Student interaction characteristics during collaborative group testing

Mauricio J. Giuliodori,1 Heidi L. Lujan,2 and Stephen E. DiCarlo2

1Catedra 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 3 July 2008; accepted in final form 30 December 2008

Giuliodori MJ, Lujan HL, DiCarlo SE. Student interaction characteristics during collaborative group testing. Adv Physiol Educ 33: 24–29, 2009; doi:10.1152/advan.90161.2008.—We used collaborative testing in a veterinary physiology course (65 students) to answer the following questions: 1) do students with individual correct responses or students with individual incorrect responses change their answers during group testing? and 2) do high-performing students make the decisions, that is, are low-performing students carried by high-performing peers? To address these questions, students first completed the exam in the traditional format as individuals. After completing the exam as individuals, students completed the same exam in groups of two. Finally, the same questions were discussed by the instructor and students (instructor feedback). We found that students with individual incorrect responses changed their answers during group testing more often than students with individual correct responses (odds ratio: 7.58, P < 0.01). Furthermore, student feedback was more beneficial when group members had different individual answers than when they had same individual answers (P < 0.05). In addition, when group members had different individual answers, more answers were changed to correct responses than to incorrect responses (77% vs. 23%, P < 0.01). It was more important to have the correct answer than to be the high-performing student, because the student with the correct response (being either the high- or low-performing student) generally prevailed (~80% of the time, P = 0.5). Finally, the positive effects of group testing (77% of total effects, P < 0.05) were due to students who changed their individual answer to the correct response after discussion with peers with the correct response and also with the incorrect individual response.

student feedback; peer instruction

THE American Association for the Advancement of Science (1) strongly recommends that “...science be taught as science is practiced...” This means that the teaching approach must be consistent with the nature of scientific inquiry (12). An essential part of scientific inquiry is collaboration (1). Why, then, do we discourage collaboration by telling our students to “do your own work” and “keep your eyes on your own paper?” This is educationally unwise because physiology is a creative science in which the synthesis of new ideas requires discussion and debate through student-to-student and student-to-teacher interactions as well as constructivism. That is, constructing knowledge by asking questions, developing answers, and interacting with peers, teachers, and the subject (6). Prompt feedback is also required to focus instruction on the most difficult concepts (3).

However, some faculty members are reluctant to incorporate collaborative testing activities in class. The reasons most often advanced for not including these collaborative activities include concerns that less-prepared students will be “carried” by the more industrious ones or that they will defer to their higher-performing classmates. However, students reported that every group member “pulled their weight” and that the level of discussion and their involvement were high during collaborative group testing as well as peer instruction activities (4, 8). Furthermore, current evidence suggests that students feel a responsibility for the group’s success and that group members tend to ensure that everyone is doing their share (11). Therefore, it is unlikely that students will be carried along in the process. Thus, this potential concern appears unwarranted.

However, the interaction characteristics that give rise to the effects of collaboration are unknown. For example, do students with individual correct answers simply convince students with individual incorrect responses to change their answer? Furthermore, which students (high performing or low performing) make the greatest impact on group responses? An understanding of the peer dynamics during group testing discussions may support the incorporation of collaborative group testing. Therefore, this study was designed to answer the following research questions. First, which students (students with individual correct answers or students with individual incorrect answers) most often change their answer during group testing? For example, do students with individual correct responses or students with incorrect responses change their answer during group testing? Second, which students (high performing or low performing) make the greatest impact on group responses? For example, do high-performing students make the decisions, that is, are low-performing students carried by their high-performing peers?

METHODS

This study was reviewed and approved by the Internal Review Board of the Faculty of Veterinary Sciences of the National University of La Plata. We used collaborative group testing in a veterinary physiology course to characterize student interactions during their discussions. 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.

Student Population

Sixty-five veterinary medical students [age: 22 (SD 3.5) yr, 43 females and 22 males] attending a physiology course (Fisiología No. 423, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata, Argentina) were included in the study. The course was lecture based and delivered by the same instructor (M. J. Giuliodori) using peer instruction (8).

Procedures

Three written exams were administered; the first two exams consisted of 20 qualitative problems, and the third exam consisted of 24
qualitative problems. Subjects tested included cell, nervous, muscular, and cardiovascular physiology (exam 1); respiratory, renal, and body fluid physiology (exam 2); and digestive, endocrine, and reproductive physiology (exam 3). The percentages of students in attendance for each exam were 68% (44 of 65 students) for exams 1 and 2 and 86% (56 of 65 students) for exam 3.

Exam questions were qualitative problems (9) that ask for a qualitative prediction (increase/decrease/no change) about the response of a system to a perturbation (8). Specifically, the questions posed conceptual problem-solving scenarios that required the integration of multiple concepts but were answered with single best multiple-choice questions (9, 10). A sample of the exam questions is shown in Table 1.

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

Traditional individual testing. For the traditional individual test, 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 individual scores.

Collaborative group testing. For the collaborative group test, students who choose to participate selected their partner, sat together in pairs, and completed the same exam over the following 30 min (~1.5 min/question). 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 group scores. Grades were based only on individual testing (no extra points were given for group testing).

Class testing. For the class test, the same exam questions were projected, and the students answered the questions with a class response system called colored letters (7). Briefly, five colored sheets of paper labeled A (red), B (white), C (blue), D (green), and E (yellow) were given to the students, and they answered 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 adequate, the instructor offered the right reasons for the correct answer. In addition, the instructor provided the reasons for the most common incorrect answers. Therefore, this testing format provided feedback provided by peers (group testing) and by the instructor (class testing).

Measurements

The numbers and percentages of correct and incorrect individual answers that changed or did not change after group testing were measured to answer the following research question: which students (students with correct answers or students with incorrect answers) most often change their answer? For example, is it more likely to change the answer of students having correct responses or students having incorrect answers?

In addition, groups were categorized based on the correctness of the individual answers of their members (to every question) as follows: 1) both individual answers correct (CC group), 2) one individual answer correct and the other incorrect (CI group), and 3) both individual answers incorrect (II group). We compared the group answers that changed positively (from incorrect to correct) or negatively (from correct to incorrect) or that did not change after group testing to answer the following research question: what happens during the discussions during group testing between two students with the same answers (correct or incorrect) or different answers (one student with the correct response and the other student with the incorrect response)?

Furthermore, group partners were classified, based on their individual testing scores, as high-performing students (the higher performer of every group) or low-performing students (the lower performer of every group). Then, when group members had different individual answers (one student that answered correctly and the other student that answered incorrectly), we compared the group answers that changed positively (from incorrect to correct) or negatively (from correct to incorrect) after group testing, depending on which member had the correct answer (high performing versus low performing), to answer the following research question: what has the greatest impact on the group answer: being the high-performing student or the student with the correct answer?

Finally, positive effects (answers that changed from incorrect to correct) and negative effects (answers that changed from correct to incorrect) during collaborative group testing were measured to answer the following research question: what types of student interactions, based on the correctness of the individual responses of the group members (both correct vs. one correct and one incorrect vs. both incorrect), are responsible for the positive and negative effects of collaborative group testing?

Statistical Analysis

All results are presented as means (SD). Significance was set at the $P < 0.05$ level.

To determine which students changed their individual response after collaborative group testing (Fig. 1), we used a $\chi^2$-test. Subsequently, an odds ratio (OR) was calculated to assess the likelihood for change.

To determine which students changed their group responses (Fig. 2), we used a Kruskal-Wallis nonparametric one-way ANOVA on ranks. Once statistical significance was established, post hoc analysis was performed with Dunnett’s method for all pair-wise multiple comparisons. The OR was also used to assess the likelihood for change.

To determine what had the greatest impact on the group answer, the high-performing student or the student with the correct answer (Fig. 3), we used a $\chi^2$-test.

Finally, to compare positive effects with negative effects of collaborative testing, we used a Kruskal-Wallis nonparametric one-way ANOVA on ranks (Fig. 4). Once statistical significance was established, post hoc analysis was performed with Dunnett’s method for all pair-wise multiple comparisons.

RESULTS

The percentages of correct and incorrect individual responses that did not change or that changed after collaborative group testing are shown in Fig. 1. Twenty-two percent (640 of 2,974) of the total individual answers were changed after the discussion, whereas the remaining 78% (2,334 of 2,974) were not changed. In addition, most of the changes were recorded in

Table 1. Sample test questions from the exams

<table>
<thead>
<tr>
<th>Question</th>
<th>Correct Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict what would happen with cardiac output if the arteriolar smooth muscle contracts?</td>
<td>A. Increase*</td>
</tr>
<tr>
<td>B. Decrease</td>
<td></td>
</tr>
<tr>
<td>C. No change</td>
<td></td>
</tr>
<tr>
<td>Predict what would happen with the resting membrane potential if the K⁺ concentration in the extracellular fluid increases (e.g., from 3 to 5 mM)?</td>
<td>A. Increase (becomes more negative)</td>
</tr>
<tr>
<td>B. Decrease (becomes more positive)*</td>
<td></td>
</tr>
<tr>
<td>C. No change</td>
<td></td>
</tr>
<tr>
<td>Predict what would happen with the milk production (in kg/day) if pharmacological doses of oxytocin are administered to a dairy cow:</td>
<td>A. Increase</td>
</tr>
<tr>
<td>B. Decrease</td>
<td></td>
</tr>
<tr>
<td>C. No change*</td>
<td></td>
</tr>
</tbody>
</table>

*Correct answer.
the groups of students with individual incorrect responses (77%, 494 of 640) compared with students with individual correct answers (23%, 146 of 640). The probability for changing answers after collaborative group testing was 7.58 times higher for students with individual incorrect responses than for students with individual correct responses [OR: 7.58, 95% confidence interval (95%CI): 6.18–9.31, \( \chi^2: 446.465, \) degree of freedom (df): 1, \( P < 0.01 \)]. In the group of individual correct answers (59% of the total answers), 92% were not changed. However, 8% were changed to incorrect responses (negative effect). In sharp contrast, in the group of individual incorrect answers (41% of the total answers), 61% were not changed and 39% were changed to a correct response after collaborative group testing (positive effect).

Group responses stratified according to the individual answers of both group members are shown in Fig. 2. A total of 1,487 group responses were computed (612 with both answers correct, 536 with one answer correct and one answer incorrect, and 339 with both answers incorrect). When the discussions were between two students with each individual having the correct answer (\( n = 612 \)), the students did not change their responses 98% of time and only 2% of the groups changed their response to incorrect answers. When the discussions were between one student with a correct response and one student with an incorrect response (\( n = 536 \)), group responses changed to correct responses 77% of the time and changed to incorrect responses 23% of the time. Finally, when the discussions were
between two individuals with incorrect answers \((n = 339)\), the students failed to change their answer 88% of the time. However, these students did change their answer to the correct response 12% of the time. A Kruskal-Wallis nonparametric one-way ANOVA on ranks showed group differences (mean ranks: 463.3, 1,193.5, and 539.9 for the CC, CI, and II groups, respectively, \(\chi^2: 1290.186, P < 0.01\)), and a post hoc analysis (Dunnett’s method) revealed significant group differences \((P < 0.05)\). The likelihood for changing answers was 4.84 times higher when both students had individual incorrect responses than when both students had individual correct answers \((OR: 4.84, 95%CI: 1.52–15.43, \chi^2: 8.461, P < 0.01)\).

A comparison of the group responses after the discussions between high- and low-performing students with different individual answers, depending on who had the correct answer (the high- vs. low-performing student), is shown in Fig. 3. The individual testing score for the high-performing students was 67 (SD 13)%; whereas the individual testing score for the low-performing students was 50 (SD 14)%). A total of 466 group responses were computed (344 high-performing correct and low-performing incorrect and 122 high-performing incorrect and low-performing correct). When the group member with the individual correct response was the high-performing student, the group responses were correct 77% of the time, and when the group member with the individual correct response was the low-performing student, the group answers were correct 80% of the time. A \(\chi^2\)-test showed no differences between groups \((\chi^2: 0.367, df: 1, P = 0.54)\).

Positive effects (individual incorrect responses changed to group correct responses) and negative effects of collaborative group testing (individual correct responses changed to group incorrect responses) are shown in Fig. 4. Specifically, a total of 640 answers changed: 77% (494 of 640) to correct and 23% (146 of 640) to incorrect. Therefore, the positive effects of collaboration were much higher than the negative effects (77% vs. 23%, respectively, \(P < 0.05\)). In addition, most of the changed responses (84%, 536 of 640) occurred when group members had different answers (the CI group). Finally, the positive effects of collaborative group testing (77% of the total effect, 494 of 640) were produced by groups with one student having a correct response and the other having an incorrect answer (64%, 412 of 640) and by groups with both members having individual incorrect responses (13%, 82 of 640). Importantly, the negative effects (23% of the total effect, 146 of 640) occurred in groups with one student with a correct answer and the other student with an incorrect response (19%, 124 of 640) and in groups with both students having correct answers (3%, 22 of 640). A Kruskal-Wallis nonparametric one-way ANOVA on ranks showed group differences (mean ranks: 1194.2, 1911.0, and 1347.3 for the CC, CI, and II groups, respectively, \(\chi^2: 838.141, df: 2, P < 0.01\)), and a post hoc analysis with Dunnett’s method revealed significant differences for all groups \((P < 0.05)\).

**DISCUSSION**

We were interested in understanding the interaction characteristics that give rise to the effects of collaborative group testing. Several important group dynamics were revealed. Specifically, 22% of the total individual answers were changed during collaborative group testing (Fig. 1). Most of the changes occurred in groups when students initially had incorrect responses. Few initial correct answers were changed to incorrect responses. Thus, it is more likely to convince someone who is incorrect than someone who is correct. These findings are consistent with recent work documenting the effect of peer instruction \((5, 8)\). We also extended these results by documenting that when the discussion was between group members with different initial answers, the student with the correct response generally convinced their partner with the incorrect answer to change (Fig. 2). Therefore, collaborative group testing is more beneficial when group members have different initial answers (e.g., one correct and the other incorrect) than when partners have the same initial answers. This may be due, in part, to the discussion required to reach an agreement, to convince partners by explaining as clearly as possible the decision and the rationale for the choice. In this sense, it is reported that the generation of explanations (for oneself and for others) is a powerful means to increase knowledge \((2)\).

Importantly, both high- and low-performing students, when they are correct, can generally convince their peers with incorrect responses to change to correct answers. Therefore, based on our data (Fig. 3), we suggest that during student discussions, it is more important to have the correct answer than to be the high-performing student of the group. Thus, educators should not be concerned that low-performing students are “carried” by their high-performing peers or may “defer” to their high-performing partners. Therefore, these results document that collaboration is important for determining the correct answer independent of the previous performance of the students. In addition, these data suggest that collaborative group testing is beneficial for all students (high performing and low performing) since the student with the correct answer prevails most of the time and the changes in responses after the discussions mostly go to the correct answer.

The positive effects of collaborative group testing were much higher than the negative effects (Fig. 4). Similarly, we \((8)\) recently reported that the magnitude of the positive effects of peer instruction were 5.6 times higher than the magnitude of the negative effects of peer instruction \((22\% vs. 4\%, P < 0.01)\) and that the positive effects were observed in the groups of students with initial incorrect answers \((8)\). The positive effects (Fig. 4) were due to students who changed their individual answers to correct responses after discussing questions with peers with individual correct and also, to a lesser extent, after discussing questions with partners with individual incorrect answers. Thus, surprisingly, two students with initial incorrect answers can get the correct response through discussion. As far as we know, this is a completely new finding concerning the benefits of collaboration.

In brief, we propose that student feedback during collaborative group testing is beneficial for most students (low performing and high performing) since the peer having the correct answer prevails most of the times. Thus, changes in responses after peer discussion go mostly to the correct response so that, as stated by Crouch and Mazur \((5)\), there would always be an increase and never a decrease in the number of correct responses after student discussion.
We walked around the room and observed the students interacting during the collaborative testing procedures. It was clear that the students were engaged and committed to the process. Students were observed listening, providing constructive feedback, and reflecting on their learning. We observed students developing leadership, decision-making, communication, and conflict resolution skills while gaining the mutual respect of peers. The level of discussion during collaborative testing was high, and the students were focused and seriously engaged in the discussion. The students were invested in the discussion, and we observed a high level of involvement from the students. Students were encouraged to provide support for their decisions during the collaborative testing, and the students discussed all the questions, not only those on which their answers differed.

One of the learning theories that provides the foundation for collaborative learning is the zone of proximal development, as developed by Vygotsky (13) and defined as follows:

The Zone of Proximal Development is the distance between the actual developmental level, as determined by independent problem solving skills, and the level of potential development, as determined through problem solving under adult guidance or in collaboration with more capable peers.

The gap between actual and potential development levels changes continuously during the learning process.

The actual development level is a retrospective view of student learning, while the potential development level is a prospective view of student learning. Specifically, the theory suggests that what students can do with assistance today, they will be able to do alone in the future. That is, self review by the student is the internalization of peer review.

Therefore, students are capable of performing at higher intellectual levels when asked to work in collaborative situations than when asked to work individually (13).

During traditional testing, students sit silently in the classroom and answer questions during the test session. The students receive their grades within a few days of the exam. In many departments, some students, especially those failing the exam, visit the instructor with a message similar to the following: “oh, I failed by only one or two questions.” And, many of the students use the valuable contact time with the instructor trying to get points needed to pass the test. Obviously, little learning results from this student-instructor interaction. In contrast, with the format described in this article, correct and incorrect answers along with the rationale supporting the answers are discussed with all the students in the classroom. In this setting, the instructor can reach all the students to provide constructive feedback. This environment is less stressful to both the students and the instructor and is much more constructive for the student-instructor relationship. Furthermore, students in this setting are eager to know if their answers and the reasons for their answers are correct. Thus, the role of the instructor, acting as a coach, facilitator, or model, is to help students uncover the rationale behind the correct and incorrect answers. In addition, it is also beneficial for the instructor because many undiagnosed misconceptions and faulty mental models are uncovered in the process.

It is advisable that the instructor provides immediate corrective feedback when collaborative group testing is implemented for many reasons. First, students may change some of their correct answers to incorrect answers (negative effect). Second, some students with the correct answer may fail to provide 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.

Limitations

There are some potential limitations that must be considered when reviewing the results from this study. First, potential bias could have entered the study since partners were self-selected. However, this may not be a significant problem since both high- and low-performing students were well represented and because all the students volunteered to participate in the study. In addition, a high proportion (~70%) of the students attended all three exams. Finally, some students may feel uncomfortable when they are required to work collaboratively with unknown partners. Therefore, we let the students choose their partners in an effort to keep our students as comfortable as possible during the exams.

From our observations, we believe that all of the questions were discussed by the students. It is possible, however, that the students focused more time on the questions in which their initial answers were different than on the questions where the students had the same responses. Finally, the results document that it is more important to get the correct answer than to be the higher performer of the group.

Second, we did not determine the effect of the different response times on the likelihood for change from the incorrect response to the correct response. However, since we asked for a prediction of a systems response (increase/decrease/no change) to a perturbation, if the causal relationship among the modified variable and the other variables of the system is known, the question can be answered quickly. Since most of the questions were answered correctly, we think that 1.5 min to discuss the response of a question with a peer is adequate. In addition, the students previously had 2 min to work alone (individual testing) on the same questions before the group testing, so they knew the questions and the three possible answers (always the same: increase/decrease/no change). In addition, similar qualitative problems were used in a peer instruction format with a similar amount of time (8).

Finally, we have no measure of the degree of confidence on the student’s answers. This certainly would have added significantly to the study and will be included in future studies.

Conclusions

Collaborative group testing is beneficial because it is much more likely to change the response of students with initial incorrect answers than to change the response of students with initial correct answers. Collaborative group testing is more beneficial when group members have different initial answers than when they have initial same answers. In addition, when group members have different individual answers, more changes go to correct responses than to incorrect responses, and when group members have the same answer (both correct or both incorrect), the likelihood for change is higher for both
students with incorrect answers than for both partners with correct answers.

Importantly, both high- and low-performing students, when they are correct, can generally convince their peers with incorrect responses to change to correct responses. Thus, educators should not be concerned that low-performing students are “carried” by or “defer” their high-performing peers.

Finally, the positive effects of collaborative group testing were much higher than the negative effects. The positive effects were due to students who change to correct responses after discussing with peers having the correct answer (major effect) and also with peers having the incorrect answer (minor effect).

ACKNOWLEDGMENTS

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

REFERENCES