

Metrics and Measures for Intelligence Analysis Task Difficulty

Frank L. Greitzer

Pacific Northwest National Laboratory
Richland, WA 99352
frank.greitzer@pnl.gov

Kelcy Allwein

Defense Intelligence Agency
Washington, DC 20340
kelcy.allwein@dia.mil

With Panelists: *John Bodnar* (SAIC), *Steve Cook* (NSA), *William C. Elm* (ManTech), *Susan G. Hutchins* (US Naval Postgraduate School), *Jean Scholtz* (NIST), *Geoff Strayer* (DIA), and *Bonnie Wilkinson* (AFRL)

Keywords: usability/habitability; evaluation; task difficulty

Abstract

Evaluating the effectiveness of tools for intelligence analysis presents several methodological challenges. One is the need to control task variables, particularly task difficulty, when conducting evaluation studies. This panel session brings together researchers and stakeholders from the intelligence community to discuss factors to consider in assessing task difficulty.

1. Introduction

An active area of research is the design and development of tools to improve intelligence analysis (IA). Evaluating the effectiveness of tools proposed for introduction into the intelligence community (IC) presents several methodological challenges. One of the most fundamental challenges is the need to control task variables when conducting evaluation studies. For example, such studies require the use of realistic IA tasks and an appropriate research methodology that controls for task difficulty. The focus of this panel session is on characterizing task difficulty and ultimately developing difficulty metrics.

There are at least two reasons for developing a better understanding of IA task difficulty. First, because of the diversity of IA tasks, it is necessary to vary task characteristics (such as difficulty) to assess the effects of a tool being studied as well as its possible interactions with task difficulty. Second, because of the limited opportunity to obtain a relatively large number of analysts to act as experimental subjects, it is not often possible to conduct evaluation studies using “between-subjects” designs that administer the same conditions to two separate groups of analysts—instead, “within-subjects” designs require the same analyst to work a problem with and without a tool. It is obvious that, under these conditions, one cannot use the same IA task for both conditions.

Therefore, there is a need to characterize the difficulty or complexity of IA tasks so it will be possible to control for the effects of task difficulty in evaluation studies. This panel session brings together researchers and stakeholders from the intelligence community to discuss and perhaps, in some cases, to debate factors that need to be considered in assessing task difficulty.

2. Task Difficulty Characterizations

An Initial Set

Several panel participants (Bodnar¹; Greitzer, 2004; Hewett and Scholtz, 2004; Hutchins, Pirolli, and Card, in press; Wilkinson, 2004) have provided opinions on task difficulty, or “What Makes IA hard.” From these, an initial characterization of IA task difficulty includes:

- *Characterization versus Prediction* (Greitzer, 2004; Wilkinson, 2004). Does the task require a description of current capabilities or does it ask the analyst to forecast future capabilities or actions?
- *Sociological Complexity* (Greitzer, 2004) or *Human Behavior* (Wilkinson, 2004). Is the focus of the analysis on an individual, group, State, or region?
- *Data Uncertainty* (Greitzer, 2004). Are the data difficult to observe or interpret? This could arise from a lack of data, from ambiguous, deceptive, or unreliable data, or because the data are of multiple types, different levels of specificity, and/or dynamic/changing over time (Wilkinson, 2004).
- *Breadth of Topic*. Wilkinson (2004) used descriptions like multiple subjects, many variables, and many organizations to describe this idea; Hewett and Scholtz (2004) described this factor in terms of the extent to which the analysis topic is narrowly focused versus broad and open-ended.

¹ Bodnar, J. Personal communication/unpublished document titled “What’s a ‘Difficult’ Intelligence Problem?” (2004)

- *Time Pressure.* The amount of time available to conduct the analysis influences the difficulty in carrying out the task, as has been observed in experimental research (Patterson, Roth, and Woods, 2001), cognitive task analyses (e.g., Hutchins et al., in press), and by the panel discussants in their respective works. It may be argued whether the time variable *per se* is a true task difficulty dimension, as opposed to a possible experimental variable to be studied.
- *Data Availability.* As pointed out by John Bodnar,² “The degree of difficulty in assessing any ... (problem) is related mainly to the data available.” He suggested that one way to assess task difficulty is to compare the amount of data that is potentially “out there” on the topic with the actual amount of data that is realistically available (i.e., possible to obtain or perhaps already obtained).
- *Problem Structure.* To what extent is the problem highly structured with a clearly defined objective, compared to the case in which the problem is ill-structured and requires the analyst to impose a structure? (Hewett and Scholtz, 2004).
- *Data Synthesis.* To what extent does the analyst need to synthesize multiple sources of information?

Discussion Points

- *Feedback.* Wilkinson (2004) suggested that *lack of feedback* is a relevant dimension (“you don’t know when you get it right”; “you don’t know why you got it wrong.”). Discussion issue: While lack of feedback is a problem for analysts that ultimately affects the quality of their products, should this not be distinguished from characteristics intrinsic to the IA tasks themselves that impact their difficulty? Is the problem more a reflection of the lack of repeatability of events and conditions?
- *Experience.* Hewett and Scholtz (2004) suggested that *analyst experience* is a relevant dimension. Discussion issue: Is this a task variable or an analyst variable? Again, this is a matter of whether task difficulty dimensions should be limited to factors that are intrinsic characteristics of tasks alone.
- *Refining the Metrics.* Possible approaches to refining the metrics are questionnaires (e.g., Hewett and Scholtz, 2004); cognitive task analyses (e.g., Hutchins et al., in press); performance-based workstation monitoring/instrumentation (e.g., Cowley, Nowell and Scholtz, 2005); or a combination of these.
- *Next Steps for Evaluation.* Evaluating the impact of tools should be conducted from a model of cognition rather than from typical outcome measures alone (Elm, 2004). The coupling between human and computer components of a system should form the focal point for any metrics or evaluation to provide direct, actionable feedback to improving the design (Woods, 2005)

² *Ibid.*

3. Conclusions

Developing a useful set of IA task difficulty metrics and associated measures is important to the IC community, researchers, and stakeholders because they are needed to assess the impact of new methods and tools that are being considered for introduction into the field. This panel focuses on difficulty metrics as an initial step in the complex process toward defining rigorous research methods and approaches to assessing the impact of IA tools. A second step in this process is the development of performance measures—i.e., measured quantities that are used to compare performance with and without tools. Finally, we must do some thinking and planning for the third step in this research process: the need to design robust and valid experimental or quasi-experimental research to support evaluation of new tools and technologies. It is hoped that this panel discussion is beneficial in laying out the overall research issues and challenges as well as laying some groundwork for the evaluation studies needed to support the deployment of tools for intelligence community.

References

- Cowley, PJ, LT Nowell, and J Scholtz. 2005. Glass Box: An instrumented infrastructure for supporting human interaction with information. *Hawaii International Conference on System Sciences*. January 2005.
- Elm, WC. 2004. *Beauty is more than skin deep: Finding the decision support beneath the visualization*. Technical Report, Cognitive Systems Engineering Center. National Imagery and Mapping Agency. Reston, VA.
- Greitzer, FL. 2004. Preliminary thoughts on difficulty or complexity of intelligence analysis tasks and methodological implications for Glass Box studies. (Unpublished). Pacific Northwest National Laboratory.
- Hewett, T and J Scholtz. 2004. Towards a Metric for Task Difficulty. Paper presented at the NIMD PI Meeting, November 2004, Orlando, FL.
- Hutchins, SG, P Pirolli, and SK Card (in press). What makes intelligence analysis difficult? A cognitive task analysis of intelligence analysts. In R. Hoffman (Ed.), *Expertise out of Context*, Hillsdale, NJ: Lawrence Erlbaum.
- Patterson, ES, EM Roth, and DD Woods. 2001. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work*, 3, 224-237.
- Wilkinson, B. 2004. What makes intelligence analysis hard? Presentation at the Friends of the Intelligence Community meeting, Gaithersburg, MD.
- Woods, DD. 2005. Supporting cognitive work: How to achieve high levels of coordination and resilience in joint cognitive systems. To appear in *Naturalistic Decision Making 7*. Amsterdam, The Netherlands, June 15, 2005.