Perceived Challenges in Primary Literature in a Master's Class: Effects of Experience and Instruction

Richard Lie,[†‡§] **Christopher Abdullah,**[†∥] **Wenliang He,**[¶] **and Ella Tour**[#*]

‡Department of Neurosciences, ∥Biomedical Sciences Graduate Program, and #Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; ¶School of Education, University of California, Irvine, Irvine, CA 92697

ABSTRACT

Primary literature offers rich opportunities to teach students how to "think like a scientist," but the challenges students face when they attempt to read research articles are not well understood. Here, we present an analysis of what master's students perceive as the most challenging aspects of engaging with primary literature. We examined 69 pairs of pre- and postcourse responses from students enrolled in a master's-level course that offered a structured analysis of primary literature. On the basis of these responses, we identified six categories of challenges. Before instruction, "techniques" and "experimental data" were the most frequently identified categories of challenges. The majority of difficulties students perceived in the primary literature corresponded to Bloom's lower-order cognitive skills. After instruction, "conclusions" were identified as the most difficult aspect of primary literature, and the frequency of challenges that corresponded to higher-order cognitive skills increased significantly among students who reported less experience with primary literature. These changes are consistent with a more competent perception of the primary literature, in which these students increasingly focus on challenges requiring critical thinking. Students' difficulties identified here can inform the design of instructional approaches aimed to teach students how to critically read scientific papers.

INTRODUCTION

Primary literature provides educators with an excellent opportunity to train students how to "think like a scientist": analyze data, draw conclusions, and identify follow-up directions to a scientific investigation. In their report *Scientific Foundations for Future Physicians*, the American Association of Medical Colleges and Howard Hughes Medical (AAMC-HHMI) Institute listed the ability of critical reading and evaluation of scientific papers as one of the competencies that students should possess before entering medical school (AAMC-HHMI, 2009, p. 26). The *Vision and Change* report by the American Association for the Advancement of Science (AAAS) highlighted C.R.E.A.T.E. (Consider, Read, Elucidate hypotheses, Analyze and interpret data, and Think of the next Experiment), a structured approach in which students analyze four papers from the same lab (Hoskins *et al.*, 2007), as an important way to introduce scientific research in nonlab courses (AAAS, 2011, pp. 32–33).

Many studies have described the use of primary literature in the classroom (e.g., Janick-Buckner, 1997; Muench, 2000; Kozeracki *et al.*, 2006; Hoskins *et al.*, 2007; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Sato *et al.*, 2014; Van Lacum *et al.*, 2014; Abdullah *et al.*, 2015). Analysis of primary literature has frequently been associated with gains in multiple aspects of students' academic self-efficacy, such as being able to understand scientific articles, science literacy, and their sense of belonging to the scientific community (Kozeracki *et al.*, 2006; Hoskins *et al.*, 2011; Gottesman and Hoskins, 2013; Van Lacum *et al.*, 2014; Abdullah *et al.*, 2015).

Some studies have also reported the positive effects of the use of primary literature on students' science literacy skills (Krontiris-Litowitz, 2013), ability to critically read research articles (Van Lacum *et al.*, 2014), and science process/critical-thinking skills (Hoskins *et al.*, 2007; Gottesman and Hoskins, 2013; Segura-Totten and Dalman, 2013; Sato *et al.*, 2014; Abdullah *et al.*, 2015).

The benefits of reading primary literature notwithstanding, instructors have detected that students struggle with some of its aspects (Gehring and Eastman, 2008; Krontiris-Litowitz, 2013; Van Lacum *et al.*, 2014). A study by Krontiris-Litowitz (2013) demonstrated that students enrolled in an introductory biology course that focused on development of scientific literacy skills had difficulties connecting the goal of investigation to the previous studies and understanding the study's hypotheses and the purpose of using a certain method to test the stated hypothesis. Data interpretation was another challenge (Krontiris-Litowitz, 2013), and using the data to justify the authors' conclusions presented a difficulty to both lower-division and upper-division biology undergraduates (Gehring and Eastman, 2008; Krontiris-Litowitz, 2013). Van Lacum and colleagues reported that 44–88% of students enrolled in a freshmen-level life sciences course could not correctly identify such elements of scientific articles as the motivation for a study, its main conclusion, and its implications (Van Lacum *et al.*, 2014). Gehring and Eastman reported that upper-division biology students had difficulties integrating an article's data with other published results (Gehring and Eastman, 2008). Among other recognized barriers to the primary literature are the unique characteristics of scientific texts: their high informational content, discipline-specific jargon, and the formal writing style that differs from textbooks or informal texts that students often find online (Gehring and Eastman, 2008; Snow, 2010). Together, these studies have revealed what instructors find their students struggling with as they begin to read the primary literature. However, students' perspectives on the challenges they face when reading scientific articles have not been systematically studied. Understanding these challenges is essential for developing effective teaching methods that help students to engage with the primary literature.

METHODS
Participants
This study focused on students enrolled in the contiguous BS/MS program of the Division of Biological Sciences in a large public university classified as RU/VH (very high research activity) in Southern California. Students enter this program during their senior year as undergraduates, complete at least six quarters of research (typically, three quarters as undergraduates, followed by at least three quarters of graduate research in the same lab), and defend a research-based thesis.

Data on students' demographics, majors, and prior experiences with primary literature were collected over four quarters from 69 students (representing 86% of students enrolled in this course during these quarters) via anonymous pre- and post-course surveys (Figure 1 and Supplemental Figure S1, B and C). The majority of students participating in this study were in their first three quarters of the master's program (88% of the students; Supplemental Figure S1A).
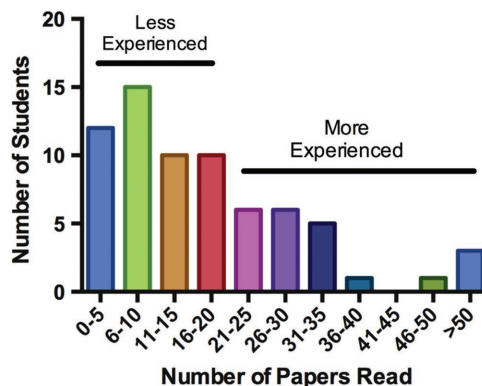

FIGURE 1. Self-reported previous exposure to primary literature, based on an anonymous survey completed before instruction. Students were asked "Approximately how many research papers have you carefully read and analyzed during your undergraduate and graduate studies?" and the response choices are shown on the X axis. $n = 69$ students. An arbitrary cutoff of 20 papers read was used to subdivide the students into "less experienced" and "more experienced" groups.

Additionally, we were interested in our students' previous experience with scientific papers and the types of instruction they had received on how to read research papers before taking this course. Only 26% of students had been explicitly taught how to read scientific articles. The distribution of self-reported number of papers read before taking this class is shown in Figure 1 ($n = 69$ students). Previous exposure to scientific literature varied among the students. More than two-thirds of the students ($n = 47$) had read 20 or fewer papers before taking this course, while the rest ($n = 22$) reported having read more than 20 articles. These two groups will be referred to as "less experienced" and "more experienced" in reading primary literature, respectively (Figure 1).

Demographic differences between the less experienced and the more experienced students were examined using Fisher's exact test (Supplemental Table S1). The two groups did not differ in terms of how long they had been in the master's program, major, ethnicity, grade point average (GPA), or gender (the data for GPA and gender were only available for the Fall 2013 and Winter 2014 quarters; Supplemental Table S1). We detected a positive but not significant relationship between belonging to the more experienced group and reporting having been taught a strategy for reading scientific papers and being more likely to participate in a journal club (Supplemental Table S1 and Supplemental Figure S1C).

Course Structure
A detailed description of this course is provided elsewhere (Abdullah *et al.*, 2015). In brief, the course was designed for master's students with the goal of providing training in critical analysis of primary literature. Rather than being focused on a specific topic, the papers discussed in this course were taken from different fields in biology, with the aim of providing students with universal skills that can be implemented in reading a paper on any topic in biology. Four articles were used: a paper that contained flaws in experimental design and data interpretation, an exemplary paper, and a pair of conflicting papers that

examined the same biological phenomenon but reached opposing conclusions. The papers are listed in the Supplemental Material, Appendix B. In selecting a flawed paper (paper 1), the course instructor (E.T.) was looking for a study in which detection of the flaws in experimental design and data interpretation did not require expertise in the topic of the paper or in the experimental procedures used in the paper. Similarly, we think that in selecting the appropriate pair of conflicting papers, it is best to select a pair in which determination of which paper presents a more convincing argument should not require specialized technical knowledge. In this sense, one of the pairs (papers 3 and 4 used in Winter 2014) was preferable to the other pair (used in Fall 2013), which required a more specialized knowledge to determine which was more convincing. Finding suitable papers for this course was serendipitous for some of the papers (papers 1 and 2 used in Fall 2013; papers 3 and 4 used in Winter 2014). The rest of the papers were identified by University of California, San Diego (UCSD), faculty after the instructor (E.T.) emailed them a brief description of the course and a request for flawed, exemplary, or conflicting papers.

Examination of each paper included five meetings consisting of the following activities performed by the students: 1) identification of unfamiliar methodology, terminology, and background; 2) presentation of unfamiliar methodology, terminology, and background; 3) critical interpretation of experimental data; 4) discussion of authors' conclusions and identification of unanswered research questions; and 5) design and presentation of a follow-up experiment and critique of experiments proposed by others. Students were graded on all of the activities described.

Surveys

Anonymous surveys were administered online via SurveyMonkey (www.surveymonkey.com), during the first and last weeks of a 10-week quarter. In addition to the question about the most challenging aspects of the primary literature, the survey included questions about students' self-efficacy in a number of skills related to the primary literature (Abdullah *et al.*, 2015;

Supplemental Material, Appendix C). The precourse (but not the postcourse) survey also included questions about students' demographics, majors, and previous exposure to primary literature (Supplemental Material, Appendix C). To allow matching between the pre and post surveys while preserving the anonymity of responses, we asked students to provide the same five-digit number in both surveys. The students received a small number of course points for completing the surveys. Overall, 69 students (86% of students enrolled in this course) completed both pre and post surveys in the Fall 2012, Winter 2013, Fall 2013, and Winter 2014 quarters.

Qualitative Analysis of Students' Surveys

Using the anonymous surveys, three raters examined 69 pairs of pre- and postcourse free responses to the question "What aspects of understanding and analyzing scientific papers do you find most challenging?" The raters (E.T., a faculty member, and R.L. and C.A., both biology graduate students) were blind to both the students' identity and the pre- or postcourse status of the responses. Our analysis was aligned with grounded theory: we examined students' responses, seeking to describe all challenges identified by the students, as opposed to approaching the data with preconceived hypotheses about the nature of these challenges (Corbin and Strauss, 2008; Andrews *et al.*, 2012). Finally, we wanted to determine whether the nature or the frequency of challenges identified by the students changed after instruction and, if so, what hypotheses could be produced to explain this phenomenon. The three raters discussed each response until a consensus on the challenges described in the response was reached. Table 1 presents examples of students' responses and the challenges identified in these responses. Many students chose to write about more than one challenge in reading and analyzing scientific papers (Table 1). The average number of challenges per response was 1.6, with a range of zero to four challenges per response. Of the responses from pre- and postcourse surveys, 19 and 17%, respectively, contained statements that could not be interpreted unambiguously,

TABLE 1. Examples of the analysis of students' responses to the question: "What aspects of understanding and analyzing scientific papers do you find most challenging?"

Student's responses[a]	Difficulty 1 (assigned Bloom's level)[b]	Difficulty 2 (assigned Bloom's level)[b]	Difficulty 3 (assigned Bloom's level)[b]
Being able to follow up an experiment[1]	Designing follow-up experiment (synthesis)		
Terminology, terms describing techniques,[1] most-widely known and used methods in the field but not known to others outside of the field.[2] Figures and images that scientists of other fields would not know immediately.[3]	Unfamiliar terminology (knowledge)	Techniques/methods (knowledge)	Data presentation (knowledge)
Interpreting figures[1] and understanding the techniques used.[2]	Experimental data (analysis)	Techniques/methods (comprehension)	
Understanding the language.[1] It is hard to grasp a paper when you are unfamiliar with the techniques used[2] and what their results mean.[3]	Scientific language/ Writing style (comprehension)	Techniques/methods (knowledge)	Experimental data (comprehension)
Figuring out what new techniques are[1] or what specific terminology mean.[2]	Techniques/methods (comprehension)	Unfamiliar terminology (knowledge)	
Papers on subjects I don't find fulfilling.[1]	Lack of interest/motivation (no Bloom's level assigned)		

[a]Representative responses and individual challenges (superscript) identified in each response. Parts of the responses that were critical in identifying each challenge are underlined.

[b]Challenges identified in each of the responses and their assigned Bloom's level.

and such statements were coded as "vague." The determination of categories of challenges was carried out after all challenges were identified. Differences between pre- and postinstruction survey results were determined using the McNemar's chi-square test, followed by the Benjamini-Hochberg false discovery rate correction for multiple comparisons. The numbers of challenges observed in each of the quarters included in this study are shown in Supplemental Figure S2.

The implied Bloom's level of the challenges identified by the students was determined by the same three raters. The raters were blind to the pre/post identity of the responses. Identification of the Bloom's cognitive levels of the challenges was based on the published descriptions of the cognitive activities associated with each of the Bloom's levels (Bloom *et al.*, 1956; Anderson *et al.*, 2001; Crowe *et al.*, 2008). A detailed description of the rating process can be found in Appendix A in the Supplemental Material. Briefly, the verbs students used to describe difficulties were very important in determining the Bloom's level of the difficulties. Table 2 presents the verbs frequently used in students' responses and their assigned Bloom's levels. However, the context in which the verb was used in students' responses was important as well, as demonstrated in the students' responses rated at application level in Table 2. In identifying the Bloom's levels of challenges related to the experimental data category, we differentiated between student responses that described difficulties with *understanding* experimental data (coded as comprehension) and *interpreting* or *analyzing* experimental data (coded as analysis). Our rationale for making this distinction is provided in Appendix A in the Supplemental Material.

The raters used students' responses from the quarters not included in this study as a training data set. During the rating of the data set included in this study, each rater separately determined the Bloom's levels of the challenges, and the interrater reliability was calculated using Fleiss's kappa statistic. Initial interrater reliability for determining Bloom's levels of

the challenges was 0.65 (substantial agreement), while the interrater reliability for lower-order cognitive skills (LOCS) versus higher-order cognitive skills (HOCS) rating of challenges was 0.74 (substantial agreement as defined in Landis and Koch, 1977). The three raters then discussed the disagreements between the Bloom's level assignments. The aim of these discussions was not to reach a consensus on the rating of each response but rather to articulate the reasoning of each rater for selecting a particular Bloom's level. This at times resulted in raters changing their initial Bloom's level assignment. After the discussion, the interrater reliability value rose to 0.89 (substantial agreement as defined in Landis and Koch, 1977). A Bloom's level was assigned to a challenge if at least two of the three raters rated the challenge at that level. Some challenges were determined to have no associated Bloom's level (e.g., lack of motivation or interest in the topic). The number of challenges assigned to each of the Bloom's levels in each of the quarters included in this study is shown in Supplemental Figure S3. Statistical analysis of the pre- and postcourse changes in the Bloom's levels of the challenges was performed using a paired t test, followed by the Benjamini-Hochberg false discovery rate correction for multiple comparisons.

Institutional Review Board

Protocols used in this study were approved by the UCSD Human Research Protections Program (project 111351SX).

RESULTS

What Do the Students Perceive as the Most Difficult Aspects of Primary Literature?

To guide us in the future design of instructional approaches aimed to teach students how to critically read and analyze scientific papers, we wished to understand what barriers our students face while reading primary literature. From our analysis of students' responses to the question "What aspects of understanding and analyzing scientific papers do you find most challenging?," we derived a comprehensive list of challenges (Table 3). We also determined the frequency of each of the challenges before and after instruction (i.e., the percentage of students who identified a particular challenge; Table 3). We then identified overarching categories that combined related challenges (indicated in bold in Table 3). Our students perceived six major categories of difficulties in the primary literature:

1. Background and terminology: challenges related to unfamiliar biological background, concepts, and terms
2. Techniques: challenges related to experimental methods and techniques
3. Experimental data: challenges related to understanding and analyzing a paper's data
4. Conclusions: challenges related to understanding conclusions, drawing independent conclusions, and evaluating authors' conclusions
5. Paper as a whole: challenges related to the overall structure of the study; for example, connection between the hypothesis and experiments, connection between individual experiments
6. Generic attributes of a scientific paper: difficulties that did not relate to the specific content of a scientific paper but to generic aspects of primary literature, such as scientific language and high density of unfamiliar information

TABLE 2. Typical verbs or expressions from students' responses used for identifying the Bloom's level of the challenges with the primary literature

Bloom's level	Typical verbs or expressions used by the students
Knowledge	Unfamiliar (being unfamiliar with)
Comprehension	Understand
	Being confused by
Application	"Understanding how small changes in methods might affect outcomes"[a]
Analysis	Analyze
	Draw conclusions
	Interpret
Synthesis	Design (think of, synthesize) a follow-up experiment
Evaluation	Evaluate
	Assess validity
	Critique
	Determine importance

[a]Note that although the student is using the verb "understanding" in this response, the context of the response indicates that the student is trying to *predict an outcome*, which is an application-level activity.

TABLE 3. Most difficult aspects of reading and analyzing scientific papers

Categories of challenges[a] Challenges comprising each category	Representative quotes from students' responses[b] *"Most difficult aspect of understanding and analyzing scientific papers is …"*	Percentage of responses before instruction[c] (*n* = 69)	Percentage of responses after instruction[c] (*n* = 69)
Background and terminology			
Unfamiliar background	"Reading papers outside my field of expertise (don't have the necessary background)"	15	13
	"Background information if I am unfamiliar with the field"		
Unfamiliar terminology	"Understanding terminology"	16	6
	"The mass of technical jargon"		
Techniques			
Techniques/methods	"Understanding the experimental methods used to produce figures (I feel that they are often not well defined in the text of the paper)."	35	23
	"Not fully understanding methods can make interpreting figures difficult as well." (Note that this response was also coded as including "experimental data" challenge)		
Limited by lack of technical expertise	"I think trying to understand the specific[s] of different methods and trying to figure out what kinds of errors can occur and what effects they might have on experimental data is difficult for me but extremely important."	3	4
Experimental data			
Experimental data	"I have difficulties in understanding the figures."	26	13
	"Analyzing all of the data and experiments"		
	"Thinking critically about figures"		
	"Probably the data interpretation"		
Data presentation	"Sometimes the analysis of the graphs or data is shown in a different way than I am used to, and may take extra time to decipher the way that the paper presents the data."	7	1
Statistics	"The statistics"	3	1
Evaluating quality of data	"Determining the validity/reliability of data"	1	4
Results (not clear whether the reference is to the data or to the *Results* section)	"Results"	0	1
Conclusions			
Drawing your own conclusions	"Drawing my own conclusions based on the data presented"	1	15
	"Coming to my own conclusions instead of blindly agreeing with the authors"		
Evaluating author's conclusions	"Determining how credible an author's conclusion is from the data they present"	3	13
	"Probably still assessing whether the conclusions are valid"		
Understanding conclusions	"Understanding how they [the authors] were able to draw a conclusion from the … experiment"	1	1
Discussion	"The discussion"	1	1
Paper as a whole			
Broad experimental design of the paper	"Connections with the hypothesis are vague."	7	6
	"Usually understanding how all experiments fit into the paper as a whole"		
Design of a follow-up experiment	"Thinking up additional/follow-up experiments"	0	6
Relationship between experiments in the paper	"Since many scientific papers will touch on several questions/conclusions, sometimes I have a hard time understanding … how they all come together."	4	1
Relating the paper's findings to the broader field	"Determining if the relatively logical conclusions reached in the paper are congruous with what is already known and what other scientists are also discovering concurrently"	1	3
Main ideas of the paper	"I… lose sight of the main idea of the paper."	3	1
Impact of the paper	"You don't particularly know enough to know … whether the results will impact the field in any way."	3	3
Assumptions	"The most challenging for me is trying think outside the paper about … potential problems with assumptions the paper makes. Basically, I tend to accept what the paper proposes without asking too many questions."	3	0

(Continued)

TABLE 3. Continued

Categories of challenges[a] Challenges comprising each category	Representative quotes from students' responses[b] *"Most difficult aspect of understanding and analyzing scientific papers is …"*	Percentage of responses before instruction[c] (*n* = 69)	Percentage of responses after instruction[c] (*n* = 69)
Generic attributes of a scientific paper			
Scientific writing style	"Deciphering the scientific language to get to the main points." "The vocabulary and sentence structure"	13	0
Excessive time required	"Timing, I struggle with reading the papers quickly and efficiently."	6	1
Getting lost in the details	"I get lost in the details."	4	0
Lack of motivation or interest in the topic	"The greatest challenge is finding the motivation to spend time and energy to read, and interpret the meaning of the paper."	1	4
The first attempt at reading	"I still have some trouble dealing with difficult concepts on a first read."	0	3
Amount of unfamiliar information	"I find it most difficult when there are large portions of text that only contain information that I am unfamiliar with. In these cases, I get overwhelmed and have trouble grasping onto any information at all."	1	0
Organization (format) of a scientific paper	"I am easily distracted when I have to flip between different pages with corresponding text, images, and supplementary endnotes."	1	0
No difficulties		0	3
Vague		19	14

[a]Challenges identified by the students, grouped into broad categories (bold).
[b]Students' responses representative of each type of challenge (in the interest of brevity, some of the quotes include only the most relevant parts of students' response).
[c]Percentage of responses that included each challenge, before and after instruction, respectively. Each response could contain more than one challenge; therefore percentages add up to more than 100%.

We also wanted to examine whether students' prior experiences with reading primary literature impacted the types and frequencies of challenges they perceived. Since, to our knowledge, no studies have been done on the relationship between the number of scientific papers read and development of expertise in reading primary literature, we decided to use 20 papers as an arbitrary threshold, with students who reported reading 20 or fewer papers before taking this course (*n* = 47 students) considered "less experienced" and students who reported reading more than 20 papers (*n* = 22 students) considered "more experienced."

Before instruction, "techniques" was one of the two most frequent categories of challenges, present in 38% of students' responses (Figure 2A). Students' responses that included this challenge often described lack of familiarity with the techniques used in the paper or difficulties understanding the techniques (Table 3). More than one-third of these students (13%) linked their lack of familiarity with techniques to their difficulty in understanding or interpreting the data (such responses were also coded as including an "experimental data" challenge):

"New techniques you have not yet encountered so it is difficult to interpret the data."

"Not fully understanding methods can make interpreting figures difficult as well."

Techniques presented a major challenge to the less experienced and to the more experienced students alike (Figure 2, B and C). After instruction, we detected no statistically significant decrease in the frequency of this category, with more than a quarter of the students still identifying techniques as

one of the most challenging aspects of reading primary literature (Figure 2, A–C).

"Experimental data" was the other most frequently identified category of challenge (Figure 2A). Responses in this category described difficulties ranging from understanding the data to data analysis and evaluation (Table 3). After instruction, the frequency of student responses that identified "experimental data" as the most challenging aspect of scientific papers decreased by 42%, but this decrease was not statistically significant. Interestingly, the only type of challenge that increased in frequency in the "experimental data" category was "evaluating quality of data" (Table 3).

The third most frequent category of challenges before instruction was "background and terminology" (Figure 2A and Table 3). After instruction, the number of students who identified challenges from this category did not decrease significantly. Among more experienced students, the frequency of the challenges in "background and terminology" category exhibited a threefold decrease after instruction (however, this decrease was not significant, possibly due to the small sample size), while remaining a persistent challenge for less experienced students (Figure 2, B and C).

"Conclusions" was the only category of challenges that exhibited a significant increase in frequency after instruction (*p* = 0.02, Cramer's V = 0.38; Figure 2A). Before instruction, only 7% of the students identified any aspect of conclusions as a major challenge, while after instruction, almost one-third (30%) of the students did so, making this category the most frequent in the postinstruction responses (Figure 2A). The largest postinstructional increase within the category of "conclusions" was seen in "drawing your own conclusions from the data," followed by "evaluating author's conclusions" (Table 3). This effect was especially pronounced among the less
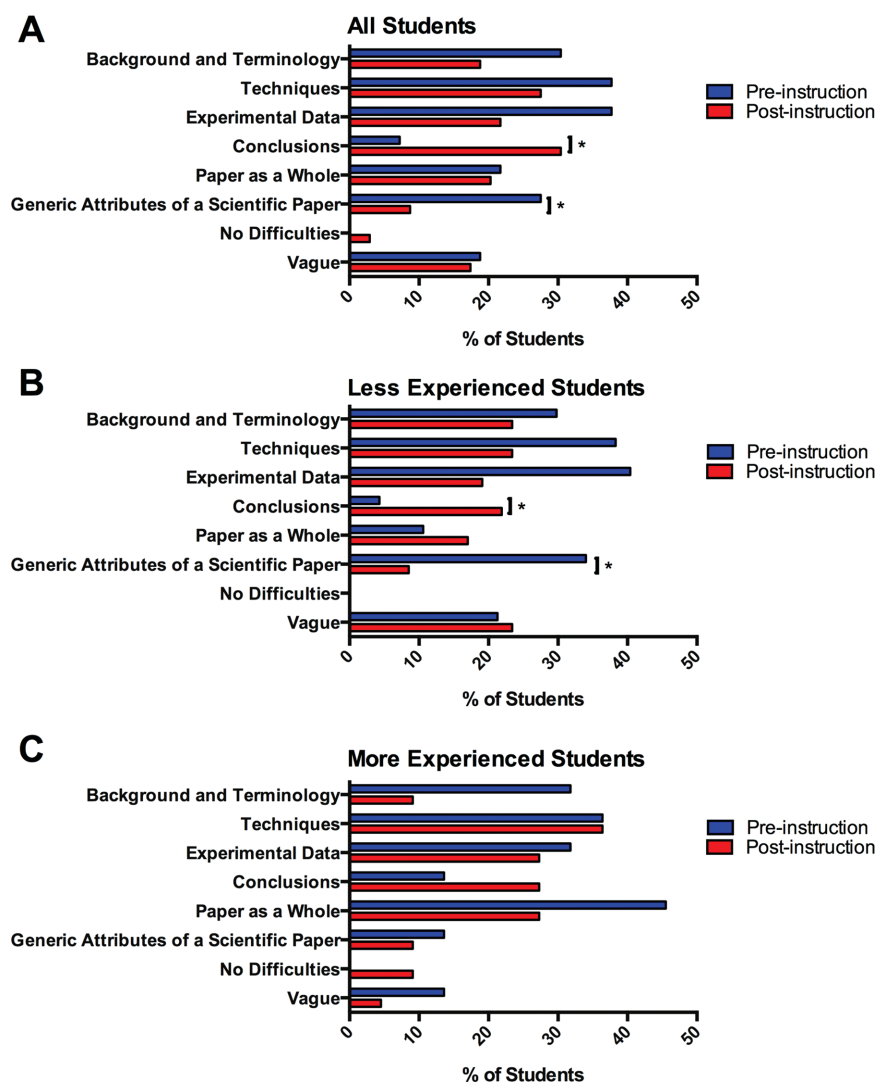
FIGURE 2. Perceived categories of challenges in the primary literature before and after instruction. (A) Frequency of each category of challenges in the pre- and postcourse responses of all students ($n = 69$ pairs of students' responses). (B) Frequency of the categories of challenges reported by less experienced readers ($n = 47$ paired responses). (C) Frequency of the categories of challenges reported by more experienced readers ($n = 22$ paired responses). *, $p < 0.05$.

students (Figure 2, B and C). Again, the small size of the more experienced group might have accounted for the lack of statistical significance in the change in this category of difficulty. A type of challenge that did exhibit an increase in the postsurvey was "designing a follow-up experiment" (from 1 to 6%; Table 3). This is not surprising, since students designed and presented follow-up experiments to the papers discussed in this course, which probably brought this aspect and its difficulty into students' awareness.

Surprisingly, more than a quarter of the students identified generic, surface features of scientific articles (e.g. "scientific writing style," "getting lost in details") or motivational aspects (lack of motivation or perceived excessive amount of time it takes to read an article) as a major difficulty in the primary literature (Table 3). Before instruction, these challenges were especially common among the less experienced students, who identified them 2.5 times more frequently than the more experienced readers (Figure 2, B and C), but this difference was not statistically significant. After instruction, the share of these difficulties decreased significantly among all students ($p = 0.032$, Cramer's V = 0.33; Figure 2A), with a fourfold drop among less experienced students ($p = 0.032$, Cramer's V = 0.33), bringing its frequency to the level observed in the more experienced group (Figure 2, B and C). The most frequent challenge in this category, "scientific writing style," decreased from 13 to 0% after instruction (Table 3).

Changes in Bloom's Cognitive Levels of Perceived Challenges Posed by Primary Literature

We also noticed a distinction in the level of cognitive skills implied in some of the challenges in students' responses. This prompted us to re-examine our data and, where appropriate, to identify the level of Bloom's taxonomy that corresponded to each challenge (Bloom *et al.*, 1956; Anderson *et al.*, 2001). Before instruction, the challenges corresponding to LOCS dominated the responses, comprising 71% of all challenges (Figure 3A). Comprehension-level challenges appeared more common among the less experienced students, while HOCS-level challenges appeared more frequently among the more experienced group (Figure 3, C and E); however, these differences were not statistically significant.

After instruction, we observed a significant overall decrease in challenges aligned with LOCS ($p = 0.001$, Cohen's $d = 0.63$; Figure 3A). This decrease was significant among the less experienced students ($p = 0.005$, Cohen's $d = 0.70$; Figure 3C), but not in the more experienced group. The overall decrease in

experienced group, where the frequency of this challenge increased more than sevenfold after instruction ($p = 0.02$, Cramer's V = 0.37; Figure 2B).

The category "paper as a whole" included challenges related to the connection between a paper's hypothesis and its experiments, the relationship between the experiments, and designing a follow-up experiment (Table 3). Before instruction, the more experienced group identified this challenge four times more frequently than the less experienced group (Figure 2, B and C), but this difference was not statistically significant. Overall, the number of responses that included this category did not change significantly between the pre- and postquarter surveys (Figure 2A); however, the two groups of students exhibited different trends: a sharp decline among the more experienced students and a small increase among the less experienced
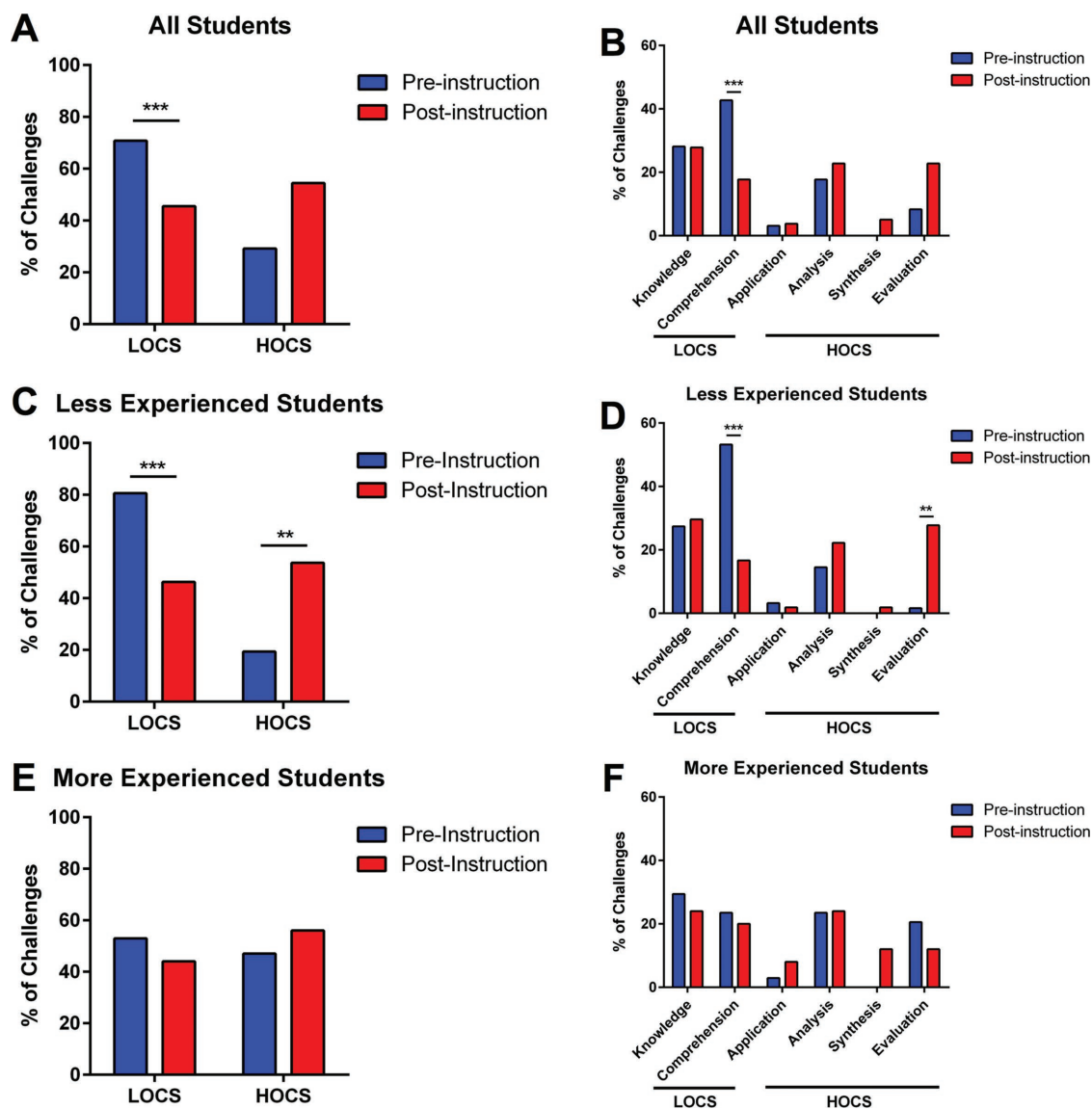
FIGURE 3. Bloom's levels of challenges students perceived in the primary literature. (A, C, and E) Frequency of challenges that corresponded to LOCS and HOCS, as reported by (A) all students, (C) less experienced readers, and (E) more experienced readers before and after instruction. (B, D, and F) Frequency of challenges corresponding to each Bloom's level, as identified by (B) all students, (D) less experienced readers, and (F) more experienced readers before and after instruction. *p* Values are reported as follows: **, *p* < 0.01; ***, *p* < 0.001. The overall change in the frequency of HOCS-level challenges (A) was not significant (*p* = 0.12, Cohen's *d* = 0.33).

LOCS-level challenges was largely driven by the drop in the frequency of challenges associated with comprehension, (*p* = 0.001, Cohen's *d* = 0.67; Figure 3B). This decrease was significant in the less experienced group (*p* = 0.001, Cohen's *d* = 0.83; Figure 3D), while no significant change was detected in the more experienced group (Figure 3F).

After instruction, the frequency of challenges that correspond to HOCS significantly increased among the less experienced students (*p* = 0.002, Cohen's *d* = 0.7; Figure 3C) but not among the more experienced group (Figure 3E). This increase in HOCS-level challenges in the less experienced group was driven primarily by the rise of evaluation-level challenges (*p* = 0.005, Cohen's *d* = 0.73; Figure 3D). Taken together, these data suggest that instruction provided in this course was associated with a significant decrease in the perceived LOCS-level challenges in the primary literature and an increased identification of HOCS-level challenges among students with less experience in reading primary literature.

DISCUSSION

Understanding the difficulties students perceive when they engage in critical reading and analysis of scientific papers is important for designing proper instructional approaches to help students in this task. To our knowledge, this study provides the first analysis of what very recent postgraduate

students perceive to be the most challenging aspects of primary literature.

Previous studies have presented the instructors' perspectives on students' difficulties with the primary literature (Gehring and Eastman, 2008; Krontiris-Litowitz, 2013; Van Lacum *et al.*, 2014). Interpretation of data presented in papers was one such difficulty (Krontiris-Litowitz, 2013), and our study shows that students also perceive understanding and interpreting the data and evaluating their quality as major challenges. Using data to draw conclusions or to justify authors' conclusions was another difficulty detected by the instructors (Gehring and Eastman, 2008; Krontiris-Litowitz, 2013). Very few of our students identified "conclusions" as a difficulty before instruction; however, it became the most frequently identified challenge after instruction.

Instructors have also found that students struggle connecting a paper's hypothesis with its experiments or placing a paper's findings in the broad context of its field (Gehring and Eastman, 2008; Krontiris-Litowitz, 2013). Our students also identified these challenges (category "paper as a whole"). More experienced students were more likely to identify this category as a challenge, implying that many less experienced students might not be even trying to make these connections and thus are unaware of these difficulties. Another previously recognized barrier is scientific language, due to its high informational density and its formal writing style (Samuels, 1983; Snow, 2010). Difficulties with scientific language were also identified by our students and were especially frequent among less experienced students, suggesting that experience helps students to overcome this barrier. Students' difficulty with terminology, or scientific jargon, has been long recognized by instructors (McDonnell *et al.*, 2016, and the references within), and our students also perceived it as a substantial challenge before instruction.

Some of the difficulties with primary literature we identify here have not been described previously. Experimental techniques were among the most frequent challenges, despite the fact that our students were enrolled in a research-based master's program. We predict that techniques are likely to present an even bigger challenge to undergraduate students with less laboratory research experience. The link that some of our students made between understanding methods and being able to interpret data indicates that, if we want our students to seriously engage with data in papers, students' difficulty with techniques needs to be addressed instructionally. Unfamiliar background was another challenge not previously documented in the literature.

Another novel finding of this study is that, before instruction, the majority of the challenges students perceived in the primary literature were LOCS-level challenges: difficulties that related to lack of knowledge or understanding. Although the instructor in this course (E.T.) wanted her students to focus on activities that involve HOCS (analyzing data, evaluating conclusions, thinking of follow-up experiments), the students were deterred by, and therefore focused on, LOCS-level challenges.

Effect of the Instruction on the Perceived Challenges

Structured instruction on how to read primary literature that was provided in this course was associated with significant changes in the frequency of two categories of challenges. The frequency of "generic attributes of scientific papers" significantly decreased, while challenges associated with "conclusions" significantly increased. This postinstructional increase in "conclusions" suggests that instruction helped our students become aware of the need for and the difficulty of critically evaluating authors' conclusions ("instead of blindly agreeing with the authors," as one student put it). We also detected changes in the cognitive level of students' difficulties. The frequency of the challenges corresponding to LOCS significantly decreased, while the frequency of challenges aligned with HOCS significantly increased among less experienced students. Together, these changes suggest a shift in less experienced students' perceptions of the primary literature toward a deeper, more critical thinking–oriented approach. The frequency of the remaining categories of challenges did not change significantly, indicating that these aspects of scientific articles require additional instructional support. Among such categories were "techniques," "experimental data," "unfamiliar background and terminology," and "paper as a whole."

Gradual Acquisition of Competency in Reading Primary Literature

The model of domain learning (Alexander *et al.*, 1997; Alexander and Jetton, 2000; Alexander, 2003) provides a useful perspective on the changes in students' perceptions of the primary literature. This model describes how text-based learning changes with acquisition of expertise and identifies three progressive stages in this process (Figure 4):

Stage 1. Acclimated learner: At this stage, the reader lacks the knowledge about the subject area of the text (such as scientific paper), has difficulty discerning between what is important and unimportant, and has mostly situational interest in the subject (such as a course requirement). When faced with a difficulty in the text, this reader employs primarily superficial strategies, such as rereading, focusing mainly on reading comprehension.

Stage 2. Competent learner: A more experienced reader, who has acquired some subject matter knowledge, has strategies to efficiently find information about unfamiliar topics and has more intrinsic interest in the subject. Less deterred by surface-level barriers (such as high informational content or the formal writing style of scientific language), a competent reader is concerned not only with comprehension, but also with analysis and evaluation of the information provided in the text.

Stage 3. Proficient learner: An expert reader, who is highly knowledgeable about the subject matter of the text and deeply interested in the subject. This proficient reader habitually employs deep strategic processing of the text: analyzing and evaluating new information, coming up with new models that integrate this information, and identifying new avenues of investigation (Figure 4).

We suggest that the process of acquisition of primary literature reading skills can be viewed as a continuum, with instruction and experience enabling readers to progress from acclimation to proficiency. In the context of our class, we do not expect our students to achieve proficiency in reading articles in just one quarter, but our data indicate that, at least in their perceptions of the challenges posed by scientific articles, our students
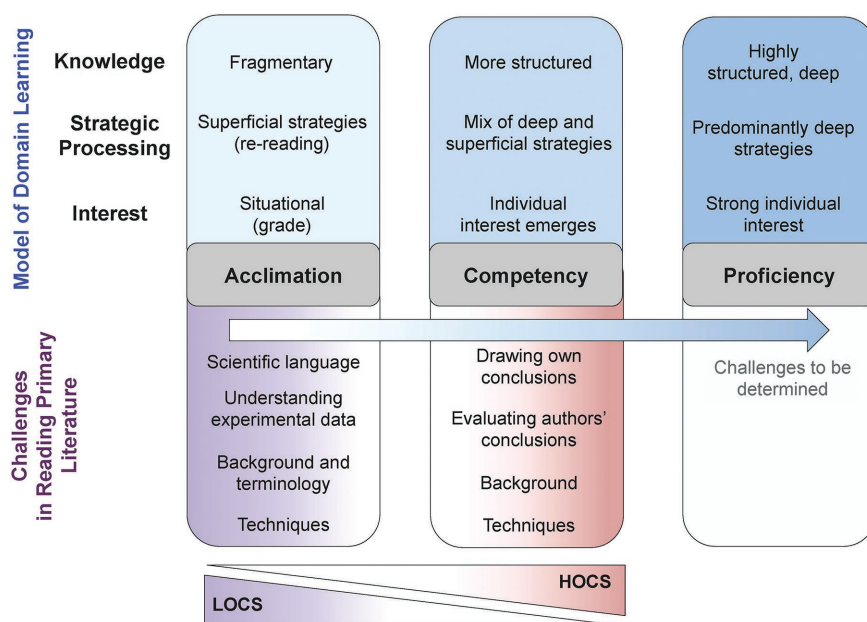
FIGURE 4. Students' difficulties with primary literature map to the acclimation and competency stages in the model of domain learning. Top, the model of domain learning, which describes how text-based learning changes as readers become more knowledgeable and skillful (Alexander *et al.*, 1997; Alexander, 2003). Three stages are identified in this process: acclimation, competency, and proficiency, with each stage characterized by the state of the reader's subject knowledge, strategic processing of the text by the reader, and the source of his or her interest and motivation (blue rectangles). Bottom, the challenges students identify in primary literature (purple-pink rectangles) are consistent with these students being in the acclimation or competency stages. The challenges perceived by the experts (proficiency stage) and the Bloom's levels of these challenges remain to be investigated. Previous exposure to scientific papers and instruction in reading primary literature help students progress along the acclimation–competency continuum (blue arrow).

are progressing toward competency. We argue that students who identified "generic attributes of a scientific paper" as a major challenge are likely to be in the acclimation stage with respect to reading primary literature (Figure 4). These readers are likely to be deterred by such surface-level barriers to reading comprehension as scientific language. Discrimination between important and unimportant concepts also presents a difficulty to readers in the acclimation stage (Alexander and Jetton, 2000), and challenges such as "getting lost in details" and "amount of unfamiliar information" reflect this difficulty. Conversely, more experienced readers were less deterred by the surface features of scientific articles. They also identified more HOCS-level challenges, although this difference was not statistically significant. These findings suggest that experience and previous instruction enabled these students to progress along the acclimation–competency continuum.

Postinstructional decrease in LOCS and increase in HOCS-level challenges, decrease in the difficulties with "generic attributes of a scientific paper," and increase in the frequency of difficulties with "conclusions" suggest that instruction in this course helped students to progress along the continuum toward competency in their perception of scientific articles (Figure 4). This effect was especially noticeable among less experienced students.

Limitations and Future Directions
One limitation of this study is that we describe what students *perceived* as the major difficulties in reading and analyzing primary literature. Our students likely had other difficulties with the primary literature, which they might not have been aware of or did not perceive as the *most* challenging. Furthermore, an increase in the awareness of HOCS-level challenges does not necessarily mean that the students are now performing better in these HOCS-level tasks. Previous studies have reported that undergraduate courses that offered structured analysis of primary literature were associated with significant increases in students' performance in questions assessing Bloom's HOCS (Hoskins *et al.*, 2007; Gottesman and Hoskins, 2013; Segura-Totten and Dalman, 2013). In a study conducted in the course we described here, we detected a significant increase in students' ability to design an experiment but not in students' ability to interpret data, draw conclusions, or evaluate a hypothesis based on the data, although our conclusions were limited by high pre-instructional performance of our students in this test (Abdullah *et al.*, 2015). A primary literature–based assessment that systematically evaluates students' skills in the various aspects of reading primary literature (i.e., identifying hypothesis, understanding experimental setup, interpreting data, placing the paper in its broader context) would be very valuable in providing an objective evaluation of students' skills and difficulties in critically reading scientific articles.

Some of the changes in students' perceptions of the challenges in primary literature could be due to factors other than the instruction provided in this course. Among such factors are students' own research, research-related literature they read, other courses they took, journal clubs, and lab meetings in which they participated during that quarter. Additionally, activities students completed in this class were likely to have an impact on what students perceived as most difficult in scientific literature: for example, while attempting to evaluate evidence from two conflicting papers or design a follow-up experiment for the first time, students might have realized that these activities are challenging.

Another limitation of this study is that, in our analysis of students' difficulties and their alignment with the Bloom's cognitive skills, we were limited by what students chose to include in their written responses. For example, it is possible that, for a master's student, "data interpretation" may mean a level of engagement with the data different from that for a more advanced graduate student, postdoc, or a faculty member. Interviews with students could provide additional insights into

students' perceptions about primary literature and the strategies they apply when reading it.

In this study, we used 20 papers as an arbitrary threshold to differentiate between the less experienced and the more experienced students. We acknowledge that such dichotomization might have resulted in a loss of information about the relationship between the number of papers read and the challenges perceived by our students. We also acknowledge that there is no reason to assume that a student who has read 21 papers has a substantially different experience from a student who has read 20 papers, and other factors (e.g., instruction they received on how to read papers, superficial vs. in-depth reading by the student) will also play a role in a student's experience. We suggest that future studies in this topic should treat the number of papers read as a continuous variable and use regression analysis to examine the relationship between students' experience with scientific papers and the challenges they perceive. Also, rather than selecting a range of numbers, as we did in this study, it would be preferable to ask the students to report the actual number of papers they have read. As in any study that assesses past behavior, a possibility of inaccurate recall of this number will remain a limitation, and students will be likely to recall more reliably the number of papers they read in the more recent past.

In this study, we assessed a population of students who are postgraduates, have demonstrated good academic achievement (undergraduate GPA of 3.0 or higher), and are engaged in active research. It will be of great interest to investigate the difficulties that undergraduate students face when approaching primary literature and how these perceptions change in the course of their studies. A sample that also includes PhD students, postdocs, and faculty can provide a better understanding of the progression from an acclimated to a competent and then to a proficient reader.

Implications for Instruction

Only 26% of the students who participated in this study reported receiving an explicit instruction on how to read scientific articles at any point of their undergraduate or graduate career. We think that such training is essential if we want to empower our students to critically engage with the primary literature. Our findings that students perceived unfamiliar techniques, background, terminology, and even scientific language as substantial challenges before instruction suggest that these difficulties will need to be explicitly addressed by instructors, both at the level of skills, by teaching students how to find information about unfamiliar concepts and techniques, and at the metacognitive level, by helping students to be aware of these difficulties, to realize that these difficulties are common and can be overcome. New online resources that can help address these challenges have recently become available. The *Journal of Visualized Experiments* (www.jove.com) provides videos of a wide variety of techniques in biology performed by experts in the field. Instructors could refer students to use these videos as a source outside class or incorporate the videos into the classroom instruction. Another tool is Science in the Classroom, a free collection of annotated articles from different fields of science, developed by a team of *Science* journal editors, which provides rich embedded descriptions of terminology, prior research background, and techniques

(AAAS, 2014; http://scienceintheclassroom.org/?tid=22). These papers could be used by instructors as a basis for introducing students to primary literature.

In addition to providing students with strategies to address the LOCS challenges, we suggest that instructors should also guide students toward engagement with the literature at a higher cognitive level: critical analysis and evaluation of a paper's data, experimental design, and conclusions and thinking about questions yet to be addressed. We suggest inclusion of papers with flaws that students can identify without relying on extensive knowledge or technical expertise. Such articles provide an opportunity for students to critically assess and evaluate the authors' claims and assertions. In addition, including papers from new, developing fields, in which many questions are still unanswered, can offer an opportunity for students to appreciate the dynamic, sometimes uncertain nature of scientific process and exercise their skills of synthesis by developing new or alternative hypotheses and proposing original follow-up experiments (Hoskins *et al.*, 2007; Abdullah *et al.*, 2015). Moving forward, it is important to develop effective teaching strategies that address the barriers that students encounter in their journey toward competence in engaging with primary literature.

REFERENCES

Abdullah C, Parris J, Lie R, Guzdar A, Tour E (2015). Critical analysis of primary literature in a master's-level class: effects on self-efficacy and science process skills. CBE Life Sci Educ 14, ar34.

Alexander PA (2003). The development of expertise: the journey from acclimation to proficiency. Educ Res 32(8), 10–14.

Alexander PA, Jetton TL (2000). Learning from text: a multidimensional and developmental perspective. In: Handbook of Reading Research, vol. 3, Abingdon, UK: Routledge, 285–310.

Alexander PA, Murphy PK, Woods BS, Duhon KA, Parker D (1997). College instruction and concomitant changes in students' knowledge, interest, and strategy use: a study of domain learning. Contemp Educ Psychol 22, 125–146.

American Association for the Advancement of Science (AAAS) (2011). Vision and Change in Undergraduate Biology Education: A Call to Action. http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf (accessed 11 September 2014).

AAAS (2014). Science in the Classroom, Washington, DC.

American Association of Medical Colleges and Howard Hughes Medical Institute (2009). Scientific Foundations for Future Physicians, Washington, DC/Chevy Chase, MD. www.aamc.org/download/271072/data/scientificfoundationsforfuturephysicians.pdf (accessed 24 September 2014).

Anderson LW, Krathwohl DR, Airiasian W, Cruikshank KA, Mayer RE, Pintrich PR (2001). A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Outcomes, complete ed., New York: Longman.

Andrews TM, Price RM, Mead LS, McElhinny TL, Thanukos A, Perez KE, Herreid CF, Terry DR, Lemons PP (2012). Biology undergraduate's misconceptions about genetic drift. CBE Life Sci Educ 11, 248–259.

Bloom BS, Krathwohl DR, Masia BB (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, New York: McKay.

Corbin J, Strauss A (eds.) (2008). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, 3rd ed., Thousand Oaks, CA: Sage, 159–193.

Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. CBE Life Sci Educ 7, 368–381.

Gehring KM, Eastman DA (2008). Information fluency for undergraduate biology majors: applications of inquiry-based learning in a developmental biology course. CBE Life Sci Educ 7, 54–63.

Gottesman AJ, Hoskins SG (2013). CREATE cornerstone: introduction to scientific thinking, a new course for STEM-interested freshmen, demystifies scientific thinking through analysis of scientific literature. CBE Life Sci Educ 12, 59–72.

Hoskins SG, Lopatto D, Stevens LM (2011). The C.R.E.A.T.E. approach to primary literature shifts undergraduates' self-assessed ability to read and analyze journal articles, their attitudes about science, and their epistemological beliefs. CBE Life Sci Educ 10, 368–378.

Hoskins SG, Stevens LM, Nehm RH (2007). Selective use of the primary literature transforms the classroom into a virtual laboratory. Genetics 176, 1381–1389.

Janick-Buckner D (1997). Getting undergraduates to critically read and discuss primary literature. J Coll Sci Teach 27, 29–32.

Kozeracki CA, Carey MF, Colicelli J, Levis-Fitzgerald M (2006). An intensive primary-literature–based teaching program directly benefits undergraduate science majors and facilitates their transition to doctoral programs. Cell Biol Educ 5, 340–347.

Krontiris-Litowitz J (2013). Using primary literature to teach science literacy to introductory biology students. J Microbiol Biol Educ 14, 66–77.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

McDonnell L, Barker MK, Wieman C (2016). Concepts first, jargon second improves student articulation of understanding. Biochem Mol Biol Educ 44, 12–19.

Muench SB (2000). Choosing primary literature in biology to achieve specific educational goals. J Coll Sci Teach 29, 255–260.

Round JE, Campbell AM (2013). Figure facts: encouraging undergraduates to take a data-centered approach to reading primary literature. CBE Life Sci Educ 12, 39–46.

Samuels SJ (1983). A Cognitive approach to factors influencing reading comprehension. J Educ Res 6, 261–266.

Sato BK, Kadandale P, He W, Murata PM, Latif Y, Warschauer M (2014). Practice makes pretty good: assessment of primary literature reading abilities across multiple large-enrollment biology laboratory courses. CBE Life Sci Educ 13, 677–686.

Segura-Totten M, Dalman N (2013). The CREATE method does not result in greater gains in critical thinking than a more traditional method of analyzing the primary literature. J Microbiol Biol Educ 14, 166–175.

Snow CE (2010). Academic language and the challenge of reading for learning. Science 328, 450–452.

Van Lacum EB, Ossevoort MA, Goedhart MJ (2014). A teaching strategy with a focus on argumentation to improve undergraduate students' ability to read research articles. CBE Life Sci Educ 13, 253–264.