

JTAR

EDITORS

GIL NAIZER

APRIL SANDERS

LAURA ISBELL

TAMI MORTON

SUSAN WILLIAMS



Journal of Teacher Action Research - Volume 7, Issue 3, 2021,
practicalteacherresearch.com, ISSN # 2332-2233

© JTAR. All Rights Reserved

JTAR

Journal of Teacher Action Research Volume 7, Issue 3, 2021

| | |
|--|----|
| Helping Struggling Writers Through Effective Spelling and Writing Strategies Xochitl Morales | 4 |
| Improving Pronunciation Skills of Grade 6 Pupils Through Oral Drills Antonio Jr. T. Enerio | 27 |
| Incorporating Autonomy in an Analytics Module: Visualizing Self-Directed Learning Hui Teng Chia | 41 |
| Spanish Vocabulary Acquisition by English Speakers Using Spaced-Repetition Algorithms and Mnemonics Jon-Erik G. Storm | 58 |
| Collaborative Testing Impacts Student Achievement and Test Anxiety for Advanced Placement Environmental Science Students Helen Bremert S. B. Boesdorfer | 66 |
| Investigating Co-Teaching for Impact on Academic Engagement: Best Practices for English Subject Learners in a Bilingual Elementary School Kyle M. Daley | 87 |



About the Journal

Founded in 2013, the Journal of Teacher Action Research (ISSN: 2332-2233) is a peer-reviewed online journal indexed with EBSCO that seeks practical research that can be implemented in Pre-Kindergarten through Post-Secondary classrooms. The primary function of this journal is to provide classroom teachers and researchers a means for sharing classroom practices.

The journal accepts articles for peer-review that describe classroom practice which positively impacts student learning. We define teacher action research as teachers (at all levels) studying their practice and/or their students' learning in a methodical way in order to inform classroom practice. Articles submitted to the journal should demonstrate an action research focus with intent to improve the author's practice.

Editorial Team

Co-Editors

Gilbert Naizer, Ph.D.
Texas A&M University-Commerce

April Sanders, Ph.D.
Texas A&M University-Commerce

Associate Editors

Laura Isbell, Ph.D.
Texas A&M University-Commerce

Tami Morton, Ph.D.
Texas A&M University-Commerce

Susan Williams
Texas A&M University-Commerce

Production Editor and Webmaster

Chase Young, Ph.D.
Sam Houston State University

www.practicalteacherresearch.com

COLLABORATIVE TESTING IMPACTS STUDENT ACHIEVEMENT AND TEST ANXIETY FOR ADVANCED PLACEMENT ENVIRONMENTAL SCIENCE STUDENTS

Helen Bremert
Sydney University

S.B. Boesdorfer
Illinois State University

Abstract Assessment is a concern in high schools because summative assessments can impact a student's potential. This researcher studied how collaborative assessments increased student achievement and reduced test-taking anxiety after observing students struggling when undertaking high stakes testing. I draw on previous research into collaboration to demonstrate the benefits of collaborative assessment. The study occurred in a suburban high school AP Environmental science class. Students undertook six assessments, three individual and three collaborative high stakes tests in heterogeneous groups. A repeated-measures ANOVA test and paired t-tests were conducted to determine the differences between the testing methods. An analysis of the findings indicates there is a significant difference between students undertaking individual and collaborative tests. In the post-test Qualtrics survey, many students showed an increased understanding of the content and reduced their test-taking anxiety. The paper argues that collaborative assessment has many benefits to students; it increases students' achievement and understanding of the concepts as they utilize reasoning and argumentation to defend their answers. Additionally, the negative impacts associated with collaborative testing can be easily addressed.

Keywords: teacher action research, collaborative assessment, high school, test anxiety, collaboration

Introduction

Education in America has changed with the recognition that students must be productive citizens in a rapidly changing world (Achieve, 2010). Scientists and other professionals worldwide require employees to collaborate, think critically, and problem-solve effectively

(Ngotngamwong, 2014). To this end, the *Next Generation Science Standards* (NGSS) has a vision of ensuring that upon graduation, students will have the necessary skills to be practical and rational thinkers (NGSS Lead States, 2013). To achieve this, the NGSS recommends inquiry, collaborative, and evidence-based instruction across a wide range of science fields (Nairman & Chrispeels, 2016; NGSS Lead States, 2013).

Collaborative learning is pedagogical method that promotes an active classroom learning environment where the students form pairs or groups to accomplish tasks (Meseke, Nafziger & Meseke, 2010). This instruction method's advantages include, increased conceptual understanding, retention, problem-solving, and critical thinking skills (Gilley & Clarkston, 2014; Pandey & Kapitanoff, 2011; Siegal, Roberts, Freyermuth, Witzig & Izci, 2015). Furthermore, collaboration promotes heightened intrinsic motivation, interpersonal skills, and students' ability to engage in evidence-based argumentation (Guiliodori, Lujan & DiCarlo, 2008; Meseke et al., 2010; Zipp, 2007), which are all necessary skills for the current workforce and supported by NGSS.

During high school, many students engage in collaborative tasks and formative assessments; however, when students undertake summative assessments, they are generally taken individually (Siegal et al., 2015). Quizzes and examinations occur because they can assess students quickly and over many learning units (Rao, Collins & DiCarlo, 2002). Additionally, this method enables the educator to determine students' academic strengths and weaknesses and holds schools accountable. Nonetheless, individual tests have several disadvantages, including lowering intrinsic motivation, using only information-recall type questions, underperformance due to outside factors, and increased test anxiety (Breedlove, Burkett & Winfield, 2004). An alternative testing method used to combat issues with individual testing is collaborative assessments. One such testing issue is the reduction test taking anxiety (Haberyan & Barnett, 2010; Ngongamwong, 2014; Siegal et al., 2015) due to students being able to "have the emotional and intellectual support" (Rao, Collins & DiCarlo, 2002, p. 38) of their peers. Moreover, group testing may improve exam performance (Gilley & Clarkston, 2014; Vogler & Robinson, 2016) and promote positive student attitudes (Haberyan & Barnett, 2010). In a group testing environment, the students must discuss questions and answers, thereby filling in knowledge gaps, leading to greater understanding and greater retention of the material (Kapitanoff, 2009; Vogler & Robinson, 2016). Much of the research on the effects of collaborative testing (Breedlove et al., 2004; Gilley & Clarkston, 2014; Meseke et al., 2010; Rao et al., 2002; Siegal et al., 2015; Vogler & Robinson, 2016) has focused on undergraduate students. Despite the positive outcomes of collaborative testing at the collegiate level, there is little data on how high school students' summative performance could improve through collaborative testing. This study adds to the literature by determining if collaborative testing enhances students' understanding and lowers test anxiety in a high school setting.

Literature Review

Assessment. Assessment is an essential part of education as it enables educators to collect information about students' academic learning, reasoning skills, and attitudes (Rao et al.,

2002). Quizzes and tests are the primary mechanisms used to determine if students have met the course's goals. The evaluation also determines student grades and their advancement into future classes (Giuliodori et al., 2008). Leight et al. (2012) describe how high stakes summative assessments like midterms, finals, and Advanced Placement (A.P.) exams help instructors know whether they have developed the required level of understanding concepts taught. They argue that testing students are the best method to ensure students retain the course material. Although this may be true, students may not understand what they did right or wrong as they do not see the exams. Therefore, students will not be able to determine if their knowledge or the strategies used were effective, reducing the assessment's pedagogical value (Leight et al., 2012; Rao et al., 2002). Additionally, individual testing does not consider disadvantages such as cultural differences, different learning styles, and additional challenges faced by English second language learners (Zapatero et al., 2012). Furthermore, traditional individual assessments do not consider social constructivist theories where students learn best in collaborative classrooms (Giuliodori et al., 2008; Zapatero et al., 2012).

Collaborative learning. Social-cultural theories of learning and teaching, developed by Bruner and Vygotsky in the late nineteen sixties and seventies, construe that students are stimulated to learn and grow through social interactions (Seifert & Sutton, 2009). Constructivism, another learning theory, explains students actively construct their knowledge out of shared experiences, which augment their metacognition (Seifert & Sutton, 2009). Collaborative learning is a pedagogical method that utilizes social-cultural and constructivism to enhance student learning. This approach to education is powerful as students are actively engaged in their learning as they converse with their peers in small groups, exchanging and defending their ideas (Ghaith, 2018; Wanzek et al., 2014). Additionally, through shared learning, students expand their cognitive skills and ideas and develop new attitudes (Meseke et al., 2010).

Furthermore, research has shown many benefits to collaborative learning such as; additional academic and social support, increased student self-esteem, positively affected student achievement, increased student motivation, improved intergroup relations, improved critical and creative thinking, and improved problem-solving skills (Baloche & Brody, 2017; Leight et al., 2012; Meseke et al., 2010; Ngotngamwong, 2014; Rao et al., 2002). For collaborative learning to be effective in the classroom, the educator should be responsible for building teams. The teacher places the students in groups to ensure that they can improve their skills and develop their knowledge (Wanzek et al., 2014). Students retain information better and enhance their understanding of the concepts when they are not homogenous (Wenzel, 2000).

Collaborative testing. Collaborative testing places the students in pairs or small groups of three or four for the test. Once in the group, the students can discuss the questions and then submit an individual test paper or a group test paper (Leight et al., 2012; Meseke et al., 2010; Weimer, 2018). There have been numerous research studies done, mainly with college students, into this assessment method's benefits. Giuliodori et al. (2008)

demonstrated an increase in collaborative test scores for high and low-performing students with the collaborative test groups. Albeit lower-performing students benefitted more. In contrast, Gilley & Clarkston (2014) discovered that all students learned from the collaborative assessments irrespective of their performances. Seigel et al. (2015) undertook a study whereby the group testing occurred first, and then the students took the remainder of the test individually. They discovered that when the exam content is conceptually more straightforward, intermediate and low performing students benefit the most from group testing. Nonetheless, when the concepts are more complicated, all students benefit from collaborative testing. Thus this method enables students to utilize higher order thinking skills, which is a critical aspect of the N.G.S.S. (Gilley & Clarkston, 2014; Meseke et al., 2010). Students' level of retention of the material in collaborative testing has had mixed results. Gilley & Clarkston (2014) found greater retention and understanding of the students' concepts when they undertook group tests instead of individual tests. The students who attempted the group tests retained the information and correctly responded when taking a written pop quiz three days later. In contrast, students who did not participate in the group exam did not increase their retention or understanding of the material (Gilley & Clarkston, 2014). Zipp (2017) undertook a collaborative assessment study where the students completed each exam individually and then were placed in groups to retake the exam. The results indicated that if the students answered correctly in the group test, they responded correctly on the final examination, two months later (Zipp, 2007). Other studies have also indicated that students who take group tests can longer retain the information (Cortright et al., 2003; Rao et al., 2002; Vogler & Robinson, 2016).

In contrast, Leight et al. (2012) and Sandahl (2010) discovered no greater retention occurred between the control group of individual test-takers than the experimental group of collaborative test takers on the final exam. An explanation for this difference may be that students who have had more exposure to the content have a sufficient prior understanding to enable new knowledge to be assimilated and retained (Leight et al., (2012). Despite not showing a difference in retention in the final, Leight et al. (2012) results indicate that cooperative testing increased students' level of engagement and confidence and enhanced their understanding of the concept.

With a better conceptual understanding and retention, collaborative classrooms have many benefits for students, including enhanced participation, increased social skills, and higher critical thinking and problem-solving skills (Baloch & Brody, 2017; Ngotngamwong, 2014). Assessment that uses collaboration between students should then have positive effects on students. Despite not showing a difference in retention in the final, Leight et al. (2012) indicate that cooperative testing increased students' level of engagement and confidence and enhanced their understanding of the concept. Likewise, Ngotngamwong's (2014) study determined most students (over 80%) felt that pair testing was enjoyable, created more outstanding teamwork and cooperation between the students, and ensured they studied harder. Hanshaw (2012) and Rao et al. (2002) and others describe many benefits from this testing method: positive influence on learning, better interpersonal, communication, conflict resolution, and critical thinking skills. Additionally, their results indicated increased metacognitive skills, increased persistence to problem solve, enhanced memory and

retention, and effective listening skills, all vital skills for students (Gilley & Clarkston; Hanshaw 2012; Kapitanoff, 2009; Rao et al. 2002).

Coupled with the above benefits, collaborative testing has been shown to diminish test anxiety in students, as described in more detail below (Breedlove et al., 2004; Cortright et al., 2003; Krispenz & Dickhäuser, 2018). Some studies indicate there are negative consequences to collaborative testing. A typical adverse claim is that students are not preparing for the assessment (Giuliodori et al., 2008). One method used to alleviate this is to ensure the students do not know they are undergoing a collaborative test until they arrive at the classroom (Rao et al., 2002). Moreover, some students have said that they could not reach a consensus on the answers or had mismatched partners (Ngotngamwong, 2014; Zipp, 2017). To overcome this effect, students could hand in an individual copy of the assessment and consequently do not need to reach a consensus (Ngotngamwong, 2014). *Test Anxiety.* Test anxiety relates to students' emotions when studying for and taking an exam (Krispenz & Dickhäuser, 2018). Test anxiety can have many consequences for students, including a cognitive component whereby the student cannot retrieve the information, the students may have difficulties in organizing and retaining information, or the students feel overcome by a fear of failure (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018). The physiological reactions of test anxiety may comprise trembling, palpitations, sweating, dizziness, and nausea, impacting student well-being (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018). Test anxiety can also affect the students' mental health as it can lower self-esteem, increase feelings of helplessness and insecurity, erode confidence, and diminish motivation (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018). Equally important is the knowledge that test anxiety impacts student academic achievement, affecting each student's educational and employment prospects (Krispenz & Dickhäuser, 2018).

Multiple studies using collaborative assessments have demonstrated that this assessment method reduces test anxiety in students (Breedlove et al., 2004; Leight et al., 2012; Meseke et al., 2010; Pandey, C. & Kapitanoff, S., 2011). Reduction in test anxiety may be brought about by students sharing prior learned information, discussing questions and answers, and having intellectual support from their peers (Breedlove et al., 2004; Rao et al., 2002).

Methodology

Statement of Purpose. This action study was undertaken in an Advanced Placement Environmental science classroom after the researcher noticed students benefited when undertaking collaborative work. The students described this method as enabling them to get different perspectives and develop their understanding of the concepts. Moreover, it was observed by the researcher that students felt very stressed when undertaking summative tests, and hence some students were not able to perform to their ability. Therefore, the researcher wanted to determine if collaborative testing could be a method to increase understanding and reduce test-taking anxiety in a high school.

The benefits of assessment in groups have been studied numerously over the years, generally in college science classes, as cited in the studies described above. However, for high schools, there have been very few studies undertaken. Therefore, this current study was conducted to determine if the impact of collaborative assessments in a high school was similar to the collegiate findings, including academic performance and reduction of test anxiety. This study, therefore, was guided by the following research question:

- Does collaborative testing in a high school A.P. Environmental Science course positively impact student test outcomes?
- Does collaborative testing in a high school A.P. Environmental Science course lower student test anxiety?
- How do students view collaborative testing in terms of preparing them for the A.P. exam?

Course Structure and Content. This study occurred over the fall, winter, and spring semesters of the 2018 -2019 school year. The participants included forty-one high school students in two sections of an Advanced Placement Environmental Science course taught by different teachers. The student population for this study consisted of thirty-eight seniors and three juniors (N=41), of which sixteen (39%) were males, and twenty-five (61%) were females. Four students missed a collaborative test; therefore, all of their testing data were excluded from analysis (N=37), but their survey results were still analyzed.

The A.P. Environmental Science is a full-year course, meeting for four sixty-minute periods per week. These periods consisted of a mix of lectures, collaborative laboratory assignments, and inquiry-based work. During November, the students participated in a group quiz on Renewable Energy, allowing them to be familiar with undertaking a collaborative examination. The groups were teacher assigned based on previous test scores, gender, and grade level. The makeup of these groups is outlined in Appendix A. During this research period, the topics covered were Renewable Energy, Climate and Biomes, Indoor and Outdoor Pollution, Climate Change, and Agriculture and Food. Summative assessments, which were identical for both classes, for each topic, were made up of 21-25 multiple choice questions, each with five answers and four to six short answer questions (20 minutes). The questions were of varying complexity and sourced from previous A.P. Environmental Science examinations run by the College Board, U.S.A.

Research Design. This study was undertaken to determine if collaborative testing could improve student achievement in a high school APES class. To accomplish this, the students end of topic tests were used to generate data to compare the individual and collaborative test results. Students attempted three individual and three collaborative tests, enabling a comparison between the student's individual and collaborative testing grades (Giuliodori et al., 2008). The classroom teacher determined the heterogeneous testing groups' assignments based on students' prior academic achievement, gender, motivation, and ability to stay on task (Wanzek et al., 2014). To ensure randomness, the students were unaware of their group assignments until entering the testing room (Meseke et al., 2010). Due to the small class size, the groups consisted of three or four participants (See Table 1). Additionally, because of gender inequality in the study group, there were more females in

some groups than males; however, every group included at least one female and one male. Table 1 describes the number of participants in each group for collaborative tests.

Table 1: Test topic and group makeup

| Test Topic | Number of groups | Group makeup | Total Students | Students in analysis |
|----------------------|------------------|-----------------------------------|----------------|----------------------|
| Renewable energy | 13 | 11 groups of 3 Two groups of 4 | 41 | 37 |
| Climate and Biomes | Individual test | | 41 | 37 |
| Indoor pollutants | Individual test | | 41 | 37 |
| Climate change | 13 | 8 groups of 3 4 groups 4 | 40 | 37 |
| Outdoor Pollutants | Individual test | | 41 | 37 |
| Agriculture and Food | 13 | 9 groups of 3 3 group 4 | 39 | 37 |

Note. The groups were determined by gender, previous test scores, and class rank.

The collaborative group test was undertaken under the same conditions as the individual test to determine any improvements between individual and collaborative testing (Wanzek et al., 2014). The students were allotted 50 to 55 minutes to answer the questions. Students did not know if a test would be collaborative or individual until they were in the classroom, ensuring they studied for the test. Each group submitted their own exam paper with all of the group names on the document (Leight et al., 2012; Nanzek et al., 2014; Wanzek et al., 2014). Moreover, the students submitted their test papers in the collaborative testing phase, thus enabling them to change their responses if they could not reach a consensus on an answer (Ngotngamwong, 2014). To ensure grading consistently between the teachers the open-ended questions were graded using the college boards APES rubric for each test. To answer the second quantitative question about whether collaborative testing affects students' test anxiety, they undertook an anonymous online survey distributed through Qualtrics two days after the final summative assessment. The survey questions included collaborative testing and test-taking anxiety (See Appendix B for survey questions) adapted for high school students from questions in Cortright et al. (2003), Hanshaw (2012), and

Ngotngamwong's (2014). Lastly, the researcher undertook an anonymous survey to determine if the students felt that collaborative testing aided their retention when undertaking their final exam (Appendix C for survey questions).

Data Analysis. To answer research question 1, descriptive statistics including averages from percentage scores and standard deviations were calculated for all six exams. The data collected only included students present at testing for both individual and collaborative tests (N=37). The data gathered was analyzed by a one-way repeated measures ANOVA test and paired-sample t-tests through IBM SPSS to ascertain the student differences between group and individual testing. The assessments' results were examined in two ways; the first was a repeated measure ANOVA that looked at the differences between the three independent and the three collaborative tests. The second test was a paired t-test to compare the averages of the individual and group tests. Additionally, a paired t-test was undertaken to investigate whether there were overall differences between the two tests' averages (Shier, 2004).

After the testing period ended, students were asked to complete an online survey (modified from Hanshaw, 2012 & Meseke et al., 2010) to determine their perceptions of collaborative testing on understanding and whether it affects their test-taking anxiety. The survey asked 19 questions and was based on the 0-5 Likert scale (Leight et al., 2012). All students (N=41) undertook the survey; a subset of the responses is represented in Table 4. Lastly, after the A.P. exam in May, students were informally questioned to determine if collaborative testing affected their understanding and retention of the concepts. Their responses are listed in Table 5.

Results

Student Achievement. For the three independent tests, the means (See Figure 1) are as follows with the standard deviation in parentheses, 78.22 (11.15), 77.41 (13.14), 78.62 (11.20). The Mauchly's sphericity tests had been met, as detailed in Table 2 (df = 2, Sig. = 0.856), $F(2, 72) = 0.156$, $p = 0.64$, indicating there is not a significant difference between the three individual test scores. Regarding the collaborative tests, there was a significant difference in the means (See Figure 1), 85.08 (11.63), 89.73 (5.47), and 94.19 (5.76), further validated by the results of the ANOVA test, whereby Mauchley's test of sphericity indicate there is a significant difference between the collaborative tests ($p < 0.000$). Due to the sphericity not being met, a Greenhouse-Geisser correction was applied (Table 2).

Table 2: ANOVA Data for Independent and Collaborative Tests

| Variable | N | Mauchley's Sphericity | df | F | p | Greenhouse-Geisser |
|-------------|----|-----------------------|----|-------|------|--------------------|
| Independent | 37 | 0.640 | 2 | 0.156 | 0.64 | |

| | | | | | | |
|---------------|----|-------|-------|--------|-------|------|
| Collaborative | 37 | 0.000 | 1.448 | 11.379 | 0.001 | 0.00 |
|---------------|----|-------|-------|--------|-------|------|

Note. The Greenhouse -Geisser test was added as the collaborative test did not meet the standard for sphericity.

To determine if there was a difference between the averages of the independent and group test scores, a paired t-test was undertaken. The t-test was between the group and individual test 1, group and individual test 2, and group and individual test 3. The results are displayed below in Table 3. There is a significant difference in test scores between collaborative and individual testing. This is also depicted in the difference in means between the tests and illustrated in Figure 1. The box plot shows that students who undertook collaborative testing generally had increased scores compared to their individual test scores.

Table 3. Paired t-test between the averages of the three individual and group tests

| Variable | Mean | Std. Deviation | t | df | Sig. (2-tailed) |
|----------------------------|--------|----------------|--------|----|-----------------|
| Pair 1- ind. Test 1- gp 1 | -6.86 | 16.13 | -2.588 | 36 | .014 |
| Pair 2 – ind Test 2 - gp 2 | -12.32 | 13.38 | -5.598 | 36 | .000 |
| Pair 3 – ind Test 3 - gp 3 | -15.57 | 12.21 | -7.756 | 36 | .000 |

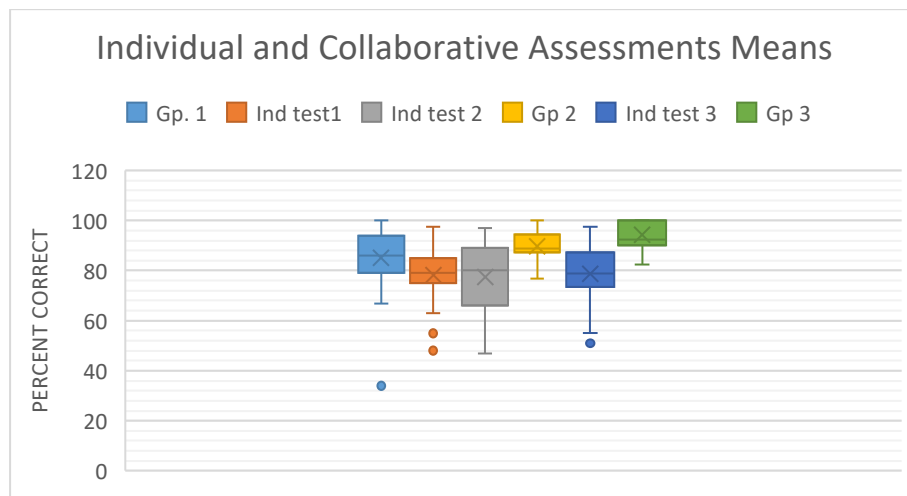


Figure 1. Individual (Ind test) and group (Gp) test results from the A.P. Environmental Science Course were displayed when they took each test.

Figure 2 below displays the averages of the three individual and collaborative tests for each student. Along with the t-test data, the graph indicates an overall increase in student test scores compared between individual and group tests, which is very apparent in individuals 3, 9, 25, 27, 30, and 36. Although individuals 7, 8, and 10 had slightly better individual test scores than group test scores.

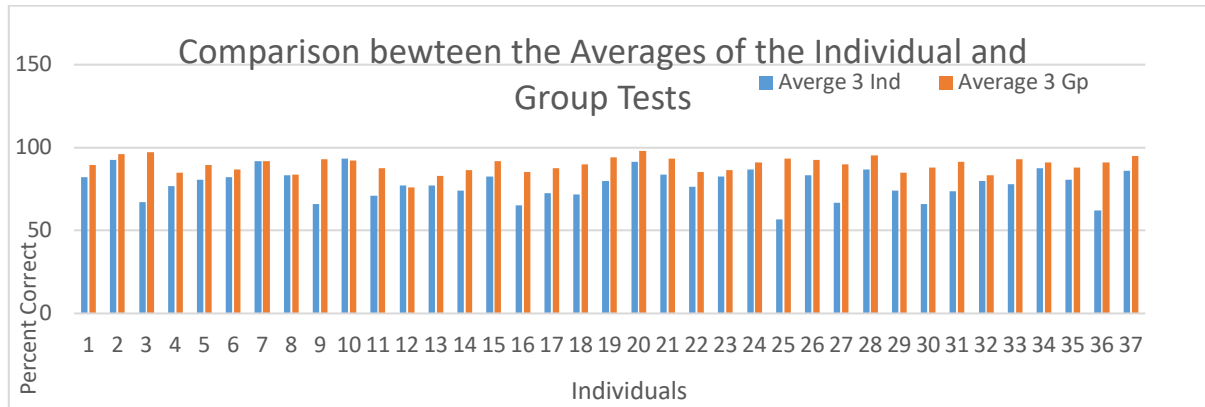


Figure 2. Comparison of each student's average individual and collaborative percent test score.

Student Test Anxiety and Preparation for A.P. Exam. To answer the second research, question all students (N=41) in the course were anonymously surveyed. The survey questions and responses are described below and in Table 4. 56% of students felt that individual testing increased their anxiety, whereas 26% said individual testing did not impact their anxiety levels. The following students' responses to the open-ended question represent most students' reasons for increased anxiety on the individual tests:

- It was very stressful because even if I knew the right answer, I second guessed it.
- Doing an individual test definitely increased my anxiety because I had no one backing up my answer or contradicting me on why the answer I picked was wrong.
- I felt that taking a test individually stresses me out more because I could be the only one getting the bad grade.
- I feel more anxious during individual tests.
- It is more stressful to take the individual section since you cannot discuss and get the opinions of your classmates.

On the other hand, the majority of students (87.88%) felt that undertaking collaborative testing lowered their anxiety levels, reasons for this from the open-ended questions included: regarding confidence in the group's answers, ability to discuss the questions, and understanding if they were not sure some member of the group would have the answer. The majority (81.81%) of students felt that working in a group was less stressful than undertaking individual tests. The student's comments below provide some of their reasoning:

- My group members were cooperative, we were all respectful and kind to each other, and if we disagreed, we just put a separate answer, and I get to interact with people and understand the material a lot more.
- I didn't feel very anxious after the test because if I didn't know the answer to a question someone else in my group did, which helped me feel like we as a group would get a good grade. I also felt that because it was a group test I wasn't the only getting the grade if it turned out to be bad."
- I feel as if it was less stressful since it was a group test and that I was more confident in my answers and more relaxed about getting my grade back.

9.09% of students said that it increased their anxiety levels; their reasons were due to always being stressed before a test. Besides, as one student states, Sometimes they disagreed with me, and I didn't want to write something different, My group was divided over many questions, and neither side was willing to accept that they may be wrong, and I find myself being able to work to a more efficient degree when coupled with my peers; however, I am often look towards as the "intelligent" one of the group, and I feel slightly stressed due to the position of power I am involuntarily elected to.

Table 4. Survey of student's anxiety levels related to individual and collaborative tests (N=41).

| Questions | Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly agree |
|---|-------------------|----------|----------------------------|-------|----------------|
| I felt anxious after individual test | 14.6% | 12.2% | 17.1% | 31.7% | 24.4% |
| I felt anxious after the group test | 61% | 24.4% | 7.3% | 4.9% | 2.4% |
| Group tests increased my anxiety | 73.2% | 14.6% | 2.4% | 9.8% | 0 |
| I found working in a group stressful | 46.3% | 36.6% | 4.9% | 4.9% | 7.3% |
| Working in a group helped my understanding of the content | 4.9% | 2.5% | 7.3% | 39% | 46.3% |

Lastly, the students were asked if collaborative testing increased their understanding of the content. Most students (87.87%) felt that their knowledge of the content increased, while 9.09% indicated it neither helped nor hindered their understanding, and one student thought it negatively impacted their ability. Representative student comments include:

- Talking about the problems helped me understand them better.
- Combining the knowledge of everyone increased my knowledge on different topics.
- By working with a group, we were able to discuss our reasoning and come to a correct answer.
- By bouncing our ideas off of one another, I felt all of the previous information I ascertained re-enter my mind and reposition itself into a more organized whole.

After the A.P. exam, 63.3% of students felt that this method of assessment aided their retention of the material, while 33.3% felt that it only helped their retention a little. Further, 66.7% of students stated a difference in retaining the information from the different testing methods (See Table 5).

Table 5: Group testing and retention of material on the A.P. exam. (N=31)

| Did you feel that you retained the information that was tested in the group tests? | Definitely | A little | No |
|--|------------|----------|----|
|--|------------|----------|----|

| | | | |
|---|------------|------------|----------|
| | 19 (63.3%) | 10 (33.3%) | 2 (6.7%) |
| Did you feel there was a difference in retention between the material from the group or individual tests? | Yes | No | Maybe |
| | 66.7% | 6.6% | 26.7% |

Note. This survey was given post-A.P. exam, only 31 students surveyed as the remaining were not in class at the time.

Limitations

The amount of time for the parental approvals to be returned impacted the time available to undertake the tests. In turn, this caused the researcher to have to speed up the testing process. Hence, students could determine when the collaborative or individual tests would be undertaken before walking into the room. It could not be as random as initially planned. Some students were also removed from the testing data but not survey data because the survey was undertaken anonymously, and the researcher did not know the individual student responses.

Another limitation may include that the study was conducted in two science classrooms with different teachers. Therefore, the other class students may have known when the group tests were going to be undertaken and the content. Knowing the test's content could influence the test scores as the students understand what subject matter to focus on, leading to inflated test scores. Additionally, knowing when the group tests were undertaken could impact the student's motivation to study. This, in effect, did occur as some students, after the last test, did indicate to the researcher that knowing the content and that there was to be a group test caused them to study less and rely more on other people in the groups. Another limitation would be biased when responding to the questions; this could reduce the results' reliability. Due to the students in the survey and last informal questioning responding with answers they think the researcher wanted to have. Lastly, this study did not measure the effects of collaborative testing on retention. Retention was informally discussed with the students.

Discussion

The purpose of this study was to determine whether there were differences between individual and collaborative test scores and levels of test anxiety in students undertaking an A.P. environmental science course. This study's findings support the thesis that most students perform better when undertaking collaborative rather than individual tests (Table 3), as shown in previous research (Giulidori et al., 2008; Haberyan & Barnett, 2010; Rao et al., 2002). The individual repeated measures test means were not substantially different from each other ($p = 0.64$), indicating there were no significant differences between the mean student scores of the three individual tests. Regarding the group tests, all three means were higher than the individual test means, although the third group test had a significantly

higher mean (94.19) than the previous two group tests (85.08 and 89.73, $p < 0.001$). There are various reasons as to why this increase may have occurred. The students in this study had not undertaken collaborative testing previously; therefore, it may have taken them some time to become accustomed to working in a group setting. Moreover, the dynamics of the different groups may have impacted the student's ability to stay on task in the first collaborative test, as they may have been distracted by working with friends or, conversely, with individuals they did not know very well. Additionally, as the students became accustomed to collaborative tests, they may have felt that if they undertook some effort, then as a group, they all could do well on the assessment.

Furthermore, when comparing the individual test to the group test, the results demonstrate a considerable difference between the two testing methods. The group tests had a significant average increase (See Table 2 and Figure 3). These and prior study results lead us to conclude that collaborative testing can facilitate student learning (Leight et al., 2012; Ngotngamwong, 2014; Rao et al., 2002; Vogler & Robinson, 2016). Collaborative testing can facilitate learning as students analyze and discuss the questions and explain their reasoning to each other. This knowledge sharing can lead to greater understanding (Rao et al., 2002). Many of the students in our survey indicate this was the case; they felt that by discussing the questions, they understood the concepts in greater depth, similar to Siegel et al. (2015) findings.

Moreover, improved testing scores could be due to students stimulating their thinking through activating prior understanding. When discussing each question, students remember what they have learned in the past related to the problem, thereby building their knowledge (Leight et al., 2012; Siegel et al., 2015). This was evident in the student's responses to the survey question regarding whether group testing enhanced their understanding of the concepts. Many students stated that discussing the group's problems gave them a deeper understanding of the ideas (See survey comments). While working in their groups, most students were engaged in discussions and felt they contributed equally. However, as noted in the student comment above, a student felt the group relied on one individual. The reliance on specific individuals to "carry the team" has been noted in other research (Rao et al., 2002; Seigel et al., 2015). In this study, as with Rao et al., (2002) study, the researcher tried to alleviate this concern by randomly selecting the individuals in each group based on the criteria mentioned above. Collaborative testing can enhance student motivation to study. In this paper, students were not told what type of test they would undertake until they were in the classroom; therefore, the researcher surmises the students studied how they typically would for a test. However, this study did not test whether students would study more or less if they were made aware of the test structure beforehand.

The survey results agree with many other researchers that collaborative testing can lead to lower anxiety levels in students (e.g., Cantwell et al., 2016; Pandey & Kapitanoff, 2011; Seigel et al., 2015). In this study, many students reported that undertaking individual tests increased their anxiety, however, most felt this was negated by working with their peers (see table 4 and student comments).

Student beliefs in how they perform on tests can negatively impact their ability to access their working memory and lower their grades (Krispenz & Dickhäuser, 2018). The students reported that working in groups enabled them to feel that if they had forgotten a fact, the other team members would know the answer; however, when working individually, they were not able to discuss the solutions, and therefore their anxiety levels would increase pre and during the test. Concerning the collaborative tests, 87% of students felt that working in collaborative groups significantly lowered their anxiety. The lessening of test anxiety documented by most students may have enabled students to focus more on the questions at hand rather than how they will perform on the test, thus increasing their participation and consequent grades (Hanshaw, 2012).

There are some concerns related to collaborative testing. Siegel et al. (2015) and Haberyan & Barnett (2012) found that students felt some groups were more unequal in ability and group dynamics. In this study, the groups changed each test, and the students did not know before the test which group they were going to be placed in. Nevertheless, during in-class observations, some groups did not work effectively, as is evident by the negative comments above. This may be due to the dynamics of high school senior social interactions and differences in personalities (Haberyan & Barnett, 2012).

Furthermore, "social loafing" may have occurred, where one or more group members do not participate in the discussions, which may lessen this assessment method (Rao et al., 2002). Social loafing was seen in this study by a few students, as mentioned in the comments above. Nonetheless, most students (89.7%) felt they each member contributed equally to the group. Another concern relates to students who understand the material in-depth compared to other students, leading to an inflation in grades (Cantwell et al., 2016; Siegel et al., 2015). Figure 2 illustrates how some students' grades increased more significantly than others compared to their tests. In this study, there were five students (12%) with notable grade increases. Grade inflation may impact the higher-performing students as their grades do not increase as much as the lower performing students. Additionally, if the students use these grades to then move into more challenging classes, they may lack the ability to perform well (Cantwell et al., 2016). Grade inflation can be mitigated by ensuring no more than 15% total worth is assigned for the collaborative test section. Moreover, Cantwell et al., (2014) proposed that if a student fails the individual portion, they do not have their group scores added or did not participate.

A more significant issue related to group testing that was not quantitatively addressed by this researcher was student retention and whether this impacts more comprehensive exams. A significant portion (96.6%) of students indicate that collaborative testing helped them retain the information. However, the literature is divided as to whether this method of assessment aids student retention. Some studies, such as those undertaken by Bloom (2009) and Cortright et al. (2003), note some improvement in students' retention. On the other hand, Leight et al. (2012) and Sandahl's (2010) studies indicate little retention by their students. Further research is required to ascertain whether collaborative testing enhances retention in students.

Conclusion

Collaborative testing provides an alternative method to assess students and has had positive results in this and other studies (Breedlove et al., 2004; Leight et al., 2012; Meseke et al., 2010; Seigel et al. 2015). This method positively impacted student test outcomes for most students in the study and provided students with opportunities to think and cultivate different viewpoints critically. Moreover, results from the survey on group assessments indicate that students' test-taking anxiety was reduced.

Currently, in secondary education, there is a reform movement whereby collaborative learning practices are being utilized more often to promote deeper engagement and understanding of the concepts (Nariman and Chrispeels, 2016). However, minimal collaborative testing is attempted in high schools. As an educator, the researcher believes this assessment method should be utilized more often in secondary education. The academic and social benefits of collaborative assessments for high school students are many. This method allowed the students to discuss questions. They had to defend their opinions and listen to other perspectives. In doing so, the students reasoned through the questions, leading them to understand the concept. Furthermore, collaboration in heterogeneous groups enabled most students to lessen anxiety related to test-taking, which can negatively impact student achievement. The issues concerning group testing are minor when compared to the overall benefits. In the future, the researcher will use this method as an assessment tool with a few modifications. The groups should be chosen from a larger cohort to ensure more heterogeneous groups, in different science classes. Additionally, the researcher believes that pacing the tests throughout the year would provide more accurate test results.

About the Authors

Helen Bremert is currently undertaking a Ph.D. at Sydney University in the School of Education and Social Work, with the topic on assessment. She recently completed a Master's in Chemistry Education at Illinois State University. Prior and during her recent graduate work, she was a science teacher in New Jersey, and Sydney, Australia. Email: hbre8322@uni.sydney.edu.au

S. B. Boesdorfer, Ph.D. is an assistant professor of chemistry education in the Department of Chemistry at Illinois State University in Normal, Illinois. Email: sbboesd@ilstu.edu

References

- Baloche, L. & Brody, C. M. (2017). Cooperative learning: Exploring challenges, crafting innovations. *Journal of Education for Teaching*, 43(3), 274-283. doi:10.1080/02607476.2017.1319513
- Breedlove, W., Burkett, T., & Winfield, I. (2004). Collaborative testing and test anxiety. *The Journal of Scholarship of Teaching and Learning*, 4(2), 33–42.
- Balkam, B., Nellessen, J., & Ronney, H. (2013). Using Collaborative testing to reduce test anxiety in elementary and middle school students, Unpublished manuscript, Retrieved from <https://files.eric.ed.gov/fulltext/ED541337.pdf>
- Cantwell, R., Sousou, J., Jadotte, Y., Pierce, J., & Akioyamen, L. (2017). Collaborative testing for improving student learning outcomes and test-taking performance in her education: A systematic review. *Campbell Systematic Reviews*, 13(1), 1–18. <https://doi.org/10.1002/CL2.186>
- Efron, S. A., & Ravid, R. (2013). Action research in education: A practical guide. The Guilford Press.
- Ghaith, G. M. (2018). Teacher perceptions of the challenges of implementing concrete and conceptual cooperative learning. *Issues in Educational Research*, 28(2), 385–404.
- Gilley, B., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching*, 43(3), 83-91. https://doi.org/10.2505/4/jcst14_043_03_83
- Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high- and low-performing students. *Advances in Physiology Education*, 32(4), 274-278. <https://doi.org/10.1152/advan.00101.2007> PMID:19047503
- Haberyan, A., & Barnett, J. (2010). Collaborative testing and achievement: Are two heads really better than one? *Journal of Instructional Psychology*, 37(1), 32–41.
- Hanshaw, L. G. (2012). Qualitative aspects of group-only testing. *College Student Journal*, 46(2), 419-426.
- Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education*, 10(1), 56–70. <https://doi.org/10.1177/1469787408100195>
- Krispenz, A., & Dickhäuser, O. (2018). Effects of an inquiry-based short intervention on state test anxiety in comparison to alternative coping strategies. *Frontiers in Psychology*, 9(201), 201. <https://doi.org/10.3389/fpsyg.2018.00201> PMID:29515507
- Kurt, A. S., Balci, S., & Kose, D. (2014). Test anxiety levels and related factors: Students preparing for university exams. *JPM. The Journal of the Pakistan Medical Association*, 64(11), 1235–1239. PMID:25831637

- Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE Life Sciences Education*, 11(4), 392–401. <https://doi.org/10.1187/cbe.12-04-0048> PMID:23222835
- Meseke, C. A., Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *The Journal of Chiropractic Education*, 24(1), 19–29. <https://doi.org/10.7899/1042-5055-24.1.19> PMID:20480011
- Nariman, N., & Chrispeels, J. (2015). P.B.L. in the era of reform standards: Challenges and benefits perceived by teachers in one elementary school. *The Interdisciplinary Journal of Problem-Based Learning*, 10(1). Advance online publication. <https://doi.org/10.7771/1541-5015.1521>
- NGSS Lead States. (2013). Next Generation Science Standards: For States, By States. The National Academies Press.
- Ngotngamwong, R. (2014). Pair tests in a high school classroom. *SAGE Open*, 4(4), 1–8. <https://doi.org/10.1177/2158244014564359>
- Nist, P., & Diehl, M. (1990). PHCC test anxiety questionnaire. Retrieved from <http://phcc.edu/ods/questionnaire.html>
- Pandey, C., & Kapitanoff, S. (2011). The influence of anxiety and quality of interaction on collaborative test performance. *Active Learning in Higher Education*, 12(3), 163–174. <https://doi.org/10.1177/1469787411415077>
- Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education*, 26(1-4), 37–41. <https://doi.org/10.1152/advan.00032.2001> PMID:11850326
- Reiser, E. (2017). Blending individual and group assessment: A model for measuring student performance. *The Journal of Scholarship of Teaching and Learning*, 17(4), 83–94. <https://doi.org/10.14434/josotl.v17i4.21938>
- Ross, K. N. (1978). Sample design for educational survey research. Pergamon Press. [https://doi.org/10.1016/0145-9228\(78\)90001-8](https://doi.org/10.1016/0145-9228(78)90001-8).
- Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives*, 31, 142–147. PMID:20635617
- Seifert, K., & Sutton, R. (2009). Educational Psychology. Retrieved from <https://courses.lumenlearning.com/educationalpsychology/chapter/major-theories-and-models-of-learning/>
- Shier, R. (2004). Statistics: 1.1 Paired t-tests. Retrieved from <http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf>
- Siegel, M., Roberts, T., Freyermuth, S., Witzig, S., & Izci, K. (2015). Research and teaching: aligning assessment to instruction: Collaborative group testing in large-enrollment science

- classes. *Journal of College Science Teaching*, 44(6), 74–81. https://doi.org/10.2505/4/jcst15_044_06_74
- Suinn, R. M. (1969). The STABS, a measure of test anxiety for behavior therapy: Normative data. *Behaviour Research and Therapy*, 7, 335–339. [https://doi.org/10.1016/0005-7967\(69\)90018-7](https://doi.org/10.1016/0005-7967(69)90018-7) PMID:5344977
- Test Anxiety Self-Assessment (N.D) - nhti.edu. Retrieved from <http://www.nhti.edu/sites/default/files/content/documents/lctestanxiety.pdf>
- Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *Journal of Experimental Education*, 84(4), 787–803. <https://doi.org/10.1080/00220973.2015.1134420>
- Wanzek, J., Vaughn, S., Kent, S., Swanson, E., Roberts, G., Haynes, M., & Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness*, 7(2), 183–204. <https://doi.org/10.1080/19345747.2013.836765>
- Wenzel, T. J. (2000). Cooperative student activities as learning devices. *Analytical Chemistry*, 72, 293A–296A. <https://doi.org/10.1021/ac0027643> PMID:10763228
- Wieman, C. E., Rieger, G. W., & Heiner, C. E. (2014). Physics exams that promote collaborative learning. *The Physics Teacher*, 52(1), 51–53. <https://doi.org/10.1119/1.4849159>
- Zabucky, K. M., & Bays, R. B. (2015). Improving student's retention of classroom material through the testing effect. *College Teaching*, 63(2), 91. <https://doi.org/10.1080/87567555.2015.1005043>
- Zapatero, E. G., Maheshwari, S. K., & Chen, J. (2012). Effectiveness of active learning environment: Should testing methods be modified? *Academy of Educational Leadership Journal*, 16(4), 101–114.
- Zipp, J. F. (2007). Learning by Exams: The Impact of two-stage cooperative tests. *Teaching Sociology*, 35(1), 62–76. <https://doi.org/10.1177/0092055X0703500105>

Appendix A: Seating Groups for Group Tests

First group test

HB (5-13-17) (1-24-25) (18-21-31-39) (5-34-40) (32-38-28) (26-7-29)

CB (19-37-8) (35-4-14) (22-2-3) (33-11-27) (37-36-9) (20-12-23)

Second group Test

HB - (28-5-1) (31-24-30) (25-6-29) (15 – 34 – 39 – 13) (7 – 38 – 32) (18-26-40)(21-17-36)

CB - (22-3-12-27) (35-4-14-8) (9-19-2) (37-20 – 16) (10-11-33-23)

Third group test

HB (28-31-30) (1-40-5) (24-38-39) (25-13-32) (7-17-15) (34-26-6)

CB (3-22-23) (4-8-12) (20-37-35-11) (33-2-14-19) (36-9-27-10)

(16, 18, 29, 21) have been removed as absent for a group test)

Appendix B: Post Test Survey Questions

1. How many hours did you study for this test?
2. I felt confident I was properly prepared for the test
3. I felt positive during the test
4. The individual test section increased my anxiety – then explain choice
5. The group test section increased my anxiety – then explain choice
6. I felt more relaxed working in a group
7. I felt relief when undertaking the individual portion of the test
8. Working in a group helped my understanding of the content – then explain my choice
9. I found working in a group stressful – then explain my choice

Appendix C: Post A.P Exam Survey Questions

1. Did you feel there was a difference in retention between the material from the group or individual tests?
2. Please explain your answer
3. Did you feel that you retained the information that was tested in the group tests?
4. Please describe the reason for your answer.