Chemical Education Research

edited by
Diane M. Bunce
The Catholic University of America
Washington, D.C. 20064

Christopher F. Bauer
University of New Hampshire
Durham, NH 03824

Departing from Lectures: An Evaluation of a Peer-Led Guided Inquiry Alternative

Scott E. Lewis and Jennifer E. Lewis*
Department of Chemistry, University of South Florida, Tampa, FL 33620; **jlewis@cas.usf.edu*

General chemistry courses at the University of South Florida have historically been taught in lecture format to accommodate the large class size, typically about 200 students.[1] To create a more student-centered learning experience, a reform method was employed based on the Peer-Led Team Learning (PLTL) developed by the National Science Foundation systemic change initiative *(1, 2)*. In this method, students work in groups of 10, with each group assigned to one peer leader. The peer leader is an undergraduate student who has successfully completed the general chemistry course.

Guided inquiry has shown promise in recent studies *(3);* we chose a guided-inquiry methodology as described by Farrell et al. *(4)* to be the basis for this reform initiative. Fortunately the PLTL scenario facilitates the implementation of the guided-inquiry methods, which can be used in smaller groups (3–4 students) within the already small groups of a typical PLTL workshop. This combination results in what we term peer-led guided inquiry (PLGI). The pedagogical focus of PLGI is on student–student interactions within small groups, with a peer leader acting as a facilitator for those interactions.

Methods

To evaluate the effectiveness of PLGI methods in terms of assisting student understanding in a college-level general chemistry course, an experiment was run comparing two sections of general chemistry. The experiment consisted of two sections of General Chemistry I that were taught concurrently by the same instructor. It was not possible to randomly assign students to each section, although during course selection students had no prior knowledge that the two sections would be different. During the first week of a semester, students are permitted to drop courses (and potentially switch sections) with no penalty. On the first day of class, students in the experimental section were told of the nature of the course, and no unusual drop rate was observed.

The control section met three times each week for 50-minute lectures. The experimental section met for two 50-minute lecture sections and one 50-minute PLGI session each week. In the PLGI sessions, students met in small groups of 10 students that were led by another undergraduate student, the peer leader, who had already successfully completed general chemistry. During the session, students worked through one or two activities from a published text *(5)*. The published activities, originally designed for "lectureless" small group learning, incorporate a learning cycle approach *(6),* in which students explore information in order to discover the need for new (to the students) concepts and subsequently "invent" and apply those concepts as the activity continues. In accordance with the learning cycle approach, the activities were chosen to precede the lectures on those topics. The role of the peer leaders, then, was not to introduce new material but rather to check for understanding of the new material as the students progressed through the activities.

Peer leaders were enrolled in a training course, taught by one of the authors, both to prepare for the guided inquiry pedagogy that was to be used and to review the topic material by working through the activities in small groups themselves prior to their meetings with students. The training course met once each week for two-hour increments. None of the ten peer leaders had any prior experience with this approach, nor had any been working for the department as tutors. Only two were majoring in chemistry (one junior and one senior), while the remainder (four seniors, one junior, and three sophomores) came primarily from interdisciplinary science majors, with one engineering student.

In prior studies, the implementation of a Peer-Led Team Learning approach has been accomplished by replacing a problem-solving recitation session led by a graduate student with a longer problem-solving session led by an undergraduate *(7)*. Since our current general chemistry course does not have a recitation session but is lecture only, the decision was made to replace one of the three weekly lectures in the experimental section with a PLGI session. As a result, this study also directly addresses the belief that interactive learning comes at the expense of content, which has been cited as a primary reason why many instructors do not employ such techniques *(8)*.

Throughout the semester, one of the authors observed the experimental section and another "regular" section (it was not possible for the author to observe the control section), taught by a different experienced instructor, in order to gather information about how the experimental section lecture content might be different from that in a typical section. The instructor's textbook chapter coverage on the syllabus, lecture notes, and presentation aids were the same for both the experimental and control sections, and the same topics were presented in the experimental section and the other observed section; they simply came faster and with fewer examples (e.g., worked problems) in the experimental section.

The general chemistry lectures are typically capped at 190 students, and generally run near capacity. This was the case for the control section. For the practical purposes of setting up the PLGI sessions, the experimental section was capped at 100 students. As a result of this, the enrollment for the control section was 178 students and the experimental section was 86 students. A comparison of average SAT and ACT scores for these sections is shown in Tables 1 and 2.

None of the differences were statistically significant at the $p = 0.05$ level. The differing sample sizes from Table 1 and Table 2 are a result of students enrolled in General Chemistry I having taken only one of the two (SAT and ACT) tests.

During the semester, students in both sections took four instructor-constructed exams and an American Chemical Society exam at the end of the course *(9)*. A significant portion (75%) of students' grades was dependent on their performance on the course exams and the ACS final exam. The students were allowed to drop their lowest course exam score. As is customary, the exams comprised contributions from each instructor teaching general chemistry for the current semester, including the instructor in this study, but not either of the authors. The tests for all sections were identical and administered at the same time, and thus serve as an appropriate basis for comparison between the two sections in the study.

Table 1. Comparison of Average SAT Scores

Student Section Assignment	Average SAT Total Score	Average SAT Mathematics Score	Average SAT Verbal Score
Control (N = 142)	1091	556	535
Experimental (N = 69)	1092	556	536

Table 2. Comparison of Average ACT Scores

Student Section Assignment	Average ACT Total Score	Average ACT Mathematics Score	Average ACT Verbal Score
Control (N = 87)	22.6	22.9	22.1
Experimental (N = 54)	22.1	22.7	22.0

Results and Discussion

Comparing the performance of these students on the course exams and final exam provides an indication of the effectiveness of the PLGI intervention. In order to maintain a similar sample of students throughout the test score comparison, student data for those who had taken fewer than three exams were removed from the analysis. Students were able to drop their lowest exam score, so this criterion removed only students who did not complete the course. Of the students who completed the course, 70 were in the experimental section and 154 in the control section. A comparison of SAT and ACT scores for these students still showed no statistically significant difference between the sections. Initial comparisons (Table 3) show the students in the experimental section to perform better than the control section.

From Table 3, it can be seen that the experimental group consistently outperformed the control group on the course exams and on the final exam. All of these differences were significant at the $p = 0.05$ level. Additional comparisons were run considering all the students who had taken each exam (not omitting those who missed more than two exams) and similar results were found, except we found no significant difference at the $p = 0.05$ level for Exam 1. In addition to the t-test results in Table 3, the effect size was also estimated using Cohen's d. This provides an indication of the extent that the two sections differ. Conventionally, a d of 0.2 is considered a small difference, while 0.5 is considered a moderate or noticeable difference *(10)*. From Table 3, Cohen's d indicated that the difference in performance between the two sections became larger as the course progressed. This may be a result of a progressive impact of the PLGI sessions: the sessions would be expected to benefit the students more as they had more exposure to this learning method. The reduced Cohen's d value for the Final Exam might be attributed to the different nature of this exam, since it was an ACS exam

Table 3. Comparison of Course Exam Scores by Student Group

Course Exam	Student Group	Mean Score	Standard Deviation	t-Test Values	p Values	Cohen's d Values
Exam 1	Control (N = 154)	48.3	16.4	2.160	0.032	0.298
	Experimental (N = 70)	53.7	19.6			
Exam 2	Control (N = 154)	55.5	16.6	2.948	0.004	0.414
	Experimental (N = 69)	63.1	19.7			
Exam 3	Control (N = 151)	55.9	21.9	3.415	0.001	0.506
	Experimental (N = 70)	66.3	19.0			
Exam 4	Control (N = 153)	42.8	19.0	3.930	0.000	0.563
	Experimental (N = 70)	53.7	19.6			
Final Exam	Control (N = 151)	50.9	16.2	2.647	0.009	0.367
	Experimental (N = 70)	57.6	20.1			

Note: See ref *10* for derivation of Cohen's d statistic.

instead of the previous instructor-created tests. While these results lead to the conclusion that the PLGI sessions were successful, they leave open the question of an alternative explanation for the difference in scores for the two groups, aside from the PLGI sessions, such as class size difference.

To investigate this possibility and to better understand the relationship between PLGI sessions and student performance, we considered attendance at the PLGI sessions. Twelve PLGI sessions were offered, and attendance constituted a very small portion (2.75%) of the students' grades in the experimental section. To determine the effects of PLGI sessions on student performance, student attendance at the sessions was included in a regression model (all control group students were assigned a 0 for PLGI session attendance) in which the dependent variable was exam score. For each regression model, attendance was considered only up to the date of the exam.

There is a problem with assigning each member of the control group a 0 for attendance. To do so is to assume that the control group is equivalent to those in the experimental group who never attended a PLGI session. Since attendance at PLGI sessions was related to preparation for college as defined by SAT scores (r = 0.319 between PLGI attendance and SAT mathematics score, r = 0.342 for PLGI attendance and SAT verbal score, both relationships $p < 0.05$), and since SAT scores have been shown to relate to student performance in chemistry *(11),* it appears that those who did not attend the PLGI sessions are not equivalent to the control group.

To account for this difference, SAT mathematics score and SAT verbal score were added to the regression models as additional predictors. This decision removes the students who did not take the SAT from the model. These students (N = 25, 16.2% for control, and N = 10, 14.3% for experimental) represent an equally small amount of the sample for both sections. Lecture section was also added to the model in order to determine whether the lecture portion of the sections could be considered to be equivalent, while considering the other parameters previously discussed. In essence this is evaluating whether the sections were different by some alternative factors that have been unaccounted for, such as class size or class meeting time. The results are presented in Tables 4–8; significant predictors are given in italics.

Table 4. Regression Model Results, Exam 1[a]

Parameters Correlated	B Values	β Values	t Values	p Values
Constant	–21.597		–2.956	0.004
SAT Verbal Score	0.0346	0.179	2.405	*0.017*
SAT Math. Score	0.0930	0.460	6.180	*0.000*
Attendance (0–3)	1.687	0.126	0.571	0.569
Section Assignment[b]	–0.274	–0.007	–0.033	0.973

[a]Exam 1 score range: 0–100; N = 189; R^2 = 0.365. Significant predictors are shown in italics. [b]Control = 0, Experimental = 1

Table 5. Regression Model Results, Exam 2[a]

Parameters Correlated	B Values	β Values	t Values	p Values
Constant	7.341		0.858	0.392
SAT Verbal Score	0.0288	0.145	1.737	0.084
SAT Math. Score	0.0565	0.272	3.272	*0.001*
Attendance (0–6)	3.681	0.509	1.983	*0.049*
Section Assignment[b]	-11.950	-0.311	-1.213	0.227

[a]Exam 2 score range: 0–100; N = 188; R^2 = 0.220. Significant predictors are shown in italics. [b]Control = 0, Experimental = 1

Table 6. Regression Model Results, Exam 3[a]

Parameters Correlated	B Values	β Values	t Values	p Values
Constant	6.563		0.627	0.532
SAT Verbal Score	0.0217	0.091	1.077	0.283
SAT Math. Score	0.0634	0.253	3.004	*0.003*
Attendance (0–9)	3.326	0.550	2.321	*0.021*
Section Assignment[b]	-13.510	-0.294	-1.243	0.215

[a]Exam 3 score range: 0–100; N = 186; R^2 = 0.203. Significant predictors are shown in italics. [b]Control = 0, Experimental = 1

Table 7. Regression Model Results, Exam 4[a]

Parameters Correlated	B Values	β Values	t Values	p Values
Constant	-22.067		-2.440	0.016
SAT Verbal Score	0.0523	0.235	2.978	*0.003*
SAT Math. Score	0.0658	0.284	3.600	*0.000*
Attendance (0–12)	0.959	0.220	1.117	0.265
Section Assignment[b]	1.506	0.035	0.180	0.857

[a]Exam 4 score range: 0–100; N = 188; R^2 = 0.300. Significant predictors are shown in italics. [b]Control = 0, Experimental = 1

Table 8. Regression Model Results, Final Exam[a]

Parameters Correlated	B Values	β Values	t Values	p Values
Constant	-21.144		-3.007	0.003
SAT Verbal Score	0.0461	0.238	3.388	*0.001*
SAT Math. Score	0.0837	0.414	5.889	*0.000*
Attendance (0–12)	1.876	0.493	2.809	*0.006*
Section Assignment[b]	-11.759	-0.315	-1.803	0.073

[a]Final exam score range: 0–100; N = 187; R^2 = 0.445. Significant predictors are shown in italics. [b]Control = 0, Experimental = 1

As can be seen from the regressions for each exam, PLGI session attendance did not make a significant difference for Exam 1. This makes sense considering the few sessions offered prior to Exam 1. By Exam 2, PLGI session attendance becomes a significant positive factor in the exam scores for the students. This effect continues for each exam except Exam 4, where attendance was not found to be a significant factor in student performance. This is most likely a result of students focusing on their final exams, since Exam 4 occurs during the final week of class, and the students did perform below their average on this fourth exam (Table 3). Also, any student satisfied with the scores received on Exams 1 through 3 would be able to drop the score for Exam 4 without a course grade penalty, providing another likely possibility for the low average and alternative findings for Exam 4.

Noteworthy from the regression models is the continued non-significance of the lecture section. This provides an indication that the students in both the control and experimental section were equivalent when controlling for PLGI session attendance and college preparedness, and discounts the possibility of alternative explanations for the difference in scores, such as class size or meeting time.

Readers may discern a somewhat large negative regression coefficient present for lecture section in some of the models, notably Exam 2 (Table 5), Exam 3 (Table 6), and the Final Exam (Table 8). While at first glance this appears to indicate that students in the experimental section performed much worse than the control section, it should be pointed out that this is only when controlling for PLGI session attendance and SAT scores. The negative regression coefficient indicates that a student in the experimental section who did not attend any of the PLGI sessions would perform this much worse than a student in the control section with a similar SAT score. This is representative of comparing a student in the experimental section who did not attend any PLGI sessions with students in the control section, provided they had similar SAT scores. Since attendance is thought to be correlated with performance, a negative relationship for lecture section is to be expected. To be clear, the negative relationship does not indicate that the experimental section performed worse than the control group, as this coefficient does not account for PLGI session attendance.

To decide which significant parameter had the greatest impact on student performance, the standardized regression weights (β) can be compared within each model *(12).* The β weight describes the expected change for the dependent variable, in standard deviations, given a change of one standard deviation for the independent variables. For example, the β weight for the SAT verbal score in the Exam 1 model is 0.179 (Table 4). That means a student who scored one standard deviation above the mean for the SAT verbal test could expect to score 0.179 standard deviations above the mean of Exam 1 (not accounting for the other parameters in the model). Like the raw regression coefficients described above, it should be pointed out that the β still represents the partial effects of each variable, or the effect of the variable when all the other variables are controlled for. (See Tables 4–8 for exam- and variable-specific data.)

We used a modified *t*-test to compare the standardized regression weights—a pair of β's—within each model. (See column 2 of Tables 4–8 for β values.) Equation 1 describes the relationships among the variables:

$$t = \frac{\beta_i - \beta_j}{SE_{\beta_i - \beta_j}} \quad (1)$$

$$df = n - k - 1 \quad (2)$$

where n represents the number of people in the sample, k represents the number of parameters in the model, and df represents the degrees of freedom. Standard error, or SE, can be found by eq 3:

$$SE_{\beta_i - \beta_j} = \sqrt{\frac{1 - R^2_{Y \bullet 12 \ldots k}}{n - k - 1}\left(r^{ii} + r^{jj} - 2r^{ij}\right)} \quad (3)$$

where r^{ii}, r^{jj}, and r^{ij} represent the elements of the inverse correlation matrix between the parameter i with itself, j with itself, and between i and j respectively *(13).* By performing this analysis for each model between each significant parameter, the relative impact of each parameter can be assessed.

Only one significant difference was found; this is between the SAT mathematics score and SAT verbal score for the model for Exam 1, which may be an indication of the focus on mathematics present in Exam 1. Among the other models, no significant difference was found between any two significant parameters. This is an indication that for the Exam 2, Exam 3 and Final Exam models, attendance at the PLGI sessions had, statistically, as much impact on student performance as the SAT mathematics score and, where it entered the model significantly (Exam 1 and the Final Exam), the SAT verbal score.

The results from this analysis indicate that a student who attends PLGI sessions can be expected to perform better on exams than another student at the same SAT level. This is especially impressive considering that students in the PLGI sessions did receive one less lecture per week than those in the control group. It should be noted that this is in agreement with previous research findings, which highlight the lack of evidence supporting the hypothesis that lecture promotes learning *(14).*

Surveying Student Responses to Peer-Led Guided Inquiry

To characterize students' responses to the PLGI session, a nine question open-ended survey was given during the last session. All 54 students in attendance responded to this survey. The results from the survey were largely positive, but this may be a result of only administering the survey to those students who attended the last session, since it is believed that a student who was strongly dissatisfied with the PLGI sessions may stop attending, since the point deductions would be minimal.

However, for the students who attended the final session, 74% thought the sessions were beneficial, with only two students (4%) giving a completely negative response. When asked if they thought the PLGI sessions made up for the missed lecture, 76% of the students responded positively.

Only seven felt that they were "shorted" due to the missed lecture. Five of those seven respondents felt that the two remaining lectures were rushed to compensate for the missing third lecture. Additionally, 20 respondents indicated that the PLGI sessions helped in understanding concepts or assisted in problem solving.

Given the chance to continue taking the PLGI sessions in the second semester of chemistry, 85% indicated that they would, with 58% of those indicating that they would attend the PLGI sessions even if they were offered as voluntary supplements to the course. Finally, indicative of the pedagogy used, 76% believed that working with groups was beneficial, with only five students who believed the group work slowed them down.

Conclusion

While the PLGI session attendance has been shown to have a significant impact on student performance, there are a number of factors open for future work. For example, PLGI influence on student retention of concepts over time was not probed, and would be another means to evaluate the effectiveness of the sessions. Also unexplored are the factors that may have contributed to students not attending the PLGI sessions. In the experimental section, 48 of the 70 students attended at least 9 of the 12 sessions, which leaves 22 of the 70 students who missed more than 3 sessions. It is interesting to note that only 1 of the 70 did not attend any of the PLGI sessions, so a large majority did give the sessions a try. The low-attendees may or may not have benefited if they had attended more PLGI sessions; this could be investigated by increasing the contribution of PLGI session attendance to the students' overall grade formula to foster increased attendance.

Finally, evidence indicates that replacement of one lecture a week with a reform-oriented teaching intervention may serve as an appropriate testing ground for the feasibility and effectiveness of the intervention. Fears that students who had less exposure to lecture would learn less proved to be groundless in this study.

Acknowledgments

The authors thank the general chemistry instructor who enabled this research to be conducted; we also thank Jeffrey Kromrey for his invaluable consultation concerning the statistical analyses we present here. This paper is based upon work supported in part by the National Science Foundation under Grant No. DUE-0310954.

Note

1. A typical student population for a first-semester general chemistry course at our institution has 54% first-year students and 26% second-year students; 57% are female; students who report Asian, Black, and Hispanic heritage constitute 10% for each group.

Literature Cited

1. Gosser, D. K.; Roth, V. *J. Chem. Educ.* **1998,** *75,* 185–187.
2. Gosser, D. K.; Strozak, V. S.; Cracolice, M. S. *Peer-Led Team Learning: Workshops for General Chemistry;* Prentice Hall: Upper Saddle River, NJ, 2001.
3. Moog, R. S. Assessment of the Guided Inquiry Approach to Chemistry Instruction at Franklin and Marshall College. Presented at the 215th American Chemical Society National Meeting: Dallas, TX, 1998.
4. Farrell, J. J.; Moog, R. S.; Spencer, J. N. *J. Chem. Educ.* **1999,** *76,* 570–574.
5. Moog, R. S.; Farrell, J. J. *Chemistry: A Guided Inquiry,* 2nd ed.; John Wiley & Sons: New York, 2002.
6. Karplus, R. *J. Res. Sci. Teach.* **1977,** *14,* 169–175.
7. Tien, L. T.; Roth, V.; Kampmeier, J. A. *J. Res. Sci. Teach.* **2002**, *39,* 606–632.
8. Cooper, M. M. *J. Chem. Educ.* **1995**, *72*, 162–164.
9. *First Term General Chemistry (Special Examination).* Examinations Institute of the American Chemical Society, Division of Education; Clemson University, Clemson, SC, 1997.
10. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences,* 2nd ed.; L. Erlbaum Associates: Hillsdale, NJ, 1988.
11. Bunce, D. M.; Hutchinson, K. D. *J. Chem. Educ.* **1993,** *70,* 183–187.
12. Rencher, A. C. *Linear Models in Statistics;* John Wiley & Sons: New York, 2000.
13. Cohen, J.; Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences,* 2nd ed.; L. Erlbaum Associates: Hillsdale, NJ, 1983.
14. Birk, J. P.; Foster, J. *J. Chem. Educ.* **1993,** *70,* 180–182.