OXFORD

BioScience, 2024, **0**, 1–11
https://doi.org/10.1093/biosci/biae001
Advance access publication date: 0 2024
Education

Group work enhances student performance in biology: A meta-analysis

Emily P. Driessen, Alan E. Wilson, Ian Hall, Peyton Brewer, Sara Odom, Sara Beth Ramsey, Sara Wood and Cissy J. Ballen

Emily P. Driessen (epd0016@auburn.edu), Ian Hall, Peyton Brewer, Sara Odom, Sara Beth Ramsey, Sara Wood, and Cissy J. Ballen are affiliated with the Department of Biological Sciences, and Alan E. Wilson is affiliated with the School of Fisheries, Aquaculture, and Aquatic Sciences at Auburn University, in Auburn, Alabama, in the United States.

Abstract

We conducted a meta-analysis to test the impacts of one active learning teaching strategy, group work, on student performance by calculating estimates across 91 studies from 53 articles. Our overall estimate indicates that the implementation of group work in biology classrooms increased student performance by 1.00 standard deviation, which we contextualized as a change greater than one letter grade. Moderator analyses revealed that this increase in performance held across all group sizes, class sizes, biology and life science majors and nonmajors, and whether the groups were assigned by the instructor. However, we did not observe increased performance in graduate level courses, in cases where group work was incorporated for only part of the course term (e.g., less than a semester or quarter) or when the group work was not graded. These results demonstrate that group work leads to impressive boosts in student performance and underscores the value of studying specific active learning strategies.

Keywords: active learning, STEM, postsecondary, cooperative learning, life sciences

Despite increasing interest in targeted teaching strategies that are effective in classroom environments (Allen and Tanner 2005, Tanner 2013, Driessen et al. 2020), robust analyses concerning the effect of single active-learning strategies on student performance across courses and institution types remain scarce. For example, current approaches to testing the effects of active learning in science, technology, engineering, and mathematics (STEM) are largely based on studies comparing courses with at least some active learning to those featuring traditional, undisrupted lecture from the instructor (Andrews et al. 2011, Freeman et al. 2011, 2014, Haak et al. 2011, Ballen et al. 2017, Barral et al. 2018, Casper et al. 2019, Theobald et al. 2020). Although such studies are valuable, for a complete understanding of the effective elements of active learning, we need to consider the individual impacts of isolated active-learning strategies on student outcomes. In the present article, we fill a gap in the literature by using meta-analytic techniques to quantitatively evaluate the effect of one common active-learning strategy: group work.

Although group work can be implemented in a variety of postsecondary subjects, we examined the effect of group work on student performance specifically in postsecondary biology classes. Group work is one of the most frequently used active-learning strategies among postsecondary biology educators (Driessen et al. 2020), and there are ample, peer-reviewed publications available on which to conduct a meta-analysis, because group work is frequently included in small-scale experimental and quasi-experimental research demonstrating positive student outcomes in postsecondary biology courses (Johnson et al. 1998, Springer et al. 1999, Knight and Wood 2005, Carmichael 2009, Chaplin 2009, Gaudet et al. 2010, Daniel 2016, Marbach-Ad et al. 2016, Yapici 2016, Donovan et al. 2018, Weir et al. 2019).

For the purposes of this meta-analysis, we broadly defined group work as students working together during class time in groups of at least two (and smaller than the total class size). This definition allowed us to capture a wide breadth of studies related to group work, encompassing specific types of group work, such as cooperative learning (Johnson et al. 1998), think–pair–share (Prahl 2017), and team-based learning (Michaelsen and Sweet 2008).

In this study, we compared the results of experiments and quasi-experiments that documented student performance in classes with at least some group work component with those documenting classes without a group work component by analyzing data from the published and unpublished literature. Gathering available data from nearly a century of research, we investigated the following question: What is the effect of group work on academic performance?

Experimental design

We followed the updated PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for best practices in quantitative reviews (Page et al. 2021). To identify studies comparing student performance in undergraduate biology classes that incorporated at least some group work component with those classes without a group work component, we searched seven online databases (Web of Science, APA PsychInfo, ERIC, PubMed, Academic Search Premier, Education Research Complete, Dissertations and Theses; supplemental table S1) using predetermined search terms (box 1) and reviewed the papers included in three previous meta-analyses on active learning and group work (Springer et al. 1999, Johnson et al. 2000, Freeman et al. 2014). As part of the systematic literature review, we

Received: February 28, 2022. **Revised:** December 26, 2023. **Accepted:** January 10, 2024

Box 1. Search terms used to identify papers.

((“Collaborative Learning” OR “Group Work” OR “Peer Instruction” or “Peer Learning” OR “PLTL” OR “Team-Based Learning” OR “Team Based Learning” OR “Team-Based-Learning” OR “Small-group” OR “Large-group” OR “Medium-group” OR “Cooperative Learning” OR “Think-pair-share” OR “Deliberative Democracy”) AND (“College Student*” or “Univ*” or “Undergraduate*” OR “Graduate*” OR “Post-secondary”) AND (“Biol*” OR “Natural Science*” OR “Health Science*) AND (“Achievement” OR “Test” OR “Performance” OR “Outcome*” OR “Learning Gain*” OR “Grade” OR “Score” OR “Summative” OR “Assessment” OR “Exam” OR “Final”))

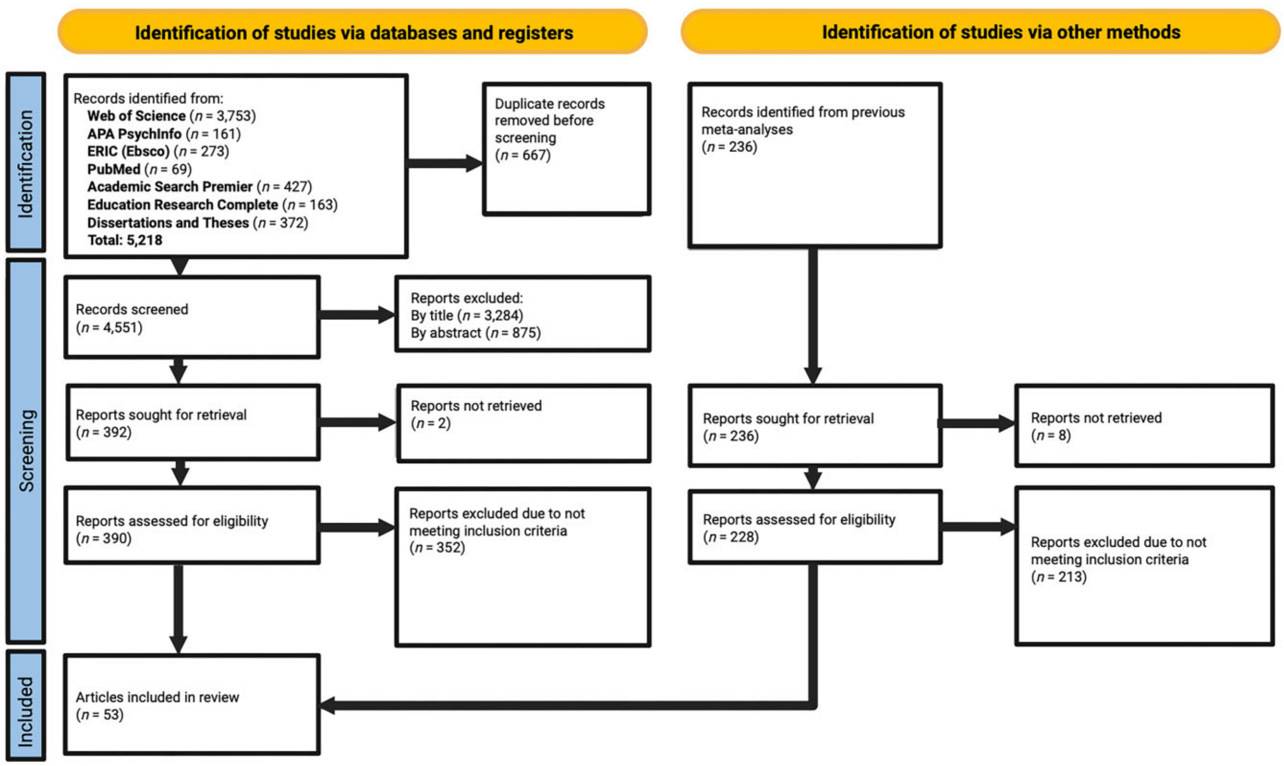

Figure 1. PRISMA flow diagram for systematic searches of databases. Source: Adapted from Page and colleagues (2021).

searched for articles published from 1 January 1924 to 16 November 2020. We selected our earliest date because it was cited in a previous article (Johnson et al. 1998) that informed a previous meta-analysis on group work (Johnson et al. 2000).

Criteria for admission

We used several criteria for admission for our study, including only studies that examined postsecondary students in a biological science class; incorporated any type of group work (defined by two or more students working or discussing topics together); and reported, at a minimum, a mean measure of academic performance before and after group work (i.e., pre- and post-treatment test) or a mean measure for the classes experiencing group work and for the classes not experiencing group work (i.e., treatment and comparison). We considered academic performance a broad measure of student learning in postsecondary biology courses, and we used student outcomes from assessments, including quiz scores, midterm exam scores, final exam scores, final grade, and concept inventories. The included studies also investigated the effect of group work on student performance in lecture courses, rather than lab courses (i.e., we did not include articles that examined the effect of group work on student performance in labs because of the inherent differences between lab and lecture courses). For this criterion, we refer to lecture as a type of class (e.g., lecture as opposed to lab class), although lecture can also be used to describe a teaching practice (i.e., passively lecturing) in other contexts. The included studies were conducted in a classroom during official class meeting times. This does not include lab recitation, quiz sections, or supplementary instruction session times. The included studies also necessarily used the same concept inventory or pre-treatment and post-treatment tests, the same exams, or final class grade to evaluate the two classes. Finally, the included studies were published or reported from 1 January 1924 to 16 November 2020.

We also used several criteria to exclude studies. Specifically, we did not include studies that compared the effect of group work with the effect of one-on-one instruction, because we did not find that representative of many postsecondary teaching environments; that used a crossover study design (i.e., studies where the same students were exposed to both individual and group work and then compared with themselves), because of the inherent confounding effect of each treatment on student performance due to the design type; that compared the effect of group work with the effect of another strategy that included group work (e.g., flipped classroom), because we were interested only in comparing student performance in courses that used group work with student performance in courses that did not use group work; or that used group work only for group exams, because we were interested in the effect of group work during class instruction rather than during summative assessments.

Review process

Three authors (ED, IH, and PB) read the titles and abstracts of the 5218 papers returned by the database searches (figure 1), as well as from the 236 citations in three previous relevant meta-analyses (Springer et al. 1999, Johnson et al. 2000, Freeman et al. 2014) to determine eligibility. We excluded articles from the study if they did not meet one or more of the criteria for admission or if they did meet one or more of the exclusion criteria (figure 1). In cases where it was unclear whether a study fit the scope of our project, we read the entirety of the paper to determine its suitability and excluded articles that did not meet our inclusion criteria or did meet our exclusion criteria. After these steps, we were left with 53 papers from which to collect data. See the supplemental materials for the full reference list of papers included in the meta-analysis. To conduct moderator analyses, we recorded the available data on group size, class size, class level, whether the class was attended by mostly majors or nonmajors, the duration of the group work (e.g., semester-long, weeklong), whether the group work was graded or not, and whether the group was assigned by the instructor or not (table 1). We also collected data on several variables for the purposes of sensitivity analyses because they are known to have an experimental impact in education research. These included experimental design type (i.e., independent or paired), whether the instructor was the same for the treatment and comparison group, and whether—in the case of independent experimental designs—the student treatment and comparison sections were academically comparable before the experiment.

Effect size

To estimate differences in student performance (e.g., average class grade, final exam score) across comparable courses that experienced group work from those that did not experience group work, we calculated a standardized mean difference for each course in the form of Hedges' g (Hedges 1981). This statistic is commonly used in meta-analyses, it estimates the effect size for the difference between means scaled to the pooled standard deviation, and it also includes a correction for small sample sizes (Hedges and Olkin 1985). A positive value indicates increased student performance in the intervention group (i.e., group work) relative to the comparison group (i.e., no group work), whereas a negative value indicates increased student performance in the comparison group (i.e., no group work) relative to the intervention group (i.e., group work; Durlak 2009).

Data collection

We categorized study designs as either paired (i.e., pre- versus post-group-work data collection) or independent (i.e., post-treatment with group work versus post-comparison without group work). The difference in study design required different approaches to data collection. Specifically, for a paired design, we collected mean student performance scores (e.g., concept inventory scores, exam scores) before and after working in groups, standard deviations, the number of students in each class both before and after experiencing group work, and a pretreatment–posttreatment correlation value. However, correlation values were not present in any of the studies, so we imputed correlation values of .9 (see the supplemental material). For an independent design, we collected mean performance scores (e.g., concept inventory scores, exam scores, final grades), standard deviations, and the number of student participants in each class (e.g., classes that experienced group work versus classes that did not experience group work) from each article. Importantly, we did not use course grade as a measure of student performance for independent research designs where the grading schemes for both the treatment and comparison sections were not the same. Some articles contributed multiple data points if multiple estimates were available.

Despite the differences between the data collected for each experimental design type, we extracted data from both types of studies directly from tables or text in the publication or from figures using WebPlotDigitizer version 4.2 (https://automeris.io/WebPlotDigitizer). In cases where a study did not provide standard deviations—or standard error values that we could convert to standard deviations—we contacted the corresponding author via email. However, if we did not get a response from the author within 3 months, we imputed values; this was the case for 15 of our 91 estimates (16.48%). According to Kambach and colleagues (2020, p. 11706), when calculating Hedges' g as the estimate, if the meta-analysis data set is missing fewer than 25% of the standard deviation values, then it is best practice to use mean-value imputation for the standard deviation values rather than exclude the data from analysis. We estimated these values using the average standard deviation from the other studies (mean-value imputation; Kambach et al. 2020; also see the supplemental material). In addition, we developed a protocol to handle cases where studies provided either multiple treatments, multiple comparisons, or multiple outcomes (see the supplemental material).

Coding data

After data collection, we categorized each study according to the following moderators: group size, class size, class level, whether the class was mostly majors or nonmajors, the duration of the group work (e.g., semester-long, weeklong), whether the group work was graded, and whether the groups were assigned by the instructor (see the supplemental material for further details on these codes). For sensitivity analyses, we also collected information on whether the instructor was the same, whether—in the case of independent experimental designs—the student treatment and comparison sections were academically comparable before the experiment, and the experimental design type used in the study. In our original approach, we planned to code for the type of group work, the amount of class time—in minutes—spent in groups, the percentage of a student's grade that was affected by graded group work, whether the group work involved high or low stakes, and what the students were specifically doing in their groups. However, this information was not commonly available in the articles or comparable among articles, if it was provided. For example, many articles simply wrote that students worked in groups, without addressing the exact amount of time spent during each class on group work, if there was a grade assigned to this work or the work the students completed within their groups. When articles did provide this information, it was difficult to compare. For example, when the authors did report the amount of time spent in groups, they reported it as a percentage, minutes of class time, or days of the week, without the context to convert this information to comparable values across articles.

Table 1. Estimates included in the meta-analyses with moderator information.

Citation	Group size	Class size	Level	Majors	Duration	Graded	Assigned
Altiparmak et al. (2009)	M	S	NA	No	<S	NA	NA
Armbruster et al. (2009)	M	M	L	Yes	S	No	Yes
Armstrong et al. (2007)	L	L	L	No	S	Yes	Yes
Armstrong et al. (2007)	L	L	L	No	S	Yes	Yes
Beers (2005)	NA	NA	U	Yes	<S	NA	NA
Brown (2016)	M	S	U	NA	S	NA	Yes
Burrowes (2003)	M	M	L	NA	S	Yes	No
Carmichael (2009)	M	L	L	NA	S	No	No
Chaplin (2009)	M	M	L	Yes	S	NA	Yes
Chuck (2011)	M	S	U	Yes	S	Yes	No
Collier (2017)	L	L	L	Yes	S	Yes	Yes
Collier (2017)	L	L	L	Yes	S	NA	Yes
Connell et al. (2016)	M	M	L	No	S	NA	Yes
Connell et al. (2016)	M	M	L	No	S	NA	Yes
Daniel (2016)	M	M	L	NA	S	NA	NA
Das et al. (2019)	L	M	G	Yes	<S	Yes	NA
Donovan et al. (2018)	M	L	L	No	S	NA	Yes
Donovan et al. (2018)	M	L	L	No	S	NA	Yes
Donovan et al. (2018)	M	L	L	No	S	NA	No
Fernández-Santander (2008)	S	S	L	NA	S	NA	NA
Fetalvero (2017)	NA	M	L	No	S	NA	Yes
Fuller et al. (2016)	L	NA	U	Yes	<S	NA	NA
Gaudet et al. (2010)	L	NA	U	NA	S	Yes	Yes
Giojalas et al. (2021)	M	M	L	NA	S	Yes	No
Hacisalihoglu et al. (2018)	S	M	L	Yes	S	NA	Yes
Heather (2008)	L	S	U	NA	S	No	NA
Huysken et al. (2019)	NA	S	L	Yes	S	NA	NA
Huysken et al. (2019)	NA	S	L	Yes	S	NA	NA
Kitchen et al. (2003)	S	M	U	Yes	S	No	NA
Klegeris et al. (2013)	NA	M	NA	NA	S	NA	NA
Knight and Wood (2005)	S	M	U	Yes	S	Yes	Yes
Koufogiannakis et al. (2005)	L	NA	G	Yes	S	Yes	Yes
Koufogiannakis et al. (2005)	L	NA	G	Yes	S	Yes	Yes
Lewis and Tamblyn (1987)	NA	S	U	Yes	S	NA	NA
Lian and He (2013)	L	NA	L	Yes	NA	No	Yes
Lyons (2008)	L	S	L	Yes	S	NA	NA
Marbach-Ad et al. (2016)	S	M	L	Yes	S	No	No
Moreno-Lopez et al. (2009)	M	S	U	NA	S	Yes	Yes
Morse and Jutras (2008)	NA	L	L	Yes	S	Yes	No
Mutlu (2018)	NA	NA	L	NA	NA	No	Yes
Mutlu (2018)	NA	NA	L	NA	NA	No	Yes
Rajappa et al. (2016)	NA	NA	G	Yes	<S	No	Yes
Randolph (1992)	NA	NA	L	NA	NA	NA	NA
Reich, and Wang (2019)	S	M	L	Yes	S	NA	NA
Reich and Wang (2019)	S	S	L	Yes	S	NA	NA
Rissanen (2018)	M	L	L	No	S	S	NA
Sangestani and Khatiban (2013)	M	S	L	Yes	S	NA	NA
Santisteban (2017)	NA	M	L	Yes	S	S	Yes
Sevening and Baron (2003)	S	S	U	Yes	S	NA	Yes
Sevening and Baron (2003)	L	S	U	Yes	S	NA	Yes
Stiles and Katene (2013)	S	NA	U	No	<S	NA	NA
Styers, Van Zandt and Hayden (2018)	NA	M	L	Yes	NA	NA	NA
Styers et al. (2018)	NA	S	L	Yes	NA	NA	NA
Styers et al. (2018)	NA	S	U	Yes	NA	NA	NA
Suhendar (2017)	M	S	NA	NA	NA	NA	NA
Szogedi et al. (2009)	S	NA	G	Yes	NA	NA	NA
Theobald et al. (2017)	NA	L	L	NA	<S	NA	Yes
Theobald et al. (2017)	NA	L	L	NA	<S	NA	Yes
Tsaushu et al. (2012)	L	M	NA	NA	NA	NA	NA
Walsh et al. (1999)	L	NA	G	Yes	<S	NA	NA
Walters (2014)	M	M	L	Yes	S	NA	Yes
Watt-Watson et al. (2004)	L	L	NA	Yes	<S	NA	Yes

Table 1. Continued.

Citation	Group size	Class size	Level	Majors	Duration	Graded	Assigned
Weasel and Finkel (2016)	L	L	L	No	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	M	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	M	NA	NA	S	NA	NA
Weir et al. 2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	M	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	M	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	L	NA	NA	S	NA	NA
Weir et al. (2019)	NA	M	NA	NA	S	NA	NA
Weir et al. (2019)	NA	S	NA	NA	S	NA	NA
Yang et al. (2019)	L	S	L	NA	S	Yes	Yes
Yapici (2016)	M	S	NA	NA	S	Yes	NA

Note: Each row represents one estimate (i.e., the difference between two means) of the effect of group work on student performance. The citation column represents the article the estimate was taken from. The full citations are provided in the supplemental materials. The group size column details whether the group was small (S, $n < 4$ students), medium (M, $4 \leq n > 6$ students), or large (L, $n \geq 6$ students). The class size column details whether the class was small (S, $n < 50$ students), medium (M, $50 \leq n < 200$ students), or large (L, $n \geq 200$ students). The level column indicates the class level the data was taken from: lower (L) is first- and second-year students, upper (U) is third- and fourth-year students, and graduate (G) is graduate students. The majors column indicates whether the students in the study were mostly biology or life sciences majors, with "yes" indicating the students were mostly majors and "no" indicating the students were mostly not majors. The duration column indicates the duration of the group work treatment: semester or quarter long treatment (S) or less than the quarter or semester length (<S). The graded column indicates whether group work was graded, with "yes" indicating group work was graded and "no" indicating group work was not graded. The assigned column represents whether groups were assigned by the instructor, with "yes" indicating the groups were assigned by the instructor and "no" indicating the groups were not assigned by the instructor. NA in any column represents the fact that this data was unavailable for the specific study. Several citations are listed in this table more than once, indicating the number of estimates they contributed to our meta-analysis. If an article contributed more than one estimate (i.e., data point), it was because the study contained multiple experiments or data from several classrooms that could be used to calculate more than one estimate. For example, Weir and colleagues (2019) is a meta-analysis of the effect of group work on student performance in which 26 different classrooms at one university were investigated, so we have 26 estimates from that article.

Calculating the effect of group work on student performance

We collected and analyzed data (i.e., student performance data, including final grades, exam scores, and concept inventory scores) from two types of studies: paired design and independent design. Specifically, we used Comprehensive Meta-Analysis software (www.meta-analysis.com) to calculate Hedges' g and the associated variance for each study. We analyzed the Hedges' g values and associated variance values in RStudio (R Core Team 2016), using the *metafor* package (Viechtbauer 2010). We used this package to conduct a multivariate linear mixed-effects model, where the fixed effect was Hedge's g and the random effect was the author of the study. We fixed our intercept for this model at zero, to test whether or not the effect of group work on student performance was significantly different from the null (i.e., group work has no effect on student performance). In addition, we fitted the model with the restricted maximum-likelihood method to find the overall effect of group work on student performance (Viechtbauer 2010).

To contextualize our estimate, we collected final grades from several large, introductory biology courses at a single institution. We took the average standard deviation for each class section and multiplied this by our overall estimate. This allowed us to calculate how group work would potentially affect a student's letter grade, at least in the case where exam scores accounted for the majority of the students' grades.

Next, we conducted several sensitivity analyses using *metafor* to determine the robustness of our results by considering the extent to which they are affected by changes in our models or assumptions (Viechtbauer 2010). Specifically, to test the sensitivity of our results to removal of particular studies, changes in correlation values, and changes in imputed standard deviation values, we conducted meta-analyses with the following changes from the original model: We removed the random effect of author from the

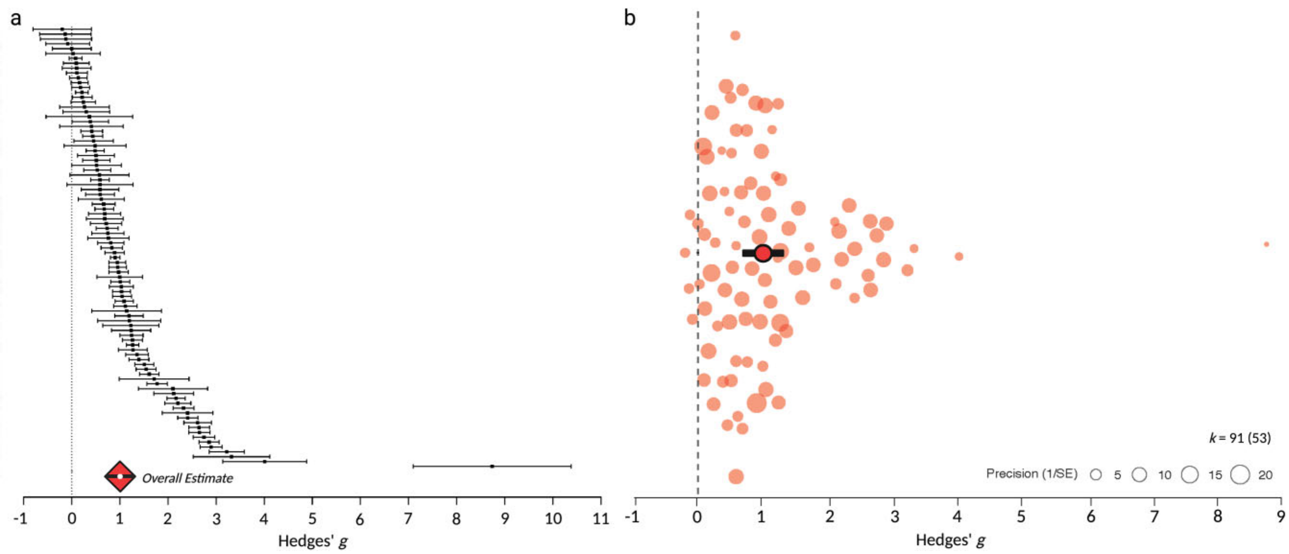

Figure 2. Caterpillar and orchard plots depicting the effect of group work on student performance. The overall effect of group work on student performance, using Hedges' ***g*** as the effect size. A positive value indicates increased student performance in the intervention group (i.e., group work) relative to the comparison group (i.e., no group work), whereas a negative value indicates increased student performance in the comparison group (i.e., no group work) relative to the intervention group (i.e., group work). (a) A caterpillar plot is a forest plot organized by estimate size. Each dot in the caterpillar plot represents one estimate in our meta-analysis, and the lines extending from each estimate represent the 95% confidence intervals. The diamond at the bottom of the plot shows the overall estimate (the white circle) and its 95% confidence intervals (the black bars). If the overall estimate plus the 95% confidence intervals lies to the right of the null (i.e., is a positive value) and does not cross the null, then the effect of group work on student performance is positive and statistically significant. (b) The orchard plot represents each estimate (one comparison of the effect of group work on student performance versus the effect of no group work on student performance) as an individual circle, and the size of each of these comparisons is inversely proportional to its standard error. The overall estimate is represented by the small circle outlined in black, and the black horizontal lines extending from either side of the circle represent the 95% confidence intervals. If the estimate, including the 95% confidence intervals, lies to the right of the null (i.e., is a positive value) and does not cross the null (i.e., zero), then the effect of group work is positive and statistically significantly different from the effect of no group work on student performance.

original analysis; we excluded one study that contributed 26 of the 91 estimates (table 1; Weir et al. 2019); we removed extreme outliers (i.e., those with estimates two or more standard deviations greater than the overall estimate); we doubled the imputed standard deviation value for those studies missing standard deviations; we removed the estimates that required imputed standard deviations; we tested whether our data set was robust to changes in the correlation value in paired research design comparisons by changing it to 0.9, 0.5, and then 0.1; we excluded data from paired design studies; in the case of studies with an independent research design, we removed all estimates obtained from studies that did not have the same instructor for the treatment and comparison groups; and in the case of studies with an independent research design, we removed all estimates obtained from studies that did not demonstrate that the treatment and comparison groups were academically comparable prior to the study. To assess publication bias, we constructed and visually inspected funnel plots (supplemental figures S3a and S4a; Sterne et al. 2005), used Duval and Tweedie's trim and fill method (supplemental figures S3b and S4b; Duval and Tweedie 2000a, 2000b), and calculated fail-safe *N*s (Rosenthal 1979).

Next, to assess whether the effect of group work on student performance held across each of the levels (e.g., lower, upper, and graduate) within each of the moderators (e.g., class level), we conducted multivariate linear mixed-effects model analyses with the restricted maximum-likelihood method for class size, class level, group size, majors or nonmajors classes, the duration of the group work (e.g., semester-long, weeklong), whether the group was graded or not, and whether the groups were assigned or not, as individual fixed effects, using the *metafor* package in RStudio (Viechtbauer 2010, R Core Team 2016). We fixed our intercept for these models at zero, to test whether the effect of group work on student performance was significantly different from the null (i.e., group work has no effect on student performance). To visualize the output as a caterpillar plot (i.e., a forest plot organized by estimate size; figure 2a) and orchard plot (figure 2b), we used the *metafor* and *orchaRd* packages in R, respectively (Viechtbauer 2010, Nakagawa et al. 2020). We then exported these plots to edit in Biorender, using the web application to combine separate R plots into one singular figure and to add aesthetic aspects, such as axis titles (BioRender.com).

Finally, we conducted pairwise analyses with the restricted maximum-likelihood method to test for differences between the levels within each of the moderators (i.e., class size, class level, group size, majors or nonmajors classes, the duration of the group work, whether the group was graded or not, and whether the groups were assigned or not) using the *metafor* package in RStudio (Viechtbauer 2010, R CoreTeam 2016).

Results reveal the staggering and positive impact of group work on student performance

From the 5454 articles and manuscripts we reviewed (i.e., 5218 from database searches + 236 used in previous meta-analyses), we compiled data for our meta-analysis from 53 articles and 91 estimates (figure 1). We analyzed these estimates in several different ways: in total, to yield an overall estimate; using several sensitivity analyses to establish reliability and validity; and using seven moderators to explore the impact of different class settings. We detail each of these analyses in the following sections.

Our results showed group work had a large positive effect on student academic performance Hedges' $g = 1.00$ (± 0.30 SD for ± 95% confidence intervals (CI), $z = 6.35$, $P < 0.0001$; figure 2, supplemental table S2a). This means, on average, students who worked in groups performed 1.00 standard deviation higher than students who did not work in groups. To put the effect size in perspective, Hedges and Hedberg (2007) asserted that Hedges' g values of 0.20 or higher on performance measures in education research should be of interest to policymakers. Kraft (2020) pointed out that effects considered small by these standards can be quite large relative to the impacts of most field-based interventions in education, particularly in larger studies with broad achievement measures. Therefore, our results (an estimate of 1.00 standard deviation) should be considered staggering and a move for policy change.

To contextualize our overall estimate of 1.00, we collected final grades from 2951 students enrolled in fourteen large, introductory biology courses at a single institution, averaging 211 students in each class. In these courses, exam scores generally accounted for the majority of students' grades. We multiplied the average standard deviation for each class section (i.e., 13.49) by the estimate (i.e., 1.00) to calculate how group work would affect a student's grade. This yielded an increase of 13.49 percentage points, a change equal to more than one letter grade.

The positive impact of group work is robust to interrogation

We then conducted several sensitivity analyses to determine the robustness of our results by considering the extent to which they are affected by changes in our models or assumptions (see the supplemental material). These sensitivity analyses created a range of estimates from 0.83 (± 0.45 SD for ± 95% confidence intervals (CI), $P < 0.001$; supplemental table S2k) to 1.15 (± 0.39 SD for ± 95% confidence intervals (CI), $P < 0.0001$; supplemental table S2g). These sensitivity analyses demonstrated our results are robust to correlation value changes and the removal of particularly large estimates, articles that contribute a large amount of estimates, imputed standard deviation changes, estimates with imputed standard deviations, estimates taken from paired design studies, estimates taken from experiments where the instructor was different for the treatment and comparison group, and estimates taken from experiments that did not demonstrate that prior preparation was similar across the treatment and comparison groups at the outset of the study (supplemental table S2b–S2h). Unpublished studies with low effect sizes are unlikely to have created sampling bias, as was demonstrated by both Duval and Tweedie's trim and fill method (Duval and Tweedie 2000a, Duval and Tweedie 2000b) and Rosenthal's fail-safe N method ($n = 179{,}395$; Rosenthal 1979; see also supplemental figures S3 and S4).

The positive impact of group work applies across a variety of moderating contexts

Next, we analyzed our data using the following moderators: group size, class size, class level, whether the classes were for majors or nonmajors, the duration of the group work, whether the group was graded or not, and whether the groups were assigned by the instructor. These analyses tested whether the positive and statistically significant effect of group work on student performance held across each of the levels within each of the moderators. Our moderator analysis of group size demonstrated group work had a significant positive effect on student performance in all group sizes, (i.e., small, medium, and large; figure 3, supplemental table S5; omnibus test of moderators, $QM(3) = 32.69$, $p < .0001$). Our moderator analysis of class size demonstrated that group work improved student performance across all class sizes (i.e., small, medium, and large; figure 3, supplemental table S6; $QM(3) = 86.65$, $p < .0001$), although the largest positive effect of group work was in large classes (i.e., 200 or more students). The moderator analysis of class level demonstrated that group work improved student performance across both lower- and upper-level classes; however, it did not have a significant effect on student performance in graduate classes (figure 3, supplemental table S7; $QM(3) = 34.67$, $p < .0001$). The moderator analysis of majors or nonmajors showed that for both majors and nonmajors classes, group work had a significant effect on student performance (figure 3, supplemental table S8; $QM(2) = 35.79$, $p < .0001$). In our moderator analysis of duration of group work, we found that group work had a significant effect on student performance in classes that used group work throughout the duration of the semester or quarter; however, group work did not have a significant effect on student performance in classes that used group work for less than the duration of the semester or quarter (figure 3, supplemental table S9; $QM(2) = 34.25$, $p < .0001$). In our moderator analysis of whether group work was graded or not, we found that group work had a significant effect on student performance in classes where the group work was graded; however, group work did not have a significant effect on student performance in classes that did not grade the group work (figure 3, supplemental table S10; $QM(2) = 12.47$, $p = .002$). Finally, in our moderator analysis of group assignment, we found for both the groups that were assigned by the instructor and those that were not assigned, group work had a significant effect on student performance (figure 3, supplemental table S11; $QM(2) = 35.46$, $p < .0001$).

These analyses reveal that the significant and positive effect of group work on student performance holds across all group sizes, class sizes, classes levels, major status, and groups assignment by the instructor status. However, we did not observe increased performance in graduate level courses, in cases where the group work was incorporated for only part of the course term (e.g., less than a semester or quarter) or when the group work was not graded.

Next, we conducted pairwise analyses (i.e., linear models where one of the levels of the moderator is the reference and each of the other levels within the moderator are compared with it) to test for differences between the levels within each moderator. On the basis of these analyses, there were no statistical differences on the impact of group work on performance outcomes between the group size ($QM(2) = 0.93$, $p = .63$), class levels ($QM(2) = 2.17$, $p = .34$), majors or nonmajors classes ($QM(1) = 0.050$, $p = .822$), whether group work was graded ($QM(1) = 0.59$, $p = .44$), or the duration of group work ($QM(1) = 1.12$, $p = .29$). However, the effect of group work on student performance did vary on the basis of class size ($QM(2) = 55.15$, $p < .0001$), with the students from the large classes performing 0.43 standard deviations better than the students from the small classes (95% CI = 0.31, $p < .01$) and 0.44 standard deviations better than the students in the medium class sizes (95% CI = 0.12, $p < .0001$). The students in the medium class sizes performed 0.02 standard deviations (95% CI = 0.30) lower than the students in the small class sizes, although this was not statistically significant ($p = .28$). In addition, the effect of group work on student performance varied on the basis of whether the group was assigned by the instructor or not ($QM(2) = 12.98$, $p = .0003$), with the students in assigned groups performing 0.55 standard deviations (95% CI = 0.30) higher than the students who were not in assigned groups ($p < .0001$).

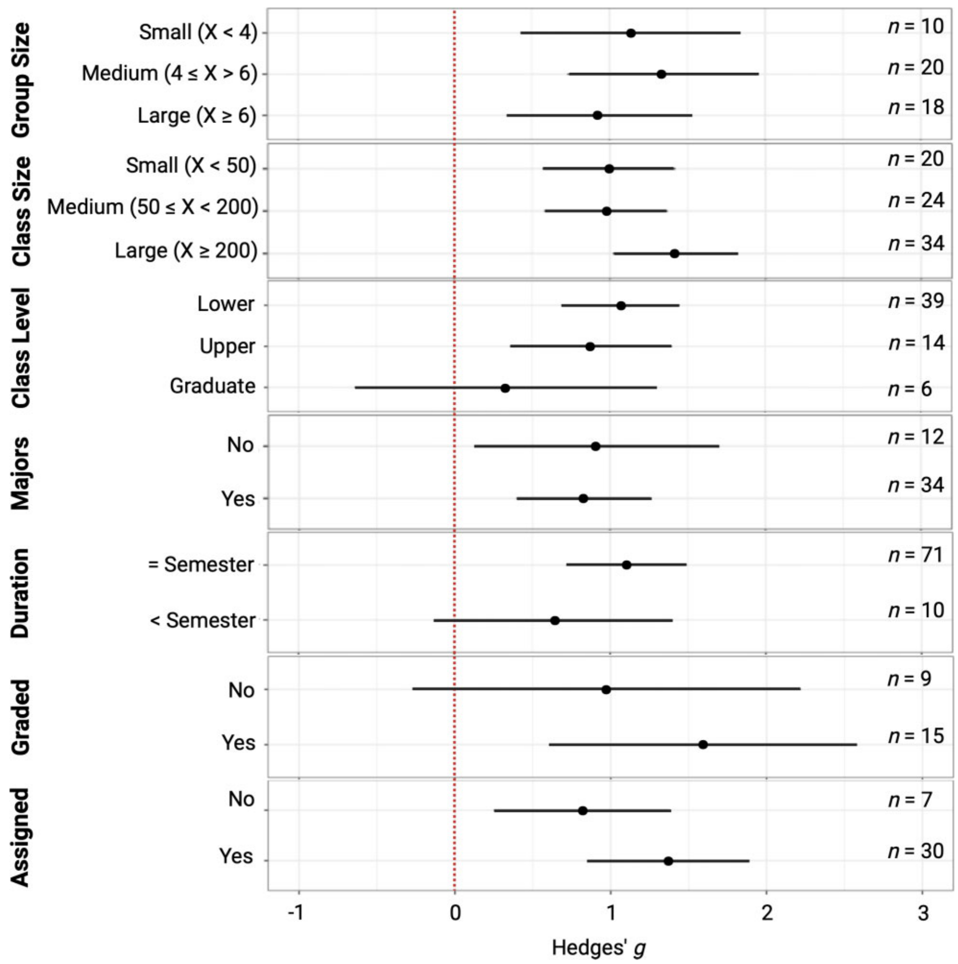

Figure 3. Mean estimates with 95% confidence intervals for the levels within each moderator: group size, class size, class level, majors versus nonmajors, duration of group work, whether the group work was graded or not, and whether the groups were assigned or not. If the estimate and the confidence intervals lie to the right of the dotted line (i.e., Hedges' $g = 0$), then the effect of group work for that level within that moderator has a positive and statistically significant effect on student performance in postsecondary biology courses. If the confidence interval crosses the dotted line, then the effect of group work on student performance for that level within that moderator is not statistically significant.

When interpreting the results of moderator analyses and pairwise analyses, we remind the reader that these are separate analyses, and sometimes one analysis will show significant effects while the other will not. This is because these two analyses compare each level of a moderator with different intercepts. Specifically, the moderator analysis compares each level of the moderator with the null (i.e., zero in this case; group work does not increase student performance), whereas the pairwise analysis compares each level of the moderator with the others (e.g., there is no statistical difference between the effect of group work on student performance when group work was graded versus ungraded).

We leverage our staggering results to encourage instructors to use group work in their courses

Using meta-analysis, we investigated the effect of group work on student academic performance in postsecondary biology courses, using broad measures of student learning, including quiz scores, exam scores, final grades, and concept inventory scores. We show that the implementation of group work increased overall student performance by 1.00 standard deviation, which we translated to approximately a 13% boost in students' grades. Given that group work is often an integral part of active learning (Driessen et al. 2020), our results support the findings of Freeman and colleagues (2014), signaling that active-learning strategies can improve student performance in the sciences. Furthermore, we found the effect of group work persisted across group sizes, class sizes, course levels, group sizes, major status, and group assignment. We also found positive effects of group work on student performance when group work was used throughout the duration of the semester or quarter and when the group work was graded. To evaluate the generality of the impacts of active-learning strategies on student outcomes, future research will benefit from both the study of a wide range of individual strategies (e.g., clicker questions, presentations) and the investigation of how different uses of the same strategy affect student outcomes (Century et al. 2010).

Our work demonstrates the significant and large effect of group work on student performance, which holds across a variety of moderating variables, and we encourage instructors to implement group work in their biology classrooms. However, on the basis of our findings that the effect of group work on student performance was significant only when group work was graded, as opposed to not graded, and when it was used throughout the quarter or semester, rather than being used for only a portion of the course, we recommend that instructors grade the group work in their class and use group work for the duration of their class.

There are plenty of resources available for instructors who are interested in maximizing the positive effects of group work, in its various forms, in their class. For example, Tanner and colleagues (2003) detailed several essential elements of effective cooperative learning (a specific type of group work), such as individual and group accountability and promoting interpersonal skills. Prahl (2017) advised instructors who wish to implement think–pair–share exercises in their classroom, offering guidance on writing questions and problems, interacting with the students, and assessing the exercises. Michaelsen and Sweet (2008) explained team-based learning and defined the four essential elements as properly formed and managed groups, student accountability for the quality of their individual and group work, frequent and timely feedback on student work, and group assignments that promote learning and team development. Finally, Smith and Knight (2020) detailed how clickers (technology used to relay student responses to the instructor) can be integrated into the biology classroom to help encourage students to apply their knowledge and analyze new scenarios. In fact, Smith and colleagues (2009) demonstrated that peer discussion during clickers improves student understanding of material through group interactions and debate, even when no one in the group knows the answer at the beginning of the exercise. They suggest a constructivist explanation for how students are arriving at their answer and advancing their understanding, underscoring the importance of coupling group work with student response systems such as clickers. Although there are many other resources available for instructors interested in implementing group work in their class, we offer these as foundational starting points.

Considerations and future directions for group work research

The present study has several limitations that inspire future directions. First, the studies we analyzed usually used a variety of group-work-oriented strategies in the treatment courses (e.g., group brainstorming, group discussion of homework questions, and group clicker questions) and a variety of teaching strategies that did not use group work in the comparison courses (e.g., lecture, lecture with clicker questions, and lecture with daily quizzes at the start of class and in-class individual assignments). Because we cannot disentangle each specific group-work strategy or factor in the type of comparison teaching strategy used (because few studies used the exact same comparison teaching strategies), we cannot identify exactly which group-work-oriented strategy produced the largest effect on student performance. Identifying the best group work strategy for increasing student performance is important in making instructor recommendations, and this requires further study.

Second, our contextualization of the overall effect of group work on student performance was based on final grades that largely consisted of points contributed from high-stakes exams. Although this is reflective of the grading structure in many undergraduate biology courses in the United States, it is not representative of the grading structure in all postsecondary biology courses. Given this, it is important to note that grading structures that incorporate features aside from high-stakes exams, such as attendance, participation, and homework, may see varying benefits from the contextualized improvement of more than one letter grade.

Third, some of the studies included in this meta-analysis lacked information for the moderators we analyzed (i.e., group size, class size, class level, majors, duration of group work usage, graded, and assigned). Missing information means we relied on subsets of data, sometimes representing data from as few as six articles as was the case for the graduate class level, to evaluate the effects of moderators on group work. Even within the moderators that we examined, sometimes the level of detail available in the papers limited our understanding. For example, we categorized biology major classes as those in which the students in the class were mostly biology or life science majors. If the only information we were provided was that 60% of the students in a class were not biology majors, we would categorize this class as for nonmajors, even if the class was required to be a biology major. Conducting follow-up studies of our moderators could certainly provide additional insights.

Fourth, few studies in our meta-analysis demonstrated a null or negative effect of group work on student performance. It may be that null or negative effects of group work are not being published because of the file drawer effect (i.e., null or negative findings are less likely to be published than is significant work; Rosenthal 1979). However, this is an issue that all meta-analyses are susceptible to, and our bias analyses demonstrate there was little concern for the file drawer effect (Rosenthal 1979).

Fifth, we did not use a forward or backward citation accumulation method. These citation accumulation methods are helpful as checks that the literature was exhaustively searched. Given that we did not use forward or backward citation accumulation methods, we may have missed several studies that fit the scope of our meta-analysis; however, we feel our search is still robust given that we searched seven large databases with many relevant search terms, checked the citations of the articles found by previous meta-analyses on active learning and group work, and found a fail-safe number of articles that suggests a minimal risk of publication bias in our meta-analysis (i.e., our fail-safe number was large, indicating it would take 179,395 articles to nullify our overall estimate result).

Sixth, future research will benefit from analyzing other moderating variables that may affect group work, such as time spent on group work, the degree of structure put in place for the groups and the group activities, whether or not students have prior experience with group work, the percentage of a student's grade that is affected by graded group work, heterogeneous versus homogeneous groups, among other variables. We were unable to analyze these variables as moderators because the literature did not consistently or commonly list these features in their experiments and quasiexperiments.

Seventh, our search terms did not include *biochemistry* or *biotechnology*, so our meta-analysis is focused on studies that examined classes labeled *biology*, *biological sciences*, *natural science* or *natural sciences*, or *health science* or *health sciences*. Although the impact of group work on biology-adjacent disciplines (such as biochemistry and biotechnology)—as well as other STEM disciplines—was out of the scope of the current study, we encourage future research to focus on this topic. The future of discipline-based education research relies on advancing our knowledge of specific strategies to make clear recommendations to instructors and curriculum developers.

Conclusions

In the present article, we demonstrated that group work is a teaching strategy that has the potential to boost student performance in postsecondary biology courses by 1.00 standard deviation, contextualized in the present article as an improvement of more than one letter grade. Given that students generally earn lower grades in STEM courses than in non-STEM courses (Koester et al. 2016) and that these grades are consequential (those who leave STEM cite the performance challenge associated with STEM courses as the primary motivation for their decision; Seymour and Hewitt 1997, Seymour and Hunter 2019), group work has the potential to benefit individuals and ultimately strengthen the STEM workforce.

Supplemental material

Supplemental data are available at *BIOSCI* online. All data and code used for this meta-analysis are available at https://github.com/EmilyDriessen/Group-work-and-student-performance-in-biology-A-meta-analysis..git.

Acknowledgments

We thank Abby Beatty, Todd Steury, and Elli Theobald for their invaluable advice. In addition, we thank Jordan Harshman, Karen McNeal, Mary Mendonca, the Ballen lab, and the Auburn University Biological Sciences Department for their thoughtful feedback. We would also like to thank the students from the Auburn University Meta-analysis course for providing useful suggestions on earlier drafts. Finally, we thank the many authors of research articles on the effect of group work who provided missing data on request. This work was supported by National Science Foundation grants no. DUE-2120934, no. DUE-2011995, and no. DBI-1919462 awarded to CJB.

Author contributions

Emily P. Driessen (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing), Alan E. Wilson (Resources, Software, Writing – review & editing), Ian Hall (Data curation), Peyton Brewer (Data curation), Sara Odom (Formal analysis), Sara Beth Ramsey (Data curation), Sara Wood (Data curation), and Cissy J. Ballen (Conceptualization, Funding acquisition, Project administration, Resources, Software, Supervision, Writing – review & editing)

References cited

Allen D, Tanner K. 2005. Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. *Cell Biology Education* 4: 262–268.

Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST. 2011. Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education* 10: 394–405.

Ballen CJ, Wieman C, Salehi S, Searle JB, Zamudio KR. 2017. Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education* 16: 56.

Barral AM, Ardi-Pastores VC, Simmons RE. 2018. Student learning in an accelerated introductory biology course is significantly enhanced by a flipped learning environment. *CBE—Life Sciences Education* 17: 129. https://doi.org/10.1187/cbe.17-07-0129.

Carmichael J. 2009. Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching* 38: 54–61.

Casper AM, Eddy SL, Freeman S. 2019. True grit: Passion and persistence make an innovative course design work. *PLOS Biology* 17: 3000359. https://doi.org/10.1371/journal.pbio.3000359.

Century J, Rudnick M, Freeman C. 2010. A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation* 31: 199–218.

Chaplin S. 2009. Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching* 39: 72.

Daniel KL. 2016. Impacts of active learning on student outcomes in large-lecture biology courses. *American Biology Teacher* 78: 651–655. https://doi.org/10.1525/abt.2016.78.8.651.

Donovan DA, Connell GL, Grunspan DZ. 2018. Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE—Life Sciences Education* 17: 283. https://doi.org/10.1187/cbe.17-12-0283.

Dougherty KJ, Jones SM, Lahr H, Natow RS, Pheatt L, Reddy V. 2014. Performance funding for higher education: Forms, origins, impacts, and futures. *ANNALS of the American Academy of Political and Social Science* 655: 163–184.

Driessen EP, Knight JK, Smith MK, Ballen CJ. 2020. Demystifying the meaning of active learning in postsecondary biology education. *CBE—Life Sciences Education* 19: 52.

Durlak JA. 2009. How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology* 34: 917–928.

Duval S, Tweedie R. 2000a. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95: 89–98.

Duval S, Tweedie R. 2000b. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56: 455–463.

Freeman S, Haak D, Wenderoth MP. 2011. Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education* 10: 175–186. https://doi.org/10.1187/cbe.10-08-0105.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111: 8410–8415. https://doi.org/10.1073/pnas.1319030111.

Gaudet AD, Ramer LM, Nakonechny J, Cragg JJ, Ramer MS. 2010. Small-group learning in an upper-level university biology class enhances academic performance and student attitudes toward group work. *PLOS ONE* 5: e15821. https://doi.org/10.1371/journal.pone.0015821.

Haak D, HilleRisLambers J, Pitre E, Freeman S. 2011. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science* 332: 1213–1216.

Hedges LV, Hedberg EC. 2007. Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis* 29: 60–87.

Hedges LV. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6: 107–128.

Hedges LV, Olkin I. 1985. *Statistical Methods for Meta-Analysis*. Academic Press.

Johnson DW, Johnson RT, Smith KA. 1998. Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning* 30: 26–35. https://doi.org/10.1080/00091389809602629.

Johnson DW, Johnson RT, Stanne MB. 2000. *Cooperative Learning Methods: A Meta-analysis*. PUBLISHER.

Kambach S, Bruelheide H, Gerstner K, Gurevitch J, Beckmann M, Seppelt R. 2020. Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecology and Evolution* 10: 11699–11712.

Knight JK, Wood WB. 2005. Teaching more by lecturing less. *Cell Biology Education* 4: 298–310.

Koester BP, Grom G, McKay TA. 2016. Patterns of gendered performance difference in introductory STEM courses. *AERA Open* 3: 2332858417743754. https://doi.org/10.1177/2332858417743754.

Kraft MA. 2020. Interpreting effect sizes of education interventions. *Educational Researcher* 49: 241–253.

Marbach-Ad G, Rietschel C, Thompson KV. 2016. Validation and application of the survey of teaching beliefs and practices for undergraduates (STEP-U): Identifying factors associated with valuing important workplace skills among biology students. *CBE—Life Sciences Education* 15: 59. https://doi.org/10.1187/cbe.16-05-0164.

Michaelsen LK, Sweet M. 2008. The essential elements of team-based learning. *New Directions for Teaching and Learning* 2008: 7–27.

Nakagawa S, Lagisz M, O'Dea RE, Rutkowska J, Yang Y, Noble DWA, Senior AM. 2020. *orchaRd: An R Package for Drawing 'Orchard' plots (and 'caterpillars' plots) from Meta-analyses and Meta-regressions with Categorical Moderators*. Ideel.org. http://www.i-deel.org/uploads/5/2/4/1/52416001/orchard_vignette.pdf.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372: 71.

Prahl K. 2017. Best practices for the think–pair–share active-learning technique. *American Biology Teacher* 79: 3–8.

R Core Team. 2016. *RStudio: Integrated Development for R*. RStudio.

Rosenthal R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86: 638.

Seymour E, Hewitt NM. 1997. *Talking about Leaving*. Westview Press.

Seymour E, Hunter A-B. 2019. *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*. Springer.

Smith MK, Knight JK. 2020. Clickers in the biology classroom: Strategies for writing and effectively implementing clicker questions that maximize student learning. Pages 141–158 in Mintzes JJ Walter EM, eds. *Active Learning in College Science*. Springer.

Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT. 2009. Why peer discussion improves student performance on in-class concept questions. *Science* 323: 122–124.

Springer L, Stanne ME, Donovan SS. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research* 69: 21–51. https://doi.org/10.3102/00346543069001021.

Sterne JAC, Becker BJ, Egger M. 2005. The funnel plot. Pages 75–98 in Rothstein HR, Sutton AJ Borenstein M, eds. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Wiley.

Tanner KD. 2013. Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE—Life Sciences Education* 12: 322–331.

Tanner K, Chatman LS, Allen D. 2003. Approaches to cell biology teaching: Cooperative learning in the science classroom: Beyond students working in groups. *Cell Biology Education* 2: 1–5.

Theobald EJ, et al. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences* 117: 6476–6483. https://doi.org/10.1073/pnas.1916903117.

Viechtbauer W. 2010. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36: 1–48.

Weir LK, Barker MK, McDonnell LM, Schimpf NG, Rodela TM, Schulte PM. 2019. Small changes, big gains: A curriculum-wide study of teaching practices and student learning in undergraduate biology. *PLOS ONE* 14: e0220900. https://doi.org/10.1371/journal.pone.0220900.

Yapici İÜ. 2016. Effectiveness of blended cooperative learning environment in biology teaching: Classroom community sense, academic achievement and satisfaction. *Journal of Education and Training Studies* 4: 269–280. https://doi.org/10.11114/jets.v4i4.1372.

Received: February 28, 2022. **Revised:** December 26, 2023. **Accepted:** January 10, 2024