Abstract
As research on the digital divide shifts away from questions about access and focuses instead on Digital Information Literacy (DIL) skills and the outcomes of productive use of such skills in digital contexts, we are faced with significant measurement challenges. To meet these challenges, complex, interactive simulation-based assessments have been developed that capture authentic learner performances. In the current study, we describe a multi-step modeling method for identifying distinct strategies captured within process data generated by students using a simulated web search tool within an inquiry task. The method considers the content, timing, and context of student actions. This approach identified meaningfully distinct strategies in students' search processes, which were associated with differences in inquiry task performance.

Clustering Student Strategies in a Simulated Web Search Environment

Objectives

As research on the digital divide shifts away from questions about access and focuses instead on Digital Information Literacy (DIL) skills and the outcomes of productive application of those skills in the context of realistic digital tools, we are faced with significant measurement challenges (Scheerder et al., 2017). To meet these challenges, complex, interactive simulation-based assessments have been developed to capture authentic learner performances and provide opportunities to explore complexities of those performances (Authors, 2018a). We developed a multi-step modeling method for identifying distinct strategies captured within process data generated from students' interactions with a simulated web search tool, to specifically consider the content and timing of student actions as well as the context in which the actions occurred. This approach affords a qualitative understanding of students' search strategies reflective of the complexity of interactions in DIL tasks.

Theoretical Framework

The construct of Digital Information Literacy (DIL) captures the complex interplay of knowledge, skills, and abilities (KSAs) needed to support individuals as they obtain, understand, evaluate, and use information within digital environments (Authors, 2016). Performance-based assessments of DIL skills must balance task realism and authenticity with valid, reliable measurement of critical aspects of the target construct(s). Such performances are often measured in terms of search term quality (Argelagos & Pifarre, 2017), location/access of target information (Coiro & Kennedy, 2011; Hahnel et al., 2017), and appropriate use of information in multiple-source synthesis tasks (Authors, 2018a; Goldman et al., 2011, 2012, 2013, 2018; Hastings et al., 2012; Shavelson et al., 2019). Some researchers argue that these product-focused approaches oversimplify the complexities of the DIL construct by focusing on discrete skills and ignoring the context of students' decisions within a larger inquiry process (Authors, 2018a; Hinostroza et al., 2018; Van Deursen & Van Dijk, 2009).

Hinostroza et al. (2018) documented various web search strategies and heuristics used by participants in a think-aloud study. They demonstrated high-level mappings between participants' strategic processes and DIL skills (e.g., searching, evaluating websites), but also observed wide strategic variability both across and within participants, concluding that naturalistic search behavior is more complex than is typically represented in DIL assessments (see Authors, 2016). These observations are consistent with cognitive science theories and models suggesting that search strategies reflect individuals' KSAs, as well as their goals, the task goals, and task design features (e.g., Britt & Gabrys, 2002; Britt et al., 2018; Coiro & Dobler, 2007; Juvina & Van Oostendorp, 2008; Rouet, 2006; Wirth et al., 2007). Assessments that narrowly focus on low-level skills may have strong technical properties but lack authenticity and construct validity; in contrast, increasingly complex, dynamic assessment tasks afford better realism, but resulting performances can be difficult to interpret and challenging to administer and score at scale (Authors, 2016; Leu et al., 2009). This warrants the development of complex approaches to modeling and interpreting the data resulting from these authentic DIL assessment environments in addition to scored response products, for valid interpretations of students' performances.

In the current study, we modeled students' search strategies as they interacted with a simulated web search tool designed to assess their ability to locate and evaluate websites relevant to the overall goals of an online inquiry task (Authors, 2018a, 2018b). We describe the use of a multi-step clustering approach that integrates the content of students' search patterns with the context of their actions to identify meaningful distinctions in search strategies. We used this method to cluster students' search sequences. We discuss initial evidence that these clusters provide descriptive information about meaningful distinctions in students' strategy use, associated with different patterns of performance.

Data Sources

We investigated data from students' interactions with a simulated web search tool from a larger online inquiry task situated within a virtual world (Authors, 2018b). The inquiry task featured an overarching narrative, a goal-driven scenario, and assessment activities that challenged students to locate task-relevant information by interacting with virtual characters and reading "print" and digital text resources. Several help functions and timer-based alert messages were embedded into the system to help students manage time-on-task and to locate required resources. Students' actions and responses were used to assess DIL subconstructs of planning, locating, evaluating, and synthesizing. Here, we focused on modeling process data from students' use of the web search tool (Figure 1), which could be visited multiple times throughout the task; we consider each visit as a unique search session.

Methods

Participants

Eighth-grade students participated (N=130, 67 females) completing the task and a post-survey in one 90-minute session. Students had on average 2.9 search sessions (SD 2.3, min=1 session, max=11), generating 319 unique sessions.

Multi-step Clustering Approach

Our aim in clustering students' search processes is to identify distinct patterns that help us characterize students' ability to plan their search, locate websites, and evaluate those websites. We used a novel, unsupervised multi-step clustering method to account for both the **content** and **context** of students' actions.

Step 1: Action Representation

Model descriptiveness is directly affected by how we represent students' search processes. We translated task log files into a series of discrete timestamped actions reflecting students' search sequences. Table 1 reports the full set of actions grouped into 5 categories. These categories include three DIL constructs: *planning, locating information*, and *evaluating sources*. We further distinguished these categories in terms of the quality of search results, websites, and student evaluations of those websites. In addition to student-initiated actions, we considered *interface actions* that result in meaningful changes to the task state. Finally, we explicitly coded significant pauses within student's action sequences. Following Authors (2020), we also assigned *pauses between actions* to ordinal categories to provide our model with the ability to distinguish actions that are completed in quick succession from those which require more interpretation and planning prior to execution (Figure 2).

Step 2: Edit Distance Clustering

After defining our action representation, we used a normalized *optimal matching* (OM) metric to calculate the distance between all sequences (TraMineR package in R; Gabadinho et al, 2011). This approach determines the dissimilarity between two sequences by calculating the number of substitutions necessary to make the sequences match while controlling for the impact of differences in sequence length. We then applied hierarchical agglomerative clustering on the resulting pairwise OM distances to identify clusters of similar search sequences. While this is a well-established method for clustering educational process data (Hao et al., 2015; Boroujeni & Dillenbourg, 2019), it does not consider the order in which actions occur, which can complicate interpretation.

Step 3: Hidden Markov Models

Unlike edit distance clustering which is very sensitive to the content of action sequences, hidden Markov model (HMMs) capture the context of actions by modeling the probabilistic transition between action states. These action states reflect a latent state that incorporates both the content and context of the action (e.g., Jeong et al., 2010; Fincham, 2020). We grouped sequences in terms of four edit-distance clusters and fit four separate HMMs, one for each cluster. We considered models with 4 to 10 states to describe the four clusters and used Bayesian Information Criterion (BIC) to determine best model fit (Figure 3).

Results

We identified four clusters of search strategies that differed in terms of the quality of search results and websites viewed, the occurrence of significant pauses within students' search processes, and the overall structure of their search processes. Figures 4 and 5 provide a visualization of the individual sequences and a node and arrow representation of the four clusters with a brief qualitative description of the different states that students experienced.

Cluster 1: Thoughtful Search

Cluster 1 (n=102 sequences, 26.8% of first sessions) appears to reflect a relatively proficient strategy use. This model identifies three separate pause states that occur in different contexts, likely representing different processing. Initially, students transitioned between a pause state and constructing a search. Searches had a high probability of yielding high-quality results, which contain the two most useful websites within the task. We observed thoughtful behavior as students iterated between short pauses and review of results and websites. This pattern ends with students deciding to save a website, followed by being prompted to evaluate its importance, usefulness, and trustworthiness. The pauses increase in length as students complete their evaluations, and from that state they either finish the session or decide to run another search.

Cluster 2: Scaffolded Search

Cluster 2 (n=93 sequences, 56.9% of first sessions) is characterized primarily by the presence of an off-topic search yielding off-topic results. When students enter an irrelevant search term, they are provided a list of unrelated, inactive results along with a hint suggesting they try entering a search term related to the inquiry topic; after five off-topic searches, students are explicitly prompted to type in a search term that will yield high-quality results. These system hints likely account for the high probability of entering a high-quality results query within this cluster. Unlike Cluster 1, this strategy features a more direct path from the results page to the selection of a website and completion of the resource evaluation prompts. This suggests that while these two strategies involve very similar content, the behaviors likely reflect different skills and decision-making processes. The prevalence of this pattern within students' first sessions suggests that it may capture acting on system hints following an initial false start.

Clusters 3 and 4: Low-Quality Searches

Clusters 3 (n=79 sequences, 11.4% of first sessions) and 4 (n=45 sequences, 4.9% of first sessions) have a similar structure containing a single pause state that connects all action states. For these clusters the model cannot distinguish any consistent patterns suggesting that the pauses reflect different processing. The primary difference between these clusters is that sessions in Cluster 3 have a high probability of running searches yielding medium-quality results (i.e., only partially-relevant to the inquiry task) and sessions in Cluster 4 have a high probability of obtaining low-quality results (e.g., results that are only tangentially relevant to the inquiry task). Both clusters have a low probability of success with only half of search sequences resulting in saving and evaluating a website (Cluster 3: 42%, Cluster 4: 50%).

Relationship Between Strategies and Performance

Finally, we conducted exploratory analyses to examine whether the strategy clusters were related to inquiry task performance. While no statistically significant differences between clusters were observed, non-parametric Spearman correlations examining the proportion of sessions within each cluster indicated distinct patterns of relationships to performance (Table 2). Specifically, a higher proportion of sessions in Cluster 1 was associated with higher performance, while a higher proportion of sessions in Cluster 2 was associated with lower performance, especially on search and evaluation activities. The proportion of sessions in Cluster 3 was positively associated with planning and questioning subscores, while the proportion of Cluster 4 sessions showed weak-to-no relationships to performance.

Educational Significance

The aim of this work is to develop a data-driven method for identifying meaningful distinctions in inquiry strategies that incorporates the quality of the materials interacted with, the time spent on different actions, and the context of those actions. Applying this approach to characterize students' interactions within a simulated web search tool, we identified four distinct strategies. The two most frequently occurring strategies resulted in similar proportions of students running high-quality searches and selecting relevant websites, however, the process models suggest students arrived at these outcomes in different ways (e.g., Cluster 2 sessions viewed high-quality results after receiving explicit hints and were associated with weaker evaluation skills). Application of different strategies within the same task this may reflect differences in students' understanding of the task, their goals, or their abilities (Van Deursen & Van Dijk, 2009). This method serves as a useful approach for qualitatively understanding student strategy use in complex, dynamic simulation-based environments. Future work will examine how these strategies can be used to complement and contextualize quantitative scores generated from complex task environments, yielding a more nuanced picture of students' DIL proficiency as estimated from complex, integrated performances within digital inquiry tasks.

References

Authors (2016)

Authors (2018a)

Authors (2018b)

Authors (2020)

Argelagós, E., & Pifarré, M. (2017). Unravelling Secondary Students' Challenges in Digital Literacy: A Gender Perspective. *Journal of Education and Training Studies*, *5*(1), 42-55.

Boroujeni, M. S., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics*, *6*(1), 16-33.

Britt, M. A., & Gabrys, G. (2002). Implications of document-level literacy skills for web site design. *Behavior Research Methods, Instruments, & Computers, 34*(2), 170-176.

Britt, M. A., Rouet, J.-F., & Durik, A. M. (2018). *Literacy beyond text comprehension: A theory of purposeful reading.* New York: Routledge.

Coiro, J., & Dobler, E. (2007). Exploring the on-line reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly, 42*(2), 214-257.

Coiro, J., & Kennedy, C. (2011). *The online reading comprehension assessment (ORCA) project: Preparing students for common core standards and 21st century literacies* [White paper]. Storrs, CT: University of Connecticut.

Fincham, O. E. D. (2020). *Analytics of student interactions: towards theory-driven, actionable insights.* [Unpublished doctoral dissertation.]

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of statistical software*, *40*(4), 1-37.

Goldman, S. R., Blair, A., & Burkett, C. M. (2018). Assessment of multiple resource comprehension and information problem solving. In J. L.G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 466-484). New York, NY: Routledge.

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. M. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly, 47(*4), 356–381.

Goldman, S. R., Lawless, K., & Manning, F. (2013). Research and development of multiple source comprehension assessment. In M. A. Britt, S. R. Goldman, & J.-F. Rouet. (Eds.), *Reading from words to multiple texts* (pp. 180-199). New York, NY: Routledge.

Goldman, S. R., Lawless, K., Pellegrino, J., Manning, F., Braasch, J., & Gomez, K. (2011). A technology for assessing multiple source comprehension: An essential skill of the 21st century. In M. C. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 173-210). Greenwich, CT: Information Age Publishing.

Hahnel, C., Goldhammer, F., Kröhne, U., & Naumann, J. (2017). Reading digital text involves working memory updating based on task characteristics and reader behavior. *Learning and individual differences*, *59*, 149-157.

Hao, J., Shu, Z., & von Davier, A. (2015). Analyzing process data from game/scenario-based tasks: an edit distance approach. *Journal of Educational Data Mining*, *7*(1), 33-50.

Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods, 44*(3), 622-633.

Hinostroza, J. E., Ibieta, A., Labbé, C., & Soto, M. T. (2018). Browsing the Internet to solve information problems: A study of students' search actions and behaviours using a 'think aloud'protocol. *Education and Information Technologies*, *23*(5), 1933-1953.

Jeong, H., Biswas, G., Johnson, J., & Howard, L. (2010, June). Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models. In *Educational Data Mining 2010*.

Juvina, I., & van Oostendorp, H. (2008). Modeling semantic and structural knowledge in web navigation. *Discourse Processes*, *45*(4-5), 346-364.

Leu, D. J., Jr., Kulikowich, J., Sedransk, N., & Coiro, J. (2009). *Assessing online reading comprehension: The ORCA project*. Research grant funded by U.S. Department of Education, Institute of Education Sciences.

Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics*, *34*(8), 1607-1624.

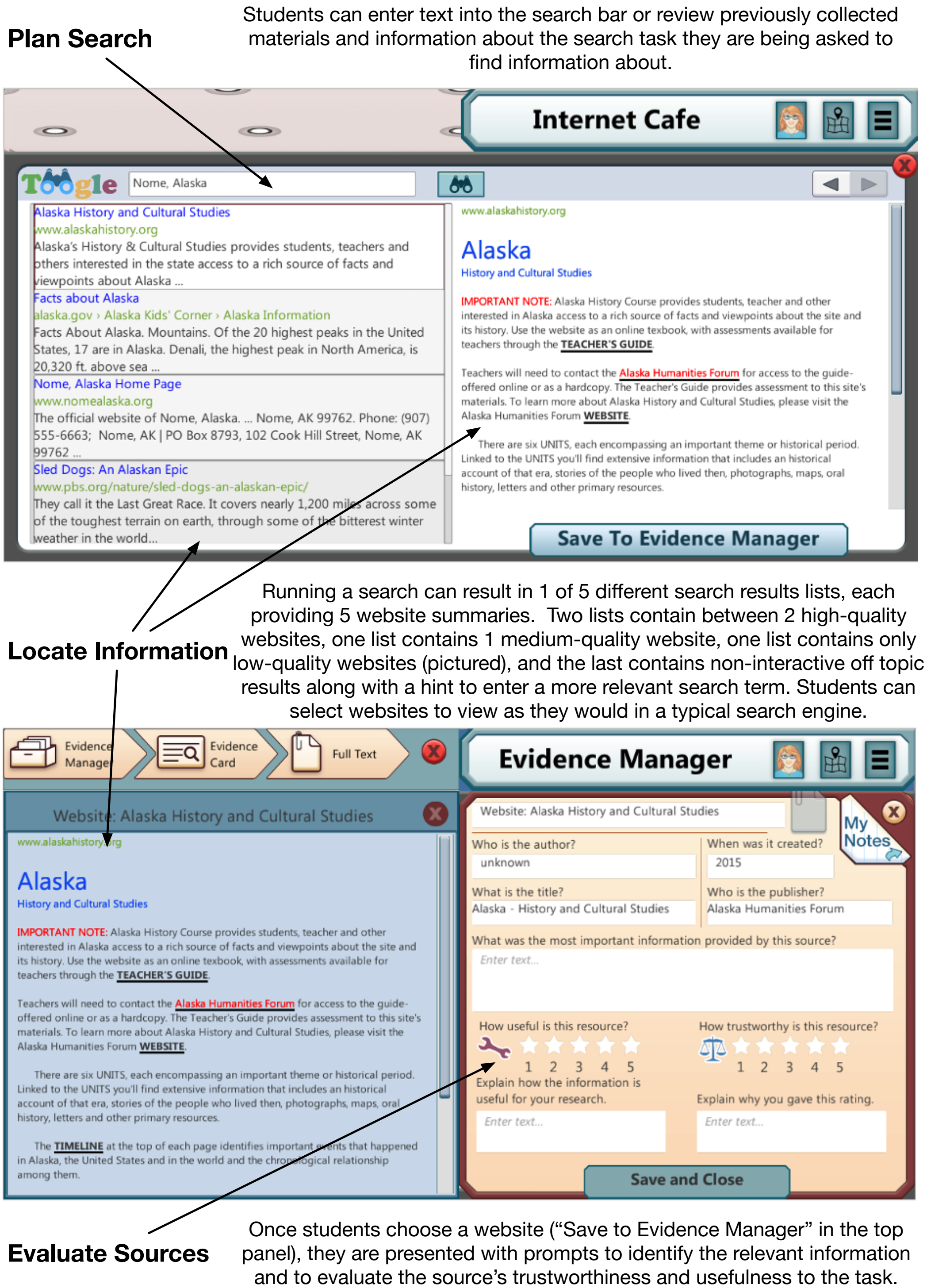
Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., & Marino, J. P. (2019). Assessment of university students' critical thinking: Next generation performance assessment. *International Journal of Testing*, *19*(4), 337-362.

Van Deursen, A. J., & Van Dijk, J. A. (2009). Using the Internet: Skill related problems in users' online behavior. *Interacting with computers*, *21*(5-6), 393-402.

Wirth, W., Böcking, T., Karnowski, V., & Von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication*, *12*(3), 778-800.

Figure 1

Digital Information Literacy Constructs Present within Simulated Web Search Task



Note. Screenshots from the web search tool illustrate three targeted DIL constructs. These search results and the website displayed reflect low-quality information, i.e., this shares some key terms with the search task but the information is only tangentially-relevant to task goals.

Table 1

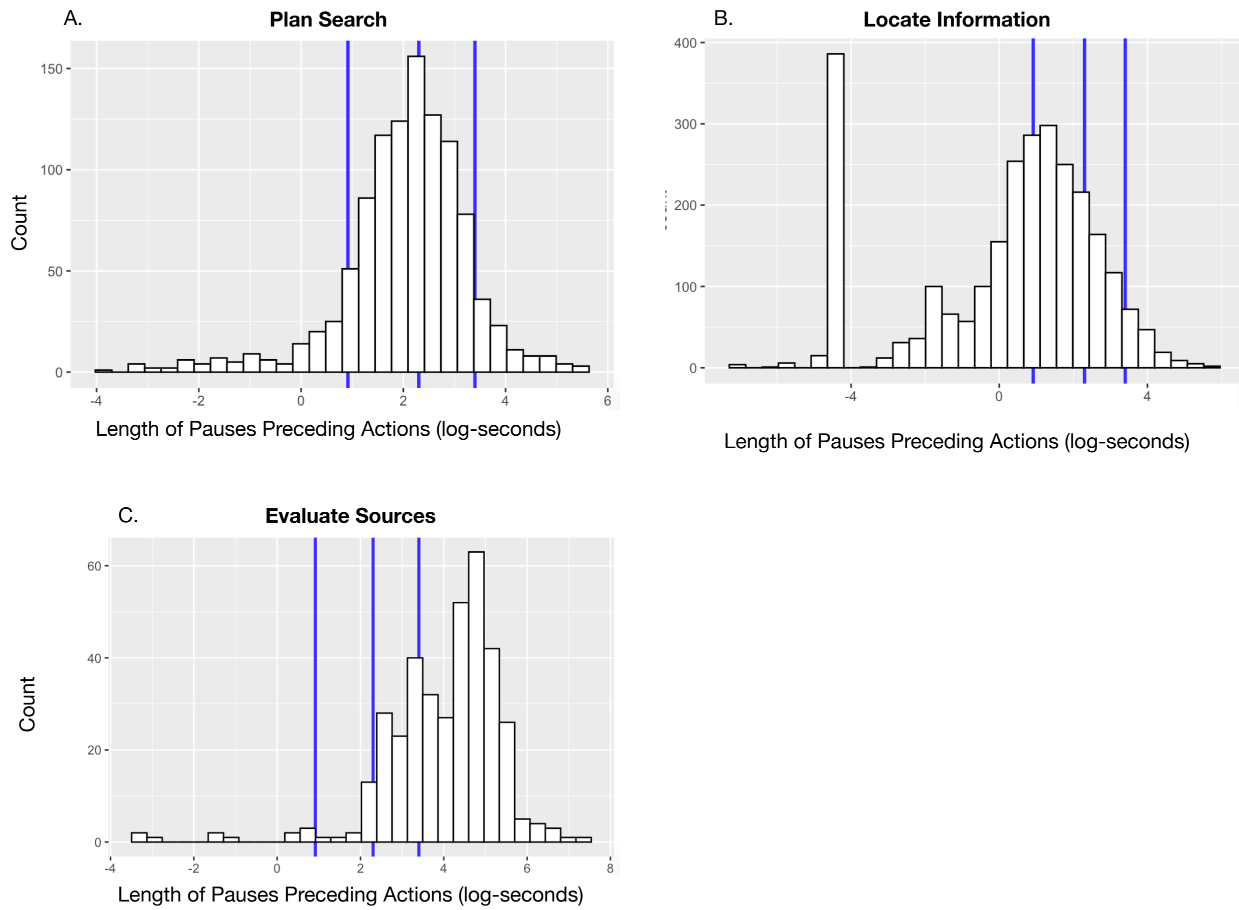
Simulated Web Search Tool Action Labels by Digital Information Literacy Construct Categories

Alignment to DIL Construct	Description	Action Label	Mean Frequency (SD)
Plan Search	Student edits their search terms or consults reference materials to generate a query for the search task	Construct Search	8.4% (3.6%)
		View Reference Materials	9.4% (8.9%)
Locate Information	When students run a search, the list of results updates	Run Search	10% (4.4%)
		View Previous Search	5% (3.1%)
	Results were coded in terms of their task relevance. Off-topic results automatically triggered a search term hint.	Results (High-quality; i.e., two useful websites)	13.1% (6.7%)
		Results (Medium-quality, i.e., one useful website)	9.6% (7.1%)
		Results (Low-quality, i.e., no useful websites)	11.1% (7.1%)
		Off-topic Results	6.8% (5.1%)
	Websites were coded in terms of their task relevance	View Website (High-quality, i.e., highly-relevant)	4.6% (2.7%)
		View Website (Medium-quality, i.e., partially-relevant)	4.4% (2.9%)
		View Website (Low-quality, i.e., tangentially-relevant)	7.4% (4%)
Evaluate Sources	Websites were evaluated in terms of their Importance, Trustworthiness, and Usefulness	Evaluate (High-quality)	5.6% (3.7%)
		Evaluate (Mixed-quality)	4.9% (3.4%)
		Evaluate (Low-quality)	4.4% (2.9%)
		Evaluate (No Rating)	3.6% (2.2%)
Interface actions	Start and End Session; these states aid in interpretation of strategies	Start Session	10% (11.1%)
		Finish Session	9.6% (11%)
	Specific help actions triggered by the system if students appear to be off-task	Timer Alert (i.e., system warns students they have 5 minutes left to search)	4.6% (4.4%)
		Key Source Resetting (i.e., system provides students with relevant resource)	4.6% (3.8%)
		Search Term Hint (i.e., system provides students with suggested search terms)	6.4% (4.5%)
	Students can ask for help and receive instructions about using the interface	Help (i.e., students can press a button to access contextual help menu)	5.6% (5.6%)
Pauses between actions	Between 2.5 and 10 seconds	Short Pause	19.9% (8.2%)
	Between 10 and 30 seconds	Medium Pause	12.8% (6.7%)
	Greater than 30 seconds	Long Pause	9.3% (6.6%)

Note. Process data logs were translated into 24 action labels alongside descriptions and alignment to the DIL construct. We also report the mean frequency with which these actions occur within students' search sequences (standard deviation in parentheses). DIL=Digital information literacy.

Figure 2

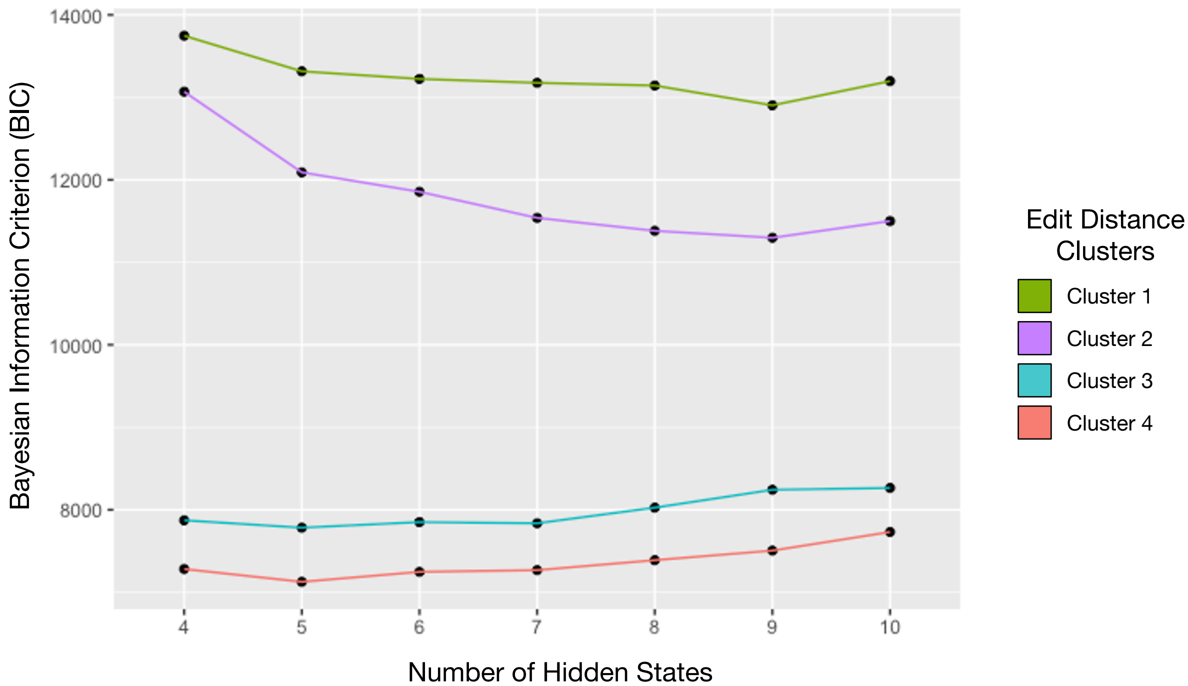
Histogram of the Duration of Pauses Preceding Actions by DIL Construct



Note. (a) Distribution of pauses preceding plan search actions (M 14.4s, SD 22.7s); (b) Distribution of pauses preceding locate information actions (M 7.4s, SD 17.7s); (c) Distribution of pauses preceding evaluation of source actions (M 103.6s, SD 134.0s). Duration is reported in log-seconds. Blue vertical bars indicate inclusion cutoff criteria. Pauses less than 2.5 seconds (the leftmost vertical line) are not coded in our action sequences. These pauses often capture time to navigate the interface. Pauses between 2.5 seconds and 10 seconds are coded as *short*, pauses between 10 seconds and 30 seconds are coded as *medium*, and pauses above 30 seconds are coded as *long*.

Figure 3

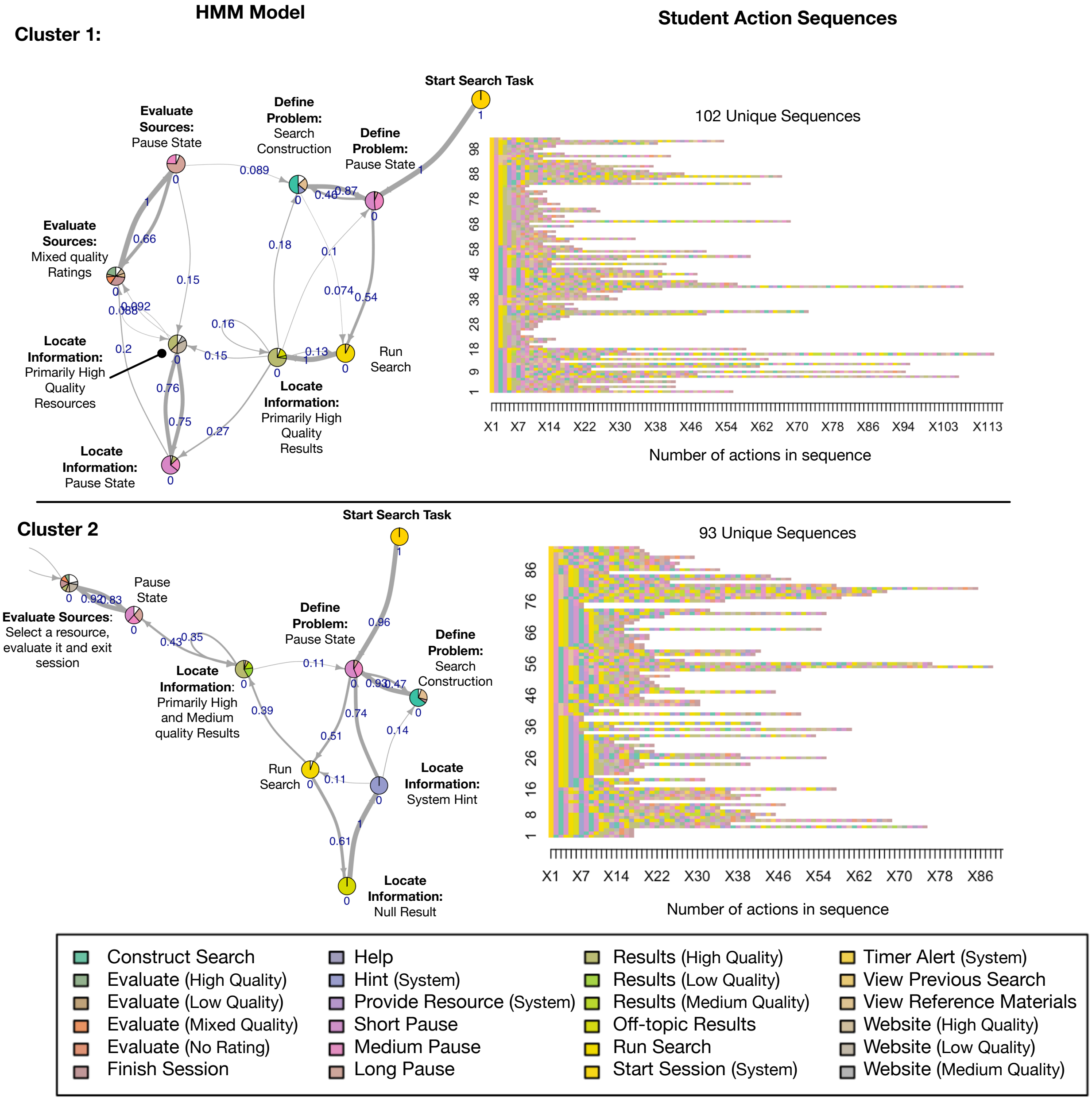
Graph of Hidden Markov Model Fits for Models with Between 4 and 10 Hidden States.



Note. Bayesian Information Criteria (BIC) indicates goodness of fit while penalizing for added parameters. Lower scores indicate better fits. Results show Cluster 1 and Cluster 2 sequences are best fit by 9 state models, whereas Cluster 3 and Cluster 4 sequences are best fit by 5 state models.

Figure 4

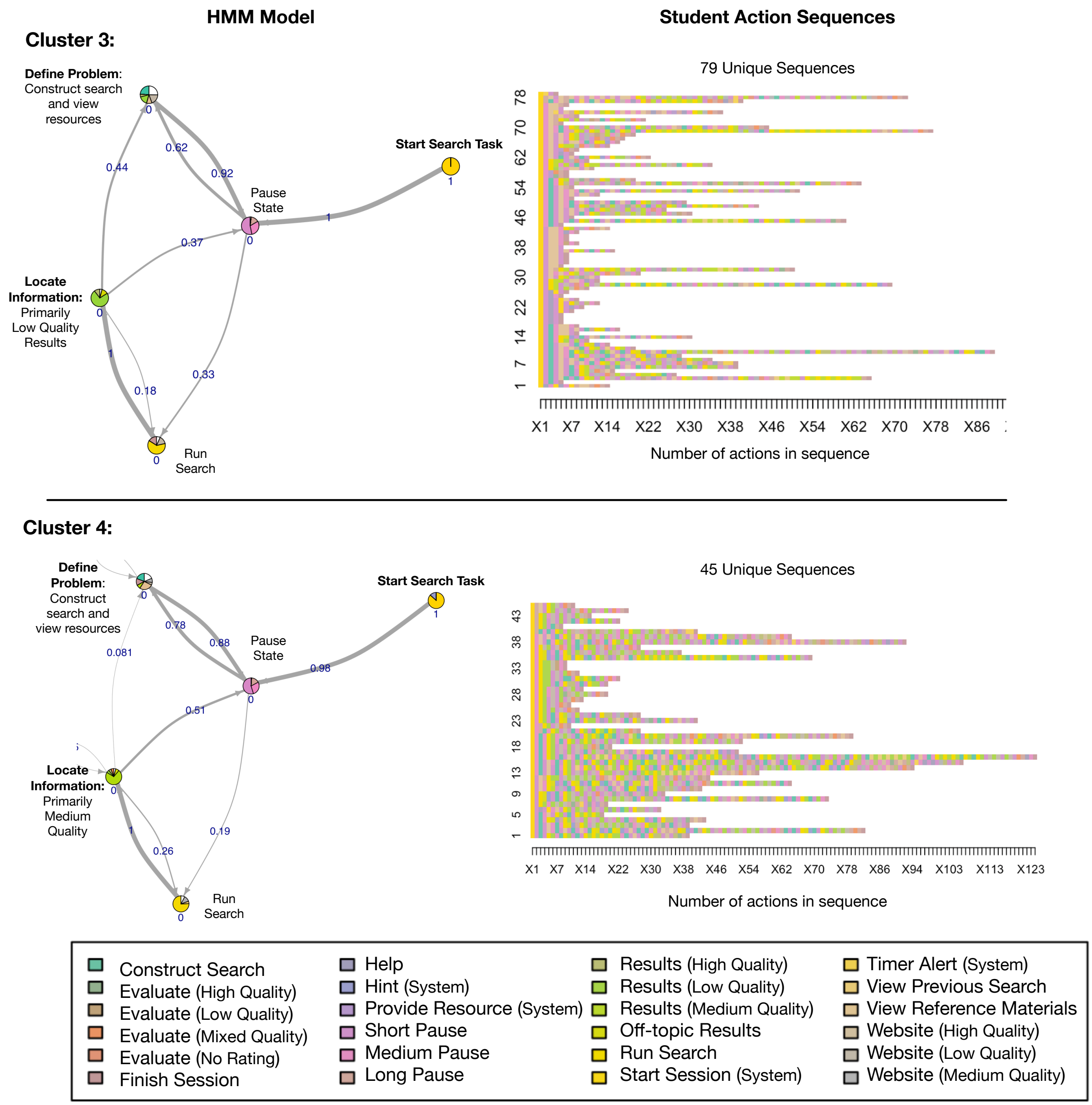
Cluster Analysis Results, Clusters 1 and 2



Note. Node and arrow representation of HMMs fit to Cluster 1 and Cluster 2 sequences are displayed on the left side, next to a visual representation of the raw action sequences that the Edit Distance Clustering approach clustered together. Nodes represent hidden states with the color reflecting the probability of the hidden state emitting action events (color coded in the legend along the bottom). Arrows represent the transition probabilities, with labels and density reflecting specific probabilities. For readability we do not display transition probabilities less than .05.

Figure 5

Cluster Analysis Results, Clusters 4 and 5



Note. Node and arrow representation of HMMs fit to Cluster 3 and Cluster 4 sequences are displayed on the left side, next to a visual representation of the raw action sequences that the Edit Distance Clustering approach clustered together. Nodes represent the hidden states with the color reflecting the probability of the hidden state emitting action events (color coded in the legend along the bottom). Arrows represent the transition probabilities, with labels and density reflecting specific probabilities. For readability we do not display transition probabilities less than .05.

Table 2

Spearman Correlations (ρ) of Inquiry Task Performance Variables with Proportion of Search Sessions in Each Cluster

	Proportion Sessions in Cluster 1	**Proportion Sessions in Cluster 2**	**Proportion Sessions in Cluster 3**	**Proportion Sessions in Cluster 4**
Total Task Score				
Inquiry Task Total Score (max: 100)	**.310**	**-.325**	.035	.019
Task Phase-Level Subscores				
Task Phase: Setup (max: 12)	**.228**	**-.251**	.066	-.017
Task Phase: Free Roam (max: 51)	**.209**	**-.260**	.044	.098
Task Phase: Conclusion (max: 37)	**.203**	**-.259**	.117	-.005
Construct Subscores				
Subconstruct: Planning (max: 6)	.100	**-.240**	**.207**	.083
Subconstruct: Locating (max: 22)	.150	-.197	.084	.015
Locating: Questioning (max: 7)	-.053	-.089	**.202**	.083
Locating: Searching (max: 4)	**.407**	**-.254**	**-.226**	-.041
Locating: Choosing Sources (max: 5)	.095	-.112	.070	-.053
Locating: Saving Sources (max: 6)	.117	-.176	.112	.001
Subconstruct: Evaluating (max: 35)	**.260**	**-.275**	-.019	.099
Evaluating: Importance (max: 7)	**.215**	**-.257**	.029	.095
Evaluating: Usefulness (max: 14)	**.247**	**-.236**	-.038	.052
Evaluating: Trustworthiness (max: 14)	**.225**	**-.245**	-.025	.118
Subconstruct: Synthesis (max: 37)	**.203**	**-.259**	.117	-.005

Note. Values exceeding |.20| appear in boldface.