

Article

Undergraduates Achieve Learning Gains in Plant Genetics through Peer Teaching of Secondary Students

H. E. Chrispeels,* M. L. Klosterman,[†] J. B. Martin,* S. R. Lundy,* J. M. Watkins,*
C. L. Gibson,* and G. K. Muday*

*Department of Biology and [†]Provost's Office of Global Affairs, Wake Forest University, Winston-Salem, NC 27109

Submitted January 20, 2014; Revised June 14, 2014; Accepted July 24, 2014
Monitoring Editor: Nancy Moreno

This study tests the hypothesis that undergraduates who peer teach genetics will have greater understanding of genetic and molecular biology concepts as a result of their teaching experiences. Undergraduates enrolled in a non-majors biology course participated in a service-learning program in which they led middle school (MS) or high school (HS) students through a case study curriculum to discover the cause of a green tomato variant. The curriculum explored plant reproduction and genetic principles, highlighting variation in heirloom tomato fruits to reinforce the concept of the genetic basis of phenotypic variation. HS students were taught additional activities related to molecular biology techniques not included in the MS curriculum. We measured undergraduates' learning outcomes using pre/postteaching content assessments and the course final exam. Undergraduates showed significant gains in understanding of topics related to the curriculum they taught, compared with other course content, on both types of assessments. Undergraduates who taught HS students scored higher on questions specific to the HS curriculum compared with undergraduates who taught MS students, despite identical lecture content, on both types of assessments. These results indicate the positive effect of service-learning peer-teaching experiences on undergraduates' content knowledge, even for non-science major students.

INTRODUCTION

Many colleges and universities have core requirements for graduation to broaden students' knowledge in a variety of content areas and expose them to ideas outside their major fields. However, the very fact that these courses are outside a student's primary area of interest creates challenges. These challenges include engaging students' interest in the subject material, motivating them to learn the material, and helping them to see how the course content is relevant to their education in both the short and long term (McClanahan and

McClanahan, 2002; Udovic *et al.*, 2002; Tessier, 2007). To address these challenges and to enable students to benefit from a core requirement of this type, class experiences need to encourage undergraduates to be creative, critical thinkers, capable of applying their knowledge and skills to broader situations and experiences. Students also need to be engaged in learning experiences that promote their interest and, subsequently, their achievement, in material outside their major fields. When students are interested in the content or teaching approach of a course, they are more likely to be engaged in learning (Hidi *et al.*, 2004) and to have increased conceptual understanding (Nieswandt, 2007). Different approaches have been used to improve students' interest in non-majors science classes, such as inquiry instruction (Udovic *et al.*, 2002), use of writing assignments (Tessier, 2006), field trips (Verderber, 1993), various active-learning activities, (McClanahan and McClanahan, 2002; Armbruster *et al.*, 2009), and same-age peer teaching (Tessier, 2007). The approach described in this paper combines service-learning with cross-age peer teaching to give undergraduate students an opportunity to increase their achievement in a

DOI: 10.1187/cbe.14-01-0007

Address correspondence to: G. K. Muday (muday@wfu.edu).

© 2014 H. E. Chrispeels *et al.* CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Non-commercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

non-majors biology course. This study examines the learning gains of undergraduates who acted as peer teachers to middle- and high-school students.

Service-Learning

While more than 147 definitions of service-learning can be found throughout the literature (Kendall, 1990), in this study we define service-learning as a pedagogical method that integrates community service with required course activities to enhance student attainment of course-specific learning objectives while simultaneously meeting the needs of community partners. Our service-learning program (SLP) most closely aligns with Bringle and Hatcher's (1995) conceptualization of the process, which describes service-learning as an educational experience that is credit bearing, organized around the needs of community members, and furthers student understanding of course content. In service-learning, students attempt to address the needs of community members or organizations that are potentially very different from their own. Ideally, students engaged in service-learning develop understanding of the content associated with a specific course as well as an awareness of the unique situation in which learning occurs. As a deeply contextualized learning experience, the perceived and actualized outcomes of service-learning experiences are thus individualized and potentially transformative.

Service-learning in science course work differs from traditional science "outreach" programs, which have the single goal of enriching K–12 science curricula. Outreach programs provide expert instructors and authentic materials and activities designed to stimulate student interest in science and potential science careers, but the activities usually are not tied to undergraduate course requirements or curricula. Often outreach takes place outside regular class hours, and most outreach facilitators are scientists. Some studies, using self-reports of participants' experiences, have found that when college or graduate students act as outreach experts, they can experience gains in content knowledge and teaching skills (Koehler *et al.*, 1999; Mitchell *et al.*, 2003; Trautmann and Krasny, 2006; Laursen *et al.*, 2007). In our SLP, the peer teaching was done by nonexperts, took place during school and university class time, was a course requirement, and related to the university course curriculum, characteristics which differentiate service-learning from outreach. Yet, like outreach, the materials and activities were developed by scientists with a goal of stimulating K–12 student interest in science, and the undergraduate instructors were well trained in the exercises.

Peer Teaching

Same-age peer teaching (students of the same age and/or grade teaching one another) and cross-age peer teaching (older students teaching younger students) have been shown to have benefits for the teaching peer, as well as the learning peer (Roscoe and Chi, 2007). Evidence of the "tutor learning effect," the learning gains of tutors compared with nontutors, has been found in multiple peer-tutoring situations. In a meta-analysis of studies of same-age and cross-age peer teaching in K–12 schools, tutors had higher academic achievement on exams compared with nontutor controls in 33 out of 38 studies, and tutors had a more positive attitude

than nontutors toward the subject matter in four out of five controlled studies (Cohen *et al.*, 1982). Thus, the act of teaching others can be a motivating factor for learning.

Most studies of learning outcomes of postsecondary peer teachers involve medical students (e.g., Buckley and Zamora, 2007; Graham *et al.*, 2008; Peets *et al.*, 2009, Gregory *et al.*, 2011), but a few studies have examined the tutor learning effect with undergraduate students. Studies have found that undergraduate students who learn material with the expectation of teaching score higher on subsequent content tests than students who are not expected to teach the material (Bargh and Schul, 1980; Benware and Deci, 1984). Annis (1983) found that students who prepared to teach and then taught scored higher on content tests than students who only prepared to teach but did not actually teach, and both groups performed better than control groups who learned material with no expectation of teaching. Similar results were found by Fiorella and Mayer (2013); however, the students who did not teach did not perform as well on a delayed content assessment compared with the students who actually taught. The above studies all involved one-on-one student tutoring or experimental situations in which the undergraduates "taught" to a camera rather than in actual course settings using course curriculum material. One study that did involve students enrolled in a course was done by Tessier (2007). In his study of same-age, small-group reciprocal peer teaching in an introductory biology course, he found that students scored significantly higher on exam material that was peer taught compared with material learned in the traditional lecture format from the professor. In the small-group settings, students acted both as teachers and learners with respect to their peers.

Given these reported benefits of same-age peer tutoring over traditional lecture, quantitative data are needed on the cognitive effects of cross-age peer teaching in actual course settings. Studies reporting cognitive benefits for cross-age undergraduate peer teachers have relied on qualitative data about the tutors' perceptions of their learning (Juel, 1991; Topping, 1996; Gafney and Varma-Nelson, 2007; Galbraith and Winterbottom, 2011), but quantitative data are lacking.

Service-Learning and Cross-Age Peer Teaching

Most studies on service-learning initiatives in undergraduate science courses have involved cross-age peer teaching in partnerships with K–12 classrooms and have relied on student self-reporting to measure benefits. In two studies, undergraduates taught chemistry concepts to elementary (K–5) students (Esson *et al.*, 2005; Kalivas, 2008). Undergraduates in both studies reported favorable reactions to these partnerships, and, in the Esson *et al.* study, they reported increased skill development specific to the tasks of problem solving and experimental design.

In service-learning partnerships with middle school (6–8) classrooms, Hatcher-Skeers and Aragon (2002) engaged students in the planning and teaching of demonstrations for middle school students related to chemistry course content. Undergraduate students reported that the process of explaining demonstrations to middle school students motivated them to learn the science content and improved their problem-solving and laboratory skills. Similarly, Hark (2008) and Robinette and Noblet (2009) found that when undergraduate

students developed instructional materials on biotechnology and entomology topics, respectively, for high school classrooms, undergraduates reported increased understanding of those concepts. In the case of Robinette and Noblet (2009), undergraduate students also reported increased awareness of service-learning benefits for self and others and increased awareness of the logistics required for community collaborations.

According to Abes *et al.* (2002), who surveyed university faculty on the factors that promote and deter the use of service-learning in undergraduate courses, one major deterrent is the perception that students' experiences may or may not result in content knowledge gains. The previous studies examined undergraduates' self-reports of knowledge gains as a result of service-learning. Self-reports are useful for understanding student knowledge gains as a result of service-learning experiences; however, what is lacking is empirical evidence of learning outcomes associated with cross-age peer-teaching and service-learning experiences embedded with science content courses. In addition, multiple measures of learning are needed to gain a complete picture of the impact of service-learning and cross-age peer teaching. Warren (2012) notes that "because student self-reported learning can be affected by a number of other variables such as teacher immediacy, liking the course and/or instructor, and so on, it is important to distinguish between measurement techniques" (p. 57).

Here we add to the literature on service-learning and cross-age peer teaching by providing quantitative data on learning gains. In this study, we explored whether participation in the SLP resulted in an increase in understanding of content by undergraduates who taught high school or middle school students. Using two types of assessments, we measured whether undergraduates scored higher on questions related to the SLP curriculum compared with the rest of the class material learned in the traditional lecture/lab format and whether these learning gains were specific to the curriculum taught by the undergraduates.

METHODS

SLP Design

Participants in this SLP were undergraduate students attending a private, liberal arts institution with an undergraduate enrollment of ~4800, who were enrolled in Biology 101: Biology and the Human Condition, a non-majors course

consisting of three 1-h lecture periods and one 3-h lab session per week. This course is an issues-based introductory biology course; the curriculum includes the basic principles of biology as well as recent advances in biology and their ethical, social, and political aspects. The course has been taught by the same faculty member (C.L.G.) every year since the inception of the SLP in 2010. Most students in this class were first- or second-year students who had not yet declared a major but who were not planning to major in a science at the time of enrollment in the course. Demographic data for the enrolled students are shown in Table 1.

Our SLP utilized the undergraduate students in Biology 101 as peer teachers to seventh-grade (MS) and high school (HS) students in a medium-sized public school district in the southeastern United States. Beginning in 2010, three of the traditional lab sessions for Biology 101 were replaced by service-learning sessions. During the first session, undergraduates were trained to teach hands-on curriculum activities by experiencing the curriculum themselves as learners. Faculty (defined as professors or graduate students associated with the project) guided the undergraduates through the activities in each level of the curriculum, using a question-and-answer method to probe students' prior knowledge and to encourage the students to ask questions and put together ideas to enhance their understanding. The lab room was set up with the same materials that were to be used in the middle and high schools. Undergraduates worked in pairs while carrying out the activities. Faculty trained the undergraduates to teach the curriculum by modeling the teaching method and answered undergraduates' questions on content as well as activity procedures. Undergraduates then repeated the activities with their lab partners to develop fluidity in explaining the concepts and carrying out the activities. Undergraduates were encouraged to spend time outside the lab session to become more familiar with the activity script and content information. Lab sections that were scheduled to teach HS students were led through additional activities that were part of an extension to the curriculum (see details on curriculum in the following section); lab sections that were scheduled to teach MS students did not learn these activities.

In the two subsequent lab sessions, undergraduates traveled to schools and taught the curriculum to MS or HS students. The schools that each lab session visited were assigned based on the correspondence of the lab time with the available class time at the middle or high school. The MS and HS schedules were unknown to the undergraduates at the time of course enrollment. Thus, assignment

Table 1. Student demographic data for 2012 and 2013 cohorts

Year	2012		2013	
	MS undergrads	HS undergrads	MS undergrads	HS undergrads
Number of students	23	41	23	31
Gender	43.5% M 56.5% F	34.1% M 65.9% F	20.8% M 79.2% F	22.6% M 77.4% F
Class	34.8% Freshmen 56.5% Sophomores 8.7% Juniors 0% Seniors	41.5% Freshmen 51.2% Sophomores 7.3% Juniors 0% Seniors	56.5% Freshmen 43.5% Sophomores 0% Juniors 0% Seniors	64.5% Freshmen 29.0% Sophomores 3.2% Juniors 3.2% Seniors

of undergraduate students into the two evaluated groups was random and based on the undergraduates' self-chosen enrollment into the lab section as determined by their own class schedules. Pairs of undergraduates led groups of five to 10 secondary students through the curriculum in the same manner as modeled in the training sessions, and teaching pairs remained the same as training pairs from the initial week whenever possible. Groups visiting high schools taught two sequential classes, while groups visiting middle schools taught either two or three sequential classes, which varied with the location of the school and the ability to coordinate with scheduled laboratory times. Project faculty, course TAs, and some third- and fourth-year biology majors accompanied the undergraduates, helped with setting up materials, introduced the undergraduates, and presented a brief overview of heirloom tomato diversity. Faculty and student assistants also helped distribute and clear away materials at the appropriate time, answered questions as necessary, and sometimes joined in teaching if there was an insufficient number of undergraduate student pairs for the size of the class.

Genetics Curriculum

The genetics curriculum taught to the MS and HS students (Klosterman *et al.*, 2014) was originally developed by two of the authors in 2010 (H.E.C. and G.K.M.) and subsequently modified to its current form over several years with feedback from other authors (M.L.K. and C.L.G.), HS and MS teachers whose classes we visited, and MS and HS student volunteers who participated in trial runs of the curriculum. The original goal in developing the curriculum was to meet and supplement the learning goals designated by state standards for genetics concepts for seventh-grade science curricula and high school biology courses. The curriculum was also consistent with the Biology 101 curriculum course goals of Mendelian and molecular genetics. The curriculum both addressed the content and course goals of Biology 101 in a way that was age-appropriate for MS and HS students and addressed the national and state science standards applicable to those grades. Therefore, some of the content was more basic than was addressed in the Biology 101 course, but all information covered in this curriculum was topically aligned. In addition, all topics were also covered in the lecture portion of the course at the college level, with the exception of plant reproduction, which was not a part of the Biology 101 course curriculum.

The genetics curriculum was designed as a multilevel case study, with each level of content increasing in complexity and dependent upon the knowledge gained in the previous level(s). By working through the levels of activities, secondary students elucidated the genetic basis of a hypothetical novel tomato mutant that produces green fruits at maturity rather than red ones, similar to known tomato mutants with the same phenotype (Barry and Giovannoni, 2006).

The first level of activities in the curriculum introduced the mutant and reviewed plant anatomy. Students then examined and evaluated 2-wk-old wild-type and mutant tomato seedlings to look for other phenotypic differences between the two varieties. This activity took advantage of a known phenotypic characteristic of green-fruited tomato mutants, increased lateral root mass (Negi *et al.*, 2010). In the second

level, students discussed the different effects of environmental and genetic factors on tomato plant appearance and then learned principles of plant reproduction using a card-sort activity. In the third level, students learned how to carry out hypothetical crosses and predict phenotypic ratios using Punnett squares. Using these tools, the students determined that the hypothetical green-fruited variety was caused by a recessive mutation. In the fourth level of activities, students extracted DNA from tomatoes to give them the opportunity to see the genetic material they had learned about in the abstract (modified from Frey, 2008). They then had the opportunity to participate in heirloom tomato tasting, using a large variety of heirloom tomatoes grown at Wake Forest University and/or purchased at local farmers' markets. The tomato tasting re-enforced the concept of the genetic basis of phenotypic variation in a fun, sensory activity. These four levels were common to both MS and HS classes.

For the HS students, an extension was added to the third and fourth levels of the curriculum, in which students were introduced to a known dominant green-fruited mutant, *Green-ripe* (Barry and Giovannoni, 2006) and completed additional crosses and Punnett squares focused on this mutant. Subsequently, students learned about the biotechnology techniques of polymerase chain reaction (PCR) and DNA gel electrophoresis, and then examined images of an electrophoresis gel that contained PCR-generated fragments of DNA from *Green-ripe* and its wild-type parent to determine that a deletion mutation is the cause of the *Green-ripe* phenotype.

Study Design

Our study used a quasi-experimental approach, as there was no control group of undergraduates who did not participate in the SLP. Although this approach limits our conclusions, our ultimate goal was to maximize the impact of the SLP on MS and HS by serving as many MS and HS students as possible. Following university protocol for human subjects' research, undergraduates agreed to participate in this study by signing a consent form on the first day of class. Participation in the research study was voluntary, and students' grades in the lab and lecture were not affected by their participation in the research study; however, all undergraduates enrolled in the course participated in the SLP, because it was a course requirement. Data were collected from the Fall semester during 2 academic years, 2012 and 2013. We obtained a varied number of participant responses for each assessment measure, which was a result of students not completing all assignments, dropping the course, or not completing the informed consent to participate in this research study. The research was approved by Wake Forest University Institutional Review Board (IRB00018129).

Assessments

Pre- and Post-SLP Test. We measured learning outcomes by analyzing scores on pre- and posttests given before and after the service-learning program. The pretest measured baseline undergraduate knowledge of curriculum-associated concepts before any SLP-related material was taught in class or in the SLP lab sessions. The posttest measured comprehension of curriculum-related concepts retained 1 wk after completion of the SLP. Short-answer questions on the tests

were written and scored to test conceptual knowledge at the college level. In 2012, the pre- and posttests consisted of the same five open-response questions (Supplemental Table S1). The test was reviewed by four authors (M.L.K., S.R.L., H.E.C., and G.K.M.) who were experts in content and/or curriculum for content and curricular validity (Messick, 1995; Moskal and Leydens, 2000).

The pretest was given to the undergraduate students in the lecture portion of the class on the first day of class, before labs had started and before any material on genetics had been presented in the lecture portion of the class. The first three questions on the test were about plant reproduction and Mendelian genetics, topics common to both the MS and HS curriculum, and the last two questions were about PCR and DNA gel electrophoresis, material that was included in the HS curriculum, but not the MS curriculum. The posttest was given during the lab session 1 wk following the final school visit and after students had heard lectures on genetics, PCR, and DNA gel electrophoresis in the lecture portion of the course. There were ~4 wk between administration of the pre- and posttests. For consistent evaluation of student knowledge, answers were scored by one author (S.R.L.) using a scoring sheet that was collaboratively developed by members of the research team (S.R.L., M.L.K., and G.K.M.) (Table S1).

In 2013, we were interested in assessing whether participation in the SLP had effects on different cognitive levels. Thus, we modified the pre/posttest to differentiate between different levels of knowledge and comprehension. Both pre- and posttests consisted of the same six questions, three of which pertained to the common curriculum taught to both MS and HS students and three of which were specific to the HS curriculum. Each question had two parts: part A consisted of a question that tested knowledge at a lower level of Bloom's revised taxonomy (Krathwohl, 2002), and part B consisted of a question that tested at a middle level, such as application or analysis. The test was reviewed by three authors (M.L.K., H.E.C., and G.K.M.) who were experts in content and/or curriculum for content and/or curricular validity (Messick, 1995; Moskal and Leydens, 2000). The pre- and posttests were administered at the equivalent time points, as in 2012. Due to the redesign of the pre/post content assessment to include multiple levels of comprehension, a more sophisticated method of scoring was warranted to more accurately measure student knowledge. A rubric for scoring the tests was developed by four of the authors (M.L.K., J.B.M., H.E.C., and G.K.M.). After these authors scored one lab section, the rubric was discussed by the authors and modified. Two authors (J.B.M. and H.E.C.) then rescored the first lab section and scored a second lab section using the modified rubric. After scores were compared, the rubric was discussed and modified again. All lab sections were rescored by the two authors (J.B.M. and H.E.C.) using the modified rubric (Table S3). Scoring using the final version of the rubric resulted in 90% interrater reliability for the pretest and 82% interrater reliability on the posttest. (The higher percentage on the pretest was likely due to the higher frequency of blank responses resulting in an automatic-zero score; on the posttest, blank responses were rare.) After discussion of scoring discrepancies, which were usually within one point value, 100% consensus on scoring was reached. Raw scores were used in subsequent statistical analysis.

Final Exam. For the 2012 cohort, we also measured learning outcomes by analyzing performance on the cumulative final exam in the lecture portion of the course, a summative assessment that measured long-term retention of curriculum-related concepts approximately 12 wk after the service-learning experience. We analyzed scores on six multiple-choice final exam questions about genetics or molecular genetic methods (Table S4) and compared scores on these questions to the total exam score, which consisted of 74 additional multiple-choice questions that covered all other course topics. The exams were graded by the instructor (C.L.G.) and then double-checked by another author (S.R.L.) to reduce grader error. The first four questions were about Mendelian genetics, a topic included in both the MS and HS curriculum taught by the undergraduate students. The last two questions focused on biotechnology techniques, particularly PCR and gel electrophoresis, and were specific to the HS curriculum material. We also examined scores on four final exam questions that covered topics learned both in lecture and in the lab portion of the course, and compared these scores with scores on the 70 remaining questions, which were about topics learned only in the lecture portion of the course. We were unable to use the 2013 course final exam for assessment purposes due to a change in the final exam format. We were working within the structure of an existing course, in which an instructor-designed final exam was used. In 2013, the short-answer format had a range of question types, formats, and point values. Additionally, the exams administered at two different times were not identical. These factors made it difficult to assess student content knowledge in selected areas.

Statistical Analysis

Statistical analysis on all assessment data was performed using R (R Core Team, 2014). A Mann-Whitney *U*-test was used to determine significant differences in scores between groups (undergraduates who taught HS students and undergraduates who taught MS students). An exact Wilcoxon signed-rank test was used to determine significant differences in scores within groups. False-discovery rate control was applied to all statistical tests (Verhoeven *et al.*, 2005). *Z*, *W*, and *p* values for all statistical tests are reported in Table S2. Significance was determined at $p < 0.05$.

RESULTS

Pre- and Post-SLP Content Knowledge Tests

We hypothesized that participation in the SLP would have a positive effect on the undergraduates' understanding of genetics and other concepts that were part of the SLP teaching curriculum. We tested that hypothesis using pre- and posttests consisting of short-answer questions directly related to the SLP curriculum. In 2012, the test consisted of five short-answer questions related to the curriculum material (Table S1): three questions common to both the curriculum taught to MS students and the curriculum taught to the HS students, and two questions specific to the HS curriculum. Figure 1 shows the results of the pre- and posttest assessments from 2012. Although the average score for each question varied, there was no significant difference between the scores of MS undergraduates and HS undergraduates on the

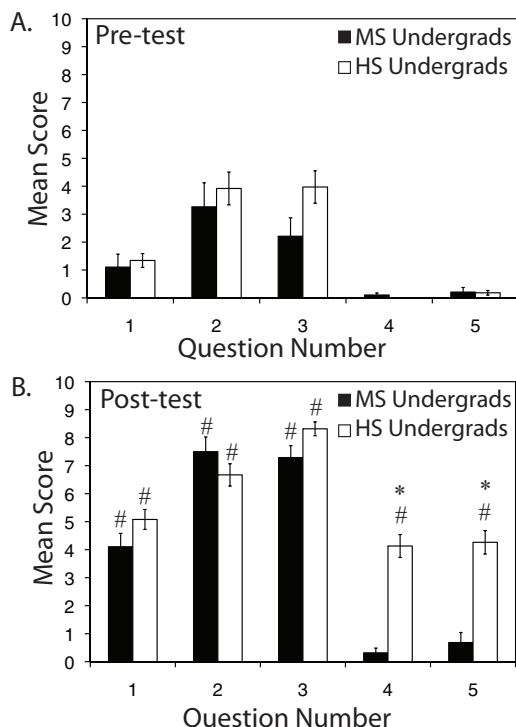


Figure 1. Comparison of mean (\pm SE) scores of MS undergraduates ($n = 19$) vs. HS undergraduates ($n = 38$) from pre- and posttests of SLP content knowledge from 2012. (A) Pretest. (B) Posttest. Asterisks (*) indicate significant difference between groups as determined by a Mann-Whitney U -test followed by false-discovery rate control. Hash marks (#) indicate significant difference between pre- and posttest score within the group, as determined by an exact Wilcoxon signed-rank test followed by false-discovery rate control. W , Z , and p values for all comparisons are reported in Table S2.

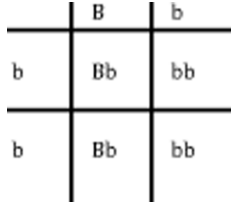
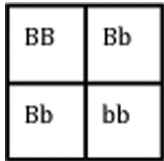
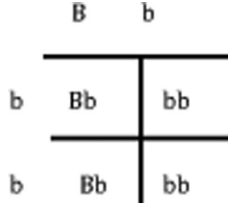
pretest for each question, as determined by a Mann-Whitney U -test (Figure 1A), suggesting that knowledge of these concepts of both groups before participation in the SLP was similar. On the posttest (Figure 1B), HS undergraduates scored significantly higher on all five questions, compared with the pretest, as determined by an exact Wilcoxon signed-rank test ($p < 0.05$). MS undergraduates scored significantly higher on the posttest compared with the pretest on the first three, common curriculum questions, but there were no significant differences between the scores on questions 4 and 5, the HS-specific questions. When scores of MS undergraduates and HS undergraduates were compared, there were no significant differences between the groups on the first three questions, as determined by a Mann-Whitney U -test. However, the HS undergraduates scored significantly higher than MS undergraduates on the last two questions, which related to PCR and gel electrophoresis, subjects that were part of the HS curriculum but not the MS curriculum (Figure 1B). These results suggest that preparation to teach and teaching done by the HS undergraduates resulted in the higher test scores on questions specific to the HS curriculum, despite both groups learning about those subjects in the lecture portion of the course during the weeks the SLP was taking place. There was no significant difference in pre- and posttest scores on those two questions for MS undergraduates, suggesting that exposure to the material during

the lecture part of the course was not sufficient for students to master this complex material.

In 2013, we changed the pre and posttests to contain six questions (Table 2), three of which pertained to the common genetics curriculum, and three of which were specific to the HS curriculum. In addition, each question had two parts: part A consisted of a question that tested the lower levels of Bloom's revised taxonomy (Krathwohl, 2002), such as remembering and understanding; part B consisted of a question that tested the middle levels, such as applying and analyzing. Using this format, we analyzed whether participation in the SLP affected cognitive gains at both levels. Figure 2A shows the results of the pre- and posttest scores for all students in 2013. For all questions except 1B, there was a significant increase in the posttest score compared with the pretest score, as determined by an exact Wilcoxon signed-rank test ($p < 0.05$). For question 1B, there was no significant difference between pre- and posttest scores. This result suggests that the undergraduates' understanding of the material increased at both lower and higher levels of comprehension for all but one of the questions. We hypothesized that teaching HS students would have a positive effect on higher levels of cognition, and thus HS undergraduates would always score higher on part B of the questions than MS undergraduates. We tested this hypothesis by comparing the posttest scores of MS undergraduates with those of HS undergraduates for part B of each question (Figure 2B). We predicted that if HS undergraduates were positively affected for higher levels of cognition, then they would always score significantly higher than MS undergraduates, even on the common curriculum questions. However, HS undergraduates only scored significantly higher than MS undergraduates on the HS-specific questions, as determined by a Mann-Whitney U -test ($p < 0.05$), and there were no significant differences between the scores of the common questions, suggesting there was no effect of participation in the HS group on higher levels of cognition. Both MS and HS undergraduates exhibited gains in lower and higher levels of cognition as a result of the SLP, with no significant difference between the two groups; thus, for subsequent comparison of MS undergraduates' and HS undergraduates' scores, the values of the A and B parts of each question were combined. Figure 3 shows the results of comparison between HS undergraduates and MS undergraduates. We found results similar to those from 2012: although the mean scores varied on each question, there were no significant differences in scores between MS and HS undergraduates on the pretest. Scores on posttest questions were significantly higher than pretest scores, for all questions for MS undergraduates, and for all but question 1 for HS undergraduates. There were no significant differences between MS and HS scores on the first three, common questions. However, the HS undergraduates scored significantly higher than the MS undergraduates on the HS-specific questions, despite all students learning about the content in the lecture portion of the course. These repeated data strengthen our findings that participation in the SLP increases student comprehension of the pertinent concepts and that the learning gains are specific to the content taught.

To provide examples of the type of knowledge gained by the undergraduates, Table 2 shows questions and sample answers to the 2013 pre- and posttests from one MS undergraduate and one HS undergraduate. These students were

Table 2. Examples of answers from one MS undergraduate and one HS undergraduate to pre- and posttest questions from 2013^a

Question	MS undergraduate		HS undergraduate	
	Pretest answer	Posttest answer	Pretest answer	Posttest answer
1A. Below are listed the events in fertilization and fruit production. List the numbers of the events in the correct order.	3-6-2-5-7-4-8-1 (4)	3-6-2-5-7-8-4-1 (3)	3-6-2-5-7-8-1-4 (3)	3-6-2-7-5-4-8-1 (3)
1B. If a plant cannot produce pollen, how would that affect its ability to produce fruit? Describe and explain.	"Plants are not asexual and a plant that cannot produce pollen cannot experience the genetic variation of a sexual organism. I suppose it could still produce fruit, but I'm not entirely sure." (0)	"It would still be able to produce fruit, it would just have to receive pollen from another plant." (1)	"I suppose it wouldn't be able to create fruit because fruit is the fertilized 'egg' ready to be planted after pollination [sic], without pollen no fruit." (1)	"It would not be able to produce fruit because it lacks the pollen that fertilizes the plant 'egg' and produces fruit." (2)
2A. Define the terms "genotype" and "phenotype."	"I know I have studied this but it has been a long time, I can't recall what they mean." (0)	"A genotype describes the set of alleles found in an organism's genes. A phenotype describes the trait that shows." (3)	"Genotype is a genetic trait while phenotypes are physical traits." (1)	"Genotype—an organisms [sic] genetic information that codes for traits. Phenotype—the physical traits an organism shows." (3)
2B. In two or three sentences, explain why two different genotypes (representing the same gene) might give the same phenotype.	No answer (0)	"It is possible for one phenotype to show through several genotypes. If the organism is heterozygous, they have a recessive allele that does not affect the genotype, or they are homozygous and have two of the dominant allele." (1)	"Because genotypes can sometime be one passive and one dominant so even if a certain genotype is present it may not show through a phenotype therefore simply producing the phenotype of the dominant genotype only." (0)	"Sometimes genotypes can have heterozygous and homozygous traits. These traits appear different in the genotype because of the presence or lack of one recessive allele. These look the same because there is a dominant allele present which always shows in the phenotype." (2)
3A. Define the terms "homozygous" and "heterozygous."	"Homozygous—has all one type of allele. Heterozygous—has different alleles." (2)	"Homozygous = two of the same allele in the genotype. Heterozygous = two different alleles in the genotype." (4)	"Homozygous—one gene. Heterozygous—two genes." (0)	"Homozygous—a trait with either both alleles being dominant or both recessive. Heterozygous—trait that has both a recessive and a dominant trait." (3)
3B. Brown eye color is dominant over blue eye color. Imagine that a heterozygous brown-eyed woman and a blue-eyed man have children. Set up a Punnett square to demonstrate the fraction of brown-eyed children from this couple. Use the letters B and/or b to designate alleles.	No answer (0)	 <p>Punnett square was correctly drawn, but the fraction of brown-eyed children was not included in the answer. (2)</p>	 <p>25% chance of blue eyes (0)</p>	 <p>Punnett square was correctly drawn, but the fraction of brown-eyed children was not included in the answer. (2)</p>
4A. List three essential components that are required to carry out a PCR experiment.	No answer (0)	"Heat resistant DNA polymerase, heat, small pieces of DNA." (2)	No answer (0)	"Primer, polymerase, DNA sample." (2)
4B. The polymerase chain reaction is dependent on multiple cycles of polymerization. Explain why multiple cycles are needed for the technique to work as it should.	No answer (0)	"The goal is to exponentially increase the rate at which the DNA is synthesized. If there aren't multiple cycles, this result will not be achieved." (0)	No answer (0)	"Because with each cycle DNA grows exponentially, the point of PCR is to have a lot of DNA to work with therefore we need a few cycles to get adequate DNA." (4)

(Continued)

Table 2. (Continued)

Question	MS undergraduate		HS undergraduate	
	Pretest answer	Posttest answer	Pretest answer	Posttest answer
5A. What is the objective of the technique of DNA gel electrophoresis? In other words, what does the technique do?	No answer (0)	"It identifies various small pieces of DNA by how fast they move through the gel." (1)	No answer (0)	"It spreads out pieces of DNA allowing one to compare sizes of DNA pieces by comparing how far down the pieces moved in the gel." (3)
5B. You are running a DNA electrophoresis gel and you discover that you accidentally reversed the positive and negative electrodes. What will happen to the DNA that you loaded at the top of your gel? Explain.	No answer (0)	"They won't move at all." (2)	No answer (0)	"It will most likely stay at the top, DNA is negatively charged therefore having the positive charge at the top would make the DNA stay at the top." (4)
6A. Some diseases are caused by mutations. What in the cell is mutated? Briefly describe three types of mutations that might occur.	"The chromosomes are damaged." (0)	"The DNA of the cell is mutated. A deletion mutation completely removes one base. An insertion mutation copies an extra base." (1)	"The DNA, sometimes an extra chromosome can be created, the 'genetic instructions' can simply be copied wrong." (1)	"The DNA is mutated, deletion—loss of a nucleotide, substitution—nucleotide being replaced by another, and addition—nucleotide is added." (4)
6B. On the gel electrophoresis image below are samples from a healthy person and someone with a mutation that causes cystic fibrosis (CF). What kind of mutation would explain this difference? Explain your answer.	No answer (0)	No answer (0)	No answer (0)	"A deletion mutation, because the DNA sample of the CF patient sits lower in the gel compared to a healthy person we can assume the DNA to be smaller and therefore contain less DNA, the product of a deletion mutation." (4)

^aEach question is worth a total of four points. The score earned for each question is shown in parentheses after the answer.

chosen as representatives because their pretests had similar scores, and their posttests had similar scores on the three common questions. However, the MS undergraduate scored much lower on the HS-specific questions than the HS undergraduate, thus representing the content assessment data as a whole. The posttest answers reveal the differences in their knowledge and understanding of the concepts tested. For example, question 4B asks why PCR requires multiple cycles in its procedure. The HS undergraduate correctly answered that the amount of DNA is increased exponentially with each cycle, whereas the MS undergraduate erroneously described that the goal of PCR is to exponentially increase the *rate* of DNA synthesis. This example, along with the other answers, suggests that the MS undergraduate partially understood the concepts but did not reach the depth of comprehension demonstrated by the HS undergraduate.

Final Exam

We hypothesized that, by teaching the genetics curriculum, undergraduates would have better retention and comprehension of the material on a delayed summative assessment, as well as on the immediate posttest. If this hypothesis is correct, we would see higher learning gains on final exam questions pertinent to the curriculum, compared with

other exam questions. Figure 4 shows the results of the final exam analysis from 2012. We analyzed scores on six multiple-choice final exam questions about concepts learned in the SLP curriculum (Table S4) and compared scores on these questions between groups. Figure 4A shows number of correct answers for each group for each of the six multiple-choice questions, represented as a percentage of number of students in each group. For the four common curriculum questions (questions 1–4), there were no significant differences between MS and HS undergraduates' scores, as determined by a Mann-Whitney *U*-test, suggesting the material was equally well understood and retained by both groups. On question 5, one of the two HS-specific questions, the HS undergraduates scored significantly higher than the MS undergraduates, suggesting the HS undergraduates had better comprehension and retention of that material than the MS undergraduates. On the other HS-specific question, (question 6), the HS students had a higher average score, but the difference was not significant between MS and HS undergraduates' scores after the false-discovery rate control was applied.

To explore the retention and understanding of the genetics principles that the undergraduates taught relative to the rest of the course content that they did not teach, we

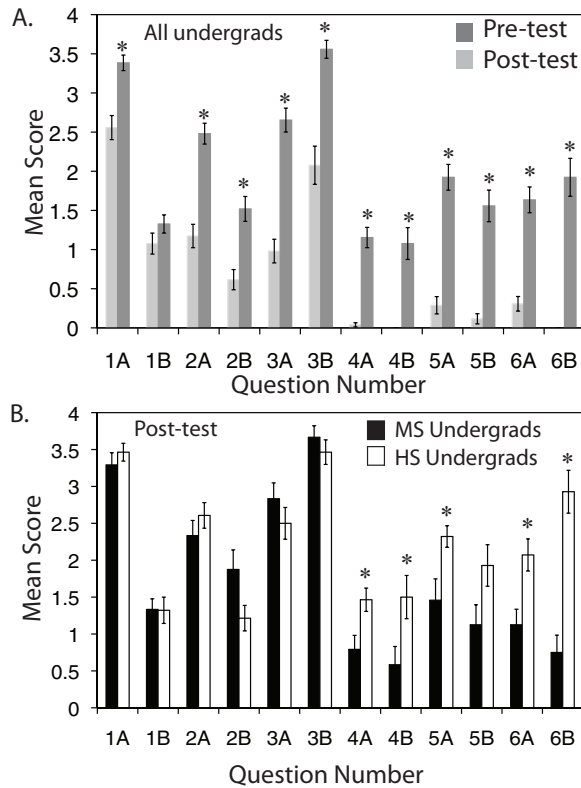


Figure 2. (A) Comparison of mean (\pm SE) pre- and posttest scores for all undergraduates ($n = 52$) from 2013. Asterisks (*) indicate significant difference between pre- and posttest scores for each question as determined by an exact Wilcoxon signed-rank test followed by false-discovery rate control. (B) Comparison of mean scores (\pm SE) from posttests of MS undergraduates ($n = 24$) and HS undergraduates ($n = 28$). Asterisks (*) indicate significant difference between groups as determined by a Mann-Whitney U -test followed by false-discovery rate control. W , Z , and p values for all comparisons are reported in Table S2.

compared responses to these questions with responses to the other final exam questions. Figure 4B shows the scores for MS and HS undergraduates on the four common curriculum questions, the two HS-specific questions, and the rest of the exam, represented as percentages. Both MS and HS undergraduates scored significantly higher on the four genetics questions compared with the rest of the exam questions, as determined by an exact Wilcoxon signed-rank test ($p < 0.05$), suggesting that students had better comprehension of this material than the rest of the course material. There was no significant difference between scores of HS undergraduates and MS undergraduates on the rest of the exam questions, as determined by the Mann-Whitney U -test, indicating that all undergraduates learned the course material equally well. MS undergraduates scored significantly lower on the two biotechnology questions than the other exam questions, while HS undergraduates scored equally well on the two biotechnology questions as on the other exam questions. In addition, the HS undergraduates who taught these conceptually complex methods scored significantly higher than the MS undergraduates. This is strong evidence that teaching this molecular genetic material had a substantial impact on student understanding.

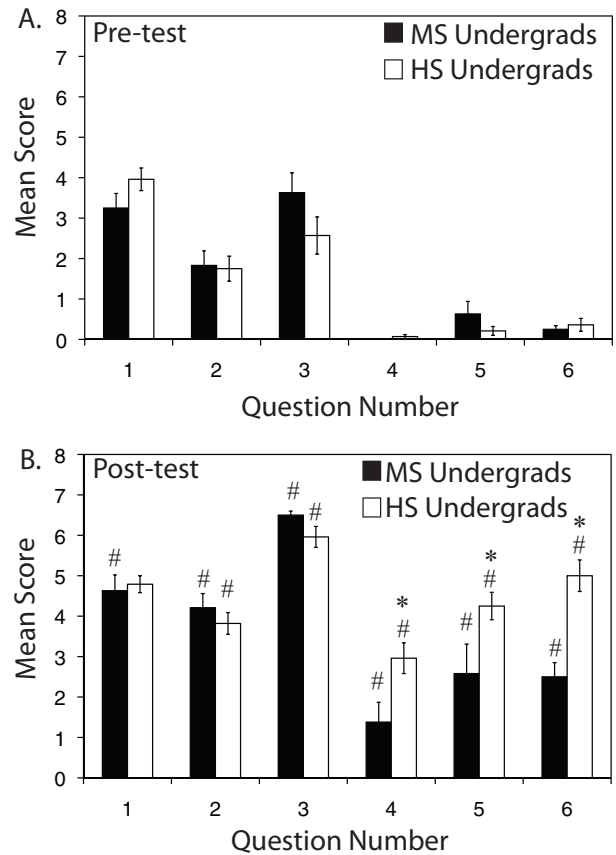


Figure 3. Comparison of mean (\pm SE) scores of MS undergraduates ($n = 24$) vs. HS undergraduates ($n = 28$) from pre- and posttests of SLP content knowledge from 2013. (A) Pretest. (B) Posttest. Asterisks (*) indicate significant difference between MS undergraduates and HS undergraduates as determined by a Mann-Whitney U -test followed by false-discovery rate control. Hash marks (#) indicate significant difference between pre- and posttest scores, within the group, as determined by an exact Wilcoxon signed-rank test followed by false-discovery rate control. W , Z , and p values for all comparisons are reported in Table S2.

To investigate the possibility that the higher scores on the SLP-related final exam questions were due to repeated exposure to the material rather than the teaching per se, we analyzed scores on four final exam questions that covered topics learned both in lecture and in certain labs of the lab portion of the course, including hypothesis generation, cell structure, and ecology, and compared these scores with scores on the 70 remaining questions, which were about topics learned only in the lecture portion of the course. (The questions, their answers, and the associated lab topics are reported in Table S5.) The lab portion of the course required undergraduates to read background information and carry out hands-on activities and experiments related to the lab topic, which represented a second exposure to these subjects in addition to what was taught in lecture. As shown in Figure 4C, we found no significant differences, as determined by an exact Wilcoxon signed-rank test, between scores of lab and lecture questions compared with scores on lecture-only questions for both MS undergraduates and HS undergraduates. This result suggests that the higher scores

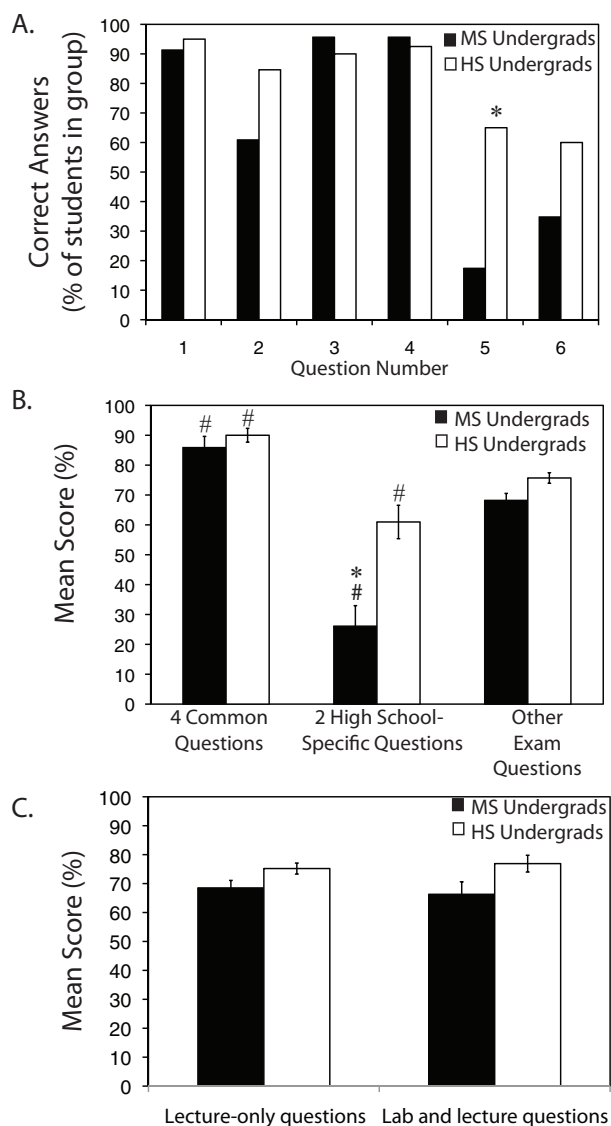


Figure 4. Comparison of final exam scores from 2012. (A) Percentage of correct answers to individual questions pertaining to material taught by undergraduate students. Questions 1–4, genetics questions common to both MS and HS curriculum; questions 5–6, questions specific to HS curriculum. Asterisks (*) indicate significant difference between MS undergraduates and HS undergraduates as determined by a Mann-Whitney *U*-test followed by false-discovery rate control. (B) Mean scores (\pm SE) represented as percentages of the four common curriculum questions, the two HS-specific questions, and the other exam questions. Asterisks (*) indicate significant difference between MS and HS for that question set, as determined by a Mann-Whitney *U*-test followed by false-discovery rate control. Hash marks (#) indicate significant difference from other exam questions, within the group, as determined by an exact Wilcoxon signed-rank test followed by false-discovery rate control. (C) Mean scores (\pm SE) represented as percentages of the 70 lecture-only questions and the four lab and lecture questions. MS, $n = 23$; HS, $n = 37$. *W*, *Z*, and *p* values for all comparisons are reported in Table S2.

on the SLP-related common curriculum questions were a result of SLP-related activities and not just repeated exposure to the material.

DISCUSSION

In this quasi-experimental study, we provide evidence to support the hypothesis that participation in this SLP increases understanding and retention of content information for the material taught by undergraduate students, as compared with material that is learned only in lecture. Our results add to the body of literature on the “tutor learning effect” by providing quantitative data, consistent over 2 yr, to suggest that learning gains occur when undergraduates participate in cross-age peer teaching with secondary students. Previous studies have examined learning gains obtained from same-age peer teaching (Annis, 1983; Tessier, 2007) or from students in artificial teaching situations (Bargh and Schul, 1980; Fiorella and Mayer, 2013); other studies have collected qualitative or anecdotal information on peer-teaching benefits (Juel, 1991; Gafney and Varma-Nelson, 2007; Galbraith and Winterbottom, 2011).

Our results suggest that the learning gains resulted from learning the SLP curriculum and teaching it to the MS and HS students and were specific to the curriculum that the undergraduates taught in the SLP. This conclusion is supported by the pre/posttest data, which show that undergraduates performed higher on test questions related to material they taught in the SLP, despite identical lecture content. These results were consistent over two consecutive years, despite changing the pre/posttest format to include multiple levels of questions. We did not determine whether the preparation to teach, the actual teaching, or the hands-on, active method of learning the material contributed to the learning gains, but all three were part of the SLP. Given that the MS undergraduates heard a lecture on the PCR and gel electrophoresis techniques and still performed significantly lower on those questions than HS undergraduates on the posttest, we can argue that exposure to the material in the lecture and personal studying was not enough to allow mastery of those concepts at the same level as the HS undergraduates. By this logic, we suggest that participation in the SLP played the predominant role in mastery of the concepts.

The results from the 2013 pre/posttests suggest that mastery of the material occurred on multiple levels of comprehension, as demonstrated by the similar scores on the subsections of these questions that focused on different levels of Bloom’s taxonomy. We found no general effect of the group of students taught (MS vs. HS) on the level of comprehension mastered by the undergraduates. This result suggests that peer teaching of students in both higher and lower grades had an equal effect on the ability of undergraduates to understand concepts at higher cognitive levels. On the other hand, evidence from student answers (Table 2) shows that, for the concepts not taught by the MS undergraduates, their comprehension gained from lecture and personal studying was not as complete as that of the HS undergraduates.

The results of the final exam for the 2012 cohort support the hypothesis that participation in the SLP had a positive effect on long-term retention of topics related to the SLP curriculum compared with other exam questions. These results are consistent with other studies showing improvement in comprehension and knowledge on delayed summative assessments through same-age peer teaching at the undergraduate level (Annis, 1983; Tessier, 2007). All undergraduates

scored higher on the common curriculum questions than on the rest of the exam questions, suggesting that participation in the SLP helped them master this material better than the topics they only heard in lecture. In addition, both MS and HS undergraduates performed at the same level on questions related to topics they learned only in lecture and topics they learned in lab and lecture, suggesting it was not simply repeated exposure to the material that was responsible for the learning gains. The HS undergraduates scored equally well on the questions related to PCR and gel electrophoresis as they did on the other exam questions, and they scored significantly higher on those questions than the MS undergraduates, suggesting that teaching these advanced topics had an effect on their understanding. That the HS undergraduates did not score higher on these questions than on the rest of the exam questions is perhaps due to the complex nature of and limited exposure to the material being tested. This is consistent with the extremely low scores on pretest questions on this content, which indicated that students had no prior exposure to this material.

There is debate in the literature over whether peer teaching is connected to short-term or long-term learning gains (Gregory *et al.*, 2011; Fiorella and Mayer, 2013). However, we suggest that both types of assessments are necessary to obtain a complete picture of learning outcomes as a result of SLP participation. Assessment designs like those advocated by Klosterman and Sadler (2010), in which assessments occur at multiple “distances” from the curriculum, have the potential to yield more powerful claims of program outcomes.

One limitation of our study was the lack of detailed assessment of a control group that did not participate in the SLP. As one of the goals of this funded project was to impact as many MS and HS students with this unique educational experience, we chose not to have a control group of students who did not teach this curriculum. However, we plan to address the issue of a control group in the future by giving assessments to students enrolled in a Biology 101 course that does not participate in the SLP, yet has the same instructor and lecture content and format.

Two different, and possibly synergistic, mechanisms could explain why undergraduates would realize learning gains from participation in the SLP. By interacting with HS and MS students in an experiential learning setting, undergraduate students had the opportunity to apply skills and knowledge in a real-world situation, which may allow for deeper understanding of the course content. By being involved in the school classroom, the undergraduates may have been more motivated to learn the material, leading to increased cognitive development (Eyler and Giles, 1999). Indeed, one undergraduate commented, “I realized that teaching is hard and I would never want to do it, but it was a good break from the ‘[College] Bubble,’” referring to the self-contained atmosphere of a university, which can often feel removed from the rest of the world. Recognizing that the HS classroom was more of a real-world situation might have motivated this student. In addition, the peer-teaching mechanism, involving learning and preparing the content, presenting the content, and subsequent interactions between teacher and learner, leads to increased cognitive development (Bargh and Schul, 1980). Our SLP had the power of both mechanisms to actuate the undergraduate students’ learning gains.

Despite the effort involved in organizing and adding a service-learning component to an undergraduate science course, our study suggests that partnering with local schools to provide teaching opportunities for undergraduates will have cognitive benefits and allow non-science major students to master subjects beyond their primary interests.

ACKNOWLEDGMENTS

We thank Patricia Friedrichsen, Janet Chrispeels, and two anonymous reviewers for helpful comments on the manuscript. We thank Michael Anderson and Daniel Griffith for help with statistical analysis. This study was funded by a grant from the Education Foundation of the American Society of Plant Biologists.

REFERENCES

- Abes ES, Jackson G, Jones SR (2002). Factors that motivate and deter faculty use of service learning. *Mich J Comm Service Learn* 9, 5–17.
- Annis LF (1983). The processes and effects of peer tutoring. *Human Learning* 2, 39–47.
- Armbruster P, Patel M, Johnson E, Weiss M (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE Life Sci Educ* 8, 203–213.
- Bargh JA, Schul Y (1980). On the cognitive benefits of teaching. *J Educ Psych* 2, 593–604.
- Barry CS, Giovannoni JJ (2006). Ripening in the tomato *Green-ripe* mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. *Proc Natl Acad Sci USA* 103, 7923–7928.
- Benware CA, Deci EL (1984). Quality of learning with an active versus passive motivational set. *Am Educ Res J* 21, 755–765.
- Bingle R, Hatcher J (1995). A service-learning curriculum for faculty. *Mich J Comm Service Learn* 2, 112–122.
- Buckley S, Zamora J (2007). Effects of participation in a cross year peer tutoring programme in clinical examination skills on volunteer tutors’ skills and attitudes towards teachers and teaching. *BMC Med Educ* 7, 20.
- Cohen P, Kulik J, Kulik C (1982). Educational outcomes of tutoring: a meta-analysis of findings. *Am Educ Res J* 19, 237–248.
- Esson JM, Stevens-Truss R, Thomas A (2005). Service-learning in introductory chemistry: supplementing chemistry curriculum in elementary schools. *J Chem Educ* 82, 1168–1173.
- Eyler JS, Giles DE Jr (1999). *Where’s the Learning in Service-Learning?*, San Francisco: Jossey-Bass.
- Fiorella L, Mayer RE (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemp Educ Psychol* 38, 281–288.
- Frey P (2008). Easy Isolation/Extraction Protocol for Isolating Tomato DNA. http://ucbiotech.org/edu/edu_aids/TomatoDNA.html (accessed 17 June 2011).
- Gafney L, Varma-Nelson P (2007). Evaluating peer-led team learning: a study of long-term effects on former workshop peer leaders. *J Chem Educ* 84, 535–539.
- Galbraith J, Winterbottom M (2011). Peer-tutoring: what’s in it for the tutor? *Educ Stud* 37, 321–332.
- Graham K, Burke JM, Field M (2008). Undergraduate rheumatology: can peer assisted learning by medical students deliver equivalent training to that provided by specialist staff? *Rheumatology* 47, 652–655.
- Gregory A, Walker I, McLaughlin K, Peets AD (2011). Both preparing to teach and teaching positively impact learning outcomes for peer teachers. *Med Teach* 33, e417–e422.

- Hark AT (2008). Crossing over: an undergraduate service learning project that connects to biotechnology education in secondary schools. *Biochem Mol Bio Educ* 36, 159–165.
- Hatcher-Skeers M, Aragon E (2002). Combining active learning with service learning: a student driven demonstration project. *J Chem Educ* 79, 462–464.
- Hidi S, Renninger KA, Krapp A (2004). Interest, a motivational variable that combined affective and cognitive functioning. In: *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development*, ed. D Yun Dai and RJ Sternberg, Mahwah, NJ: Erlbaum, 89–115.
- Juel C (1991). Cross-age tutoring between student athletes and at-risk children. *Read Teach* 45, 178–186.
- Kalivas J (2008). A service-learning project based on a research supportive curriculum format in the general chemistry laboratory. *J Chem Educ* 85, 1410–1415.
- Kendall JC (1990). *Combining Service and Learning: A Resource Book for Community and Public Service*, vol. 1, Raleigh, NC: National Society for Experiential Learning.
- Klosterman ML, Chrispeels HE, Reagan B, Lundy S, Browne C, Munday GK (2014). That's a tomato? Using a familiar food to explore genetic variation. *Science Activities* 51, 1–16.
- Klosterman ML, Sadler TS (2010). Multi-level assessment of scientific content knowledge gains associated with issues-based instruction. *Intl J Sci Educ* 32, 1017–1043.
- Koehler BG, Park LY, Kaplan LJ (1999). Science for kids outreach programs: college students teaching science to elementary school students and their parents. *J Chem Educ* 76, 1505–1509.
- Krathwohl D (2002). A revision of Bloom's taxonomy: an overview. *Theory Pract* 41, 212–218.
- Laursen S, Liston C, Thiry H, Graf J (2007). What good is a scientist in the classroom? Participant outcomes and program design features for a short-duration science outreach intervention in K–12 classrooms. *CBE Life Sci Educ* 6, 49–64.
- McClanahan EB, McClanahan LL (2002). Active learning in a non-majors class: lessons learned. *Coll Teach* 50, 92–96.
- Messick S (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 50, 741–749.
- Mitchell J, Levine R, Gonzalez R, Bitter C, Webb N, White P (2003). *Evaluation of the National Science Foundation Graduate Teaching Fellows in K–12 Education (GK–12) Program*, Chicago: American Educational Research Association.
- Moskal BM, Leydens JA (2000). Scoring rubric development: validity and reliability. *Pract Assess Res Eval* 7–10.
- Negi S, Sukumar P, Liu X, Cohen JD, Munday GK (2010). Genetic dissection of the role of ethylene in regulating auxin-dependent lateral and adventitious root formation in tomato. *Plant J* 61, 3–15.
- Nieswandt M (2007). Student affect and conceptual understanding in learning chemistry. *J Res Sci Teach* 44, 908–937.
- Peets AD, Coderre S, Wright B, Jenkins D, Burak K, Leskosky S, McLaughlin K (2009). Involvement in teaching improves learning in medical students: a randomized cross-over study. *BMC Med Educ* 9, 55.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing* Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org> (accessed 25 April 2014).
- Robinette MS, Noblet R (2009). Service-learning in entomology: teaching, research, and outreach domestically and abroad. *J High Educ Outreach Engagement* 13, 135–153.
- Roscoe RD, Chi MTH (2007). Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Rev Educ Res* 77, 534–574.
- Tessier J (2006). Writing assignments in a nonmajor introductory ecology class. *J Coll Sci Teach* 35, 25–29.
- Tessier J (2007). Small-groups peer teaching in an introductory biology classroom. *J Coll Sci Teach* 36, 64–69.
- Topping KJ (1996). The effectiveness of peer tutoring in further and higher education: a typology and review. *High Educ* 32, 321–345.
- Trautmann NM, Krasny ME (2006). Integrating teaching and research: a new model for graduate education? *BioScience* 56, 159–165.
- Udovic D, Morris D, Dickman A, Postlethwait J, Wetherwax P (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology course. *BioScience* 52, 272–281.
- Verderber GW (1993). Lasting impact: Johnson State College students learn tropical ecology through experience. *Am Biol Teach* 55, 412–416.
- Verhoeven KJF, Simonsen KL, McIntyre LM (2005). Implementing false discovery rate control: increasing your power. *Oikos* 108, 643–647.
- Warren J (2012). Does service-learning increase student learning? A meta-analysis. *Mich J Comm Service Learn* 18, 56–61.

HIGHLIGHTS:

This work describes the learning gains of undergraduates who participated in a service-learning program involving teaching genetics to middle and high school students. Results show increases in content knowledge of topics taught by undergraduates, compared with other course content, assessed by both an immediate posttest and a delayed course final exam.