11-2022

Optimizing a Library Website for Student Research: Comparing User Metrics Between Encore and Google Scholar

Lindsay Ozburn
Utah State University

Ryan Bushman
Utah State University

Kade Stevenson
Utah State University

Margaret Winward
Utah State University

Liz Woolcott
Utah State University

Optimizing a Library Website for Student Research: Comparing User Metrics between Encore and Google Scholar

Lindsay Ozburn, Ryan Bushman, Kade Stevenson, Margaret Winward, and Liz Woolcott
Utah State University, USA

Abstract

The paper addresses the methods and general conclusions portion of an experiment that evaluated user preference and search experience using Google Scholar and Utah State University (USU) Libraries' Encore discovery layer as a starting point for research. USU's 2019 Ithaka S+R Faculty survey highlighted that our faculty utilize Google Scholar more as a starting point for their research. To triangulate these findings, the experiment attempts to identify which search methods undergraduates prefer.

Introduction

In pursuit of a search environment balancing the needs of users and the viability of available vendor solutions, Utah State University (USU) Libraries has, for several years, researched the efficacy and structure of the Libraries' existing discovery layers. Tension is manufactured when user expectation, conditioned by Google's simplified search environment, meets the aggravatingly complex and jerry-rigged library discovery systems. It creates a continuous rotation of discussions among the Libraries' discovery committees about whether updating, replacing, or maintaining existing systems is the best course of action. This research project attempts to investigate one of the primary questions in that rotation: What would happen if the Libraries did not provide an overarching discovery platform?

For context, USU is a mid-sized R1 institution serving approximately 28,000 undergraduate, graduate, masters, and doctoral students on 9 campuses and 23 education centers statewide. USU Libraries uses the Innovative Interfaces Inc. (III) Sierra catalog with Encore Duet as the primary discovery layer and WebPac as an optional public catalog interface. In 2019, USU Libraries conducted an Ithaka S+R Faculty survey that examined, among other aspects, the starting point for research. The findings of this survey indicated that a significant number of faculty started their research in Google Scholar. Prompted by the frequent upheavals in the library discovery vendor landscape and the possibility of generalizing the Ithaka Survey results to undergraduate populations, the research team considered the impact of replacing the single search box discovery layer (currently Encore; single search environment) with a dual-tab search interface. In this interface, Google Scholar would be the primary search for article-level data, and WebPac would be the primary search for books and media.

The research team examined the following research questions:

1. What is the **average completion time** for tasks performed in the dual-tab interface versus tasks performed in the currently implemented single-search interface?

2. On average, how many **actions** does it take to perform tasks in the dual-tab interface versus the currently implemented single-search interface?
3. What are the **benefits and/or drawbacks** that users perceive when using the dual-tab interface versus the currently implemented single search interface to search for information?

To investigate the questions, the research team recruited 25 undergraduate students using a random stratified sampling method to test USU's current search environment (single search interface) versus a test platform that contained a dual-tab interface. To test these interfaces, researchers developed a list of information-searching tasks comprised of ten questions split into two categories (Part 1 and Part 2) for participants to answer. The tasks asked participants to search for various item types on the USU Libraries' current website (single search) and a test platform (dual-tab). The information-searching tasks covered: known items, topical ideas, book reviews, journal articles, e-book titles, call numbers, and more. Researchers generated the search tasks based on a previous research experiment that looked deeply at information-searching pathways that proved problematic for users in the past.[1]

The research team expected that a dual-tab search interface would facilitate a curated and more popular information searching process as opposed to a single-search box because the interface explicitly labeled the articles tab as "Google Scholar," and indications of Google Scholar brand awareness were high. Performance indicators of task speed and pages visited provided quantitative evidence of efficiency, while surveyed feedback on interface perception offered qualitative evidence of user preference. Findings proved surprising, with an overall trend of users preferring the single-search interface that was most like Google (in that it was a single search bar) but not branded as such.

This paper will address the methods and general conclusions portion of the experiment described above. The researchers hope the paper offers readers some starting points for evaluating their users' information search practices and sharing helpful knowledge on research design and population sampling methodology.

Recruitment

As mentioned in the introduction, USU's previous fielding of the Ithaka S+R Faculty survey identified that, like peers across the nation, many faculty begin their research in Google Scholar. National survey trends point to Google Scholar usage increasing over time in this context. These findings prompted the researchers to consider whether the same was true with other populations—namely, undergraduates. As a research population, undergraduates tend to be the least familiar with digging deep into a library catalog system. The researchers built this experiment on the assumption that straightforward, uncomplicated search interfaces would likely produce the best information-searching results for undergraduate students. With this in mind, researchers identified USU's undergraduates for the experiment participant pool.

Researchers conducted a power calculation to determine how many undergraduate participants were required to have sufficient statistical power and discern significant relationships between factors such as discipline, familiarity with the existing library search interface, order of interface interaction, etc. The power test—a test that takes a

significance threshold, estimated mean difference, estimated standard deviation, and desired statistical power to determine appropriate sample size—concluded that 12 participants per group would be sufficient.

Instead of sampling the entire undergraduate population, the research team conducted a stratified random sample. This sampling method increased the chances of obtaining a representative sample, allowing researchers to generalize findings. After receiving a protocol exemption from USU's Institutional Review Board (IRB) Office, the team coordinated with USU's Central Information Technology (IT) and the campus Registrar's Office to obtain a dataset of the undergraduate population. The dataset included the following demographics: first name; last name; email address; grade level; primary college; primary department; primary major; and primary campus.[2]

As previously mentioned, researchers designed this experiment based on a previous study that found problematic information-searching pathways. These problematic pathways were primarily non-STEM based topics. This fact prompted the team to wonder if there were design needs and information-searching habits relevant to discipline. In particular, the team wondered if different student disciplines correlated with specific search behaviors and/or preferences that could impact their performance or satisfaction with USU's search interface. To test this hypothesis, the research team divided the participants into the following strata (i.e., subsets of the population): STEM and non-STEM.

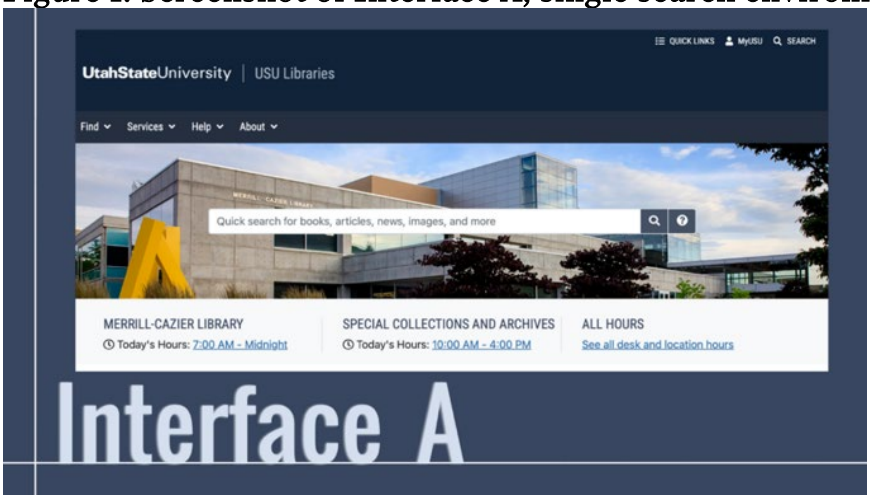
Since some disciplines straddle the varying definitions of STEM fields, researchers predetermined which colleges qualified as STEM and non-STEM based on our local context and sorted students into groups based on their assigned colleges. After manually separating the population, the team took note of what proportion of the entire population each group represented. We found that at the time, about 59.19% of the USU undergraduate population belonged to the STEM group and 40.81% to the non-STEM group. Our goal was to attempt to match the proportions present in our sample to those in the population.

The team used a random number generator to select the appropriate number of students from each group, where the generated number corresponded to the row number in a spreadsheet of potential participants. Knowing that randomized sampling is challenging and often yields fewer participants than desired, we sampled twice the number of needed students, 112, anticipating that most would not respond. This method was conducted several times until we had sufficient numbers for our two groups. Upon completing the experiment, we recruited 12 participants in one group and 13 in the other.

Experiment Execution and Methodology

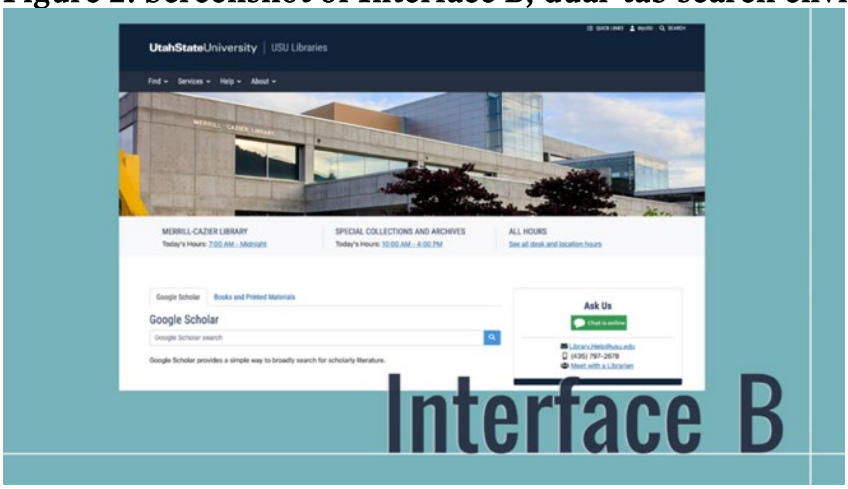
The research team utilized several tools to design, execute, and analyze the experiment results. The study consisted of a pre-survey, the experiment, and a post-survey. The pre-survey was designed to gauge students' research habits and preferences. The experiment implemented A/B testing in the form of a cross-over designed experiment. The participants were asked to complete common research tasks on two different platforms called Interface A (Figure 1) and Interface B (Figure 2). One platform was the standard library interface that implements an Encore discovery layer (Interface A).

Figure 1: Screenshot of Interface A; single search environment



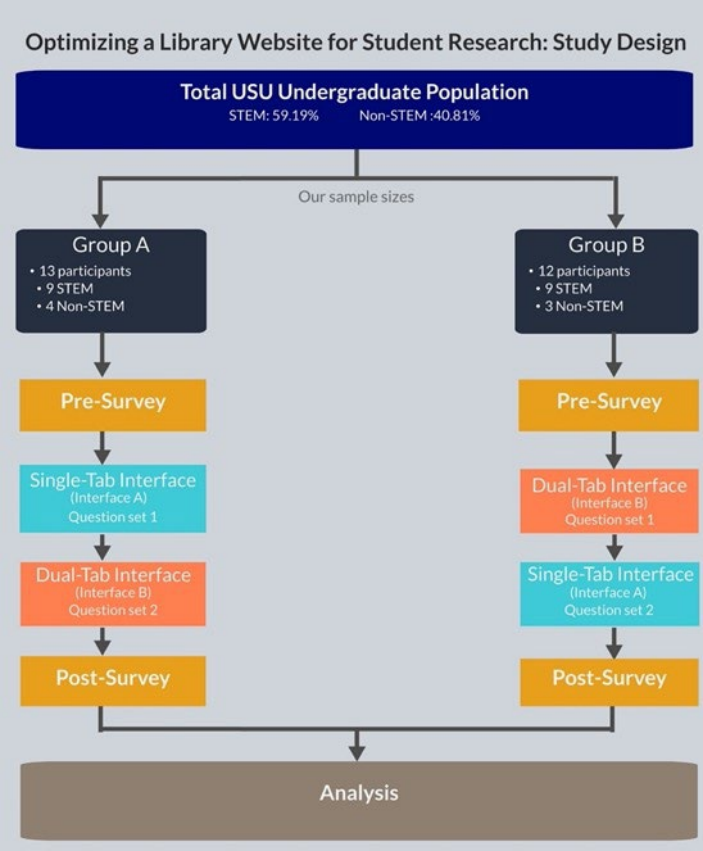
The second platform was a dual-tab interface that allowed the user to complete the research tasks in either Google Scholar or WebPac.

Figure 2: Screenshot of Interface B; dual-tab search environment



The two groups completed two sets of research tasks. They completed the same tasks in the same order but interacted with the interfaces in the opposite order (Figure 3).

Figure 3: Flow chart detailing participant group assignments



As previously mentioned, the research team developed the information-searching tasks based on data from a previous research project conducted by the USU Libraries in 2019.[3] This project pulled web logs from the Encore discovery layer and analyzed search trends. All information-searching tasks for this project were based on actual user searches. In some cases, though, the research team supplied some context to the questions to frame a typical narrative that users might encounter. For instance, instead of just asking users to search for the term "Anaïs Nin," the tasks included contextual clues such as the need to find a book BY the author instead of ABOUT the author.

Table 1: Information Searching Tasks
Part 1

Question Number	Task Order	Question	Search Type	Search Cue	Anticipated Resource
1	1	A friend recommended you read some books by Anaïs Nin. Determine if the library has any books by this author.	Known Item	Author	Book
2	2	Please research the topic of prescription medication disposal. Find the title and author of two	Topical	Topic	Article

Question Number	Task Order	Question	Search Type	Search Cue	Anticipated Resource
		articles you could use to write a paper.			
3	3	Your professor asked you to include the following source in your paper: "Ten principles of grammar facilitation for children with specific language impairments" by Marc E. Fay, Steven H. Long, and Lizbeth H. Finestack. It was published in 2003 in the American Journal of Speech-Language Pathology. Determine if the library has access to a copy.	Known Item	Citation information	Article
4	4	Your professor gives you the following citation. "Baker, N. D., & Nelson, K. E. (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. First Language, 5, 3–22." Determine whether the library provides access to this article.	Known Item	Citation information	Article
5	5	Find two e-books (electronic books) on the topic of Buddhist meditation.	Topical	Topic, Format	e-Book

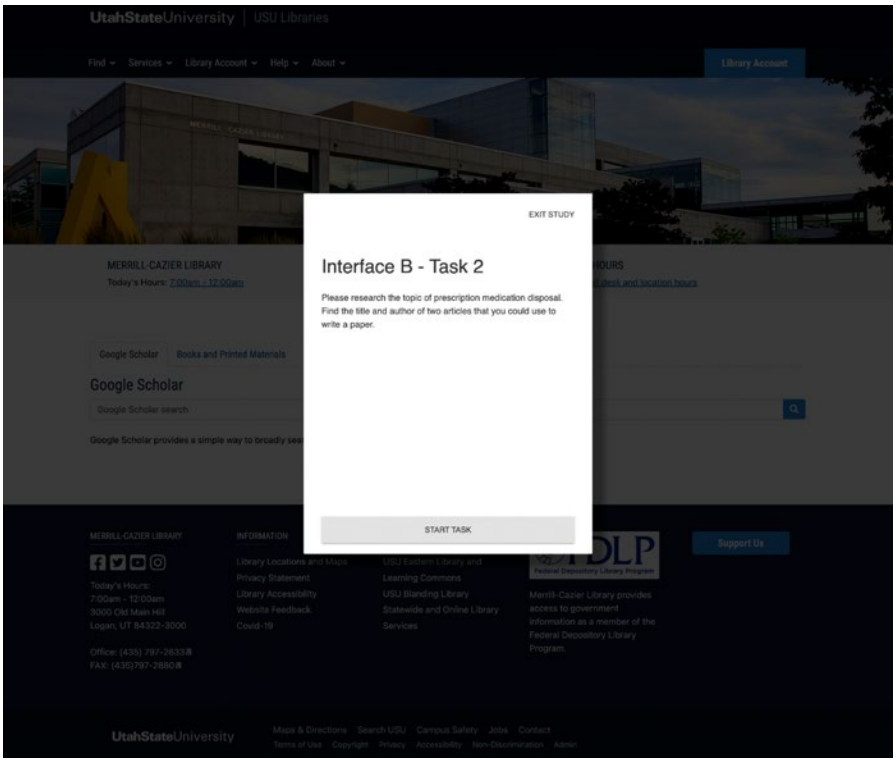
Part 2

Question Number	Task Order	Question	Search Type	Search Cue	Anticipated Resource
1	6	Please find the call number for the book "The Devil in the White City: Murder, Magic and Madness at the Fair That Changed America" by Erik Larson. (Example of a call number: LC 345.N54 2001)	Known Item	Title	Book

Question Number	Task Order	Question	Search Type	Search Cue	Anticipated Resource
2	7	Find out who won the Nobel Prize in Physics in 2018.	Topical	Topic	Unspecified
3	8	Your class acquaintance recommended a book for you. They said it was called something like teaching cues for sports skills – but you can't remember the author's last name. Please find the full title and author.	Known Item	Title	Book
4	9	Determine if you can find an e-book (electronic book) of The Brothers Karamazov by Fyodor Dostoevsky.	Known Item	Format, Title, Author	e-Book
5	10	You are investigating the effect of technology on mental illness. Please find a review article highlighting the most recent research on this topic.	Topical	Topic	Article

The task list was presented to participants via a pop-up window in the Loop11 software interface. Participants were directed through the platform to one of the two interfaces based on their group assignment. For instance, the second task for all participants was finding two articles on prescription medication disposal. Loop11 displayed the same information-searching task to all participants but directed Group A to the single search interface and Group B to the dual-tab interface (Figure 4).

Figure 4: Screenshot of information search task window presented to Group B for the dual-tab search environment



Once the participant read the task, they could click “Start Task” and begin their search. Once they were finished, they clicked on a “Show Task” button on the bottom left of their screen and were presented with three choices (Figure 5): 1) to “Hide” the task box so they could continue searching, 2) to “Abandon” the task and move onto the next one, or 3) “Task Completed.”

Figure 5: Screenshot of information search task box options

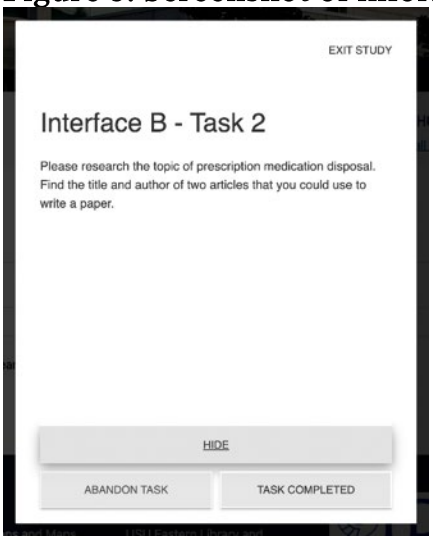
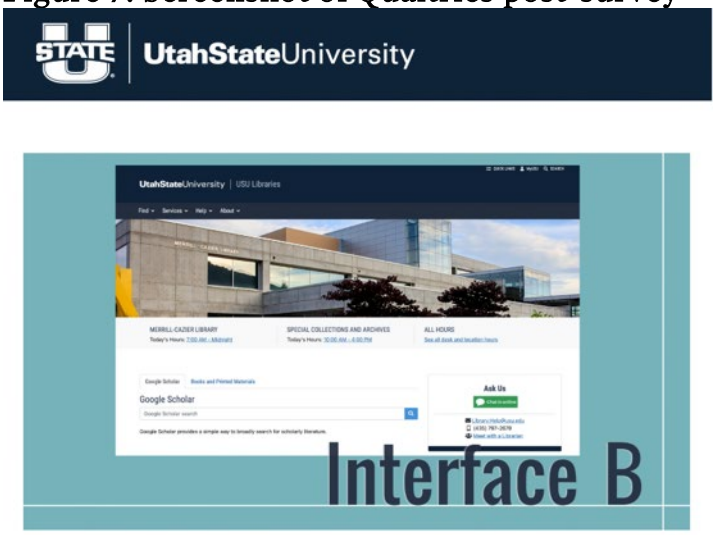


Figure 6: Screenshot of information search task answer box

Once they chose “Task Completed”, an answer box was triggered (Figure 6) where they could input what they found. Usually, this was in the form of a citation for articles or answering a question.

Data such as completion time, pages visited, and screen recordings of participant interactions with the interfaces were collected using Loop11 software. Loop11 provided a dashboard with each participant’s activities, including time-stamped videos, heat maps, clickstreams, and statistics. The latter was readily downloadable in a CSV format. The visuals could be downloaded one at a time or screen captured for analysis. The post-survey (Figure 7), conducted in Qualtrics, asked the students what they liked and disliked about each platform and their overall preferred platform.

Figure 7: Screenshot of Qualtrics post-survey



Analysis and Observations

Quantitative Analysis

The data collected in this study were run through a statistical model using SAS OnDemand software to determine whether any significant relationships existed between the explanatory variables—Oi, Sj, and Pk—and the response variable—Yijk.

$$Y_{ijk} = \mu + O_i + S_j + P_k + \varepsilon_{ijk}$$

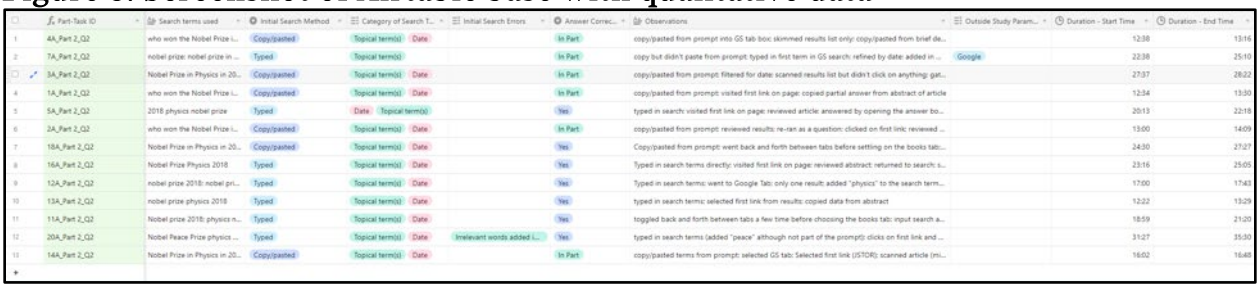
- Yijk := Difference in average completion time
- μ := Overall mean
- Oi := Effect due to the order participant interacted with the two interfaces
- Sj := Effect due to the participants' stratum

- Pk:= Effect due to participants' prior research preference between Google Scholar and the current library discovery layer
- Ɛijk := Random error

Qualitative Analysis

Qualitative analysis was also conducted on the data to look for user behavior and preference trends. Data from both Loop11 and Qualtrics were imported into Airtable (Figure 8). The project team viewed each task video and coded participants' actions for such things as search terms used, search errors, correctness of answers, whether the participants went outside the study parameters, the duration of the task (due to an error in the Loop 11 timing configuration), how many times participants saw login requests and/or opted to authenticate, facets and features of each system used, which tab was selected (for the dual tab tasks), and general observations.

Figure 8: Screenshot of Airtable base with qualitative data



Research team members used a code book created in Confluence to guide the coding process. This codebook was developed incrementally, with the research team doing some rough preliminary coding on a select number of items and then meeting to discuss issues or observations during our experience that we could feed back into the codebook to develop it more robustly.

Observations

The team began with three main research questions:

Research Question 1: What is the average completion time for tasks performed in each interface?

The team discovered that there **was** a difference in task completion time when taking into account the order in which a participant interacted with an interface. Average participant completion times improved in the second interface they encountered regardless of the interface. The team attributes this finding to familiarity with the experiment process improving completion times. Additionally, when considering the different strata, STEM students performed tasks with similar speed regardless of the interface used, while non-STEM performed faster in the single search interface.

Research Question 2: How many actions does it take to perform tasks in each interface?

Overall, there was a minimal statistical difference between interfaces regarding how many actions it took to perform tasks. Screen capture software highlighted skimming behaviors that did not correlate to answer correctness or how many actions

participants took to find the answer they submitted. Some participants read pages deeply and still presented incorrect answers, while others rapidly skimmed and still gave correct answers.

Research Question 3: What are the benefits and drawbacks participants perceived in each interface?

Researchers found participant feedback the most interesting. Overall, participants preferred the single search environment more, citing that they liked the simplicity of the single search bar. Their preferences did not necessarily correlate to a correct answer. Second, participants noted that filtering was difficult in both interfaces, and our WebPac catalog (dual-tab destination) felt old and clunky. WebPac has an antiquated interface with robust but complicated search options. Last, the team noticed that authentication prompts were a major deterrent in the single search environment. As noted above, participants exhibited extreme skimming behaviors. The authentication system was counterproductive for this behavior.

Conclusion

While the research team anticipated that the dual search interface would create a better experience for users, the results of the experiment did not support this theory. Both interfaces performed similarly in looking at the performance indicators of task speed without considering other factors such as prior search interface preference, order of interface use, and educational background. However, when the educational background is taken into account, we saw that those with a STEM background performed about as well on either platform. In contrast, non-STEM background participants performed much better in the single search environment. This information will be valuable as USU Libraries consider possible changes to its discovery layer and auxiliary search interfaces.

Additionally, given the brand awareness of Google Scholar, the research team assumed the dual-tab interface would be preferable to the single-tab search experience. However, participants showed a strong preference for the single-search interface. Complicating factors, however, include the look and feel of the WebPac interface (dual-tab) in which results were displayed. Participants found this interface clunky and difficult to both navigate and interpret results. They factored the interaction of these products into their overall experience in the dual-tab interface. Lastly, USU Libraries will have to consider the barrier that authentication prompts cause in the search process when considering skimming behaviors exhibited by their current undergraduate population.

The researchers hope that the paper offered readers some starting points for evaluating their users' information search practices and sharing helpful knowledge on research design and population sampling methodology.

Endnotes

[1] Robert Heaton & Liz Woolcott, "Unraveling the (search) string: Assessing library discovery layers using patron queries," in *Proceedings of the 2020–2021 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment, October 29, 2020–March 17, 2021*, virtual conference, ed. Sue Baughman, Jackie Belanger, Emery Durnan, Elizabeth Edwards, Martha Kyrillidou, Klara Maidenberg, Angela Pappalardo, and Maurini Strub (Washington, DC: Association of Research Libraries, 2021), https://www.libraryassessment.org/wp-content/uploads/2021/06/261-Heaton-Unraveling-the-Search-String.pdf.

[2] The research team did not sample students outside of the main Logan main campus. This decision was due to logistics; however, future experiments may include these populations outside the Logan campus.

[3] Heaton & Woolcott, "Unraveling the (search) string," 2021.

Bibliography

Robert Heaton & Liz Woolcott, "Unraveling the (search) string: Assessing library discovery layers using patron queries," in *Proceedings of the 2020–2021 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment, October 29, 2020–March 17, 2021*, virtual conference, ed. Sue Baughman, Jackie Belanger, Emery Durnan, Elizabeth Edwards, Martha Kyrillidou, Klara Maidenberg, Angela Pappalardo, and Maurini Strub. Washington, DC: Association of Research Libraries, 2021. https://www.libraryassessment.org/wp-content/uploads/2021/06/261-Heaton-Unraveling-the-Search-String.pdf.