction, opportunity to improve grades, and preference for team over individual testing.

Some of the general disadvantages to cooperative testing that students noted were changing their correct answer after team discussion, time constraints, and partners who did not fully participate. Zimbardo et al. (2003) also found benefits to cooperative testing compared to solitary testing such as (a) improved student performance, (b) decreased test anxiety, (c) increased enjoyment of the course, and (d) enhanced interpersonal skills such as deliberating with others.

Moreover, cooperative testing is a versatile technique that can be used during class time with positive results. Rao and DiCarlo (2000) paused after 15-20 minutes of class instruction and asked students to answer questions, first individually, and then in a team. Student answers to factual and conceptual questions were significantly better in a team compared to the individual condition.

A common complaint of team assessment approaches is that members ride on the coattails of the most knowledgeable team member. For instance, the answer may simply be provided by one member of the team with no explanation given to the less well-informed member. Contrary to this possibility, participants in the Zimbardo et al. (2003) study reported that the most frequent strategy used to decide a final answer was members fully discussing their differences of opinion.

Competitive testing is a third assessment method that may identify what students know about a content area. Like cooperative testing, competitive assessment involves a team but adds a competitive element between teams. Students cooperate within their team to determine their answer and then compete against other teams for the most correct and/or swiftest answer. Formats can include question and answer games, board games, or simulations of complex phenomenon. Student motivation may be enhanced and performance increased as a result of the competitive environment. On the other hand, students may feel pressured to respond quickly or hesitant to discuss answers, lessening the benefits that may otherwise be accrued when students interact in a competitive format. Little research has compared the effectiveness of competitive groups to other assessment methods.

In one of the few studies that examined competitive testing, Desrochers, Pusateri, and Fink (2007) compared a competitive game assessment to solitary testing (the student's answer alone). In this study, students' knowledge of course material in both conditions was measured using multiple choice items. Team scores were found to be superior to individual test scores. Also, team answers were more often correct than the initial answers of team members before group discussion. An analysis of team decision making styles showed that team answers were rarely the product of a single team member dictating the team answer (authoritarian). Instead of authoritarian rule, unanimity (everyone agreed on the decision) and majority rule (two of three team members agreed on the decision) were the basis of team decision-making. The participants in this study liked the competitive testing condition more and perceived it as a more accurate measure of their course knowledge compared to participants' ratings in the solitary testing condition.

In sum, some of the difficulties associated with using a solitary testing approach (e.g., low motivation, incorrect question interpretation) may be diminished with group testing (cooperative or competitive). Additional benefits may accrue from use of the group testing approach could be learning through peer interactions and more positive affect toward the course material.

Instructors may find a group testing approach one way to alleviate assessment concerns and more accurately measure student learning compared to individual testing methods. Not only could the group approach be used for summative course assessment, but also as a formative measure of students' knowledge of course material gathered during classroom review sessions designed to bolster student learning.

Our main purpose in the current study was to empirically examine students' knowledge of course material using three different methods: solitary, cooperative, and competitive testing. Additionally, we examined the following: (a) differences in students' confidence in their multiple choice answers in the three assessment conditions, and (b) how the three assessment formats compare as to students' affect (preference). Also, the manner in which team decisions were reached – unanimity, majority rule, or authoritarian rule – in the cooperative versus competitive conditions was investigated.

Method

Participants

Seventy-seven students from an introductory psychology course at a medium-size, liberal arts New York State college voluntarily participated. Participants received extra credit for participating in the study as is standard in American colleges. There were 24 students (15 women, 9 men) with an average age of 19 years in the solitary testing condition, all of whom were Caucasian. There were 27 students (16 females, 11 males) with an average age of 19 years in the cooperative testing condition; 93% were Caucasian and 7% African American. There were 27 students (19 women, 7 men, 1 undeclared) with an average age of 19 years in the competitive condition; 89% were Caucasian, 7% African American, and 4% other. Across conditions, most participants were in their first

(43%) or second year (45%) of college and majoring in psychology (21%) or Criminal Justice (14%). Fifty-seven percent of the students had grade point averages (GPA) between 2.6 and 3.5. Groups did not differ regarding age, year in college, or general grade point average ($p > .05$).

Procedure

Students were randomly assigned to one of three conditions: solitary, cooperative or competitive testing. Their knowledge of the course material in the unit covering learning and genetics was measured. Following the collection of demographic information, the same 16 multiple choice questions were presented in each condition. Eight items were factual (i.e., statements, definitions) and eight items were conceptual (i.e., application, synthesis, or integration of information). The study was held outside of class time the day before the exam on the content area tested.

Solitary testing condition. The solitary testing condition was held in a small classroom. The experimenter reviewed the procedure in this condition with the participants before the 16 questions were presented. The researcher orally and visually presented each of the 16 questions. After a question was read, students were informed that they had up to 20 seconds to individually answer and write down their answer on a sheet of paper. Then, each participant rated how confident they were that the answer was correct along a five-point scale labeled 1=*Extremely confident*, *Very*, *Moderately*, *Somewhat*, and 5=*Not at all confident.* Each participant was instructed to place their answer sheet in an envelope to prevent answers from being changed. Then, the researcher provided the correct answer and presented the next question and so on until all 16 items had been presented and answered.

Cooperative testing condition. In the cooperative testing condition, students were randomly assigned to a group of three members, introduced themselves to one another, and came up with a team name. There were three teams in each session, each working in a separate room with a different researcher.

The experimenter reviewed the procedure in this condition with the participants before the 16 questions were presented. The researcher orally and visually presented a multiple choice question. After the question was read, participants were informed that they had 20 seconds to answer the question alone, write down their answer, and rate their confidence in the correctness of the answer on a sheet of paper. After 20 seconds, participants were told to place their answer in an envelope to prevent them from changing it. The researcher then instructed the students that they had up to one minute to discuss amongst themselves which alternative is the correct answer. Each team member wrote down the team's answer, rated one's confidence in the correctness of the team answer, and placed the team answer in an envelope. Thus, each team member wrote an answer twice: the individual answer first and, then, the team's answer. Similarly, each group member's confidence in an answer was measured twice: first, confidence in one's own answer and, then, confidence in the group's answer. Again, the answer sheet was placed in an envelope to prevent changes to the student's answer. The researcher presented the correct answer and went onto the next question until all 16 items had been answered.

Competitive testing condition. The competitive testing condition was similar to the cooperative testing condition with the addition that all teams were in the same room and competed against one another to obtain points for correct answers. In the competitive testing condition students were randomly assigned to one of three groups of three members. In each session there were three teams sitting in different areas of the same room. Team members were instructed to introduce themselves and decide on a team name to foster group identity. The experimenter reviewed the procedure in this condition with the participants before the 16 questions were presented.

The researcher orally and visually presented a multiple choice question. Individually, students were informed that they had 20 seconds to write down their answer, rate their confidence in the accuracy of their answer on a sheet of paper, and place it in an envelope. Following individual answers, the experimenter announced that team discussion time was allowed for up to one minute to decide which answer to present. As soon as the group had an answer, they were asked to write it down on the answer sheet, complete the confidence rating for their group answer, place it in an envelope, and hit the light button. A member of each group hit the light button to signal when their group decided on a correct answer. The lit button allowed the experimenter to visually determine the order in which the three groups decided upon an answer. A member of the team who hit the button first answered the question. The team that gave the correct answer received a point, publicly displayed on their team name card. If the answer was incorrect, the team who hit the light button second was given an opportunity to answer the question and so on. After 16 questions had been presented, the team with the most correct answers was deemed the winner.

In all three conditions following the completion of the 16 test questions, participants rated along a 7-point scale the perceived accuracy of that assessment of their knowledge (1 = *Very accurate* to 7 = *Very inaccurate*) and the degree to which they liked their assessment approach (1 = *Strongly like* to 5 = *Strongly dislike*). Participants in the cooperative and competitive testing

conditions also rated preference for individual versus team format along a seven-point scale (1 = *Strongly prefer individual participation* to 7 = *Strongly prefer team participation*). Group participants were asked if the study were repeated would they like to work again with the same team and answered using a seven-point scale (1 = *Strongly like* to 7 = *Strongly dislike*).

Data Analysis

We calculated participants' percent correct answers and ratings on the 16-item multiple choice test; then, we compared mean scores for the solitary condition to individual (before team discussion) and team answers (following team discussion) for the cooperative and competitive conditions. Participants' ratings of confidence, accuracy, and liking for their condition were also analyzed. Any differential effect of type of question—factual versus conceptual—in each condition was examined. An analysis of the decision-making in the group conditions over all questions, both correct and incorrect, was performed based on a comparison of individual answers to team answers. The three categories of decision making were (a) unanimity—all individual answers were the same as the team answer, (b) majority—two of three team members had the same decision as the team, and (c) authoritarian—only one person in the team had the same answer as the team answer.

Results

Correct Answers: A Comparison Across the Three Conditions

Correct answers on the 16 item multiple choice test were compared across the three conditions: solitary testing and team answers in the cooperative and competitive conditions. Correct answers significantly differed between conditions, $F(2,75) = 7.6$, $p < .001$). A Scheffé post hoc test ($p < .05$) showed that cooperative testing ($M = 64.8\%$, $SD = 13.2$) and competitive testing team scores ($M = 60.2\%$, $SD = 17.0$) were significantly greater than the solitary testing score ($M = 47.7\%$, $SD = 18.1$). Students working in teams were more often correct than students working alone.

Correct Answers: A Comparison of Alone Versus Team Answers

In the cooperative testing condition, correct team answers ($M = 64.8\%$, $SD = 13.2$) were greater than participants' individual pre-discussion correct answers ($M = 54.9\%$, $SD = 11.5$), $t(26) = 5.7$, $p < .001$). In the competitive testing condition, correct team answers ($M = 60.2\%$, $SD = 17.0$) were also superior to participants' individual pre-discussion correct answers ($M = 49.8\%$, $SD = 15.9$), $t(26) = 3.4$ $p < .01$). For respondents in both group conditions, cooperative testing condition and competitive testing condition, the team answer was more often correct than the average individual answer from the same participants before entering group discussion. The inferiority of the solitary testing condition to the group testing conditions was also supported by examining answers to each of the 16 items. Participants' answers in one of the group conditions (competitive or cooperative) were always more correct than those in the solitary condition.

Correct Answers: A Comparison of Answering Alone in all Three Conditions

Individual answers (percent correct before group discussion) for participants in the cooperative testing condition (54.9%, $SD = 11.5$) and competitive testing condition ($M = 49.8\%$, $SD = 15.9$) did not significantly differ from those for participants in the solitary testing condition ($M = 47.7\%$, $SD = 18.1$) across the 16 multiple choice items, $F(2,75) = 1.5$, p = ns. As expected by random assignment, participants' initial knowledge of the course material in the three conditions was statistically equivalent at the beginning of the study.

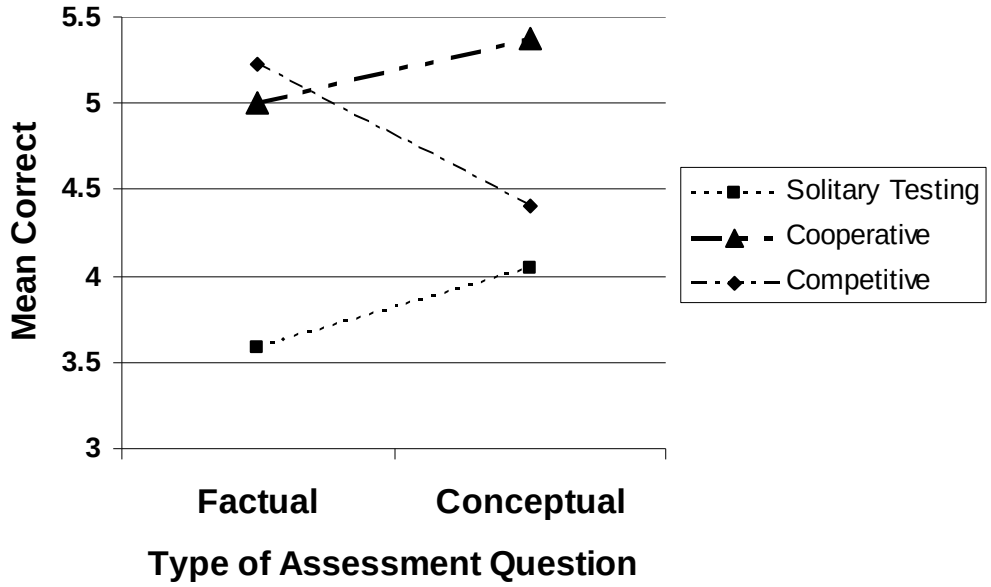
Correct Answer: Factual Versus Conceptual Items

Did the superiority of the cooperative and competitive testing conditions over the solitary testing condition depend on whether the items were factual or conceptual? There were 8 factual and 8 conceptual items on the 16 item multiple choice test. A 2 X 3 mixed ANOVA was calculated for item type (factual or conceptual) by condition (solitary, cooperative, competitive). The main effect of condition was significant $F(2,75) = 7.55$, $p < .001$. Group answers were more often correct than answers given by participants in the solitary testing condition as shown in a post hoc Scheffé ($p < .05$). The main effect of item type was not significant, $F(2,75) = 4.6$, p = ns. The interaction was significant, $F(2,72) = 5.74$, $p < .01$. As seen in Figure 1, only in the competitive condition did participants do worse on the conceptual items relative to the factual items.

Accuracy in Assessment

We measured students' belief in the accuracy of their assessment method (1 = *Very accurate* to 7 = *Very inaccurate*) since one's beliefs may affect performance. There was a significant difference between the three conditions in the participants' rating of how accurately they believed their assessment condition measured their

FIGURE 1
Mean Correct to Factual and Conceptual Assessment Questions in the Solitary Testing, Cooperative Testing, and Competitive Testing Conditions



course knowledge, $F(2,75) = 6.08$, $p < .01$. A Scheffé post hoc test showed that participants in the cooperative testing condition ($M = 2.1$, $SD = 0.7$) perceived their assessment condition's manner of measuring their knowledge as more accurate than did participants in the solitary testing condition ($M = 3.3$, $SD = 1.6$). There were no significant differences involving the competitive testing condition ($M = 2.7$, $SD = 1.3$).

Confidence in Answers

Participants' confidence that their answers were correct was rated along a 5-point scale anchored by 1 = *Extremely* and 5 = *Not at all*. Participants assigned to the group conditions rated their confidence in their answer twice: first, after their individual answer and, second, after their team answer. Participants' confidence that their final answer was correct differed between the three conditions ($F(2,75) = 22.23$, $p < .01$). The solitary testing ($M = 3.4$, $SD = 0.6$) participant ratings were between *Moderately* and *Somewhat confident.* The confidence in the team answer among participants in the cooperative testing condition ($M = 2.3$, $SD = 0.6$) was close to *Very confident.* The confidence in the team answer among participants in the competitive testing condition ($M = 2.6$, $SD = 0.7$) was between *Very* and *Moderately confident.* A post hoc Scheffé indicated a significant difference between the solitary testing condition and both group conditions with participants from both group conditions displaying more confidence that their answers were correct.

We compared confidence ratings for the same participant in their alone versus team answers for each group condition. Paired comparison *t*-tests showed that participants were more confident of their answers in the team compared to individual situation for the cooperative testing ($M = 2.9$; paired $t(15) = 10.80$, $p < .001$) and competitive testing conditions ($M = 3.1$; paired $t(15) = 6.69$, $p < .001$). For the same individual, participants were more confident that the collaborative answer was correct than their individual answer before interacting with others.

Assessment Method Preference

Preference for a particular assessment method may influence its use. Most participants in each condition rated (on a 5-point scale with 1 = *Strongly like* to 5 = *Strongly dislike*) liking for their condition close to the *Moderately liked* category (solitary testing $M = 2.4$, $SD = 0.9$; cooperative testing $M = 1.7$, $SD = 0.5$; competitive testing $M = 2.2$, $SD = 1.0$). There was a significant difference between conditions in participants' liking for their assessment method ($F(2,75) = 4.37$, $p < .01$). A post hoc Scheffé indicated that a significant difference ($p = .05$) existed between the solitary testing and cooperative testing conditions. Undergraduates liked participating in the cooperative assessment approach significantly more than they liked the testing alone approach.

Participants in the cooperative and competitive testing conditions rated their preference for participating in this study as a member of a team or as an individual along a 7-point rating scale (1 = *Strongly prefer individual participation*, 7 = *Strongly prefer team participation*). There was no significant difference between participants in the cooperative testing condition (M = 5.0, SD = 1.5) and the competitive testing condition (M = 5.0, SD = 1.7), $t = 0.0$, p = ns. Participants in both conditions *Somewhat prefer* group participation.

Lastly, participants in the cooperative and competitive testing conditions were asked to rate their degree of like/dislike (1 = *Strongly like*, 7 = *Strongly dislike*) to work again with the same team. There was no significant difference between the two conditions (cooperative testing, $M = 2.1$, $SD = 1.1$; competitive testing, $M = 2.1$, $SD = 1.5$), t (52) = -.2, p = ns. Participants *Moderately liked* the idea of working with the same team again.

Decision making in groups

How were decisions made in the group—by unanimity (all three members agreeing), majority rule (two of three members agreeing), or authoritarian (one member decided on the team answer)? Most frequently, majority rule prevailed in both cooperative (36% unanimous, 43% majority, 21% authoritarian) and competitive conditions (23% unanimous, 53% majority, 24% authoritarian). Unanimity plus majority rule produced more than 75% of decisions in groups. Authoritarian rule, wherein a single individual dictates the team answer, occurred in less than 25% of the decisions.

Discussion

Compared to the solitary testing method, our results showed that the group format produced superior student knowledge on the 16-item multiple choice test. Participants' team scores were higher than their individual answers prior to discussion on the assessment instrument as a whole as well as on most individual items. Students enjoyed the team format more and perceived it as a more accurate measure of their knowledge. Moreover, better team decisions were not simply due to one knowledgeable member in the group providing the correct answers.

These findings of improved performance by teams are consistent with previous research. Evidence for superior performance for students working in groups compared to answering alone (Riggio, Fantuzzo, Connelly, & Dimeff, 1991; Stockdale & Williams, 2004) was replicated in our study. Our research extends this literature by comparing cooperative and competitive testing conditions.

Team Versus Individual Performance

Why do teams arrive at better answers than individuals? It is possible that team members may stimulate and encourage each other through their discussion, termed a synergy effect by Zimbardo et al. (2003). Additionally, error correction procedures may occur in groups to effectively weed out incorrect answers. Through active participation such as verbalizing a reason for one's answer, a student's misconception of the course material may be clarified by fellow students. It is possible that by using a group testing approach instructors are structuring their courses so that students assist each other in mastering the course material, an approach called peer tutoring.

There are other possible interpretations of the positive group effects. Perhaps while working alone a student carelessly reads an item and thus misinterprets the question resulting in an incorrect answer. During group interaction the student's misreading is corrected by one's peers. In this case, rather than better understanding of the course material, the improved performance is due to accurate comprehension of the test question. Research is needed to elucidate the reasons for superior group performance over solitary testing.

Another difference between team versus solitary testing is that students perceived the team format as preferable and more accurate than the individual format to assess their knowledge of course material. When others confirm your initial answer, the recipient's confidence in their answer is bolstered and accompanied by an increase in positive affect due to receiving support by another (Rubin, 1973). Positive affect in the group condition may boost motivation and thus foster learning.

Team Decision Making

Did the team output reflect a group effort or was one individual toting the load? We analyzed the individual answer compared to the team answer to address this question. We found that it was more often the case that the majority of team members governed the team decision. There were no cases where, for a particular team, one person made most of the decisions. Whether this result is an accurate reflection of the discussion that took place, though, is unknown given that our analysis compared individual to team written answers rather than a recording of the team discussion.

The use of a computer-mediated environment may be the next frontier for analysis of how team answers

are determined since automated means would provide a log of transactions between members (Rummel & Spada, 2005). A record of the interactions between team members would allow ready identification and categorization of participation among members as well as provide a basis for determining the category of group interaction (i.e., authoritarian, majority, unanimity). Perhaps with training, the appropriate collaborative behaviors between team members (e.g., discussing differences of opinions, providing rationales, encouraging correct answers) could be facilitated (Prichard, Stratford, & Bizo, 2006).

An identification of authoritarian rule may occur under other conditions. Our study was quite short in duration: only 16 items were administered. Possibly, had a longer assessment been performed, a "knowledgeable" member of the group would have been identified in each team resulting in authoritarian rule affecting the team decision-making (Bonner, 2004; Bonner, Baumann, & Dalal, 2002).

Type of Team Assessment

Given that team format appears more effective in terms of learning than the individual format, which type of team—cooperative or competitive—is better? In most comparisons, the outcomes from the cooperative and competitive groups were equivalent—in terms of correct answers, confidence in answers, learning the course content, and decision making.

Conclusion

Placing students in teams to assess their knowledge of a subject matter appears to be a viable strategy according to our results and may provide instructors with an attractive option to solitary testing. Students perform well in group situations, perceive it as an accurate measure of their knowledge, like the experience, and appear to work together to arrive at answers to questions. This group assessment approach may also generate student learning through peer interaction and help students comprehend test items.

References

Birenbaum, M. (2007). Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education, 53*(6), 749-768.

Bonner, B. L. (2004). Expertise in group problem solving: Recognition, social combination, and performance. *Group Dynamics: Theory, Research, and Practice*, *8*(4), 277-290.

Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes, 88*(2), 719-736.

Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*(1), 37-50.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education*, *27*(3), 102-108.

Desrochers, M. N., Pusateri, M., & Fink, H. (2007). Game assessment: Fun as well as effective. *Assessment and Evaluation in Higher Education, 32*(5), 527-539.

Hinsz, V. B., & Nickell, G. S. (2004). Positive reactions to working in groups in a study of group and individual goal decision making. *Group Dynamics: Theory, Research, and Practice, 8*(4), 253-264.

Hogan, T. P. (2007). *Psychological testing: A practical introduction* (2nd ed.). Indianapolis, IN: John Wiley & Sons.

Ingram, E. L., & Nelson, C. E. (2006). Using discussions of multiple choice questions to help students identify misconceptions and reconstruct their understanding. *The American Biology Teacher, 68*(5), 275-279.

Jensen, M. (1996). Cooperative quizzes in the anatomy and physiology laboratory: A description and evaluation. *Advances in Physiology Education, 16*(1), 48-54.

Jensen, M., Moore, R., & Hatch, J. (2002). Electronic cooperative quizzes. *American Biology Teacher, 64*(3), 169-174.

Johnson, R. L., Penny, J., & Gordon, B. (2001). Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication, 18*(2), 229-250.

Longford, N. (1994). Reliability of essay rating and score adjustment. *Journal of Educational and Behavioral Statistics, 19*(3), 171-200.

Lu, C., & Suen, H. K. (1995). Assessment approaches and cognitive styles. *Journal of Educational Measurement, 32*(1), 1-17.

McGimsey, J. F., Greene, B. F., & Lutzker J. R. (1995). Competence in aspects of behavioral treatment and consultation: Implications for service delivery and graduate training. *Journal of Applied Behavior Analysis, 28*(3), 301-315.

Mitchell, N., & Melton, S. (2003). Collaborative testing: An innovative approach to test taking. *Nurse Educator*, *28*(2), 95-97.

Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science, 65*(3), 221-244.

Norton, L. S., Tilley, A. J., Newstead, S. E., & Franklyn-Stokes, A. (2001). The pressures of

assessment in undergraduate courses and their effect on student behaviours. *Assessment & Evaluation in Higher Education, 26*(3), 269-284.

Prichard, J. S., Stratford, R. J., & Bizo, L. A. (2006). Team-skills training enhances collaborative learning. *Learning and Instruction, 16*(6), 256-265.

Rao, S. P., & DiCarlo, S. E. (2000). Peer instruction improves performance on quizzes. *Advances in Physiology Education, 24*(1), 51-55.

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26*(1), 37-41.

Riggio, R. E., Fantuzzo, J. W., Connelly, S., & Dimeff, I. A. (1991). Reciprocal peer tutoring: A classroom strategy for promoting academic and social integration in undergraduate students. *Journal of Social Behavior & Personality, 6*(2), 387-396.

Rubin, Z. (1973). *Liking and loving: An invitation to social psychology*. New York: Holt, Rinehard and Winston.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201-241.

Stockdale, S. L., & Williams, R. L. (2004). Cooperative learning groups at the college level: Differential effects on high, average, and low exam performers. *Journal of Behavioral Education, 13*(1), 37-50.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.

Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative college examinations: More gain, less pain when students share information and grades. *The Journal of Experimental Education, 71(2)*, 101-125.

MARCIE N. DESROCHERS is an assistant professor of psychology at the State University of New York, College at Brockport. Her research and teaching interests are in behavior analysis.

HERBERT C. FINK is an assistant professor of psychology at the State University of New York, College at Brockport. His research and teaching interests are in social psychology.

ANDREA THOMAS, JOE KIMMERLING, and WENDY TUNG are all graduate students in the Psychology Master's program at State University of New York, College at Brockport.