How Real is Real Enough?: Participant Feedback on a Behavioral Simulation Used for Information-Seeking Behavior Research

Authors

Amy G. Buhler, University Librarian, University of Florida, Marston Science Library, PO Box 117011, Gainesville, FL 32611-7011, abuhler@ufl.edu ORCID: 0000-0002-0511-8273

Brittany Brannon, Research Support Specialist, OCLC Research, 6565 Kilgour Place, Dublin, OH 43017, brannonb@oclc.org ORCID: 0000-0001-7090-3342

Tara Tobin Cataldo, Associate University Librarian, University of Florida, Marston Science Library, PO Box 117011, Gainesville, FL 32611-7011, ttobin@ufl.edu ORCID: 0000-0003-4569-6374

Ixchel M. Faniel, PhD, Senior Research Scientist, OCLC Research, 6565 Kilgour Place, Dublin, OH 43017, fanieli@oclc.org ORCID: 0000-0001-7302-5936

Lynn Silipigni Connaway, PhD, Director of Library Trends and User Research, OCLC Research, 6565 Kilgour Place, Dublin, OH 43017, connawal@oclc.org ORCID: 0000-0001-9783-675X

Joyce Kasman Valenza, PhD, Assistant Teaching Professor, Rutgers University, SC&I, 1277 Cox Road, Rydal, PA 19046, joyce.valenza@rutgers.edu ORCID: 0000-0002-9951-2772

Rachael Elrod, EdD, Associate University Librarian, University of Florida, Education Library, PO Box 117016, Gainesville, FL 32611-7016, relrod@ufl.edu ORCID: 000-0002-1840-4789

Christopher Cyr, Associate Research Scientist, OCLC Research, 6565 Kilgour Place, Dublin, OH 43017, cyrc@oclc.org ORCID: 0000-0001-5246-6900

Note

This is the accepted manuscript of an article published in the *Journal of Librarianship and Information Science*, copyright 2022. This accepted manuscript is deposited on the OCLC Research website under the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

Suggested Citation

Buhler, A. G., Brannon, B., Cataldo, T. T., Faniel, I. M., Connaway, L. S., Valenza, J. K., Elrod, R., & Cyr, C. (2022). How real is real enough? Participant feedback on a behavioral simulation used for information-seeking behavior research. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006211067799

Accepted Manuscript

How Real is Real Enough?: Participant Feedback on a Behavioral Simulation Used for Information-Seeking Behavior Research

Abstract
A challenge of studying information-seeking behavior in open web systems is the unpredictability of those systems. One solution to counteract this issue is employing a simulation to ensure experimental control. However, concerns arise over the realism of such an environment. This paper assesses the realism of a behavioral simulation used to study the evaluation behavior of 175 students from 4th grade through graduate school. We assess realism through the examination of targeted participant feedback about what would have made the simulated environment and tasks more realistic to these participants. Based on this feedback, we reflect on decisions made in designing the simulation and offer recommendations for future studies interested in incorporating behavioral simulation in their research design. We find that a thoughtfully designed simulation can elicit naturalistic behavior when the controlled environment is designed to be realistic in meaningful ways. Because the simulation does not have to perfectly match reality to elicit these behaviors, designing a simulation that is real enough is an effective method to study information-seeking behavior.

Keywords: Information-seeking behavior, information evaluation, simulation methods, simulated environment, realism, participant feedback

Introduction
One of the challenges of studying students' information-seeking behavior in open web systems is the unpredictability of those systems. The use of a behavioral simulation allows for the observation and capture of rich, comparable data, both quantitative and qualitative, while providing experimental control. A major concern with using simulated environments to facilitate data collection is whether or not the environment is realistic enough to the participant to elicit natural behaviors. This paper examines feedback from student participants on the realism of a simulation used in a study of point-of-selection behavior (i.e., the point at which an information seeker determines a resource meets their information need) of 175 students from 4th grade through graduate school (Buhler et al., 2019). Based on this feedback, we reflect on decisions made in designing the simulation and offer recommendations for future studies.

Background
Simulation is employed by many academic disciplines, with the meaning of the term differing widely. The modern definition of the term arose in the 1940s as "The technique of imitating the behaviour of some situation or process (whether economic, military, mechanical, etc.) by means of a suitably analogous situation or apparatus, esp. for the purpose of study or personnel training" (*OED Online*, n.d.). However, a consistent definition of simulation that extends across the academic literature or even within a single discipline is lacking. In part, this is because the term simulation describes a range of varying techniques that tend to be used in different disciplines for different purposes.

In the social and behavioral sciences, Jones (1985: 327) states that the main purpose of a simulation is to "uncover the essence of the phenomenon being modeled, to see how it works." Gist et al. (1998) identify and define three major categories of simulation in an effort to distinguish between the variations that exist:

- *Physical* – Using or interacting with a scaled down model of a physical object (often used in the fields of engineering and design).

Accepted Manuscript

- *Computer* – Using computational algorithms to test or simulate various outcomes (often used in the sciences and engineering). Human behaviors can be modelled using this technique; however, there is no human interaction with the simulation itself.
- *Behavioral* – Modelling particular aspects of the social environment for the purposes of observing human interaction within the model (often used in the social and behavioral sciences).

Landriscina (2013) collapses the types of simulation into two categories: model-based (Gist et al.'s computer simulation) and experiential (encompassing both physical and behavioral simulation). The simulation discussed in this paper best aligns with the category of experiential simulation and more specifically Gist's behavioral simulation, in replicating specific elements of the environment to observe how people interact with the model to better understand a phenomenon of user behavior.

Gist et al. (1998: 253) define behavioral simulation as "… a constructed (versus natural) research setting in which human participants interact with each other and/or a confederate [ex. actor] in an experience that is characteristic of one found naturally in organizations, because important contextual factors (e.g. task, physical setting, etc.) have been construed realistically." Realism is a critical component of behavioral simulation and the discussion of realism as a theme associated with simulation occurs repeatedly in the behavioral science literature. Its basis is in Aronson and Carlsmith's (1968) discussion of realism within the context of psychological research methods. They divide realism into two separate (although not mutually exclusive) concepts: experimental and mundane. Experimental realism denotes how realistic the situation is to the participant, while mundane realism refers to the likelihood that the events that occur in the laboratory setting would occur in a natural setting.

Abelson (1968) characterizes the balance between realism and reality itself as the *simulation gap*. This concept is echoed by others (Breakwell et al., 2000; Gravetter and Forzano, 2006) who note a main concern of the use of simulated situations in psychological research is ecological validity (i.e., how much the observed behavior mirrors the participant's actual behavior). According to Sommer and Sommer (2002), the validity of the simulation hinges on the researcher capturing the major components of the setting or experience being simulated. There also is a fine line between simulation and experiment that can be crossed should the simulation create too much separation between events that would naturally occur together. They suggest that realism can be gauged through comparison of simulation results with behaviors in the natural environment.

DiFonzo et al. (1998) argue that even noncomplex microworlds (i.e., computer-generated simulated environments that participants interact with) possess higher levels of experimental and mundane realism when compared to typical laboratory experiments, thus making them an ideal research tool. Gravetter and Forzano (2006) also make an important point: no matter how realistic a simulation may be, participants will still know that the simulation is an experiment, and their acceptance of that and willingness to engage with the simulation is critical. DiFonzo et al. (1998) claim the realism of a simulation can be assessed through participant feedback via self-reported or observational methods. The purpose of this paper is to assess the realism of our behavioral simulation instrument from the users' perspective. This assessment will be accomplished using targeted participant feedback collected as part of a larger study examining students' selection of online resources.

Use of Behavioral Simulation in Library and Information Science

The use of behavioral simulation is limited in the field of library science, most often leveraged as a teaching tool rather than a research tool. Ullah and Ameen's (2018) systematic review of applied methodologies and methods found a single instance of field simulation as a research method within LIS. This finding aligns with Walsh's (2009) article reviewing assessment methods for information literacy that feature two articles employing simulation, one paper-based and one online. The latter was an online simulation-based information literacy instructional approach used to guide middle school students through the information-seeking process (Newell, 2010). An additional search of the literature revealed two more studies that explicitly used online simulation. The first evaluates a prototype of a computer-supported

Accepted Manuscript

learning environment (The Information Retrieval Game) used to show students the overall effectiveness of their search strategies (Halttunen and Sormunen, 2000). The second discusses an assessment of a mixed-reality simulation environment to help instruct and develop preservice librarians' understanding of multicultural pedagogies (Underwood et al., 2015).

Perhaps most closely aligned is a pocket of literature in information science relating to user-centered interactive information retrieval (IIR) which takes a cognitive viewpoint and shifts study focus to the user's role as opposed to the system (Belkin, 1990). These studies often centralize around understanding user behavior and experience, evaluating IR system or interface features, and meta-evaluation of evaluation metrics (Liu and Shah, 2019). For a comprehensive overview of IIR current study designs and methods, readers should refer to Kelly (2009), Kelly & Sugimoto (2013) and Liu and Shah (2019). User-centered IIR often employs various methodological components that align with Gist et al.'s definition of behavioral simulation (1998). IIR seeks to assess a system by making use of real users as testers who "state personal need interpretations to the system and judge the relevance of the retrieved documents, under controlled circumstances." (Borlund, 2000: 74). There are two aspects of behavioral simulation that can clearly be seen in this segment of literature.

First, and perhaps most common, is the use of simulated work task situations, leveraging Borlund's (2003) IIR evaluation model. Borland and Ingwersen (1997: 228) define the function of the simulated work task situation as twofold: "Firstly, it serves as a description of the 'universe' of the information need situation in which the user is supposed to see himself, and based upon which the user formulates the search statement to the system. It consists of an indicative request, a definition, and a simulated work task situation. Secondly, the concept serves as a foundation for the user when assessing situational relevance of the retrieved documents." Borlund (2000) argues that the simulated work task situation use enables experimental control while allowing for the collection of comparable cognitive and performance data for information retrieval techniques. This model serves as a framework for assessing users' interactions with systems in both experimental and naturalistic research settings. Many studies employ simulated work task situations to study components of information-seeking behavior in a variety of search systems. An early example applying this model is seen in Blomgren et al.'s (2004) examination of newspaper journalists' use of the database NewsLink. A more recent example is Gossen's (2014) study comparing adults and children's interactions with search engine results pages.

Second, behavioral simulation methods can also be seen in the use of simulated environments to create a research setting for observing user behavior. These environments range in approach from the simple to complex. For example, Novin and Meyers (2017) used an HTML-based "mock-SERP" to study university students' sense making of conflicting science information. Haas and Unkel (2017) created an HTML-manipulated SERP to assess the influence of page rank and source reputation on the credibility judgements of university students. Complex environments can serve dual purposes: an interactive setting for study participants and a means to collect data. Some examples of the tools that contrive a realistic environment to capture web-based user behavior are YASFIIRE (Wei et al., 2014), SCAMP (Renaud and Azzopardi, 2012), and WiIRE (Toms et al., 2004).

There are several recent instances where both of these behavioral simulation components are employed to study information-seeking with real subjects. For example, Bilal and Gwidzka's (2016, 2018) studies employed both simulated work tasks and an experimental interface to examine how children formulate search engine queries and read search engine results pages (SERPs). Their experimental environment consisted of a page that both displayed the task the user was to complete along with an embedded modified Google search interface, allowing them to manipulate the results seen by users and restrict the number of results per page. Wildemuth et al. (2019) examined the usefulness of different surrogates (i.e., snippets or descriptions) for making relevance judgements of online videos. The simulated environment in this project limited results to a particular subset of videos, but all search functionality was intact to ensure a realistic searching context for participants.

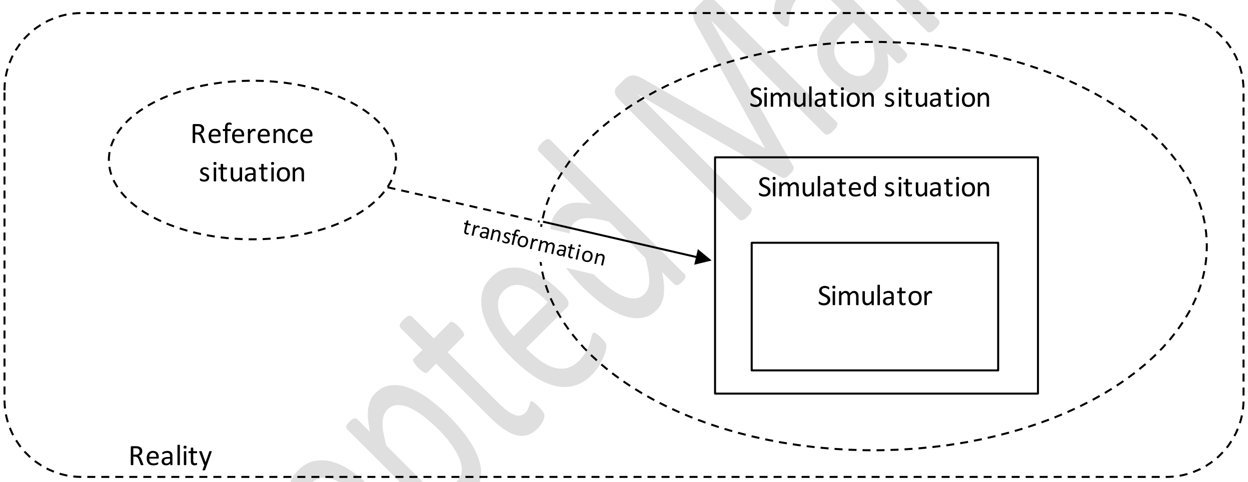
Accepted Manuscript

The user-centered IIR literature provides the context for the use of behavioral simulation components in experimental design. The main purpose of many human-focused studies is to either refine the systems being tested or better understand information-seeking behavior. However, they do not assess the participants' perception of the experimental environment itself. This paper complements those studies found in the existing literature, in that it focuses on assessing the realism of our study's experimental environment through the analysis of targeted participant feedback. Through this analysis, it seeks to discover what components of behavioral simulation are critical to attend to when constructing a realistic experimental environment to elicit naturalistic information-seeking behavior.

Framework

To structure our analysis we employed Vidal-Gomel and Fauquet-Alekhine's (2016) behavioral simulation framework. They contend that a behavioral simulation comprises three parts that must be carefully designed to ensure realism: the simulator, the simulated situation, and the simulation situation. Figure 1 illustrates how the three parts of simulation relate according to Vidal-Gomel and Fauquet-Alekhine. They define the *simulator* as "the artifacts that simulate (partially or completely) the operation or the behavior of a technical system, facility, or a natural phenomenon" (Vidal-Gomel and Fauquet-Alekhine, 2016: 2). The *simulated situation* is the reference situation that has been replicated. The *simulation situation* includes all elements and activity of the participants and facilitators engaging with the simulation. In the following sections, we will use Vidal-Gomel and Fauquet-Alekhine's framework to discuss our simulation by looking in depth at the decisions made and the feedback received for each part.

Figure 1: Visual adaptation of Vidal-Gomel and Fauquet-Alekhine's components of behavioral simulation.



Simulation Development

The Researching Students' Information Choices (RSIC) study used a mixed methods approach to investigate online *point-of-selection behavior*. The objectives of the project are to learn how students determine the credibility of online information resources and whether they differentiate among different types of online information resources for a school research project.

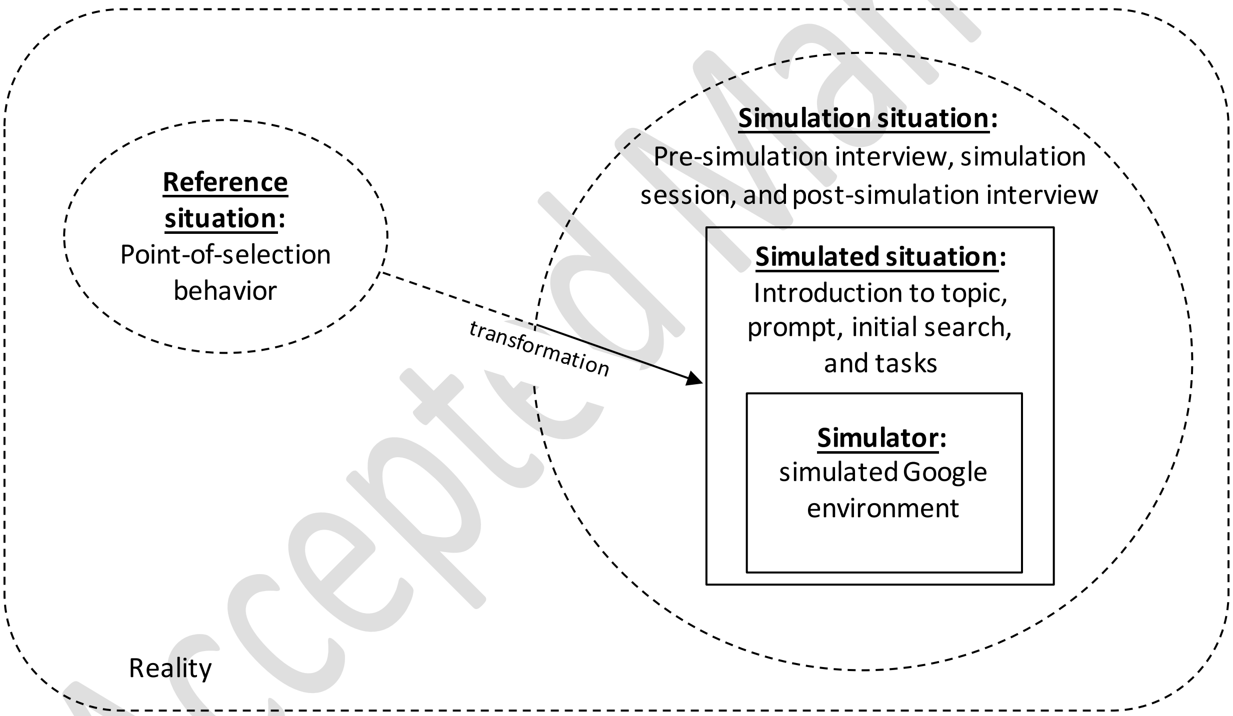
In order to study these behaviors, we wanted to create an environment where students could interact naturally with online information resources, while controlling the environment enough to ensure students encountered the same circumstances. Controlling the circumstances produces data that are more comparable and reduces the possible explanations for participants' decisions (Cataldo et al., 2020). To accomplish this, we designed a simulation. While it is not a data collection tool, a simulation creates a naturalistic web environment where data can be collected. In our study, we were able to use the simulation software to collect quantitative data and used a think-aloud protocol to gather qualitative data on students' thought processes.

Accepted Manuscript

While a major limitation of simulations is their artificialness, we felt that the benefits outweighed this negative aspect. Three major benefits for this project were the avoidance of pitfalls associated with live web search (e.g., 404 errors, changes in resources); the elimination of variability (e.g., different results/order of results depending on IP address); and the ability to compare within and across groups of participants.

The study included 175 students: 30 graduate, 30 undergraduate, 30 community college, 26 high school, 30 middle school, and 29 elementary school. Each student was shown a short video news clip to introduce them to the topic of their research (the Burmese python in the Florida Everglades), given a research project prompt (i.e. simulated work task) appropriate to their educational stage, and asked to conduct an initial search within the simulated Google interface. A set of simulated Google results pages were returned and each participant proceeded through five tasks in which they (1) selected helpful resources for the research topic, (2) judged the citability of their selected resources, (3) explained why they did not select other resources, (4) assessed the credibility of their selected resources, and (5) identified the container (i.e., publication type, see (Brannon et al., In Press)) of a preselected list of resources. A short demonstration of the simulation used for the post-secondary students can be viewed at https://ufdc.ufl.edu/IR00010570/00001/videos/0. Figure 2 overlays Vidal-Gomel and Fauquet-Alekhine's framework with elements specific to the RSIC project simulation. These elements will be discussed in more detail in the following sections.

Figure 2: RSIC project components within Vidal-Gomel and Fauquet-Alekhine's framework.



The Simulator

As the artifact that must convincingly replicate a real-life system, the simulator can be one of the most complex, time-consuming, and expensive portions of developing a simulation. The RSIC project created and utilized what Vidal-Gomel and Fauquet-Alekhine (2016) call a full-scale simulation, or a simulation that attempts to replicate the entirety of the technical interface with which a participant interacts. In a full-scale simulation, both the technical functioning of the system and the perceptual aspects of interaction with the system must be convincing enough to engage the participant in the reality of the simulation (Vidal-Gomel and Fauquet-Alekhine, 2016).

Accepted Manuscript

To do this, we used Articulate's Storyline[1], an instructional design software, to emulate Google results pages and linked information resources. We used the font size, colors, and sequence of the different elements in the Google result for each resource. When possible, we also used the snippet that appeared with the resource in a real Google search. When we could not find a Google link to a particular resource, we adapted what we saw in other Google snippets to create one ourselves.

When determining which results to return, we selected resources that were topically aligned but otherwise did not attempt to perfectly replicate the type of results that Google returns. The research team worked with an advisory panel of teachers, librarians, and professors to find and evaluate potential resource options that would be suitable for each educational stage (Cataldo et al., 2020). We selected resources that varied in nature and amount of information, resource type, and container type. We included results, such as scholarly journal articles and books, which would normally be behind paywalls and not necessarily appear in open web search results. Based on the typical ordering of results on a Google search, we placed a "sponsored" link at the top of the first page, followed by a link to a relevant Wikipedia page.

We recreated each resource using screenshots and other capture methods. These resources were then connected to the appropriate URL in the results pages, which allowed participants to click through each link on the results pages as they typically would in Google. To make the resource webpages more realistic, we also created active links within the resources themselves. The decision to do so was made on a resource-by-resource basis, determined by what the team thought students most likely would want to click on (e.g., PDF, about the author, etc.).

For each task, with the exception of the Not Helpful Task, we included task-specific buttons and behaviors which can be seen in Figures 3-6. The buttons not only prompted the participants to make explicit task decisions, they also recorded values for different variables within the task. The Helpful Task displayed the entire list of resources and had a simple checkbox next to each result. The subsequent tasks displayed only a subset of the total results and had buttons specifically suited to the decision that we were asking participants to make in that task. The Cite Task used yes and no buttons, the Credible Task used buttons on a Likert scale, and the Container Task used drag-and-drop labels. For these tasks, making it easy for participants to record their task decisions was more important than making the pages look identical to Google search results pages, particularly because these tasks already diverged from natural evaluation behavior.

Accepted Manuscript

---

[1]For more details about how the software was utilized and the technical build process, please see Cataldo et al. (2020), documentation for simulation build (https://ufdc.ufl.edu/IR00011300/00001/pdf), and post in Storyline forum (https://community.articulate.com/discussions/articulate-storyline/freebie-google-search-result-template).

Figure 3: Helpful Task

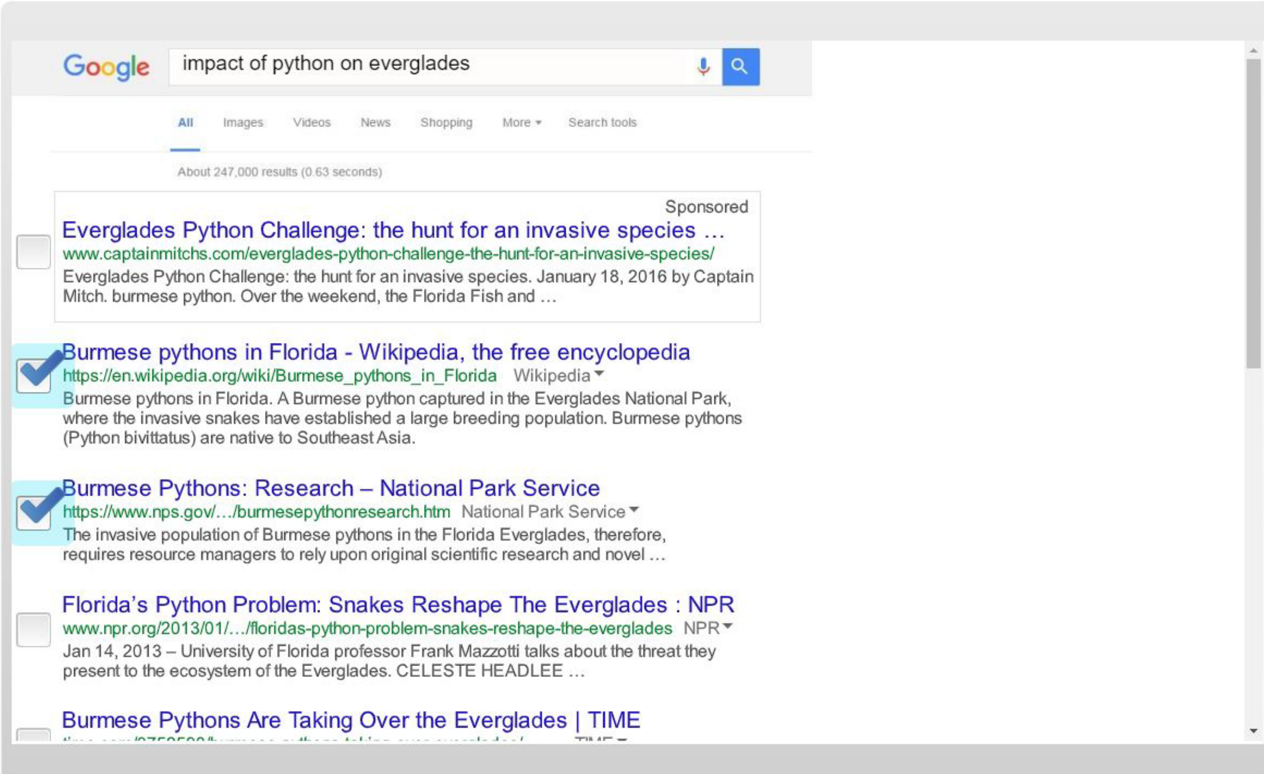


Figure 4: Cite Task

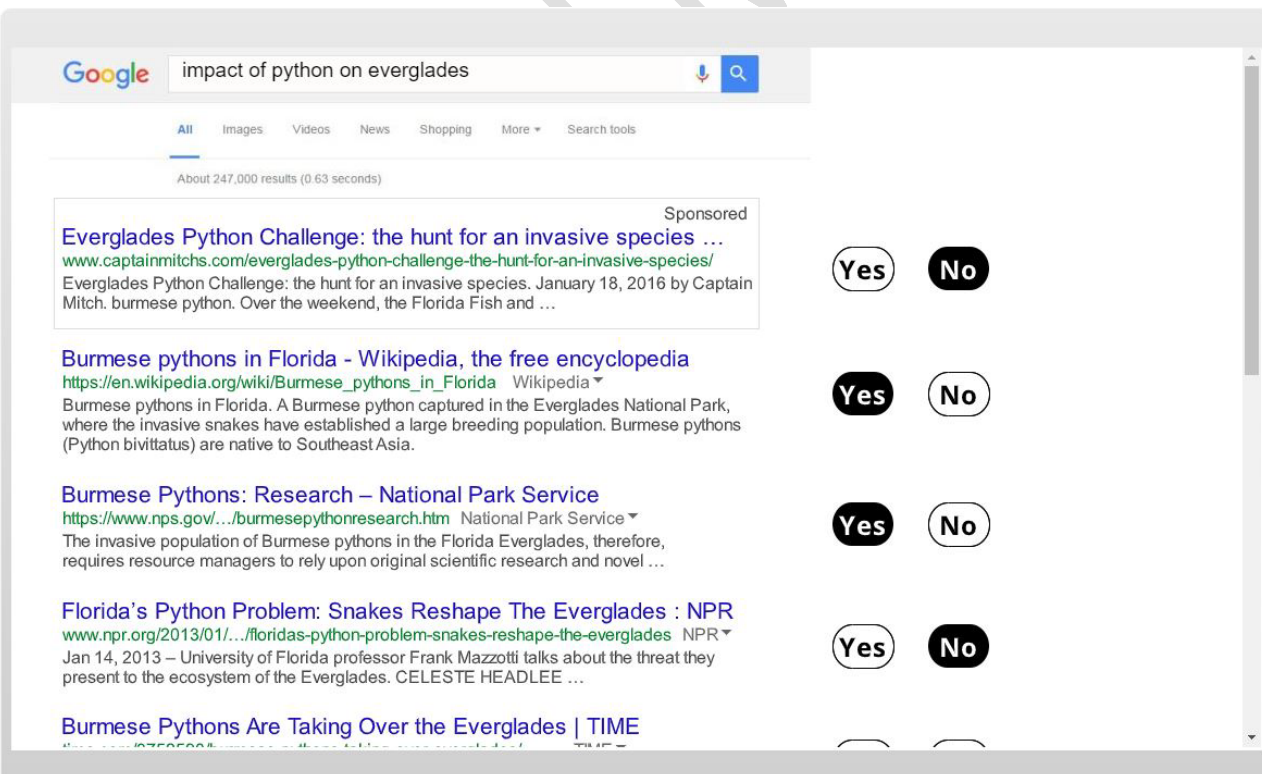


Figure 5: Credible Task

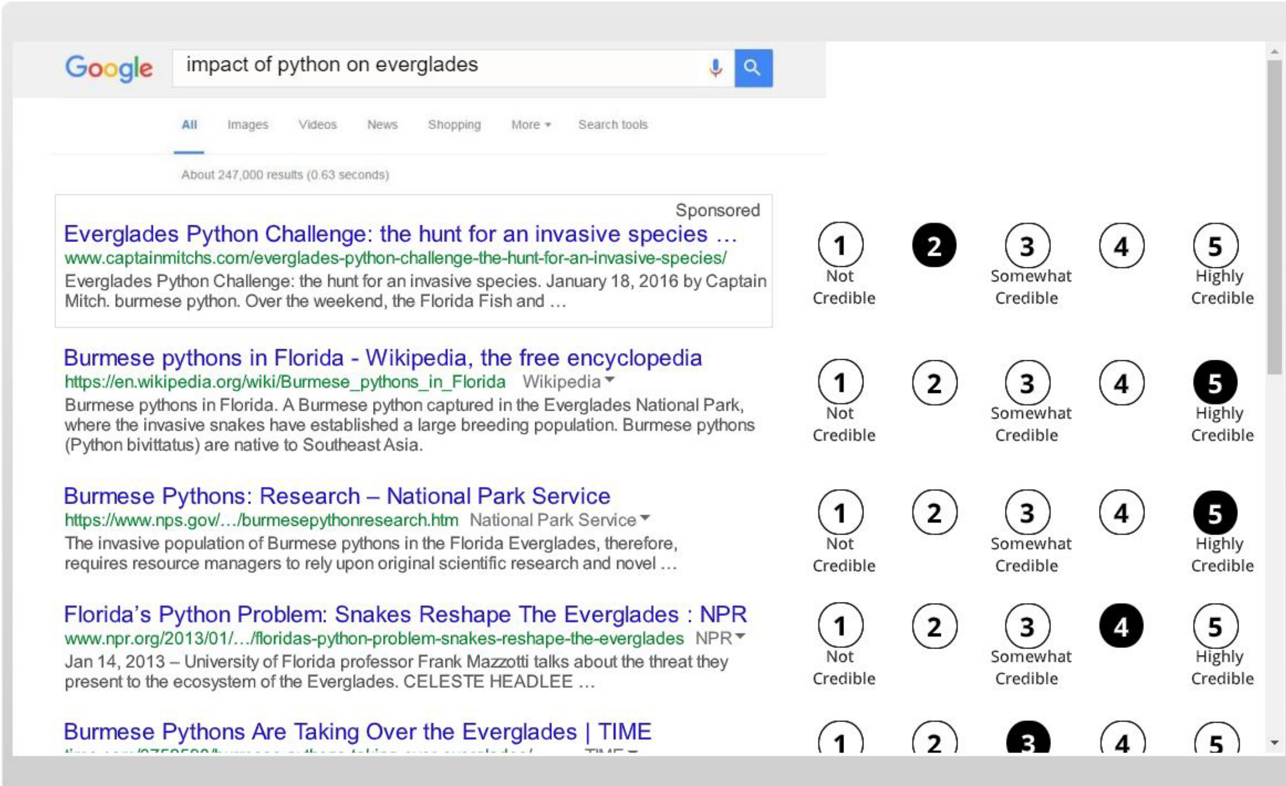
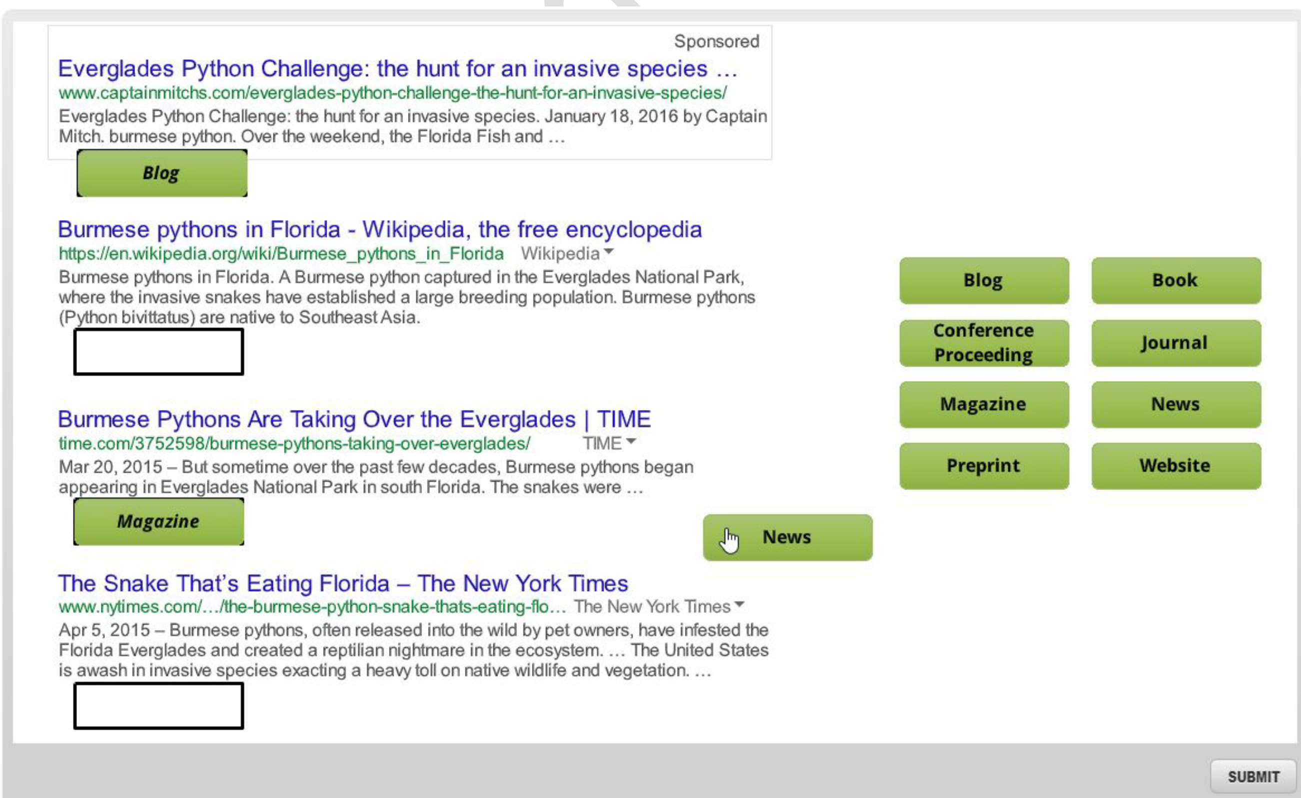


Figure 6: Container Task



The Simulated Situation

Constructing the simulated situation requires transposing the real or reference situation into a sequence of events and circumstances that can be meaningfully completed in the simulation (Vidal-Gomel and Fauquet-Alekhine, 2016). Vidal-Gomel and Fauquet-Alekhine (2016: 4) define the reference situation as "the category of situations which trainees must be able to manage at the end of the training." In the case of a simulation built for research, rather than instruction, the reference situation can be better understood as the category of situations that the researcher wishes to capture data about.

It is important to note that we are talking about a *category* of situations. The reference situation is an abstraction based on the researchers' understanding of the numerous lived experiences of people engaged in the behavior of interest. The implications of this are twofold. First, it means that researchers must have a good understanding of the situation they are studying in order to collect useful data. Second, it means that the simulation will not be able to recreate a situation that exactly matches each participant's expectations.

To study point-of-selection behavior, we had to replicate the reference situation in which this behavior would occur. To make that situation as realistic as possible, we included the introductory activities of being introduced to the topic, receiving an assignment prompt (i.e., simulated work task), and conducting an initial search. We worked with our advisory panel to select a topic that could be used across all six educational stages and to develop appropriate prompts for each stage based upon that topic.

In order to capture targeted data, the reference situation can be transformed in several different ways as it is recreated in the simulation, including dividing, decoupling, and focusing (Vidal-Gomel and Fauquet-Alekhine, 2016). In other words, when creating the simulated situation, researchers can divide tasks that the participant would normally complete together, decouple variables from one another to ask participants to assess them separately, and/or focus the participants' attention on the researchers' area(s) of interest. One benefit of these types of transformations is that they reduce cognitive load for participants (Vidal-Gomel and Fauquet-Alekhine, 2016). They also allow researchers to isolate behaviors or thought processes that may otherwise be lost when observing in natural settings. However, researchers also have to retain the balance between dividing, decoupling, and focusing variables with maintaining a realistic simulated situation.

In our simulation, students were asked to complete five total tasks with the simulated results that their search returned. The first was the Helpful Task, which most closely replicated the reference situation and behavior under study. This was the least-defined task and asked participants to select a predetermined number of results that they found the most helpful for their assignment. No criteria or definitions of *helpful* were given; participants determined what helpful meant to them.

Our research objectives focused our work on particular types of resource evaluation and parts of the information-seeking process. To ensure that we collected data that would address these questions, we transformed the reference situation based upon the research team's collective experience working with students and prior research. All of the tasks following the Helpful Task moved away from realism to focus students' judgments on a particular element of interest to the research team.

- The Not Helpful Task showed participants the results that they did not select and asked why they were not helpful, further exploring what makes resources helpful, or not, to students at the point of selection.
- The Cite Task showed participants the resources they selected as helpful and asked whether they would cite them in their assignment.
- The Credible Task again showed participants their helpful resources and asked them to rate the credibility of those resources.

Accepted Manuscript

- The Container Task asked students to apply the best of eight possible container labels to a set of pre-selected resources.

Separating these tasks isolated students' strategies for making each judgement, ensuring we could study the different types of evaluation whether or not students discussed them during the Helpful Task.

The Simulation Situation

The simulation situation encompasses all of the activity that takes place to conduct the simulation and is divided into three phases: the briefing, the simulation session, and the debriefing (Vidal-Gomel and Fauquet-Alekhine, 2016). In order to ensure that the simulation situation was consistent for each participant, we developed a set of guides for our facilitators. These included preparation checklists, instructions for setting up the technology, and a script to follow when conducting the simulation. Cataldo et al. (2020) provides more details about these materials.

Our briefing included greeting the participant, obtaining informed consent, and conducting a pre-simulation interview. The debriefing included explaining to the participant that they had been interacting with a simulation and conducting a post-simulation interview. The bulk of our simulation situation was comprised of the simulation session.

During the simulation session, the facilitator's role was to ensure that the simulator and simulated situation progressed according to the research design. In order to keep the simulation session running smoothly and ensure that the appropriate data was collected, the facilitator had to answer participants' questions and intervene to resolve problems or remind the participant to think aloud. To keep the participant immersed in the simulation, facilitators intervened only when necessary.

The simulations were conducted in conjunction with a think aloud protocol (Ericsson and Simon, 1993; van Someren et al., 1994). Participants were asked to say aloud everything that went through their minds as they completed the simulation. This thinking aloud was a deviation from the realism of the simulation, but it enabled us to collect rich data about how and why participants were making decisions, which was crucial for addressing our research questions. All of our simulations took place in private, reserved library conference rooms to protect participant confidentiality and help participants be more comfortable and relaxed.

Methods

Our goal in developing the simulation was threefold: to create an environment realistic enough to elicit natural behaviors, to keep the participants immersed in the simulation tasks, and to collect comparable data. In order to gauge our success, we included the question "Thinking about the activity we just did, what would have made it more realistic to the way you look for information for a research project?" in the post-simulation interview. This question assesses the experimental realism of the simulation by directly asking participants to reflect on their perceptions while completing the simulation. It also allows us to roughly gauge the mundane realism of the simulation because it gives us a proxy for understanding which elements of the simulation were unrealistic enough to attract participants' notice and thereby potentially disrupt their normal behaviors.

As mentioned above, the RSIC study collected data from a total of 175 participants: 30 graduate, 30 undergraduate, 30 community college, 26 high school, 30 middle school, and 29 elementary school students. Participants were all students in the same county and ranged in age from 9 to 35. All post-secondary students were Science, Technology, Engineering, and Mathematics majors.

For this paper, we reviewed all 175 participant responses to this question for themes, agreed on theme definitions, categorized responses by theme, and met to discuss ambiguous examples. We divided the participant feedback themes according to the three parts of the simulation, grouping together themes that related to each part. For each of the three parts, we analyze and reflect upon how well our simulation created a realistic environment for our participants.

Accepted Manuscript

Findings and Discussion

Overall, 39 participants (22%) said the simulation was *realistic* to their usual behavior. This was tied as the most commonly mentioned theme (Figure 7). Of the participants who said the simulation was realistic, 12 were high school, 11 were elementary school, eight were community college, six were middle school, and one each were graduate and undergraduate students. This suggests that in attempting to create a simulation that was consistent across educational stages, we may have been more successful in meeting the expectations of some groups than others. Figure 8 shows the number of times the three most common themes were mentioned in each educational stage. Many of them only said that the simulation was realistic to their normal behavior and had no suggestions or criticism to offer. Others went on to mention ways the simulator, simulated situation, and simulation situation could have been more realistic, such as the ability to conduct multiple searches or start their search from a different point. We will talk about these and other themes in the paragraphs that follow.

Figure 7: Most Mentioned Themes in Participant Feedback (N=175)

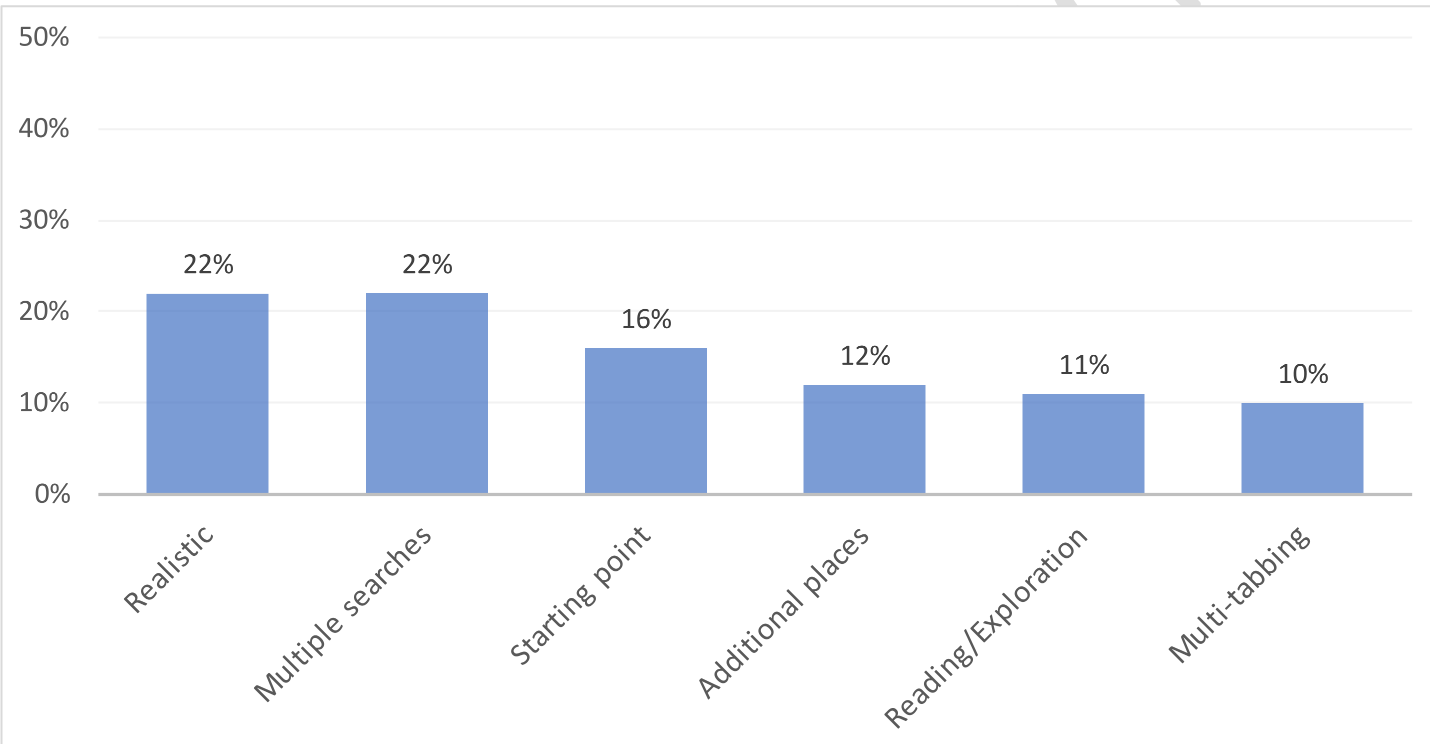
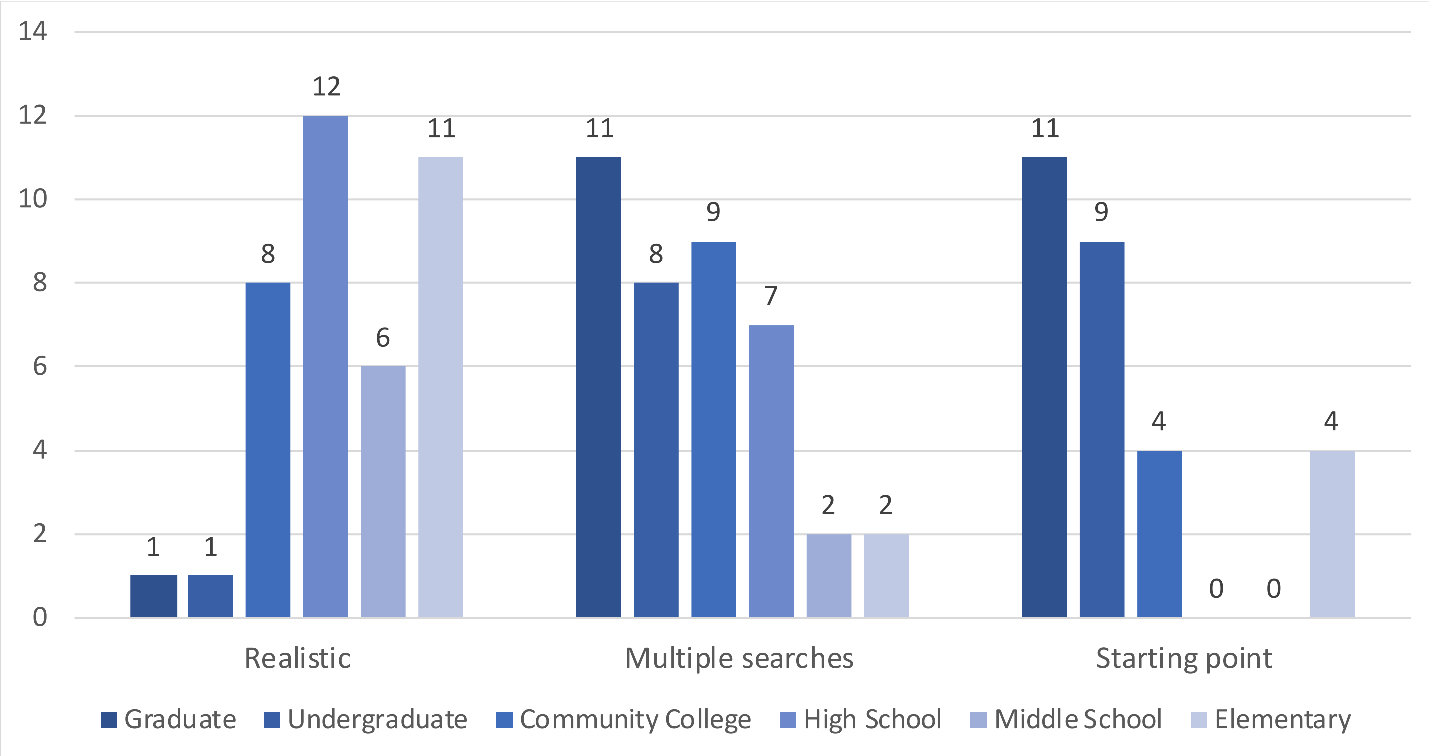


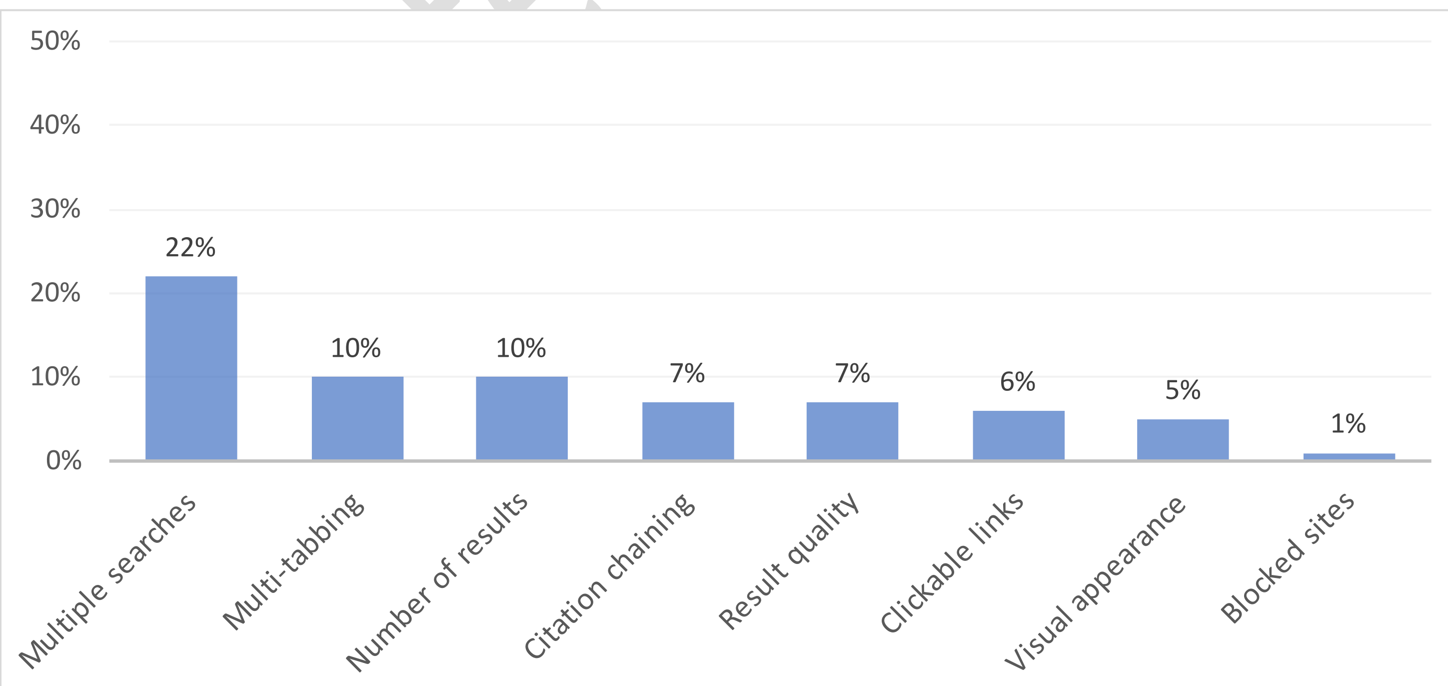
Figure 8: Three Most Commonly Mentioned Themes Divided by Educational Stage (N=175)



The Simulator

Only nine participants (5%) mentioned the *visual appearance* of the simulator, suggesting that the research team successfully accomplished replicating the overall look of the Google results pages (Figure 9). Most of these comments were about the Storyline module itself, which created a window-within-a-window appearance. This suggests that the task-related buttons included on each page were not disruptive. Because the buttons allowed for the capture of quantitative data as the simulation progressed, we highly recommend using a similar technique in research simulations.

Figure 9: Student Feedback on the Simulator's Realism (N=175)



Participants' most common frustrations were with the function of the simulator, particularly being confined to and working from a single set of results. The simulator was built to only allow one search query and participants had to navigate entirely within the same window. Thirty-nine participants (22%) mentioned that they would have conducted *multiple searches*, 18 participants (10%) mentioned *multi-tabbing*, 12 participants (7%) mentioned *citation chaining*, and 10 participants (6%) mentioned *clickable links*. These frustrations occurred most often at the higher educational stages.

The mention of *multiple searches* included participants mentioning that they would have conducted multiple keyword searches, changed or refined their search terms, or conducted additional searches to verify an author, source, or piece of information. Eleven graduate students mentioned this, as well as eight undergraduate, nine community college, seven high school, two middle school, and two elementary school students.

Multi-tabbing included participants mentioning that they would have had multiple tabs or browser windows open, usually to conduct multiple searches but often also because they wanted to open links in a new tab. *Citation chaining* referred to participants mentioning that they would normally use a resource's reference list to gather more resources, often by clicking on hyperlinked references or opening new tabs to search for listed references. *Clickable links* comprised any mention of wanting to be able to click on more links within the resources. For each of these themes, there were a similar number of mentions among graduate, undergraduate, community college, and high school students, but no mentions by middle or elementary school students.

During the simulation build, we made several deliberate decisions to limit the number of resources that participants could access. Part of our motivation for using a simulation, rather than observing behavior in natural settings, was that a simulation gave more control over the research environment, ensuring that data was comparable across participants. This included controlling which and how many resources the participants could interact with by limiting the number of pages of results and results per page, as well as limiting the ability for participants to conduct additional searches or to click on a link within a resource that would take them to a different resource. There also were practical considerations in adding additional resources and links within resources. Each additional page built added significantly to the workload of the designer constructing the simulation and, consequently, the costs and timeline for the project.

We recognized that limiting the number of resources that participants could access would be less realistic, but it was necessary to achieve reliable and comparable data that would allow us to perform the analyses we needed to address our research questions. It is, however, a limitation of the study. The fact that relatively few participants mentioned each of these themes suggests that it was a worthwhile tradeoff for our purposes.

Another consideration to note is the pattern of simulator-related themes across educational stages. These themes were mentioned more by participants in the higher educational stages, suggesting that older students may be accustomed to employing more complex search strategies and, therefore, are more comfortable with the iterative and lateral nature of searching. Students in K12 stages—particularly elementary and middle school—seemed fine with conducting a single search, selecting a few resources from the results, and calling their task complete. Students in higher educational stages were more likely to mention wanting to adapt their tactics according to what they found, including modifying their search, verifying authors or information, and following references or links to find additional relevant resources. Interestingly, those at the high school level tended to align with post-secondary participants. Given a different population or different area of focus, it would be worth investigating ways to allow participants to conduct multiple searches, open multiple tabs, and follow in-resource links to different resources.

The other aspect of the simulator that participants commented on were the results returned, including the number of results returned and the nature of those results. Seventeen participants (10%) mentioned that the *number of results* was not quite right, either because there were not as many pages of results as Google normally returns or because they had to review more results pages than they normally would. Thirteen participants (7%) mentioned the theme of *result quality* – the results in the simulation did not match their expectations of what they normally see in Google in terms of

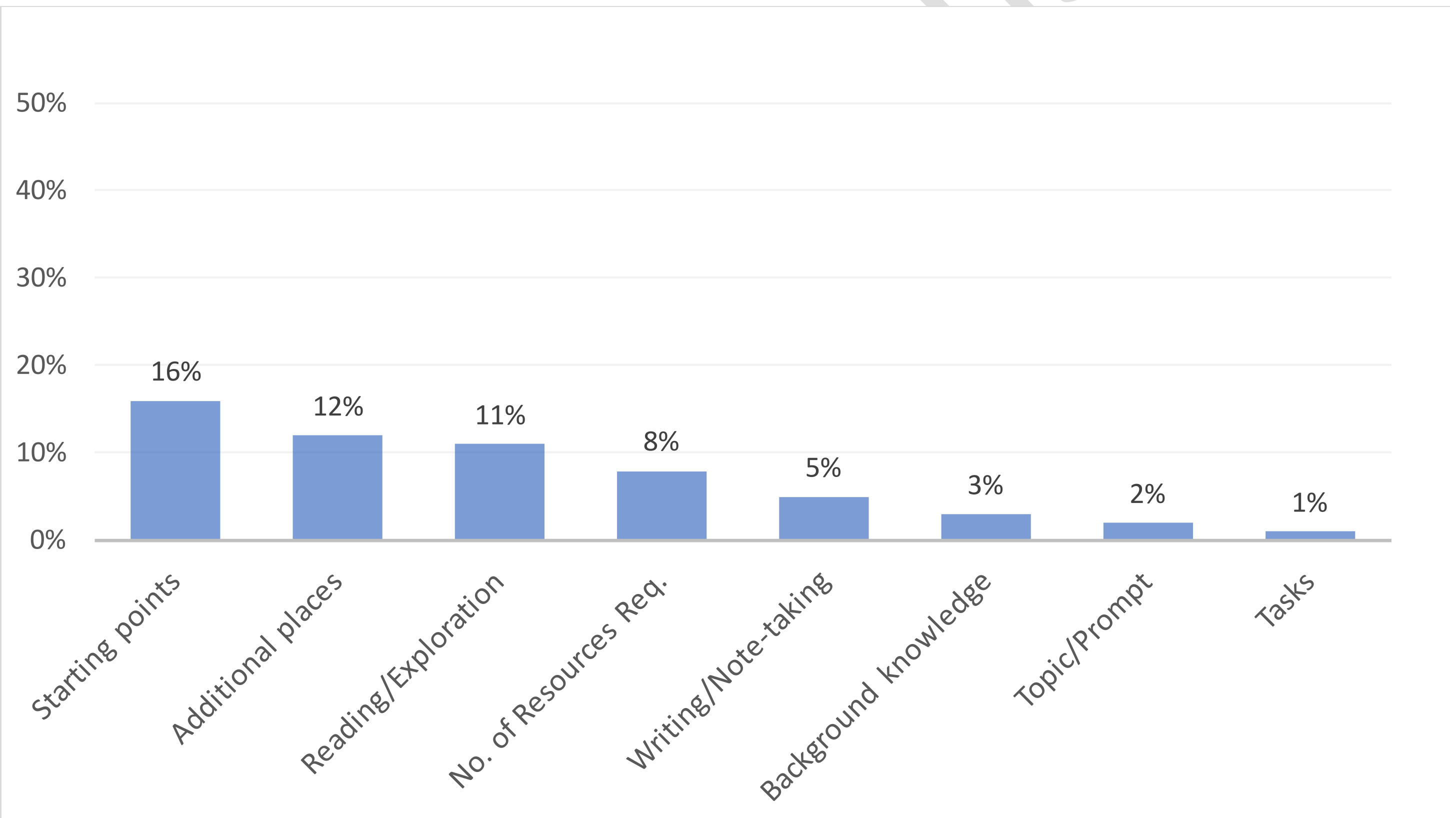
Accepted Manuscript

result type, quality, and/or relevance. Two participants (1%), both elementary school students, mentioned that they normally would not have been able to see some of the results that appeared in the simulation because those *sites are blocked* by their school. For these themes, there were no clear patterns across cohorts. Overall, the low number of participants who mentioned these issues (less than 10% for each) and the lack of pattern across cohorts suggests that the research team succeeded in determining how many and what type of results to present to each cohort.

The Simulated Situation

Figure 10 displays the themes mentioned in association with what would have made the simulated situation more realistic. Only two students (1%) mentioned that the *tasks* were not realistic. One graduate student said they would not assign a credibility score to resources they were interested in, and one elementary school student said that they would not try to identify the container. As we built the simulation, we were concerned that separating out the tasks would make the simulation less realistic, reducing the fidelity of students' behaviors and thus the mundane realism. This result suggests that did not happen. Sequencing the tasks did not disrupt students' perceptions of the realism, suggesting that this transformation of the reference situation, which was important for collecting data that would address our research questions, is a valid research approach.

Figure 10: Student Feedback on Simulated Situation's Realism (N=175)



The majority of frustrations students reported with the simulated situation revolved around the assumptions we made about where students go to search. We chose Google for the environment because it would be familiar to all of our participants, intuitive to navigate, and because many students use Google for school assignments (Perruso, 2016). When asked what would make the simulation more realistic, 28 participants (16%) mentioned *alternate starting points* and 21 participants (12%) mentioned *additional places to search*. These were the third and fourth most-mentioned themes overall, but they were still mentioned by only a small proportion of the participants.

Participants who mentioned *alternate starting points* stated that they would have started their search with a different search tool or resource. Of the participants who mentioned this theme, 11 were graduate, nine were undergraduate, four were community college, and four were elementary students. The fact that this theme was mentioned far more often by post-secondary students (24 in total) than K12 students (four in total) suggests that Google as a search tool may be a more common starting point for younger students who are just beginning to develop search strategies and skills. It also may be the case that because younger students do not yet have disciplinary specialties, the information-seeking strategies that they are learning rely on more generalized tools and resources. This could also account for the lack of students in a transitional stage, like community college and high school participants, mentioning this theme. More advanced students, on the other hand, may have developed search routines based upon their disciplinary knowledge and thus rely on more specialized tools and resources.

Participants who mentioned *additional places to search* discussed other search tools or resources that they would have used in addition to Google. For this theme, there was no clear pattern among the education stages. Both alternate starting points and additional places to search had overlap in the type of resources that participants mentioned wanting to use. A few mentioned other generic open web search engines such as Bing, but most discussed using either Google Scholar or specialized library databases that they would have had access to through their school. Several participants also mentioned using non-digital resources, including the library, books, other print items, and people. Despite the ubiquity of digital search, we have to be cognizant of the fact that people's information-seeking behavior is complex and involves digital, physical, and human sources (Cyr et al., 2021). Although our study was specifically focused on how students evaluate search engine results at the point of selection, and not their search strategies, a full understanding of information-seeking behavior must account for these other approaches.

Other responses that reflect on the simulated situation clustered around how well we replicated participants' experience of the school context, including the *number of resources required*, the *topic/prompt*, and *background knowledge*. Fourteen participants (8%) mentioned that the number of resources they were required to select did not meet their expectations. Only three participants (2%) mentioned that the assignment topic or prompt could have been more realistic.

Six total participants (3%), two each from high school, middle school, and elementary school, mentioned that they typically receive *background knowledge* about the topic through their class work or teacher to direct their assignment. This potentially represents an interesting distinction between the K12 and post-secondary groups. Post-secondary students often are encouraged to research topics not directly covered in class and consequently will not always have prior knowledge of the topic. This may result in these students being more comfortable with independent inquiry and beginning a search with a vaguely defined information need. By contrast, K12 students may be more likely to consider teachers and class content important sources of information at the beginning of a research project.

Finally, participants mentioned that their information use would have been different in a natural setting. Nineteen participants (11%) mentioned that their *reading or exploration strategy* would have been different, meaning that they would have consumed information from the resources differently. Eight participants (5%) mentioned that their *writing or note-taking strategy* would have been different, meaning that they would have employed a different strategy for using the information in the resources to prepare themselves for writing. There were no clear patterns among the educational stages for these themes. Many of the behaviors that participants mentioned wanting to engage in were things they could have done during the simulation, but others required a technology setup that wasn't available to them (e.g. having a word processing document open in a different tab to take notes).

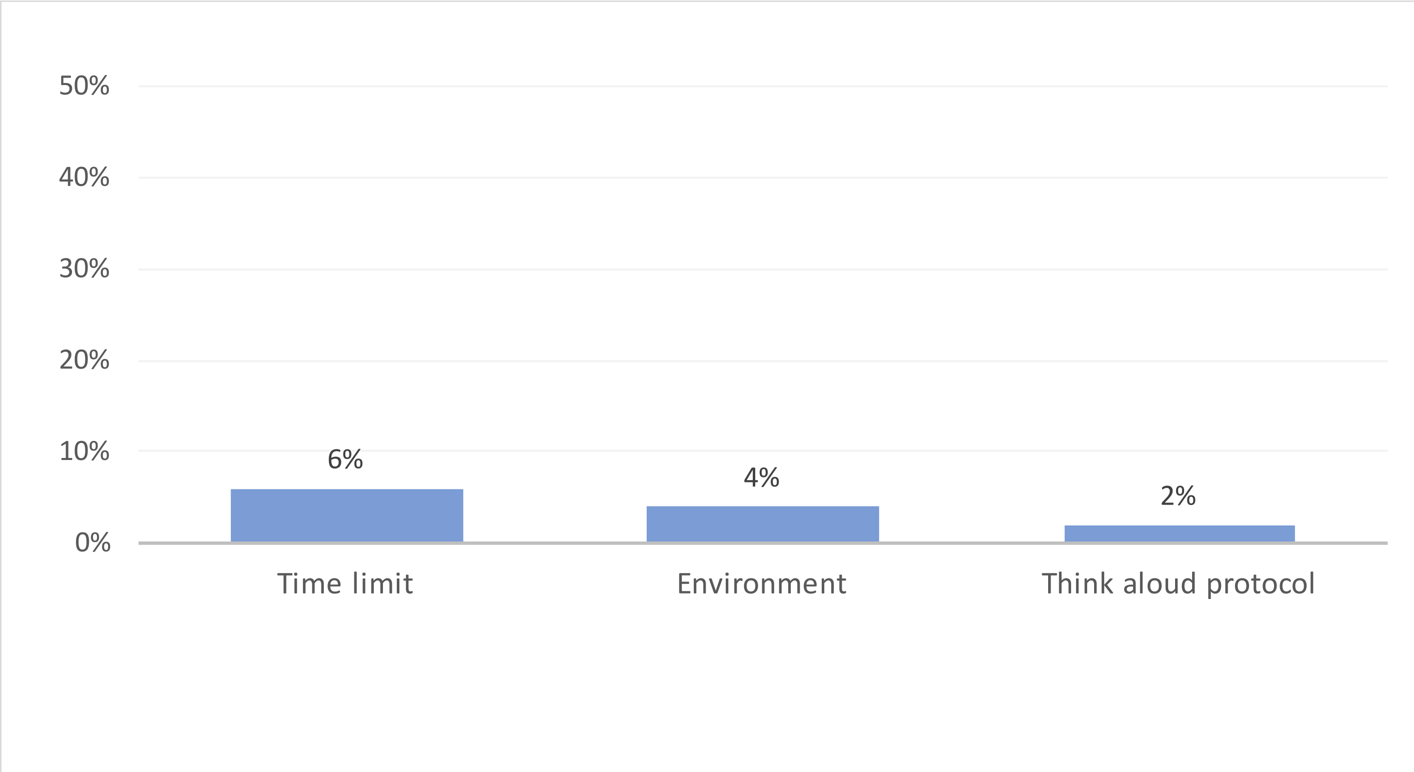
The Simulation Situation

Very few participants commented on elements of the simulation situation in response to the question of what would have made the simulation more realistic. Figure 11 displays the themes mentioned. Only three participants (2%) mentioned that *thinking aloud* while working was unnatural or abnormal. This suggests that the think aloud protocol is

Accepted Manuscript

an effective method when paired with a simulation because it gathers rich data without significantly disrupting the realism of the simulation.

Figure 11: Student Feedback on Simulation Situation's Realism (N=175)



Seven participants (4%) mentioned elements of the *environment*, external to the simulator that did not meet their expectations. Many of these responses were humorous, including one participant who wanted "a cup of coffee [laughter]" (undergraduate student) and another who was used to "having all of my siblings yelling [laughter]" (elementary school student). While it is impossible to perfectly replicate the environment that each participant is accustomed to while doing schoolwork, the infrequency of mentions of this element suggests the environment we selected was neutral enough to not be distracting.

Ten participants (6%) mentioned the theme of *time limit* – that they wanted to or would normally spend more time on their search. There was no time limit explicitly imposed on the participants, but the situation of being in a booked conference room with a facilitator may have made participants feel that they needed to complete the session within a shorter timeframe. Additionally, a few of these participants alluded to the more iterative nature of search—for example, that they would have spent more time over the course of several days or weeks gathering resources. This is outside of the scope of our project but would be interesting to study using a different combination of methods.

Conclusions & Reflections

The use of behavioral simulation as a research method allows for the observation and capture of rich, comparable data while allowing researchers greater control over the research environment. This method has been employed in some parts of the library and information science discipline. However, there is a lack of information on participants' perception of such research approaches. This paper provides a better understanding of how research participants view the realism of a behavioral simulation to help researchers make more informed choices while designing and deploying this method.

In the course of creating and deploying the simulation, we had to make many decisions that weighed the realism of the simulation against the quality and type of data we could collect. This analysis shows that several of the decisions that prioritized the data were worthwhile and things that we would recommend to other researchers. Sequencing the different types of judgments in separate tasks, particularly with the more open-ended task at the beginning, did not disrupt students' perceptions of the realism and was important in allowing us to collect data that directly addressed our research questions. Similarly, the use of buttons on the task pages, although a divergence from the appearance of the Google results pages, was not disruptive enough for students to comment on. These buttons allowed for the capture of precise quantitative data about students' evaluation of the resources for later data analysis. Finally, the very low number of comments about the think aloud protocol suggests that it is an effective data collection technique when paired with a simulation because it gathers rich data without significantly disrupting the realism of the simulation. As a consequence, we highly recommend all of these strategies, either in isolation or combined, as ways to collect quality data in behavioral simulation studies.

Some of the tradeoffs, however, remain limitations of our study. This project is focused on resource selection, and therefore evaluation rather than search behaviors. In making decisions that would enable us to better study selection behaviors, we made the search experience less realistic. In particular, limiting participants to a single search, preventing them from opening new tabs or windows, and not creating many clickable links affected the experimental realism of the simulation for some participants. It is difficult to say how much that may have affected the mundane realism of the simulation for our full sample. Given the low number of mentions of each theme, it is reasonable to conclude that the alteration of the search experience did not significantly disrupt students' evaluation behaviors.

One of the largest benefits of our study was the use of a single research design across a wide variety of educational stages because it collected data that allows us to draw direct comparisons among students that aren't often studied together. A notable—though not surprising—finding was the differences in perceived realism across educational stages. The limitations discussed in the above paragraph, for example, were hardly mentioned by elementary and middle school participants. These students seemed fine with conducting a single search, selecting a few resources from the results, and calling their task complete. Students in higher educational stages were more likely to mention wanting to adapt their tactics according to what they found, including modifying their search, verifying authors or information, and following references or links to find additional relevant resources. Interestingly, those at the high school level tended to align with post-secondary participants on these themes.

Only K12 students mentioned that they would have had background information about their assignment to guide their inquiry. Post-secondary students often are encouraged to research topics not directly covered in class and consequently will not always have prior knowledge of the topic. This may result in these students being more comfortable with independent inquiry and beginning a search with a vaguely defined information need. By contrast, K12 students may be more likely to consider teachers and class content important sources of information at the beginning of a research project. Since this is a part of the simulated situation, and not necessarily the simulator itself, one change we could have made was to provide more context and background information to participants in K12 stages prior to asking them to conduct their search and begin the tasks. This would have kept the data we collected consistent and comparable while potentially providing these students with a more realistic experience. Overall, however, far more K12 students said that the simulation was realistic and fewer mentioned things that would have made it more realistic than their post-secondary counterparts. This suggests that in trying to create a simulation that worked across all educational stages, we better met the expectations of students in earlier educational stages than those in later educational stages.

Based on our findings, we recommend a tightly controlled Google environment for students in elementary and middle school but suggest that researchers consider giving these students sufficient background information to help inform their information seeking. For undergraduate and graduate students, we recommend that researchers carefully consider their research goals to determine 1) whether the data comparability of a tightly controlled environment is worth the loss

Accepted Manuscript

in realism, 2) whether a different search system such as an academic database would be familiar enough to all participants to serve as a more realistic simulated environment, and 3) whether it's feasible and worthwhile to construct a simulator that allows students to search iteratively and across multiple systems. High school and community college students seem to represent an interesting transitional group in that their perspectives on the realism of the simulation sometimes aligned more closely with the elementary and middle school students and sometimes with the undergraduate and graduate students. Additional research is needed on the information-seeking behaviors of these groups to better understand their needs and expectations when conducting research for school projects.

Collecting participant feedback on the simulation allowed us to analyze how effective our simulation was at replicating a natural research environment. The relatively low number of participants who mentioned that different elements of the simulation were not realistic indicates that none of the design choices significantly hampered the experimental realism of the simulation. This suggests that the simulation also supported mundane realism. Overall, the simulation for this project succeeded in creating a naturalistic environment in which to capture rich, real-time quantitative and qualitative data. The use of behavioral simulation as a library and information science research method can provide researchers with a realistic view into information behavior.

Acknowledgements

The authors wish to acknowledge the RSIC research team and advisory panel, which made this work possible. We would also like to thank Samuel Putnam for his review and editorial comments.

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].

References

Abelson RP (1968) Simulation of Social Behavior. In: Lindzey G and Aronson E (eds) *The Handbook of Social Psychology*. 2nd ed. Reading, Mass: Addison-Wesley Pub. Co., pp. 274–356.

Aronson E and Carlsmith JM (1968) Experimentation in Social Psychology. In: Lindzey G and Aronson E (eds) *The Handbook of Social Psychology*. 2nd ed. Reading, Mass: Addison-Wesley Pub. Co., pp. 1–79.

Belkin NJ (1990) The cognitive viewpoint in information science. *Journal of Information Science* 16(1). SAGE Publications Ltd: 11–15. DOI: 10.1177/016555159001600104.

Bilal D and Gwizdka J (2016) Children's eye-fixations on google search results. In: *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology*, USA, 14 October 2016, pp. 1–6. ASIST '16. American Society for Information Science.

Bilal D and Gwizdka J (2018) Children's query types and reformulations in Google search. *Information Processing & Management* 54(6): 1022–1041. DOI: 10.1016/j.ipm.2018.06.008.

Blomgren L, Vallo H and Byström K (2004) Evaluation of an Information System in an Information Seeking Process. In: *Research and Advanced Technology for Digital Libraries* (eds R Heery and L Lyon), Berlin, Heidelberg, 2004, pp. 57–68. Lecture Notes in Computer Science. Springer. DOI: 10.1007/978-3-540-30230-8_6.

Borlund P (2000) Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation* 56(1): 71–90. DOI: 10.1108/EUM0000000007110.

Accepted Manuscript

Borlund P (2003) The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research-an International Electronic Journal* 8(3). Available at: http://informationr.net/ir/8-3/paper152.

Borlund P and Ingwersen P (1997) The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation* 53(3). MCB UP Ltd: 225–250. DOI: 10.1108/EUM0000000007198.

Brannon B, Buhler AG, Cataldo TT, et al. (In Press) Genre Containers: Building a Theoretical Framework for Studying Formats in Information Behavior. *Journal of the Association for Information Science and Technology*.

Breakwell GM, Fife Schaw C and Hammond S (2000) *Research Methods in Psychology*. 2nd ed. London ; Thousand Oaks, Calif: Sage Publications.

Buhler AG, Faniel IM, Brannon B, et al. (2019) Container Collapse and the Information Remix: Students' Evaluations of Scientific Research Recast in Scholarly vs. Popular Sources. In: *Proceedings of the Association of College and Research Libraries* (ed. Dawn M. Mueller), Cleveland, OH, 2019, pp. 654–667. Available at: https://ufdc.ufl.edu/IR00010779/00001 (accessed 3 June 2020).

Cataldo TT, Buhler AG, Faniel IM, et al. (2020) 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(4). University of Borås. Available at: https://doi.org/10.47989/irpaper881 (accessed 19 January 2021).

Cyr C, Brannon B and Connaway LS (2021) "I still go ask someone I enjoy talking to." The use of digital and human sources by educational stage and context. Article in press at Library & Information Science Research.

Difonzo N, Hantula DA and Bordia P (1998) Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, & Computers* 30(2): 278–286. DOI: 10.3758/BF03200656.

Ericsson KA and Simon HA (1993) *Protocol Analysis : Verbal Reports as Data*. Cambridge, Mass: MIT Press.

Gist ME, Hopper H and Daniels D (1998) Behavioral Simulation: Application and Potential in Management Research. *Organizational Research Methods* 1(3): 251–295. DOI: 10.1177/109442819813001.

Gossen T, Höbel J and Nürnberger A (2014) A comparative study about children's and adults' perception of targeted web search engines. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 26 April 2014, pp. 1821–1824. CHI '14. Association for Computing Machinery. DOI: 10.1145/2556288.2557031.

Gravetter FJ and Forzano LAB (2006) *Research Methods for the Behavioral Sciences*. 2nd ed. Australia ; Belmont, CA: Thomson/Wadsworth.

Haas A and Unkel J (2017) Ranking versus reputation: perception and effects of search result credibility. *Behaviour & Information Technology* 36(12): 1285–1298. DOI: https://doi.org/10.1080/0144929X.2017.1381166.

Halttunen K and Sormunen E (2000) Learning Information Retrieval through an Educational Game. Is Gaming sufficient for learning? *Education for Information* 18(4): 289. DOI: 10.3233/EFI-2000-18403.

Jones RA (1985) *Research Methods in the Social and Behavioral Sciences*. Sunderland, Mass: Sinauer Associates.

Kelly D (2009) *Methods for Evaluating Interactive Information Retrieval Systems with Users*. now. Available at: https://ieeexplore.ieee.org/document/8187092 (accessed 13 December 2019).

Accepted Manuscript

Kelly D and Sugimoto CR (2013) A systematic review of interactive information retrieval evaluation studies, 1967–2006. *Journal of the American Society for Information Science and Technology* 64(4): 745–770. DOI: 10.1002/asi.22799.

Landriscina F (2013) An Introduction to Simulation for Learning. In: *Simulation and Learning: A Model-Centered Approach*. New York: Springer, pp. 1–12.

Liu J and Shah C (2019) Interactive IR User Study Design, Evaluation, and Reporting. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 11(2). Morgan & Claypool Publishers: i–93. DOI: 10.2200/S00923ED1V01Y201905ICR067.

Newell TS (2010) Learning in Simulations: Examining the Effectiveness of Information Literacy Instruction Using Middle School Students' Portfolio Products. *Evidence Based Library and Information Practice* 5(3): 20–38. DOI: 10.18438/B85K7T.

Novin A and Meyers E (2017) Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, New York, NY, USA, 7 March 2017, pp. 175–184. CHIIR '17. Association for Computing Machinery. DOI: 10.1145/3020165.3020185.

OED Online (n.d.) simulation, n. Oxford University Press. Available at: https://www.oed.com/view/Entry/180009 (accessed 2 November 2020).

Perruso C (2016) Undergraduates' Use of Google vs. Library Resources: A Four-Year Cohort Study. *College & Research Libraries* 77(5): 614–630. DOI: 10.5860/crl.77.5.614.

Renaud G and Azzopardi L (2012) SCAMP: a tool for conducting interactive information retrieval experiments. In: *Proceedings of the 4th Information Interaction in Context Symposium*, New York, NY, USA, 21 August 2012, pp. 286–289. IIIX '12. Association for Computing Machinery. DOI: 10.1145/2362724.2362776.

Sommer R and Sommer BB (2002) *A Practical Guide to Behavioral Research : Tools and Techniques*. 5th ed. New York: Oxford University Press.

Toms EG, Freund L and Li C (2004) WiIRE: the Web interactive information retrieval experimentation system prototype. *Information Processing & Management* 40(4): 655–675. DOI: 10.1016/j.ipm.2003.08.006.

Ullah A and Ameen K (2018) Account of methodologies and methods applied in LIS research: A systematic review. *Library & Information Science Research* 40: 53–60. DOI: 10.1016/j.lisr.2018.03.002.

Underwood J, Kimmel S, Forest D, et al. (2015) Culturally Relevant Booktalking: Using a Mixed Reality Simulation with Preservice School Librarians. *School Libraries Worldwide; Edmonton* 21(1): 91–107. DOI: http://dx.doi.org/10.14265.21.1.006.

van Someren MW, Barnard YF and Sandberg JacobijnAC (1994) *The Think Aloud Method: A Practical Approach to Modelling Cognitive Processes.* LondenAcademic Press. Available at: https://dare.uva.nl/search?identifier=7fef37d5-8ead-44c6-af62-0feeea18d445 (accessed 9 December 2019).

Vidal-Gomel C and Fauquet-Alekhine P (2016) Reflections and Theoretical Contributions Regarding Trainers' Practice and Simulation. In: Fauquet-Alekhine P and Pehuet N (eds) *Simulation Training: Fundamentals and Applications Improving Professional Practice Through Simulation Training*. Cham: Springer International Publishing, pp. 1–30.

Walsh A (2009) Information literacy assessment: Where do we start? *Journal of Librarianship & Information Science* 41(1): 19–28.

Accepted Manuscript

Wei X, Zhang Y and Gwizdka J (2014) YASFIIRE: yet another system for IIR evaluation. In: *Proceedings of the 5th Information Interaction in Context Symposium*, New York, NY, USA, 26 August 2014, pp. 316–319. IIiX'14. Association for Computing Machinery. DOI: 10.1145/2637002.2637051.

Wildemuth BM, Marchionini G, Fu X, et al. (2019) The usefulness of multimedia surrogates for making relevance judgments about digital video objects. *Information Processing & Management* 56(6): UNSP 102091. DOI: 10.1016/j.ipm.2019.102091.

Accepted Manuscript