

A Novel Assessment Tool for Quantitative Evaluation of Science Literature Search Performance: Application to First-Year and Senior Undergraduate Biology Majors

Jason M. Blank, Karen J. McGaughey, Elena L. Keeling, Kristen L. Thorp, Conor C. Shannon, and Jeanine M. Scaramozzino

Expertise in searching and evaluating scientific literature is a requisite skill of trained scientists and science students, yet information literacy instruction varies greatly among institutions and programs. To ensure that science students acquire information literacy skills, robust methods of assessment are needed. Here, we describe a novel tool for longitudinal, crossover assessment of literature-searching skills in science students and apply it to a cross-sectional assessment of literature-searching performance in 145 first-year and 43 senior biology majors. Subjects were given an open-ended prompt requiring them to find multiple sources of information addressing a particular scientific topic. A blinded scorer used a rubric to score the resources identified by the subjects and generate numerical scores for source quality, source relevance, and citation quality. Two versions of the assessment prompt were given to facilitate eventual longitudinal study of individual students in a crossover design. Seniors were significantly more likely to find relevant, peer-reviewed journal articles, provide appropriate citations, and provide correct answers to other questions about scientific literature. This assessment tool accommodates large numbers of students and can be modified easily for use in other disciplines or at other levels of education.

Jason M. Blank is Associate Professor and Elena L. Keeling is a Professor in the Department of Biological Sciences; Karen J. McGaughey is Assistant Professor and Conor C. Shannon is Undergraduate Research Assistant in the Department of Statistics; and Jeanine M. Scaramozzino is Associate Librarian in Robert E. Kennedy Library; all at California Polytechnic State University; e-mails: jmblank@calpoly.edu, ekeeling@calpoly.edu, kmcgaugh@calpoly.edu, conorcshannon@gmail.com, jscaramo@calpoly.edu. Kristen L. Thorp is Youth Services Manager at Eugene Public Library; e-mail: Kristenlthorp@gmail.com. © 2016 Jason M. Blank, Karen J. McGaughey, Elena L. Keeling, Kristen L. Thorp, Conor C. Shannon, and Jeanine M. Scaramozzino, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/3.0/>) CC BY-NC.

Introduction

Trained scientists in diverse disciplines share certain fundamental skills, and educators in science, technology, engineering, and mathematics (STEM) aim to train students in these cross-disciplinary skills in addition to the skills and knowledge of specific disciplines. Several of these fundamental skills are enumerated within the concept of science information literacy, which has been defined “as a set of abilities to identify the need for information, procure the information, evaluate the information and subsequently revise the strategy for obtaining the information, to use the information and to use it in an ethical and legal manner, and to engage in lifelong learning.”¹ Trained scientists must apply these skills routinely, whereas high school students typically evaluate sources and information on a more limited basis.² To develop these skills in undergraduate students, STEM curricula must incorporate information literacy.³ However, information literacy instruction varies greatly among institutions and programs.⁴ Likewise, assessment of information literacy varies widely in format, information skills and knowledge tested, time frame over which students are monitored, and degree to which student inputs are standardized and measurements are comparable between studies (see table 1).⁵

To address this variation, a definition of science information literacy that clearly enumerates its multiple components is essential. The concept of science information literacy advanced by the Association of College and Research Libraries (ACRL) includes a number of components that are directly relevant to undergraduate science coursework. These components include defining information needs, identifying potential sources, constructing search strategies, retrieving information, extracting information, and synthesizing concepts from the information. These skills are echoed in the standards of scientific disciplinary organizations and other agencies.⁶ Application of these skills often requires use of specialized scientific concepts and vocabulary. Thus several studies have emphasized the benefits to science majors who practice accessing and interpreting the scientific literature through assignments and instruction embedded in their major courses⁷ or provided at an appropriate career stage to complement major courses.⁸ Librarians or science faculty may deliver this discipline-specific information literacy instruction, which may be matched with assignments that require students to access the scientific literature and practice skills as they are introduced.⁹ Many science faculty assume that undergraduate students receive information literacy instruction at some point in the curriculum. However, this assumption is not always valid; when information literacy instruction is spread across disciplinary courses, some students may receive formal library instruction repeatedly while others never encounter it.¹⁰

Robust assessment methods may help to ensure adequate information literacy training amid the variety of curricular approaches and individual student experiences. Several studies have monitored the success of stand-alone library instruction courses or modules embedded within introductory science courses, primarily using assessments containing multiple-choice items measuring recognition of library tools, citation formats, and so on.¹¹ Other researchers have developed broader information literacy assessment tools requiring analysis of graphics, documents, and simulations,¹² but these assessments are not tailored to the sciences. Conversely, the Test of Scientific Literacy Skills (TOSLS)¹³ is clearly science-focused. The TOSLS includes a section that assesses the ability to evaluate the validity of sources along with measurements of a broad set of skills such as the use of statistics, interpretation of graphical data, and quantitative problem solving. However, the TOSLS does not directly examine the ability to search the published literature.

A number of studies have emphasized “authentic assessment” of information skills by examining portfolios of student work such as papers written for classes.¹⁴ As these studies point out, portfolio analysis is complicated by wide variation in assignments,

TABLE 1
Selected Published Assessments of Information Literacy

Reference	Doi	Science-Specific	Comparison	Population/Level	Type of Assessment
Multiple-choice questions on recognition of citation features, library catalog, databases, and the like.					
Greer et al., 1991	10.5860/crl_52_06_549	N	Cross-sectional user survey	Survey of library visitors, undergraduate	Multiple-choice and self-assessment
Ferrer-Vinent & Carello, 2008	10.1080/01942620802202352	Y	Pre-post test, semester course	Introductory biology course	Multiple-choice questions
Dunnington & Strong, 2010	n/a	N	Pre-post test, half-semester course	Library instruction course, introductory	Multiple-choice questions
Hufford, 2010	10.5860/0710139	N	Pre-post test, semester course	Library instruction course, introductory	Multiple-choice questions
Porter et al., 2010	10.1187/cbe.10-01-0006	Y	Pre-post test, semester course	Introductory biology course	Multiple-choice questions
Ferrer-Vinent & Carello, 2011	10.1080/0194262X.2011.592789	Y	Follow-up, 1 to 3 years	Introductory biology course	Multiple-choice and fill-in-the-blank
Assessments of information skills involving a range of question formats					
O'Connor et al., 2002 (SAILS)	10.5860/crl.63.6.528	N	Instrument development, no results	Undergraduate students	Multiple-choice questions, not science-specific
Cameron et al., 2007 (ILT)	10.5860/crl.68.3.229	N	Instrument development, limited results	1st- and 2nd-year undergraduates	Multiple-choice questions, not science-specific
Katz, 2007 (iSkills)	10.6017/ital.v26i3.3271	N	Instrument development, no results	Undergraduate, introductory and advanced	Multiple-choice questions, not science-specific
Gormally et al., 2012 (TOSLS)	10.1187/cbe.12-03-0026	Y	Pre-post test, semester courses	Undergraduates, various levels, 3 institutions	Multiple-choice questions, science-specific
Portfolio-based assessment					
Scharf et al., 2007	10.1016/j.acalib.2007.03.005	N	Descriptive, single time point	Humanities capstone course, undergraduate	Writing portfolio

TABLE 1
Selected Published Assessments of Information Literacy

	Reference	Doi	Science-Specific	Comparison	Population/Level	Type of Assessment
	Knight-Davis & Sung, 2008	10.5860/cr1.69.5.447	N	Cross-sectional	Variety of courses across academic years	Writing portfolio, citation analysis
Papers, annotated bibliographies, or other assignments within disciplinary classes						
	Wang, 2006	n/a	N	Took IL course or did not take IL course	Library instruction course, upper-level undergraduate	Citation analysis
	Flaspohler et al., 2007	10.1187/cbe.07-04-0022	Y	Descriptive, several years of same course	Upper-division immunology course	Annotated bibliographies
	Gehring & Eastman, 2008	10.1187/cbe.07-10-0091	Y	Pre-post, semester course	Developmental biology, upper level undergraduate	Written and oral assignment
	Samson, 2010	10.1016/j.acalib.2010.03.002	N	First-year vs. senior capstone	Non-science courses, undergraduate	Citation analysis in written assignment
	Gilbert et al., 2012	10.1017/S1049096511001788	N	Course section with library lab vs. no lab	Intermediate undergraduate political science course	Annotated bibliographies
Self-efficacy or self-report						
	DeBurman, 2002	10.1187/cbe.02-07-0024	Y	Pre-post, quarter or semester course	2nd-year undergraduate cell biology students	Self-assessment of skills
	Callinan, 2005	10.1108/00242530510583039	Y	First-year vs. final year	Undergraduate biology students	Self-reported e-resource usage
	Kozeracki, 2006	10.1187/cbe.06-02-0144	Y	Pre-post, two-year research experience	3rd- and 4th-year undergraduate biology students	Self-assessment of skills
	Shanahan, 2008	10.1016/j.ijmedinf.2007.10.004	Y	Pre-post-follow-up	Undergraduate radiography students, second year	Self-reported search behavior
Usage logs or observation						
	Judd & Kennedy, 2011	10.1111/j.1467-8535.2009.01019.x	Y	Cross-sectional	1st- to 3rd-year medical students	Usage logs and user questionnaires

instructions, and faculty guidance. The analysis of student papers or annotated bibliographies produced within defined disciplinary classes offers a more controlled approach. This method has been used to demonstrate the efficacy of library research courses and library instruction embedded in upper-level social sciences¹⁵ and biology courses.¹⁶ Despite the variety of approaches employed to date, few studies have compared performance in procuring and evaluating information between students of different academic-year classes. Studies that have tested performance have involved either reading the work of students in small classes¹⁷ or portfolio analysis with a wide variety of student inputs.¹⁸

This indicates a need for methods of information literacy assessment that are (1) quantitative, (2) capable of assessing performance, (3) readily scalable to large numbers of students, and (4) adaptable to a wide variety of scientific disciplines and other fields. Here, we demonstrate a tool for measuring the proficiency of students at searching and selecting scientific resources, a subset of information literacy that aligns with ACRL standards #2 and #3.¹⁹ Students respond to a prompt that requires them to find multiple sources of information addressing a particular scientific question. The resources identified by the students are then scored by blinded scorers according to a rubric. Two versions of the assessment prompt are given to facilitate a crossover design for eventual long-term follow-up of individual students. Here, we present data generated with this tool to compare scientific information searching skills of first-year and senior undergraduate biology majors at a single institution at a single point in time. However, the assessment framework is readily adaptable to other science majors or to populations outside the sciences.

Methods

Subjects

Participants included first-year and senior majors in biological sciences at California Polytechnic State University (Cal Poly). First-year students were recruited during fall 2013 from BIO 160, the first course in the biological sciences curriculum, during their first quarter of university enrollment. Our analysis omitted data from students enrolled in BIO 160 who were not first-year students (such as those who had transferred or changed majors) or were not biological sciences majors. Seniors, defined as students in their fourth or fifth year of college, were recruited throughout the 2013–2014 academic year from BIO 461, which is one of two options for a required senior capstone course. In both BIO 160 and BIO 461, assessments were assigned as homework for either regular course credit or extra credit. Decisions to award regular credit or extra credit were left to the discretion of the course instructors, who were not the authors of the study. Completion of the assessment was necessary to earn homework or extra credit points, but students were given the choice to opt in or opt out of the research study at the beginning of the assessment. Eleven students opted out, and their responses were deleted prior to data analysis. A total of 145 first-year students and 43 seniors completed the assessments and opted to participate in the research project. All students who opted to participate completed all of the assessment questions, although one subject's answers to questions 15–21 were omitted from data analysis due to an obvious failure to comply with instructions. Project methods were approved by the Institutional Review Board of California Polytechnic State University.

Assessments

Assessments requiring subjects to find and select relevant scientific literature on a designated topic were administered via a Moodle-based online course management system (CMS). Subjects completed the assessments as homework assignments outside

of class. Each assessment began with a prompt stating a scientific issue and asking subjects to find three peer-reviewed articles that could help address that issue. The prompt also instructed subjects to “keep note of all the search terms and sources you use.” Next, subjects entered a citation and URL for each source in text boxes provided in the CMS. They then answered a series of additional questions (15 through 21) meant to gather information about their search process and their awareness of what constitutes peer-reviewed scientific literature and the distinction between primary and secondary sources. Each response was entered in a text box or chosen from a menu of possible responses. The complete assessment text is provided in appendix A.

Two versions of the assessment prompt were created to enable eventual longitudinal assessment of individual student performance spanning each student’s undergraduate career using a crossover design. Each student was arbitrarily assigned to complete one of the two assessment versions based on the first letter of the student’s last name. The scientific issue presented in the prompt differed between the two versions, but the structure of the assessment and the text of all subsequent tasks and questions were identical.

Assessments were administered early in each 10-week academic quarter. BIO 461 includes significant bibliographic instruction; therefore, the assessment due date was in the first or second week of the quarter, before any formal instruction on literature searching was provided by either the instructor or the College of Science and Mathematics librarian. The assessment due date for subjects enrolled in BIO 160 was during the third week, so some first-year students may have received bibliographic instruction in other nonscience classes during the period that the assessment was available. In general, students were provided one to two weeks to complete the assessment after it was made available in the CMS. Other than the due date, no time limit was imposed on the assessment. For several reasons, we were unable to measure the time that students spent on the assessment. The CMS provides limited detail on student activities, and students conducted online literature searches outside of the CMS. After accessing the assessment, students may have completed the searching tasks and questions immediately, postponed them until later, or multitasked or taken breaks while remaining logged in to the CMS. Due to the large class size in BIO 160, it would not have been feasible to conduct the assessment in class.

Scoring and Statistical Analysis

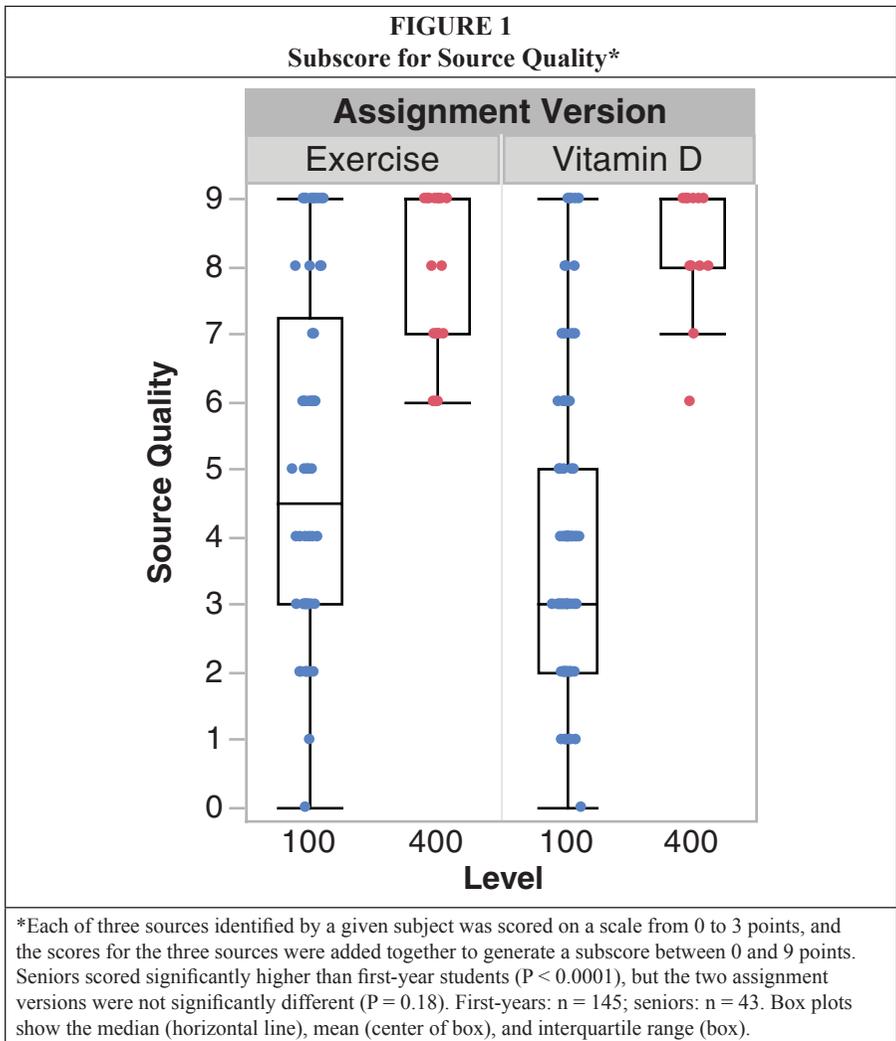
Each participant’s assessment response was assigned a unique numerical code and stripped of subject-identifiers as well as any other information that would reveal the academic year of the subject. A single individual who was blinded to subject identity and class level then scored all responses according to a rubric that provides score values for each category of response to each task or question (see appendix B).

To evaluate rubric reliability, two additional scorers rescored a subset of 30 responses, balanced across assessment versions and subject levels. Cronbach’s alpha, a common measure of interrater reliability, was calculated for each question or question type (such as citation quality). For this analysis, no attempt was made to calibrate scorers or adjudicate differences in scores. Following IRR analysis, a second scorer rescored all responses to question 18 and any divergent scores were replaced (see Discussion). With the exception of question 18, scores from the two additional scorers were used only for interrater reliability measurements and are not included in other data analysis.

Each subject’s total score for the assessment was calculated along with a set of subscores focusing on particular aspects of the assessment. The scorer used each citation and URL together to identify the corresponding source, making it possible to identify 100 percent of the sources selected by subjects. Each source was then scored separately for its quality and for its relevance to the assessment prompt. Peer-reviewed journal articles received

maximum scores for quality, whereas personal or commercial websites received the minimum score (see appendix B). Maximum relevance scores were given for articles that clearly addressed the topic of the prompt. Some sources received high scores for quality and low scores for relevance or vice versa. The citation was scored for completeness of information and appropriate format. Scores for source quality, source relevance, and citation quality were added to generate a sources subscore that provided a measure of how well the subject completed the task given in the prompt. Scores for questions 15 through 21 were added to generate a nontask subscore for awareness of peer-reviewed literature that was not directly dependent on the sources selected by the subject.

Multivariate analysis of variance (MANOVA) was carried out in SAS²⁰ to test the effects of student level (in other words, academic year), assignment version, and level by version interaction on the total score and subscores. Univariate two-way analysis of variance was then run for each subscore. *P*-values less than 0.05 were considered to be significant. Unless otherwise stated, scores are listed as means ± S.D.

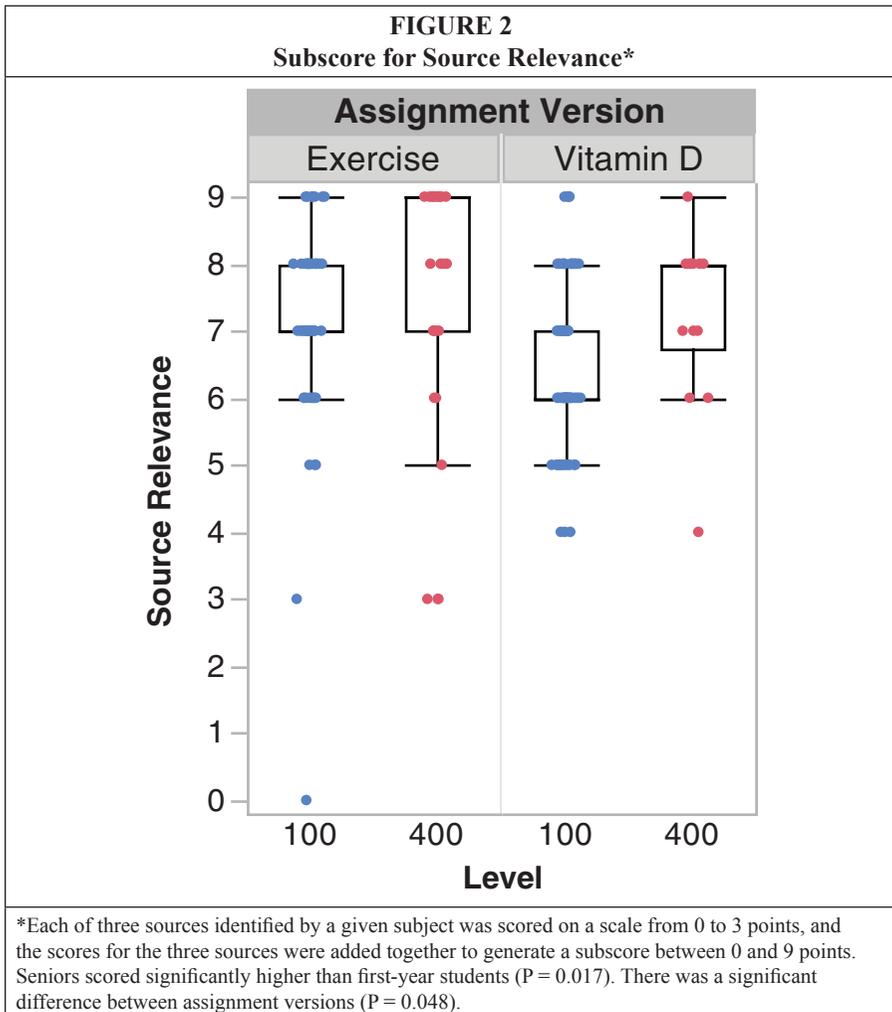


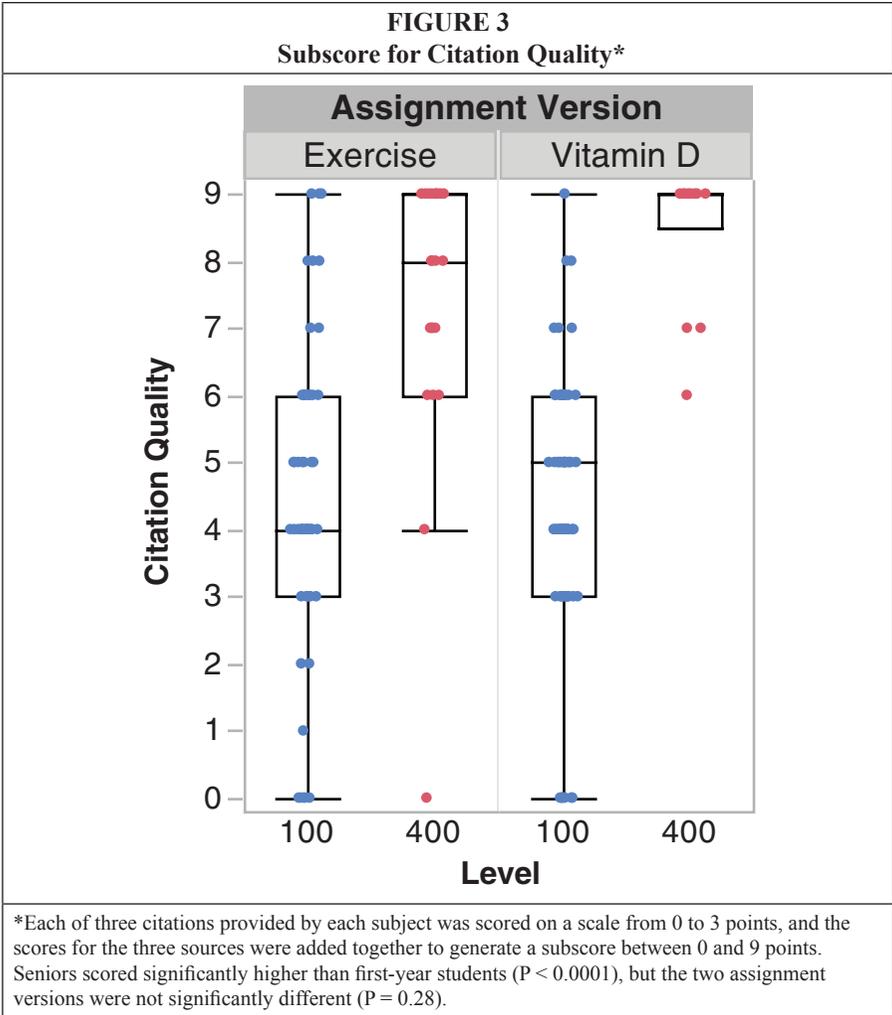
Results

Sixty-five per cent of first-year students and 61 percent of seniors in participating class sections completed the assessments. Results of the MANOVA indicated significant effects of student level (Wilks' $\lambda = 0.54$, $F = 38.36$, $P < 0.0001$) and assessment version (Wilks' $\lambda = 0.95$, $F = 2.55$, $P = 0.041$). However, there was no interaction between student level and assessment version (Wilks' $\lambda = 0.98$, $F = 0.99$, $P = 0.412$).

Scores for the quality of each of the three sources found by each subject were added together to produce that subject's subscore for source quality. Seniors found sources of significantly higher quality than first-year students (8.2 ± 1.0 vs. 4.4 ± 2.4 , $P < 0.0001$; see figure 1). While some first-years found high-quality sources, 68 percent of first-year students scored below 6 out of 9 points. All seniors scored at least 6 points, and 58 percent of seniors scored the maximum of 9 points. A score of 9 points indicates that all three selected sources were peer-reviewed journal articles. Source quality did not differ significantly between the two assessment versions ($P = 0.18$).

Scores for the relevance of each of the three sources found by each subject were added together to produce that subject's subscore for source relevance. Sources iden-





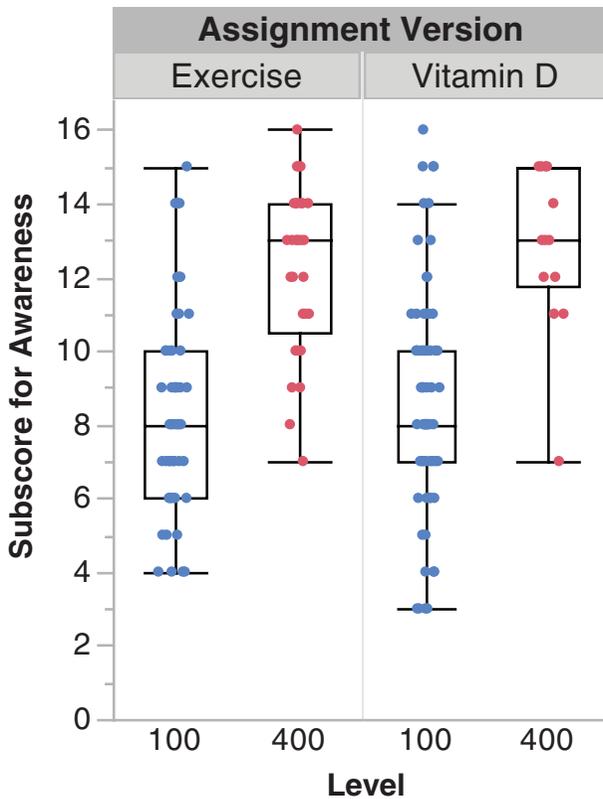
tified by senior students were rated as moderately, but significantly, more relevant to the scientific topic posed in the prompts (7.5 ± 1.8 vs. 6.8 ± 1.4 , $P = 0.017$; see figure 2). In general, students of both class levels were able to find relevant sources, with 86 percent of first-year students and 88 percent of seniors scoring at least 6 out of 9 points. Relevance scores were 11 percent higher for the exercise version of the assessment than the vitamin D version (7.3 ± 1.7 vs. 6.6 ± 1.2 , $P = 0.048$).

Seniors also produced significantly better citations than first-year students (7.7 ± 1.8 vs. 4.4 ± 2.0 , $P < 0.0001$; see figure 3); however, differences between the two versions were not significant ($P = 0.28$). Whereas 72 percent of first-years scored below 6 out of 9 points and 9 percent scored 0 points, 95 percent of seniors scored at least 6 points, and 49 percent of seniors scored the maximum of 9 points, indicating the use of a proper format for all three citations. Scores for source quality, source relevance, and citation quality were added to generate the source subscore. Seniors scored significantly higher than first-year students (23.5 ± 3.0 vs. 15.5 ± 4.6 , $P < 0.0001$), and there was no significant difference between assessment versions ($P = 0.40$).

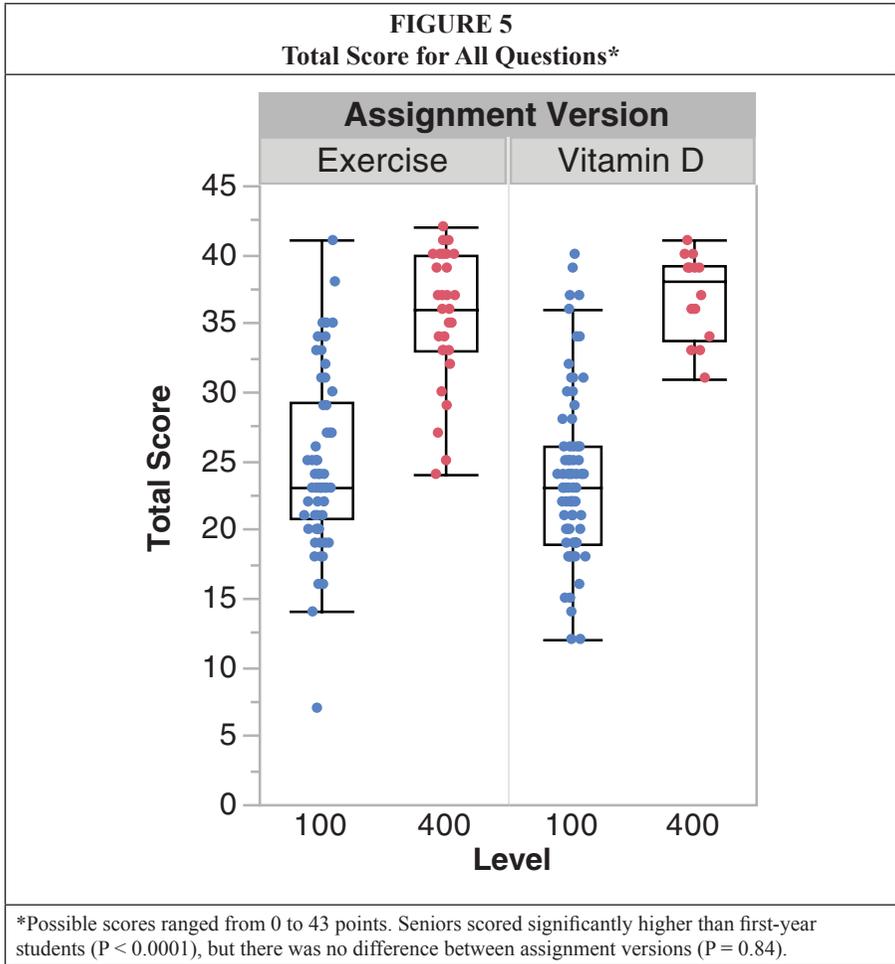
After finding sources, subjects answered a series of questions about the sources they selected and scientific literature in general (questions 15 through 21). Seniors earned significantly higher scores than first-year students on these questions (12.4 ± 2.2 vs. 8.5 ± 2.8 , $P < 0.0001$; see figure 4). There was no difference in scores of students completing the two assessment versions ($P = 0.40$), for which questions 15 through 21 were identical. Total scores for all scored questions were 49 percent higher for seniors than first-year students (35.9 ± 4.4 vs. 24.1 ± 6.1 , $P < 0.0001$; see figure 5), while scores for the two assessment versions did not differ ($P = 0.84$).

In addition to the questions that were scored numerically, subjects answered a series of unscored questions intended to evaluate search behavior and prior training in literature searching. Whereas 81 percent of first-year students found their articles using Google, only 2 percent of seniors used Google to find articles. Conversely, 74 percent of seniors used Google Scholar or an article database such as PubMed, Web of Science, or ScienceDirect, as compared to 13 percent of first-years (see figure 6). Among first-year students, those who started their searches with Google earned significantly lower scores for source quality than those who began with a specialized scientific database (3.8 ± 2.0 vs. 7.7 ± 1.9 , $P < 0.0001$). First-year and senior students attempted similar numbers

FIGURE 4
Subscore for General Awareness of Peer-Reviewed Literature*



*Each subject's scores for questions 15 through 21 were added together to generate a subscore between 0 and 16 points. Seniors scored significantly higher than first-year students ($P < 0.0001$), but there was no difference between assignment versions ($P = 0.40$).

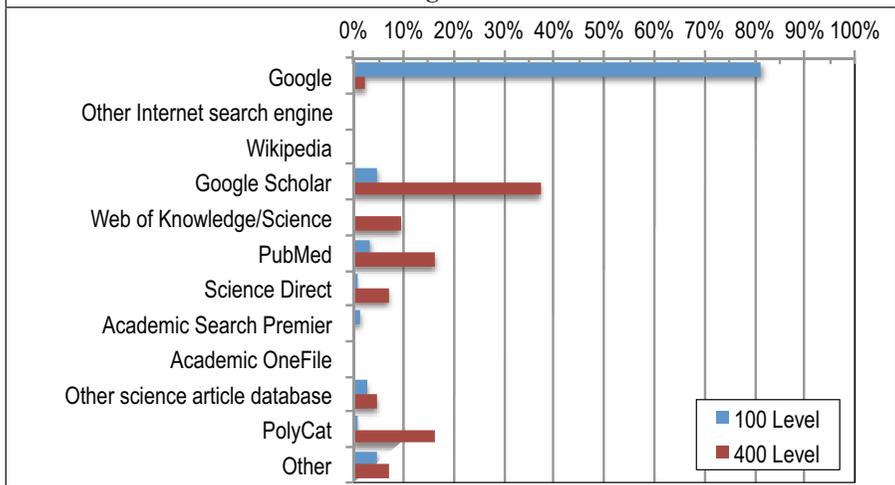


of search strings, with 44 percent of first-years and 40 percent of seniors using only a single combination of search terms. In contrast, 21 percent of first-years and 30 percent of seniors used 3 or more search strings.

First-year and senior students indicated different levels of prior training in finding peer-reviewed articles, with 67 percent of first-years indicating no prior training as compared to only 9 percent of seniors. A total of 84 percent of seniors reported prior instruction in a Cal Poly course, while 29 percent of first-years indicated that they had received training in a high-school course (see figure 7).

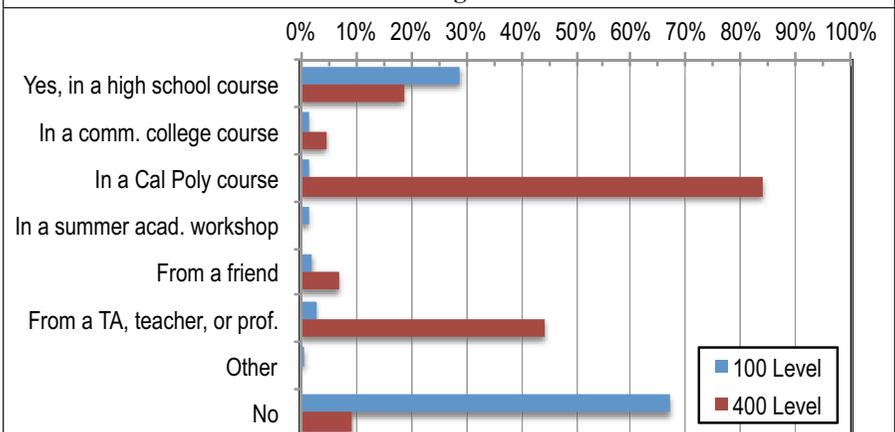
Interrater reliability was high (Cronbach's alpha > 0.85) for subscores associated with the primary task of finding and citing articles (see table 2) and three of the additional questions. However, reliability was lower (Cronbach's alpha < 0.7) for responses to three of the additional questions (those were questions 15, 18, and 19). These three questions addressed criteria that subjects used to decide whether an article would be useful, identification of primary vs. secondary literature, and the definition of a peer-reviewed article.

FIGURE 6
Databases and Search Engines Used to Locate Articles*



*Bars indicate the percentage of students selecting each method in answer to the question, "Where did you locate the articles?" Each subject could choose only one response.

FIGURE 7
Prior Instruction in Finding Peer-Reviewed Articles*



*Subjects responded to the question, "Have you had previous instruction on finding peer-reviewed articles?" by choosing from a menu. Bars indicate the percentage of subjects selecting each option.

Discussion

The aim of this study was to develop a malleable assessment instrument for quantitative evaluation of science literature search performance and apply it to first-year and senior undergraduate biology majors. A number of studies have described subsets of information literacy skills in particular undergraduate populations, but few have compared skills between students of different academic standing (that is to say, year of study), and the study populations, methods, and results of these studies have varied widely. Several studies have indicated that upper-division students either improved on tests of basic library skills,²¹ used greater numbers of specialized databases and

peer-reviewed sources,²² or had greater confidence in their library skills²³ compared to lower-division undergraduates. To our knowledge, however, there are no published studies comparing performance of science literature searching tasks undertaken by beginning and advanced science students. A number of papers have described the development of assessment instruments that might be appropriate for cross-sectional or longitudinal comparisons of information literacy, but these papers have not included results of longitudinal assessment of students or cross-sectional comparison across student levels.²⁴

Our assessment tool is intended to fill two gaps in the assessment of scientific information literacy skills. It is amenable to quantitative, longitudinal, or cross-sectional comparison of student populations, and it serves as a “performance assessment” (as described in Oakleaf, 2008)²⁵ that can elicit natural academic information-searching behavior from subjects without prompting the use of specific search strategies or tools. To find appropriate sources in response to the prompt, subjects were compelled to make an active and independent choice to use appropriate scientific databases, then recall or investigate how to access them and search them, and finally evaluate search results to choose appropriate sources. Thus, the assessment requires the integration of information competencies in a realistic search process. Likewise, the requirement that subjects cite sources using “an appropriate format” was intended to measure the ability of subjects to understand and apply the essential features of a scientific citation rather than simply looking up and following a specified format. Because the assessment did not require the use of a specific search process or database, some students without prior training in the use of scientific databases or citations may have been able to succeed in the assigned task by careful and logical exploration.

The tool described in this study does not attempt a deep assessment of how well subjects understand the scientific resources that they select. The subjects in our study selected approximately 320 unique resources ranging from primary research articles to popular health web pages. Both the number and variety of resources selected would complicate attempts to ask detailed questions about methods, results, or conclusions presented within the resources to assess understanding at a deep level. Such assessment would be more straightforward if students were assigned to read particular articles that the study personnel had also read in detail. Because that approach would be incompatible with the open-ended search task employed in the study, the searching assignment could be paired with a separate assignment focusing on comprehension of specified scientific articles. It is conceivable that subjects could have scored well on the rubric used in this study without reading the articles in depth. However, it is also conceivable that, at the undergraduate level, the ability to recognize a relevant resource based on the title and abstract requires a certain level of information literacy in and of itself.

In this cross-sectional comparison, seniors clearly outperformed first-year students in finding relevant, peer-reviewed scientific journal articles, citing the articles appropriately and answering general questions about the nature of scientific literature. The large differences in scores between seniors and first-year students suggest that most

Variable	Cronbach's Alpha
Source Quality	0.959
Source Relevance	0.851
Citation Quality	0.887
Question 15	0.607
Question 16	0.853
Question 17	0.897
Question 18	0.684
Question 19	0.633
Question 20	0.817
Question 21	0.863

* Based on 3 independent scorers

senior participants had received effective basic training in finding and evaluating scientific literature. This skill is explicitly included as a program learning outcome of the bachelor of science degree in Biological Sciences at our institution: "Students will demonstrate proficiency in searching, reading, and evaluating the scientific literature."²⁶ The senior seminar class is an important capstone experience supporting this objective. Notably, seniors in this study completed the assessment within the first two weeks of the senior seminar class, prior to the explicit bibliographic instruction and practice in literature searching that are included in that course. This timing suggests that seniors had already acquired scientific literacy skills through other science coursework or other experiences prior to taking the senior seminar class. Corroborating this inference, 84 percent of seniors reported that they had received previous instruction on finding peer-reviewed articles in a Cal Poly course. It is likely that seniors would earn higher (or more consistent) scores if they completed the assessment after taking the senior seminar course. Thus, the improvement in performance measured in the present study may be a conservative estimate of gains in science information literacy from matriculation to graduation.

Since the present study involves cross-sectional comparison of first-year vs. senior performance rather than longitudinal assessment of individual students, certain caveats are in order. While all first-year biology majors enroll in BIO 160 in fall quarter, seniors have two options for the required senior capstone experience. All of the data in this study came from students enrolled in BIO 461, a seminar course in which students write and present a research proposal but do not conduct the proposed experiments. For logistical reasons, we were unable to recruit students enrolled in the alternative capstone course, BIO 462, which is an independent study course in which students carry out field or laboratory research under the supervision of a faculty member. It is possible that students who opt for the independent, hands-on project differ systematically in awareness of scientific literature from those who choose to take the proposal writing course. If the two classes attract distinct populations, the available data may not fully represent the entire cohort of senior biology majors. Because completion of a hands-on senior project requires specific lab or field skills, many faculty encourage students to complete one or more quarters of supervised research (in other words, special problems courses, BIO 200 and BIO 400) before registering for BIO 462. Nearly three-quarters (73%) of students enrolled in BIO 462 in 2013–2014 had completed at least one quarter of a "special problems" course (mean of 2.05 quarters per student), whereas only 24 percent of BIO 461 students had taken such a course (mean of 0.48 quarters per student) before taking BIO 461. If these independent study experiences provided exposure to relevant primary journal articles in addition to hands-on research, BIO 462 students may have gained a greater awareness of scientific literature than their counterparts who took BIO 461. If so, the observed differences in performance between first-year and senior students might tend to underestimate rather than overestimate changes in capability of the entire cohort of students.

Within participating class sections, response rates of first-year and senior students were similar (65% vs. 61%). Response rates appeared to vary between course sections; students in some course sections may have chosen not to complete the assignment because their instructors offered extra credit points for completion instead of requiring it for normal course credit. This decision was left to the discretion of the instructors.

Any comparison of first-quarter students with seniors may be subject to survivorship bias, as some first-year students leave the major and/or the university. Students who complete the assessment as seniors finishing the biology major may be more interested in science or more capable academically than those who change majors or discontinue enrollment. The most recent institutional data indicate that 78 percent

of Cal Poly first-years entering in fall 2008 graduated within 6 years.²⁷ If the lowest-scoring 22 percent of first-years, based on total score, are excluded from the data set, significant differences between first-years and seniors are still observed for total score, source quality, citation quality and general awareness of literature (ANOVA, all $P < 0.05$). Thus, survivorship bias is unlikely to play a major role in the patterns observed in this study, though it is an important consideration for assessment efforts in general.

To address potential sources of bias in this cross-sectional comparison, we intend to assess the current first-year subjects again when they are seniors in 2016–2017. This repeat assessment will allow longitudinal comparison of individual performance. In anticipation of conducting a longitudinal, crossover assessment, we gave two versions of the assessment prompt; each first-year student who completed the assessment reported in this study will complete the alternate version as a senior. One concern for this experimental design is that the alternate versions might not be equivalent in difficulty. Relevance scores were significantly lower for the “vitamin D” version of the prompt than the “exercise” version ($P = 0.048$), and it is likely that other subscores differ slightly between the two versions despite P -values greater than 0.05. However, it is reassuring that such differences did not preclude the identification of significant differences in score between first-year and senior students. The fact that large groups of pseudorandomly selected students achieved fairly similar scores on the two versions indicates that the versions are sufficiently similar in difficulty to provide effective assessment of improvement in student skills over time. Furthermore, the crossover design built into the study provides an inherent control for small differences in difficulty between versions.

To maximize scoring consistency, a single blinded scorer generated all scores used in the primary analysis. For evaluation of the rubric, two additional scorers rescored the responses from 30 subjects (see table 2). The three scorers had varied training, including one nonscientist graduate student in library and information sciences (KLT), one biology MS student, and one biology faculty member (JMB). Good interrater reliability (IRR) is indicated by Cronbach’s alpha scores > 0.85 for the core subscores for source quality, source relevance, and citation quality that relate directly to the literature-searching task assigned to the subjects.²⁸ Cronbach’s alpha scores fell below 0.7 for three of the seven questions testing general awareness of primary literature. The answers to two of these questions (15 and 19) involved open-ended descriptions; thus their scoring was inherently subjective. Subjects typed their responses to questions 15 and 19 in text boxes using an Internet browser. Therefore, it was difficult to determine whether short, vague answers from many subjects indicated limited understanding or merely reflected a tendency to reply briefly in the online format. The low IRR for question 18 (“Is this article a primary or secondary literature article?”) is in part an artifact of the compressed scoring range of zero or one point for this question. It also revealed discrepancies between the library graduate student and the biologists in identification of primary and secondary journal articles. For question 18 only, all responses were rescored by the biology faculty member, and divergent scores were replaced prior to further data analysis. In the future, consultation among scorers to adjudicate questionable scores or use of a calibration set of sample responses might improve scoring reliability substantially.

One criterion used in developing prompts for the assessment was the ability to discriminate between novice and expert search behaviors. Prompt topics were deliberately selected so that a general Internet search would turn up some misleading, low-quality sources and would be unlikely to lead subjects directly to relevant peer-reviewed literature. However, the topics allowed subjects who searched a scholarly database or Google Scholar using appropriate search parameters to find numerous

resources of varying relevance. Results for Google search strings that include technical terminology often include a prominent link to “Scholarly articles for (*the search string*),” so it was important to avoid technical terminology in the prompt. A prompt topic on the evolutionary relationship of birds and reptiles may have suffered from this drawback during a 2012–2013 pilot test of the assessment model. Thereafter, the authors performed a series of searches on terms related to each potential topic before choosing effective prompts. Changes in search-engine algorithms and functions, scientific databases, and the published literature may necessitate continued revision of the assignment prompts and tasks to maintain an appropriate level of difficulty for both novice and expert searchers in the future.

Timeliness of the topic is another consideration. A scientific topic that is prominently featured in the media may increase student interest in the assignment, but a new or trendy topic may complicate long-term assessment of subjects if there are rapid changes in the number of available scientific resources related to the topic. One potential hurdle in accessing subscription-based databases is the requirement that students be on campus or logged in to the university library system to authenticate access. At our institution, 95 percent of first-year students live on campus, while nearly all seniors live off campus. Thus, seniors may have faced the minor additional challenge of logging in to the system if they completed the assessment at home.

While the present study measures the improvement of one institution’s undergraduate biological sciences majors at finding, selecting, and citing relevant peer-reviewed journal articles, its broader aim is to demonstrate a generalizable model for quantitative assessment of differences in information-seeking behavior between groups or over time. Merely altering the prompt topics could make the assessment useful for students in a broad range of fields at undergraduate or graduate/professional school levels. To adapt the assessment for K–12 students or adult populations with varying levels of education, question phrasing and/or scoring rubrics could be altered to deemphasize peer-reviewed primary and secondary articles and allow for a broader range of valid resources. For instance, an assessment of a community health education initiative might ask subjects to search for “trustworthy” sources to answer a health question and then give top scores to those who found hospital or government websites.²⁹ In its present form, the quantitative assessment model can enable programmatic evaluation of information literacy training modules or other curricular interventions, with repeated assessment over various time scales. More generally, we expect that multiple variations of the general assessment approach presented here may be developed for examining information searching skills of a variety of student populations in the sciences and other disciplines.

Acknowledgements

We thank Cameron Anvari for assistance in copy editing the manuscript and wording and formatting assignments in Moodle. Ben Ruttenberg, John Perrine, Lisa Needles, Christina Vasquez, Marie Yeung, Chris Kitts, Pat Fidopiastis, and Gita Kolloru assisted with subject recruitment. Michael Black participated in discussions of experimental design, and Greg Scott assisted with data analysis and interpretation. Rachel Wilson assisted in rescoring responses for determination of interrater reliability. We are grateful to Rosemary Klein for helpful comments on the manuscript. Funding was provided to JMS by the Cal Poly Extramural Funding Initiative.

Appendix A. Assessment Text

Prompt 1a: The ability of exercise to produce or assist in weight loss in humans is controversial. Imagine that you are trying to understand the evidence for the role of exercise in weight loss in order to give a short presentation to your classmates. Locate 3 relevant, electronically available, peer-reviewed articles that address this topic. Please keep note of all the search terms and sources you use in conducting your research.

Prompt 1b: The health benefits of vitamin D supplements have been a matter of debate in the last decade. Imagine that you are trying to understand whether vitamin D supplements are beneficial in order to give a short presentation to your classmates. Locate 3 relevant, electronically available, peer-reviewed articles that address this topic. Please keep note of all the search terms and sources you use in conducting your research.

(Note that the following questions are the same for the two versions—only the topic of the prompt differs.)

* Subjects answer by choosing one or more answers from a set of predetermined options

Subjects answer by entering text in a text box

2. Provide a complete citation for one of the three articles using an appropriate format.

3. Paste the URL of the article that you just cited.

(Questions 4–7 repeat questions #2 and #3 for the remaining two articles found)

* 8. Where did you start your search for these articles? (please select one answer)

9. If other, please specify:

* 10. Where did you locate the articles? (please select all that apply)

11. If other, please specify:

12. Which search terms were successful?

* 13. How many combinations of search terms did you try before finding your three articles?

Please select one of the three articles you found and answer the following questions.

14. Please provide the citation of the article for which you are answering the questions.

15. What criteria did you use to decide that this article would be helpful in your research?

16. Please describe in your own words how this article is relevant to the topic.

17. How did you determine that this was a peer-reviewed article?

* 18. Is this article a primary or secondary literature article?

19. Please describe in your own words what a peer-reviewed (scholarly) article is.

20. Please describe in your own words what a primary literature article is.

21. Please describe in your own words what a secondary literature article is.

* 22. Have you had previous instruction on what a peer-reviewed article is? (select all that apply)

* 23. Have you had previous instruction on finding peer-reviewed articles? (select all that apply)

* 24. Have you had any previous library instruction at Cal Poly?

25. If you answered yes to the previous question, please do your best to list the class(es) or department(s) that provided the instruction.

(Additional demographic questions are omitted here.)

Appendix B. Scoring Rubric

1. Informed consent. Nothing to score.

2. Provide a complete citation for one of the three articles using an appropriate format. (0 to 3 pts. per article)

Citations must include the following for full credit: Authors (all or first et al.), Year, Article title, Journal name, Volume, Page numbers (or article number or DOI for e-journals). The overall format should match any acceptable format used in the biological sciences. Authors should remain in the order given on paper or formatting points are lost.

(3) Full citation in appropriate format.

(2) Citation is missing an item or format is mildly incorrect.

(1) Citation is missing multiple items or format has severe deficiencies.

(0) Citation is insufficient to identify the item or bears little resemblance to a properly formatted citation.

3. Paste URL of article

Relevance (0 to 3 points per article)

(3) Obviously relevant and useful for the proposed assignment

(2) Maybe relevant and useful in combination with other sources

(1) Questionable or tangential relevance and utility (for example: may share a few keywords, but isn't really on topic)

(0) No apparent relevance

Source type (0 to 3 points per article)

(3) Peer-reviewed primary or secondary literature

(2) Other scholarly material (such as book chapter, government report, dissertation) or reliable, citable source (such as government, museum, or other official website, conference proceeding)

(1) Other article type (such as major newspaper or periodical with high journalistic standards)

(0) Unreliable/invalid source (such as miscellaneous commercial or personal website)

(Questions 4–7 repeat questions 2–3 for two additional articles, and are scored identically.)

8, 9. Where did you start your search for these articles?

No score for this question. Answers are interesting from an instructional perspective.

10, 11. Where did you start your search for these articles?

No score for this question. Answers are interesting from an instructional perspective.

12, 13. Which search terms were successful?

No score for this question. Answers are interesting from an instructional perspective.

14. Provide citation.

No score for this question. Repeats previous question to frame subsequent answers.

15. What criteria did you use to decide that this article would be helpful in your research? (0 to 3 points)

(3) Criteria addressing specific aspects of both the content of the article and the quality of the source

- (2) Criteria addressing both the content of the article and the quality of the source, but only generally
- (2) Criteria addressing specific aspects of either the article content or the source quality, but not both
- (1) Criteria addressing either the article content or source quality in a general way
- (0) No criteria stated or don't know

16. Describe in your own words how this article is relevant to the topic. (0 to 3 points)

- (3) Multiple, specific details of the article are related to the original topic and the findings are mentioned.
- (2) Specific details of the article are related to the original topic.
- (1) The topic of the article is stated in general terms.
- (0) The relevance is unclear from the answer.

17. How did you determine that this was a peer-reviewed article? (0 to 3 points)

- (3) Logical reasoning based on source of article or database
- (2) Presence of citations
- (2) Evidence for editing or revision process
- (1) Identity of authors
- (1) Unreliable reasoning based on source of article or database
- (0) Source was not peer-reviewed
- (0) Unsure or don't know

18. Is this article a primary or secondary literature article? (0 to 1 point)

1 point for correct selection

19. Please describe in your own words what a peer-reviewed (scholarly) article is.

- (0 to 2 points)
- (2) The answer clearly states that the article has been reviewed by other scholars prior to acceptance for publication.
- (1) The answer vaguely mentions something about review by others, validity of the content, or credentials of authors.
- (0) Unsure or don't know

20. Please describe in your own words what a primary literature article is. (0 to 2 points)

- (2) The answer clearly explains that primary research involves presentation of original experimental or observational data by the researcher who conducted the experiment.
- (1) The answer says something about presentation of data or presentation by the researcher, but is unclear or includes misconceptions.
- (0) Unsure or don't know

21. Please describe in your own words what a secondary literature article is. (0 to 2 points)

- (2) The answer clearly explains that secondary literature involves review and synthesis of previously published data or primary literature.
- (1) The answer says something about review or summary, but is vague or unclear or includes misconceptions.
- (0) Unsure or don't know

22–30. Demographic questions and feedback on survey—no score.

Notes

1. Association of College and Research Libraries (ACRL), "Information Literacy Standards for Science & Engineering/Technology," 2006, available online at www.ala.org/acrl/standards/infolitsci [accessed 20 February 2014].
2. Heidi Julien and Susan Barker, "How High-School Students Find and Evaluate Scientific Information: A Basis for Information Literacy Skills Development," *Library and Information Science Research* 31, no. 1 (Jan. 2009): 12–17; Amber Walraven, Saskia Brand-Gruwel, and Henry P.A. Boshuizen, "How Students Evaluate Information and Sources When Searching the World Wide Web for Information," *Computers and Education* 52, no. 1 (Jan. 2009): 234–46.
3. Eleanor M. Smith, "Developing an Information Skills Curriculum for the Sciences," *Issues in Science & Technology Librarianship* 37 (Spring 2003), doi:10.5062/F48P5XGT; Jeanine M. Scaramozzino, "Integrating STEM Information Competencies into an Undergraduate Curriculum," *Journal of Library Administration* 50, no. 4 (Apr. 2010): 315–33.
4. Robin N. Sinn, "Library Instruction for Biology Courses: A Literature Review and Survey," *Research Strategies* 16, no. 2 (Jan. 1998): 103–15.
5. Andrew Walsh, "Information Literacy Assessment: Where Do We Start?" *Journal of Librarianship and Information Science* 41, no. 1 (Feb. 2009): 19–28.
6. Association of College and Research Libraries (ACRL), "Information Literacy Standards for Science & Engineering/Technology"; Accreditation Council for Graduate Medical Education "ACGME Common Program Requirements," (2013): 8, available online at www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRS2013.pdf [accessed 31 July 2014]; American Chemical Society, "Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs," (2008): 14, available online at www.acs.org/content/dam/acsorg/about/governance/committees/training/acsapproved/degreetprogram/2008-acgme-guidelines-for-bachelors-degree-programs.pdf [accessed 31 July 2014]; California State Board of Education, "Next Generation Science Standards for California Public Schools, Kindergarten through Grade Twelve," (2013): 5, available online at www.cde.ca.gov/pd/ca/sc/documents/ngss-ca-gr9-12-dci.doc [accessed 31 July 2014].
7. Cecilia Brown and Lee R. Krumholz, "Integrating Information Literacy into the Science Curriculum," *College & Research Libraries* 63, no. 2 (Mar. 2002):111–23; Shubhik K. DebBurman, "Learning How Scientists Work: Experiential Research Projects to Promote Cell Biology Learning and Scientific Process Skills," *CBE Life Sciences Education* 1, no. 4 (Winter 2002):154–72; Molly R. Flaspohler, Erika M. Rux, and John A. Flaspohler, "The Annotated Bibliography and Citation Behavior: Enhancing Student Scholarship in an Undergraduate Biology Course," *CBE Life Sciences Education* 6, no. 4 (June 2007):350–60.
8. Nancy Goebel, Paul Neff, and Angie Mandeville, "Assessment within the Augustana Model of Undergraduate Discipline-Specific Information Literacy Credit Courses," *Public Services Quarterly* 3, no. 1/2 (2007): 165–89.
9. Robin N. Sinn, "A Comparison of Library Instruction Content by Biology Faculty and Librarians," *Research Strategies* 17, no. 1 (Spring 2000): 23–34; Jason R. Porter, "Information Literacy in Biology Education: An Example from an Advanced Cell Biology Course," *CBE Life Sciences Education* 4, no. 4 (Winter 2005): 335–43.
10. Gloria J. Leckie and Anne Fullerton, "Information Literacy in Science and Engineering Undergraduate Education: Faculty Attitudes and Pedagogical Practices," *College & Research Libraries* 60, no. 1 (Jan. 1999): 9–29; Steven J. Bell, "Stop IAKT Syndrome with Student Live Search Demos," *Reference Services Review* 35, no. 1 (2007): 98–108; Brian Winterman, "Building Better Biology Undergraduates through Information Literacy Integration," *Issues in Science and Technology Librarianship* 58 (Summer 2009), doi:10.5062/F4736NT6; Aditi Bandyopadhyay, "Measuring the Disparities between Biology Undergraduates' Perceptions and Their Actual Knowledge of Scientific Literature with Clickers," *Journal of Academic Librarianship* 39, no. 2 (Nov. 2013): 194–201.
11. Jessame E. Ferguson, Teresa Y. Neely, and Kathryn Sullivan, "A Baseline Information Literacy Assessment of Biology Students," *Reference & User Services Quarterly* 46, no. 2 (Winter 2006): 61–71; Ignacio J. Ferrer-Vinent and Christy A. Carello, "Embedded Library Instruction in a First-Year Biology Laboratory Course," *Science and Technology Libraries* 28, no. 4 (Oct. 2008): 325–51; Angela Dunnington and Mary L. Strong, "The Rocky Road to Assessment: The Creation of an Assessment Tool for an Information Literacy Credit Course," *Codex: The Journal of the Louisiana Chapter of the ACRL* 1, no. 2 (Aug. 2010): 53–79; Jon R. Hufford, "What Are They Learning? Pre- and Post-Assessment Surveys for LIBR 1100, Introduction to Library Research," *College & Research Libraries* 71, no. 2 (Mar. 2010): 139–58; Jason A. Porter, Kevin C. Wolbach, Catherine B. Purzycki, Leslie A. Bowman, Eva Agbada, and Alison M. Mostrom, "Integration of Information and Scientific Literacy: Promoting Literacy in Undergraduates," *CBE Life Sciences Education* 9, no.

4 (Winter 2010): 536–42; Ignacio J. Ferrer-Vinent and Christy A. Carello, “The Lasting Value of an Embedded, First-Year, Biology Library Instruction Program,” *Science and Technology Libraries* 30, no. 3 (Sept. 2011): 254–56.

12. Lynn Cameron, Steven L. Wise, and Susan M. Lottridge, “The Development and Validation of the Information Literacy Test,” *College & Research Libraries* 68, no. 3 (May 2007): 229–37; Irvin R. Katz, “Testing Information Literacy in Digital Environments: ETS’s iSkills Assessment,” *Information Technology and Libraries* 26, no. 3 (2007): 3–12.

13. Cara Gormally, Peggy Brickman, and Mary Lutz, “Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates’ Evaluation of Scientific Information and Arguments,” *CBE Life Sciences Education* 11, no. 4 (Winter 2012): 364–67.

14. Davida Scharf, Norbert Elliot, Heather A. Huey, Vladimir Briller, and Kamal Joshi, “Direct Assessment of Information Literacy Using Writing Portfolios,” *Journal of Academic Librarianship* 33, no. 4 (July 2007): 462–77; Stacey Knight-Davis and Jan S. Sung, “Analysis of Citations in Undergraduate Papers 1,” *College & Research Libraries* 69, no. 5 (Sept. 2008): 447–58.

15. Rui Wang, “The Lasting Impact of a Library Credit Course,” *Libraries and the Academy* 6, no. 1 (Jan. 2006): 79–92; Julie K. Gilbert, Katherine Knutson, and Christopher P. Gilbert, “Adding an Integrated Library Component to an Undergraduate Research Methods Course,” *PS: Political Science & Politics* 45, no. 1 (Jan. 2012): 112–18.

16. Flaspohler, Rux, and Flaspohler, “The Annotated Bibliography and Citation Behavior”; Kathleen M. Gehring and Deborah A. Eastman, “Information Fluency for Undergraduate Biology Majors: Applications of Inquiry-Based Learning in a Developmental Biology Course,” *CBE Life Sciences Education* 7, no. 1 (Spring 2008): 54–63.

17. Gehring and Eastman, “Information Fluency for Undergraduate Biology Majors.”

18. Scharf et al., “Direct Assessment of Information Literacy”; Knight-Davis and Sung, “Analysis of Citations in Undergraduate Papers 1.”

19. ACRL, “Information Literacy Standards.”

20. SAS Institute Inc, “SAS 9.1.3 Help and Documentation,” Cary, N.C.: SAS Institute Inc., 2002–2004, available online at <http://support.sas.com/onlinedoc/913/docMainpage.jsp> [accessed 27 May 2014].

21. Cameron, Wise, and Lottridge, “The Development and Validation of the Information Literacy Test.”

22. Madeleine C. Shanahan, “Transforming Information Search and Evaluation Practices of Undergraduate Students,” *International Journal of Medical Informatics* 77, no. 8 (Aug. 2008): 518–26; Sue Samson, “Information Literacy Learning Outcomes and Student Success,” *Journal of Academic Librarianship* 36, no. 3 (May 2010): 202–10; Terry Judd and Gregor Kennedy, “Expediency-Based Practice? Medical Students’ Reliance on Google and Wikipedia for Biomedical Inquiries,” *British Journal of Educational Technology* 42, no. 2 (Mar. 2011): 351–60.

23. Arlene Greer, Lee Weston, and Mary L. Alm, “Assessment of Learning Outcomes: A Measure of Progress in Library Literacy,” *College & Research Libraries* 52, no. 6 (Nov. 1991): 549–57.

24. Lisa G. O’Connor, Carolyn J. Radcliff, and Julie A. Gedeon, “Applying Systems Design and Item Response Theory to the Problem of Measuring Information Literacy Skills,” *College & Research Libraries* 63, no. 6 (Nov. 2002): 528–43; Katz, “Testing Information Literacy in Digital Environments”; Gormally, Brickman, and Lutz, “TOSLS.”

25. Megan Oakleaf, “Dangers and Opportunities: A Conceptual Map of Information Literacy Assessment Approaches,” *Libraries and the Academy* 8, no. 3 (July 2008): 233–53.

26. California Polytechnic State University, “2015–2017 Catalog,” available online at catalog.calpoly.edu/collegesandprograms/collegeofsciencemathematics/biologicalsciences/bsbiologicalsciences [accessed 7 December 2014].

27. California Polytechnic State University, “Cal Poly Retention and Graduation Reports,” available online at www.ir.calpoly.edu/content/publications_reports/ret_grad/index [accessed 24 July 2015].

28. Jum C. Nunnally and Ira H. Bernstein, *Psychometric Theory*, 3rd ed. (New York: McGraw-Hill, 1994).

29. Rosalie Van der Vaart, Alexander J.A.M. van Deursen, Constance H.C. Drossaert, Erik Taal, Jan A.M.G. van Dijk, and Mart A.F.J. van de Laar, “Does the eHealth Literacy Scale (eHEALS) Measure What It Intends to Measure? Validation of a Dutch Version of the eHEALS in Two Adult Populations,” *Journal of Medical Internet Research* 13, no. 4 (Fall 2011): e86.