# Expert–Novice Comparison Reveals Pedagogical Implications for Students' Analysis of Primary Literature

**April A. Nelms[†] and Miriam Segura-Totten[‡*]**
[†]Department of Teacher Education and [‡]Department of Biology, University of North Georgia, Dahlonega, GA 30597

## ABSTRACT

**Student engagement in the analysis of primary scientific literature increases critical thinking, scientific literacy, data evaluation, and science process skills. However, little is known about the process by which expertise in reading scientific articles develops. For this reason, we decided to compare how faculty experts and student novices engage with a research article. We performed think-aloud interviews of biology faculty and undergraduates as they read through a scientific article. We analyzed these interviews using qualitative methods. We grounded data interpretation in cognitive load theory and the ICAP (interactive, constructive, active, and passive) framework. Our results revealed that faculty have more complex schemas than students and that they reduce cognitive load through two main mechanisms: summarizing and note-taking. Faculty also engage with articles at a higher cognitive level, described as constructive by the ICAP framework, when compared with students. More complex schemas, effectively lowering cognitive load, and deeper engagement with the text may help explain why faculty encounter fewer comprehension difficulties than students in our study. Finally, faculty analyze and evaluate data more often than students when reading the text. Findings include a discussion of successful pedagogical approaches for instructors wishing to enhance undergraduates' comprehension and analysis of research articles.**

## INTRODUCTION

When students read primary literature, they have opportunities to encounter new material and vocabulary and analyze data and arguments. The process of reading and analyzing scientific articles results in gains in critical thinking and the ability to evaluate data, as well as gains in the understanding of how science is done (Hoskins *et al.*, 2007; Snow, 2010; Krontiris-Litowitz, 2013; Segura-Totten and Dalman, 2013). Another learning gain associated with reading primary literature is an increase in scientific literacy (Choe and Drennan, 2001; Hoskins *et al.*, 2007, 2011; Gottesman and Hoskins, 2013; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Abdullah *et al.*, 2015). Scientific literacy is defined as the understanding of the methods that produce scientific knowledge as well as those skills involved in analyzing and interpreting scientific data (Gormally *et al.*, 2012). Scientific literacy skills have been touted as necessary for the success of undergraduates in the sciences (American Association for the Advancement of Science [AAAS], 2011), and there is agreement that the development of these skills should be a priority in undergraduate biology education (National Research Council [NRC], 2003).

Given these results and calls for greater integration of primary literature reading in the undergraduate biology curriculum, faculty and researchers have developed and studied strategies for helping students learn to read research articles (e.g., Janick-Buckner, 1997; Kozeracki *et al.*, 2006; Hoskins *et al.*, 2007; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Segura-Totten and Dalman, 2013; Sato *et al.*, 2014; Marsh *et al.*, 2015; Shorbagi and Ashok, 2016). Approaches to reading primary

**Erin L. Dolan,** *Monitoring Editor*
Submitted May 17, 2018; Revised Aug 20, 2019; Accepted Aug 23, 2019
CBE Life Sci Educ December 1, 2019 18:ar56
DOI:10.1187/cbe.18-05-0077
*Address correspondence to: Miriam Segura-Totten (mstotten@ung.edu).

literature range widely. For instance, some approaches involve a high level of scaffolding that includes supplying a glossary, explicitly describing methods, and providing tips on how to read a research article (AAAS, 2017; Wenk and Tronsky, 2011). Other approaches walk students through the scientific process (e.g., Janick-Buckner, 1997; Hoskins *et al.*, 2007; Sato *et al.*, 2014). Most approaches include components that prompt students to analyze and critique data and involve giving particular assignments to help students prepare for discussion about an article (Janick-Buckner, 1997; Hoskins *et al.*, 2007; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Segura-Totten and Dalman, 2013). In our experience, preassignments are important for students to fully engage with the reading (Segura-Totten and Dalman, 2013).

Research on how high school, college, and master's students read research articles has contributed to the understanding of how students view scientific texts. For example, a comparison of the sections of articles that faculty, senior and first-year college students highlight found that faculty had a high level of agreement on important sections of the text, while this agreement decreased for college seniors and even more for college freshmen (Gallo and Rinaldo, 2012). This suggests that, as scientists progress in their careers, they are better able to distinguish important information in scientific articles. Marsh and colleagues (2015) also looked at student markings on articles and found that upper-level students, who had a higher level of comprehension of the text, highlighted papers more often and highlighted more sections of papers than introductory-level students. A survey of scientists at different career stages may help explain why novices and experts mark articles differently—Hubbard and Dunbar (2017) found that undergraduates had difficulty with the methods and results sections of primary literature. Further, undergraduates placed less value on methods and experimental results and their interpretation than individuals at later career stages (Hubbard and Dunbar, 2017). Similarly, when master's students were asked at the beginning of a course about their perceived difficulties with reading primary literature, they reported techniques and experimental data as challenges (Lie *et al.*, 2016).

Research on the analysis of scientific texts has also shed light on factors that aid in student understanding. Several studies have found that college students' ability to correctly evaluate the trustworthiness of a scientific source correlates with their comprehension of a topic (Bråten *et al.*, 2009; Wiley *et al.*, 2009). Self-explanation of concepts, monitoring comprehension of the text, and annotations of the text are also correlated with high learning gains in college and high school students (reviewed in Greenleaf *et al.*, 2011; Goldman *et al.*, 2012). The format of an article can also have an effect on comprehension: high school students who read texts adapted to facilitate student understanding had a better comprehension of an article, while those who read the versions that most closely resemble primary literature thought more critically about the topic (Norris *et al.*, 2012).

Expert–novice comparisons can provide useful insights into how to aid novices in developing more expert-like behaviors and thinking (NRC, 2012). For instance, research on expertise has shown that experts are different from novices in the following ways. One main difference between experts and novices is the vast domain knowledge of experts. Experts can organize the large amount of domain knowledge they possess to create patterns that facilitate skill application (Newell and Simon, 1972; Glaser and Chi, 2014; Feltovich *et al.*, 2018). How experts and novices approach problem-solving also differs: while novices will typically rush into solving a problem, experts first create mental models about the problem before attempting to solve it (Glaser and Chi, 2014; Feltovich *et al.*, 2018). Experts also consider a problem at a deeper level than novices. For example, studies of how experts and novices approach ill-structured problems in social science suggest that experts use their larger schemas involving the topic of the problem to decompose it into major factors that can then be addressed to find a solution (Voss and Post, 2014). Experts also have stronger monitoring skills than novices, which allows them to determine more often when they make an error (Chi *et al.*, 1982; reviewed in Glaser and Chi, 2014; Feltovich *et al.*, 2018). Experts' monitoring skills are tied to their greater domain knowledge (Simon and Simon, 1978; Chi, 1987; Glaser and Chi, 2014).

Many of the differences in how novices and experts approach tasks are linked to how individuals organize and retrieve information from memory. For this reason, research on cognitive load theory is also useful for thinking about why experts and novices might differ in how they go about a cognitively demanding task like reading primary literature. Complex readings like primary research articles have a high level of difficulty, or intrinsic cognitive load, which places a high demand on a reader's short-term memory. This is especially important to consider because individuals can hold a limited amount of information in their short-term memory (Miller, 1956; Sweller *et al.*, 2019). However, new information can be organized into schemas and stored in long-term memory, thus circumventing the limit on short-term memory (Paas *et al.*, 2003; Sweller *et al.*, 2019). Items that are stored in long-term memory can be recalled during a complex task without increasing cognitive load (Sweller *et al.*, 2019). Thus, experts' larger access to schemas stored in long-term memory frees up space in short-term memory (Chase and Ericsson, 1982; Glaser and Chi, 2014). Coupled with experts' more complex schemas is the fact that they have more practice in domain-specific skills. These two factors allow experts to retrieve information from memory faster than novices (Posner, 2014). Another way in which cognitive load can be managed is by bypassing the short-term memory limit through the chunking of bits of information into larger units, a process termed "recoding" (Miller, 1956; Paas *et al.*, 2003). Studies contrasting master chess players with experts and novices found that master players were better able to remember domain-specific knowledge through recoding of chessboard configurations (Chase and Simon, 1973; reviewed in Posner, 2014; Feltovich *et al.*, 2018). Another insight from studies of experts in diverse fields like chess, music, sports, and science is the importance of building expert-level skills through years of practice and study (reviewed in Posner, 2014; Sweller *et al.*, 2019).

Although the level of intrinsic cognitive load of a text cannot be modified, instructors can find ways of presenting the material in an easily digestible way, thus lowering the extrinsic cognitive load associated with the reading (Sweller *et al.*, 1998). According to cognitive load theory, if instructors convey material in a way that lowers the extrinsic cognitive load, learners can dedicate more of their short-term memory to the creation of new schemas (Sweller *et al.*, 1998). However, as learners progress, their schemas become more complex, and supports designed for novices can actually be counterproductive to the learning of more advanced students (Kalyuga *et al.*, 2003, 2012).

The evidence on expertise suggests that novices develop domain-specific skills over time through practice and changes in cognition. For this reason, it is important to describe the cognitive engagement of students in domain-specific tasks as they progress toward expertise. The ICAP (interactive, constructive, active, and passive) framework is particularly suited for this because it defines the levels of student cognitive engagement during tasks through specific observable behaviors (Chi, 2009; Chi and Wylie, 2014). The ICAP hypothesis divides active learning into four hierarchical modes: 1) passive, or receiving; 2) active, or manipulating; 3) constructive, or generating; and 4) interactive, or dialoguing (Chi and Wylie, 2014). For example, while reading a research article, a student might show the following levels of engagement: 1) passive, if he or she reads text passages without interacting with the reading in any other way; 2) active, if he or she performs mechanical actions like underlining seemingly important sections of the text; 3) constructive, if he or she generates novel outputs like notes in his or her own words that synthesize sections of text; and 4) interactive, if he or she debates the findings of an article with other students. According to ICAP, greater engagement during a certain task leads to higher learning and more successful integration of new information within existing schemas (Chi and Menekse, 2015; Wiggins *et al.*, 2017; reviewed in Chi and Wylie, 2014).

In this study, we sought to explore, compare, and contrast how faculty members and undergraduate students engage with a scientific article. We accomplished this by conducting think-aloud interviews of six biology faculty and 11 undergraduate students while they read a research article (Kuusela and Paul, 2000; Meijer *et al.*, 2006). We analyzed the interview data inductively using standard qualitative content analysis procedures. We then grounded the interpretation of our data in cognitive load theory and the ICAP framework. We found that faculty had more complex schemas than students and that they reduced cognitive load through two main mechanisms: summarizing and note-taking. Faculty also engaged with the article at a higher cognitive level, described as constructive by the ICAP hypothesis, when compared with students. Finally, faculty used certain dimensions of scientific literacy skills more often than students when reading the text.

## METHODS

Because our study contains 17 participants (11 students and six faculty), it is considered a small-*N* study. As noted by Gouvea (2017), insights from small-*N* studies can provide an in-depth look into how students learn science. Here, we describe our research methods, context, and participant demographics in detail to assist readers with determining how the study results apply to readers' context.

### Context

This study was conducted within the biology department at a 4-year, master's-granting university situated in the Appalachian region of the southeastern United States. The student population is ~20,000 and spans five campuses. The research conducted in this study occurred on one campus, which is the only residential undergraduate campus. Approval to conduct this study (expedited status, application 2014116) was granted by the University of North Georgia Institutional Research Board.

### Participants

Participants were 11 students who were taking courses in the biology department and six faculty members in the biology department. To identify student participants, students who were enrolled in biology department coursework were emailed about the research study and invited to participate. Participation in the study was not connected to any coursework and was completely voluntary. Students who volunteered were provided with additional study information and the informed consent form. To identify faculty participants, M.S.-T. extended an open call to her colleagues in person in the biology department. They were presented with additional study information and the informed consent form.

Student participants completed a demographic survey (Supplemental Appendix 1) that also included questions about their experiences with reading scientific articles. Because we determined that comparing student novices among themselves was beyond the scope of this study, we did not use this portion of the survey. The ethnic composition of the convenience sample included one Asian, one Latinx, and 15 Caucasians. Eight of the participants were female, eight were male, and one participant did not identify a gender. Of the 11 students, six were upper-division students and five were lower-division students. Nine of the students majored in biology, and two were business majors, one who was pursuing a biology minor and one who had taken the introductory biology course sequence. The average of the self-reported grade point average (GPA) was 3.09, and all participants reported GPAs that reflected good academic standing. The demographics of the student participants were similar to the demographics of biology department student population (departmental student demographics: 56% female, 44% male; ~10% Latinx), except for African-American students (none in the study vs. 4.6% in the department).

Of the six faculty participants, one held a master's degree and five held PhDs. One participant had been a faculty member for 1 year, and the rest had 4–19 years of experience as faculty members. None of the participants conducted research directly related to the field of the chosen article. Five had expertise in molecular biology, and the sixth was an ecologist. All faculty participants had expertise in research involving animals, and two had experience in behavioral studies with mice, which the study involved. The individual who had been faculty for only 1 year had extensive experience with behavioral, molecular, and neurobiology research using rodents, so she could be considered an expert in the general area of the article chosen.

### Think-Aloud Interviews

We wanted to choose a primary research article that would not be too complicated for first- and second-year student participants so as not to impede their thought processes. To select the article, we polled faculty in the biology department for examples of scientific articles that introductory students had read in class. From the articles nominated by faculty, we selected "Fatal Attraction in Rats infected with *Toxoplasma gondii*" (Berdoy *et al.*, 2000), because it was relatively straightforward and contained only two figures, reducing the time commitment for participants. See Supplemental Appendix 2 for the version of the article with line numbers that we used in the study.

We conducted think-aloud interviews with participants as they read the paper. Think-aloud interviews are a metacognitive exercise deriving from the information-processing tradition (Newell and Simon, 1972; Ericsson and Simon, 1993) in which participants perform a task and, while doing so, verbalize their internal thoughts (Boren and Ramey, 2000; Kuusela and Paul, 2000). The purpose of the think-aloud protocol is to ascertain an individual's cognitive processes and provide insights into participants' reasoning, self-talk, and feelings.

We asked participants to verbalize their thoughts as they read, and we reminded them to think aloud if they forgot to do so. As they read, we prompted participants with discussion questions (see Supplemental Appendix 3) that we designed to enable us to observe the participants' ability to evaluate, analyze, and interpret data and to comprehend the material in the text. Approximately 80% of the codes came from the reading of the text, while ~20% of the codes came from responses to the questions. The interviews lasted about an hour on average and were recorded for transcription.

### Analysis

The data consisted of transcripts of more than 880 minutes of the audio portion of video recordings, which were analyzed using the constant comparative method (Glasser and Strauss, 1967). We employed a purely inductive coding process because we envisioned that pursuing this study in an exploratory manner could yield novel insights to the body of research on the analysis of primary literature. To initially form codes, we individually read the transcripts in small batches to categorize the text. After forming the initial list of codes, we met monthly to discuss our work and what we understood each code to mean. From the discussions, categories, or themes, emerged that were used to sort the initial codes. Using the collaboratively developed code list, we related categories of information to one another and to the problem being investigated, and finally, we analyzed the transcripts and revised the categories and codes until we were able to produce no new codes or categories. To reach a consensus as to what code best described the data being coded, we tried to avoid instances of double coding to establish consensus validation, but at times it was unavoidable when both codes applied. While we used NVivo (QSR International) software for its organizational components and ease of determining code frequency, we coded manually within the software. Through the analysis process, we identified three categories, which were further refined as themes, and 27 codes that captured all of the strategies being used or being demonstrated by participants (the most prevalent codes are found in Table 1, and a list of all codes is shown in Supplemental Appendix 4).

Although our goal with this study was to describe and qualitatively compare how faculty as experts and undergraduate students as novices read primary literature, we thought it would be useful to calculate how often particular strategies were used by each group. Thus, we used the number of times each code was observed to calculate the average usage of a code by group and the percentage of participants who used a specific subtheme (Table 2).

## RESULTS

Our analysis revealed three themes in how experts and novices read primary literature: *thinking tools*, *scientific literacy and process skills*, and *comprehension difficulties* (Table 1). Here, we present the themes and associated subthemes that were most prevalent in our data set (a complete list of subthemes can be found in Supplemental Appendix 4).

### Key Theme 1—Thinking Tools

Students and faculty employed a variety of thinking tools to understand what they read, including rereading the text one or more times, summarizing and underlining sections of text, activating their prior knowledge, and taking notes. Thinking tools made up the majority of the codes we encountered for faculty and students. Rereading was the tool most often employed, and all faculty and student participants reread text during the interview (Table 2). However, faculty reread the text on average three times as often as students (average instance in Table 2). When rereading, participants paused after reading a section or sentence, went back to the beginning, and read the section or sentence again. All participants also summarized portions of the text during the interview, but faculty did it on average three times more often than students (Table 2). For an instance to be coded as summarizing, a participant had to paraphrase aloud what the authors wrote in the text. Often, participants summarized several sentences into a statement. For instance, one faculty participant summarized part of the experimental setup for the article, highlighting seemingly important areas, "Okay, they did it with a video camera, they had, they were illuminated and the rats had been habituated, that's really important. Alright, so they had habituated the rats to the, but they don't show any of that data."

A second faculty participant summarized the life cycle of *T. gondii* after reading the introduction of the article:

> But it looks like while they can infect everybody, the life cycle can't be completed by everybody. Cats, for this species, are what are needed.... So essentially, the progeny have to come out of cats. They won't come out of anybody else despite infections occurring in other animals.... So cats, they can reproduce out of cats, but they can infect all mammals. So everybody gets sick, but only cats can allow them to complete the life cycle.

All faculty activated prior knowledge when reading through the article, compared with 55% of students (Table 2). On average, faculty also activated prior knowledge more often than students (Table 2). As exemplified below, some of the prior knowledge of faculty participants pertained to their experience as scientists:

> The wild animal, um, it's really hard to collect and have any kind of consistency with wild animals because they come from so many different unknown social backgrounds. And when you're studying behavior, that's a really important thing to consider.

On the other hand, students' prior knowledge consisted mostly of facts, some of which they commented on having learned in biology courses, like "Well, different things can be parasitic, it just means that it takes nutrients from another."

Students also relied on the definition provided for terms in the article (45% of students; Table 2), while faculty did not use this tool while reading.

Similar numbers of faculty and student participants underlined parts of the text (67% of faculty and 64% of students; Table 2). Participants used this tool for different reasons, such as to remind themselves of important information or to note words or phrases that were unfamiliar. At other times, they underlined sections of the text without indicating a specific reason. On the other hand, a larger percent of faculty took

TABLE 1. Themes encountered during qualitative analysis[a]

| Subtheme (+/– indicates correct/incorrect) | Working definition | Example |
|---|---|---|
| **Theme 1: Thinking tools** | | |
| Rereading text one or more times "Rereading" | The participant commented that s/he reread a portion of the text. | I'm gonna go back to the last sentence.—Student |
| Summarizing or recapping "Summarizing" | The participant summarized a portion of the text. | So cats, they can reproduce out of cats, but they can infect all mammals. So everybody gets sick, but only cats can allow them to complete the life cycle.—Faculty |
| Using a reference point/prior knowledge "Prior knowledge" | The participant exhibited prior knowledge or used a reference point in the text while thinking aloud. | The wild animal, um, it's really hard to collect and have any kind of consistency with wild animals because they come from so many different unknown social backgrounds. And when you're studying behavior, that's a really important thing to consider.—Faculty |
| Underlining a key piece of information "Underlining" | The participant underlined a portion of the text. | And whenever I'm reading papers I like to underline like the summary sentences.—Student |
| Taking notes | The participant wrote down notes. | So, I'm gonna write on the side, uhh, let me see, parasites … found … are transmitted through food … transmitted … through … food … exhibit … uhh, manipulation hypotheses.—Student |
| Relying on definition of term provided in article "Relying on definition provided" | In the event that a term was described in the text, a participant indicated that he or she either understood it or noticed it. | Oh, so that's what they mean by laboratory–wild hybrids.—Student |
| **Theme 2: Scientific literacy and process skills** | | |
| Understanding research design (+/–) "Research design" | The participant indicated understanding or lack of understanding of research design. | I thought, we'll get to the t test later then won't we to compare the two corners. So they did it the way I would have done it, which is a factorial design. [*Tilts head to read Figure 1*.] And you've got infected versus noninfected and you've got the four corners. And it is a repeated measure in that case. [*Nods head*.] Sure, because each rat is going to invest[igate], could potentially go into all four corners and if they don't go into a corner, they get, they just get a zero.—Faculty |
| Evaluating a scientific argument "Evaluating" | The participant judged the quality of the research or methods in the article and provided a justification. | No. Uh, and, and to explain my answer, there was not enough concrete behavioral evidence to support it. They make statements about studies without really providing any of the evidence that is in those papers. So I don't have enough to go on to actually make that call. In fact I'm a little, little bit, I'm a little suspicious of the whole, of the whole thing. I think that was obvious when I was talking about the lab rats that they used.—Faculty |
| Analysis (+/–) | The participant verbalized at least one of the following: thoughts indicating that s/he understood relationships in the information presented in the article, analysis of the data the graph depicted, or understanding and interpretation of statistical analysis. | 'Kay, so it's saying the *T. gondii* infected cats had a preference for the cat side is the uh, as opposed to the rabbit side. [*Pause*] I mean I see what it's saying, but that graph for some reason isn't, doesn't really help me too much. I think the, the wording was best.—Student |

*Continued*

**TABLE 1. Continued**

| Subtheme (+/– indicates correct/incorrect) | Working definition | Example |
|---|---|---|
| **Theme 3: Comprehension difficulties** | | |
| Due to unknown vocabulary/jargon “Jargon” | Participants did not understand the reading because they were unfamiliar with the vocabulary or jargon being used. | I don’t know what “sorties” is.—Faculty |
| Due to lack of knowledge/incorrect knowledge “Lack of knowledge” | Participants expressed that they did not know something and/or speculated about the meaning of it. | So I-V-E-R-M-E-C-T-I-N ivermeectin A-N-T-E-H-E-L-M-I-N-T-I-C, anthelmintic. Uh, MSD-Agvet limited [*inaudible segment*]. I have, it’s clearly some type of chemical agent. I do not know what it is.—Student |
| Participant becomes distracted focusing on a small detail “Distracted” | Instead of continuing their reading, participants would become distracted or focused on a small detail that would cause them to not follow through with expressing their understandings aloud. | Student participant: “Parasitic… gondiai [/gondii/]…I don’t know how to say it.” Interviewer: “Gondii is how I say it. Would say it.” Student participant: “Gondii.” Interviewer: “Yeah. Gondii? Gondiai? I don’t know.” |
| Due to wording/sentence structure “Wording” | The wording and/or sentence structure of the article created comprehension difficulties for participants. | Including cat, okay. That’s fine, it just started with a bunch of sources and it like threw me off.—Student |

[a]The themes arising from qualitative analysis are shown in bold. The most prevalent subthemes encountered during the analysis are arranged below the themes. Shortened versions of the subtheme names used in the text are shown in quotation marks.

notes when compared with students (67 vs. 45%, respectively; Table 2). In addition, there were noteworthy differences between faculty and student notes. Faculty mostly wrote about their evaluation or analysis of a passage, as shown in these examples from two individuals:

> So they’re saying, at least in this figure legend [*points at figure legend with finger on page*], they’re making the same assumption that cat urine equals high predation risk [*writes “cat urine = high predation risk” below figure legend*].

> Uh, um, also making a note so the next question I’m supposed to answer at the end of the results is predict how the results would change if the author had used lab rats instead of wild hybrids and they address that in the methods saying that they make this hybrid so they show these innate behavior[s] or they’re arguing at least they’re showing these innate behavior responses.

In the first example above, the faculty participant is analyzing the data in the graph and the figure legend text to make a connection between avoidance of the cat urine smell and the

**TABLE 2. Think-aloud theme frequencies[a]**

| Themes/subthemes | Faculty: No. out of 6 | Faculty: Percent | Faculty: Average instance ± SEM | Students: No. out of 11 | Students: Percent | Students: Average instance ± SEM |
|---|---|---|---|---|---|---|
| Thinking tools | | | | | | |
| Rereading text one or more times | 6 | 100 | 33 ± 8.7 | 11 | 100 | 10 ± 3.1 |
| Summarizing or recapping | 6 | 100 | 14.7 ± 4.8 | 11 | 100 | 4.6 ± 1.1 |
| Using a reference point/prior knowledge | 6 | 100 | 7.5 ± 1.9 | 6 | 55 | 2 ± 0.86 |
| Underlining a key piece of information | 4 | 67 | 8 ± 3.4 | 7 | 64 | 6.9 ± 3.2 |
| Taking notes | 4 | 67 | 5.2 ± 2.7 | 5 | 45 | 2 ± 0.89 |
| Relying on definition provided | 0 | 0 | 0 | 5 | 45 | 0.45 ± 0.16 |
| Science literacy and process skills | | | | | | |
| Understanding research design + | 6 | 100 | 10.9 ± 2.1 | 11 | 100 | 3.1 ± 0.73 |
| Evaluating a scientific argument | 6 | 100 | 9.2 ± 2.1 | 2 | 18 | 0.55 ± 0.37 |
| Analysis + | 6 | 100 | 13.6 ± 0.71 | 8 | 73 | 5.6 ± 0.97 |
| Understanding research design – | 3 | 50 | 0.50 ± 0.22 | 9 | 82 | 1.7 ± 0.38 |
| Participant comprehension difficulties | | | | | | |
| Due to unknown vocabulary/jargon | 4 | 67 | 1.50 ± 0.73 | 8 | 72 | 3.40 ± 1.2 |
| Due to lack of knowledge/incorrect knowledge | 4 | 67 | 0.67 ± 0.21 | 6 | 55 | 0.73 ± 0.24 |
| Participant becomes distracted focusing on a small detail | 2 | 33 | 0.33 ± 0.21 | 2 | 19 | 0.27 ± 0.19 |
| Due to wording/sentence structure | 1 | 17 | 0.17 ± 0.17 | 6 | 55 | 0.73 ± 0.31 |

[a]Subthemes are listed below themes in order of prevalence for faculty. The number and percentage of participants who demonstrated a subtheme as well as average instance are shown. $N = 6$ (faculty); $N = 11$ (students); SEM = standard error of the mean; + means that it was done correctly, and – means that it was done incorrectly.

*y*-axis of the figure, which is shown as cumulative preference. This analysis leads the participant to determine that cat urine represents high predation risk. In the second example, the faculty participant sounds skeptical of the authors' claim that the hybrid rats show innate behavior, which suggests that the participant is evaluating this claim before accepting it.

In contrast, most of students' notes were written as reminders of what the text stated, such as, "So, I'm gonna write on the side, uhh, let me see, parasites … found … are transmitted through food … transmitted … through … food … exhibit … uhh, manipulation hypotheses."

Because of the qualitative differences we observed between faculty and student tool usage, we further classified thinking tools into active or constructive according to the ICAP framework (Chi and Wylie, 2014). Passive engagement with the text involves reading without taking other actions, while interactive engagement involves dialogue between two or more individuals, so our codes reflect only active or constructive behaviors. Rereading, underlining, and relying on the definition provided for a term are active behaviors, because the reader is interacting with the text but is not producing new information or inferences as a result. Summarizing, using prior knowledge, and taking notes can be used actively or constructively, depending on whether the participant derives new information or inferences while using the tool. For faculty, 26% of instances of activating prior knowledge were active, and 74% were constructive, with faculty either explaining what the text meant to them using words not mentioned in the text, or generating new knowledge (Chi and Wylie, 2014; Table 3). On the other hand, 38% of instances of prior knowledge activation in students were active and 62% were constructive (Table 3). Seventy-five percent of notes taken by faculty were constructive and 25% were active, while 21% of student notes were constructive and 79% were active (Table 3). Constructive notes contained new ideas, the synthesis of several ideas in the article, or the evaluation of information presented, while active notes were verbatim from the text. Finally, 73% of the time faculty generated summaries that were constructive, containing new inferences and explanations of the material (also called "self-explaining"; Chi and Wylie, 2014; Table 3). Students generated constructive summaries 60% of the time, while 40% were "copy-and-delete" summaries, which are verbatim summations of the text (Chi and Wylie, 2014). Overall, for activation of prior knowledge, note-taking, and summarizing, 74% of faculty instances were constructive and 26% were active compared with 52% constructive and 48% active for students.

In summary, faculty and students used a variety of thinking tools as they read, but they used these tools differently. Faculty summarized portions of the text, activated their prior knowledge, and took notes more often than students. Furthermore, faculty notes often included the results of analysis and evaluation of information, while student notes reflected material taken directly from the text. Finally, faculty used thinking tools in a constructive manner more often when compared with students.

**TABLE 3. Frequency of active and constructive usage of thinking tools[a]**

| | Faculty | | Students | |
|---|---|---|---|---|
| **Tool** | **Constructive** | **Active** | **Constructive** | **Active** |
| Prior knowledge | 74 | 26 | 62 | 38 |
| Note taking | 75 | 25 | 21 | 79 |
| Summarizing | 73 | 27 | 60 | 40 |

[a]Each instance of tool usage was classified as either active or constructive, as defined by Chi and Wylie (2014). The percent of the total number of times a tool was used is shown.

## Key Theme 2—Science Literacy and Process Skills

Faculty and students displayed science literacy and process skills as they read, including analyzing information, evaluating arguments, and making sense of the research design.

All faculty and students displayed an understanding of the research design at some point while reading the article (Table 2). However, on average, faculty displayed this ability almost four times as often as students (average instance in Table 2). This faculty participant demonstrates understanding of how the study was set up. In addition, the passage illustrates how the participant incorporated the research design described in the article into his existing schema of factorial design and quantification of rat behavior:

> I thought, we'll get to the t test later then won't we to compare the two corners. So they did it the way I would have done it, which is a factorial design. [*Tilts head to read Figure 1.*] And you've got infected versus noninfected and you've got the four corners. And it is a repeated measure in that case. [*Nods head.*] Sure, because each rat is going to invest[igate], could potentially go into all four corners and if they don't go into a corner, they get, they just get a zero.

Students were unable to understand elements of the research design more often than faculty (82% of students and 50% of faculty; Table 2). For instance, this student demonstrated lack of understanding of experimental design in response to the prompt "Predict how the results would change if the authors had used laboratory rats instead of laboratory–wild rat hybrids." Specifically, the student failed to put the selection of laboratory rats in the context of the rest of the experimental setup described:

> So they probably would respond a lot differently and most likely die from the parasite. Um, secondly you wouldn't really be able to compare them to an environmental, like factor, I guess like an environmental situation if they grew up all their life not living in the wild. Um, so, they might, they would just all, like the parasite would probably affect them a lot differently.

All faculty evaluated a scientific argument while reading the article, compared with only 18% of students (Table 2). Moreover, on average, faculty demonstrated the skill of evaluation 17 times more often than students (Table 2). For example, in response to the question "Given the information presented in the Introduction, do you think that *T. gondii* will interfere with the rat's innate reaction to potential predation risk by cats?," this faculty participant is skeptical of the authors' claims given the data they presented:

> No. Uh, and, and to explain my answer, there was not enough concrete behavioral evidence to support it. They make statements about studies without really providing any of the evidence that is in those papers. So I don't have enough to go on to actually make that call. In fact I'm a little, little bit, I'm a

> little suspicious of the whole, of the whole thing. I think that was obvious when I was talking about the lab rats that they used.

Faculty and students engaged in three different types of analysis that are encompassed by the analysis subtheme: analyzing the relationship between different pieces of information, the analysis of graphical data, and the interpretation of statistical analyses (Table 1). All faculty and 83% of students engaged in analysis while reading the article (Table 2). On average, faculty correctly analyzed information twice as often as students (average instance in Table 2). This student correctly interpreted a figure in response to one of the questions embedded in the article:

> Okay, let's see there is a significant difference with the noninfected rats and the infected rats when it came to the smell of the cat. And it looks like that was their activity. So the infected rats had a higher activity level and seemed to go out more where the smell was than the uninfected who probably stayed in their little pen things. And let's see the other ones were pretty close to each other, so it doesn't look like a big difference.

In summary, faculty used scientific literacy skills more often than students when analyzing the article. Specifically, faculty were more capable of making sense of the study design, they more frequently evaluated the claims presented, and they analyzed information more often than students.

### Key Theme 3—Comprehension Difficulties

Comprehension difficulties arose for students and faculty as they read through the research article. These included jargon, lack of knowledge, distraction, and sentence structure. Not surprisingly, faculty had a lower level of comprehension difficulties (2% of the total code instances), while students voiced five times more comprehension difficulties than faculty (10% of the total code instances).

In this study, we defined "jargon" as a technical word or phrase that is unknown to the participant. If participants noted that they did not know a particular word and did not look the word up or otherwise try to understand its meaning, that occurrence was placed in this subtheme. For both faculty and students, jargon comprised the largest subtheme within comprehension difficulties, with 67% of faculty and 72% of students encountering jargon at least once while reading the article (Table 2). Additionally, students noted coming across jargon twice as often as faculty on average (average instance in Table 2). Students frequently made statements such as "never heard that word before" or "I don't know that word either" while reading and thinking aloud.

The second most prevalent reason why faculty had trouble comprehending the text was due to lack of knowledge (67%; Table 2). Although the percentage of students who reported lack of knowledge was lower than that of faculty (55%; Table 2), both groups voiced lack of knowledge at a similar rate (average of $0.67 \pm 0.21$ for faculty and $0.73 \pm 0.24$ for students; Table 2). In this usage, participants expressed that they did not know something, or they speculated about the meaning of it. Lack of knowledge is more extensive than just the meaning of a word (i.e., jargon). For example, in the following passage, the faculty participant does not know how *T. gondii* is transmitted to humans:

> Which also, I don't understand why the parasite would be in humans, even though it said that it affected all mammals, but if the parasite had to lay its oocysts … that does—I don't understand because there's no predator of humans so I don't know, unless humans contracted them from meat that they were eating. That would be the only thing that would make sense.

The third most prevalent reason for faculty having difficulty in comprehending the text was distraction caused by focusing on a detail in the text; this affected 33% of faculty compared with 18% of students (Table 2). A distraction would manifest in a manner that would prevent a participant from progressing in the reading, ultimately delaying their reading comprehension. For example, this faculty participant was hindered by the phrase "The rising line for uninfected cats":

> [*Participant rereads*] "The rising line for uninfected rats." [*Pauses, continues paraphrasing*] indicates a prolonged avoidance of cat-scented areas. [*Continues reading, starting with the legend of Figure 2.*]
>
> [*Participant finishes line 17 of Figure 2 legend and draws line adjacent to x-axis, separating third and fourth sortie.*]
>
> [*Participant rereads*] "The rising line for uninfected rats." [*Circles "The rising line for uninfected rats" on Figure 2 legend.*] "The rising line of uninfected rats." Something is wrong there, either with me or with this paper. "The rising line for uninfected rats." Okay, I don't think that can be right.

The least prevalent reason for faculty not comprehending the text was "Due to wording/sentence structure" (17% of faculty; Table 2). In contrast, 55% of students did not understand the text because of wording and sentence structure (Table 2). Although published papers are written by experts and reviewed by editors before publication, a reader may still have difficulty with the wording or the sentence structure of the text. When this occurs, it could delay or even prevent comprehension. For example, this student expresses frustration with the dense construction of a sentence: "[*Rereads sentence that starts in line 3*] It's like, a lot of words, haha, that I don't, I don't know, don't feel like they go together."

## DISCUSSION

The inclusion of primary literature analysis in the classroom has grown in significance over the past two decades, as evidenced by the number of published approaches to reading the literature (e.g., Janick-Buckner, 1997; Kozeracki *et al.*, 2006; Hoskins *et al.*, 2007; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Segura-Totten and Dalman, 2013; Sato *et al.*, 2014; Marsh *et al.*, 2015; Shorbagi and Ashok, 2016). To better understand how expertise in reading primary literature develops, we compared how faculty and students went about reading a research article during think-aloud interviews. Our study reveals processes and techniques that biology undergraduates and faculty employ as they read through primary literature. Our research also shows that faculty and students engage with research articles dissimilarly, suggesting areas of student development that could improve their analysis of scientific papers.

### Interpreting How Experts and Novices Read Primary Literature Using Cognitive Load Theory and the ICAP Framework

*Cognitive Load Theory.* Perhaps not unexpectedly, our study revealed that faculty experts have fewer comprehension difficulties than student novices when they read a research article. Three reasons grounded in cognitive load theory may help explain why faculty demonstrated fewer comprehension difficulties: 1) experts have more complex and extensive schemas related to scientific techniques, knowledge, and process (as evidenced by the higher number of instances of faculty comments and their complexity within the subthemes of understanding of research design and prior knowledge), 2) experts encounter unfamiliar technical words less often than students (perhaps because of their more extensive schemas), and 3) experts reduce their cognitive load by recoding information.

Faculty experts' more complex schemas for technical terms, techniques, and scientific process allow them to retrieve prior knowledge relevant to the reading more often than students, thus placing fewer demands on short-term memory. This may lead experts to have a higher cognitive-processing capacity (as has been shown for problem-solving in mathematics; Sweller, 1988) and, thus, a deeper understanding of the textual material. Faculty may also use their complex schemas to make more connections between the information presented in the article and their prior knowledge, thus facilitating analysis and evaluation of data.

To reduce cognitive load while reading the article, faculty used recoding through summarizing and note-taking. Faculty summarized passages from the text and took notes on average about three times as often as students (Table 2). Recoding by summarizing has also been described as self-explanation of material when it contains new information generated by the individual (Chi *et al.*, 1989; Chi and Wylie, 2014). This finding is particularly interesting to us, because self-explanation improves students' understanding of a text as well as their course performance, particularly for students with low levels of initial knowledge (Chi *et al.*, 1994; McNamara, 2017). Also, summaries that are more complete and connect the material to prior knowledge are linked with better student performance (Bednall and James Kehoe, 2011). Because faculty notes in many cases contained the analysis or evaluation of a passage, these notes could also be considered the simplifying, or recoding, of a large amount of information. On the other hand, student notes were often reminders of what the text stated without significant analysis or evaluation. Thus, while it seems that students were still attempting to reduce cognitive load through note-taking, they did not do it as frequently as faculty, and they did not take notes of recoded information. Interestingly, a study of high school graduates who were asked to take notes while reading a text showed that participants who took notes that summarized the text and those who took more notes better comprehended the text when compared with students who took fewer notes and to those whose notes contained verbatim information from the text (Slotte and Lonka, 1999).

On the basis of our findings, we recommend reducing the extrinsic cognitive load of a research article to facilitate student understanding. One way to do this is by including resources to supplement student prior knowledge. For example, the use of a glossary could help to reduce student issues related to jargon. Several published approaches for the analysis of the primary literature feature ways of supplementing student knowledge on techniques (Janick-Buckner, 1997; Jacques-Fricke *et al.*, 2009; Segura-Totten and Dalman, 2013), and one of these resulted in students who were better able to select the correct technique to answer a particular scientific question, an important component of research design (Jacques-Fricke *et al.*, 2009). The annotated research articles created by the American Association for the Advancement of Science (AAAS, 2017; McCartney *et al.*, 2018) and the approach described in Abdullah *et al.* (2015) both target student issues with unknown terms and techniques and may serve as effective platforms for literature discussions. Additionally, the process of annotating leads to a more readable version of the research article, which in turn may lead to better student understanding (McCartney *et al.*, 2018).

Instructors could also decrease the extrinsic cognitive load of a research article through focused instruction. Master's students who were asked about the perceived challenges associated with reading research articles reported issues associated with jargon and lack of knowledge of scientific methods (Abdullah *et al.*, 2015). After focused instruction on reading scientific articles, including discussion of terms and methods found in the texts, the less experienced student participants (who may align better with the undergraduates in our study) reported fewer issues with scientific terms and information in scientific articles (Lie *et al.*, 2016). They also reported fewer problems understanding the writing style of scientific articles, an issue that surfaced for students in our study (Table 2, "Participant Comprehension Difficulties: Due to Wording/Sentence Structure"). A third way to lower cognitive load is by having students summarize and take notes as they read an article. We will discuss these two techniques in the context of the ICAP hypothesis in the next section.

*ICAP Hypothesis.* The ICAP framework describes how an individual's knowledge is changed by different types of engagement during a cognitive task: 1) new information is stored in isolation (passive engagement); 2) new information leads to the activation of prior knowledge and is integrated into existing schemas (active); and 3) after integration, additional new knowledge is inferred based on the activated and integrated knowledge (constructive). We will not address the interactive mode of ICAP, because it requires constructive dialogue between two or more individuals, something that was not captured in our think-aloud interviews. ICAP predicts that active engagement during a task allows students to fill in gaps in their schemas so they may more easily retrieve this information when they encounter similar cognitive tasks (Chi and Wylie, 2014). Constructive engagement during a task leads to deeper learning than active engagement, because the knowledge obtained can be transferred and applied in different contexts (Chi and Wylie, 2014). Analysis of published studies and the results of studies designed to test the ICAP framework show that a higher level of engagement leads to higher learning gains and student understanding (Chi and Wylie, 2014; Chi and Menekse, 2015;

Wiggins *et al.*, 2017). In our study, faculty engaged with the article at a constructive level more often than students (Table 3), better inferred new knowledge, and created outputs more often (e.g., rich notes that include data analysis and evaluation). Thus, the higher level of cognitive engagement of faculty while reading the research article may help explain why this group encountered fewer comprehension difficulties than students in our study.

We hypothesize that instructors can enhance student comprehension of scientific articles by encouraging them to engage with the material in constructive ways. For example, we predict that prompting students to summarize information constructively while they read an article will lead to better comprehension of the text. Summarizing allows readers to connect different pieces of the text to form a cohesive idea, rather than evaluating separate pieces of information in isolation (Dunlosky *et al.*, 2013). This, in turn, may be helpful in distilling the big picture of a scientific study. We recommend that instructors train students in how to summarize information and support them in doing so, for example, by showing them how to draft summaries, creating assignments that involve students in practicing this skill or by using approaches to reading primary literature that engage students in summarizing (e.g., Round and Campbell, 2013; Sato *et al.*, 2014). Another way for students to engage constructively while reading a research article is by creating notes that capture data analysis and evaluation. Because the cognitive engagement of faculty and students differed the most during note-taking (Table 3), we foresee that this will be an area where students will require a lot of support and practice. Two ways of teaching students how to take richer notes are to use an approach to reading primary literature that contains a note-taking component (Hoskins *et al.*, 2007; Round and Campbell, 2013) and to model expert behavior in class.

### Development of Expertise in Scientific Literacy

In our study, faculty experts analyzed and evaluated scientific information more often than student novices. It is very possible that experts in our study acquired these skills through years of reading scientific articles. Previous studies show that the analysis of research articles increases dimensions of scientific literacy and knowledge of the science process in undergraduate and graduate students (Choe and Drennan, 2001; Hoskins *et al.*, 2007, 2011; Gottesman and Hoskins, 2013; Krontiris-Litowitz, 2013; Round and Campbell, 2013; Abdullah *et al.*, 2015). In fact, thousands of hours of practice are required to attain expertise in a certain domain (Chase and Simon, 1973; reviewed in Posner, 2014; Ericsson, 2018). Alternatively, students' epistemological beliefs on the nature of scientific knowledge may lead them to view information in scientific articles as immutable facts (Schommer, 1990). This may in turn prevent them from attempting to analyze or evaluate the text. It would be interesting to determine whether changes in students' epistemological beliefs go hand in hand with increases in their ability to analyze and evaluate scientific sources and whether students with more expert-like epistemological beliefs are better able to analyze and evaluate scientific data. It is worth noting that the two explanations we posit are not mutually exclusive and that the development of student epistemological beliefs may be another facet of the progression from novice to expert.

Students in our study also had more difficulty than faculty in understanding research design. Instructors could help students make sense of research designs by dedicating one or more class sessions to the nature of scientific research design in the context of the articles that students will be discussing. Alternatively, lecture activities could highlight the research design of classical experiments or of those related to the content of the lecture. The textbook described by Barsoum and colleagues (2013) provides a good model to follow for the incorporation of experimental design into lectures. Instructors could also employ approaches that prompt students to consider the research design of an experiment (Janick-Buckner, 1997; Hoskins *et al.*, 2007; Krontiris-Litowitz, 2013; Sato *et al.*, 2014). Although it is unlikely that undergraduates will achieve mastery of research design in one semester, exposing students to the approaches detailed above over the course of several semesters, especially early in their college careers, may accelerate their ability to understand this important facet of scientific literacy (Coil *et al.*, 2010). Instructors may choose to further develop student scientific literacy skills through a first-year course that trains students to read primary literature and exposes them to the nature of science or, alternatively, throughout students' college careers by adding activities to courses that involve data analysis and evaluation and expose students to research design.

### Limitations of Our Study

While our study involved a small group of faculty and students, the demographic composition of our participants is representative of the overall biology department population at our institution. It will be interesting to repeat our study in other types of institutions and with other individuals to determine whether there are variations in the way students read and analyze primary literature that we did not observe in our student population.

It is important to note that, while faculty participants can be considered experts in the analysis of research articles, they are not experts in the topic of the research study for the article they read. Thus, the skills we observed in this study denote those that are transferable across domains of knowledge, at least within biology. As has been shown for expert–expert comparisons in other processes (e.g., Roth and Bowen, 2003), we may find other dimensions or a different organization for the construction of knowledge if we examine faculty as they analyze articles in their subjects of expertise. A comparison of faculty who are experts in the content area of a research article with students who are not familiar with the content area might yield further insight into how experts and novices differ in their analysis of primary literature.

### Future Directions

The results of our study suggest different approaches that would help support students' analysis of research articles. It would be interesting to test whether including elements in literature discussions that lower the extrinsic cognitive load of the text, such as glossaries and information on experimental design, leads to better student comprehension. Looking at student comprehension of a text in its original format compared with an annotated version, like those available through the Science in the Classroom initiative (AAAS, 2017), could shed light on whether scaffolds to lower extrinsic cognitive load aid in student understanding. This study focused on the commonalities rather than differences in how undergraduates approach reading the literature by looking at the student participants as a group instead of comparing participants at different points of the college experience. However, it would be interesting to investigate how student methods for reading articles vary, especially as they progress through their college studies. Determining how students' approaches to reading research articles change during their college career through longitudinal think-aloud interviews of a cohort of students could allow us to better describe the transition from novice to expert in this process.

## ACKNOWLEDGMENTS

We are thankful to Kristen N. Redmon, Kelsee Ryan, and Emma K. Stapleton for their assistance with the transcription of student and faculty interviews. We are grateful to Dr. T. Jameson Brewer, Dr. Frank Corotto, Dr. Kelly L. McFaden, Dr. Margaret Smith, and Dr. Julie Dangremond Stanton for their critical reading of the article. We are also grateful to Dr. Erin Dolan and Dr. Stephanie Gardner for their guidance during the manuscript review process. This study was funded in part through the Harry B. Forrester Fund–UNG Foundation and a UNG Presidential Award to A.A.N. and M.S.-T.

## 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 Sciences Education*, *14*(3), ar34. doi: 10.1187/cbe.14-10-0180

American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action. Final report*. Washington, DC. Retrieved August 5, 2019, from http://visionandchange.org/finalreport/

AAAS. (2017). *Science in the classroom*. Retrieved August 5, 2019, from www.scienceintheclassroom.org/

Barsoum, M. J., Sellers, P. J., Campbell, A. M., Heyer, L. J., & Paradise, C. J. (2013). Implementing recommendations for introductory biology by writing a new textbook. *CBE—Life Sciences Education*, *12*(1), 106–116. doi: 10.1187/cbe.12-06-0086

Bednall, T. C., & James Kehoe, E. (2011). Effects of self-regulatory instructional aids on self-directed study. *Instructional Science*, *39*(2), 205–226. doi: 10.1007/s11251-009-9125-6

Berdoy, M., Webster, J. P., & Macdonald, D. W. (2000). Fatal attraction in rats infected with *Toxoplasma gondii*. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1452), 1591–1594. Retrieved August 5, 2019, from www.ncbi.nlm.nih.gov/pmc/articles/PMC1690701]

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261–278. doi: 10.1109/47.867942

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28. doi: 10.1598/RRQ.44.1.1

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In Bower, G. H. (Ed.), *Psychology of learning and motivation* (Vol. 16, pp. 1–58). Cambridge, MA: Academic Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi: 10.1016/0010-0285(73)90004-2

Chi, M. T. H. (1987). Representing knowledge and metaknowledge: Implications for interpreting metamemory research. In Weinert, F. E., & Kluwe, R. (Eds.), *Metacognition, motivation, and understanding* (pp. 239–266). Hillsdale, NJ: Erlbaum.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. doi: 10.1207/s15516709cog1302_1

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. doi: hTtps://doi.org/10.1016/0364-0213(94)90016-7

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In Sternberg, R. J. (Ed.), *Advances in the psychology of human intelligence* (Vol. *1*). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns in peer collaboration that promote learning. In Resnick, L. B., Asterhan, C. S. C., & Clarke, S. N. (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 253–264). Washington, DC: American Educational Research Association.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. doi: 10.1080/00461520.2014.965823

Choe, S. W. T., & Drennan, P. M. (2001). Analyzing scientific literature using a jigsaw group activity: Piecing together student discussions on environmental research. *Journal of College Science Teaching*, *30*(5), 328–330.

Coil, D., Wenderoth, M. P., Cunningham, M., & Dirks, C. (2010). Teaching the process of science: Faculty perceptions and an effective methodology. *CBE—Life Sciences Education*, *9*(4), 524–535. doi: 10.1187/cbe.10-01-0005

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. doi: 10.1177/1529100612453266

Ericsson, K. A. (2018). The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts. In Williams, A. M., Kozbelt, A., Ericsson, K. A., & Hoffman, R. R. (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 745–769). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2018). Studies of expertise from psychological perspectives: Historical foundations and recurrent themes. In Williams, A. M., Kozbelt, A., Ericsson, K. A., & Hoffman, R. R. (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 59–83). Cambridge, UK: Cambridge University Press.

Gallo, M., & Rinaldo, V. (2012). Towards a mastery understanding of critical reading in biology: The use of highlighting by students to assess their value judgment of the importance of primary literature. *Journal of Microbiology & Biology Education*, *13*(2), 142–149. doi: doi:10.1128/jmbe.v13i2.493

Glaser, R., & Chi, M. T. H. (2014). Overview. In Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.), *The nature of expertise* (pp. xv–xxviiix). London: Psychology Press.

Glasser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research* (2nd ed.). Chicago: Aldine.

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, *47*(4), 356–381. doi: 10.1002/rrq.027

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, *11*(4), 364–377. doi: 10.1187/cbe.12-03-0026

Gottesman, A. J., & Hoskins, S. G. (2013). CREATE Cornerstone: Introduction to scientific thinking, a new course for STEM-interested freshmen, demystifies scientific thinking through analysis of scientific literature. *CBE—Life Sciences Education*, *12*(1), 59–72. doi: 10.1187/cbe.12-11-0201

Gouvea, J. (2017). Insights from small-*N* studies. *CBE—Life Sciences Education*, *16*(3), fe4. doi: 10.1187/cbe.17-06-0110

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., … & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, *48*(3), 647–717. Retrieved August 5, 2019, from www.jstor.org/stable/27975305

Hoskins, S. G., Lopatto, D., & Stevens, L. M. (2011). The C.R.E.A.T.E. approach to primary literature shifts undergraduates' self-assessed ability to read and analyze journal articles, attitudes about science, and epistemological beliefs. *CBE—Life Sciences Education*, *10*(4), 368–378. doi: 10.1187/cbe.11-03-0027

Hoskins, S. G., Stevens, L. M., & Nehm, R. H. (2007). Selective use of the primary literature transforms the classroom into a virtual laboratory. *Genetics*, *176*(3), 1381–1389. Retrieved August 5, 2019, from www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17483426

Hubbard, K. E., & Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PLoS ONE*, *12*(12). doi: 10.1371/journal.pone.0189753

Jacques-Fricke, B., Hubert, A., & Miller, S. (2009). A versatile module to improve understanding of scientific literature through peer instruction. *Journal of College Science Teaching*, *39*(2), 24–32.

Janick-Buckner, D. (1997). Getting undergraduates to critically read and discuss primary literature. *Journal of College Science Teaching*, *27*, 340–347.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. doi: 10.1207/S15326985EP3801_4

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313–337. doi: 10.1007/s10648-012-9195-x

Kozeracki, C. A., Carey, M. F., 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 Biology Education*, *5*(4), 340–347.

Krontiris-Litowitz, J. (2013). Using primary literature to teach science literacy to introductory biology students. *Journal of Microbiology & Biology Education*, *14*(1), 66–77. doi: 10.1128/jmbe.v14i1.538

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology*, *113*(3), 387–404.

Lie, R., Abdullah, C., He, W., & Tour, E. (2016). Perceived challenges in primary literature in a master's class: Effects of experience and instruction. *CBE—Life Sciences Education*, *15*(4), ar77. doi: 10.1187/cbe.15-09-0198

Marsh, T. L., Guenther, M. F., & Raimondi, S. L. (2015). When do students "learn-to-comprehend" scientific sources? Evaluation of a critical skill in undergraduates progressing through a science major. *Journal of Microbiology & Biology Education*, *16*(1), 13–20. doi: 10.1128/jmbe.v16i1.828

McCartney, M., Childers, C., Baiduc, R. R., & Barnicle, K. (2018). Annotated primary literature: A professional development opportunity in science communication for graduate students and postdocs. *Journal of Microbiology & Biology Education*, *19*(1). doi: 10.1128/jmbe.v19i1.1439

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*, *54*(7), 479–492. doi: 10.1080/0163853X.2015.1101328

Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation*, *12*(3), 209–237. doi: 10.1080/13803610500479991

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi: 10.1037/h0043158

National Research Council (NRC). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.

NRC. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Eglewood Cliffs, NJ: Prentice Hall.

Norris, S. P., Stelnicki, N., & de Vries, G. (2012). Teaching mathematical biology in high school using adapted primary literature. *Research in Science Education*, *42*(4), 633–649. doi: 10.1007/s11165-011-9215-8

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. doi: 10.1207/S15326985EP3801_1

Posner, M. I. (2014). Introduction: What is it to be an expert? In Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.), *The nature of expertise* (pp. xv–xxviii). London: Psychology Press.

Roth, W.-M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction*, *21*(4), 429–473. doi: 10.1207/s1532690xci2104_3

Round, J. E., & Campbell, A. M. (2013). Figure facts: Encouraging undergraduates to take a data-centered approach to reading primary literature. *CBE—Life Sciences Education*, *12*(1), 39–46. doi: 10.1187/cbe.11-07-0057

Sato, B. K., Kadandale, P., He, W., Murata, P. M. N., Latif, Y., & Warschauer, M. (2014). Practice makes pretty good: Assessment of primary literature reading abilities across multiple large-enrollment biology laboratory courses. *CBE—Life Sciences Education*, *13*(4), 677–686. doi: 10.1187/cbe.14-02-0025

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, *82*(3), 498–504.

Segura-Totten, M., & Dalman, N. E. (2013). The CREATE method does not result in greater gains in critical thinking than a more traditional method of analyzing the primary literature. *Journal of Microbiology & Biology Education*, *14*(2), 166–175. doi: 10.1128/jmbe.v14i2.506

Shorbagi, S., & Ashok, A. (2016). Designing an audiocast assignment: A primary-literature-based approach that promotes student learning of cell and molecular biology through conversations with scientist authors. *Journal of Microbiology & Biology Education*, *17*(3), 472–474. doi: 10.1128/jmbe.v17i3.1110

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In Siegler, R. S. (Ed.), *Children's thinking: What develops?* (pp. 325–348). Hillsdale, NJ: Erlbaum.

Slotte, V. V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, *21*(4), 1–20.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*(5977), 450–452. doi: 10.1126/science.1182597

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. doi: 10.1007/s10648-019-09465-5

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Voss, J. F., & Post, T. A. (2014). On the solving of ill-structured problems. In Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.), *The nature of expertise* (pp. 261–285). London: Psychology Press.

Wenk, L., & Tronsky, L. (2011). First-year students benefit from reading primary research articles. *Journal of College Science Teaching*, *40*(4), 60–67.

Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences. *AERA Open*, *3*(2), 2332858417708567. doi: 10.1177/2332858417708567

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106. Retrieved August 5, 2019, from www.jstor.org/stable/40284747]