

# Evaluating Information Visualizations with Working Memory Metrics



Alisa Bandlow, Laura E. Matzen, Kerstan S. Cole, Courtney C. Dornburg  
Charles J. Gieseler, John A. Greenfield, Laura A. McNamara, Susan M. Stevens-Adams

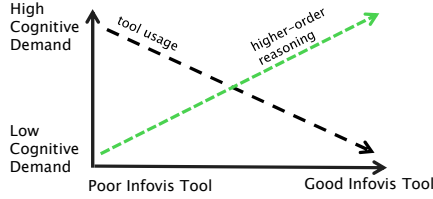


## Do infovis tools help analysts do their jobs better?

### Evaluation Based on Cognitive Resources

Analytical tasks are cognitively demanding and human cognitive resources are finite

- Effective visualizations should minimize the cognitive demands associated with data-driven reasoning and tool usage.



Evaluations based on cognitive resources would provide a standardized, quantitative measure for assessing infovis tools

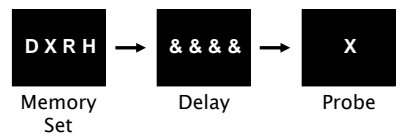
- Measures that have important implications for a tool's impact and effectiveness.
- Could be used as a general metric to compare tools across different analysts, tasks, and data sets.
- Would make comparisons easier, more transparent, and more cost-effective.

### Working Memory Evaluation

Working memory (WM) is a mental workspace: the ability to hold task-relevant information in mind and manipulate it in order to perform a cognitive task

- WM is involved in managing information processing, sustaining attention, and engaging other cognitive processes as needed. It is essential to complex cognition!
- It has a limited, measurable capacity.

A classic working memory test: (Sternberg, 1969)



**Was the probe in the memory set?**  
Participant must hold and manipulate items in memory to answer the question.  
Memory set size can be increased to test limits of working memory.

## Study Design

### Primary Task

- Present three versions of each graph, with 20, 40 and 80 elements
  - Questions that are a "good fit" for the vertex-edge graph are a "bad fit" for the ring graph, and vice versa.
- Answer the NASA-TLX questionnaire after each question.

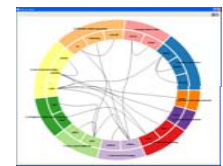
### Example "good fit" questions for vertex-edge graph

- Find the person "Jill".
- Which person has the highest number of first-degree connections to other people?
- What is the shortest path between Jill and Jack?



### Example "good fit" questions for ring graph

- How many communities are in this graph?  
How many members does it have?
- Which is the largest community?
- Which community has the greatest number of connections among its members?



### Secondary Task

Memory set of three letters presented before every task. Random string of letters presented over computer speakers. Click the mouse when a letter from the memory set is heard.

**D X O**

**Z H X A K U V O A D J**

**Z H X A K U V O A D J**

## Hypothesis

### Primary Task

- Will take more time to answer "bad fit" questions than "good fit" questions.
- As graph size increases, will take more time to answer questions and will make more errors.
- As graph size increases, performance difference between "bad" and "good fit" questions will increase.

Secondary task results will mirror primary task results.

## Study Results

### NASA-TLX subjective evaluations

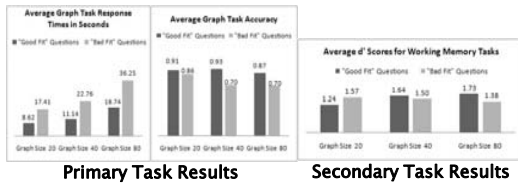
- Primary task rated as more difficult for (a) larger graphs and (b) questions that were a "bad-fit" to the graph type.

### Primary Task

- Performance declined as graph size increased.
- Decline was greater for questions that were a "bad fit" for the graph type.

### Secondary Task

- Performance on the working memory task reflected the difficulty of the primary task.
  - Responses were scored as correct if the participant responded to the target letters before the next letter was presented.
- d' scores:** a measure of the participants' ability to discriminate between the target and distractor items. Calculated using hit rates and false alarm rates.



## Discussion

- Results indicate that a secondary working memory task could be useful for evaluating infovis tools in cases where it is difficult or impossible to assess primary task performance.
- Lesson learned: Select primary task questions that cannot be answered too easily.
  - Some primary task questions were too easy for the size 20 graphs.
  - Participants were able to answer the questions before any targets in the memory set were presented in the secondary task, which led to sparse data.