

The Benefits of Peer Review and a Multisemester Capstone Writing Series on Inquiry and Analysis Skills in an Undergraduate Thesis

K. F. Weaver,^{†*} V. Morales,[†] M. Nelson,[†] P. F. Weaver,[†] A. Toledo,[†] and K. Godde[‡]

[†]Department of Biology, [‡]Academic Success Center, and [§]Department of Sociology and Anthropology, University of La Verne, La Verne, CA 91750

ABSTRACT

This study examines the relationship between the introduction of a four-course writing-intensive capstone series and improvement in inquiry and analysis skills of biology senior undergraduates. To measure the impact of the multicourse write-to-learn and peer-review pedagogy on student performance, we used a modified Valid Assessment of Learning in Undergraduate Education rubric for Inquiry and Analysis and Written Communication to score senior research theses from 2006 to 2008 (pretreatment) and 2009 to 2013 (intervention). A Fisher-Freeman-Halton test and a two-sample Student's *t* test were used to evaluate individual rubric dimensions and composite rubric scores, respectively, and a randomized complete block design analysis of variance was carried out on composite scores to examine the impact of the intervention across ethnicity, legacy (e.g., first-generation status), and research laboratory. The results show an increase in student performance in rubric scoring categories most closely associated with science literacy and critical-thinking skills, in addition to gains in students' writing abilities.

INTRODUCTION

Undergraduate research is a high-impact practice that provides students the opportunity to meaningfully contribute to knowledge building in their fields (Hu *et al.*, 2008). The literature on the effects of undergraduate research indicates an overall improvement in students' communicative, critical-thinking, and problem-solving abilities and increased student satisfaction and frequency of faculty-student interactions, which also correlate with student success (Brownell and Swaner, 2010).

As part of the movement promoting undergraduate research in the sciences, science education experts call for the meaningful integration of writing into science courses to help students develop the science reasoning skills necessary for successful research (Yore *et al.*, 2003; Prain, 2006; Tytler and Prain, 2010). While some conceive of writing in the sciences primarily as a means to communicate findings (e.g., Morgan *et al.*, 2011; Brownell *et al.*, 2013), courses informed by a write-to-learn pedagogy recognize the higher-order science reasoning abilities that students develop through writing. The skills required for effective writing—planning, organization, creating coherence, reviewing, and revision—overlap in large measure with those fundamental to successful learning (Bangert-Drowns *et al.*, 2004; Mynlieff *et al.*, 2014), and write-to-learn curricula capitalize on the unique value of writing as a learning tool. Through writing, students build connections between prior knowledge and new knowledge and articulate connections between areas of a discipline that often otherwise remain isolated for them as distinct sets of course materials (Lankford and vom Saal, 2012). Science writing assignments also enable deep learning, because they push students beyond recall of technical vocabulary into more demanding cognitive tasks, including synthesis, evaluation, and analysis, and likewise require them to articulate their own understandings

Kathryn E. Perez *Monitoring Editor*

Submitted January 16, 2016; Revised June 18, 2016; Accepted June 18, 2016

CBE Life Sci Educ December 1, 2016 15:ar51

DOI:10.1187/cbe.16-01-0072

*Address correspondence to: K. F. Weaver (kweaver@laverne.edu).

© 2016 K. F. Weaver *et al.* CBE—Life Sciences Education © 2016 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–Noncommercial–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.

of scientific concepts (Prain, 2006; Kalman, 2011). One effect of integrating writing into course work is an improvement of students' performance on course exams, especially those requiring higher-level conceptual understanding of the material (Günel *et al.*, 2007), a difference explained in part by findings that show that students who receive well-designed science writing instruction employ more metacognitive strategies in their problem solving than do students who have not received this instruction (van Opstal and Daubenmire, 2015).

Well-designed science writing assignments involve students in the active construction of scientific knowledge, including the analysis of findings and the ability to persuasively present them to varied audiences (Hand *et al.*, 1999; Yore *et al.*, 2003; Prain, 2006; Schen, 2013). For biology students, to take one example, substantive inquiry-driven writing assignments have been shown to aid the development of the capacity to craft strong research foci and appropriate inquiry methodologies (Stanford and Duwel, 2013). Other studies show that students in courses with integrated writing assignments demonstrate a better understanding of their projects and their significance and an improved ability to assess how their hypotheses are supported (or not) by findings (Quitadamo and Kurtz, 2007). These gains can be explained in part by the fact that strong science writing curricula involve students more authentically in the "practice of scientists," including high-level critical thinking, than do "traditional laboratory experiences" (Grimberg and Hand, 2009, p. 507).

In addition to the effects of writing-intensive courses on critical thinking, studies also show the impact on students' analytical and evaluative capabilities. Quitadamo and Kurtz (2007), for instance, compared critical-thinking performance of students who experienced a laboratory writing treatment with those who had experienced a traditional quiz-based laboratory in a general biology course. The results showed that the writing group demonstrated more advanced critical thinking than the control group; more specifically, analysis and inference skills improved significantly, along with the ability to evaluate findings. This impact on students' analytical abilities has been seen to extend to areas that would be traditionally considered "quantitative." Research on integrating writing in statistics courses suggests that writing bolsters students' capacity for statistical thinking, most notably again in higher-order conceptual skills: the ability to compare, analyze, and synthesize information (Delcham and Sezer, 2010).

One of the main impediments of developing and implementing writing-centric science courses is the considerable extra time and effort expended by instructors on designing assignments, teaching writing-related content, and most crucially, offering feedback (Theoret and Luna, 2009; Delcham and Sezer, 2010; Stanford and Duwel, 2013; Dowd *et al.*, 2015b). Integrating well-designed peer-review activities, however, can help to mitigate these challenges (Timmerman and Strickland, 2009; Sampson *et al.*, 2011; Walker and Sampson, 2013; Mynlieff *et al.*, 2014; Dowd *et al.*, 2015a,b). Done right, peer reviewing itself becomes a learning tool, one that is often more impactful for the students giving than receiving the feedback (Guilford, 2001; Akcay *et al.*, 2010). For example, Timmerman and Strickland (2009) compared lab reports from introductory biology courses that incorporated peer review against the writing of upperclassmen enrolled in an upper-level course that had

no peer review; the writing of the students in the intro-level course evinced stronger data selection and presentation and use of primary literature. To achieve these benefits, however, students must be taught the characteristics of effective science writing and how to read effectively for offering useful feedback (Morgan *et al.*, 2011). Lu and Law (2012) and Bird and Yucel (2013) stress the importance of instructing students to see how effective peer review goes beyond the ability to recognize substandard work and instead helps writers advance higher-order aspects of their writing such as logic, coherence, and evidence selection (Glaser, 2014). By thinking beyond "error" as peer reviewers, students are able to gain insight on their own science reasoning and analysis (Morgan *et al.*, 2011; Glaser, 2014).

As with peer review specifically, writing in general must be integrated into the science classroom with an intentional design, one that uses "structured scaffolding and ... assessment tools explicitly designed to enhance the scientific reasoning in writing" (Dowd *et al.*, 2015b, p. 39). To this end, practitioners and scholars highlight the need for science writing rubrics in classes that integrate writing (Reynolds and Thompson, 2011; Bird and Yucel, 2013; Mynlieff *et al.*, 2014; Dowd *et al.*, 2015b). Science writing rubrics articulate for instructors and students the aspects of a writing project most salient to science reasoning—clear research focus, well-reasoned data analysis, logically drawn implications for future research, and so on—while deprioritizing surface-level errors (Morgan *et al.*, 2011; Glaser, 2014). Rubrics also help instructors work effectively and efficiently with student writers (Hafner and Hafner, 2003; Reynolds and Thompson, 2011; Bird and Yucel, 2013; Dowd *et al.*, 2015b).

One example of a science writing rubric that has been shown to impact student learning is the Biology Thesis Assessment Protocol, as well as its close cousin, the Chemistry Thesis Assessment Protocol. Modeled after professional peer-review guidelines, these rubrics outline departmental expectations alongside guidelines for both giving and receiving constructive feedback, an approach that has been shown to promote growth in writing and critical-thinking skills (Reynolds *et al.*, 2009). Others include the Developing Understanding of Assessment for Learning model (Bird and Yucel, 2013) and the VALUE (Valid Assessment of Learning in Undergraduate Education) rubrics put forward by the Association of American Colleges and Universities (AAC&U). Rubrics aid departments and institutions in accurately assessing students' scientific reasoning and writing skills by providing a common metric that can be used across a number of courses (Timmerman *et al.*, 2011), longitudinally across a series (Dowd *et al.*, 2015b), or across a program or institution (Rhodes, 2011; Rhodes and Finley, 2013; Holliday *et al.*, 2015).

Despite the unique value that rubrics offer for making standardized assessments of students' writing, including longitudinal assessments, few researchers use them to generate data for student performance on authentic science writing tasks. Reynolds and Thompson (2011) and Dowd *et al.* (2015a,b), who use TAP rubrics to measure and compare students' science reasoning on writing assignments, are notable exceptions; most studies on writing in the sciences lack concrete data on students' writing. Some researchers focus instead on how adding writing to science classes improves retention of information and performance on exams, neglecting its greater value in

developing students' ability to meaningfully participate in the scientific inquiry process (Günel *et al.*, 2007; Kingir *et al.*, 2012; Gingerich *et al.*, 2014; van Opstal and Daubenmire, 2015). More studies stress the value of writing to students' overall ability to conceptualize and integrate their knowledge in the discipline but lack evidence that demonstrates this effect in students' writing. Many have only self-reported metrics to gauge the effectiveness of these interventions; these include their perceived success in posing novel questions and experimental design (Stanford and Duwel, 2013; see also Kalman, 2011), understanding and communicating science concepts (Brownell *et al.*, 2013), and attitudes toward science writing and writing instruction (Morgan *et al.*, 2011).

The most meaningful studies are those that substantively assess students' writing according to their performance on pre-identified areas relevant to scientific reasoning and presentations. Sampson *et al.* (2011) measured the effects of a writing-intensive intervention for student performance on a writing task through a pre- and postassessment writing task designed for the study. In contrast, Dowd *et al.* (2015b) used TAP rubrics to examine the correlation between writing-focused supplemental support and students' writing on undergraduate thesis projects (p. 14). Dowd and coauthors evaluated more than a decade's worth of student work to conclude that "students who participated in structured courses designed to support and enhance their research exhibited the strongest learning outcomes" (p. 14).

Many previous studies also focus on the effects of writing instruction on one assignment type, such as lab reports (e.g., Morgan *et al.*, 2011), grant proposals (Cole *et al.*, 2013), reflective writing (Kalman, 2011), or senior capstone assignments (Lankford and vom Saal, 2012). Others have looked at writing performance across course levels (Schen, 2013) or at science writing in large undergraduate courses (Mynlieff *et al.*, 2014; Kelly, 2015). However, few studies have examined the results of writing interventions across a multisemester curriculum on student achievement in science inquiry and analysis across all students in a program. Here we focus on the impact of multicourse, scaffolded writing assignments with peer review on science inquiry and analysis skills as measured in the senior thesis research projects required of all undergraduate biology majors at the University of La Verne. The aims governing our longitudinal study were threefold: 1) to examine gains in each of the six AAC&U rubric dimensions applied to the capstone projects; 2) to investigate whether students experienced the changes differently across ethnicity, Hispanic identity, legacy, and research laboratories; and 3) to observe whether this focus on writing in science courses correlated with changes in on-time thesis completion.

METHODS

The University of La Verne is a private, doctoral-granting university located in Southern California. The university offers a mission-driven liberal arts education for a student population that is nearly 50% Latino, with a high proportion of students who are the first in their families to attend college. All students at the University of La Verne complete an undergraduate capstone requirement (Appendix A in the Supplemental Material). In biology, the capstone includes undergraduate research, a 25- to 30-page written thesis, and a 15-minute oral presentation at

a research conference or the department's senior symposium. On-time completion of senior theses—"on-time" meaning the final project was submitted by the end of the summer following the second senior seminar course—was tracked over 8 years for the current study. On-time completion rates were consistently low before 2008, with only ~25% of students completing their capstone projects in time to graduate within 4 years. Much of this deficit was attributed to differences in students' writing abilities coming into the senior year. These disparities, however, were not being addressed in the curriculum, as students did not take any classes with formal writing instruction after their first year. A lack of faculty interest in integrating writing and writing instruction into biology courses may also have contributed to this deficit.

In addition to contributing to low on-time completion rates, this lack of instruction in research writing contributed to students' failure to consistently engage in an authentic science writing and inquiry process; faculty teaching in the senior seminar noted that students were attempting to develop, analyze, and write their theses all in the final month of their senior year. Beyond a pragmatic failure to plan and develop their projects, students' procrastination indicated a larger failure in their ability to engage in meaningful scientific research. Absent explicit instruction guiding them through a thorough, iterative process for developing and revising their writing projects, students lacked the opportunity to develop full understanding of their projects' significance, data and findings, and implications for future research.

Course Description and Design

Two treatment groups were identified from the data generated by the introduction of the multisemester writing-centered biology course work: pretreatment (2006–2008) and intervention (2009–2013). Pretreatment groups were composed of students completing their research in 2006, 2007, and 2008 with a two- to four-unit senior seminar course. These units could be completed in two face-to-face courses with the biology department chair or as a single directed study with an individual faculty member. Through 2008, the senior seminar course was intended to monitor thesis completion and provide a grade for the research; the curriculum was not focused on writing to learn, nor did it contain formalized instruction in science writing or use peer review to promote continuous improvement in students' writing. In 2008, we piloted adding one formalized writing assignment, including peer review, to the second-semester senior seminar course; however, no formal junior series existed, nor did the intervention span the semester.

In an attempt to standardize and improve the instruction students received in writing, critical thinking, and data analysis, the authors, with the support of faculty in the biology department, developed a two-course series for the junior year; these courses, along with the senior seminar courses, were mandated in the curriculum for all biology majors. In 2008, Research Methods and Biostatistics was added to the junior year curriculum (for the graduating class of 2009 forward). This junior series of two two-unit courses in research methods and biostatistics, was created to teach statistical analysis and experimental design and to help students begin the writing process. The series starts students on their senior thesis projects by asking them to work on specific science writing and reasoning skills

through manageable writing assignments that focus on the skills necessary for effective writing in the sciences as outlined by Gopen and Swan (1990). Students from across research laboratories were enrolled in mixed groups and were taught by a professor from the department who may or may not have been their research mentor. The writing assignments included the following: a five-page literature review (Fall semester of junior year); a five-page grant proposal, including a summary of methods and budget (Spring semester of junior year); and a draft of the introductory five pages of their thesis project (Spring semester of junior year). Students were encouraged to work with their faculty research mentors on topic selection so their writing pieces could dovetail easily into their senior theses.

In 2009, senior seminar was further modified to include training on effective peer review and on presenting data and results. The redesigned senior seminar series integrated four peer-review sessions per semester. Peer-review submissions were five-page sections of the larger thesis project, which the course designers felt was a manageable unit for submission, critique, and revision. In addition, students developed and continuously modified writing goals and a work plan.

For peer-review sessions, the course instructor formed critiquing groups of three to four students, alternating between inter- and intralab members to provide a range and depth of perspective. The intralab peer groups were effective in providing feedback on accuracy of science, as students from the same lab were often given common readings at lab meetings. Interlab groups were helpful for students in developing their abilities to write for broader scientific audiences. Before review, writers were encouraged to give their peers guidance on their papers' goals and the areas where they wanted the most feedback (Appendix B in the Supplemental Material). Peer reviewers were given 1 week to critique and provide written feedback on the draft based on the guidelines provided by the professor (Appendix C in the Supplemental Material). In class, groups would meet over a 50-minute period to share their written comments. The professor or a postbaccalaureate teaching assistant were available to model the peer-review process and to provide clarification when disagreements or confusion arose. Writers then had a week to incorporate the peer feedback, after which time the workshop would be repeated with a new group.

Participants

All senior research theses were collected from 2006 to 2013 across research laboratories, for a total of 84 students with nine faculty instructors across eight years. The following sample size was reported for each of the academic years: 2006 ($n = 9$), 2007 ($n = 10$), 2008 ($n = 15$), 2009 ($n = 9$), 2010 ($n = 5$), 2011 ($n = 11$), 2012 ($n = 11$), and 2013 ($n = 14$). The research foci were dependent on the faculty members and their area of expertise. Examples of research topics included examining the effects of endocrine disruptors on the developing immune system; biogeography and toxicology studies of the land snail *Oreohelix* as a model organism; effects of free radicals on mitochondrial energy production in aging, obesity, and neurodegenerative models; and delaying programmed cell death during the development of *Zea mays* to produce more efficient crops and optimize overall yield. While research laboratories could be determined based on content, students' identities were protected. All identifying information was removed from the theses before

they were delivered to the raters in order to develop a blind study, with each thesis then assigned an identification code.

As a Hispanic-serving institution committed to providing equitable education to all students, including those from educationally underserved backgrounds, we also felt it was important to collect demographic information for each student writer so that these factors could be included in the data set: ethnicity (Hispanic, white, Black, Asian/Pacific Islander, and other), Hispanic/non-Hispanic identity, and legacy (first generation, first generation at La Verne, second generation at La Verne, and other). A student writer's demographic information was then assigned the same identification code linked to his or her thesis. A third party collated all data by student code after the rubric scoring was complete.

Rubric and Norming Committee

A scoring rubric was developed to measure the success of the intervention on the quality of students' science reasoning and writing as demonstrated in their senior thesis projects. The rubric was adapted from the AAC&U's VALUE rubrics. These rubrics were originally developed for AAC&U by an interdisciplinary team of faculty from a wide variety of institutions who together determined the "core expectations" of student learning in key areas, including science reasoning, and coded these as scoring categories in the rubrics (Rhodes, 2011). Subsequent research has demonstrated the face and content validity and high degrees of reliability of these rubrics for assessing student learning (Finley, 2011).

The current study relied on the AAC&U VALUE Rubric for Inquiry and Analysis, with one additional scoring category drawn from the Written Communication VALUE rubric (see Appendix D in the Supplemental Material). In the rubric-development process, tweaks were made to the wording of a small number of the scoring dimensions to clarify the distinctions between scoring categories and to better fit the local context (e.g., "uses graceful language that skillfully communicates meaning to readers" (Association of American Colleges and Universities, 2011) was changed to read "uses expert language that skillfully communicates meaning to readers." These kinds of modifications are recommended for institutions adopting AAC&U VALUE rubrics (Rhodes, 2011), and as the purpose of the rubric is to make reliable intrainstitutional assessments of student learning possible (as opposed to interinstitutional comparisons), they do not affect the validity or reliability of the rubrics.

The final University of La Verne scoring rubric contains five dimensions taken from the Inquiry and Analysis rubric: 1) existing knowledge, research, and/or views (EK); 2) design process (DP); 3) analysis (A); 4) conclusion, sources, and evidence (C); and 5) limitations or implications (LI). An additional scoring category was added from the Written Communication rubric: 6) coherence, control of language, and readability (CCR). Each dimension was scored on a scale of 0–4, which corresponds with below benchmark (0), benchmark (1), milestone (2 and 3), and mastery (4). Although the rubric allows for a student's thesis to score a 0 on a given dimension, multiple dimensions scoring at or below 1 were rare after the intervention, as faculty had agreed on a standard required for completion of the thesis requirement or course. Following best practices for rubric adoption (Finley, 2011), scoring of the thesis projects was normed by a norming committee made up of four postbaccalaureates from

different research laboratories, including one laboratory outside the University of La Verne; three faculty members; and the director of the Center for Academic and Faculty Excellence on campus. Scoring methods followed Dowd *et al.* (2015b). Norming sessions were held before the reading and marking of the test group, using senior theses outside the test years (2003–2005). The norming committee assessed and discussed these theses to ensure consistency and fairness in scoring. Each thesis in the test group was assessed independently by two raters over the course of a 3-week period in the Summer of 2015. The raters then met to discuss the final scores and establish a consensus score, which was a discussion-based final score rather than an average of both scores. As suggested by Brown (2010), in situations in which an agreement could not be met by both raters and/or disagreement between raters exceeded 1 point, a third rater served as a tiebreaker. Consensus scores were used in all analyses.

Assessment Methods

To assess the efficacy of the multicourse writing emphasis intervention, we tested composite scores across the treatment groups. The result of the intervention was also assessed across student ethnicity, student Hispanic identity, student legacy, and individual research lab. Some data were missing from the sequence of courses individual students completed (both in the pretreatment and intervention groups), as were some demographic variables (e.g., ethnicity); the affected students were dropped from analyses using those factors. An examination of the composite scores for the eliminated students reveals they were in line with the greater sample, and their exclusion from further analysis did not bias the results.

Each dimension (e.g., analysis) was tested for differences pretreatment versus intervention using a Fisher-Freeman-Halton test for R×C tables, which allows for comparison between the observed and expected values for ordinal (rubric dimension) and nominal (pretreatment or intervention) data. The small sample size prohibited further categorical data analysis to detail the student demographic and research laboratory influences on the dimension score across the treatment groups. To facilitate an examination of these variables, given the small sample size, we developed a composite score for each thesis that summed all of the scores for the individual rubric dimensions so that the composite scores could be treated as interval data for analysis in a blocked analysis of variance (ANOVA). The composite score was developed assuming that all rubric areas are of equal value; a sentiment derived from the identification of these as key areas for excellence in scientific inquiry by the AAC&U. Composite scores were normally distributed, so a two-sample Student's *t* test was performed on the composite score (pretreatment versus intervention) to verify whether it effectively modeled the increase in student performance among the six dimensions.

A randomized complete block design ANOVA was run on the composite score as the dependent variable with an independent (treatment) variable representing pretreatment and intervention categories and an independent (block) variable of a particular student demographic or research laboratory. The ANOVA analyses were carried out to examine ethnicity, Hispanic identity, legacy, and individual research laboratory association on composite scores in the treatment groups. However, not all fac-

ulty research laboratories were represented in both the pretreatment and intervention for various reasons: 1) the department chair and some senior faculty take a lower undergraduate research load in the department, 2) the timing of sabbaticals interrupted some faculty members' participation, and 3) student interest in faculty research areas varies. Three new faculty were hired during the course of the study, which further complicates the pre/post comparisons.

A one-way ANOVA was implemented to examine change over time of the composite scores within the intervention group. For all analyses, determination of significance was based on an aggregation of evidence against the null (Wasserstein and Lazar, 2016). Here we reach our conclusions of significance on *p* values from multiple tests, trends depicted in the figures, and considerations suggested by Wasserman and Lazar, such as research design (randomized complete block design), validity and reliability of the variables (vetted rubrics, normalized data-collection procedures), and assumptions (normality of the composite variable, external factors considered such as institutional changes). Additional data that support the conclusions (e.g., graduation rates) were also considered. The 0.10 and 0.05 *p* value cutoffs (α) were selected in light of this evidence (Wasserstein and Lazar, 2016). All statistical analyses were conducted in NCSS (Hintze, 2007), while figures were generated in R (R Core Team, 2013) and Microsoft Excel.

RESULTS

Interrater equal scoring within one point of the rubric was determined at 89%. The chi-square results from the individual rubric dimensions were all significant at the 0.10 level or lower (Table 1), with all but LI significant at the 0.05 level. Thus, students scored significantly higher than expected in each rubric dimension once the intervention was applied (EK: $\chi^2 = 12.402$, $df = 3$, p value = 0.0061; DP: $\chi^2 = 13.261$, $df = 3$, p value = 0.0041; A: $\chi^2 = 19.254$, $df = 3$, p value = 0.0002; C: $\chi^2 = 9.0942$, $df = 3$, p value = 0.0281; LI: $\chi^2 = 9.3525$, $df = 4$, p value = 0.0529; CCR: $\chi^2 = 11.888$, $df = 3$, p value = 0.0078). The two-tailed *t* test demonstrates that composite scores were significantly different in pretreatment and intervention groups (t value = -4.4465 , $df = 79$, p value < 0.0001), confirming the significant chi-square results (Table 1). An examination of the composite scores plotted across the pretreatment and intervention

TABLE 1. Student's *t* test and Fisher-Freeman-Halton tests for pretreatment and intervention against composite score and each rubric class

Variable ^a	<i>t</i> value/chi-square	<i>df</i>	<i>p</i> value
Composite	-4.4465	79	<0.0001*
EK	12.402	3	0.0061*
DP	13.261	3	0.0041*
A	19.254	3	0.0002*
C	9.0942	3	0.0281*
LI	9.3525	4	0.0529
CCR	11.888	3	0.0078*

^aVariables were as follows: Inquiry and Analysis rubric: Composite, sum of all rows; EK, existing knowledge, research, and/or views; DP, design process; A, analysis; C, conclusion, sources, and evidence; and LI, limitations or implications; Written Communication Rubric: CCR, coherence, control of language, and readability.

*Significant at 0.05 level or lower.

groups shows they increased in the intervention group, which mirrors the rubric dimension findings and portrays the outcome of the intervention on student performance (Figure 1).

The blocked ANOVA that tested whether students experienced the same correlation of the intervention across ethnicities (Table 2) revealed a significant trend across treatment groups (pretreatment and intervention) but not across the block (ethnicity). It can be concluded that there was no significant correlation with ethnicity in these data (treatment F value = 22.84, $df = 1$, p value ≤ 0.0001 ; block F value = 0.48, $df = 4$, p value = 0.7526), and therefore, students of different ethnicities experienced the research core treatment similarly. The pattern of statistical significance with the treatment variable and a lack of block association was the same for Hispanic identity of a student (treatment F value = 22.85, $df = 1$, p value ≤ 0.0001 ; block F value = 0.19, $df = 1$, p value = 0.6682) and legacy (treatment F value = 16.15, $df = 1$, p value = 0.0002; block F value = 0.03, $df = 3$, p value = 0.9917), indicating the research capstone changes affected student performance similarly regardless of a student's Hispanic identity or legacy status (first generation, first generation at La Verne, second generation at La Verne, or other) as a college student.

The research lab block variable was significant when considered with the treatment variable (pretreatment vs. intervention: treatment F value = 11.44, $df = 1$, p value = 0.0011; block F value = 4.65, $df = 8$, p value = 0.0001), demonstrating that student performances differed across treatment groups and research laboratories (Table 2), and was a source of the variation in these results. Finally, the one-way ANOVA was not significant (F value = 0.1211, $df = 4$, p value = 0.974), demonstrating the composite score gains were stable within the intervention group from implementation to the end of the data-collection time frame.

The institutional data on graduation rates for the students in the study support the statistical findings of significance, wherein the intervention years saw an increase in graduation rates (pretreatment years = 36% average completion, with a low of 0% in 2006 and a high of 75% in 2008; intervention years = 80% average completion with a low of 63% in 2010 and a high of 91% in 2009). The high in the pretreatment correlates with the pilot introduction of writing in the senior seminar course in 2008, and the low in the intervention years corresponds with

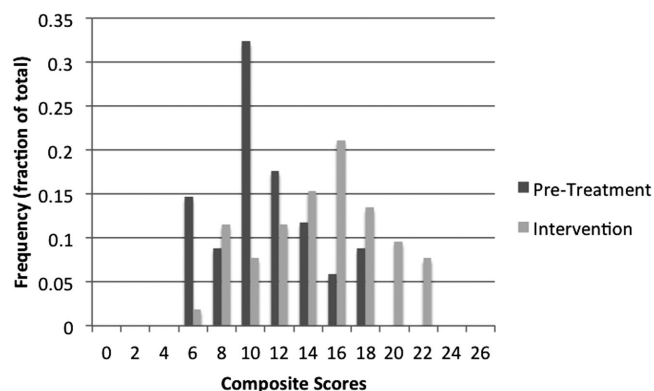


FIGURE 1. Count distribution of the composite scores in the pretreatment and intervention groups.

TABLE 2. Composite score changes across treatment groups

Test (block df)	F value treatment ($df = 1$)	F value block	p value treatment	p value block
Ethnicity (4)	22.84	0.48	<0.0001*	0.7526
Hispanic identity (1)	22.85	0.19	<0.0001*	0.6682
Legacy (3)	16.15	0.03	0.0002*	0.9917
Research lab (8)	11.44	4.65	0.0011*	0.0001*

*Significant at 0.05 level or lower.

the class in which every student missed at least one part of the intervention (due to transfer or directed-study courses).

DISCUSSION

Results from this study indicate that science inquiry and analysis skills significantly improved with the addition of a series of writing-centered biology courses in the junior and senior years that integrate scaffolded writing assignments to culminate in a senior thesis and a formal and iterative peer-review process. Undergraduate theses in the intervention group demonstrated improvements not only in their composite scores but also in the individual dimensions identified by the scoring rubric, including EK, DP, A, C, and CCR.

Most significantly, our results demonstrate that write-to-learn pedagogies in science classes result in substantive improvements in key measures of science literacy. The aim of a write-to-learn approach is to improve students' *thinking*, along with their writing, and on this metric, students who received the intervention in our study excelled. In addition to the score related to error-free writing, students' performance on the dimensions from the Inquiry and Analysis rubric that correspond to critical thinking and science literacy (EK, DP, C, and most critically, A) showed significant improvement. These gains recommend the use of write-to-learn pedagogies to foster deep learning and critical-thinking skills in science students. Using research writing as a vehicle for learning requires students to understand and address a specific audience, to identify a logical research gap, to analyze and synthesize information from relevant sources, to develop a theoretical framework for a valid methodology, and to uncover and explain insightful data patterns related to the research focus, all of which are essential components of science reasoning (Hand et al., 1999; Yore et al., 2003; Prain, 2006; Grimberg and Hand, 2009; Schen, 2013). Our results indicate that science curricula that include multisection, writing-centered course work with a carefully constructed peer-review component will result in deeper gains for student learning when compared with those that do not. For institutions looking to improve their STEM students' performance, write-to-learn classes suggest themselves as a high-impact, low-cost intervention.

In addition, our results support the importance of undergraduate research for developing inquiry and analysis skills in all students, notably including equal gains for first-generation college students and students from underrepresented populations. More specifically, results from this study suggest that write-to-learn pedagogies that stress scaffolded writing assignments and iterative feedback lead to improvements in science literacy for students regardless of ethnicity or legacy status.

While the benefits of building meaningful writing assignments into science courses are deep and wide ranging, as supported by the results presented earlier, significant pragmatic barriers keep more departments from adopting the practice. The most significant impediment is faculty attitudes, a point captured by Lankford and vom Saal (2012), who cite previous studies showing that “faculty equated the time and effort required for teaching a single, writing-intensive course to teaching two courses instead of one” (p. 21). Other researchers report similar findings on faculty perceptions (Theoret and Luna, 2009; Delcham and Sezer, 2010), while still others highlight the increased time demands and costs associated with writing-intensive courses (Stanford and Duwel, 2013; Dowd *et al.*, 2015b). In writing-intensive classes: “Faculty are required to expend additional time and energy in the development of course work designed to engage students with critical thinking, peer collaboration, formal papers, and rubric development for nontraditional assessments” (p. 21). These demands, plus the time and energy required to provide frequent, substantive, and timely feedback on students’ writing, necessitates what Delcham and Sezer (2010) label “magnum doses of patience and nurturing on the part of the instructor” (p. 612).

To mitigate the workload requirements for La Verne’s intervention, faculty instructors in the capstone series were given an extra unit toward their teaching load (three units of credit instead of two units per course). Once the series was established, the courses were shared across the faculty, and the descriptions for new faculty hires listed the series as part of the workload requirements. Finally, a small amount of departmental funds was allocated for postbaccalaureate teaching assistants. These assistants helped facilitate peer-review sessions, which, combined with the peer reviewing itself, also decreased faculty workload for first and second drafts. Overall, these high-impact practices were given a high priority by the depart-

ment and were listed prominently in the program learning outcomes, so resources were committed to their implementation.

Though peer review was stressed in the curriculum across the series as a means of increasing learning, it also had positive effects on instructor workload. This is in keeping with previous studies that have shown how incorporating peer review into writing-intensive classes helps mitigate workload obstacles (Walker and Sampson, 2013; Mynlieff *et al.*, 2014; Dowd *et al.*, 2015a,b). Peer review offers an avenue for student writers to get substantive, regular feedback without instructors becoming overwhelmed by time and effort demands (Timmerman and Strickland, 2009; Sampson *et al.*, 2011; Dowd *et al.*, 2015a,b). Because students gained writing and research help from one another outside the mentor/mentee relationships, faculty mentors reported a workload reduction. They also noted an increase in the quality of student learning: students in the intervention worked more independently of their faculty mentors and with more success. As long as courses are shared among the faculty, the workload decrease appears to balance the increase in effort required of course faculty. Still, issues of faculty workload and resistance can be seen in our data. While the results show a significant trend of improvement in the students’ performance corresponding to the intervention across all research labs, there is variation in the degree of effectiveness of the various labs. This reflects a regrettable but perhaps unavoidable difference in faculty investment in the write-to-learn pedagogy. It suggests the need for both department-wide buy-in and robust training for faculty who are new to write-to-learn teaching.

Whatever the pragmatic difficulties, the impact on student success recommends the implementation of a multicourse write-to-learn sequence for all science students. In addition to student gains in inquiry, analysis, and writing skills, the biology department also saw an improvement in on-time completion of the senior thesis projects, from 36 to 80% (Figure 2). In the

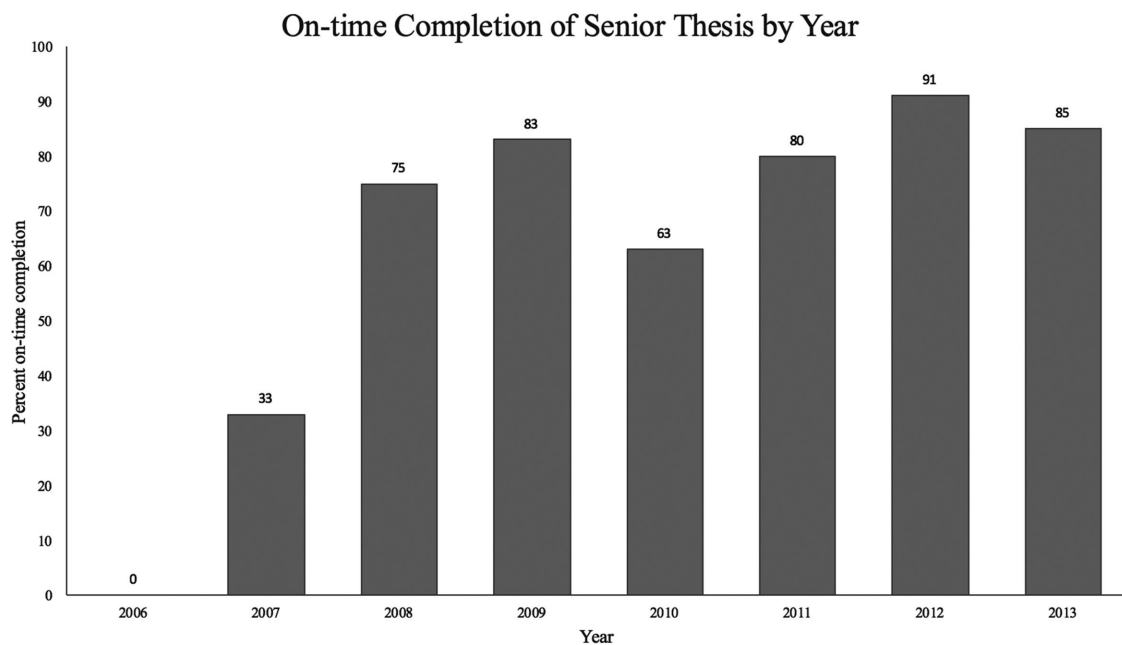


FIGURE 2. Bar graph showing the change in on-time thesis completion over time.

remaining years of the intervention, on-time completion rates were all greater than 89%.

One caveat for the study is that students in the intervention groups were given access to the modified VALUE rubric during the senior seminar courses. Because the use of rubrics in writing courses alone are shown to help faculty improve student writing (Hafner and Hafner, 2003; Reynolds and Thompson, 2011; Bird and Yucel, 2013; Dowd *et al.*, 2015b), we cannot in this study differentiate between the positive effect of rubrics, peer review, and the addition of new courses to support writing and statistical analysis. It may be an individual component or a combination of effects that led to the improvement in students' inquiry and analysis skills; however, a quality design for an efficient and effective write-to-learn curriculum would integrate rubrics almost by necessity, with peer review as a pragmatic and resource-efficient complement.

A second caveat should note that, in addition to the intervention described in this study, other departmental and university-wide changes may also have contributed to an increase in graduation rates and student success during this time period. In 2008, Cell Biology and Developmental Biology were redesigned into class-based undergraduate research experiences (CUREs; Madhuri and Broussard, 2008). In addition, in 2011, the University of La Verne began its focus on the La Verne Experience, a series of high-impact practices including first-semester learning communities, community-engagement experiences, and common intellectual experiences. Although this group of students will not graduate until Spring 2015, the university-wide focus on teaching strategies and student success also may have influenced the successes in the biology department and the outcomes seen in this study.

Despite the complicating factors, the results of the intervention recommend it as a promising practice for science education. The improvement demonstrated in the current study in students' inquiry and analysis skills for students of all ability ranges and backgrounds and the gains in on-time completion of the thesis projects warrant a multisection write-to-learn curriculum as worth examining for any program looking to improve its students' success.

ACKNOWLEDGMENTS

Thanks to professors Margie Krest, Jerome Garcia, Christine Broussard, and Anil Kapoor, who provided feedback and helped implement the series through thoughtful instruction. Additional thanks to Sammy Elzarka, Amanda Todd, Bradley Blackshire, and Jeremy Wagoner for their scoring assistance. Support for this work was provided by a Title V, HSI-STEM grant. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Education. IRB approval was obtained for this work (2012-CAS-20-Weaver).

REFERENCES

- Akcaay H, Hand B, Norton-Meier L (2010). Creative writing and promoting understanding in science: alternative ways to interest students in writing about science. *School Sci Rev* 92(339), 49–53.
- Association of American Colleges and Universities (2011). VALUE Rubrics Case Studies. www.aacu.org/value/casestudies; www.aacu.org/value-rubrics (accessed February 2015).

- Bangert-Drowns RL, Hurlley MM, Wilkinson B (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Rev Educ Res* 74, 29–58.
- Bird FL, Yucel R (2013). Improving marking reliability of scientific writing with the developing understanding of assessment for learning programme. *AEHE* 38, 536–553.
- Brown GT (2010). The validity of examination essays in higher education: issues and responses. *Higher Educ Q* 64, 276–291.
- Brownell SE, Price JV, Steinman L (2013). A writing-intensive course improves biology undergraduates' perception and confidence of their abilities to read scientific literature and communicate science. *Adv Physiol Educ* 37, 70–79.
- Brownell JE, Swaner LE (2010). Five High-Impact Practices: Research on Learning Outcomes, Completion and Quality, Washington, DC: Association of American Colleges and Universities.
- Cole KE, Inada M, Smith AM, Haaf MP (2013). Implementing a grant proposal writing exercise in undergraduate science courses to incorporate real-world applications and critical analysis of current literature. *J Chem Educ* 90, 1316–1319.
- Delcham H, Sezer R (2010). Write-skewed: writing in an introductory statistics course. *Education* 130, 603–615.
- Dowd JE, Connolly MP, Thompson RJ Jr., Reynolds JA (2015a). Improved reasoning in undergraduate writing through structured workshops. *J Econ Educ* 46, 14–27.
- Dowd JE, Roy CP, Thompson RJ Jr., Reynolds JA (2015b). "On course" for supporting expanded participation and improving scientific reasoning in undergraduate thesis writing. *J Chem Educ* 92, 39–45.
- Finley AP (2011). How reliable are the VALUE rubrics? *Peer Rev* 13(4/1), 31–33.
- Gingerich KJ, Bugg JM, Doe SR, Rowland CA, Richards TL, Tompkins SA, McDaniel MA (2014). Active processing via write-to-learn assignments: learning and retention benefits in introductory psychology. *Teach Psychol* 41, 303–308.
- Glaser RE (2014). Design and assessment of an assignment-based curriculum to teach scientific writing and scientific peer review. *J Learn Design* 7, 85–104.
- Gopen GD, Swan JA (1990). The science of scientific writing. *Am Scientist* 78, 550–558.
- Grimberg BI, Hand B (2009). Cognitive pathways: analysis of students' written texts for science understanding. *Int J Sci Educ* 31, 503–521.
- Guilford WH (2001). Teaching peer review and the process of scientific writing. *Adv Physiol Educ* 25, 167–175.
- Gunel M, Hand B, Prain V (2007). Writing for learning in science: a secondary analysis of six studies. *Int J Sci Math Educ* 5, 615–637.
- Hafner J, Hafner P (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *Int J Sci Educ* 25, 1509–1528.
- Hand B, Lawrence C, Yore LD (1999). A writing in science framework designed to enhance science literacy. *Int J Sci Educ* 21, 1021–1035.
- Hintze J (2007). NCSS, PASS and GESS, Kaysville, UT: NCSS.
- Holliday W, Dance B, Davis E, Fagerheim B, Hedrich A, Lundstrom K, Martin P (2015). An information literacy snapshot: authentic assessment across the curriculum. *Coll Res Libr* 76, 170–187.
- Hu S, Kuh GD, Li S (2008). The effects of engagement in inquiry-oriented activities on student learning and personal development. *Innov High Educ* 33, 71–81.
- Kalman CS (2011). Enhancing students' conceptual understanding by engaging science text with reflective writing as a hermeneutical circle. *Sci Educ* 20, 159–172.
- Kelly L (2015). Effectiveness of guided peer review of student essays in a large undergraduate biology course. *Int J Teach Learn High Educ* 27, 56–68.
- Kingir S, Geban O, Gunel M (2012). How does the science writing heuristic approach affect students' performances of different academic achievement levels? A case for high school chemistry. *Chem Educ Res Pract* 13, 428–436.
- Lankford D, vom Saal F (2012). Converting a biology course into a writing-intensive capstone course: using collaboration between a professor and graduate teaching assistant. *J Coll Sci Teach* 41, 14–22.

- Lu J, Law N (2012). Online peer assessment: effects of cognitive and affective feedback. *Instr Sci* 40, 257–275.
- Madhuri M, Broussard C (2008). "Do I need to know this for the exam?" Using popular media, inquiry-based laboratories, and a community of scientific practice to motivate students to learn developmental biology. *CBE Life Sci Educ* 7, 36–44.
- Morgan W, Fraga D, Macauley WJ (2011). An integrated approach to improve the scientific writing of introductory biology students. *Am Biol Teach* 73, 149–153.
- Mynlieff M, Manogaran AL, St Maurice M, Eddinger TJ (2014). Writing assignments with a metacognitive component enhance learning in a large introductory biology course. *CBE Life Sci Educ* 13, 311–321.
- Prain V (2006). Learning from writing in secondary science: some theoretical and practical implications. *Int J Sci Educ* 28, 179–201.
- Quitadamo IJ, Kurtz MJ (2007). Learning to improve: using writing to increase critical thinking performance in general education biology. *CBE Life Sci Educ* 6, 140–154.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing.
- Reynolds J, Smith R, Moskovitz C, Sayle A (2009). BioTAP: a systematic approach to teaching scientific writing and evaluating undergraduate theses. *BioScience* 59, 896–903.
- Reynolds JA, Thompson RJ (2011). Want to improve undergraduate thesis writing? Engage students and their faculty readers in scientific peer review. *CBE Life Sci Educ* 10, 209–215.
- Rhodes TL (2011). Emerging evidence on using rubrics. *Peer Rev* 13(4/1), 4.
- Rhodes TL, Finley AP (2013). *Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment*, Washington, DC: Association of American Colleges and Universities.
- Sampson V, Grooms J, Walker JP (2011). Argument-driven inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: an exploratory study. *Sci Educ* 95, 217–257.
- Schen M (2013). A comparison of biology majors: written arguments across the curriculum. *J Biol Educ* 47, 224–231.
- Stanford JS, Duwel LE (2013). Engaging biology undergraduates in the scientific process through writing a theoretical research proposal. *Bioscene: J Coll Biol Teaching* 39, 17–24.
- Theoret JM, Luna A (2009). Thinking statistically in writing: journals and discussion boards in an introductory statistics course. *Int J Teach Learn High Educ* 21, 57–65.
- Timmerman B, Strickland D (2009). Faculty should consider peer review as a means of improving students' scientific reasoning skills. *JSC Acad Sci* 7, 1–7.
- Timmerman BEC, Strickland DC, Johnson RL, Payne JR (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assess Eval High Educ* 36, 509–547.
- Tytler R, Prain V (2010). A framework for re-thinking learning in science from recent cognitive science perspectives. *Int J Sci Educ* 32, 2055–2078.
- van Opstal MT, Daubenmire PL (2015). Extending students' practice of metacognitive regulation skills with the science writing heuristic. *Int J Sci Educ* 37, 1089–1112.
- Walker JP, Sampson V (2013). Learning to argue and arguing to learn: argument-driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *J Res Sci Teach* 50, 561–596.
- Wasserstein RL, Lazar NA (2016). The ASA's statement on p values: context, process, and purpose. *Am Stat* 70, 129–133.
- Yore L, Bisanz GL, Hand BM (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *Int J Sci Educ* 25, 689–725.