Do the benefits of collaborative group exams extend beyond just improved student learning?

Jillian Arzoumanian, University of Tampa, Tampa, FL, jillian.arzoumanian@spartans.ut.edu
Suann Yang, SUNY Geneseo, Geneseo, NY, yang@geneseo.edu
Michelle Roux-Osovitz, University of Tampa, Tampa, FL, mosovitz@ut.edu
Jeffrey Grim, University of Tampa, Tampa, FL, jgrim@ut.edu

Subject/Problem

Modern pedagogical approaches are rooted in current, data-driven practices and incorporate various tools and techniques designed to promote active, student-centered learning. The incorporation of student-centered learning improves student performance and closes the education debt (Freeman et al., 2014; Theobald et al., 2020), and therefore, not surprisingly, a push has been made to integrate these approaches in classrooms. The focus of this study examines the implementation of a student-centered assessment tool classified as "two-stage testing" or collaborative group exams (CGEs). CGEs appear to have universal benefits, regardless of institution, department, and course (*e.g.*, Knierim et al., 2015). Karkhanis and Turowski (2015, as cited in Kirkland and Karkhanis, 2017) concluded that psychology students at a small community college benefited from the opportunity to participate in collaborative group exams, regardless of how they were structured or executed and that collaborative group exams are "good for performance and also for learning."

With many students prioritizing attaining a degree that would enable them to qualify for a good job, academic and social integration also play a role in students' commitment to their institutions (Davidson et al., 2015). As a result, career readiness competencies have become more meaningful to help graduates be productive contributors in the workplace (Robles, 2012), with 85.5% of employers and students/graduates placing a greater emphasis on competencies relating to communication, teamwork, and learning in the current decade (Succi & Canovi, 2020). Additionally, student retention at educational institutions, especially in STEM majors, is supported by student-centered learning approaches to promote student engagement, peer interaction, and a sense of belonging (Crosling et al., 2009; Watkins & Mazur, 2013).

Our study quantifies the effects of large-scale adoption of CGEs on student performance, learning, and group dynamics across all levels of the biology curriculum. Students at all levels realized significant benefits from CGEs, including increased performance as represented by earned course grades. Students also attributed their learning in the CGE courses to positive group dynamics and meaningful peer interactions. We present data for both quantitative and qualitative evaluation of CGEs to make recommendations for instructors interested in adopting CGEs within their classrooms. In addition, we will discuss how the positive group dynamics associated with CGEs are likely to benefit students' career preparation and promote student retention.

Design/Procedure

Participants and Course Demographic

In the 2022-2023 academic year, 13 instructors adopted collaborative group exams (CGEs) in 31 sections of 15 distinct undergraduate biology courses at a medium-sized private university. These courses included three levels; general education, introductory, and upper division, with a total enrollment of approximately 954 students. Of these, 87% (n=834) consented to participate in the study. Student parameters documented were course-category, academic year (i.e., first-year, sophomore, etc.), academic major (i.e., biology, psychology, etc.), and prior experience with CGEs. Data collection was approved by UT IRB (IRB #22-032).

Study Design

The CGE process consisted of two stages: 1) an individual exam and 2) a CGE. Each student took an exam-style assessment individually first, and then in the subsequent class period, they repeated the exact same or similar exam assessment a second time but with a collaborative peer group of two to five students. Participating instructors converted all summative exam assessments to the CGE format and weighted those exams 75% and 25% for individual and collaborative, respectively. Beyond these grading parameters, instructors had complete autonomy to tailor their implementation of CGEs in ways best aligned with their individual teaching philosophies and course designs, including how groups were formed. After each CGE, students were asked to complete an online survey to self-report their perceptions and experiences with CGEs. Survey questions included 11 Likert-type questions and six open-ended qualitative questions. Survey completion was 64% (last exam). This repeated-measures design continued for each exam within a course.

<u>Exam Performance Data</u> - Student performance on each component of the exams (individual and collaborative) was scored as a percentage out of 100 and reported by instructors. We modified the calculation for normalized learning gain (Hake, 1998) to create a normalized performance gain (NPG) metric. Briefly, based on previous literature, a minimum benchmark of 30% was set to assess the effectiveness of CGEs as an educational tool (*e.g.,* Hake, 1998).

$$NPG = \frac{(collaborative\ \%\ score\ - individual\ \%\ score)}{(100\% - individual\ \%\ score)}$$

Quantitative Data Analysis

<u>Normalized Learning Gain (NPG)</u> - To explore the factors that may contribute to NPG, we used generalized linear mixed effect models (GLMMs) on positive (beta family) and negative (gamma family) NPG values separately. Positive NPGs represent student performance gains, whereas negative NPGs show student performance losses from their individual exams to CGEs. Since NPG is bounded on the upper end by 1, and can have negative values, it does not follow a normal distribution. Therefore, positive, and negative NPGs were analyzed separately. For both analyses, academic year, academic major, course category, exam number (first or last), and prior participation in the study were fixed effects, and student identifier and course number were random effects. We only included students for whom we could calculate NPGs for first and last exams (n=762).

<u>Student Surveys of Perception and Experience</u> - Six Likert item questions focused on group dynamics and perceived learning after the last CGE were analyzed through ordinal regressions with cumulative link mixed models (CLMMs). Academic year, course category, exam number (first or last), and prior participation in the study were fixed effects, and student identifier and course number were random effects. We analyzed only the specific perceptions of students who completed both an individual exam and CGE and had corresponding NPG values (n=487-490 as not all students answered all questions).

All statistical analyses were performed using R (v4.3.0; R Core Team, 2023) and (RStudio Team, 2023). NPG GLMMs were performed using the *glmmTMB* package (Brooks et al., 2017) and ordinal regressions using the *ordinal* package (Christensen, 2022). NPG model selection (identified by lowest AIC) was conducted through the *MuMIn* package (Bartoń, 2023). Tests for the significance of predictors were obtained using the *RVAideMemoire* package (Hervé, 2023), and R^2 values for the GLMMs and CLMMs were produced using the *performance* package (Lüdecke et al., 2021). Statistical significance was identified as $p < 0.05$.

Findings/Analysis

Exam Performance

Overall, students showed improvement from individual to CGEs, with a mean NPG of 43.4% for the first and last CGEs within courses, exceeding the 30% efficacy threshold. The overwhelming majority of students (89%) earned a positive NPG, with a mean positive NPG score on the last exams being 58%. The remaining 11% of students earned a negative NPG (performance loss from individual to CGE), but the relatively large mean (-83%) is disproportionately weighted by a few extremely negative values (median=-29%). The best-fit GLMM models for both the positive (AIC=-340.8, marginal R^2<0.001 conditional R^2=1.169) and negative values (AIC=192.6, marginal R^2<0.001, conditional R^2=0.561) indicated that none of the investigated fixed effects were significant predictors of NPGs. Therefore, CGEs are not influenced by academic year, academic major, course category, exam number, or prior participation but rather by a student's individual experience progressing through a course.

Student Surveys of Perception and Experience

A significant majority of students self-reported positive perceptions and experiences with CGEs while working with their peers.

Group Dynamics - The majority of students self-reported that their group dynamics were positive during the CGEs as indicated by responses on two survey questions (Figure 1A). 93% of students also reported in response to question two (Q2), "My group worked collaboratively to arrive at a final answer when members of the group had different answers", either "most of the time" or "always". In response to question three (Q3), "All members participated and collaborated equally during collaborative group exams", a mean of 83% of students responded either "most of the time" or "always", with students in General Education courses significantly more likely to respond higher than peers in introductory courses (est marginal mean = 0.7670 +/- 0.309 SE, z = 2.484, p = 0.0347).

The results indicating positive group dynamics and collaboration were further explored by qualitative responses. Students offered a variety of perspectives on their experiences within their groups. For example, one student said, "It honestly makes you work in a team. Everyone hears other's opinions and together find a common answer. I would say you learn teamwork and get to meet people that could make a study group for the next exam." Another linked collaboration with collegiality stating, "For me, when I am working with my peers in group exams, I am able to build a bridge of understanding for anything I may not understand. On top of this, there are certain things I am able to help my peers with, creating a better collaborative effort and a welcoming and friendly environment."

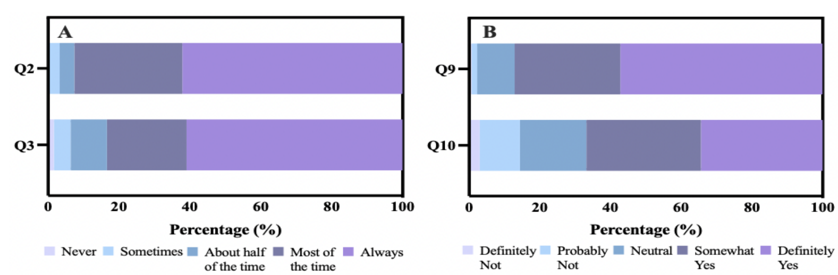

Figure 1: Total student responses to 4 Likert item questions following their participation in CGEs. Items in Panel A represent Group Dynamics (n=489), while those in Panel B indicate Perceived Learning (n=487).

Perceived Learning - Additionally, the majority of students self-reported that CGEs promoted their learning within their courses (Figure 1B). In response to question nine (Q9), "Collaborative group exams improved my overall learning in this course", 90% of students responded either "somewhat yes" or "definitely yes". Statistical models indicate that first-year and sophomore students were more likely to report that CGEs improved their overall learning (i.e., towards the somewhat yes and definitely yes end of the scale) compared to juniors (first-years vs. juniors: est marginal mean = 1.247 ±0.421 SE, z = 2.964, $p = 0.0253$; sophomores vs. juniors: est marginal mean = 0.973 ± 0.343 SE, z = 2.836, $p = 0.0369$). Fewer students attributed learning to the conversations with peers (mean=66%) by responding either "somewhat yes" or "definitely yes" in response to question ten (Q10), "I will change the way I study and learn in this class based on conversations with my peers". First-year and sophomore students are more likely to respond higher than senior peers (first-years vs. seniors: est marginal mean = 1.311 ± 0.414 SE, z = 3.169, $p = 0.0133$; sophomores vs. seniors: est marginal mean = 1.056 ± 0.337 SE, z = 3.133, $p = 0.0149$).

Students also indicated positive learning experiences in their qualitative responses. Some students noted, "I learned more about a lot of the concepts I didn't understand on the individual exam. For example, when me and my group mates were stuck on an answer, we would discuss which helped me learn new information from my peers", and "My group suggested alternatives and ideas that I had forgotten from lecture or overlooked in my own studying."

Contribution

Collaborative group testing is an effective pedagogical tool benefiting students in various disciplines (*e.g.*, Knierim et al., 2015). Given the benefits of CGEs to student learning, the natural question to ask is, "what mechanism(s) underlie these benefits?" Our results support the hypothesis that the second exposure to the exam material during the CGE increases student learning, in part, by allowing for a comparison of ideas, peer feedback, and helping students practice self-assessment of their own knowledge (metacognition) (Karkhanis and Turowski 2015, as cited in Kirkland and Karkhanis, 2017). Additionally, CGEs appear to enhance student exam performance regardless of academic year, academic major, course category, and prior experience with CGEs.

Our study identifies novel potential benefits of CGEs that extend well beyond discipline-specific academic performance gains. The self-reported group dynamics responsible for the positive performance associated with CGEs are also well-aligned with factors that enhance career readiness and student retention. For example, Succi and Canovi (2020) conveyed that respondents rank communication, teamwork, learning skills, and tolerance to stress as part of the top five most important of these career competencies. Student retention is linked to student emotional well-being (Thomas et al., 2021), which is directly supported by a sense of belonging/community underpinned by social connectedness and belonging (Travia et al., 2021). Sentiments of positive and constructive peer interactions and collaborative and productive group dynamics were overwhelming and consistent in both quantitative and qualitative student responses. Consequently, we propose that CGEs, and any other pedagogical tool that facilitates positive peer interactions, are also likely to help students develop interpersonal skills crucial for their future careers, as well as benefit student well-being and retention.

General Interest

Our research explores the adoption of CGEs at all levels of a biology curriculum, evaluating their impact on exam performance, perceived learning, and group dynamics. While our study focused on the adoption of CGEs in higher education, our findings are valuable to all educators at all levels, especially those teaching biology, who are looking for new approaches to ensure the academic achievement, overall well-being, and success of their students.

Acknowledgments

This study was supported by the University of -------- Office of Undergraduate Research and Inquiry (OURI) Grant No. 3273 and the Summer Undergraduate Research Fellowship (SURF) Grant. We would also like to acknowledge the wonderful instructors that agreed for their courses to participate in this study.

References

Bartoń K. (2023). MuMIn: Multi-Model Inference. R package version 1.47.5. https://CRAN.R-project.org/package=MuMIn

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Smaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378. https://doi.org/10.32614/RJ-2017-066

Christensen, R. H. B. (2022). ordinal - Regression Models for Ordinal Data. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal.

Crosling, G., Heagney, M., & Thomas, L. (2009). Improving student retention in higher education: Improving teaching and learning. *The Australian Universities' Review, 51*(2), 9-18

Davidson, W. B., Beck, H. P., & Grisaffe, D. B. (2015). Increasing the institutional commitment of college students: Enhanced measurement and test of a nomological model. *Journal of College Student Retention: Research, Theory & Practice, 17*(2), 162-185. https://doi.org/10.1177/1521025115578230

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences - PNAS, 111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74. https://doi.org/10.1119/1.18809

Hervé, M. (2023). RVAideMemoire: Testing and Plotting Procedures for Biostatistics. R package version 0.9-83. https://CRAN.R-project.org/package=RVAideMemoire

Kirkland, T., & Karkhanis, D. (2017). Want Your Students to Learn More? Test Them in Groups! [Blog post]. Retrieved from http://noba.to/akdch4mu

Knierim, K., Turner, H., & Davis, R. K. (2015). Two-stage exams improve student learning in an introductory geology course: Logistics, attendance, and grades. *Journal of Geoscience Education, 63*(2), 157-164. https://doi.org/10.5408/14-051.1

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Posit Team (2023). RStudio: Integrated development environment for R. Posit Software, PBC, Boston, MA.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in Today's workplace. *Business Communication Quarterly, 75*(4), 453-465. https://doi.org/10.1177/1080569912460400

Succi, C., & Canovi, M. (2020). Soft skills to enhance graduate employability: Comparing students and employers' perceptions. *Studies in Higher Education, 45*(9), 1834-1847. https://doi.org/10.1080/03075079.2019.1585420

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E.,…Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences - PNAS, 117*(12), 6476-6483

Thomas, N.S., Barr, P.B., Hottell, D.L., Adkins, A.E., & Dick, D.M. (2021). Longitudinal influence of behavioral health, emotional health, and student involvement on college student retention. Journal of College Student Development, 62(1), 2-18. https://doi.org/10.1353/csd.2021.0001

Travia, R.M., Larcus, J.G., Thibodeau, K.R., Hutchinson, C.R., Wall, A., & Brocato, N. (2021). Measuring well-being in a college campus setting. American College Health Foundation. https://www.acha.org/documents/Resources/Guidelines/Measuring_Well-Being_In_A_College_Campus_Setting_White_Paper.pdf

Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *Journal of College Science Teaching, 42*(5), 36-41. https://doi.org/10.2505/4/jcst13_042_05_36