A Framework for Search as Learning Experiments: Design, Implementation, and Usability Insights

Joel H. N. de O. Silva[1], **Alfredo Neto**[1], **Breno Rosado**[1], **Marcelo Machado**[1,2], **Jairo F. de Souza**[1] **Sean W. M. Siqueira**[2]

[1]LApIC Research Group – Department of Computer Science
Federal University of Juiz de Fora – Juiz de Fora — MG – Brazil

[2]Postgraduate Program in Informatics
Federal University of the State of Rio de Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brazil

{joel.henrique, alfredo.lucas, breno.rosado}@estudante.ufjf.br
jairo.souza@ufjf.br, marcelo.oc.machado@gmail.com
sean@uniriotec.br

Abstract. *Search as Learning (SAL) explores how users engage with search systems to acquire knowledge and develop understanding. Despite advances in SAL, the lack of general-purpose tools hinders reproducibility and standardization in experimental studies. This paper presents a framework to support researchers in designing SAL experiments, encompassing task creation, data collection, and learning assessment. To evaluate the proposal, we conducted a usability study with 12 participants, which yielded a score of 83.07, indicating excellent usability. Feedback of the participants also provided suggestions for improvement, guiding future development. This work contributes to strengthening methodological practices and fostering reproducibility in SAL research.*

1. Introduction

The increasing reliance on web search (and chatbot interfaces) as a tool for learning has gained significant attention across information science, education, and human-computer interaction communities. This phenomenon, commonly referred to as Search as Learning (SAL), explores how users engage with search engines not merely to retrieve facts but to construct understanding, acquire new knowledge, and develop skills [Rieh et al., 2016, Vakkari, 2016]. Understanding user behavior in SAL contexts is crucial for designing more effective search systems, supporting users' learning processes, and addressing broader educational challenges in the digital age [Machado et al., 2024b].

Despite the growing body of research on interactive information retrieval (IIR), many of the methodologies and systems created may not be suitable for the SAL context [Urgo and Arguello, 2022]. Therefore, general-purpose systems for creating experiments are scarce, culminating in the destructuring of the body of research [Machado et al., 2024a]. In other words, existing empirical experimentation in the field remains confined to isolated platforms, hindering reproducibility and collaborative progress within the field. This limitation underscores a critical need for a more unified approach to experimentation.

To address this gap, we have proposed a framework to assist researchers in designing experiments aimed at understanding user search behavior in SAL tasks. This framework guides key decisions in task construction, behavioral instrumentation, and

DOI: 10.5753/sbie.2025.12265

evaluation of learning outcomes. We divided its implementation and evaluation into two cycles. The first cycle focused on the user/learner experience of using an instance of the framework in experimentation scenarios [Machado et al., 2024b]. Now, we are including the researcher's perspective in creating real experiments using our framework. Therefore, the main objective of this work is to evaluate the researcher's experience when using our proposed artifact. Our ultimate goal is to provide a structured approach that improves consistency, transparency, and reproducibility in SAL research.[1]

To evaluate this second cycle, we conducted an experimental usability evaluation study with 12 participants who assumed the role of researchers in creating SAL experiments within the system. Upon completing their tasks, participants responded to the System Usability Scale (SUS) survey [Brooke, 1996]. The system achieved a SUS score of 83.07, reflecting a high level of usability. Feedback from open-ended questions revealed suggestions for new features, as well as identified pain points and areas for improvement.

The remainder of this paper is organized as follows: Section 2 provides a concise background on SAL and related works; Section 3 details the proposed artifact; Section 4 presents how we conducted an experimental usability study to evaluate our proposal, while Section 5 presents and discusses the results; Finally, Section 6 concludes the paper and offers insights into future work.

2. Background and Related Work

Using the Internet to access information is a common activity in daily life, and search systems are strong allies in this process. Traditional search systems, such as Google, are designed and optimized for direct informational searches on specific issues. However, most web search activities do not merely consist of simply looking up a specific piece of information; most often, they are complex and exploratory in nature [Marchionini, 2006].

This phenomenon is captured by the concept of Search as Learning, which refers to the process in which search behavior is directly linked to learning outcomes. SAL views the search process not merely as information retrieval but as a cognitive and constructive learning activity where users acquire, process, and integrate new knowledge [Ghosh et al., 2018]. Throughout SAL tasks, individuals typically engage with a variety of information sources, take notes, decompose the main learning objective into manageable subgoals, and revisit subjects to deepen and verify their comprehension [Urgo and Arguello, 2022]. Research in SAL emphasizes that the act of searching goes beyond simply locating information, it involves the active construction of new knowledge and meaningful understanding [Rieh et al., 2016].

SAL is a multidisciplinary research area that considers a wide range of research questions. To name a few, some research focus on understanding the real-world contexts in which people search for the purpose of learning, e.g., [Gimenez et al., 2020]; Other might focus on developing tools to encourage and support learning during search [Machado et al., 2024a]. Studies might also focus on discovering search behaviors that predict learning during search [Hoppe et al., 2025, Machado et al., 2024b, Tibau et al., 2022]. Finally, an important branch of SAL is concerned with methodological de-

[1]The framework is open-source and can be accessed here `https://github.com/orgs/Framework-for-Search-as-Learning/repositories`.

velopment, aiming to advance how SAL phenomena are studied [Machado et al., 2020, von Hoyer et al., 2022].

When it comes to tools for conducting experiments, some studies have pointed out the absence of general-purpose platforms that could support diverse research needs. For instance, Costa et al. [2025] highlighted significant challenges in designing, orchestrating, and managing user studies, including limited support for task creation, user management, and data collection. Although focused on Software Engineering, these findings are also applicable to SAL, where similar methodological gaps exist.

In the field of IIR, tools such as SCAMP [Renaud and Azzopardi, 2012] and RAT [Sünkler et al., 2023] have been developed to support the design and execution of experiments. SCAMP facilitates lab-based IIR experiments by providing infrastructure for participant registration, consent, and data collection. RAT focuses on evaluating search engine results and includes scraping tools and interfaces for judge-based assessments. However, both tools fall short for SAL experiments: SCAMP supports only within-subject designs, where all participants are exposed to the same condition, and lacks capabilities for implicit logging – such as the automatic capture of user interactions (e.g., mouse movements and clicks) – which is crucial for non-intrusive behavioral tracking in SAL studies. RAT, in turn, is designed for post hoc evaluation of search results using pre-collected data and does not track users' behavior during the search process. As a result, it does not support the full experimental cycle typical of SAL, which includes task design, in-situ behavioral data collection, and learning outcome assessment.

In SAL, the challenges are even greater. Capturing natural user behavior during learning-oriented search is critical, and in the absence of appropriate tools, many studies relied on observational methods, such as Think-Aloud protocols (e.g., [Tibau et al., 2021, Yang et al., 2025]). Although Think-Aloud is a valid method, particularly useful in exploratory studies, it can interfere with natural user search behavior [Fox et al., 2011]. Therefore, reliance on methods that potentially alter user behavior also highlights the need for a flexible and reusable system that can capture rich behavioral data unobtrusively while supporting the design and execution of SAL experiments.

To address these limitations, we introduce a framework designed to support researchers conducting SAL and IIR experiments. It facilitates defining experiments with various study designs (i.e., within-subject and between-subject), assigning experimental groups using different rules, managing participants, collecting interaction data, deploying questionnaires at multiple stages, selecting search engines, and defining evaluation metrics. Additionally, all data can be easily exported and imported across instances of the framework, enhancing reproducibility and enabling further comparisons.

3. Proposal

Inspired by Design Science Research (DSR) epistemology [Pimentel et al., 2020], we divided the design of our artifact into iterative cycles. The first cycle describes the initial version of the framework, which was designed with a focus on the user/learner view. The second cycle details the current version of the framework, which shifts the focus to the user/researcher.

3.1. First Cycle: learner-centered

In [Machado et al., 2024a], we introduced a framework designed to support the execution of experiments on search behavior within the context of SAL. The requirements were derived from comprehensive literature reviews that covered both SAL studies and research on cognitive biases in search. This latter area is particularly concerned with understanding user search behavior when interacting with search engines (e.g., see [Azzopardi, 2021]), therefore, it is a rich source of features. The goal was to understand how experiments in these domains were typically conducted and to identify common requirements in existing approaches.

As shown in Figure 1, the proposed framework is structured into layered components that delineate the interaction between users, the *front-end* interface and the core logic of the system. At the center of the architecture lies the *Application Layer*, which serves as the primary interface between users and the system. This layer, implemented as the front-end, is responsible for capturing user interactions and relaying them to the *back-end* services. The *Experiment Layer*, positioned on the back-end, handles key operations such as data validation, experimental logic orchestration, including the scheduling of activities and participant group separation, and the semantic structuring of experiment data.

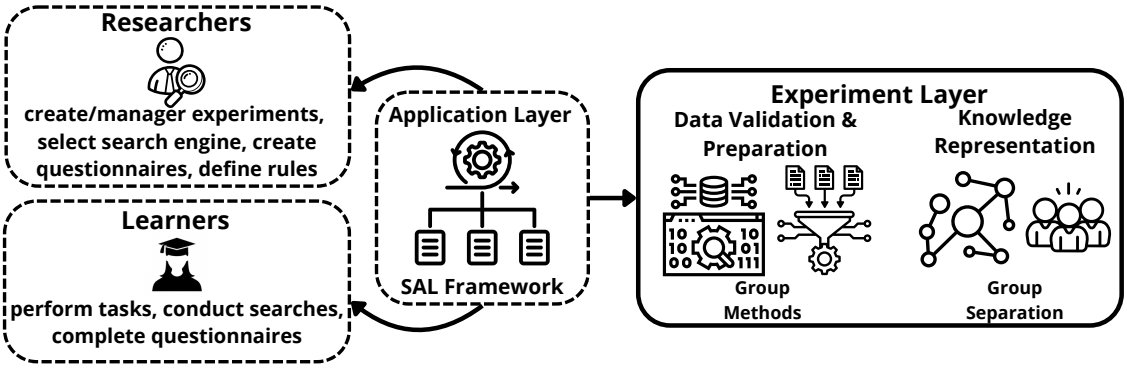

Figure 1. Layered architecture of the system. The Application Layer (front-end) mediates user interaction and triggers back-end processes in the Experiment Layer, which is responsible for data validation, experimental logic, and group structuring. Researchers and learners access the system through the Application Layer to either manage or participate in experiments. Dashed lines indicate front-end components, while solid lines represent back-end operations.

Two user roles interact with the system: Learner and Researcher. Learners engage with the experimental tasks by conducting searches and completing questionnaires. During the learner's interaction, the system unobtrusively logs behavioral data–such as query length, number of queries, scrolling activity, and interaction patterns–alongside learners' explicit responses to questionnaires. On the other hand, researchers are able to create and manage experiments, select search engines, define rules for group separation, monitor learner progress, view collected data, and analyze experimental outcomes directly within the system.

The evaluation of the framework in the first cycle was conducted through a real-world study to investigate search behavior in SAL contexts and collect user behavior data [Machado et al., 2024b]. The framework proved effective, supporting more than 100

learners whose interaction data was systematically captured and structured to facilitate subsequent analysis by researchers.

3.2. Second Cycle: researcher-centered

In the second cycle, we seek to provide an intuitive API and a graphical interface that allows researchers to create experiments in a flexible manner, accommodating different types of experimental design, as illustrated in Figure 2.

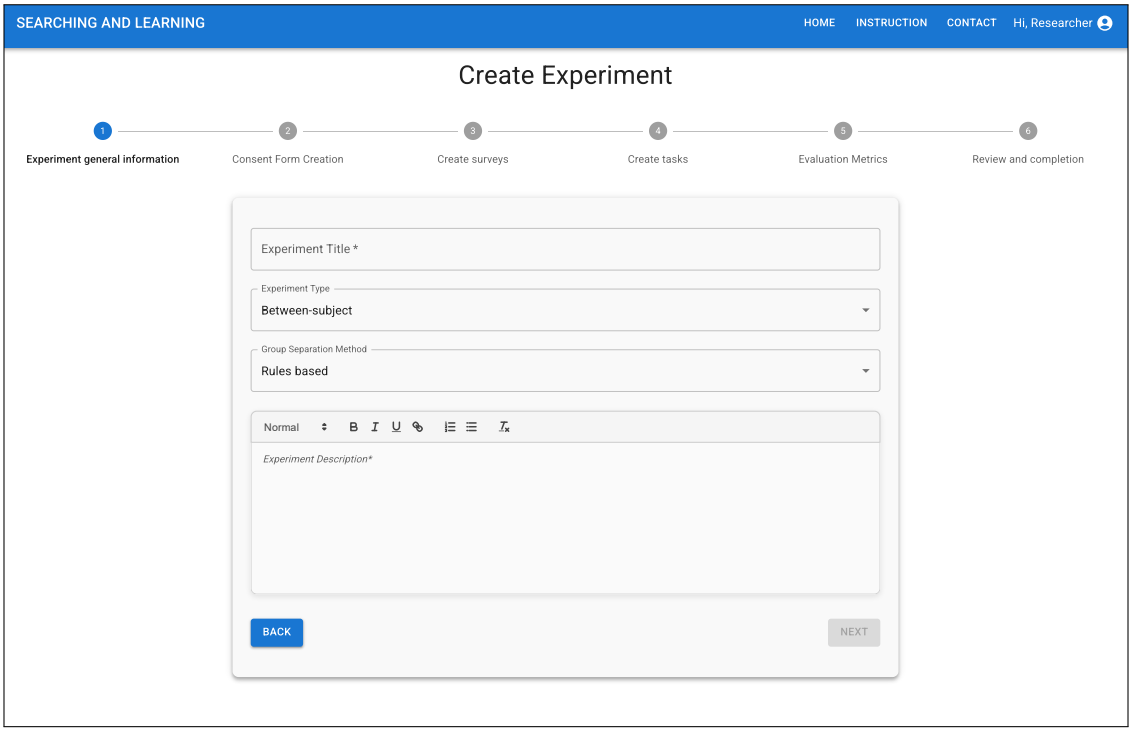

Figure 2. Graphical interface of the framework, illustrating the initial step in configuring a new experiment.

As shown in Figure 3, creating a new experiment is done in sequential steps. This initial model was conceived to encourage the researcher to structure their objectives well from the start, reducing the need for later rework. Even so, all steps are editable: it is possible to return to previous steps and make adjustments as needed. To advance from one step to the next, it is required that all mandatory fields be duly filled in. Below, we describe each of these steps in detail, explaining the purpose and the required inputs at each stage. We start with the first step, which establishes the foundation of the experiment and guides the subsequent steps.

Figure 3. Diagram of the sequential steps involved in creating a new experiment through the interface

In the first step, the researcher is prompted to establish the research question, which corresponds to the phenomenon to be investigated through the manipulation and

observation of specific variables and constitutes the central objective of the experiment, and may then choose between two experimental designs: *within-subject* or *between-subject*. In a within-subject design, all learners are exposed to every condition being investigated. For instance, the researcher can compare two search interfaces where each learner might use both interfaces to complete equivalent learning tasks, allowing the researcher to directly compare performance and user experience across all conditions within the same individuals. Whereas, in choosing between-subject, the researcher will be attesting that each group of learners will be exposed to a single condition of the independent variable. For example, one group might use a search system with query suggestions enabled, while another uses the same system without suggestions. In between-subject, group separation can occur in three ways: manual, random, and rule-based. Finally, the researcher prepares a brief overview of what the experiment entails for the learners, providing initial context and clarifying the study's purpose.

The second step consists of defining the Informed Consent Form (ICF). Here, the researcher drafts the text that will be presented to the learner before the experiment is conducted, guaranteeing that they are aware of the study's objectives and conditions.

In the third step, the researcher creates and configures questionnaires (a.k.a. surveys). Each questionnaire is defining by its title, description, and whether it will be used as a demographic or pre- or post-questionnaire. A demographic questionnaire serves to capture general information that characterizes the participants (e.g., gender, age, occupation). The pre-questionnaire is typically used to collect variables that aid in group separation (e.g., position on a controversial topic, knowledge about a subject), while the post-questionnaire seeks to gather data that answer the research question (e.g. new position on a controversial topic, level of knowledge acquired). Regarding the items, it is possible to choose among types such as open-ended, multiple-choice, and checkboxes. It is also possible to mark them as mandatory and decide if they will be scored, assigning different weight to each option of a given question as illustrated in Figure 4.

In the fourth step, the researcher defines the search tasks that participants will perform, with the aim of collecting data to validate the research question. In creating the task, researcher can select the search engine to be used, according to the focus and objectives of the study. The options are Google, Bing, or a conversational agent as ChatGPT, in future work we intend to allow researchers to configure their own search engine 5.

In a scenario in which the researcher chooses the between-subject study-design, they must specify what the tasks are for each of the groups of participants. In manual separation, the researcher directly determines the number of tasks (groups), and manually selects which participants will be part of each group. In random separation, participants are automatically assigned to different tasks, observing only basic balancing criteria to avoid disproportions. Finally, rule-based design includes quasi-experiments, where the researcher decides to separate the groups according to some variable captured in the pre-questionnaires. For instance, the researcher may decide to separate the groups based on the profession of the participants, or according to a position on a controversial topic, or even based on their previous knowledge about a topic. The researcher may choose to use the score of an entire questionnaire or just some questions 5.

Finally, the researcher must select evaluation metrics that reflect participants'

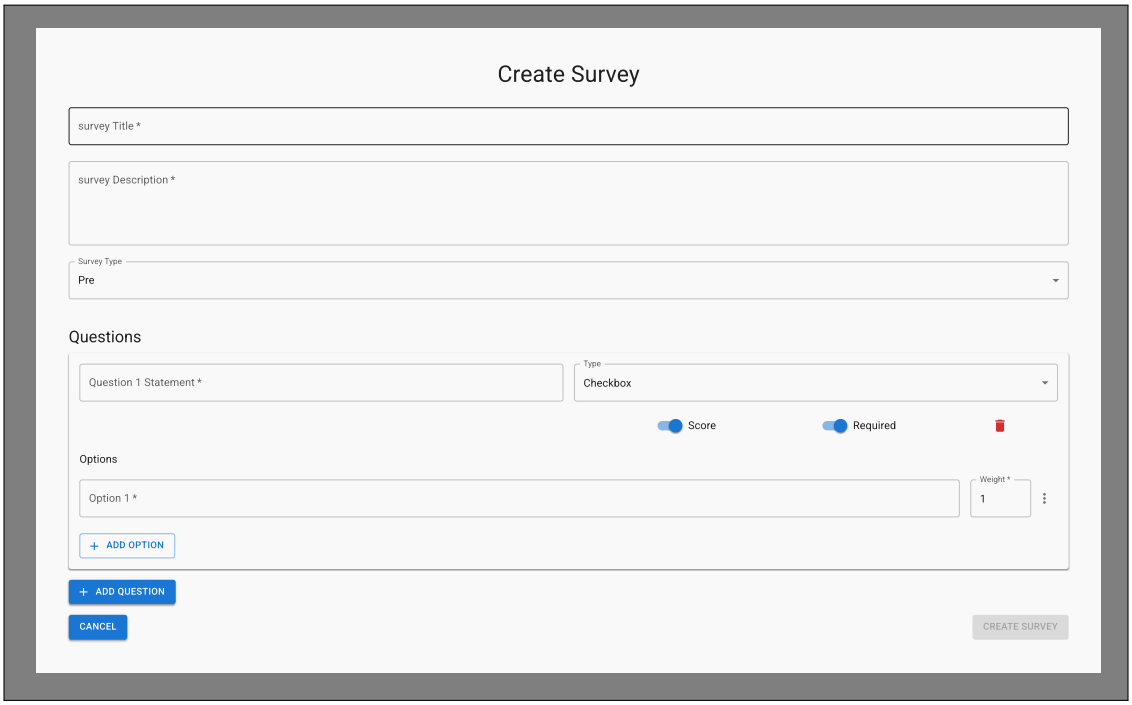

Figure 4. Graphical interface of the framework illustrating the step for creating a new survey.

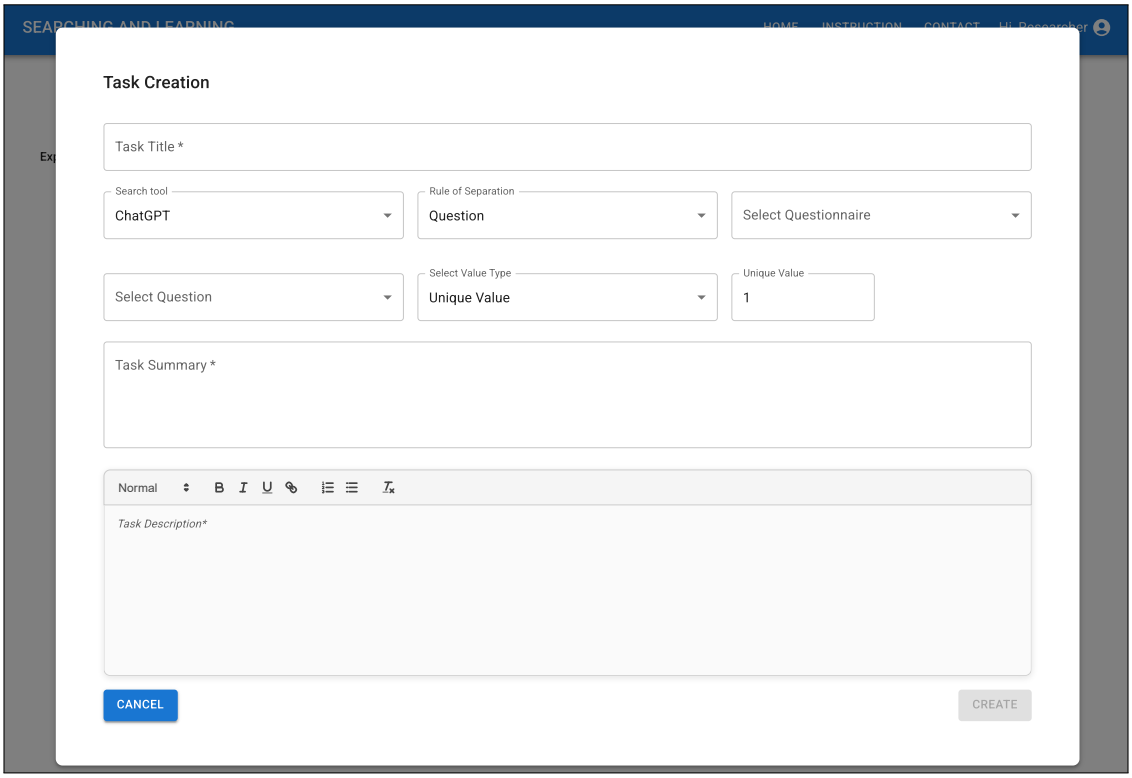

Figure 5. Graphical interface of the framework illustrating the step for creating a new task.

search behavior during the experiment. Examples include the number of queries issued, the number and rank of accessed results, and the number of chat messages exchanged. The choice of metrics may vary depending on the search engine used.

Once the tasks are defined and the participants assigned, the researcher can review all the details of the experiment and make any necessary edits.

4. Evaluation

We recruited 12 Brazilian students who are majoring in Computer Science, Letters, and Information Systems – without associated costs or benefits. All participants signed the ICF and received a short training about observational and experimental research methods. After the training, participants were required to design an experiment on SAL and then add the experiment specification on the framework.

Participants were divided into four groups. Each group was responsible for designing an experiment based on a different type of learner allocation. The *Within-Subjects* group designed an experiment in which all participants were exposed to the same conditions. The *Between-Subjects* groups designed experiments where participants were divided either randomly, manually, or based on rules. After completing the experiment setup, participants were instructed to edit the experiment details. Once finished, they were directed to complete a questionnaire via Google Forms.

To guide this research towards its objective, a survey was conducted to understand participants' impressions after using the system to create an experiment. Survey-based research allows the collection of information by questioning a group of people in a systematic manner, within a specific domain, enabling the acquisition of knowledge and relevant information from respondents quickly and at low cost.

The survey, created using Google Forms, aimed to answer the main research question: "What is the perception of researchers regarding the use of our system for creating experiments on SAL?". The data analysis followed a mixed-methods approach: quantitative analysis, based on counts and percentages, presented through graphs and tables for the closed-ended questions; and qualitative analysis, involving the interpretation of open-ended responses.

The System Usability Scale (SUS), introduced by Brooke [1996], provides a simple and reliable way to measure system usability. It has become widely adopted for its balance of scientific rigor and ease of use, enabling quick assessments by both researchers and users. Designed to evaluate user interfaces, SUS focuses on effectiveness, efficiency, and user satisfaction. In our study, we follow Bangor et al. [2009], who confirmed the scale's reliability and enhanced its interpretability by adding an adjective rating scale. These qualities make SUS well-suited for evaluating systems in practical settings, such as our framework implementation.

Our questionnaire includes 16 statements rated on a 5-point Likert scale, where 1 means "strongly disagree" and 5 means "strongly agree". The survey is administered immediately after participants attempt to complete a predefined set of tasks within the system.[2]

[2]The questionnaire can be accessed here: `https://docs.google.com/forms/d/e/1FAIpQLSfy2G_jWqeXJ93XxASVWk2g7aKSOLpmLNgCOl7jyHoon48o8Q/viewform`

5. Results and Discussion

To assess the usability of the developed system, we conducted a mixed-method evaluation combining quantitative metrics with qualitative feedback.

Participants were asked to interact with the system through tasks that reflected its full set of features. Following this hands-on experience, they completed the SUS questionnaire and provided open-ended feedback. The results, shown in Table 1, offer insights into the system's usability performance and identify concrete areas for refinement in subsequent design iterations.

As shown in Table 1, the system achieved a SUS score of 83.073, surpassing the threshold of 80 proposed by Lewis and Sauro [2018], which is commonly interpreted as indicative of good usability. In addition, participants described the user interface as intuitive in their qualitative feedback, but suggested minor adjustments to further improve clarity. These findings support the artifact evaluation phase of the DSR methodology and will guide improvements for the next design cycle.

Table 1. SUS Statistics by group

Group	*N*	Mean SUS Score	Standard Deviations
Within Subjects	4	79,297	10,55
Between Subjects, Random	4	77,344	1,56
Between Subjects, Rules Based	2	96,875	4,44
Between Subjects, Manual	2	88,281	1,10
All	12	83,073	9,06

In addition to the quantitative results obtained, we collected open-ended qualitative feedback to further contextualize users' experiences and perceptions of the system. Although the overall SUS score of 83.073 indicates good usability [Lewis and Sauro, 2018], the comments of the participants provide valuable insight into specific challenges in interaction and opportunities for improvement.

Several users described the interface as intuitive and positively acknowledged the system's general functionality. However, recurring suggestions highlighted areas where usability could be improved. These include the lack of clarity in the required fields, limitations in navigation, such as the inability to return to previous steps by clicking on the progress indicators, and visibility and placement of key interface elements during the creation process.

Participants also reported issues such as: (1) inconsistent behavior when editing task descriptions that include bullet points, (2) difficulty in locating or removing subitems within tasks, (3) confusion caused by the position of critical fields (e.g., experiment type selectors), (4) limited ability to reorder questionnaires or tasks after creation, (5) unintuitive flow after completing core actions (e.g., not being redirected to the home screen), and (6) access errors when editing consent forms or viewing user-assigned experiments.

These qualitative insights complement the SUS findings and emphasize the importance of iterative design in DSR-based development and evaluation. The reported issues, although not critical, will inform refinements in both the interaction flow and visual hierarchy to improve clarity, control, and user confidence in future development cycles.

6. Concluding Remarks

This paper presented an evolution of our proposed framework created to support the design and execution of experiments in the SAL domain. Motivated by the lack of general-purpose tools for this type of research, our framework addresses key challenges related to the reproducibility, consistency, and transparency of SAL experiments. It guides researchers through critical stages, including task design, behavioral instrumentation, and the assessment of learning outcomes.

Inspired by the DSR methodology, we structured the design, development, and evaluation of the framework into two iterative cycles. While the first cycle focused on supporting the user/learner, this study extended the evaluation to the researcher's perspective. Our usability study with 12 participants yielded an SUS score of 83.07, indicating excellent usability and alignment with the framework's intended purpose.

Beyond usability validation, the study also highlighted relevant functional and methodological contributions. Participants reported that the framework facilitated the creation of SAL experiments by offering an integrated environment that streamlines common tasks such as group management, task configuration, consent form generation, and questionnaire elaboration. These findings reinforce the framework's potential to reduce the technical barriers often associated with experimental research in this field.

Nonetheless, the feedback also surfaced limitations that will guide future improvements. Some participants pointed to the need for greater flexibility in task design, more intuitive navigation between steps, and enhanced support for customizing data collection instruments. These suggestions have been incorporated into the development backlog, forming the foundation for what we envision as the third design cycle of this framework.

By offering a reusable and extensible platform, this work contributes to advancing methodological practices in SAL research. We believe that fostering reproducibility and standardization in experimental design can significantly accelerate cumulative knowledge building in the field. Moreover, the framework promotes collaboration among researchers, reducing the need to develop ad-hoc experimental systems for each study.

For future work, we plan to extend the framework with features that support more complex experimental designs, including adaptive tasks, robust integration with generative AI tools, and real-time monitoring of learner interactions. Additionally, we intend to conduct longitudinal studies to assess the framework's impact on research productivity and the quality of SAL investigations over time. We also envision creating a repository of reusable experiment templates to further support the SAL research community.

Acknowledgment

This study was partially funded by the 'National Council for Scientific and Technological Development – CNPq' (Process 305436/2021-0) and by the Coordination of Superior Level Staff Improvement (CAPES)—Support Program for the Dissemination of Scientific and Technological Information (PADICT) and CAPES Periodical Portal)—Finance Code 001.

References

Azzopardi, L. (2021). Cognitive biases in search: a review and reflection of cognitive biases in information retrieval. In *CHIIR 2021 - Proceedings of the 2021 Conference*

on Human Information Interaction and Retrieval, CHIIR 2021 - Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, page 27–37.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123.

Brooke, J. (1996). Sus: A quick and dirty usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, I. L., editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis, London. Capítulo de livro.

Costa, L., Barbosa, S., and Cunha, J. (2025). Mind the gap: The missing features of the tools to support user studies in software engineering. *Journal of Computer Languages*, 84:101345.

Fox, M. C., Ericsson, K. A., and Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? a meta-analysis and recommendations for best reporting methods. *Psychological bulletin*, 137(2):316.

Ghosh, S., Rath, M., and Shah, C. (2018). Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, CHIIR '18, page 22–31, New York, NY, USA. Association for Computing Machinery.

Gimenez, P., Machado, M., Pinelli, C., and Siqueira, S. (2020). Investigating the learning perspective of searching as learning, a review of the state of the art. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 302–311, Porto Alegre, RS, Brasil. SBC.

Hoppe, A., Yu, R., Liu, J., and Bhattacharya, N. (2025). Iwilds'25: The 5th international workshop on investigating learning during web search. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, WSDM '25, page 1116–1117, New York, NY, USA. Association for Computing Machinery.

Lewis, J. R. and Sauro, J. (2018). Item benchmarks for the system usability scale. *J. Usability Studies*, 13(3):158–167.

Machado, M., Assis, E. C., Souza, J. F., and Siqueira, S. W. M. (2024a). A framework to support experimentation in the context of cognitive biases in search as a learning process. In *Proceedings of the 20th Brazilian Symposium on Information Systems*, SBSI '24, New York, NY, USA. Association for Computing Machinery.

Machado, M., Gimenez, P., and Siqueira, S. (2020). Raising the dimensions and variables for searching as a learning process: A systematic mapping of the literature. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1393–1402, Porto Alegre, RS, Brasil. SBC.

Machado, M., Souza, J., and Siqueira, S. (2024b). Identifying confirmation bias in a search as learning task: A study on the use of artificial intelligence in education. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 1208–1221, Porto Alegre, RS, Brasil. SBC.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46.

Pimentel, M., Filippo, D., and Dos Santos, T. M. (2020). Design science research: pesquisa científica atrelada ao design de artefatos. *RE@ D-Revista de Educação a Distância e eLearning*, 3(1):37–61.

Renaud, G. and Azzopardi, L. (2012). Scamp: A tool for conducting interactive information retrieval experiments. *IIiX 2012 - Proceedings 4th Information Interaction in Context Symposium: Behaviors, Interactions, Interfaces, Systems*.

Rieh, S. Y., Collins-Thompson, K., Hansen, P., and Lee, H.-J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, 42(1):19–34.

Sünkler, S., Yagci, N., Sygulla, D., Mach, S., Schultheiß, S., and Lewandowski, D. (2023). Result assessment tool: A software toolkit for conducting studies based on search results. *Proceedings of the Association for Information Science and Technology*, 60:1143–1145.

Tibau, M., Siqueira, S. W. M., and Nunes, B. P. (2021). Think-aloud exploratory search: Understanding search behaviors and knowledge flows. In Visvizi, A., Lytras, M. D., and Aljohani, N. R., editors, *Research and Innovation Forum 2020*, pages 303–315, Cham. Springer International Publishing.

Tibau, M., Siqueira, S. W. M., and Nunes, B. P. (2022). The impact of non-verbalization in think-aloud: Understanding knowledge gain indicators considering think-aloud web searches. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, HT '22, page 107–120, New York, NY, USA. Association for Computing Machinery.

Urgo, K. and Arguello, J. (2022). Understanding the "pathway" towards a searcher's learning objective. *ACM Trans. Inf. Syst.*, 40(4).

Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1):7–18.

von Hoyer, J., Hoppe, A., Kammerer, Y., Otto, C., Pardi, G., Rokicki, M., Yu, R., Dietze, S., Ewerth, R., and Holtz, P. (2022). The search as learning spaceship: Toward a comprehensive model of psychological and technological facets of search as learning. *Frontiers in Psychology*, Volume 13 - 2022.

Yang, Y., Urgo, K., Arguello, J., and Capra, R. (2025). Search+chat: Integrating search and genai to support users with learning-oriented search tasks. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, CHIIR '25, page 57–70, New York, NY, USA. Association for Computing Machinery.