

Design Activity Worksheet: Identifying Data Sources for Developing a Research Plan

Vetria L. Byrd
Computer Graphics Technology
Purdue University
West Lafayette, USA
0000-0002-0733-2062

Abstract— This research to practice paper presents results of students' perception of a worksheet method for identifying appropriate data sources when developing research questions. Identifying reliable and trustworthy data sources takes time and experience. The purpose of this study is to help students plan, generate, and determine the availability of data for research. The worksheet was completed by 32 undergraduates enrolled in a data visualization course. Results show 53% of undergraduates who utilized the worksheet indicate the worksheet helped to identify primary data sources for their topic, 66% indicate the worksheet helped to identify secondary data sources, 59% indicate the worksheet helped to identify tertiary data sources, and 69% indicate the worksheet helped to assess if there is enough data for their research topic. This research informs scaffolding practices for undergraduate and graduate students as they learn and implement research methodology. Implications of this work will help students understand different types of data sources, the role each type of data source plays in developing a research plan, and how to determine the type of data needed to support their research.

Keywords— *student perception, information literacy, research to practice*

I. INTRODUCTION

Engineers must continuously renew their knowledge beyond the fundamental concepts and skills traditionally taught in undergraduate programs in order to keep up with rapidly evolving technologies and job responsibilities [1]. To stay competitive, students must also develop data and information literacy in the context of technical programs. Information literacy is a blend of self-directed learning and reflective judgment that broadly encompasses the ability to plan and pursue information searches, as well as the skills necessary to evaluate the accuracy of information and the quality of information sources [2]. In some situations, information seekers will readily sacrifice content for convenience [3]. In this work, we introduce an activity worksheet designed to motivate students to think more critically about data and their process for identifying data sources.

A. Theoretical Frameworks

Theoretical frameworks previously examined for assessing information literacy skills for engineering design tasks [1] are utilized in this work. Kuhlthau's [4] information search process (ISP) model and the Society of College, National, and

University Libraries (SCONUL) pillars of information literacy model (Society of College, National, and University Libraries) [5] frameworks support this work. The ISP model is widely respected, particularly within information science communities, as being a research-based, holistic model of information seeking behavior [6]. The ISP model starts with the recognition for the need for information and progresses to a critical point where the individual becomes more aware of the full range of information available within topics explored [1]. When developing research questions and designing research plans, identifying data sources is the step that connects the two. This is where people tend to feel most doubtful, frustrated, and anxious as a result of finding too much, not enough, or conflicting information [1].

To address this issue, three aims are identified for this research. The first aim of the study is to introduce students to different types of data sources. The second aim is to introduce relevance and reliability of data sources. A third aim is to assess student's perception of the usability of an activity worksheet in helping to identify relevant and reliable data sources. In this paper results are presented from undergraduates who used the "From a Problem to Data Sources" Activity Worksheet to motivate critical evaluation of data sources and information identified for a semester research project.

This research is driven by two research questions:

R1: How do students perceive the usability of the worksheet for identifying data sources.

R2: Do students' perception of the worksheet differ between undergraduates and graduate students?




B. Implications of This Work

This research is significant because it will help to better inform educational practice and pedagogy in research methods. Outcomes from this work will inform scaffolding practices for undergraduate and graduate students as they learn and implement research methodology. Implications of this work will help students understand different types of data sources (primary, secondary and tertiary), the role each type of data source plays in developing a research plan and how to determine the type of data needed to support their research.

Although presented as a mechanism to help facilitate research, the worksheet can also be used in a regular course for developing course projects that require the acquisition of data. The

worksheet can also be used as a starting point for exploratory study and developing ideas to build on based on the availability and identification of appropriate data sources.

TABLE I. TOPICS TO PROBLEM STATEMENT ACTIVITY WORKSHEETS (PRIOR WORK #1)

	 Identifying Topics	 Topics to Questions	 Problem Statement
Goals	Generate research topics.	Generate research questions.	Develop a sound problem statement.
Objectives	Students will identify a list of topics of interest.	Students will transition from a broad topic to a specific research question.	Students will transition from a broad topic to a specific question.
Outcomes	Students will create a list of ranked topics of interest and identify who the intended audience is.	Students will answer who/what/when /where questions to provide an in depth description of their topic.	Students will identify their research problem as a practical or conceptual problem and explain its significance.

II. BACKGROUND

A. Previous Work

In previous work ([7], [8], [9], [10], [11]), activity worksheets were developed to provide guided instructions for identifying topics, identifying unexplored questions about topics and developing a sound problem statement to answer those questions. Table I shows goals, objectives and outcomes for each worksheet.

B. Relevance and Reliability Literature

Often times, for students, the search for data sources begin and end with a generic Google search with surface level results. The ease of locating information on the internet impacts students' motivation to look beyond the internet for data sources. The ease of access to information also solidifies the internet as a first choice of students when identifying and locating data. Watson [12] identified five themes regarding relevance and reliability criteria students employ when searching for data and information to support research questions: a) socio-academic context, b) convenience or pragmatic approach, c) relevance priming, d) relevance chaining, and e) knowledge building. Watson [12] summarizes the five elements as:

a) *Socio-academic context.* The use of the reputation of a source to judge reliability is commonly derived from the views of others in the social and academic spheres of a student's life [12].

b) *Convenience or pragmatic approach.* Convenience encompasses the information source chosen by people, their level of satisfaction with it

and how easy they find it to use , and the time they have available [3].

c) *Relevance priming.* To find a prime source, students depend heavily on the first or early cooccurring results returned by a search engine; the information in the article is presumed by the student to be accurate, but only provisionally, and then forms the basis on which to build a fuller picture of the topic under investigation [12].

d) *Relevance chaining.* The process-oriented behavior that occurs when a participant builds on prior knowledge, generally after first having developed an overview and then finding subsequent sources that relate to the framework built from that overview [12].

e) *Knowledge building.* Student's knowledge building takes three major forms: filtering information, matching information, and adding information [12].

C. Utility of the Worksheet Relative to Other Guidelines

There are numerous tools and guidelines for identifying data sources, typically available through library resources. For example, the "Guide to Academic Research," provided by [13] discusses data as sources, ways of obtaining data, examples of data specific data repositories, as well as criteria for evaluating data for relevance and credibility. This resource provides example scenarios for evaluating data as sources. Iowa State University's LIB 160: Information Literacy webpage [14] lists "three major tools for finding information: web search engines, library discovery tools and article indexes." The web site provides a list of "finding tool features," where to find different types of information and an opportunity to self-check how well a viewer understands the differences between the three types of information finding tools. The University of Cincinnati Libraries' website [15] provides a library research guide for undergraduate programs with along with strategies for successful searching, choosing the right tool for the task and general tips for developing search terms to yield usable results.

Completion of the worksheet described in this work helps students to articulate what their topic is, describe the research goal, and describe the intention for the reader (what the reader should better understand as a result of completing the described research). Having a clear understanding of the 'What' and 'Why' of the research question provides a better idea of what to look for when using tools and guidelines for identifying data sources.

III. METHODOLOGY

A. Problems to Data Sources Worksheet

The "Problems to Data Sources" worksheet consists of four questions designed to motivate students to think critically about data sources for their research project. The goal of the worksheet is to help students identify appropriate data sources to facilitate the transition from a broad topic to a specific question. The objective of the worksheet is to plan, generate, and determine the availability of data for the research project/topic. Outcomes

from the worksheet include a list of appropriate data sources and description of data required to complete the project.

From a Problem to Data Sources ○ ○ ○ ● ○

Goal: To identify appropriate data sources and transition from a broad topic to a specific question.

Objectives: To plan, generate, and determine the availability of data for the research project.

Outcomes: List of appropriate data sources and data required to complete the project.

1) Describe your semester topic below

- I am working on the topic of: *describe your topic*
- Because I want to find out: *describe research goal*
- In order to help my reader understand better: *describe intention for reader*

I need data to support my topic.

2) What are the data you will need for your research topic? Consider talking to a Librarian to help identify appropriate data sources.

Identify and list a minimum of 3 data sources. You should identify a variety of data sources. If all of your data sources are internet web sites, this page will be returned to you to find additional data sources.

Primary Sources: *Raw Data; where application*

Secondary Sources: *Literature; must have at least 1 secondary source, best secondary sources are book from university press and/or articles that have been peer reviewed.*

Tertiary Sources: *Books and Articles that synthesize secondary data for general readers*

3) How did you evaluate the data sources provided above for relevance and reliability?

4) If the data does not exist yet, describe the data (variables, information) you will need for your research.

Fig. 1. From a Problem to Data Sources Activity Worksheet.

Fig. 1 shows a view of the “From a Problem to Data Sources” Activity Worksheet. The worksheet consists of four questions. The first question asks students to describe the semester topic. In order to determine what data is needed, it is important to know what question(s) are being asked. It is assumed, at this point, students have identified a topic [8], transformed the topic into a question [9], and utilized the question to clearly state the problem [10] to be addressed by the proposed research.

The second question asks students what data is needed for the research. Students are encouraged to talk with a librarian to help identify appropriate data sources. Students are asked to identify a minimum of three unique data sources. A variety of data sources is preferred. Content covered in the worksheet can be used as talking points for in-class discussions. For example, students are asked to identify primary, secondary and tertiary data sources to support their research. To ensure students understand differences between the three types of data sources, in this research, class discussions included time dedicated to explaining the differences and providing examples.

The third question asks, “How did you evaluate the relevance and reliability of the data source(s)?” This question sets the stage for a discussion about relevance and reliability of information as described by relevance and reliability literature. The last question asked students “If the data does not exist yet, describe the data (variables, information) needed.

In this study, the “From a Problem to Data Sources” Activity Worksheet is introduced in parallel with discussions about relevance and reliability of data sources. The internet makes an abundance of information available almost instantly. It is important to discuss with students the reliability of internet sources and stress availability does not imply reliability or accuracy of content.

B. Participants

The worksheet was completed by 32 undergraduates enrolled in two sections of a data visualization course offered in

fall 2019. The course is a gateway course for the undergraduate data visualization major at a research-intensive university in the Midwest United States. Students were sophomore or higher-level status (junior or senior) in their academic programs. Students enrolled in the course had little to no experience in identifying and locating data for research. Undergraduates completed the worksheet in week four of the semester after completing three previous worksheets in the series: Identifying Topics [7], From Topics to Questions [8], and the From Questions to a Problem Statement [9] worksheets. Each worksheet was initially presented as part of an in-class assignment to allow students to ask questions and the instructor to provide clarification if needed. Students were given one week to submit the completed worksheet.

In a previous study, a different approach was taken with graduate students. In a graduate seminar course, graduate students were given access to all of the worksheets (for undergraduates, the worksheets were introduced one at a time, one worksheet a week). Graduate students were asked to complete the worksheets as they developed their developed their research questions. Graduate students were at liberty to submit the worksheets incrementally as they completed them but not later than the last week of class. Both undergraduate and graduate students were encouraged to update the worksheets as needed based on information gained as they looked for data sources.

C. Data Collection

The data for this research consisted of the completed “From a Problem to Data Sources” Activity Worksheet, and students’ self-assessment feedback on the usability of the worksheet. Students completed the worksheet after completing and submitting the first three worksheets in the “From Topics to Data Sources” series of worksheets (see Related Work section). After completing the “From a Problem to Data Sources” Activity Worksheet students were given a self-assessment survey. The survey consisted of four Likert-scale questions with responses ranging from strongly disagree (SD) to strongly agree (SA). Each question is followed by a written response question to allow students to provide a brief explanation to responses that were disagree (D) or strongly disagree. The self-assessment questions are provided in Fig. 2.

D. Data Analysis

Analysis of the data aligns with the first and third aims of the worksheet: 1) to introduce primary, secondary and tertiary data sources and 2) to assess students’ perception of the usability of the activity worksheet in helping to identify relevant and reliable data sources [10]. The activity worksheet was completed by 32 undergraduates enrolled in two sections of a data visualization course offered in fall 2019. The course is a gateway course for the undergraduate data visualization major at a research-intensive university in the Midwest United States. Students were sophomore or higher-level status (junior or senior) in their academic programs. Students enrolled in the course had little to no experience in identifying and locating data for research.

Using the Likert scale below indicate your level of agreement.

Q1. The worksheet helped to identify primary data sources for the project topic.

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

Q1a. If you answered Strongly Disagree or Disagree, please provide a brief explanation.

Q2. The worksheet helped me to identify secondary data sources for the project topic.

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

Q2a. If you answered Strongly Disagree or Disagree, please provide a brief explanation.

Q3. The worksheet helped me to identify tertiary data sources for the project topic.

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

Q3a. If you answered Strongly Disagree or Disagree, please provide a brief explanation.

Q4. The worksheet helped to assess if there enough data sources and data for the project topic.

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

Q4a. If you answered Strongly Disagree or Disagree, please provide a brief explanation.

Fig. 2. Self-assessment questions.

IV. RESULTS

A. Overall Responses

Students were asked to indicate their perception of the usability of the worksheet in helping to identify data sources using a five-point Likert scale to answer four questions. The Likert scale consisted of: 1 - strongly agree, 2 - agree, 3 - neutral, 4 - disagree, and 5 - strongly disagree. Table II shows the overall Likert responses and basic statistics for 32 student responses. The calculated average response for all questions, as seen in Table II, is two. A mean value of two corresponds to “2 - agree” on the Likert-scale.

TABLE II. OVERALL LIKERT RESPONSES

	Assessment questions and basic statistics		
	Assessment Questions (n = 32)	Mean	SD
Q1	The worksheet helped me to identify primary data sources.	2.343	0.7749
Q2	The worksheet helped me to identify secondary data sources.	2.187	0.6343
Q3	The worksheet helped me to identify tertiary data sources.	2.343	0.7335
Q4	The worksheet helped me to assess if there was enough data sources and data for the project topic.	2.218	0.7799

B. Response Rate Per Question

Fig. 3 shows the response rate for each of the four questions (see Table II). As seen in Fig. 3, the “(A)gree” Likert-response has the highest percentage response for each question, indicating

students “agreed” with each statement (see Table II). Fifty-three percent “agree” the worksheet helped to identify primary data sources, 66% “agree” the worksheet helped to identify secondary data sources, 59% “agree” the worksheet helped to identify tertiary data sources, and 69% “agree” the worksheet helped to assess if there was enough data sources and data for the project.

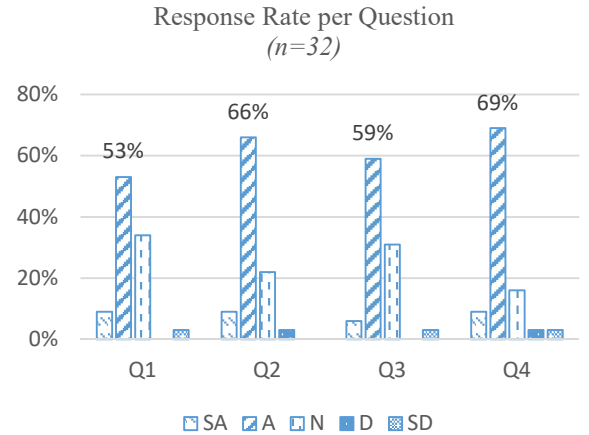


Fig. 3. Response rate for each Likert-response question.

C. Time to Complete the Activity Worksheet

Although not the primary focus of this study, we were interested in how much time students spent completing the worksheet. As a preliminary step towards the next iteration of the study, students were asked to self-report how much time was spent completing the worksheet by choosing the time range that most closely reflected how many hours were spent. As seen in Fig. 4, 56% of students reported spending 30 minutes to 1 hour

to complete the worksheet. Twenty-five percent spent up to 30 minutes, 13% reported spending 1.5 to 2 hours, and 6% reported spending more than 2.5 hours.

We acknowledge we do not have sufficient evidence to support a statement that more time is needed for identifying appropriate data sources; however, we are motivated to refine future research designs to capture appropriate data to address the research question, “Do students who spend more time completing the worksheet find more appropriate data sources compared to students who spend less or no time completing the worksheet?”

V. DISCUSSION

In this research, we examined students’ perception of the “Problem to Data Sources” Activity Worksheet. Our examination also addressed specific aims outlined for the project: 1) introduction of different types of data sources (primary, secondary and tertiary), 2) introduction of the concepts of relevance and reliability of data sources, and 3) assessment of students’ perception of the usability of the “Problem to Data Sources” Activity Worksheet in helping to identify relevant and reliable data sources. The overall response to the worksheet was positive. Results of this work address the first research question:

R1: How do students perceive the usability of the worksheet for identifying data sources.

Overall responses (Table II) and response rates per question (Fig. 3) show the most frequent response was “agree.” Fifty-three percent of students found the worksheet to be helpful in identifying primary data sources. Sixty-six percent agree the worksheet helped to identify secondary data sources, 59% agree the worksheet helped in identifying tertiary data sources, and 69% agree the worksheet helped to assess if there was enough data sources and data for the proposed project.

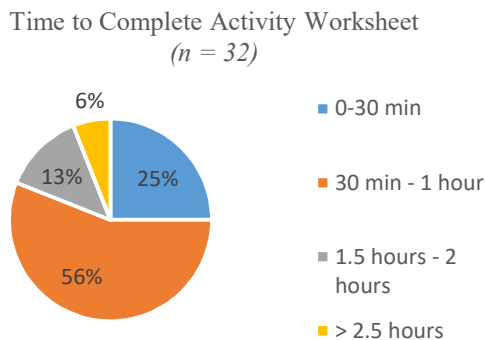


Fig. 4. Students’ reported time to complete the activity worksheet.

The prior work of [11] was reviewed to address the second research question:

R2: Do students’ perception of the worksheet differ between undergraduates and graduate students?

In our prior work [11], we examined the perception of graduate students who completed the worksheet as part of a graduate research seminar. Graduate students self-identified as

first-, second-, and third-semester masters and PhD level students. Each academic level was characterized as a cohort to facilitate ease of analysis of responses. Similar to the approach implemented in this work, the “Problem to Data Sources” Activity Worksheet was completed as part of a four-part series of worksheets [10].

The relation between graduate student cohorts and students’ perception of the “From a Problem to Data Sources” worksheet was examined using a chi-square test and was found to be significant, $X^2(8, N=100) = 50.60, p = 3.13 \cdot 10^{-8}$. Since the critical value ($p = 3.13 \cdot 10^{-8}$) is less than the confidence level (.05) the null hypothesis is accepted, there is a relationship between cohorts and students’ perception of the usability of the worksheet. The data indicate 41% of the responses were “agree” and of those responses, the first-semester cohort found the “From a Problem to Data Sources” worksheet to be most helpful. [10].

Prior results show there is a relationship between graduate cohorts and students’ perception of the usability of the activity worksheet. The results further suggest first-semester graduate cohort found the “Problem to Data Sources” Activity Worksheet to be most helpful. Undergraduates, with little to no research experience, also found the worksheet to be helpful. These preliminary results the activity worksheet is most helpful for students with little to minimal experience locating and identifying data sources, irrespective of graduate or undergraduate status.

The results of this work show 56% of undergraduates who completed the worksheet reported spending 30 minutes to an hour completing the worksheet. It is unclear how students gauged time spent (i.e., does the reported time include actual time spent looking for data sources or time to answer questions on the worksheet. Twenty-five percent reported spending two or more hours to complete the worksheet.

The practice of research takes careful planning and time. The “Problem to Data Sources” Activity Worksheet can be used as an outline for class discussion, for introducing research concepts like the data life cycle for example. Identification of relevant and reliable data sources will help students make sound engineering arguments and build a practice of lifelong learning. The worksheet can also be used by research faculty to assess students’ knowledge of project data as well as identify students’ data sources. The worksheet can also be used in graduate and undergraduate courses as independent study, in course projects where students are required to identify data and developing thesis and dissertation topics.

A. Implications for Scaffolded Learning

Encouraging students to complete self-assessments after completing the worksheet will inform student’s perception of the worksheet and identify parts of the process students find challenging. If a student responds “disagree” or “strongly disagree” to Likert-scaled questions their response could serve to facilitate scaffolding of the process, and the need for additional information or content on concepts students disagree with in Likert questions.

Instructional scaffolding is an essential tool to support students during problem centered instruction ([16], [17], [18], [19]). Instructional scaffolding can take on many forms: one-to-one, peer scaffolding or computer-based scaffolding. The approach adopted in this work is one-to-one utilizing the activity worksheet as the mechanism for identifying opportunities to support meaningful engagement between students and faculty. In this work we adopt [20] outline of the term scaffolding:

The first key feature that distinguishes scaffolding from other forms of instructional support is that it is temporary support that is provided as students are engaging with problems ([21], [22], [23]). Next, scaffolding needs to lead to skill gain such that students can function independently in the future ([21], [23], [24]). Third, scaffolding not only simplifies tasks, but also highlights complexity therein ([18], [23]). This is because struggling while attending to certain complexities inherent in a particular task can lead to robust learning ([18], [25]). Fourth, to qualify as scaffolding, students need to meaningfully participate in the target task and have an understanding of what success at the task means ([23], [26]). If the tool does all or most of the work or if students do not know how to recognize successful performance of the target skill, then the possibility of skill gain is compromised ([24], [27]).

The search for scientific data sources is an ongoing challenge as the number of scientific sources continue to grow at an exponential pace. The search for scientific data sources is deserving of an extensive review and as such is outside of the scope of this paper; however, we acknowledge and recommend readers with interest to consider literature on sharing of research data. In previous work [28] sharing research data as well as missed opportunities in the long-tail of data is discussed. The long-tail of data is defined by [29] as small granular data sets, collected in the course of day-to-day research (i.e., alternative endpoints, results from pilot studies, metadata, etc.) that may be considered supportive but not the true focus of the study and are not published. There are tremendous possibilities for discoveries and supplemental work that can benefit from sharing of these types of data. Although the type of data may vary and data dictates the type of tools used to explore and visualize it, the activity worksheet described in this work is data type agnostic making the worksheet flexible for use regardless of data type. In the worksheet presented in this work, the focus is on having clarity on what data is needed, to better inform the data search.

B. Research Agenda

The worksheet examined in this work is presented as a stand-alone worksheet; however, this activity can be used in parallel with other data-driven processes and methodologies. The worksheet can also serve as a preliminary step towards preprocessing of data for later use.

For example, this worksheet is the final worksheet in a series of worksheets designed help students develop sound research questions and identify data needed to support a research plan. This work will help students understand different types of data sources, the role each type of data source plays in developing a research plan, and how to determine the type of data needed to

support their research. The next step in this process is data acquisition and data visualization. *Acquiring* data is a crucial step in many problem-solving processes like engineering design and data visualization [11], yet students get very little training in identifying viable data sources. This in turn impacts research plans and visualizations created to support research outcomes. It is not enough to generate a visual representation of data. An understanding the data lifecycle and knowledge of data origins can provide a more complete view of the problem and what is needed to solve it. More research is needed in the area of knowledge building, data and information literacy in the context of engineering design, data visualization, research processes and students' ability to successfully implement these processes.

VI. CONCLUSION

The data from this research shows that students perceived the worksheet to be helpful. The aim of the worksheet is to help students think critically about what data is needed to support their research question and to identify appropriate data sources. This research did not examine or assess the quality of data sources identified. The concept of identifying "better" data sources was not explored. This worksheet can be used as an instructional guide for faculty teaching methodologies with a focus on data-driven research as well as a precursor to existing tools and guidelines for identifying data sources. At a time when heterogeneous data from different sources are combined and utilized to inform research methods and outcomes, data provenance is becoming more important. The term data provenance refers to a record trail that accounts for the origin of a piece of data (in a database, document or repository) together with an explanation of why and how it got to the present place [30].

The availability and access to data continues to evolve as more data is shared across organizations, nations, and digital platforms [31]. The availability of data impacts the demand for data-driven solutions, tools and resources that allow for deeper insight into approaches to analyze and assess data. The worksheet presented in this work, when completed, could contribute to meta-data for data sharing, as well as document what data currently exist and what data remains to be acquired. In this manner, the worksheet could contribute to the identification of data gaps or areas where additional research focus is needed. The ubiquitous nature of data and its widespread applicability can benefit from the "*Problem to Data Sources*" Activity Worksheet.

ACKNOWLEDGMENT

The author would like to thank all students who completed the worksheet and provided invaluable feedback and Isha Ashish Mahadalkar for assisting with formatting.

REFERENCES

- [1] R. E. Wertz, Ş. Purzer, M. J. Fosmire, and M. E. Cardella, "Assessing information literacy skills demonstrated in an engineering design task," *Journal of Engineering Education*, vol. 102, no. 4, pp. 577-602, 2013.
- [2] Association of College and Research Libraries (ACRL), 2000. Framework for Information Literacy for Higher Education. Retrieved from http://www.ala.org/acrl/sites/ala.org/acrl/files/content/issues/infolit/Framework_ILHE.pdf

- [3] L. S. Connaway, T. J. Dickey, and M. L. Radford, "If it is too inconvenient I'm not going after it: Convenience as a critical factor in information-seeking behaviors," *Library & Information Science Research*, vol. 33, no. 3, pp. 179-190, 2011.
- [4] C. C. Kuhlthau, *Seeking meaning: a process approach to library and information services*. 2nd. ed. Westport, CT: [Libraries Unlimited](#), 2004.
- [5] Society of College, National, and University Libraries (SCONUL). (2011). The SCONUL seven pillars of information literacy: Core model for higher education. Retrieved from https://eprints.ncl.ac.uk/file_store/production/192827/156279C4-30EE-4356-A2ED-A48D9D474286.pdf
- [6] E. K. Owusu-Ansah, "Debating definitions of information literacy: enough is enough!," *Library Review*, vol. 54, no. 6m pp. 366-374, 2005.
- [7] A. Zheng, and V. L. Byrd, "Students' Perception of a Method for Identifying Topics for Research Questions." In *2020 IEEE Frontiers in Education Conference (FIE)*, October 20, 2020.
- [8] N. R. Kanter, and V. L. Byrd. "A Method for Transforming a Broad Topic to a Focused Topic for Developing Research Questions." In *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1-7, 2020.
- [9] D. Pecher, Z. Chu, and V. L. Byrd (2020, October). A method for transforming a question into a problem statement. In *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1-8. IEEE, pp. 1-8, 2020.
- [10] V. L. Byrd, and J. D. Camba. "A Worksheet Method for Developing Research Questions: An Examination of Three Graduate Student Cohorts." In *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1-7, 2020.
- [11] V. Byrd, and J. D. Camba. "Design Activity Worksheets for Developing Research Questions." In *ASCE Virtual Annual Conference*, 2020.
- [12] C. Watson, "An exploratory study of secondary students' judgments of the relevance and reliability of information," *Journal of the Association for Information Science and Technology*, vol. 65, no. 7, pp. 1385-1408, 2014.
- [13] *Choosing & Using Sources: A Guide to Academic Research*, The Ohio State University Pressbooks <https://ohiostate.pressbooks.pub/choosingsources/chapter/data-as-sources/>
- [14] Iowa State University Library (2019). LIB 160: Information Literacy, Three Tools for Finding Information, Iowa State University, Digital Press. <https://iastate.pressbooks.pub/lib160/chapter/three-tools-for-finding-information/>
- [15] University of Cincinnati Library (2021). Undergraduate Research Programs: Library Research Guide. <https://guides.libraries.uc.edu/undergradresearch/startresearch>
- [16] B. R. Belland, K. D. Glazewski, and J. C. Richardson, "A scaffolding framework to support the construction of evidence-based arguments among middle school students," *Educational Technology Research and Development*, vol 56, no. 4, pp. 401-422, 2008. <http://doi.org/10.1007/s11423-007-9074-1>.
- [17] J. Lu, S. P. Lajoie, and J. Wiseman, "Scaffolding problem-based learning with CSCL tools," *International Journal of Computer-Supported Collaborative Learning*, vol. 5, no. 3, pp. 283-298, 2010. <http://doi.org/10.1007/s11412-010-9092-6>.
- [18] B. J. Reiser, "Scaffolding complex learning: The mechanisms of structuring and problematizing student work," *Journal of the Learning Sciences*, vol. 13, no. 3, pp. 273-304, 2004. http://doi.org/10.1207/s15327809jls1303_2.
- [19] H. G. Schmidt, J. I. Rotgans, and E. H. Yew, "The process of problem-based learning: What works and why. *Medical Education*," vol. 45, no. 8, pp. 792-806, 2011. <http://doi.org/10.1111/j.1365-2923.2011.04035.x>.
- [20] B. R. Belland, *Instructional scaffolding in STEM education: Strategies and efficacy evidence* (p. 144). Springer Nature, 2017.
- [21] B. R. Belland, "Scaffolding: Definition, current debates, and future directions," in J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th edn., pp. 505-518). New York: Springer, 2014.
- [22] A. Collins, J. S. Brown, and S. E. Newman, "Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics," in L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1989.
- [23] D. Wood, J. S. Bruner, and G. Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89-100, 1976. <http://doi.org/10.1111/j.1469-7610.1976.tb00381.x>.
- [24] R. D. Pea, "The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity," *Journal of the Learning Sciences*, vol. 13, no. 3, pp. 423-451, 2004. http://doi.org/10.1207/s15327809jls1303_6.
- [25] K. D. Simons, and P. A. Ertmer, "Scaffolding disciplined inquiry in problem-based learning environments," *International Journal of Learning*, vol. 12, no. 6, pp. 297-305, 2006.
- [26] J. W. Mahardale and C. B. Lee, "Understanding how social and epistemic scripts perpetuate intersubjectivity through patterns of interactions," *Interactive Learning Environments*, vol. 21, no. 1, pp. 68-88, 2013. <http://doi.org/10.1080/10494820.2010.547204>.
- [27] M. T. H. Chi, "Constructing self-explanations and scaffolded explanations in tutoring," *Applied Cognitive Psychology*, vol. 10, no. 7, pp. 33-49, 1996. [http://doi.org/10.1002/\(SICI\)1099-0720\(199611\)10:73.0.CO;2-E](http://doi.org/10.1002/(SICI)1099-0720(199611)10:73.0.CO;2-E).
- [28] V. Byrd, "Sharing Research Data, Data Standards and Improving Opportunities for Creating Visualizations." Chapter 9 in. *Partners for Preservation: Advancing digital preservation through cross-community collaboration*. J. Kramer-Smyth, Eds. Facet Publishing. 2019.
- [29] A. R. Ferguson, J. L. Nielson, M. H., Cragin, A. E. Bandrowski, and M. E. Martone, "Big data from small data: data-sharing in the 'long tail' of neuroscience," *Nature Neuroscience*, vol. 17, no. 11, pp. 1442-1447, 2014.
- [30] A. Gupta, "Data Provenance," In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA., 2009. https://doi.org/10.1007/978-0-387-39940-9_1305.
- [31] A K. Kar and Y. K. Dwivedi. "Theory building with big data-driven research-Moving away from the "What" towards the "Why". *International Journal of Information Management*, vol. 54, no. 102205, 2020.