Collaborative group testing benefits high- and low-performing students

Mauricio J. Giuliodori,1 Heidi L. Lujan,2 and Stephen E. DiCarlo2
1Cátedra de Fisiología, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata, Argentina; and 2Department of Physiology, Wayne State University School of Medicine, Detroit, Michigan

Submitted 7 November 2007; accepted in final form 8 August 2008

Giuliodori MJ, Lujan HL, DiCarlo SE. Collaborative group testing benefits high- and low-performing students. Adv Physiol Educ 32: 274–278, 2008; doi:10.1152/advan.00101.2007.—We used collaborative group testing in a veterinary physiology course (65 students) to test the hypothesis that all students (e.g., high-performing and low-performing students of each group) benefit from collaborative group testing. In this format, students answered questions in the traditional format as individuals. Immediately after completing the exam as individuals, students answered the same questions in groups of two, and, finally, the same questions were discussed by the instructor and students. We measured two learning outcomes for every student: individual and group test scores. Based on individual test scores, students were categorized as “high performing” (students with higher individual scores) or “low performing” (students with lower individual scores). Finally, student evaluations of the format were collected. Collaborative group testing enhanced student performance. Specifically, group scores were higher than individual scores (collected. Collaborative group testing enhanced student performance. Finally, student evaluations of the format were higher individual scores) or “low performing” (students with lower individual scores). Finally, student evaluations of the format were collected. Collaborative group testing enhanced student performance. Specifically, group scores were higher than individual scores (P < 0.001). Importantly, the size of the collaborative testing effect was large for the population and for the low-performing students; however, the collaborative testing effect was small for the high-performing students. Finally, student evaluations of this testing format were very positive. In conclusion, collaborative group testing was beneficial for all students; however, collaborative testing was significantly more beneficial for low-performing students.

feedback; test scores

PHYSIOLOGY is a creative, constructivistic science that requires discussion and debate through student-to-student and student-to-teacher interactions (7). Constructivist learning theory is based on the idea that knowledge is actively constructed by the learner. Social constructivism suggests that knowledge creation is a shared experience rather than an individual experience (23). Teachers can encourage constructivism by carefully considering the instructional strategies that they use (7, 14, 22, 27). For example, exams, when administered in a collaborative format, can be used as teaching and learning tools promoting constructivism (5, 24). Specifically, adding a collaborative component to the traditional individual exam format transforms traditional evaluation into a formative type of assessment. Formative assessment is assessment for learning, that is, to inform and improve student performance, whereas summative assessment is assessment of learning, that is, to evaluate student performance (30). However, the pedagogical value of traditional exams is reduced because students usually complete the exam individually. In this situation, students work in isolation from classmates. Unassisted student learning is pedagogically unsound, since most learning is assisted and promoted by the efforts of a wide variety of individuals within and outside the school (12). Furthermore, because feedback from traditional exams is delayed and often insufficient, students are often unclear about the correct answer and whether the thought process behind the answer is adequate (11). It would be educationally sound for students to discuss the individual questions immediately after completing the exam (24).

In contrast, in the collaborative testing setting, students become more engaged in learning by discussing material, facilitating understanding, and encouraging hard work. With collaborative testing, students have the emotional and intellectual support that allows them to go beyond their present knowledge and skills and accomplish shared goals (28). In addition, cooperative learning has positive effects on race relations, self-esteem, and a willingness to cooperate in other settings (29).

For students to be competitive in the global economy, they need to learn to collaborate and solve problems in small groups and apply what they’ve learned in the real world. These lessons also teach students to show respect for others as well as to be punctual, responsible, and work well in teams. These skills, part of the hidden curriculum, are missing in many graduates. The hidden curriculum consists of cognitive skills, attitudes, and behaviors that are not made explicit to students but are acquired because of the way the teacher organizes and conducts the class (13).

Collaborative activities are associated with several positive outcomes, such as student satisfaction (10), academic performance (4, 13, 25, 30), patient-centered interviewing skills (20), and professional behaviors (13). However, it is unknown if all students receive equal benefits from collaborative procedures. Furthermore, some educators may be concerned that less-prepared students will defer to better-prepared students or that strong students will carry their partners. Therefore, this study was designed to test the hypothesis that all students (e.g., high-performing and low-performing students) benefit from collaborative group testing.

METHODS

This study was approved by the Internal Review Board of the Faculty of Veterinary Sciences of National University of La Plata. We used collaborative group testing in a veterinary physiology course (65 students) to test the hypothesis that all students (e.g., high-performing and low-performing students) benefit from collaborative group testing. In this format, students initially answered questions in the traditional format as individuals. Immediately after completing the exam as individuals, students answered the same questions in groups of two, and, finally, the same questions were discussed by the instructor and students. We measured two learning outcomes for every student: individual and group test scores. The objectives were 1) to determine the effect of collaborative group testing on the test performance of high-performing and low-performing students; 2) to determine the proportion of students who increased, decreased, or had no change in their individual test scores with collaborative group testing.
as well as the magnitude of those changes; and 3) to record student evaluations of this testing methodology.

**Student Population**

The collaborative group testing methodology was implemented during the Physiology course (Fisiología no. 423) at the Facultad de Ciencias Veterinarias (Universidad Nacional de La Plata, La Plata, Argentina). The course was lecture based with peer instruction activities (9). The student population consisted of 65 veterinary medical students [age: 22 (SD 3.5) yr, 43 women and 22 men].

**Testing Procedures**

The collaborative group testing procedures consisted of three steps: 1) traditional individual testing, 2) collaborative group testing, and 3) class testing (Table 1).

*Traditional individual tests.* Each student initially completed the exam individually for the first 40 min (~2 min/question). After completion, all of the exams were collected and used for the individual scores.

*Collaborative group tests.* Students who choose to participate selected their partner, sat together in pairs, and completed the same exam over the next 30 min (~1.5 min/question). For the collaborative group tests, only one exam was given to each group, and both members of the group wrote their name on the answer sheet. At the end of this step, all of the exams were collected and used for the group scores (grades were based only on individual testing, such that no extra points were awarded by group testing).

*Class tests.* During this step, the same exam questions were projected, and the students answered the questions with a class response system called “colored letters” (8). Briefly, five colored sheets of paper, labeled A (red), B (white), C (blue), D (green), and E (yellow), were given to the students, and students answered the questions by holding up the colored letter representing their selected option. This step required 20 min (~1 min/question). The instructor selected representative students, based on the distribution of the responses, to discuss the answers. The intent of this step was to emphasize the reasoning for each correct response. When the rationale for the correct answer given by the selected student was not accurate, the instructor provided the reasons for the correct answer. In addition, the instructor gave the reasons for the most common incorrect answers. This testing format provided time for feedback provided by peers (group testing) and by the instructor (class testing).

**Measurements**

Two learning outcomes, individual test scores and group test scores (numbers and percentages of correct responses), were measured for every student (Fig. 1).

**Table 1. Sample test questions**

Predict what would happen with the resting membrane potential if the K\(^+\) concentration in the extracellular fluid increased (e.g., from 3 to 5 mM):
- A. Increase (becomes more negative)*
- B. Decrease (becomes less negative)*
- C. No change

Predict what would happen to cardiac output if arteriolar smooth muscle constricts:
- A. Increase*
- B. Decrease
- C. No change

Predict what would happen with milk production (in kg/day) if oxytocin is administered to a dairy cow:
- A. Increase
- B. Decrease
- C. No change*

*Correct answer.

**Fig. 1.** Individual scores and group scores for the three written examinations (exams 1, 2, and 3) from a veterinary physiology course where collaborative group testing was used. *P < 0.001, individual scores vs. group scores.

Group scores were also compared with individual scores and classified as higher, lower, or the same as individual scores to determine the percentage of students who improved, decreased, or stayed the same after collaborative group testing. In addition, individual scores of students who increased or decreased their scores after collaborative group testing together with the number of questions gained or lost were also measured (Table 2).

Students were classified as the high-performing or low-performing student of the group based on their individual scores. For example, the student with the higher individual score of every group was classified as the high-performing student, and the student with the lower individual score of every group was classified as the low-performing student. The individual scores were recorded and compared with group scores to test the hypothesis that collaborative group testing enhances the performance of all students (Table 3). Groups with both members having the same individual scores were not included in the analysis since we were unable to categorize them as high- or low-performing students. The effect size (ES) of collaborative group testing on student performance for the population (all students combined) as well as for the high- and low-performing students was also measured (Table 4).

Finally, student evaluations regarding this testing technique were obtained with a 17-item questionnaire given to students on the day of the second examination. Student responses to these statements were on a 5-point scale (from 1 (completely disagree) to 5 (completely agree); Table 5).

**Statistical Analysis**

All results are presented as means (SD), and significance was set at the P < 0.05 level. To determine if collaborative group testing increased student performance [test score (Fig. 1)], we used a Student’s paired t-test on the raw data to compare responses obtained when the students answered questions as individuals with responses obtained when the students answered questions in groups.

To determine if collaborative group testing enhanced the performance of high- and low-performing students, we used a Student’s paired t-test on the raw data to compare individual scores versus group scores from high- and low-performing students (Table 3). To determine the impact of collaborative group testing on student performance (population, high-performing students, and low-performing students), we estimated the ES with the Hedges’s method [mean difference/SD (population, high-performing students, and low-performing students), we estimated the ES with the Hedges’s method [mean difference/SD pooled; Table 4 (21)]. Finally, student responses to the evaluation were expressed as means (SD) (Table 5).
How We Teach  

COLLABORATIVE GROUP TESTING INCREASES SCORES

Table 2. Individual scores and questions gained or lost by students who increased or decreased their scores after collaborative group testing

<table>
<thead>
<tr>
<th>No. of Students for the 3 Exams</th>
<th>Individual Scores</th>
<th>No. of Questions Gained or Lost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD), %</td>
<td>95% CI</td>
</tr>
<tr>
<td>Increased scores</td>
<td>106</td>
<td>55.9 (14.7)</td>
</tr>
<tr>
<td>Decreased scores</td>
<td>11</td>
<td>68.3 (13.5)</td>
</tr>
</tbody>
</table>

CI, confidence interval.

RESULTS

The percentages of correct responses when the students completed exams as individuals and when the students completed the same exams in groups are shown in Fig. 1. As individuals, students responded correctly 58.7 ± 15.6% of the time. In sharp contrast, in groups, students answered the same questions correctly 70.2 ± 13.7% of the time [t-value: 11.4, degree of freedom (df): 137, P < 0.001, mean difference (raw data): 2.28, 95% confidence interval (CI): 1.88–2.67]. Group scores were higher than individual scores 76.8% (106 of 138) of the time for the three examinations (71.4%, 73.8%, and 83.3% for exams 1, 2 and 3, respectively). In sharp contrast, group scores were lower than individual scores only 7.9% (11 of 138) of the time for the three exams (7.1%, 11.9%, and 5.5% for exams 1, 2, and 3, respectively). Finally, group scores were the same as individual scores 15.2% (21 of 138) of the time for the three examinations (21.4%, 14.2%, and 11.1% for exams 1, 2, and 3, respectively). Data from students who increased or decreased their scores after collaborative group testing are shown in Table 2. Students who increased their scores during group testing had a mean individual score of 55.9 ± 14.7% and answered 3.1 ± 2.0 additional questions correctly during group testing. In contrast, students who decreased their scores during group testing had a mean individual score of 68.3 ± 13.5% and answered 1.5 ± 0.8 fewer questions correctly. These changes were significant [paired t-test (individual scores vs. group scores) t-value: 16.3, df: 105, and P < 0.001 for students increasing their scores; and t-value: 5.9, df: 10, and P < 0.001 for students decreasing their scores].

Individual scores (in %) and group scores (in %) from high- and low-performing students are shown in Table 3. The high-performing student of every group responded correctly 66.7 ± 13.3% of the time as individuals and 69.7 ± 13.7% of the time during group testing. The low-performing student of every group responded correctly 49.7 ± 14.1% of the time as individuals and 69.7 ± 14% of the time during group testing.

Standardized ESs of collaborative group testing on the performance of students are shown in Table 4. ESs and 95% CIs were 0.78 (95% CI: 0.53–1.02) for the student population, 0.22 (95% CI: –0.14 to 0.57) for high-performing students, and 1.38 (95% CI: 0.99–1.77) for low-performing students.

Student evaluations of this testing activity were very positive and are shown in Table 5.

DISCUSSION

In this work, we document that veterinary students improved their test scores during collaborative testing. These results are consistent with previous work. For example, collaborative testing improved test scores for medical students (24) and nursing students (15). Similarly, Crouch and Mazur (6) observed significant increases in conceptual problem-solving skills involving physics scenarios over a 10-yr period of peer instruction experience.

The new finding in this work is that although all students benefited from collaborative testing (Table 3), low-performing students increased their scores more than high-performing students. To the best of our knowledge, this work represents the first report documenting an impact of collaborative group testing for high- and low-performing students.

These are important findings because some educators are concerned that less-prepared students may defer to better-prepared students and/or that high-performing students may “carry” low-performing students during group testing. However, it remains to be determined whether the higher improvement for the low-performing students resulted from students being persuaded by stronger classmates or on students being convinced by someone who had the correct answer and the appropriate rationale.

This study extends our recent work (9) documenting that peer instruction enhanced veterinary medical student performance on similar qualitative problem scenarios (Table 1). These problems require the integration of multiple concepts and necessitate that students explore the underlying concepts and generate connections to other information. This complex cognitive activity of information integration requires time (3). Therefore, some students fail to answer these types of ques-

Table 3. Individual scores and group scores from high- and low-performing students for 3 written examinations in a veterinary physiology course where collaborative testing was used

<table>
<thead>
<tr>
<th>No. of Groups Having Students With Unequal Scores for Each Exam</th>
<th>High-Performing Students</th>
<th>Low-Performing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Scores</td>
<td>Group Scores</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 1</td>
<td>19</td>
<td>63.9 (11.8)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>59.2 (13.3)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>74.2 (10.7)</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>66.7 (13.3)</td>
</tr>
</tbody>
</table>

Values are means (SD) (in %). *P < 0.05; †P < 0.01; ‡P < 0.001 (individual scores vs. group scores).
Table 4. Hedges’ ES and 95% CI of collaborative group testing on student performance in a veterinary physiology course

<table>
<thead>
<tr>
<th>Student Population</th>
<th>High-Performing Students</th>
<th>Low-Performing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ES 95% CI</td>
<td>ES 95% CI</td>
</tr>
<tr>
<td>Exam 1</td>
<td>0.72 0.28–1.16</td>
<td>0.24 −0.40 to 0.87</td>
</tr>
<tr>
<td>Exam 2</td>
<td>0.78 0.33–1.22</td>
<td>0.13 −0.53 to 0.78</td>
</tr>
<tr>
<td>Exam 3</td>
<td>1.08 0.68–1.48</td>
<td>0.38 −0.18 to 0.94</td>
</tr>
<tr>
<td>Total</td>
<td>0.78 0.53–1.02</td>
<td>0.22 −0.14 to 0.57</td>
</tr>
</tbody>
</table>

ES, effect size.

Table 5. Student evaluations of the collaborative group testing methodology

<table>
<thead>
<tr>
<th>Question</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The purpose of and rationale behind the educational process was fully explained.</td>
<td>4.4 (0.6)</td>
</tr>
<tr>
<td>2. The process was not too lengthy or complex in its format.</td>
<td>3.9 (0.6)</td>
</tr>
<tr>
<td>3. The peer discussions improved my level of confidence in the answers.</td>
<td>4.0 (1.1)</td>
</tr>
<tr>
<td>4. The level of peer (group) discussions was very high.</td>
<td>4.1 (0.9)</td>
</tr>
<tr>
<td>5. The immediate feedback given by peer discussions was very positive.</td>
<td>4.0 (0.8)</td>
</tr>
<tr>
<td>6. The peer (group) discussions enhanced my understanding of the concepts.</td>
<td>4.3 (1.0)</td>
</tr>
<tr>
<td>7. My level of involvement during discussions was high.</td>
<td>4.0 (0.6)</td>
</tr>
<tr>
<td>8. The level of discussions with the instructor was very high.</td>
<td>4.0 (0.6)</td>
</tr>
<tr>
<td>9. The immediate feedback given by discussions with the instructor was very positive.</td>
<td>4.1 (0.9)</td>
</tr>
<tr>
<td>10. The discussion with the instructor enhanced my understanding of the concepts.</td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td>11. This testing methodology provided a more constructive classroom environment.</td>
<td>4.6 (0.6)</td>
</tr>
<tr>
<td>12. This testing methodology provided the opportunity to discuss incorrect answers and fill in knowledge gaps.</td>
<td>4.4 (0.8)</td>
</tr>
<tr>
<td>13. Knowing my test score before leaving the classroom was very positive.</td>
<td>4.5 (0.9)</td>
</tr>
<tr>
<td>14. This testing methodology was educationally attractive due to the novelty of this style and format.</td>
<td>4.3 (0.9)</td>
</tr>
<tr>
<td>15. This testing methodology was less stressful than traditional testing methods.</td>
<td>3.9 (0.9)</td>
</tr>
</tbody>
</table>

Responses are presented as means (SD); 33 of 44 students returned the completed questionnaire (75% response rate). Students were given the following instructions to respond to the questions: please circle the number (1–5) that most accurately defines the way you feel regarding each statement, where 1 = completely disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; and 5 = completely agree.

It is advisable that the instructor provides corrective feedback when collaborative group testing is implemented for many reasons. First, a low percentage of students will decrease their group scores. Second, some students with the correct answer may fail to give the right reasons for their correct response. Third, some questions remain unanswered by many students (even for the higher performers). Therefore, the instructor, as the authority, should monitor the learning that is taking place in the classroom during collaborative testing. To put it simply, the instructor provides the right reasons for the correct answers to all the test questions.

Student evaluations of collaborative group testing were very positive and were in agreement with previous studies on collaborative learning activities (4, 5, 24, 25). We observed students listening, providing constructive feedback, and revisions (9). However, under this challenging situation, collaborative group testing increased the performance of the majority of students while it decreased the scores of a small proportion of the population. These results are in agreement with the results of Rossignol (26), who reported that 73.5% of nursing students (25 of 34 students) gained points, 14.7% lost points (5 of 34 students), and 11.7% had no change (4 of 34 students) with collaborative testing. In that study (26), the students who improved their scores answered 1–12 additional questions correctly (3.9 ± 3.8), whereas the students who decreased their scores with collaborative testing answered 1–2 fewer questions correctly.

Importantly, there is always a population increase and never a decrease in the number of correct answers after discussion with classmates, and it is much easier to convince someone who is wrong than it is convince someone who has selected the correct answers for the right reasons (16). In this sense, the maximum benefit of collaborative testing occur when the percentage of correct individual responses is between 35% and 70% (16). In the present study, the percentage of correct individual responses was within the range for optimal improvement.

Individual work on test questions during individual testing is very important. It provides the time for students to order their thoughts and consider their understanding of the concepts at hand. Subsequently, group testing provides the time and opportunity for their mental models to be revised (and possibly corrected) by peer discussion. Thus, the beneficial effect of collaborative group testing could be due to the generation of explanations for themselves and for their classmates (2). That is, the best way to learn something is to teach it, since teaching requires the generation of explanations, both for oneself and the learner (17).

We estimated the ES (95% CI) for collaborative testing from previous reports (5, 26). The estimated ESs from previous reports were 2.05 (95% CI: 1.26–2.83) (5), 1.21 (95% CI: 0.69–1.73) (26), and 1.46 (95% CI: 0.68–2.24) (23a). These effect sizes were slightly higher than the ES reported in this study (Table 4). Furthermore, low-performing students had larger ESs than high-performing students (1.38 vs. 0.22 for low- and high-performing students, respectively; Table 4). The zone of proximal development, a learning theory by Vygotsky, proposes that a student can perform better (e.g., solve problems) working with the assistance of a more capable peer than working alone (3). Furthermore, skills that a student can do with assistance today can be achieved independently in the future. Therefore, the fact that the ES was larger for the low-performing students than for the high-performing peers was expected based on the work of Vygotsky, where a more capable peer helps a lower-skilled partner to the point where the low performer significantly improves.

We estimated the ES (95% CI) for collaborative testing from previous reports (5, 26). The estimated ESs from previous reports were 2.05 (95% CI: 1.26–2.83) (5), 1.21 (95% CI: 0.69–1.73) (26), and 1.46 (95% CI: 0.68–2.24) (23a). These effect sizes were slightly higher than the ES reported in this study (Table 4). Furthermore, low-performing students had larger ESs than high-performing students (1.38 vs. 0.22 for low- and high-performing students, respectively; Table 4). The zone of proximal development, a learning theory by Vygotsky, proposes that a student can perform better (e.g., solve problems) working with the assistance of a more capable peer than working alone (3). Furthermore, skills that a student can do with assistance today can be achieved independently in the future. Therefore, the fact that the ES was larger for the low-performing students than for the high-performing peers was expected based on the work of Vygotsky, where a more capable peer helps a lower-skilled partner to the point where the low performer significantly improves.
reflecting on their learning. The interaction between the instructor and students, as well as students and students, is one of the most powerful factors in promoting learning. In fact, interacting with peers and with the instructor strongly correlates with improvements in grade point average, graduating with honors, critical thinking, problem solving, and interpersonal skills (1). However, immediate feedback is not possible during traditional testing, where an individualistic and competitive atmosphere predominates. Furthermore, educators too often view examinations only as a way of assigning grades, and little emphasis is placed on using exams to help instructors teach and students learn.

Exams should be used as a mechanism for evaluating performance as well as enhancing learning (19). During traditional testing, students sit silently in the classroom and answer questions during the test session. Afterward, they get their grades a few days later. Some students, especially those failing the exam, visit the instructor with a message similar to this: “Oh, I failed by only one or two questions.” And, most of the students spend the contact time with the instructor trying to get the points needed to pass the test. Obviously, little or no learning results from this student-instructor interaction. However, with the testing methodology described here, correct answers with the rationale for them are discussed with all students in the classroom. In this context, the role of the instructor, acting as a coach, is to help students discover the rationale behind correct and incorrect answers. At the same time, it is also beneficial for the instructor because many undiagnosed misconceptions are uncovered in the process.

In conclusion, collaborative group testing enhanced the performance of high- and low-performing students; however, the benefits were greater for low-performing students than for their high-performing peers.

ACKNOWLEDGMENTS

We thank Veronica Rebollo Pratz and Silvina Raffo for the expert technical assistance with data analysis.

REFERENCES