Analyzing the success of a student support program beyond the target class

Lauren Perry, California State University, Sacramento, Department of Mathematics and Statistics, Sacramento, CA, USA

Corey Shanbrom*, California State University, Sacramento, Department of Mathematics and Statistics, Sacramento, CA, USA

Matthew Krauel, California State University, Sacramento, Department of Mathematics and Statistics, Sacramento, CA, USA

*Corresponding author: corey.shanbrom@csus.edu

ABSTRACT

A wealth of research supports the efficacy of peer support in bolstering student success in undergraduate courses, but little is known about the impact on outcomes in future courses. Here we investigate the existence of a dosing effect on students participating in a peer-led academic support program at a large public university. Our analysis includes 26,731 unique students and 62,222 observations in 20 STEM courses. Due to the voluntary nature of the treatment, we use propensity score weighting to control for observed confounding. We find a significant dosing effect, where increased exposure to the program treatment is associated to higher grades across all of the courses studied, not only those targeted by the intervention. Compared with students with no exposure, students with one, two, and three or more cumulative support courses had adjusted estimated grade differences of approximately 0.13 grade points, 0.23 grade points, and 0.32 grade points, respectively. We also find that this dosing effect holds across various academic and demographic subpopulations. Our results add to the growing literature supporting peer instruction as a powerful tool for improving student outcomes and provide new evidence for positive secondary effects in concurrent and future courses.

Keywords: Dosing effect, peer support, STEM, propensity score analysis

INTRODUCTION

While student success in a single course is often measured to evaluate the effectiveness of an academic support program, this practice can obscure the effects of repeated program exposure on student success. When a student receives one *dose* of a program, that is, one intervention targeting a focal course, the link between a program's influence and student success is relatively direct. However, when students experience a higher *dosage*, representing cumulative cross-course exposure to the program, assessing a program's effectiveness on student outcomes is more complex. Effective programs may promote *transfer*, whereby something learned in one context helps in another (Perkins & Salomon, 1992). In this case, additional program doses targeting different courses may influence student success metrics in other courses. Research on program

dosage is limited, especially in the higher education setting, as is research on a program's effects on student success beyond focal courses. In this study, we find that increased dosages of a peer learning program at a large regional university is associated with higher cumulative grade point average (GPA) across a set of historically low-pass-rate science, technology, engineering, and mathematics (STEM) courses, extending beyond focal course performance.

The central challenge in this analysis is estimating cumulative and transferable exposure to a peer learning program using administrative course data. Student-course observations provide rich information on students' academic histories, allowing program exposure in one course to be related to outcomes in another. These data also allow us to treat program dosage as a cumulative, multi-level exposure rather than as a binary, course-specific treatment, and to distinguish direct exposure in the focal course from non-direct exposure in other courses. While this analysis uses techniques commonly employed in observational causal inference (see Gonzalez Canche (2019) for an example with a binary outcome), the research question and data require an extension of typical course-level evaluation designs. Propensity score weighting is used to address self-selection into different program dosage levels, and overlap weights are applied to avoid relying on poor comparisons between students with differing probabilities of receiving each dosage level (Li et al., 2018; Rosenbaum & Rubin, 1983). This framework provides a reusable approach for evaluating opt-in academic support programs where exposure is repeated, cumulative, and potentially transferable across courses.

In addition to the main analysis, we construct exploratory decomposition models examining degrees of dose intensity (direct versus indirect), the effects of dosage on student subgroups (underrepresented minority (URM) status, gender, and transfer status), and changes in GPA relative to students' pre-dose performance. Although exploratory, these analyses provide additional evidence that greater exposure to the peer learning program is associated with improved academic outcomes. They also suggest that the program may attenuate the association between baseline GPA and course grades, consistent with reduced disparities in performance across student groups.

BACKGROUND

The Peer Assisted Learning Program

This study examines the Peer Assisted Learning (PAL) Program in the College of Natural Sciences and Mathematics at California State University, Sacramento, also known as Sacramento State. Sacramento State enrolls over 31,000 undergraduate students annually and is one of 22 campuses in the California State University system. It is a nationally recognized Asian American and Native American Pacific Islander-Serving and a Hispanic Serving Institution. It is also a Carnegie R2 classified institution, offering many graduate degrees, including some doctorates. Most of Sacramento State's students are from the diverse Sacramento Metropolitan area, and large portions are categorized as transfer students, first-generation, or Pell grant eligible.

PAL originated in 2011 and has grown to now serve, depending on the year, 17 low-pass STEM major focal courses in biology, chemistry, mathematics, physics, and statistics. As of the start of

this analysis, PAL has served 20 distinct focal courses since its inception. The program's primary method of serving these courses is through its offering of 1-unit opt-in supplementary PAL workshops specific to one of the focal courses. These PAL classes are graded Credit/No Credit based entirely on attendance and participation, meet twice a week for 50 minutes, and are run by an undergraduate student employed by the program, called a Facilitator. Facilitators are interviewed in-person for the position, with empathy being a desired skill, and must have been successful in the focal course, earning a B or higher. Once employed, they also strengthen or refresh their skills in the focal course by attending its lectures for the duration of the semester they serve as a Facilitator.

PAL classes typically have 10 to 15 students enrolled and attendance is mandatory (with up to three unexcused absences to receive credit). Content is not taught in PAL classes. Instead, enrolled students are placed into groups of three to four to work on worksheets that have been created by Sacramento State faculty who have taught the focal course. The job of the Facilitator is to guide this group work by following the PAL Model, a framework originally based on the Peer-Led Team Learning (PLTL) model of peer learning, but which has been significantly altered.

Hallmarks of the PAL Model include Facilitators never confirming or denying student answers, never explaining material, and facilitating group work on whiteboards. PAL is not tutoring and it is not remedial. Students enrolled collaboratively solve problems with their peers and engage in productive struggle in order to build confidence in their own ideas and work. Indeed, the PAL Model is designed to promote these skills with the intention of facilitating the transfer of learning. By withholding explanations and confirmation of answers, Facilitators encourage students to articulate their reasoning, evaluate peers' work, establish ownership of their problem-solving strategies, and think critically about their process and answers. In addition, Facilitators are near-peers, who also discuss ways to navigate college as a student, campus resources, and best practices for studying. Further details can be found in Shanbrom and Perry (2025) or Shanbrom et al. (2023).

To manage their classroom learning environments, Facilitators are extensively trained. They undergo eight hours of paid pre-semester training. They also participate in a two-unit, graded, upper-division course that meets once a week for two hours and which immerses them in a variety of additional and continuous professional development. This includes expertise in using the PAL Model, as well as leadership, communication, cultural competency, and other socio-emotional skills. The course also requires Facilitators to engage in year-long research projects surrounding pedagogical aspects of their own PAL classes. Other works outline the program's unique structure and facilitator experience (Akhavan et al., 2024; Lundmark et al., 2017) and demonstrate that Facilitators develop important career readiness skills, including the eight National Association of Colleges and Employers competencies (Krauel & Shanbrom, 2026).

Concerning PAL's effectiveness in influencing student success, students who earn credit in PAL demonstrate higher grades in the corresponding focal course. Indeed, Shanbrom et al. (2023) used propensity score matching involving 11 PAL classes to find participation in PAL, on average, corresponds to a 23% increase in a student's focal course GPA. More broadly, a large

body of research exists supporting the effectiveness of PLTL and other peer learning programs. See the bibliography of Arendale (2022) for a comprehensive, though incomplete, list of works.

Dosing and the Transfer of Learning

The medical and pharmaceutical literature carefully distinguishes between various subtle variables associated to dosing (Conn & Chan, 2015; Conn & Groves, 2011). In our context, however, we consider passing a PAL course to be a single dose of the treatment that is the same for all participants, rendering the notions of intensity, frequency, and duration irrelevant. The primary discrete variable of interest is PAL dosage. We are primarily interested in the existence of a *dosing effect*, by which we mean the phenomenon where multiple doses are more effective than a single dose.

The effects of dosage on student success have been examined across a two-semester chemistry course sequence for a PLTL program by Mitchell et al. (2012) and for the Sacramento State PAL program by Shanbrom and Perry (2025). The outcomes measured in these studies were students' pass rate and GPA, respectively, and a comparison of the results can be found in the discussion of Shanbrom and Perry. In PLTL programs where attendance is not required, participating students may receive vastly different treatments according to the time they invest. For such programs it would make sense to consider dosage as a measure of the total number of hours participating, and indeed this has been studied, but only in settings quite different than PAL, such as elementary schools, and the results have been mixed (Cohen et al., 1982; Ginsburg-Block et al., 2006; Johnson et al., 1981; Rohrbeck et al, 2003;). In other educational programs, dosage has been considered in high-dosage tutoring (Kortecamp & Peters, 2024; Roberts, 2021) and high-school dual enrollment programs (Karp et al., 2007; Lee & Villarreal, 2022), where Liu et al. (2025) also note gaps in the literature concerning dosage.

Meanwhile, research on transfer is vast and comprehensive (Barnett & Ceci, 2002; Haskell, 2001). In PLTL, transfer research is often qualitative, attempting to isolate the skills that a PLTL program develops. For example, skill development in groupwork, complex idea comfort, creativity, and more were studied using surveys for a PLTL program by Utschig and Sweat (2008). Quantitative studies measuring student outcomes in courses are more limited, but the work of Mitchell et al. (2012) and Shanbrom and Perry (2025) do support evidence of transfer. Mitchell et al. (2012), for example, explicitly describe the desire to probe PLTL's role in promoting transfer as motivation for their work. To our knowledge, no research exists that relates dosage to transfer.

METHODS

Data

We analyze student-level administrative records from a four-year regional public university to evaluate whether PAL has benefits that transfer beyond the course in which it was originally taken. The raw dataset contained 74,544 student-course observations on 27,791 unique students in 20 lower- and upper-division undergraduate STEM courses across 13 academic years. The

courses are listed in Table A1 of the Appendix. The outcome of interest is course grade on a four-point scale. Courses taken for Credit/No Credit, as well as courses marked "incomplete," are not assigned a numeric grade and so were excluded. Student-course observations that resulted in withdrawals were not included in the original dataset. We further restricted the data to the first attempt for each student-course combination. Repeated attempts were excluded because repeat enrollment reflects a different academic context: students have prior exposure to the course content, prior grade experience, and potentially prior PAL exposure attached to the same course. Including repeats would make it difficult to distinguish transferable PAL exposure from the consequences of retaking a course. The first-attempt restriction accounted for most of the reduction from the raw dataset to the final analytic sample.

Students in categories with very small cell sizes were excluded when there was no substantively appropriate comparison group into which they could be combined: nonbinary gender categories were not collapsed into binary gender categories, and post-baccalaureate students were not combined with undergraduates. These exclusions were made to avoid unstable propensity score estimates and inappropriate category aggregation. The dataset was also restricted to the covariates required for propensity score weighting and outcome modeling. Finally, observations with missing values on variables required for the analytic models were excluded from the primary analysis, except for covariates with nontrivial missingness, which were retained using the imputation-and-indicator approach described below. The final analytic sample included 62,222 observations on 26,731 unique students. See Table 1.

Sample Construction Step	Observations Remaining	Students Remaining	Notes
Raw administrative records	74,544	27,791	Initial student-course records
Remove course repeats	64,181	27,403	
Remove nonbinary, unknown gender, and post-baccalaureate students	63,639	26,967	Excluded because cell sizes were too small for stable subgroup adjustment
Remove invalid course grades	62,331	26,744	Outcome available on four-point scale
Subset to complete cases	62,222	26,731	Final analytic sample

Table 1. Sample construction process showing number of observations and number of students remaining at each step.

Measures and Exposure Construction

Our primary measure of interest is cumulative PAL dosage, which includes all PAL courses taken up to the point the grade was assigned for the focal course. This includes all prior terms, as well as any PAL classes taken in the current term. This definition reflects the hypothesis that PAL may support both later coursework and other concurrent STEM courses through transferable learning strategies. This variable was limited to zero, one, two, or three-or-more cumulative PAL courses, as cell sizes for four or more PAL courses taken were too small for meaningful analysis. A PAL dose was counted only when the student both enrolled in and attended the PAL course.

We also constructed decomposed exposure variables to examine the separate effects of PAL for the focal course and transfer exposure from other PAL courses. For each student-course observation, the focal course is the course whose grade is being analyzed. Direct PAL exposure indicates the student took PAL attached to that focal course. Same-term transfer exposure indicates the student took PAL attached to a different course in the same semester. These are not mutually exclusive: a student could have direct PAL exposure for the focal course and have transfer PAL exposure from another course taken in the same term. Prior-term cumulative exposure describes the total number of PAL courses taken by the student in previous semesters, coded as 0, 1, 2, or 3+. Finally, non-direct PAL dosage combines same-term transfer exposure and prior-term cumulative exposure (and does not include any PAL for the focal course), and is coded as 0, 1, 2, or 3+.

The propensity score model included 21 covariates related to academic performance, academic load and enrollment status, prior credits/remedial coursework, demographic characteristics, socioeconomic indicators, time and cohort variables, and missingness indicators. See Table A2 in the Appendix for details. For covariates with appreciable missingness, we retained observations by using mean-imputed values together with missingness indicators in the propensity score model (Allison, 2009; Groenwold et al., 2012). This allowed missingness patterns to contribute to the estimated weights without treating imputed values as observed data. For covariates with less than 1% missingness, observations with missing values were excluded because the loss of observations was minimal, and separate missingness indicators would have been too sparse to contribute meaningfully to the weighting model.

For the outcome models, we controlled for additional covariates which are relevant to a student's grade, but which do not represent baseline measures which would be appropriate in the propensity score models for the number of previous PAL courses a student has taken. Included as covariates in the model were course, instructor, and major; each student was treated as a cluster to account for repeated observations.

For the post hoc models, we also created a variable for URM status. This variable was defined from the Integrated Postsecondary Education Data System ethnicity variable: Black, Hispanic, Native American, and Pacific Islander students were classified as URM; White and Asian students were classified as non-URM; and students identified as Two or More Races or Unknown were classified as Other/Unknown.

Analysis

Our analysis uses an observational causal-inference framework and our estimates can be interpreted as observational causal estimates under the assumptions of adequate adjustment for observed confounding, sufficient overlap, and no unmeasured confounding (Austin & Stuart, 2015; Rosenbaum & Rubin, 1983; Rubin, 2008). We address each of these assumptions in our analysis: observed confounding is addressed through propensity score weighting and covariate balance; overlap is addressed by generating overlap weights and examining balance diagnostics and effective sample sizes; and unmeasured confounding is addressed through a sensitivity analysis.

Because students self-select into PAL, we used propensity score weighting to adjust for observed differences across PAL dosage groups. As PAL dosage was categorical with four levels, we estimated generalized propensity scores for the multi-level treatment (Imai & Van Dyk, 2004; Imbens, 2000). Propensity scores were estimated using a gradient boosting machine (GBM) with 21 covariates capturing prior academic performance, missingness in academic records, academic load, student status, demographic and socioeconomic characteristics, and time in the institution (Greifer (*WeightIt*), 2026; McCaffrey et al., 2004). This flexible approach was selected because the covariate set included several high-cardinality variables and because simpler modeling approaches did not achieve adequate balance. The GBM was used as a design-stage tool, and its performance was evaluated by post-weighting covariate balance rather than predictive accuracy (Ho et al., 2007). We used overlap weights to target the population of students with comparable observed characteristics across PAL dosage conditions; the primary estimand was the average treatment effect in the overlap population (Li et al., 2018). That is, we target the population of students who might plausibly fall into any of the PAL dosage groups. Balance was assessed using standardized mean differences across all treatment pairs, with values below 0.10 treated as acceptable (Austin & Stuart, 2015; Griefer (*cobalt*), 2026; Stuart, 2010). We also examined effective sample sizes after weighting to assess the amount of information retained in each dosage group.

We estimated PAL dosage effects using weighted linear models to predict course grade on a four-point scale (Austin & Stuart, 2015; Ho et al., 2007). We used linear models because the primary quantities of interest were adjusted mean differences in course grade points, which are directly interpretable on the institutional grading scale. The propensity score model adjusted for pre-treatment student characteristics, while the outcome model additionally adjusted for course, instructor, and academic plan to account for grading context, course material, instructional approach, and student program at the time of the focal course. These variables were included in the outcome model to improve precision and account for grading context but were not used to define the propensity weights because they may be downstream of course-taking patterns and PAL availability. Standard errors were clustered by student to account for repeated observations from the same student, and robust standard errors were used to reduce dependence on homoskedasticity assumptions (Blair et al., 2025; Cameron & Miller, 2015; Liang & Zeger, 1986; White, 1980). We then used estimated marginal means to compare adjusted course grades across PAL dosage groups, using a Holm p-value adjustment for multiple comparisons (Holm, 1979; Lenth & Piaskowski, 2026). Course, instructor, and academic plan were treated as nuisance factors when estimating marginal means.

Our models proceed in three parts. First, a primary cumulative-dosage model examining total PAL effect. Second, we construct exploratory decomposition models: (A) direct versus non-direct PAL dosage and (B) direct, same-term transfer, prior-term dosage. Finally, we construct exploratory heterogeneity models: GPA, URM status, transfer status, and gender.

For the primary cumulative dosage model, the PAL dosage covariate was total PAL exposure. This was used to examine the overall PAL exposure effect across multiple dosage levels. This model is our main weighted causal analysis. We then generated two exploratory decomposition models to separate course-specific effects from transfer exposure from other PAL courses. The

first model separated direct PAL exposure from non-direct PAL exposure, including both in the model as separate covariates. Direct PAL was included in the outcome model because it represents another concurrent exposure component rather than a baseline confounder. This model was designed to separate course-specific support from transfer exposure and is our primary decomposition model because it provides the clearest test of whether the cumulative PAL association reflects primarily focal course support or support from PAL exposure outside the focal course. The second decomposition model separated out direct PAL exposure, same-term transfer PAL exposure, and prior-term PAL dosage, including all three as separate covariates in the model. This second model was designed to distinguish concurrent transfer from prior-term transfer exposure. Because the decomposition variables created smaller and more complex exposure cells, separate propensity score models for those component exposures produced less stable balance diagnostics. We therefore retained the primary cumulative-dosage weights for the decomposition analyses and interpret these models as exploratory analyses of the components of the primary PAL dosage association. These models are intended to clarify whether the overall cumulative association appears to be driven more by direct focal-course PAL, same-term transfer exposure, or prior-term PAL exposure, rather than to provide separately weighted causal estimates for each component.

After examining the primary PAL dosage models, we estimated exploratory post hoc heterogeneity models using the primary cumulative-dosage weights. These models examined whether the association between total PAL exposure and course grade differed by baseline GPA, gender, transfer status, and URM status. The GPA heterogeneity model included an interaction between PAL dosage and baseline local GPA. The gender, transfer, and URM models included interactions between PAL dosage and the subgroup indicator, with baseline local GPA retained as an additional covariate to adjust for prior academic performance. Students classified as Other/Unknown for URM status were excluded from the binary URM subgroup analysis.

Finally, because propensity score methods can only adjust for observed confounding, we conducted a sensitivity analysis for potential unmeasured confounding. We conducted this sensitivity analysis for the primary cumulative-dosage contrasts, converting estimated grade differences to standardized mean differences and calculating E-values to assess how strong an unmeasured confounder would need to be to explain away the estimated PAL effects (Linden et al., 2020; VanderWeele & Ding, 2017).

Analyses were conducted in *R* (R Core Team, 2026) using packages *WeightIt* for propensity score weighting (Greifer, 2026), *cobalt* for balance assessment (Greifer, 2026), *estimatr* for weighted linear models with robust clustered standard errors (Blair et al., 2025), *emmeans* for adjusted marginal means and contrasts (Lenth & Piaskowski, 2026), and *EValue* for sensitivity analyses (VanderWeele & Ding, 2017). *R* code for this analysis is available as a supplementary file.

RESULTS

Our results are organized in several parts. First, we describe the unweighted distribution of grades and PAL dosage in the analytic sample. Second, we report covariate balance and effective

sample sizes after weighting. Third, we present the primary cumulative-dosage model and exploratory decomposition models. Finally, we report exploratory heterogeneity analyses and sensitivity analyses for unmeasured confounding.

Unweighted Sample Statistics

Initial, unweighted descriptive statistics show the mean course grade increase from 2.35 for students with a 0 PAL dosage to 2.72 for students with a 3+ PAL dosage. The sample sizes for the PAL dosage groups ranged from 43,865 for 0 PAL dosage to 2,593 for 3+ PAL dosage. These descriptive differences are not adjusted for differences in student characteristics across dosage groups and are presented only to characterize the analytic sample. See Table 2.

PAL Dosage	Mean Grade Point	Standard Deviation	n
0	2.35	1.34	43,865
1	2.42	1.22	11,782
2	2.50	1.16	3,982
3+	2.72	1.10	2,593

Table 2. Unadjusted summary statistics (mean, standard deviation, and sample size) for each PAL dosage level on the final analytic sample.

Covariate Balance and Effective Sample Sizes

After applying the propensity score analysis for the four PAL treatments, all covariates were balanced at the 0.1 threshold; the largest adjusted absolute standardized mean difference was 0.0945. A Love plot of these results may be found in Figure A1 of the Appendix. Effective sample sizes showed good information retention, with an adjusted sample size of 1,988 in the 3+ PAL dosage group. See Table 3.

	PAL Dosage			
	0	1	2	3+
Unadjusted	43,865	11,782	3,982	2,593
Adjusted	14,768.79	7,368.94	3,533.75	1,988.03

Table 3. Effective sample sizes for each dosage level after weighting.

Primary Cumulative-Dosage Model

In the weighted outcome model, PAL dosage was positively associated with subsequent STEM course grades. Our main result shows that, compared with students with no PAL exposure, students with one, two, and three or more cumulative PAL doses had adjusted estimated grade differences of approximately 0.13 grade points, 95% CI (0.08, 0.17); 0.23 grade points, 95% CI (0.18, 0.29); and 0.32 grade points, 95% CI (0.24, 0.39), respectively. The consecutive contrasts were positive but smaller at higher dosage levels, with estimated increments of 0.105 grade

points from 1 to 2 PAL exposures, 95% CI (0.05, 0.16); and 0.087 grade points from 2 to 3+ exposures, 95% CI (0.01, 0.16). See Table 4.

Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
0 vs 1	0.127	0.0219	<0.0001
0 vs 2	0.232	0.0284	<0.0001
0 vs 3+	0.319	0.0385	<0.0001
1 vs 2	0.105	0.0282	0.0004
2 vs 3+	0.087	0.0395	0.0275

Table 4. Primary cumulative dosage results; contrasts were performed on each dosage level relative to the baseline (no PAL) and consecutively.

Exploratory Decomposition Models

In exploratory decomposition models using the primary cumulative-dosage weights, the direct PAL estimate was larger than the same-term transfer and non-direct dosage estimates, suggesting that the cumulative dosage pattern was driven substantially by PAL attached to the focal course. However, non-direct exposure estimates were also significantly positive, indicating that PAL exposure outside the focal course may contribute to the overall cumulative association. Consecutive contrasts suggest the clearest difference is between no non-direct exposure and any non-direct exposure. See Table 5.

Exposure	Estimate	Standard Error	Holm-Adjusted P-Value
Direct PAL	0.2890	0.0199	<0.0001
Non-Direct PALs: 1	0.0494	0.0203	0.0147
Non-Direct PALs: 2	0.0983	0.0313	0.0050
Non-Direct PALs: 3+	0.1315	0.0442	0.0058

Table 5. Estimated grade point difference for each exposure setting (compared to no PAL exposure).

In the secondary decomposition model, we separated non-direct exposure into same-term transfer exposure and prior-term PAL dosage. Same-term transfer PAL was associated with a 0.101 increase in course grade (SE = 0.0343, p = 0.0032). The pattern of results for prior-term PAL dosage was substantively similar to those for non-direct PAL dosage. Complete secondary decomposition model results, along with consecutive contrasts for the first model, are reported in Tables A3 and A4 of the Appendix.

In general, as total PAL exposure increases, the proportion of students in that category receiving direct PAL exposure also increases. This is expected: as students accumulate more PAL exposure, they are increasingly likely to be observed in a course for which they have direct PAL exposure. This is reflected in the estimates when PAL dosage is broken down into direct and non-direct components. See Table 6.

PAL Dosage	n	n Without Direct PAL	n With Direct PAL	% Without Direct PAL	% With Direct PAL
0	43,865	43,865	0	100	0
1	11,782	5,460	6,322	46.3	53.7
2	3,982	1,609	2,373	40.4	59.6
3+	2,593	983	1,610	37.9	62.1

Table 6. Sample sizes broken down into number of students with and without direct PAL exposure for each dosage level of total PAL exposure. Counts refer to student-course observations. Direct exposure indicates PAL attached to the focal course; non-direct exposure indicates PAL exposure from same- or prior-term course(s).

Exploratory Heterogeneity Models

Exploratory heterogeneity models using the primary cumulative dosage weights were also used to examine baseline GPA, as well as differences in transfer status, URM status, and gender. Because the heterogeneity and decomposition models were exploratory and used the primary cumulative-dosage weights, these results are interpreted as evidence about patterns in the primary association rather than as separate confirmatory causal estimates.

The exploratory GPA model examined whether the association between baseline local GPA and STEM course grade differed by cumulative PAL dosage. Baseline local GPA was measured before any PAL exposure. In this model, STEM course grades continued to increase significantly with each additional PAL dosage level, even after explicitly accounting for students' baseline GPA; these consecutive dosage contrasts are reported in Table A5 of the Appendix. Table 7 reports the estimated baseline-GPA slope within each PAL dosage level. These slopes can be interpreted as the expected increase in course grade associated with a one-unit increase in baseline GPA. Larger slopes indicate a stronger association between prior academic performance and focal-course grade. The estimated GPA slope was smaller at higher PAL dosage levels, suggesting that baseline GPA was less strongly associated with course grades among students with greater PAL exposure. This pattern is consistent with possible attenuation of the relationship between prior academic performance and focal-course grades, although this exploratory result should be interpreted cautiously.

PAL Dosage	Estimated GPA Slope (95% Confidence Interval)
0	0.890 (0.837, 0.943)
1	0.744 (0.680, 0.809)
2	0.660 (0.545, 0.776)
3+	0.654 (0.542, 0.765)

Table 7. Estimated GPA slope for each PAL dosage level.

The GPA-controlled models for transfer status, URM status, and gender were designed primarily to examine whether the associations seen in the primary cumulative dosage model are consistent across different student groups. They are not intended as direct tests of differences between student groups, although numerical differences across groups are described where relevant. The exploratory transfer status model examines the relationship between cumulative PAL dosage and

transfer status. Estimated cumulative-dosage contrasts were numerically similar between transfer students and native freshmen. See Table 8.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Native Freshmen			
0 vs 1	0.145	0.0231	<0.0001
0 vs 2	0.252	0.0300	<0.0001
0 vs 3+	0.343	0.0395	<0.0001
Transfer Students			
0 vs 1	0.132	0.0353	0.0002
0 vs 2	0.279	0.0541	<0.0001
0 vs 3+	0.303	0.0691	<0.0001

Table 8. Cumulative PAL dosage contrasts (treatment versus control) for native freshmen versus transfer students.

The exploratory URM model examines the relationship between cumulative PAL dosage and URM status. Estimated cumulative-dosage contrasts were numerically larger for URM students than for non-URM students, especially at higher dosage levels. See Table 9.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Non-URM Students			
0 vs 1	0.116	0.0279	<0.0001
0 vs 2	0.218	0.0361	<0.0001
0 vs 3+	0.268	0.0467	<0.0001
URM Students			
0 vs 1	0.183	0.0314	<0.0001
0 vs 2	0.319	0.0408	<0.0001
0 vs 3+	0.435	0.0525	<0.0001

Table 9. Cumulative PAL dosage contrasts (treatment versus control) for URM versus non-URM students.

The exploratory gender model examines the relationship between cumulative PAL dosage and gender. Estimated cumulative-dosage contrasts were numerically larger for female students than for male students, particularly for the 3+ dosage contrast. See Table 10.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Female			
0 vs 1	0.141	0.0265	<0.0001
0 vs 2	0.246	0.0347	<0.0001
0 vs 3+	0.349	0.0422	<0.0001
Male			
0 vs 1	0.127	0.0292	<0.0001
0 vs 2	0.244	0.0386	<0.0001

0 vs 3+	0.287	0.0575	<0.0001

Table 10. Cumulative PAL dosage contrasts (treatment versus control) for female versus male students.

Consecutive contrasts for the transfer, URM, and gender models showed the same general pattern observed in the primary cumulative-dosage model: additional PAL exposure was associated with positive but diminishing grade increments. All consecutive contrasts were statistically significant except the 2 versus 3+ contrast among non-URM students, the 2 versus 3+ contrast among male students, and the 2 versus 3+ for transfer students. Full consecutive contrast results are reported in Tables A6-A8 of the Appendix.

Sensitivity Analysis

For sensitivity analysis, we converted the primary cumulative dosage estimates from grade points to standardized mean differences using the observed standard deviation of course grades in the analytic sample, SD = 1.296. This yielded standardized mean differences of 0.098, 0.179, and 0.246 for the 1, 2, and 3+ PAL contrasts, respectively. We then calculated E-values for each primary contrast to assess the minimum strength of association that an unmeasured confounder would need to have with both PAL exposure and course grades, conditional on the observed covariates, to explain away the estimated association. The largest contrast, the 3+ PAL contrast, had an E-value of 1.812, with a lower-bound E-value of 1.658. This indicates that an unmeasured confounder would need to be associated with both high-dosage PAL participation and subsequent course grades by risk-ratio-scale associations of approximately this magnitude, above and beyond the observed covariates, to explain away the estimated effect. See Table 11.

Comparison	Estimate	Standardized Mean Difference	E-Value (95% Lower Limit)
1 vs 0	0.127	0.098	1.412 (1.315)
2 vs 0	0.232	0.179	1.634 (1.519)
3 vs 0	0.319	0.246	1.812 (1.658)

Table 11. Sensitivity analysis results for the primary cumulative-dosage contrast.

DISCUSSION

Conclusion

Our main result is found in Table 4, in which the primary cumulative-dosage model demonstrates significant dosing effects at each level, relative to both the baseline and consecutively. That is, the more PAL sections a student takes, the stronger the positive effect. We conclude that participation in the Sacramento State PAL program can promote student success in both later STEM coursework and other concurrent STEM courses.

This result appears to be the first of its kind in the higher education literature, and has far-reaching implications for both researchers and practitioners. First, our results add to the growing literature supporting peer instruction as a powerful tool for improving student outcomes. It seems safe to assume that while the primary goal of most peer instruction programs is focal success in the focal course, positive secondary effects in concurrent and future courses is undoubtedly desirable and potentially even more impactful on broader long-term success of the participating student.

Second, while this study focused on one specific student success program at one institution, our methods readily apply to programs across the intervention spectrum. Most obviously, our results beg the question of whether other peer support programs (including, but not limited to, Peer-Led Team Learning, Supplemental Instruction, and Learning Assistant programs, not necessarily restricted to STEM disciplines or higher education) enjoy similar dosing effects. However, similar studies could also analyze the effectiveness of indirect academic support programs, from advising to mentoring to study skill development interventions. Third, knowledge of a significant dosing effect in any of these programs could serve as a valuable recruitment tool for students (in voluntary programs) and leverage for program directors to convince funders and administrators of the lasting value and impact of investment in their program. Finally, as discussed below, our study suffers from many limitations which we view as an invitation to researchers to dive deeper into what we believe is a rich, interesting, and understudied field of education scholarship.

In addition to our main result, the secondary exploratory results of Tables 5-10 both answer important questions and give rise to new ones. The decomposition model (Table 5) confirms what one would expect: that most of the effect in Table 4 is driven by direct PAL exposure. But importantly, it also confirms that there are significant non-direct effects as well. In other words, PAL exposure does indeed transfer as beneficial in future and concurrent non-focal courses.

As described in the Methods, since baseline GPA is a well-known predictor of student success, we used high school, transfer, and local GPA as covariates in our propensity score model. However, we also designed the exploratory GPA model to tease out whether the established dosing effect varied by baseline GPA. Interestingly, we found that higher PAL dosages were associated to lower GPA slopes, suggesting that baseline GPA could be less predictive of student success for students with more PAL exposure. That is, as PAL participation increases, prior GPA matters less and less.

The other exploratory heterogeneity models were first and foremost designed to determine whether the observed dosing effect was consistent across different socio-demographic groups, rather than driven by specific subpopulations. Indeed, Tables 8-10 show that native freshmen and transfers, URM and non-URM, and female and male students all enjoyed significant dosing

effects. While the tables show some interesting differences among the various subpopulations, one must exercise caution when interpreting the results from any of our exploratory models; see below for details.

Our primary statistical contribution is a reproducible approach for estimating cumulative and transferrable associations in administrative course data. Stronger causal claims would require experimental or quasi-experimental variation in PAL participation.

Limitations and Future Research

Our study suffers from many limitations, some of which can be taken as invitations to further research. First, we studied a single program at a single institution; it would be interesting to investigate whether this effect holds for similar peer programs elsewhere. Perhaps more importantly, our dataset was limited to courses served by the PAL program: the 20 courses listed in Table A1 of the Appendix. Therefore, we can say nothing about transferability of learning to courses outside this list, including any non-STEM courses. Further, while the 20 courses served by PAL cover a reasonably broad swath of the lower-division STEM curriculum, they were determined non-scientifically by local, logistical, and financial factors outside the control of this analysis. A study which managed to account for outcomes in all student coursework would be a significant improvement on our work here.

Moreover, we have conducted a technical statistical analysis which is unable to explain exactly what about the PAL exposure is responsible for the established dosing effect. While our exploratory models examine same-term and prior-term exposure as an approximation of transfer, they do not directly examine the mechanisms of transfer. In other words, we don't know precisely why we are seeing this effect. We have some ideas, ranging from improved study skills, to strengthened foundational knowledge, to sense of belonging in STEM. However, rigorously studying this question would require theoretical and qualitative research, which is outside the realm of this quantitative study.

As mentioned above, the decomposition and heterogeneity models are exploratory because they use the primary cumulative-dosage weights rather than separately reweighting on each new set of components. Consequently, the results of our exploratory models are subtle, as we did not set out a priori to compare these specific academic (baseline GPA) and demographic (URM status, transfer status, gender) subgroups. In particular, the observed differences among the various subgroups are interesting but inconclusive; carefully exploring these differences would require designing a propensity score analysis for this express purpose, which could potentially be difficult to balance. While our purpose was to demonstrate that there was in fact a dosing effect in each of these groups of students, it would be interesting to explore any differences further in future research.

In terms of statistical limitations, the observational causal inference framework results in estimates that depend on assumptions of adequate observed confounding adjustment, sufficient overlap in the treatment populations, and no unmeasured confounding. Propensity score weighting improves comparability on observed covariates, but it cannot address unmeasured factors such as motivation, scheduling conflicts, or awareness of PAL. By using overlap weights, we examine students who are plausible members of multiple PAL dosage groups. This improves group comparability, but results should not be generalized uncritically to all students. For all models, higher dosage levels have smaller sample sizes. Estimates for these groups are less precise and more dependent on the overlap-weighted sample. The E-value sensitivity analysis quantifies the strength of unmeasured confounding needed to explain away the primary analysis, but it does not demonstrate that such confounding is absent.

Finally, since direct, same-term, and prior-term PAL exposures can occur simultaneously, decomposition models should be interpreted as evidence of the cumulative association rather than as independent causal effects.

Acknowledgments

We thank Jennifer Lundmark and Vincent Pigno for their support.

Declarations

All three authors have been affiliated with the PAL program in some way as part of their academic appointment. The authors receive no additional compensation in these roles.

This work was supported in part by NSF award number 2524510.

This project was approved by the Research Integrity and Compliance Officer in the Offices of Research, Innovation, and Economic Development at Sacramento State under the Human Subjects Research Institutional Review Board 15−16−143.

APPENDIX

Courses Studied

Course Code	Course Title
BIO 22	Introductory Human Anatomy
BIO 25	Human Anatomy and Physiology I
BIO 26	Human Anatomy and Physiology II
BIO 121	Molecular Cell Biology
BIO 131	Systemic Physiology
BIO 139	General Microbiology
BIO 184	General Genetics
CHEM 1A	General Chemistry I
CHEM 1B	General Chemistry II
CHEM 4	Chemical Calculations
CHEM 6A	Chemistry for Nurses
CHEM 24	Organic Chemistry I
CHEM 124	Organic Chemistry II
MATH 12	Algebra for STEM Courses
MATH 29	Pre-Calculus Mathematics
MATH 30	Calculus I
MATH 31	Calculus II
MATH 32	Calculus III
PHYS 5A	General Physics: Mechanics, Heat, Sound
STAT 1	Introduction to Statistics

Table A1. Courses used in analysis.

Covariates Used in Propensity Score Model

Covariate	Explanation
High school GPA	High school GPA with mean imputation on missing values
Transfer GPA	Transfer GPA with mean imputation on missing values
Local GPA	Baseline local GPA with mean imputation on missing values
High school GPA sissing	Missingness indicator for high school GPA
Transfer GPA missing	Missingness indicator for transfer GPA
Local GPA missing	Missingness indicator for baseline local GPA
COH	Indicates transfer or native freshman
Units	Number of units the student is enrolled in this term
Remedial units	Number of remedial units the student has taken
New at baseline	Whether the student was newly enrolled at their earliest instance in the data
Home income	Median income in student's home Zip Code at time of application (based on Census 5-year American Community Survey data released in the year before the application year) with mean imputation on missing values
School income	Median income in student's high school Zip Code at time of application, with mean imputation on missing values
Home income missing	Missingness indicator for home income
School income missing	Missingness indicator for school income
Gender	Gender (restricted to male/female due to small cell sizes)
Pell flag	Indicator for student Pell eligibility
Ethnicity	IPEDS ethnicity
Mother’s education	Mother’s highest level of education
Father’s education	Father’s highest level of education
Base time	Number of terms elapsed since first term as a degree-seeking student at Sac State.
Term age	Student's age at the beginning of the term

Table A2. All covariates used in the propensity score model.

Balance Diagnostics

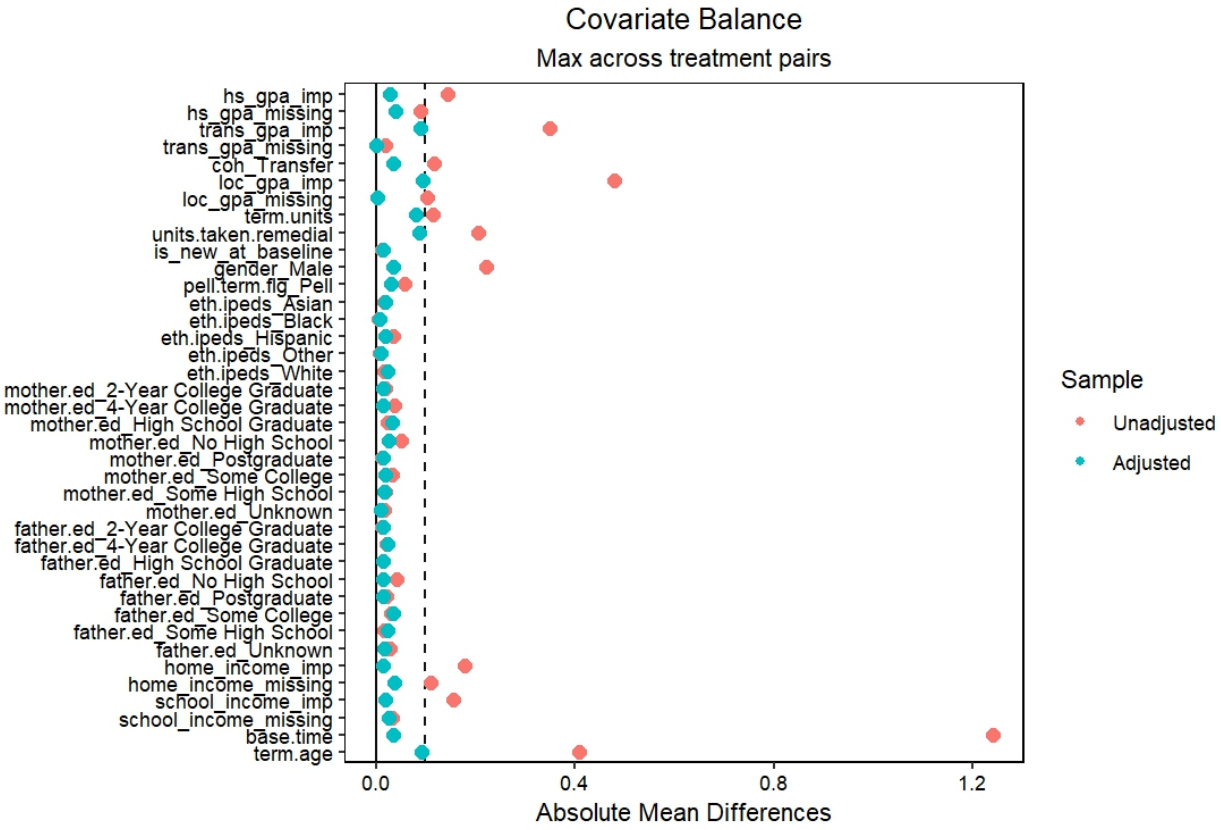

Figure A1. Love plot showing covariate balance. The dashed line shows an absolute mean difference of 0.1, with absolute mean differences for all covariates falling below this value.

Complete Decomposition Models Results

Non-Direct PAL Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
0 vs 1	0.0494	0.0203	0.0442
1 vs 2	0.0489	0.0300	0.2061
2 vs 3+	0.0332	0.0429	0.4386

Table A3. Consecutive contrasts of grade differences across non-direct PAL dosage levels.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Treatment vs Control			
0 vs 1	0.0510	0.022	0.0396
0 vs 2	0.0843	0.032	0.0255
0 vs 3+	0.0975	0.047	0.0396

Consecutive Contrasts			
0 vs 1	0.0510	0.0220	0.0594
1 vs 2	0.0330	0.0323	0.6133
2 vs 3+	0.0133	0.0467	0.7762

Table A4. Treatment-versus-control and consecutive contrasts of grade differences for prior-term PAL dosage levels.

Post Hoc Models

Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
0 vs 1	0.1518	0.0200	<0.0001
1 vs 2	0.1212	0.0277	<0.001
2 vs 3+	0.0853	0.0400	0.0328

Table A5. Consecutive contrasts of grade differences for PAL dosage levels in a model with baseline local GPA included as a covariate allowed to interact with cumulative PAL dosage.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Native Freshmen			
0 vs 1	0.1447	0.0231	<0.0001
1 vs 2	0.1068	0.0297	0.0007
2 vs 3+	0.0910	0.0412	0.0272
Transfer Students			
0 vs 1	0.1322	0.0353	0.0006
1 vs 2	0.1465	0.0554	0.0164
2 vs 3+	0.0239	0.0782	0.7601

Table A6. Decomposition model contrasts (consecutive) of grade differences for PAL dosage levels, broken down into native freshmen versus transfer students.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Non-URM Students			
0 vs 1	0.1164	0.0279	<0.0001
1 vs 2	0.1016	0.0373	0.0130
2 vs 3+	0.0502	0.0511	0.3252
URM Students			
0 vs 1	0.1833	0.0314	<0.0001
1 vs 2	0.1359	0.0406	0.0016
2 vs 3+	0.1154	0.0543	0.0335

Table A7. Decomposition model contrasts (consecutive) of grade differences for PAL dosage levels, broken down into URM versus non-URM students.

Decomposition Dosage Contrast	Estimate	Standard Error	Holm-Adjusted P-Value
Female			
0 vs 1	0.1407	0.0265	<0.0001
1 vs 2	0.1049	0.0340	0.0041
2 vs 3+	0.1039	0.0447	0.0199
Male			
0 vs 1	0.1272	0.0292	<0.0001
1 vs 2	0.1171	0.0406	0.0079
2 vs 3+	0.0427	0.0611	0.4844

Table A8. Decomposition model contrasts (consecutive) of grade differences for PAL dosage levels, broken down into female versus male students.

REFERENCES

Akhavan, N., Davison, J., Taylor, E., Hill, J., Mosley, V., & Salimo, K. J. (2024). The Community and Structure of Sacramento State's Peer Assisted Learning Program. *Advances in Peer-Led Learning*, *4*, 19–27. https://doi.org/10.54935/apll2024-01-03-19

Allison, P. D. (2009). Missing data. *The SAGE handbook of quantitative methods in psychology*, *23*, 72-89. https://doi.org/10.4135/9780857020994.n4

Arendale, D. R. (2022). *2022 annotated bibliography of postsecondary peer cooperative learning programs*. University Digital Conservancy. https://hdl.handle.net/11299/254251

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, *34*(28), 3661-3679. https://doi.org/10.1002/sim.6607

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Blair G., Cooper J., Coppock A., Humphreys M, Sonnet L (2025). *estimatr: Fast Estimators for Design-Based Inference*. doi:10.32614/CRAN.package.estimatr. R package version 1.0.6.

Colin Cameron, A., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, *50*(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*, *19*(2), 237–248. https://doi.org/10.3102/00028312019002237

Conn, V. S., & Chan, K. C. (2015). How Much, How Often, How Long? Addressing Dosage in Intervention Studies. *Western Journal of Nursing Research*, *38*(1), 3–4. https://doi.org/10.1177/0193945915605067

Conn, V. S., & Groves, P. S. (2011). Protecting the power of interventions through proper reporting. *Nursing Outlook*, *59*(6), 318–325. https://doi.org/10.1016/j.outlook.2011.06.003

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*(4), 732–749. https://doi.org/10.1037/0022-0663.98.4.732

González Canché, M. S. (2019). Community College Students Who Attained a 4-Year Degree Accrued Lower Student Loan Debt than 4-Year Entrants Over 2 Decades: Is a 10 Percent Debt Accumulation Reduction Worth the Added "Risk"? If So, for Whom?. *Research in Higher Education*, *61*(7). https://doi.org/10.1007/s11162-019-09565-9

Greifer, N. (2026). *cobalt: Covariate Balance Tables and Plots*. doi:10.32614/CRAN.package.cobalt. R package version 4.6.2.

Greifer, N. (2026). *WeightIt: Weighting for Covariate Balance in Observational Studies*. doi:10.32614/CRAN.package.WeightIt. R package version 1.7.0.

Groenwold, R. H., White, I. R., Donders, A. R. T., Carpenter, J. R., Altman, D. G., & Moons, K. G. (2012). Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. *Canadian Medical Association Journal*, *184*(11), 1265-1269. https://doi.org/10.1503/cmaj.110977

Haskell, R. E. (2001). *Transfer of Learning*. Academic Press.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199-236. https://doi.org/10.1093/pan/mpl013

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70. https://www.jstor.org/stable/4615733

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*(467), 854-866. https://doi.org/10.1198/016214504000001187

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*(3), 706-710. https://doi.org/10.1093/biomet/87.3.706

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, *89*(1), 47–62. https://doi.org/10.1037/0033-2909.89.1.47

Karp, M. M., Calcagno, J. C., Hughes, K. L., Jeong, D. W., & Bailey, T. R. (2007). *The postsecondary achievement of participants in dual enrollment: An analysis of student outcomes in two states.* National Research Center for Career and Technical Education, University of Minnesota. https://ccrc.tc.columbia.edu/publications/dual-enrollment-student-outcomes.html

Kortecamp, K. & Peters, M. L. (2024). The Impact of a high-dosage tutoring program on reading achievement of beginning readers: A multi-level analysis. *Journal of Education for*

Students Placed at Risk, 29(3), 291–309. https://doi.org/10.1080/10824669.2023.2179056

Krauel, M. & Shanbrom, C. (2026). The influence of a Peer Assisted Learning program on the career readiness of its student facilitators. In review.

Lee, H. B., & Villarreal, M. U. (2022). Should students falling behind in school take dual enrollment courses? *Journal of Education for Students Placed at Risk (JESPAR), 28*(4), 439–473. https://doi.org/10.1080/10824669.2022.2100994

Lenth, R. & Piaskowski, J. (2026). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. doi:10.32614/CRAN.package.emmeans. R package version 2.0.3.

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, *113*(521), 390-400. https://doi.org/10.1080/01621459.2016.1260466

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *biometrika*, 13-22. https://doi.org/10.1093/biomet/73.1.13

Linden, A., Mathur, M. B., & VanderWeele, T. J. (2020). Conducting sensitivity analysis for unmeasured confounding in observational studies using E-values: the evalue package. *The Stata Journal*, *20*(1), 162-175. https://doi.org/10.1177/1536867x20909696

Liu, V. Y. T., Chelliah, B., Saunders, T., Ju, H., & Viera, C. (2025). The More the Merrier? The Dosage Effect of Dual Enrollment Credits Earned in High School on College Enrollment and Degree Completion? *Research in Higher Education*, *67*(1). https://doi.org/10.1007/s11162-025-09872-4

Lundmark, J., Paradis, J., Tashiro, L., Kapp, M., & Lowe, E. (2017). Development and Impact of a Training Program for Undergraduate Facilitators of Peer-Assisted Learning. *Journal of College Science Teaching*, *046*(06). https://doi.org/10.2505/4/jcst17_046_06_50

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, *9*(4), 403. https://doi.org/10.1037/1082-989x.9.4.403

Mitchell, Y. D., Ippolito, J., & Lewis, S. E. (2012). Evaluating peer-led team learning across the two-semester general chemistry sequence. *Chemistry Education Research and Practice, 13*(3), 378–383. https://doi.org/10.1039/c2rp20028g

Perkins, D. N., & Salomon, G. (1992). Transfer of Learning. In T. Husén, & T. N. Postlethwaite (Eds.), The International Encyclopedia of Education (2nd ed., pp. 425-441). Oxford: Pergamon.

Roberts, G. J., Dumas, D. G., McNeish, D., & Coté, B. (2021). Understanding the Dynamics of Dosage Response: A Nonlinear Meta-Analysis of Recent Reading Interventions. *Review of Educational Research*, 003465432110514. https://doi.org/10.3102/00346543211051423

R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. doi:10.32614/R.manuals.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.2307/2335942

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3), 808-840. https://doi.org/10.1214/08-aoas187

Shanbrom, C., Norris, M., Esgana, C., Krauel, M., Pigno, V., & Lundmark, J. (2023). Assessing Student Success in a Peer Assisted Learning Program Using Propensity Score Matching. *The Journal of College Science Teaching*, *52*(7), 129–136. https://doi.org/10.1080/0047231x.2023.12315888

Shanbrom, C. & Perry, L. (2025). The Dosing Effect of Peer-Assisted Learning in Undergraduate Organic Chemistry. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.5c00485

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1. https://doi.org/10.1214/09-sts313

Utschig, T. T., & Sweat, M. (2008). Implementing Peer Led Team Learning in first-year programming courses. *2008 38Th Annual Frontiers in Education Conference*, F3C–13-F3C-18. https://doi.org/10.1109/fie.2008.4720641

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of internal medicine*, *167*(4), 268-274. https://doi.org/10.7326/m16-2607

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, 817-838. https://doi.org/10.2307/1912934