

# Students' Perceptions of Preprints Discovered in Google: A Window into Recognition And Evaluation

Tara Tobin Cataldo, Ixchel M. Faniel, Amy G. Buhler, Brittany Brannon, Lynn Silipigni Connaway, and Samuel Putnam\*

Preprints play an important role in scholarly conversation. This paper examines perceptions of preprints through the lens of students using a simulated Google environment. Data were collected from 116 high school, community college, undergraduate, and graduate students with attention toward the helpfulness, credibility, and identification of preprints. Findings show preprint and peer-reviewed cues play little to no role in judging helpfulness or citability, but peer-review does when judging credibility. Further, most students did not recognize these resources as preprints. Implications and recommendations are discussed surrounding awareness and use of these openly available sources of scientific information.

## Introduction

Preprints can be described as scholarly manuscripts “posted by the author(s) in an openly accessible platform, usually before or in parallel with the peer review process.”<sup>1</sup> Preprints have been a part of scholarly communication for decades. [ArXiv.org](https://arxiv.org/), a preprint server that primarily archives research in physics, mathematics, and computer science, was created in 1991. Other preprint servers such as bioRxiv for the life sciences and ChemRxiv for the chemical sciences, were created in the past decade. Fraser (2021) surmised that the growth of preprints could be disruptive to the online information environment, and this came to fruition during the COVID-19 pandemic.<sup>2</sup> The easy access to and enormous output of preprints has sparked debates about their positives and negatives,<sup>3</sup> their use in journalism,<sup>4</sup> and journal policies and guidelines.<sup>5</sup> The publishing world is by no means united in its view of preprints. Journals such as *Anesthesiology* prohibit references to non-peer-reviewed scientific articles.<sup>6</sup> Others, such as *eLife*, require authors to submit their manuscripts to a preprint server.<sup>7</sup> Klebel (2020) details the ambiguity of

---

\* Tara Tobin Cataldo is the Biological Sciences Librarian at the University of Florida, email: [ttobin@ufl.edu](mailto:ttobin@ufl.edu); Ixchel M. Faniel is a Senior Research Scientist at OCLC Research, email: [fanieli@oclc.org](mailto:fanieli@oclc.org); Amy G. Buhler is an Engineering Librarian at the University of Florida, email: [abuhler@ufl.edu](mailto:abuhler@ufl.edu); Brittany Brannon is a Senior Research Specialist at OCLC Research, email: [brannonb@oclc.org](mailto:brannonb@oclc.org); Lynn Silipigni Connaway is the Executive Director, Research at OCLC Research, email: [connawal@oclc.org](mailto:connawal@oclc.org); and Samuel Putnam is the Director of Bern Dibner Library at New York University, email: [samuel.putnam@nyu.edu](mailto:samuel.putnam@nyu.edu). ©2023 Tara Tobin Cataldo, Ixchel M. Faniel, Amy G. Buhler, Brittany Brannon, Lynn Silipigni Connaway, and Samuel Putnam, Attribution-NonCommercial (<https://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

many journals' preprint and peer review policies.<sup>8</sup> Some studies have compared preprints to the published versions of the articles, and most find the differences markedly small.<sup>9</sup> Many are concerned that without peer review, poor quality research in preprints will clutter the scholarly discourse. Others argue that even with peer review, questionable studies are still published. As of this writing, the Retraction Watch service listed 190 COVID-19 articles as having been retracted.<sup>10</sup> Only forty-one (22%) of the articles are preprints, and some of the retractions are from prestigious journals such as the *New England Journal of Medicine* and *The Lancet*.

Despite the controversy over preprints and the value of the peer review process, researchers are citing preprints in their work. Lariviere et al.'s study showed articles indexed in Web of Science cite arXiv preprints in a range of 0.2% to 6.6% depending on the discipline.<sup>11</sup> Pagliaro used the databases Scopus and Dimensions to study preprint citation patterns from three servers, amassing evidence that preprints are regularly cited in peer-reviewed journal articles, books, and conference proceedings.<sup>12</sup> Students are using preprints in academic work as well. Flynn reported that preprints represented 4.6% of citations from an analysis of math and statistics dissertations.<sup>13</sup> Flynn expressed concern that the average age of the preprints was 8.4 years, but the study did not investigate if the preprints had published versions available at the time. In Dotson and Frank's study of civil engineering, computer science, mathematics, and physics dissertations, up to 4.2% of the citations were preprints.<sup>14</sup>

The lack of formal peer review is an important characteristic of preprints and the key differentiator between preprint articles and journal articles. A review of the literature finds no studies examining the relationship between peer review and preprints amongst students. However, there are studies of students' perceived value of peer review and application of this cue in the information seeking process. For example, Komissarov's survey revealed that undergraduate students value peer-reviewed sources, although this attribute ranked fifth in importance, after subject relevance, full-text availability, understandability, and currency, respectively.<sup>15</sup> In an observational study of twenty undergraduates' evaluation of online information, Rempel discovered that students were attracted to descriptions mentioning "peer review" or "scholarly" when choosing their search tool.<sup>16</sup> However, once students began selecting resources within a tool, they mostly favored content or their ability to understand and apply the information from a resource, with peer-review status not necessarily serving as a deciding factor. Several students in the study also assumed that the library databases would provide only peer-reviewed sources, and none used the peer-review facet available in one of the search tools. This is contrary to the Kliewer et al. study, which found that when undergraduates did apply a facet to their search, they applied the peer-reviewed facet about a third of the time.<sup>17</sup> Blummer et al. also found that graduate students mentioned using the peer-reviewed facet, looking at peer-reviewed journals only and peer-reviewed articles endorsed by a trusted professional association.<sup>18</sup>

The literature on students' perception of preprints and their lack of peer review is sparse. One exception is a recent study by Soderburg et al., who surveyed researchers (including graduate students) regarding the cues they value as important to the task of assessing the credibility of preprint articles.<sup>19</sup> The study found that researchers who view preprints favorably (including 80% of graduate students) place less value on 'traditional' metrics like author and peer-review information. Instead they tend to value open science (e.g., links to analysis scripts and data) and independent verification indicators (e.g., links about independent reproductions). However, participants were not specifically asked about peer review, but instead if they valued the user comments about a preprint or even simple thumbs up/thumbs down

ratings by other users. Their low value of this cue could be because researchers didn't equate this type of open peer-review with the formal process conducted by a journal.

In the discussion and studies around preprints, there has been little consideration to date about whether students recognize or contextualize preprints and how that influences their information choices and evaluations. Disintermediated search tools like Google and Google Scholar are ubiquitous in the search for information, including academic research, and preprints are discoverable in the search results, especially in Google Scholar. In recognition of the growing importance of preprints, the National Institutes of Health (NIH) launched a preprint pilot wherein preprints from NIH-funded studies would be ingested into PubMed Central and consequently the PubMed database, and then picked up by Google Scholar.<sup>20</sup> The project took significant pains to ensure preprints were clearly labeled both visually and within the bibliographic record metadata. However, there is no evidence whether clear labelling affects recognition and use of preprints. The Researching Students' Information Choices (RSIC) study begins to address the gap in preprint research by examining students' perceptions of preprints. Using data from our study of students searching Google for a school research assignment, this paper will highlight how students perceive preprints and whether they can distinguish them from other publication types such as peer-reviewed journal articles.

## Methods

The data for this paper were collected from 116 students in the RSIC study (IRB Protocol #2016-U-0131) from four educational cohorts. These cohorts were high school (n=26) and three post-secondary cohorts: community college (n=30), undergraduate (n=30), and graduate (n=30). Each student completed an initial demographic questionnaire, followed by a facilitated task-based session using a simulated search engine environment. The simulated search engine environment allowed students to navigate and interact in a realistic way with a controlled set of Google search results and captured their real-time clicks and decisions. A think-aloud protocol was used during the task-based simulation to capture what students were thinking during each task. The think-aloud audio files were transcribed and a codebook was created to capture the cues and common judgments. Each student's transcript was coded in the NVIVO software. Cataldo et al. details the full research methodology.<sup>21</sup>

In the task-based sessions, students were presented with a research assignment appropriate for their educational cohort about the Burmese python's impact on the Florida Everglades. They started the simulation by performing a Google search. Then they completed five tasks using the simulated search results. They were asked to determine the resources' helpfulness, citability, credibility, and container identity, in that order. Helpfulness encompassed two tasks: one in which they selected resources they deemed helpful for their research assignment, and one in which they described why the unselected resources were not helpful. Students only judged citability and credibility on the resources they found helpful. Students used a yes-no option to decide whether their helpful resources were citable and a scale of 1 (not credible) to 5 (highly credible) to rate credibility. For the container task, students in the same cohort were shown the same set of resources, whether they found them helpful or not. The students were presented with eight containers to choose from, and preprint was one of the choices. The others were blog, book, conference proceeding, journal, magazine, news, and website. The students were not given definitions for the containers. A short video demonstration of the simulation session can be viewed at <http://ufdc.ufl.edu/IR00010570/00001/video?search=rsic>.

**FIGURE 1**  
**bioRxiv Google Snippet and Front Page**

**Building realistic assemblages with a Joint Species ... - bioRxiv**  
[biorxiv.org/content/early/2014/04/09/003947](https://www.biorxiv.org/content/early/2014/04/09/003947)  
 by DJ Harris – 2014 – Cited by 1  
 Apr 9, 2014 – Building realistic assemblages with a Joint Species Distribution Model. View ORCID Profile David J. Harris. doi: <http://dx.doi.org/10.1101/> ...

The screenshot shows the bioRxiv front page for the article. At the top, there is a navigation bar with links for HOME, ABOUT, SUBMIT, and ALERTS / RSS. Below this is a search bar and a 'Preview PDF' button. The main content area displays the article title, author name (David J. Harris), and DOI. A notice states: 'This article is a preprint and has not been peer-reviewed [what does this mean?]'.

**Abstract**

1. Species distribution models (SDMs) can be used to predict how individual species---and whole assemblages of species---will respond to a changing environment. Until now, these models have either assumed (1) that species' occurrence probabilities are uncorrelated, or (2) that species respond linearly to preselected

**Subject Area**

Ecology

**Subject Areas**

All Articles

This paper analyzes qualitative and quantitative data pertaining to two preprint articles included in the RSIC study and discoverable in the simulated Google search results. Both articles were from preprint servers, a bioRxiv preprint article (hereafter referred to as bioRxiv) and a PeerJ preprint article (hereafter referred to as PeerJ). A few visual details about each resource will be helpful in understanding the students' perceptions. The images were captured for the study in 2017. The bioRxiv Google snippet had no mention of preprint while the PeerJ snippet had three. The bioRxiv front page had two preprint cues and one peer-reviewed cue (figure 1). PeerJ had seven preprint cues and three peer-reviewed cues (figure 2). In addition, one of the seven PeerJ peer-reviewed cues linked to the peer-reviewed version of the article that students could open, and another cue noted users could "browse peer-reviewed articles instead." Students were not able to click to a peer-reviewed version of the article for bioRxiv as none existed.

## Results

BioRxiv appeared in the search results for the post-secondary students (n=90), while PeerJ appeared for the post-secondary and high school students (n=116). Overall, our findings show few mentions of preprint regardless of the educational cohort, even though there were more preprint cues than not-peer-reviewed cues. This may be due in part to students not knowing

**FIGURE 2**  
PeerJ Google Snippet and Front Page

Potential problems of removing one invasive species at a time - PeerJ  
<https://peerj.com/preprints/1651/>  
 by SA Ballari – 2016  
 Jan 15, 2016 – Potential problems of removing one invasive species at a time: ... PeerJ  
 PrePrints 4:e1651v11 <https://doi.org/10.7287/peerj.preprints.1651v1>

PeerJ, Inc. [US] | <https://peerj.com/preprints/1651/>

PeerJ go to journal home ▾ ARTICLES PREPRINTS More ▾ SUBMIT ARTICLE Login Search

**NOT PEER-REVIEWED**  
 "PeerJ Preprints" is a venue for early communication or feedback before peer review. Data may be preliminary.  
 Learn more about preprints or browse peer-reviewed articles instead.

Download Follow preprint

See PeerJ's Benefits ↕

Or Sign up for free and we'll keep you up to date on the latest fee waiver offers and research.

**Potential problems of removing one invasive species at a time: Interactions between invasive vertebrates and unexpected effects of removal programs**

Animal Behavior Biodiversity Conservation Biology Ecology Zoology

Sebastián A Ballari<sup>1</sup>, Sara E Kuebbing<sup>2</sup>, Martin A Nuñez<sup>3</sup>  
 January 15, 2016

A peer-reviewed article of this Preprint also exists

Meta  
 Citations in Google Scholar  
 Feedback  
 Questions  
 Links  
 Visitors 200  
 Views 244  
 Downloads 117  
 Outline

what preprint means. Few students made the connection that part of what makes a preprint is that it has not been peer-reviewed. Some students did not mention preprint until it was given as a choice in the container task. The same can be said of the not-peer-reviewed cue, which was mentioned most frequently during the credibility task.

### *Helpfulness*

Results show that most students did not choose bioRxiv or PeerJ as helpful. Fifteen of the ninety post-secondary students (17%) chose bioRxiv as helpful, whereas twenty-eight of the 116 post-secondary and high school students (24%) chose PeerJ as helpful. Very few students mentioned the articles' preprint cues during this task. More students attended to whether the articles were peer-reviewed when evaluating PeerJ (25%) than when assessing bioRxiv (7%). Although the majority of students who attended to the not-peer-reviewed cue concluded that the lack of peer review made the articles less desirable, it was not used as their primary determinant when deciding whether the articles were helpful.

### *bioRxiv*

Students who did not select bioRxiv as helpful (n=75/90) did so for a variety of reasons. The top three were lack of relevance, its conceptual focus (i.e., aboutness), and lack of keywords of

interest within the resource. One student expressed “This one also looks good, but it doesn’t really talk about the Burmese pythons anywhere in this little info” (community college 22). Only a few mentioned cues associated with preprint and/or peer review. Some were confused by the term preprint because they thought the article was from a journal: “I actually don’t know what this is.... I’d have to look if this is a journal” (graduate 16).

Some noticed bioRxiv was not peer-reviewed and used that as their sole reason for not choosing it, but this was rare. More often students used it as one of several reasons for not selecting it. Other reasons were that it was too complicated or not necessary given other journal articles already chosen.

I did not use this one because I didn’t know what that was.... Some sort of a preprint server for biology. Some sort of laboratory. But it’s not saying that it’s peer-reviewed or anything, so I’m probably going to skip it because I have my PLoS ONE and I have my Springer (community college 11).

No clear theme emerged among the students who selected bioRxiv as helpful, though reasons for selecting it loosely revolved around the content, such as its relevance or aboutness.

Building realistic assemblages with a joint species. I would possibly check on this one. It probably provides a solution. What happens after an invasive species enters the biodiversity of the Florida Everglades (undergraduate 1).

Very few noticed that it was a preprint and not peer-reviewed. Interestingly, those that did were not deterred, because they were familiar with preprints. One graduate student equated bioRxiv to arXiv and concluded it was a research paper. Another contextualized preprints by the peer-reviewed process that was expected to follow.

They’re preprints, so they’re not peer-reviewed yet. But, ideally, it says that it’s been accepted for publication somewhere but I don’t see anything saying that. So it’s a little questionable but that’s okay (graduate 26).

## PeerJ

The students’ reasons for not choosing PeerJ as helpful (n=88/116) were similar to those given for bioRxiv (e.g., lack of relevance, its aboutness, and lack of matching keywords). Others also mentioned not choosing it because they did not recognize it.

I don’t know what this is, either. PeerJ.com. I mean, it could be a journal article. I see DOI on here (undergraduate 12).

Similar to bioRxiv, few students mentioned that PeerJ was a preprint. However, 25% mentioned cues associated with peer review. A few simply mentioned the term peer review. Others specifically mentioned the not-peer-reviewed cue and quickly rejected the article: “Not peer-reviewed. Goodbye” (high school 7). Others were more tentative but questioned the credibility of the resource based on the lack of peer review: “I don’t know if that’s going to be credible

information at the moment" (graduate 12). Some who mentioned the preprint and not-peer-reviewed cues in conjunction rendered similar judgements: "This preprint is not peer-reviewed. So probably not a reputable source" (community college 5). A few others provided explanations that described their perceptions of the peer-review process, which were not always accurate.

peerj.com. Not peer-reviewed... and it even warned you, "Early communication or feedback before peer review...." Usually, peer review means that experiment was done over with similar results, you would think. I don't know. Not sure (community college 14).

Similar to bioRxiv, the majority of students who thought PeerJ was helpful based their decisions on article content. For instance, some thought it was interesting or could help them provide a counterargument for their project. A couple mentioned that the article was not peer-reviewed, but they drew different conclusions. One had second thoughts about using it: "So maybe I wouldn't use that" (undergraduate 2), whereas the other thought "it might still have good information" (community college 7). Only one graduate student mentioned both preprint and not-peer-reviewed cues: "That's interesting, not peer-reviewed. So this is another preprint but it's more recent. Oh, cool. Peer reviewed version exist" (graduate 26). This student and a couple of others who mentioned peer review noticed the link to the peer-reviewed version.

### *Citability*

Twelve of the fifteen post-secondary students (80%) who chose bioRxiv as helpful also thought it was citable. For PeerJ it was nineteen out of twenty-eight high school and post-secondary students (70%). Of those that decided not to cite the articles, only a few attributed the lack of peer review as the reason. None of the students mentioned the articles being from preprint servers as a reason.

### **bioRxiv**

No clear theme emerged as to why students found bioRxiv citable, though a few mentioned it was a "paper" and therefore could be cited. Others referred to different aspects of the article's content that made it citable. For instance: "it uses computation to look at if pythons are actually causing any problems in the ecosystem" (graduate 32) and "it gives me a different approach than most of them" (undergraduate 16). An undergraduate who said they would cite it was decidedly confused by the website and potential access:

I think this is a book, but the website. If it's a journal, if I don't have access to the full journal I wouldn't cite it, but if I did have access, if it's a journal or an article, then I would not cite the website, I would cite the article (undergraduate 26).

Although the three students who chose not to cite bioRxiv did not mention preprint as the reason, one did mention lack of peer review as a deciding factor. That student also noted they "didn't see that [not-peer-reviewed cue] earlier" (community college 32) during the helpful task.

## PeerJ

Although students could click on the peer-reviewed version of the article for PeerJ, none of them mentioned this when describing why they would cite it. Instead, similar to bioRxiv they discussed different aspects of the content.

I would also use this as the first one that I've seen, like I said earlier, about the problems of removing the species. Not just why we should remove the species. I'd use that because that's just new information (undergraduate 5).

Only a few who chose not to cite PeerJ mentioned the lack of peer review as the reason: "So I wouldn't use this just because, I mean, it says on the top it's not peer review" (undergraduate 2). Other reasons for not citing PeerJ were that they didn't think they had access, the content was not right, and they were not familiar with the source.

## *Credibility*

The students who chose bioRxiv (n=15) and PeerJ (n=28) as helpful also had to rate the articles' credibility. When rating the credibility of bioRxiv, very few students mentioned the preprint and not peer-reviewed cues. Both cues were more prominent for PeerJ and half of the students mentioned the not-peer-reviewed cue, but very few students mentioned the preprint cue. The average credibility rating dropped as a result with bioRxiv a 4.27 out of 5 and PeerJ only 3.28.

## bioRxiv

Most of the students rated bioRxiv's credibility highly for reasons typically associated with credibility. For instance, several mentioned that it was from a journal or was an article or paper. Others noted that it was research or mentioned the research approach used (e.g., distribution model, modeling process) as reasons for their credibility ratings. One who used a combination of these factors said "Since this is a paper from a laboratory. I mean research laboratory I would give it high credibility" (graduate 10).

Many students who did mention the preprint and/or not-peer-reviewed cues drew different conclusions and rated bioRxiv as somewhat credible. They also provided additional reasoning for their ratings, noting "it was cited by only one person, and it's been three years since they published it" (graduate 11) and "I don't know who the authors are... [and] it doesn't list their qualifications" (community college 32). Another student noticed the article's DOI link, assumed they could find a "peer-reviewed version of this paper" (graduate 26), and rated the article very credible as a result.

## PeerJ

Students attending to the not-peer-reviewed cue in PeerJ indicated this was the primary reason for the lower credibility rating. More than half of the students rated it between not credible and somewhat credible. A few questioned other aspects of the article such as the website, authors and their affiliation, and the journal. The students who rated it highly did so for reasons similar to the students who rated bioRxiv highly, noting it was a paper, article, or journal. A couple of others attributed their ratings to these and/or some aspect of the content, such as the supplemental information or the diagrams. In one case the student didn't even understand



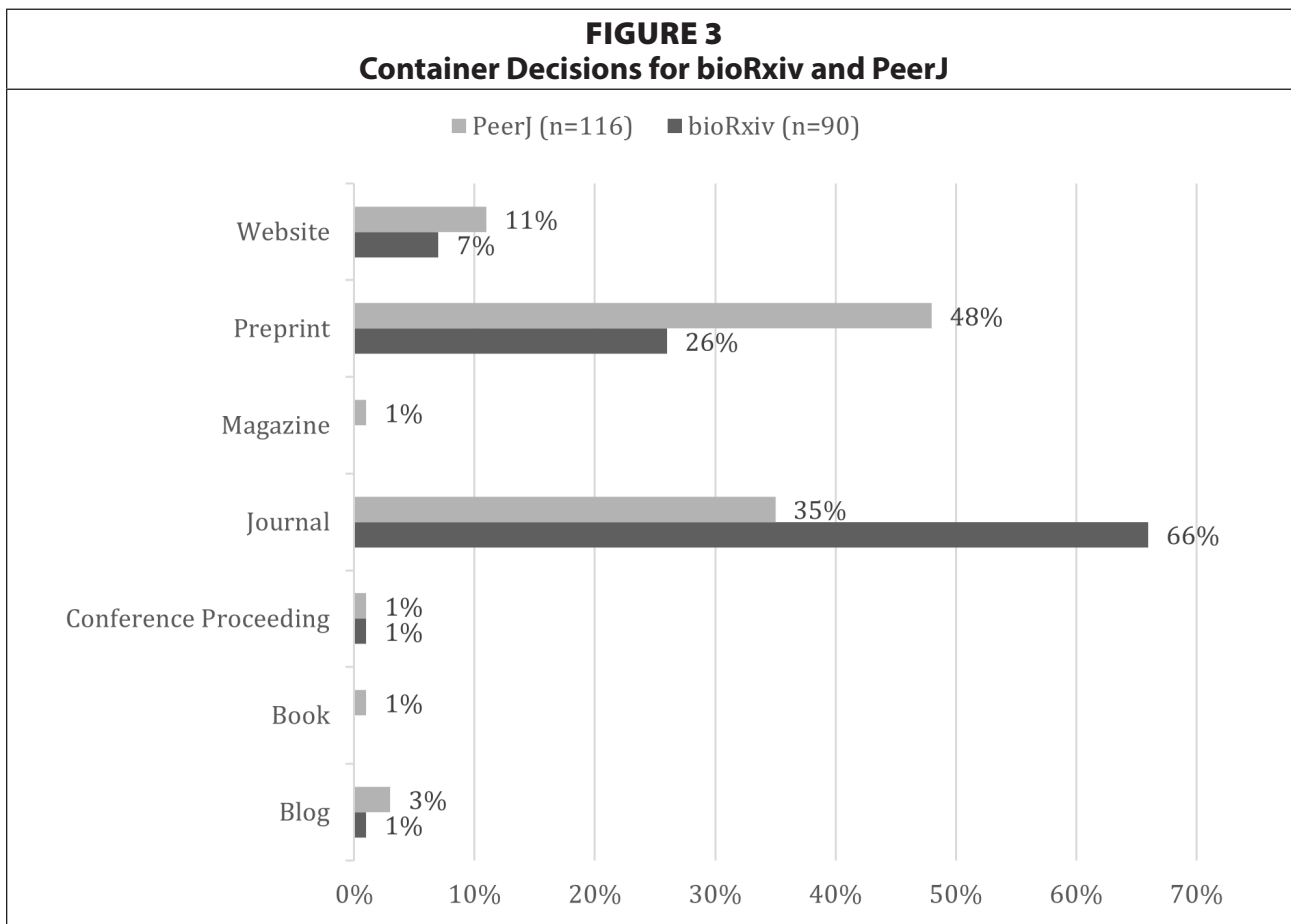
the diagrams but thought they were evidence of the authors' expertise.

Peer reviews [are] usually pretty good [and] once again the diagram and stuff which don't even make any sense to me to make [me] feel like they know what they are talking about (undergraduate 5).

One of the students who noticed that a peer-reviewed version of PeerJ existed during the cite task rated it very credible. This is the same student mentioned above who incorrectly assumed a peer-reviewed version of bioRxiv existed upon seeing the DOI. After comparing the two versions of PeerJ to identify changes made, the student concluded that the changes were minor but still wanted to cite the peer-reviewed version: "So this resource by itself is non-citable, but the peer-reviewed version is" (graduate 26).

**Container**

All the post-secondary students were presented with bioRxiv during the container task (n=90), and all post-secondary and high school students were presented with PeerJ (n=116). Students mentioned preprint the most during this task. Close to 30% of students (n=26/90) mentioned the preprint cue when discussing bioRxiv, and nearly 50% (n=57/116) did so when discussing PeerJ. These percentages roughly align with the percentage of students who correctly identified the articles as preprints in the task (figure 3). However, it cannot be ignored that at least half of the students incorrectly identified the articles' containers, with most picking journal.



### bioRxiv

The students who thought bioRxiv was contained within a journal reasoned that it mimicked the look and structure of a journal article. Some based this decision on it being an article: “This is an actual article. There’s actual research here. That’s a journal” (community college 26). It looked like a research or scientific paper—“Looking at it, it’s a scientific paper. It’s from a journal. That’s all I know” (graduate 27), and was associated with a lab: “Cold Spring Harbor Laboratory, okay, this looks like a journal” (undergraduate 3).

Others based this decision on seeing abstract and methodology sections or charts and diagrams: “So this is a journal for sure, because it has proper structure, like abstract and main content” (graduate 31). Another used the DOI as an indicator: “Well, it has a DOI, so. So, I guess, that’s a journal” (graduate 16).

Even those who chose preprint did not come by the decision easily. Attending to the preprint cue helped redirect some students.

Oh, Cold Spring Harbor. So yes, I would. This is a journal.... Well, it’s Cold Spring Harbor so that’s actually—I’m from New York so this is—I know the research place this is done. And I’m just looking to see if this is journal or—oh, this says preprint on it so I guess that’s a preprint (community college 31).

Some students who attended to the preprint cue did not always know what it meant or how it differed from the other options they were considering.

The preprint server for biology... I’m going to say that this is a blog. Although I don’t think it’s a blog [laughter]... I’m very confused as to what this is. It could be a blog. Book, conference proceeding. I don’t think so. Preprint? I don’t even know what a preprint is. I don’t think it’s a journal. So what would I do? I’m going to go with preprint. I saw a PDF version (community college 6).

BioRxiv. What’s that? This looks like a journal. It’s an abstract. [silence] It says this article [is] a preprint and has not been peer-reviewed. What does this mean, so that would say that it’s not a journal. I guess website (graduate 25).

### PeerJ

Similarly, some students simply used the preprint cue in PeerJ to correctly identify the article, regardless of whether they knew what it meant. Some who also attended to the not-peer-reviewed cue used it to contextualize the term preprint.

It’s a preprint; it says it right there. Not peer-reviewed. I guess that’s what preprint is (undergraduate 10).

I don’t know what a preprint is. Does that mean that it’s not peer-reviewed [laughter]? Assuming, because it says preprint [laughter]. So I’m just going to assume that that’s that [laughter] (graduate 15).

For others however, the not-peer-reviewed cue, when used in conjunction with the “PeerJ go to journal home” cue caused confusion.

Potential problems of removing one invasive species at a time. So this is an article. Let's see. So author and article information. Published, where was it published? PeerJ preprints. So I guess it's still... I mean, it's saying preprints, but I think it's still a journal because I think PeerJ preprints is the journal because it has the 4:e1651v1. So I think it's still a journal. And it says, “PeerJ, go to journal home” (graduate 3).

I don't even want to put journal for that because it's not peer-reviewed. But I mean, it just says journal so I'm just going to (community college 11).

## Discussion

### *Implications for Practice*

Our findings indicate familiarity with cues and how they relate can influence whether and how they are used. The term preprint was unfamiliar to most students who encountered it, which likely contributed to its limited use across tasks. Mentioned the most during the container task, the students who correctly labeled bioRxiv and PeerJ as preprints did so by matching the articles' cues to the label options, often asking what a preprint was or saying they were unfamiliar with the term. These findings suggest several implications for practice.

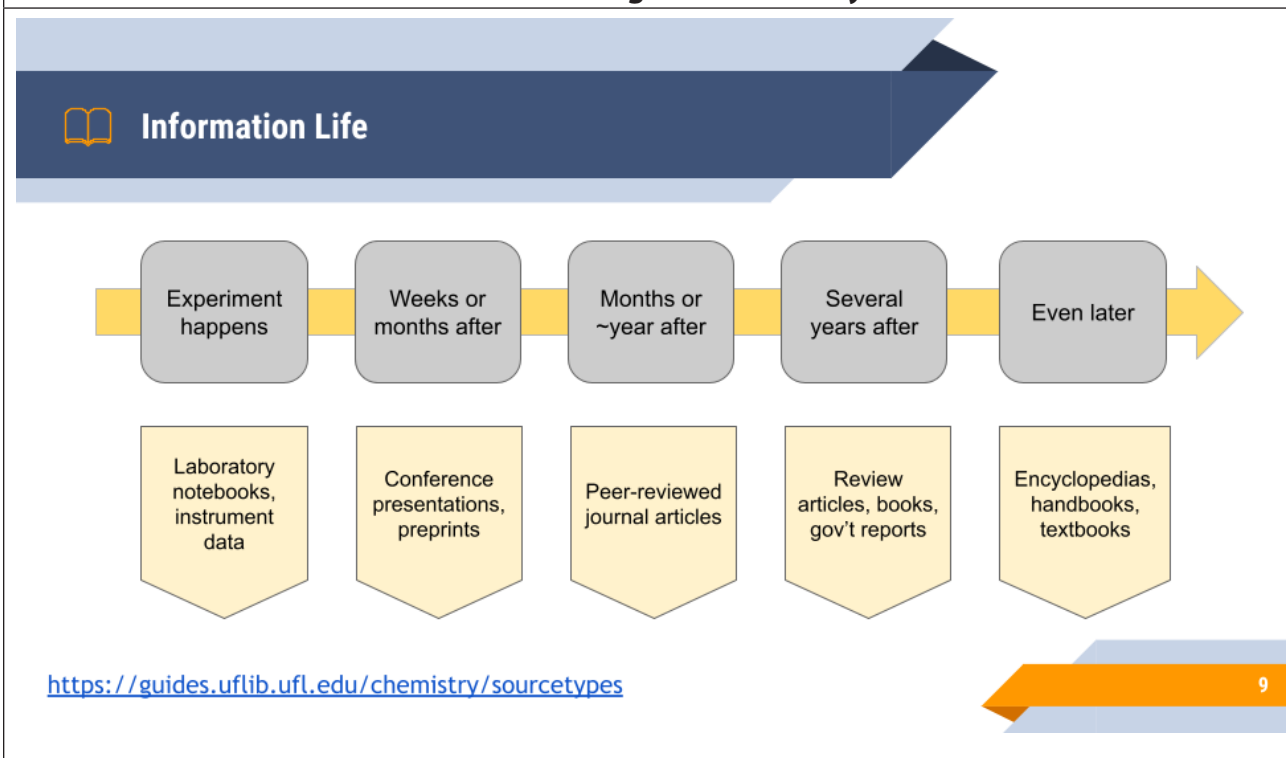
First, in order to use cues, students must know what they mean. Many students treated bioRxiv and PeerJ as journal articles, because they looked like published journal articles despite cues to the contrary. The scholarly community recognizes that it is important for preprints to be labeled clearly so their status as a non-peer-reviewed publication is explicit. Ravinetto et al. (2021) make very specific recommendations, including statements that not-peer-reviewed works should be on the first page of preprints, and each page should include a red watermark stating “Caution – Not Peer Reviewed.”<sup>22</sup> This is work for the content providers, but changes also need to happen in the educational settings. Given the increasing availability of freely accessible preprint articles on the internet and debates about their use, information literacy instruction needs to build awareness about the existence of preprints, what they do and do not provide, and guidelines for their use. For example, when teaching PubMed, the new filter for preprints can be demonstrated and can open up a discussion on preprints. Additionally, preprints can be woven into instruction on citation styles and creating bibliographies. Organizations like NIH require preprint citations to be labeled as such and calls for more standardization are increasing, but to date, preprints are not showing up as a format in most citation management software. This makes preprints a good example of when you have to create or edit a citation manually.

Second, the cues students mention and use relate to the type of evaluation. Relevance was the primary factor for most students when deciding whether the articles were helpful and citable, aligning with some of the findings by Komissarov and Rempel.<sup>23</sup> There were few mentions of the preprint and not-peer-reviewed cues during these tasks. Students mentioned the not-peer-reviewed cue most during the credibility task and many of these students knew what it meant, which impacted their ratings. This was particularly true for PeerJ because the cue was displayed more prominently. Yet, this prominently placed cue was only mentioned

and used by half of the students. These findings suggest information literacy instruction should build awareness even for well-established cues such as peer review and provide a more nuanced discussion of the circumstances under which the cues should be used, especially in an information landscape that is growing in complexity. Discussions incorporating preprint and peer review cues could include lectures on the information life cycle (figure 4) or lessons that introduce the use of Google Scholar that show the versioning of articles, including preprints.<sup>24</sup> Another way to help students contextualize preprints within the information lifecycle is through the use of experiential learning like the scaffolded approach employed by Scheifele, Tsoதாக, and Wolyniak (2021) where students compared and contrasted a preprint article with the final version, reflecting upon the differences.<sup>25</sup>

Third, the helpfulness of a resource goes beyond relevant content. For instance, a preprint article that has not been peer-reviewed may be helpful if it cites relevant resources, whereas a peer-reviewed journal article that has been retracted is likely to be viewed as not helpful. Citable resources on the other hand are expected to be credible as well as relevant. Yet students did not consider the credibility of resources during the cite task. Moreover when they did evaluate the credibility, findings showed students who did not use the not-peer-reviewed cue evaluated the preprints like journal articles, because the preprints presented research conducted in labs, data analytics techniques, and diagrams. In other words, they relied on what Rempel et al. called superficial criteria.<sup>26</sup> Information literacy instruction can help students go beyond choosing and using resources that look relevant, citable, and credible on the surface by providing them with ways to critically examine and correctly apply the content within the context of their educational stage. This includes providing guidelines for students that help them determine whether or not they have the expertise necessary to vet and correctly use a resource and choosing not to use it if they do not.

**FIGURE 4**  
**Information Life as Taught to Chemistry Students**



## Changes in bioRxiv and PeerJ

The challenges that our findings illuminate need to be addressed not only by educators but also by those creating and hosting preprints. Since some time has passed since the preprint images used in the simulations were captured, it is worth discussing what, if anything, has changed. New images were captured on December 2, 2021. The bioRxiv Google snippet has changed slightly but has not added any preprint or not-peer-reviewed cues. PeerJ has two Google snippets, one for the HTML and one for the PDF, and these changed more significantly. They both went from having three preprint cues to only two, but the PDF snippet added a peer-reviewed cue where there were none before. On its webpage, BioRxiv now has a large, yellow box above the article which focuses on COVID-19 and emphasizes that the articles “have not been formally peer-reviewed” (figure 5). The statement adds one more peer-reviewed cue to the page, bringing it to a total of two. Otherwise, the article appears the same. PeerJ’s webpage reduced the number of preprint cues from seven to five and increased its peer-reviewed cues from two to four (figure 6). This movement towards emphasizing peer review instead of preprint is a positive one as the students in our study were more likely to attend to and understand peer review as opposed to preprint.

## Limitations and Future Research

The task-based simulation used for this study is not without limitations, the details of which can be found in Cataldo et al.<sup>27</sup> A couple of limitations are discussed here in the context of future research. First, the simulation provided a controlled environment that allowed for comparisons within and across student cohorts, but created a level of artificiality.<sup>28</sup> Part of this

**FIGURE 5**  
**bioRxiv Google Snippet and Front Page in 2021**

The screenshot displays the bioRxiv website interface. At the top, a Google snippet shows the URL <https://www.biorxiv.org/content> and the article title "Building realistic assemblages with a Joint Species ... - bioRxiv". Below the snippet, the article's date "Apr 9, 2014" and a brief description are visible. The main content area shows the bioRxiv logo and a prominent yellow warning box: "bioRxiv posts many COVID-19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive." The article title "Building realistic assemblages with a Joint Species Distribution Model" is displayed, along with the author's name "David J. Harris" and the DOI "https://doi.org/10.1101/003947". A note states "This article is a preprint and has not been certified by peer review (what does this mean?).". The page includes navigation links, a search bar, and social media sharing options. A sidebar on the right contains a "Subject Area" section.

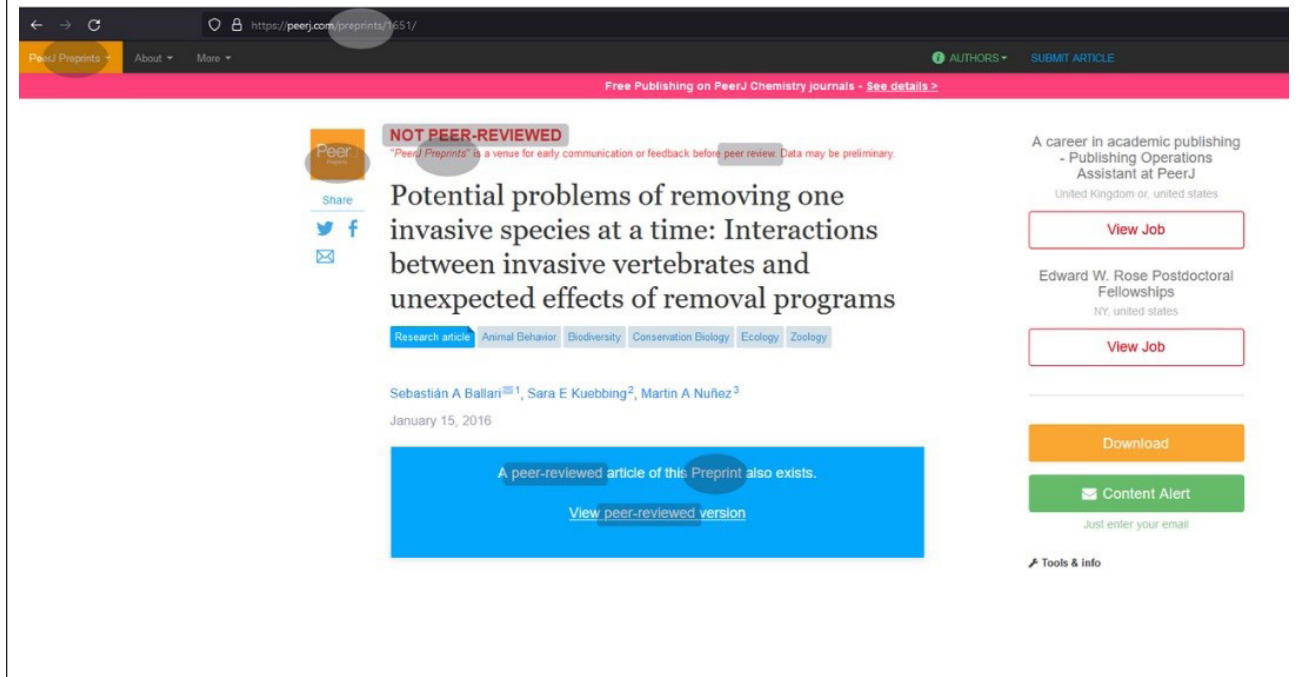
**FIGURE 6**  
PeerJ Google Snippets and Front Page in 2021

https://peerj.com > preprints ⋮

**Potential problems of removing one invasive species at a time**  
by SA Ballari · 2016 · Cited by 5 — "PeerJ Preprints" is a venue for early communication or feedback before ... **Potential problems of removing one invasive species at a time:** ...

https://peerj.com > preprints PDF ⋮

**Potential problems of removing one invasive species at a time**  
by SA Ballari · 2016 · Cited by 5 — A peer-reviewed version of this preprint was published in PeerJ ... **Potential problems of removing one invasive species at a time:** ...  
22 pages



Free Publishing on PeerJ Chemistry journals - See details >

**NOT PEER-REVIEWED**  
"PeerJ Preprints" is a venue for early communication or feedback before peer review. Data may be preliminary.

Share  
Twitter  
Facebook  
Email

**Potential problems of removing one invasive species at a time: Interactions between invasive vertebrates and unexpected effects of removal programs**

Research article · Animal Behavior · Biodiversity · Conservation Biology · Ecology · Zoology

Sebastián A Ballari<sup>1</sup>, Sara E Kuebbing<sup>2</sup>, Martin A Nuñez<sup>3</sup>  
January 15, 2016

A peer-reviewed article of this Preprint also exists.  
[View peer-reviewed version](#)

Download  
Content Alert  
Just enter your email

Tools & info

A career in academic publishing  
- Publishing Operations  
Assistant at PeerJ  
United Kingdom or, united states  
[View Job](#)

Edward W. Rose Postdoctoral  
Fellowships  
NY, united states  
[View Job](#)

artificiality is that students are not likely to separate and order their evaluation of resources by task (e.g., helpful, citable, credible). They also are not likely to evaluate resources only during point of selection, but rather reserve and/or revisit their evaluations as they engage with resources over the course of their research assignments. Future research in a naturalistic setting that collects longitudinal data over the course of a research assignment would balance the findings from our task-based simulation study. This would provide an understanding about whether students' information behavior changes over the course of their research assignments and if so, when, how, and why. In addition, the simulation did allow students to click and explore resources, but not all of the links within and outside of the resources were made clickable in order to manage simulation design complexity. Research that examines

students' interactions with preprints through eye tracking or heat mapping software to see what they are looking at in addition to what they are saying could serve to complement this work. Providing these kinds of insights would create a holistic picture of students' behaviors in relation to the discovery, analysis, and use of preprints.

## Conclusion

The role of preprints in scholarly communication continues to evolve. Publishers are reviewing and revising their policies. Funding agencies are determining the appropriateness of their appearance in grant applications, with some for (NIH) and some against (Australian Research Council).<sup>29</sup> Members of standards organizations are calling for revisions to recommended practices (the National Information Standards Organization's 2008 Journal Article Versions Recommended Practices<sup>30</sup>). Conversations are happening within professional organizations, especially those involved in open access such as the Open Access Scholarly Publishing Association.<sup>31</sup> However, these conversations are rather siloed, since librarians, publishers, and the hosts of repositories have one type and researchers and editors have another. Our findings on students' judgments of preprints suggest that a more collaborative approach is needed to make preprints more recognizable and understood. Given the importance of preprint servers as an openly available source of scientific information, publishers, scholars, librarians, journalists and teachers should work together to ensure standardized, universal approaches. The perceptions of students, as seen in this study, can help illuminate the way.

## Acknowledgments

The authors wish to acknowledge the RSIC research team and advisory panel, which made this work possible.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Institute of Museum and Library Services [grant number LG-81-15-0155].

---

## Notes

1. COPE Council, "COPE Discussion Document: Preprints" (2018). <https://doi.org/10.24318/R4WByao2>
2. Nicholas Fraser et al., "The Evolving Role of Preprints in the Dissemination of COVID-19 Research and Their Impact on the Science Communication Landscape," *Plos Biology* 19, no. 4 (2021): e3000959, <https://doi.org/10.1371/journal.pbio.3000959>.
3. Andrea Chiarelli et al., "Preprints and Scholarly Communication: An Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers [Version 2; Peer Review: 3 Approved, 1 Approved with Reservations]," *F1000Research* 8 (2019): 971, <https://doi.org/10.12688/f1000research.19619.2>; Annette Flanagan, Phil B. Fontanarosa, and Howard Bauchner, "Preprints Involving Medical Research—Do the Benefits Outweigh the Challenges?" *JAMA* 324, no. 18 (2020): 1840, <https://doi.org/10.1001/jama.2020.20674>; Claire Hopkins, "Preprints-Expediting Access or Compromising Quality?" *International Forum of Allergy & Rhinology* 11, no. 5 (2021), <https://doi.org/10.1002/alr.22732>.
4. Anderson, Rick, "Journalism, Preprint Servers, and the Truth: Allocating Accountability," *The Scholarly Kitchen* (blog), December 14, 2020, <https://scholarlykitchen.sspnet.org/2020/12/14/journalism-preprint-servers-and-the-truth-allocating-accountability/>; Alice Fleerackers et al., "Communicating Scientific Uncertainty in an Age of COVID-19: An Investigation into the Use of Preprints by Digital Media Outlets," *Health Communication*, (2021): 1–13, <https://doi.org/10.1080/10410236.2020.1864892>. Massarani, Luisa, and Luiz Felipe Fernandes Neves. "Reporting COVID-19 Preprints: Fast Science in Newspapers in the United States, the United Kingdom and Bra-

zil." *Ciência & Saúde Coletiva* 27 (M2022): 957–68. <https://doi.org/10.1590/1413-81232022273.20512021> ; Schalkwyk, François van, and Jonathan Dudek. "Reporting Preprints in the Media during the COVID-19 Pandemic." *Public Understanding of Science*, (2022), <https://doi.org/10.1177/09636625221077392> ..

5. Jaime A. Teixeira da Silva and Judit Dobránszki, "Preprint Policies among 14 Academic Publishers," *The Journal of Academic Librarianship* 45, no. 2 (2019): 162–70, <https://doi.org/10.1016/j.acalib.2019.02.009>; Jan Higgins and Robert D. Steiner, "Author Preprint Behaviour and Non-Compliance with Journal Preprint Policies: One Biomedical Journal's Experience," *Learned Publishing* (2021), <https://doi.org/10.1002/leap.1376>; Kazuki Ide et al., "Guidelines Are Urgently Needed for the Use of Preprints as a Source of Information," *Journal of Epidemiology* 31, no. 1 (2021): 97–9, <https://doi.org/10.2188/jea.JE20200506>; Dorothy S. Massey et al., "Assessment of Preprint Policies of Top-Ranked Clinical Journals," *JAMA Network Open* 3, no. 7 (2020): e2011127, <https://doi.org/10.1001/jamanetworkopen.2020.11127>.

6. Evan D. Kharasch, "Non–Peer-Reviewed Preprint Articles as References in Anesthesiology: Reply," *Anesthesiology* 134, no. 5 (2021): 821, <https://doi.org/10.1097/ALN.0000000000003715>.

7. "The Editorial Process," eLife Author Guide (n.d.), <https://reviewer.elifesciences.org/author-guide/editorial-process>.

8. Thomas Klebel et al., "Peer Review and Preprint Policies Are Unclear at Most Major Journals," *PLoS One* 15, no. 10 (2020): e0239518, <https://doi.org/10.1371/journal.pone.0239518>.

9. Clarissa F. D. Cameiro et al., "Comparing Quality of Reporting between Preprints and Peer-Reviewed Articles in the Biomedical Literature," *Research Integrity and Peer Review* 5, no. 1 (2020): 16, <https://doi.org/10.1186/s41073-020-00101-3>; Martin Klein et al., "Comparing Published Scientific Journal Articles to Their Pre-Print Versions," *International Journal on Digital Libraries* 20, no. 4 (2019): 335–50, <https://doi.org/10.1007/s00799-018-0234-1>; Lisa Bero et al., "Cross-Sectional Study of Preprints and Final Journal Publications from COVID-19 Studies: Discrepancies in Results Reporting and Spin in Interpretation," *BMJ Open* 11, no. 7 (2021): e051821, <https://doi.org/10.1136/bmjopen-2021-051821>; Mario Pagliaro, "Preprints in Chemistry: An Exploratory Analysis of Differences with Journal Articles," *Publications* 9, no. 1 (2021): 5, <https://doi.org/10.3390/publications9010005>; Xiaoting Shi et al., "Assessment of Concordance and Discordance among Clinical Studies Posted as Preprints and Subsequently Published in High-Impact Journals," *JAMA Network Open* 4, no. 3 (2021): e212110, <https://doi.org/10.1001/jamanetworkopen.2021.2110>; Fang Xu et al., "The Consistency of Impact of Preprints and Their Journal Publications," *Journal of Informetrics* 15, no. 2 (2021): 101153, <https://doi.org/10.1016/j.joi.2021.101153>; Liam Brierley, Federico Nanni, Jessica K. Polka, Gautam Dey, Máté Pálffy, Nicholas Fraser, and Jonathon Alexis Coates, "Tracking Changes between Preprint Posting and Journal Publication during a Pandemic." *PLOS Biology* 20, no. 2 (2022): e3001285. <https://doi.org/10.1371/journal.pbio.3001285> .

10. "Retracted Coronavirus (COVID-19) Papers," *Retraction Watch* (blog), April 29, 2020, <https://retractionwatch.com/retracted-coronavirus-covid-19-papers/>.

11. Vincent Larivière et al., "ArXiv E-Prints and the Journal of Record: An Analysis of Roles and Relationships," *Journal of the Association for Information Science and Technology* 65, no. 6 (2014): 1157–69, <https://doi.org/10.1002/asi.23044>.

12. Mario Pagliaro, "Did You Ask for Citations? An Insight into Preprint Citations En Route to Open Science," *Publications* 9, no. 3 (2021): 26, <https://doi.org/10.3390/publications9030026>.

13. Kathleen Hutchinson Flynn, "Citation Analysis of Mathematics and Statistics Dissertations and Theses from the University at Albany," *Science & Technology Libraries* 39, no. 2 (2020): 142–54, <https://doi.org/10.1080/0194262X.2020.1727396>.

14. Daniel S. Dotson and Tina P. Franks, "Winning the Popularity Contest: Researcher Reference When Selecting Resources for Civil Engineering, Computer Science, Mathematics and Physics Dissertations," *Issues in Science and Technology Librarianship*, no. 81 (2015), <https://doi.org/10.5062/F4N29TZK>.

15. Sloan Komissarov and James Murray, "Factors That Influence Undergraduate Information-Seeking Behavior and Opportunities for Student Success," *Journal of Academic Librarianship* 42, no. 4 (2016): 423–29, <https://doi.org/10.1016/j.acalib.2016.04.007>.

16. Hannah Gascho Rempel, Stefanie Buck, and Anne-Marie Deitering, "Examining Student Research Choices and Processes in a Disintermediated Searching Environment," *Portal: Libraries & the Academy* 13 no. 4 (2013): 363–84, <https://doi.org/10.1353/pla.2013.0036>.

17. Greta Kliewer et al., "Using Primo for Undergraduate Research: A Usability Study," *Library Hi Tech* 34, no. 4 (2016): 566–84, <https://doi.org/10.1108/LHT-05-2016-0052>.

18. Barbara Blummer, Sara Lohnes Watulak, and Jeffrey Kenton, "The Research Experience for Education Graduate Students: A Phenomenographic Study," *Internet Reference Services Quarterly* 17, no. 3–4 (2012): 117–46, <https://doi.org/10.1080/10875301.2012.747462>.

19. Courtney K. Soderberg, Timothy M. Errington, and Brian A. Nosek, "Credibility of Preprints: An Inter-



- disciplinary Survey of Researchers," *Royal Society Open Science* 7, no. 10 (2020): 201520, <https://doi.org/10.1098/rsos.201520>.
20. "NIH Preprint Pilot," (2021), <https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/>.
21. Tara Tobin Cataldo et al., "Mixed Methods Data Collection Using Simulated Google Results: Reflections on the Methods of a Point-of-Selection Behaviour Study," *Information Research: An International Electronic Journal* 25, no. 4 (2020), <https://doi.org/10.47989/irpaper881>.
22. Raffaella Ravinetto et al., "Preprints in Times of COVID19: The Time Is Ripe for Agreeing on Terminology and Good Practices," *BMC Medical Ethics* 22, no. 1 (2021): 106, <https://doi.org/10.1186/s12910-021-00667-7>.
23. Sloan Komissarov and James Murray, "Factors That Influence Undergraduate Information-Seeking Behavior and Opportunities for Student Success," *Journal of Academic Librarianship* 42, no. 4 (2016): 423–29, <https://doi.org/10.1016/j.acalib.2016.04.007>.
24. Michelle Nolan, *Information Life*, (2021). <https://guides.uflib.ufl.edu/chemistry/sourcetypes>.
25. Lisa Z. Scheifele, Nikolaos Tsoakos, and Michael J. Wolyniak, "Preprint Articles as a Tool for Teaching Data Analysis and Scientific Communication," *PLOS ONE* 16, no. 12 (2021): e0261622, <https://doi.org/10.1371/journal.pone.0261622>.
26. Hannah Gascho Rempel, Stefanie Buck, and Anne-Marie Deitering, "Examining Student Research Choices and Processes in a Disintermediated Searching Environment," *Portal: Libraries & the Academy* 13, no. 4 (2013): 363–84, <https://doi.org/10.1353/pla.2013.0036>.
27. Cataldo et al., "Mixed Methods Data Collection Using Simulated Google Results."
28. Amy G. Buhler et al., "How Real Is Real Enough?: Participant Feedback on a Behavioral Simulation Used for Information-Seeking Behavior Research," *Journal of Librarianship & Information Science*, (2022) <https://doi.org/10.1177/09610006211067799>.
29. National Institutes of Health, "Reporting Preprints and Other Interim Research Products," 2017, <https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html>; Clare Watson, "Preprint Ban in Grant Applications Deemed 'Plain Ludicrous,'" *Nature*, August 25, 2021, <https://doi.org/10.1038/d41586-021-02318-8>.
30. "NISO Voting Members Approve Work to Update Journal Article Versions (JAV) | NISO Website," (2021), <http://www.niso.org/press-releases/2020/11/niso-voting-members-approve-work-update-journal-article-versions-jav>.
31. "OASPA | Open Access Scholarly Publishers Association," (2021), <https://oaspa.org/>.