

<i>Appendix A. Demographic Questionnaire and Full description of analysts</i>	2
<i>Appendix B. Scenario Descriptions</i>	6
<i>Appendix C. Report Template</i>	11
<i>Appendix D: Plenary Checklist</i>	12
<i>Appendix E: IRB Submission Form and Consent Form</i>	13
<i>Appendix F: Workshop Schedule</i>	19
<i>Appendix H. Observer Protocol</i>	22
<i>Appendix I. Observation Worksheet</i>	27
<i>Appendix J. SmiFro Console</i>	28
<i>Appendix K. Status Questionnaire</i>	29
<i>Appendix L. NASA TLX</i>	30
<i>Appendix M. Post-Scenario Questionnaire</i>	31
<i>Appendix N. Post-Session Questionnaire</i>	33
<i>Appendix O. Post-Session Debriefing Interview Schedule (Note: Space between questions deleted in this version)</i>	36
<i>Appendix P. NASA TLX Weighting</i>	37
<i>Appendix Q. Post-System Questionnaire</i>	39
<i>Appendix R. Post-System Debriefing Interview Schedule</i>	44
<i>Appendix S. Cross-Evaluation Interface</i>	45
<i>Appendix T. Scenario Difficulty Assessment</i>	46
<i>Appendix U. Focus Group on Systems Schedule</i>	49
<i>Appendix V. Focus Group of Entire Workshop Schedule</i>	51
<i>Appendix W. Post-Session Questionnaire Results</i>	53
<i>Appendix X. Post-System Questionnaire Results</i>	56
<i>Appendix Z. Post-Scenario Questionnaire Results</i>	60

## Appendix A. Demographic Questionnaire and Full description of analysts

### Demographic Questionnaire

In order to help us interpret results of our experiments, we'd like to know a little more about you. Please respond to the following questions.

1. User ID
2. Age
3. Highest degree
4. What is your full-time job?
5. Military rank
6. Length of military service
7. Length of time doing intel analysis 
  - 7a. If possible, can you tell us a little about the areas you do or did analysis in?
8. Do you have a computer at home?  Yes  No
  - 8a. If so, what tools do you most frequently use?
9. Do you use a computer at work?  Yes  No
  - 9a. If so, what tools do you most frequently use?
10. Do you rely on a computer for analysis work?  Yes  No
  - 10a. If so, what tools do you most frequently use?
11. How would you rate your computer expertise?  Expert  Medium  Novice
12. Do you have any experience in doing data base queries?  Yes  No
  - 12a. If so, how would you rank your expertise?  Expert  Medium  Novice
13. Your Myers-Briggs type (if known)

**Table 1: Full description of demographics of the analysts**

	<b>ch1</b>	<b>ch2</b>	<b>ch3</b>	<b>ch4</b>	<b>ch5</b>	<b>ch6</b>	<b>ch7</b>	<b>ch8</b>
<b>analyst age</b>	44	31	30	40	44	49	34	54
<b>highest degree</b>	Master	MS (MBA Pending)	BA	MS	Ph.D.	B.S.	High School	Masters
<b>civilian job</b>	Hospital Administrator	Portfolio Manager	Investigative Analyst	HOMEMAKER	Intelligence Officer-ARDA Rep	Executive Officer, 10th Civil Support Team (WMD), WA Air National Guard	Supervisor	Law Enforcement
<b>milrank</b>	LCDR	LTjg	CTT1	LCDR	CDR	MAJOR/O-4	E6	Commander (O-5)
<b>milservice (yrs)</b>	18	2.5	12	19	23	31	16	25
<b>analysis (yrs)</b>	0	2.5	12	18	23	5.5	3	19
<b>area of expertise</b>		HUMINT	Criminal Backgrounds, SIGINT, All Source Analysis	Active duty 1985-1989 watch floor air, surf, subsurface analysis. 1995 POW/MIA Korean War analysis. 1996-97 Op South Watch air defence analysis. 2000 Naval Order of Battle analysis.	Scientific, Sigint, General military Intel- US Southern Command, Counterdrug Intel- US Southern Command	Criminal and terrorist activity; identification of chemical/biological/radiological/nuclear weapons of mass disruption/ destruction; homeland defense	DOXX	Primarily basic intell to brief Special Agents and supervisors in current intell and/or a specific topic.



<b>computer expertise - recoded</b>	2	3	3	3	3	3	3	3
<b>experience query</b>	no	yes	yes	yes	yes	yes	yes	yes
<b>query expertise</b>	None	novice	medium	medium	medium	novice	medium	novice

## Appendix B. Scenario Descriptions

The full set of scenarios delivered by AFRL are included in this listing. The ones labeled A-H were used in the evaluation; those labeled 1-6 were only used during the Scenario Complexity Assessment activity. The order of presentation also reflects the order in which the scenarios were reviewed during the Complexity Assessment. Scenarios were used in alphabetic order during the series of block, i.e. during Block 1, analysts worked on Scenario A followed by Scenario B and during the final block, they worked first on Scenario G and lastly Scenario H.

ID	Topic
A	Indian Chemical Weapons Production and Delivery Systems
B	Libyan Chemical Weapons Program
C	Iranian Chemical Weapons Development and Impact
D	North Korean Chemical and Biological Weapons Research
E	Pakistani Chemical Agent Production
F	Current Status of Russia's Chemical Weapons Program
G	South African Chemical Agents Program Status
H	Assessment of Egypt's Biological Weapons
1	Al-Qaeda Terrorist Group and Chemical Weapons Accessibility
2	Taiwanese Chemical Weapons Development
3	Islamic Movement of Uzbekistan's Possession of Chemical Weapons
4	Status Report of Uzbekistan's Chemical Weapons
5	Indonesia's Chemical Weapon Advancements
6	Report on Biological Agents and Warfare

### Scenario F: Current Status of Russia's Chemical Weapons Program

As a new member of the Russian analysis team, the following scenario has been developed to assist in your familiarization with your area of concern and to assist in the updating of the Russian Chemical Weapon database. Since its collapse in the 1990's, there has been much speculation as to the current status of Russia's chemical weapons program. Please research Russia's chemical weapons program with a concentration on chemical agents produced, facilities, scientists, current research, weaponization, and delivery systems. Also, provide information on any countries that have benefited from Russian chemical weapons research and the implications of these findings on the United States. You may add any other relevant information to your report.

- Customer: [doesn't matter]
- Role: Country desk – Russia
- What they want: Status on Russia CW Program [very well defined]

### Scenario B: Libyan Chemical Weapons Program

Before a U.S. military presence is reestablished in Libya, a current, thorough study of Libya's chemical weapons program must be developed. Your task is to produce a report for the Secretary of the United States Navy regarding general information on Libya and the production of chemical weapons. Provide information regarding Libya's access to chemical weapons research, their current capabilities to use and deploy chemical weapons, reported stockpiles, potential development for the next few years, any assistance they have received for their chemical weapons program, and the

impact that this information will have on the United States. Please add any other related information to your report.

- Customer: Secretary of U.S. Navy
- Role: Country desk – Libya
- What they want: General report on Libya and CW production

#### **Scenario H: Assessment of Egypt's Biological Weapons**

As terrorist activity in Egypt increases, the Commander of the United States Army believes a better understanding of Egypt's Military capabilities is needed. Egypt's biological weapons database needs to be updated to correspond with the Commander's request. Focus your investigation on Egypt's access to old technology, assistance received from the Soviet Union for development of their pharmaceutical infrastructure, production of toxins and BW agents, stockpiles, exportation of these materials and development technology to Middle Eastern countries, and the effect that this information will have on the United States and Coalition Forces in the Middle East. Please incorporate any other related information to your report.

- Customer: Commander CENTCOM
- Role: Country Desk – Egypt
- What they want: Updating subject matter report on biological weapons

#### **Scenario G: South African Chemical Agents Program Status**

Though the South African government has disavowed their chemical weapons program in the 1990's, the Joint Forces Command (JFC) needs a complete and up to date analysis of South Africa and its chemical weapons program. Please direct your study on the current chemical weapons capabilities of South Africa, any current or future development of advanced types, chemical weapons developed under Project Coast, relocation of scientists and research efforts under this project, and any implications that this will have on the United States and other countries. Please attach any other supplemental information to your report.

- Customer: JFC
- Role: Country Desk – South Africa
- What they want: Threat assessment of South Africa chemical weapons

#### **Scenario A: Indian Chemical Weapons Production and Delivery Systems**

In 1997, India acknowledged that they had developed a chemical weapons program. The Defense Intelligence Agency is interested in India's current chemical weapons program and general infrastructure for a country study. Please center your report on current deploy and usage capabilities, toxins and BW agents produced, delivery systems and their potentials, location of facilities, future chemical weapons development, assistance or partnerships from civilian companies or countries, and the significance of India's chemical weapons development on neighboring countries and the United States. Your report should contain answers to this information, as well as any other associated information.

- Customer: DIA
- Role: Country desk – India
- What they want: CW program/infrastructure ... included in larger country study

### **Scenario C: Iranian Chemical Weapons Development and Impact**

The Central Intelligence Agency (CIA) is examining chemical weapons information that could effect current United States and coalition operations in the Iraq area of operations. To address the CIA's needs, focus your efforts on chemical weapons access, current capabilities, production of chemical weapons, location of facilities, scientists, intentions, future projections, and the implications that this information has on the United States and Allied Forces, as well as the War on Terror. Please incorporate any other supplemental information to your summary.

- Customer: CIA
- Role: Threat Analyst (I&W)
- What they want: Report of situational awareness; background/info report

### **Scenario D: North Korean Chemical and Biological Weapons Research**

The Defense Intelligence Agency has requested a detailed update of North Korea's country study and their chemical and biological weapons programs that could affect current United States forces in South Korea and Japan. As part of your research, please investigate North Korea's access to weapons research, countries that are believed to have assisted North Korea, current weapons capabilities, old technology, types of chemical munitions produced, stockpiles, locations of facilities, intentions, and the implications of this information on the United States, South Korea, Russia, and Japan. Also report on the future projection of North Korean chemical and biological research and the confidence of the information found. Please include any other related information in your report.

- Customer: DIA
- Role: Country desk – North Korea
- What they want: Chem/Bio weapons country update
  - Very broad
  - Must consider a comparison of 'old' vs 'new' since this is an 'update'.
  - To narrow the problem, the strategy is to go broad rather than deep.

### **Scenario E: Pakistani Chemical Agent Production**

The Defense Intelligence Agency (DIA) has requested a detailed update of Pakistan's country study information. The DIA is also looking for chemical weapons information that could effect current United States and Coalition operations in the Afghanistan area of operations. In your report include production capabilities, ability to deploy and utilize chemical weapons, delivery systems, location of research and storage facilities, partnerships, sponsorships, or assistance received from countries and organizations, and the implications of these findings on the United States Military. Please add any other related information to your summary.

- Customer: DIA
- Role: Country desk – Pakistan
- What they want: Update to country study info including any threat info

### **Scenario 6: Report on Biological Agents and Warfare**

The Biological Warfare Command is currently updating their databases on the agents, uses, weaponization, and proliferators of biological weapons. As an intelligence analyst at the command your task is to help update our information on Anthrax and Smallpox. Please include information on the type of toxins and/or agents these are, production and weaponization techniques, countries and organizations that produce and possess these agents, access to these agents, use of these agents with a concentration on terrorist activity, repercussions of these uses, physical symptoms of exposure, and the impact of this information on different state and non-state actors around the world. You may incorporate any other related information to your report.

### **Scenario 2: Taiwanese Chemical Weapons Development**

There are suspicions that the Republic of Taiwan has been developing chemical weapons. The Department of Defense is requiring the following information on Taiwan's chemical weapons program. In your report, include Taiwan's access to chemical weapons development, current capabilities to deploy chemical weapons, stockpiles, locations of facilities, potential research and development of advanced chemical weapons, and the potential ramifications of these developments on surrounding countries and the United States. Include any other relevant information to your report

### **Scenario 3: Islamic Movement of Uzbekistan's Possession of Chemical Weapons**

After government forces raided and killed 19 Islamic militants in Tashkent, Uzbekistan, the Central Intelligence Agency is requesting a briefing on the terrorist group Islamic Movement of Uzbekistan (IMU) and their possible use of chemical weapons. Focus your study on IMU's access to chemical weapons, current capabilities, intentions of development, countries supporting IMU's chemical weapons program, estimated strength of the IMU and its chemical weapons program, and the implications this information has on this region. Also, briefly reflect on the level of confidence of the information found. Please incorporate any other relevant information to your report.

### **Scenario 5: Indonesia's Chemical Weapon Advancements**

As a member of the Far East Threat Operations Center, Indonesian Desk, the following information on chemical weapons must be provided for the development of a briefing for the Commander of the United States Navy. Please provide information on Indonesia's possession of chemical weapons, development of chemical agents, capabilities of their chemical weapons, delivery systems, locations of facilities and stockpiles, institutions assisting their program, any support from countries, and the potential ramifications of this information on the surrounding area. Include any other related information to your report.

### **Scenario 1: Al-Qaeda Terrorist Group and Chemical Weapons Accessibility**

Since American Forces have been activated to perform operations in the Middle East and Eastern Africa, the Commander of the United States Army would like a security brief on our area of operations. A detailed analysis of the terrorist groups in our area of interest is needed. Your task is to gather analysis on the Al-Qaeda terrorist network

including possible use of chemical weapons, access to other countries' or organizations' chemical weapons technology, current capability to deploy and use chemical weapons, countries that have supported Al-Qaeda's weapons development, prospects for future development of chemical weapons by Al-Qaeda in the upcoming years, and the implications of this information for the United States and other countries that are fighting terrorism. Please comprise your report of this information along with any other related information.

**Scenario 4: Status Report of Uzbekistan's Chemical Weapons**

After the fall of the Soviet Union many of the former republics were left with the legacy of the USSR's chemical weapons program. Uzbekistan has made an agreement with the international community that they will dismantle their weapons and research efforts. The United States State Department has requested a report on the current country study information and Uzbekistan's effort to destroy their chemical weapons program. In your report, include access to chemical weapons development, current weapons and development capabilities, production of chemical agents, status of chemical weapons cleanup, location of facilities and stockpiles, scientists, research, and the potential impact of the dismantling on neighboring countries and the United States. Please incorporate any other supplementary information to your summary.

## Appendix C. Report Template

Rep <ID #>

**Title:** <Scenario title>

**Summary/Background** [use whichever fits the scenario and/or your style]

<Insert

- your best summary description [**Source: doc\_x.ext**] and/or
- write it yourself or
- insert a note that you aren't ready to summarize since you are still actively searching for key information to support this section.>

[All but key evidence removed to approximate size of final report.]

### **Point 1**

Short summary of thrust [# of pieces of evidence in backup]

- Snippet 1 **Source:**
- Snippet 2 **source:**
- Snippet 3 **source:**

### **Point 2**

Short summary of thrust [# of pieces of evidence in backup]

- Snippet 1 **Source:**
- Snippet 2 **source:**
- Snippet 3 **source:**

### **Point 3**

Short summary of thrust [# of pieces of evidence in backup]

- Snippet 1 **Source:**
- Snippet 2 **source:**
- Snippet 3 **source:**

### **Point n**

Short summary of thrust [# of pieces of evidence in backup]

- Snippet 1 **Source:**
- Snippet 2 **source:**
- Snippet 3 **source:**

## Appendix D: Plenary Checklist

### Introductory Session

- Review agenda and schedule
- Explanation of compressed schedule; lunch for today
- Systems providers
  - Stefano – Ginko
  - Andy – Ferret
  - Sharon & Hilda – HITIQA
  - Emile & Joe – GNIST
- Observers
  - Robert
  - Aleksandra
  - Diane
  - Emile
- PNNL Support staff
  - Antonio
  - Ben
  - Troy
  - Katie Johnson
  - Frank Greitzer
- Analysts
- Context – discussion with Katie Johnson
- Deliverable – continued group discussion on quality measures
- Creating deliverable
  - Outline or other organizational scheme
  - Make your strategy clear
  - Handling document identifiers
- User ID/password – card in packet
- SmiFro Console
- Glass Box
- Scenario format – filenames and report IDs
  - Other temporary or supporting files should be kept in proper subdirectory (/My Documents/ch[#])
- Refrain from discussing systems with other participants
- Consent Form
- Heads up about demographic information survey

#### Packets:

- Username and Password
- Agenda
- Schedule
- Consent form
- Context and deliverable quality criteria
- SmiFro Console instructions
- Glass Box instructions
- Scratch paper

# Appendix E: IRB Submission Form and Consent Form

## 1) IRB Form

**Date of Submission:** 2/26/2004

### MEMORANDUM TO: THE NIST HUMAN ETHICS RESEARCH COMMITTEE

#### Through:

Sharon Laskowski \_\_\_\_\_ Date: \_\_\_\_\_  
(Group Manager – Visualization and Usability Group)

Martin Herman \_\_\_\_\_ Date: \_\_\_\_\_  
(Division Chief, Information Access and User Interfaces Division)

#### Submitted by:

Jean Scholtz, Ph.D. – NIST employee  
Building A225, Room A205  
Phone: (301) 975-2520  
[jean.scholtz@nist.gov](mailto:jean.scholtz@nist.gov)  
FAX: (301) 975-5287

Emile Morse, Ph.D. – NIST employee  
Building A225, Room A225  
Phone: (301) 975-8239  
[emile.morse@nist.gov](mailto:emile.morse@nist.gov)  
FAX: (301) 975-5287

1. **Title of Project:** ARDA AQUAINT Challenge Workshop -- Evaluation Methods, Metrics, and a Benchmark
2. **Site:** Northwest Regional Research Center, Pacific Northwest Labs, Richland, WA
3. **Estimated Duration of Study:** 2 weeks (June 14-25, 2004)
4. **Number of Subjects:** 8
5. **Types of Subjects:** Naval reservists
6. **Purpose of Project:** To develop methods and metrics for evaluating interactive software systems and to create a benchmark against which other interactive systems, including question-answering systems, can be evaluated.
7. **Source of Funding:** ARDA Challenge (Cost Center: TBD)
8. **Background and Objectives:**

1. The ARDA AQUAINT (Advanced Question-Answering for INTElligence) program “intends to address a scenario in which multiple, inter-related questions are asked in a focused topic area by a skilled, professional information analyst who is attempting to respond to larger, more complex information needs or requirements”. To further its goals ARDA supports Workshops via a competitive Challenge process. NIST won the 2004 Challenge for AQUAINT and acts as the leader of a group of academic, corporate and government/military researchers.

Organizations represented include: State University of New York at Albany, Language Computer Corporation, and Air Force Research Lab (Rome NY).

2. A key to determining the effectiveness of the analyst interaction is well-planned evaluation of the interaction. “User-centered” evaluations facilitated by NIST have contributed to such efforts as the DARPA IC&V, IM, Communicator and ARDA NIMD programs.

3. User-centered evaluations are done *in the context of use*. Not only must technology work, but also it must work for the users in a real work environment. As a system evaluation, this approach assesses overall system utility, where the system includes the tools and the analysts. Hence, the evaluation is based on an assessment of the analysis product produced, the analysis process used, and the analyst’s perceived usability of the tools.

## **9. Summary of Human Subjects Protocol:**

4. Subjects will be recruited through a contact in the Naval Reserves and the experiments will be conducted during the course of their usual duties. The contract is issued by NSA/ARDA specifically for use in evaluation of ARDA-sponsored projects. Subjects will report to the test facility of the Northwest Regional Research Center at Pacific Northwest Laboratories in Richland, WA each day during the 10-day study period. The first day will provide an overall orientation to the study; consent forms will be signed during the initial session. In addition, a demographic questionnaire will be collected.

5. The next 8 days will be broken into 4 2-day blocks; during a block an individual analyst will be paired with one of three question-answering systems or the baseline (Google) system. Each block has four phases:

- The first half-day of each block will be devoted to training on the system before using it. In addition, a usability test will be given to ensure that the analysts can accomplish the basic tasks with each system before engaging in any of the analytic scenarios.
- During the afternoon of the first day, the first scenario will be presented to the analysts. They will work for up to 4 hours to analyze the problem and develop their reports. Questionnaires will be administered and some debriefing may occur at this time.
- The next morning we will present the analysts with the second scenario and they will use the same system as on the previous day to analyze the problem and prepare a report. Questionnaires will be administered and some debriefing may occur at this time.
- The final half-day of each block will be spent in cross-evaluation. We will split the group into two groups of four analysts. One group will be asked to evaluate the reports from the first scenario, and the other group will evaluate reports from the second scenario. The method for moderator-led cross-evaluation has been used previously by the NIST team as well as other researchers in the Workshop group.

6. On the final day of the data collection phase of the Workshop, the analysts will meet with the system developers and other researchers in a final round of debriefing on their experiences about the Workshop.

7. One of the NIST investigators will be available during the entire period. Training in use of the software will be done by a researcher involved with the development of the particular system. In addition, the computer which the subject uses will be instrumented to capture the subject’s interactions with the system and the computer display will be captured and stored as a movie file as appropriate for the evaluation. The screen capture program also automatically captures voice. The subjects will not be videotaped, but they will be audiotaped. Morning and

afternoon breaks and lunch breaks will be provided. The total time commitment for each subject will be 80 hours.

8. The data that will be collected include: quantitative measures such as time on task and number of errors, qualitative measures such as satisfaction, screen capture with audio, post-scenario debriefing.

10. **Potential Risks and Benefits:** There are no risks to the subjects. There are no direct benefits to the subjects. Longer-term potential benefits include the development of better methods for evaluating software systems used by Federal intelligence analysts. Such evaluations are likely to lead to software that enables analysts to do their jobs more effectively, more efficiently and with greater satisfaction. These evaluations will also allow the ARDA Program Managers to make more informed decisions about resource allocation.

11. **Protection of Subjects against Potential Risks and Protection of the Subjects' Privacy:** Data will be coded to protect the privacy of subjects. Data will be statistically manipulated before reporting mean values, so that individual performance cannot be matched to individuals. Even during the cross-evaluation of reports, individual reports will be coded so that no analyst will know the author of any particular report except his/her own. The data will become part of a benchmark that will be shared with the ARDA community, but only coded data will become part of the shared collection. Only participants in this Workshop will know the identities of the individual analysts.

12. **Certification:** "I hereby certify that the information furnished concerning the procedures to be taken for the protection of human subjects is correct. I will seek and obtain prior Internal Review Board (IRB) approval for any substantive modification of this protocol and will report any unexpected or otherwise significant adverse effects encountered in the course of this study."

---

Signature of Principal Investigator(s) and date

---

Signature of Principal Investigator(s) and date

See attached copy of informed consent form.

## 2) Consent Form

### INFORMED CONSENT

#### ARDA AQUAINT Challenge Workshop – Evaluation Methods, Metrics, and a Benchmark

**Research Description:** ARDA is an “IC center for conducting advanced research and development related to information technology.” It manages programs such as AQUAINT and NIMD to support research into how software tools can assist intelligence analysts to do their jobs more effectively, more efficiently and with greater satisfaction. To further its goals ARDA supports Workshops via a competitive Challenge process. NIST won the 2004 Challenge for AQUAINT and acts as the leader of a group of academic, corporate and government/military researchers. The goal of this workshop is to develop better ways of evaluating software; software is judged by many criteria, such as how fast it is and how accurate it is, but the most important factor is whether people can use the tool to work better. We call this user-centered testing.

You have been asked to participate in this project because you perform analysis as part of your work assignment with the Navy Reserves. You have acquired skills through your training that we are interested in having you apply today. Your role in this study is to perform analysis of up to eight scenarios. You will be performing your work in a specialized computer environment. Two things make the environment special:

1. There is a piece of software that you will be helping to evaluate. There are four systems available for you to use in your analysis work. Three of these systems are Question-Answering systems developed in the AQUAINT Program; the fourth system is a baseline based on Google. Although the researchers have a great deal of expertise, their software has not necessarily undergone the kind of development that leads to a product that is found in a retail outlet. Part of your job is to tell us when the software is behaving poorly so that we can relay this information to the developers.
2. A piece of software called the GlassBox will be capturing your interactions with the computer. The same software will be recording your voice. You will be shown how to ‘Pause’ the software so that you will not be recorded. You may turn off the recording when you are doing work that is not a part of the evaluation we are asking you to perform.

This evaluation period is 2-weeks long and it is divided into four blocks of 2 days plus this day of orientation and a final day of debriefing. Today you will be shown how to use the Glass Box software which will be running on each of the computers that you will be using. We will also allow you time to familiarize yourself with the other software on the computers that you will be using. Let us know if you need something that doesn’t appear to be available. Next Friday you will be asked to give the researchers feedback about your experiences during the workshop. Your comments are vital to the development of more robust tools for aiding the intelligence analysis process.

During the morning of the first day of a block, you will receive training in using one of the software systems. The training will consist of lecture-style instruction as well as hands-on experience. To assess whether the training is adequate, we may administer questionnaires or have you perform short scenarios. You will be encouraged to ask questions and to ask for additional time so that you are confident before the rest of the testing begins.

You will be given scenarios to analyze on the afternoon of the first day of a block and on the morning of the second day. The tasking will describe the context and the deliverable. You will have up to 4 hours to perform the analysis. Your report will be printed and copied so that it can be ranked by your colleagues in the next phase of the block. In order for the ranking to be

anonymous, do not include your name in any part of the briefing report. When you are finished or when time expires, you will be asked to fill in an on-line survey questionnaire and to respond to questions from the researchers.

The ranking mentioned above is one part of the final phase of each block. Our goal is to make measurements about the process that you use to do analysis; we gain insights by looking at the database that the Glass Box creates as well as by analyzing some of the results of questionnaires that you have filled in. However, there is much to be learned by looking at the product of analysis, that is, the briefing outlines that you will prepare. We will be using a method called Cross-Evaluation. One of the workshop researchers will act as the moderator and you will be asked to join in a discussion with other analysts regarding the set of reports that were generated during the block. Two different scenarios will need to be considered. Although it may appear that you are being evaluated, we want you to know that we are less concerned with the individual reports than we are with the various software systems.

During subsequent 2-day blocks, you will be rotated through each of the systems and will work on 2 new scenarios with each of them. On each day when a block is in progress, you will be asked to work for up to 8 hours. You make take breaks when you need to and there will be a break at lunch time. You are free to withdraw at any time.

**Risks/Benefits:** There are no risks associated with your performance in this study. Although there are no direct benefits to you for participating in this study, longer-term potential benefits include the development of better methods for evaluating software systems used by Federal intelligence analysts. Such evaluations are likely to lead to software that enables analysts to do their jobs more effectively, more efficiently and with greater satisfaction.

**Confidentiality:** All user-supplied data will be stored using coded values rather than your name. The study's investigators will control access to the data. To the extent permitted by law including the Freedom of Information Act (FOIA), the key that relates names and codes will not be shared outside the team of researchers who are participating in this Workshop; a full listing of the Workshop participants can be found in your Workshop packet. Publications and presentations based on the data gathered in this study will be structured so as to hide the identity of individuals participating in the study.

**Contact:** If you have questions about the research study or about your role and rights within the context of the study, you may contact either of the Principal Investigators:

Jean Scholtz, PhD  
Building 225, Room A205  
Phone: (301)975-2520  
mailto: jean.scholtz@nist.gov  
FAX: (301)975-5287

Emile Morse, PhD  
Building 225, Room A225  
Phone: (301)975-8239  
mailto: emile.morse@nist.gov  
FAX: (301)975-5287

**“I have read the above description of this research project. I have also spoken to Jean Scholtz and/or Emile Morse, who answered any questions I had about this project. I acknowledge that I have received a personal copy of this form. I agree to participate in this research and I understand that I may withdraw at any time.”**

**Subject's Signature:** \_\_\_\_\_ **Date:** \_\_\_\_\_

**Investigator's Signature:** \_\_\_\_\_ **Date:** \_\_\_\_\_

**Witness Signature:** \_\_\_\_\_ **Date:** \_\_\_\_\_

## Appendix F: Workshop Schedule

### **ARDA Metrics Challenge Workshop Northwest Regional Research Center, June 9<sup>th</sup>-June 25<sup>th</sup> Agenda**

June, Wednesday 9th	08:30	Arrival (technical folks)
	09:30-17:00	Systems setup
June, Thursday 10th	08:30-17:00	Systems setup
June, Friday 11th	08:30-17:00	Systems setup
June, Monday 14th	07:30-08:30	Arrival (analysts)
	08:30-10:30	Intro plenary sessions
	10:30-18:00	Block 1
June, Tuesday 15th	08:30-17:00	
June, Wednesday 16th	08:30-17:00	Block 2
June, Thursday 17th	08:30-17:00	
June, Friday 18th	08:30-17:00	Block 3
June, Monday 21st	08:30-17:00	
June, Tuesday 22nd	08:30-17:00	Block 4
June, Wednesday 23rd	08:30-17:00	
June, Thursday 24th	08:30-17:00	Wrap-up (ALL) <ul style="list-style-type: none"> <li>• Scenario Difficulty assessment</li> <li>• “Meet the developers”</li> <li>• Final focus groups (2)</li> </ul>
	• 08:30-10:30	
	• 11:00-13:00	
	• 14:00-17:00	
June, Friday 25th	08:30-17:00	Wrap-up (technical folks)

#### **Rooms**

<b>Appendix H. ISB1</b> 115: NIST, Tel.: 372-6378 550: SUNY@Albany, Tel.: 375-6891 139: Cycorp, Tel.: 375-6765	<b>Appendix H. ISB2</b> 107: Common Room, Tel.: 375-4434 409: LCC, Tel.: 375-3981
---	---

## Block 1

Day 1	Start Time	Duration	Location	Activity
	10:30	1:00	1-139; 2-409; 1-115; 1-550	System Training
	11:00			Check Ride
	11:30	30 min		Exploration
	12:00	30		
	12:30	0:50	Conference 2-107	LUNCH
	1:20	10	Conference 2-107	Scenario Discussion
	1:30	2:30	1-139; 2-409; 1-115; 1-550	Scenario Task
	2:00			
	2:30			
	3:00			
	3:30			
	4:00	10		NASA TLX Online Instrument
	4:10	10		Post-Scenario Online Instrument
	4:20	10		Post-Session Online Instrument
	4:30	30		Post-Session Debriefing (15min each)
5:00			Quitting Time!	
Day 2	Start Time	Duration	Location	Activity
	8:15	15 min	Conference 2-107	Scenario Discussion
	8:30	2:30	1-139; 2-409; 1-115; 1-550	Scenario Task
	9:00			
	9:30			
	10:00			
	10:30			
	11:00	10		NASA TLX Online Instrument
	11:10	10		Post-Scenario Online Instrument
	11:20	10		Post-Session Online Instrument
	11:30	30		Post-Session Debriefing (15min each)
	12:00	1:00	Conference 2-107	LUNCH
	12:30			
	1:00	10		NASA TLX Weighting Online Instrument
	1:10	20		Post-System Online Instrument
	1:30	30		Post-System Debriefing (15min each)
	2:00	2:00	1-139; 2-409; 1-115; 1-550	Cross-Evaluation Independent
	2:30			
	3:00			
	3:30			
	4:00	20		Cross-Evaluation Ranking
4:20	40	1-139; 1-115	Group Discussion of Rankings (2 groups)	
5:00			Quitting Time!	

## Blocks 2, 3 & 4

Day 1	Start Time	Duration	Location	Activity		
	9:30	1:00	1-139; 2-409; 1-115; 1-550	System Training		
	10:00			Check Ride		
	10:30	30 min		Exploration		
	11:00	30				
	11:30	1:15	Conference 2-107	LUNCH		
	12:00					
	12:30					
	12:45	15	Conference 2-107	Scenario Discussion		
	1:00	2:30	1-139; 2-409; 1-115; 1-550	Scenario Task		
	1:30					
	2:00					
	2:30					
	3:00					
	3:30				10	NASA TLX Online Instrument
	3:40				10	Post-Scenario Online Instrument
	3:50				10	Post-Session Online Instrument
4:00	30	Post-Session Debriefing (15min each)				
4:30			Quitting Time			
Day 2	Start Time	Duration	Location	Activity		
	8:15	15 min	Conference 2-107	Scenario Discussion		
	8:30	2:30	1-139; 2-409; 1-115; 1-550	Scenario Task		
	9:00					
	9:30					
	10:00					
	10:30					
	11:00	10		NASA TLX Online Instrument		
	11:10	10		NASA TLX Weighting Online Instrument		
	11:20	10		Post-Scenario Online Instrument		
	11:30	10		Post-Session Online Instrument		
	11:40	30		Post-Session Debriefing (15min each)		
	12:10	1:00	Conference 2-107	LUNCH		
	12:30					
	1:10					
	1:10	20		Post-System Online Instrument		
	1:30	30		Post-System Debriefing (15min each)		
	2:00	1:45	1-139; 2-409; 1-115; 1-550	Cross-Evaluation Independent & Ranking		
	2:30					
3:00						
3:45	40	1-139; 1-115	Group Discussion of Rankings (2 groups)			
4:20			Quitting Time!			

# Appendix H. Observer Protocol

## ARDA Challenge Workshop

### OBSERVER PROTOCOL

**Date:** \_\_\_\_\_

**Observer:** \_\_\_\_\_

**Block:** \_\_\_\_\_

After you complete each item below, place an *X* through the number.

### SCENARIO

Before analysts arrive:

1. Verify that you have the correct observer packets and that all materials are labeled correctly.
2. Place observer packets in appropriate places on the observation desk.
3. Verify that you have two copies of each of the two scenario worksheets.
4. Verify that the tutorial materials are next to each machine.
5. Verify that system developer's phone number is available.
6. Verify that writing pens are available at each station.

When analysts arrive:

1. Welcome analysts and remind them of your name.
2. Indicate to each analyst their appropriate work station (Station 1 [left] = Analysts ch1, ch2, ch3, ch4; Station 2 [right] = Analysts ch5, ch6, ch7, ch8).
3. Attach lapel microphone to the plastic necklace part of each analyst's badge, somewhere near the collarbone.
4. Indicate the relevant icons on the computer screen:
  - Glass Box
  - System
  - MS Word
  - Questionnaire Index
  - SmiFro Console
5. *Briefly* remind analysts of the purpose of the SmiFro Console and that you will interrupt them at 30-minute intervals to ask them to complete one of the Status Questionnaires. Show them how each button corresponds to a different point in time (30, 60, 90, 120).
6. Indicate tutorial materials (these should be next to the stations).
7. Tell analysts where you will be in the room and what you will be doing; stress your role as an *observer*.
8. Tell analysts that they can take a break at any time during the session, and that they should just leave everything on their machine running while they are away from the station including the Glass Box.
9. Instruct analysts to start the Glass Box and select the appropriate *task* and *phase* from the drop-down menu.
  - The analyst should remember his/her user name and password, but as a reminder, the *user id* is the analyst's code name and the *password* is this code name plus the next two sequential numbers.

- If the analyst is completing a scenario, then the *task* corresponds to the specific scenario number.
  - If the analyst is doing anything other than a scenario (questionnaire, interview or cross-evaluation), then the task corresponds to Post-Analysis Data Collection.
  - The *Phase* is always Analysis.
10. Analysts should know how to start, pause and stop the Glass Box, but as a reminder:
    - Click Red Button to START recording (button will turn Black)
    - Click Black Button to PAUSE recording (button will turn Red)
    - Click Black Button and then X in the upper left corner to STOP recording
    - ALWAYS PAUSE unless it is the end of the day
  11. Indicate the time to analysts and remind them that they have 2.5 hours to work. Tell them that you will let them know when there are only 30 minutes and 15 minutes remaining.
  12. Give analysts the scenario worksheet and tell them to start; tell them that they can write on the scenario worksheet if they like.
  13. Remind analysts to save their work early and often.
  14. Take note of the time on the observer worksheet and identify appropriate times to interrupt for the 30-minute status questionnaires.

While analysts are working:

1. Observe in silence.
2. If analysts ask a question about the system, refer them to the tutorial.
3. If they cannot find an answer to their question, try and answer the question for them.
4. If you cannot answer their question, call the system developer.

If analysts take break:

1. Record the start and end times of the break on the Observer Worksheet.

When analysts finish the scenario:

1. Remind analyst to PAUSE the Glass Box.
2. Remind analyst to SAVE deliverable.
3. Wait until BOTH analysts are finished before you instruct them to click on the Questionnaire Index icon to complete the NASA TLX. If one analyst finishes before the other, indicate to the early finisher that they are free to take a short break.

NASA TLX

1. Reattach lapel microphone if necessary.
2. Instruct analysts to UNPAUSE Glass Box and select the appropriate task and phase from the drop-down menu.
3. Instruct analysts to access the instrument by clicking on the Questionnaire Index icon on their desktops and following the NASA TLX link.
4. Ask analysts to leave Glass Box running when they finish and to turn around in their chairs and face you (this way you know when both are finished with the TLX). Wait until both analysts are finished before you instruct them to complete the Post-Scenario Questionnaire.

POST-SCENARIO and POST-SESSION QUESTIONNAIRES

1. Instruct analysts to return to the Questionnaire Index page, select the Post-Scenario Questionnaire link, and complete the questionnaire. Tell analysts when they finish this questionnaire they should go back to the Questionnaire Index page, select the Post-Session Questionnaire link, and complete this questionnaire.

2. Indicate to analysts that their results from the second questionnaire (the Post-Session Questionnaire) will display on the computer screen after they press the Submit button. Ask them to leave this information displayed on the screen.
5. Ask analysts to PAUSE the Glass Box when they finish both questionnaires and turn around in their chairs and face you (this way you know when both are finished with the questionnaires). Wait until both analysts are finished before you instruct them further.
6. Identify which analysts you should interview first: If it is the first session of the block, then ask the analyst at Station 1 to go first. If it is the second session of the block, then ask the analyst at Station 2 to go first.
7. Instruct the analyst who will be involved in the first interview to take a quick break, and return to the room in a few minutes (instruct the analyst that if the door is closed upon return, then he should wait outside).
8. Instruct the analyst who will be involved in the second interview to take a break and to return in 15 minutes.

#### POST-SESSION DEBRIEFING INTERVIEW

1. Review the first analyst's Post-Session Questionnaire and identify the questions that you would like to use in the interview.
2. Record the numbers of these questions on the Interview Sheet in the appropriate place.
3. Review the analyst's Observation Worksheet and identify the observations that you would like to use in the interview.
4. Write these numbers on the Interview Sheet in the appropriate place.
5. Repeat Steps 2-4 for the second analyst.
6. Organize two chairs around the appropriate analyst's workstation and ask this analyst come back into the room and sit down with you.
7. Attach lapel microphone.
8. Ask the analyst to UNPAUSE the Glass Box and select the appropriate task and phase.
9. Announce the interview by stating the information at the top of Post-Session Debriefing Interview worksheet.
10. Conduct interview.
11. When you finish the interview:
  - a. If this is Day 1 of the block, then tell the analyst that this is the end of the day. Ask them to STOP the Glass Box and instruct them about where and when to report on the following day.
  - b. If this is Day 2 of the block, then ask the analyst to PAUSE the Glass Box, and direct them to the Conference Room 107 for lunch. NOTE: During lunch you need to fetch *two sets* of all deliverables that were produced by all analysts during the block. These will be used in the Cross-Evaluation, which will take place after lunch.

**We will rely on Ying to get these deliverables printed. Pick them up from the Conference Room.**

Repeat steps 6-11 with the second analyst.

In Situation 11b above, do the following after lunch:

#### NASA TLX WEIGHTING

1. Attach lapel microphone.
2. Instruct analysts to UNPAUSE Glass Box and select the appropriate task and phase from the drop-down menu.
3. Instruct analysts to access the instrument by clicking on the Questionnaire Index icon on their desktops and following the NASA TLX WEIGHTING link.

4. Ask analysts to leave Glass Box running when they finish and to turn around in their chairs and face you (this way you know when both are finished with the TLX).
5. Wait until both analysts are finished before you instruct them to complete the Post-System Questionnaire.

#### POST-SYSTEM QUESTIONNAIRE

1. Instruct analysts to return to the Questionnaire Index page, select the Post-System Questionnaire link and complete the Questionnaire.
2. Indicate to analysts that their results from this questionnaire will display on the computer screen after they press the Submit button. Ask them to leave this information displayed on the screen.
3. Ask analysts to PAUSE the Glass Box when they finish both questionnaires and turn around in their chairs and face you (this way you know when both are finished with the questionnaires). Wait until both analysts are finished before you instruct them further.
4. For the Post-System Debriefing Interview, Analyst 1 will always go first.
5. Instruct Analyst 1 to take a quick break, and return to the room in a few minutes (instruct the analyst that if the door is closed upon return, then he should wait outside).
6. Instruct Analyst 2 to take a break and to return in 15 minutes.

#### POST-SYSTEM DEBRIEFING INTERVIEW

1. Review Analyst 1's Post-System Questionnaire and identify the questions that you would like to use in the interview.
2. Record the numbers of these questions on the Interview Sheet in the appropriate place.
3. Repeat Steps 1-2 for Analyst 2.
4. Organize two chairs around the appropriate analyst's workstation and ask this analyst come back into the room and sit down with you.
5. Attach lapel microphone.
6. Ask the analyst to UNPAUSE the Glass Box and select the appropriate task and phase.
7. Announce the interview by stating the information at the top of Post-System Debriefing Interview worksheet.
8. Conduct interview.
9. When you finish the interview, ask the analyst to PAUSE the Glass Box and take a 15-minute break. Ask the analyst to return to the room for the Cross-Evaluation when finished with the break, but to wait outside if the door is closed.
10. Repeat steps 4-9 with Analyst 2.
11. When you are finished with Analyst 2, ask Analyst 1 to come back into the room for Cross-Evaluation.

#### CROSS-EVALUATION

1. Give each analyst a copy of all deliverables for the first scenario of the block.
2. Attach lapel microphone.
3. Remind analysts to UNPAUSE the Glass Box and select the appropriate task and phase.
4. Tell analysts that they have 1:10 minutes to do the Cross-Evaluation for the first scenario: the first hour is dedicated to the Cross-Evaluation Questionnaire (independent ratings), while 10 minutes is dedicated to ranking. Tell them that you will indicate when they have only 15 and 5 minutes remaining in the hour, and when they should start on the ranking. If they finish the Cross-Evaluation Questionnaire before the hour expires, then they can go on to the ranking. **HOWEVER, BE SURE AND KEEP TRACK OF THE TIME.** Analysts **ONLY** have 10 minutes to complete the ranking, even if they finish the Cross-Evaluation Questionnaire early.
5. Instruct analysts to do the following:
  - a. Return to the Questionnaire Index page, select the Cross-Evaluation link and complete the Questionnaire using each of the deliverables.

- b. When the analyst finishes the Cross-Evaluation Questionnaire, ask them to rank order the deliverables from 1 to 8, with 1 being the BEST. Ask them to write the rank number on the upper left-hand corner of the first page of each deliverable.
      - c. Instruct them to follow the Cross-Evaluation Rank link from the Questionnaire Index page and enter their rankings.
      - d. Don't forget to record the start time of the Cross-Evaluation Questionnaire and Rankings.
6. When BOTH analysts have finished Step 4-5 for the first set of deliverables, provide them with the set of deliverables for the second scenario of the block, and repeat the process. Analysts should start the Cross-Evaluation of the second set of deliverables at the same time.
7. When analysts finish with the Cross-Evaluation Questionnaire and Rankings for the second scenario of the block, they will participate in the group discussion of ranking criteria with two other analysts.
  - a. Group discussions will take place in the following rooms, 1-139 and 1-115. Analysts 1 (ch1, ch2, ch3 and ch4) will report to 1-115. Analysts 2 (ch5, ch6, ch7 and ch8) will report to 1-139.
  - b. Group discussion will be led by Observer B and Observer C. Observer B will take RM 1-115 and Observer C will take RM 1-139. During this time Observers A and D can take a break in Conference Room 107.

# Appendix I. Observation Worksheet

## ARDA Challenge Workshop Observation Worksheet

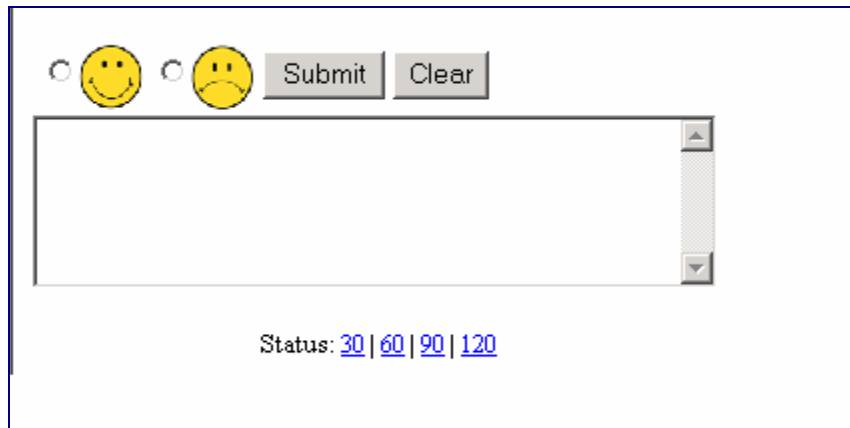
Date: \_\_\_\_\_  
Start Time: \_\_\_\_\_  
End Time: \_\_\_\_\_  
Observer: \_\_\_\_\_



Analyst: \_\_\_\_\_  
System: \_\_\_\_\_  
Scenario: \_\_\_\_\_

TIME	OBSERVATION	NOTES

## Appendix J. SmiFro Console

### Instructions for Using SmiFro Console



○  ○  Submit Clear

Status: [30](#) | [60](#) | [90](#) | [120](#)

#### HOW TO ACCESS:

Before every one of the eight analysis sessions, you will need to place the above webpage on your desktop.

1. Find the short-cut on your desktop that leads to the list of important URLs and questionnaires; double-click it.
2. Select 'SmiFro Console' from the list. This will make the form appear in a browser.
3. If the page is too big, you should adjust the size of the browser frame. If the page is too small – either you can't see the row of blue links at the bottom or you think the comment box is too narrow or too wide – adjust it appropriately.

#### HOW TO USE:

This page has two parts – a mood indicator and a link to intermittent status reports. The **Mood Indicator** should be used to tell the system when you are particularly elated or particularly unhappy or frustrated.

- Just click the appropriate radio button next to the smiling or frowning face.
- If typing a short explanation for you feeling seems natural, do that. If it would take too much time or would be distracting, don't.
- In either case, just click 'Submit' to send you report.
- You may also send plain comments with no mood rating. This is a good way to communicate with the researchers about 'notable events' that happen.

Every 30 minutes your observer will prompt you to choose a **Status** link. Just click on the appropriate hyperlink; a page will pop up in another browser. Answer the questions you find there and Submit your results.

## **Appendix K. Status Questionnaire**

Analysts were prompted to answer online questionnaires every 30 minutes during each analytic session. The content of the questions varied slightly across the time period. The actual wording of the questions is shown below.

### **Status 30**

How is it going so far?

Do you think that you and this system will be able to complete the task?

Is there anything else about your experiences so far that you would like to tell us?

### **Status 60**

How is it going, now?

How do you feel now about whether you and this system will complete the task?

Is there anything else about your experiences so far that you would like to tell us?

### **Status 90**

How is it going, now?

How do you feel now about whether you and this system will complete the task?

Is there anything else about your experiences so far that you would like to tell us?

### **Status 120**

How is it going, now?

How do you feel now about whether you and this system will complete the task?

This is the last mid-search questionnaire for this task; is there anything else about your experiences so far that you would like to tell us?

## Appendix L. NASA TLX

### NASA TLX Instrument

Please rank these six factors in terms of analyzing the returned results and providing feedback for the program you were testing. 1 means little and 7 means much.

Mental demand: whether this searching task affects a user's attention, brain, and focus.

1     2     3     4     5     6     7

Physical demand: whether this searching task affects a user's health, makes a user tired, etc.

1     2     3     4     5     6     7

Temporal demand: whether this searching task takes a lot of time that a user can't afford.

1     2     3     4     5     6     7

Performance: whether this searching task is heavy or light in terms of workload.

1     2     3     4     5     6     7

Frustration: whether this searching task makes a user unhappy or frustrated.

1     2     3     4     5     6     7

Effort: whether a user has spent a lot of effort for this searching task.

1     2     3     4     5     6     7

## Appendix M. Post-Scenario Questionnaire

### Post-Scenario Evaluation Questionnaire

Day and Time: Sun Aug 1 21:46:17 EDT 2004

Scenario ID:

User ID:

**We would like to ask you a few questions about the scenario that you just completed. Please read each of the following questions carefully, and indicate your answers by clicking on the appropriate button.**

1. How similar was this scenario to tasks you typically perform at work?

*dissimilar* *similar*

2. How would you assess the amount of time that you were given to perform this task?

*too little* *just right* *too much*

3. How did the scenario compare in difficulty to tasks that you normally perform at work?

*less difficult* *more difficult*

4. Given you were performing this scenario outside of your standard work environment, without many of your standard resources, how comfortable were you with the process of preparing your deliverable?

*uncomfortable* *comfortable*

5. How realistic was the scenario?

*unrealistic* *realistic*

6. Given that you were performing this scenario outside of your standard work environment, with access to a restricted set of documents, how satisfied were you with the quality of the deliverable that you produced?

*unsatisfied*

*satisfied*

Reset

**Thank you very much for your input!**

## Appendix N. Post-Session Questionnaire

### Post-Session Evaluation Questionnaire

Day and Time: Sun Aug 1 21:49:39 EDT 2004

Scenario ID:

User ID:

**Now, we would like to ask you a few questions about your experiences using the [X] system to complete the scenario. Please read each of the following questions carefully, and indicate your answers by clicking on the appropriate button.**

1. How confident were you of your ability to make the [X] system work to accomplish the assigned task?

*unconfident*

*confident*

2. In comparison to other systems that you normally use for work tasks, how would you assess the length of time that it took to perform this task using the [X] system?

*less time*

*about the  
same*

*more time*

3. The [X] system helped me to better understand the scenario by the information that it provided.

*strongly  
disagree*

*strongly  
agree*

4. If you had to perform a task like the one described in the scenario at work, do you think that having access to the [X] system would help answer specific questions that you currently have trouble answering?

*not at all*

*a lot*

5. In general, how did you find formulating questions that resulted in useful responses from the system?

*easy*

*difficult*

6. In general, how appropriate were the answers that the system provided for the questions that you asked?

*inappropriate*

*appropriate*

7. The [X] system helped me to think about the scenario in new ways.

*strongly  
disagree*

*strongly  
agree*

8. In general, were the answers that the system provided helpful in meeting the goals set forth in the scenario?

*unhelpful*

*helpful*

9. I felt that the system showed me all of the available relevant information about this scenario.

*strongly  
disagree*

*strongly  
agree*

10. After using the [X] system, it is clear to me that there is more information that I need to know about the scenario that I cannot find in the collection.

*strongly  
disagree*

*strongly  
agree*

11. The [NAME] system stimulated my thinking about the scenario.

*strongly  
disagree*

*strongly  
agree*

12. If you had to perform a task like the one described in the scenario at work, do you think that having access to the [X] system would help you find information that you usually have trouble finding?

*not at all*

*a lot*

13. If you had to perform a task like the one described in the scenario at work, do you think

that having access to the [NAME] system would help increase the speed with which you find information?

*not at all*

*a lot*

14. It was easy to gather relevant information about this scenario with the [NAME] system.

*strongly  
disagree*

*strongly  
agree*

15. The [NAME] system expanded my understanding of the scenario.

*strongly  
disagree*

*strongly  
agree*

[Reset](#)

**Thank you very much for your input!!**

## Appendix O. Post-Session Debriefing Interview Schedule (Note: Space between questions deleted in this version)

### ARDA Challenge Workshop

#### POST-SESSION DEBRIEFING INTERVIEW

Date: \_\_\_\_\_  
Start Time: \_\_\_\_\_ Analyst: \_\_\_\_\_  
End Time: \_\_\_\_\_ System: \_\_\_\_\_  
Observer: \_\_\_\_\_ Scenario: \_\_\_\_\_

Re-attach the lapel microphone to the analyst. Ask the analyst to unpause the Glass Box. State the information that is at the top of this form, omitting the end time before you start the interview.

1. How did you feel about using this system during this session?
2. Do you think that you've got the hang of the system?
3. Looking over these answers (from the Post-Session instrument) I see that your most negative thoughts are about [\_\_\_\_\_]. Can you tell me a little more about this? [WRITE DOWN QUESTION NUMBER]  
  
(repeat if there is another negative outlier – use back of this paper if necessary).
4. I see that you gave a pretty high score here on \_\_\_\_\_. Can you tell me more about that? [WRITE DOWN QUESTION NUMBER]  
  
(repeat if there is another positive outlier – use back of this paper if necessary).
5. While you were using the system, I noticed that you \_\_\_\_\_ (from Observation Notes). Can you please tell me more about that? [WRITE DOWN OBSERVATION INDEX NUMBER]  
  
(repeat if there is another observation that you want to ask the analyst to discuss -- use back of this paper if necessary).
6. Is there anything else I should have asked you about your use of the system during this session?
7. Do you have any other comments about the session?

**THANK THE ANALYST :O)**

**If the next activity is LUNCH, ask the analyst to PAUSE the Glass Box and direct the analyst to the Conference Room.**

**If the next activity is QUITTING TIME, ask the analyst to STOP the Glass Box and instruct the analyst about when and where they should report tomorrow.**

## Appendix P. NASA TLX Weighting

### NASA TLX Weighting

User ID      Scenario





<choose> ▼	<choose> ▼
------------	------------

#### C.1.1 Definitions

- Mental demand: whether this searching task affects a user's attention, brain, and focus.
- Physical demand: whether this searching task affects a user's health, makes a user tired, etc.
- Temporal demand: whether this searching task takes a lot of time that a user can't afford.
- Performance: whether this searching task is heavy or light in terms of workload.
- Frustration: whether this searching task makes a user unhappy or frustrated.
- Effort: whether a user has spent a lot of effort for this searching task.

For each pair of workload types in the table below, select the one that was a more important factor when you performed tasks with the current system.

<input checked="" type="checkbox"/> Mental	<input type="checkbox"/> Physical
<input type="checkbox"/> Mental	<input type="checkbox"/> Temporal
<input checked="" type="checkbox"/> Mental	<input checked="" type="checkbox"/> Performance
<input type="checkbox"/> Mental	<input type="checkbox"/> Frustration
<input checked="" type="checkbox"/> Mental	<input checked="" type="checkbox"/> Effort
<input type="checkbox"/> Physical	<input type="checkbox"/> Temporal
<input checked="" type="checkbox"/> Physical	<input checked="" type="checkbox"/> Performance
<input type="checkbox"/> Physical	<input type="checkbox"/> Frustration
<input checked="" type="checkbox"/> Physical	<input checked="" type="checkbox"/> Effort
<input type="checkbox"/> Temporal	<input type="checkbox"/> Performance
<input checked="" type="checkbox"/> Temporal	<input checked="" type="checkbox"/> Frustration
<input type="checkbox"/> Temporal	<input type="checkbox"/> Effort
<input checked="" type="checkbox"/> Performance	<input checked="" type="checkbox"/> Frustration

 Performance	 Effort
 Frustration	 Effort





*strongly  
disagree*

*strongly  
agree*

14. My skill at using the system improved over the course of the workshop.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

15. The system allowed me to easily change my search strategy.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

16. The system would be a useful addition to the tools that I already have at work.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

17. The system would let me stop using some of the tools that I currently use at work.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

18. I couldn't find enough documents with relevant information.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

19. The system is hard to use.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*

20. The system did a good job of helping me to integrate the materials that I found while searching.

1       2       3       4       5

*strongly  
disagree*

*strongly  
agree*



*strongly  
disagree*

*strongly  
agree*

29. The system was inflexible.

1     2     3     4     5

*strongly  
disagree*

*strongly  
agree*

30. It was easy to re-examine my previous searching activities with the system.

1     2     3     4     5

*strongly  
disagree*

*strongly  
agree*

31. The training materials are hard to understand.

1     2     3     4     5

*strongly  
disagree*

*strongly  
agree*

32. The training materials contain most of the information I needed to learn to use the system.

1     2     3     4     5

*strongly  
disagree*

*strongly  
agree*

33. The system often presented me with redundant information.

1     2     3     4     5

*strongly  
disagree*

*strongly  
agree*

Reset

Thank you very much for your input!!

# Appendix R. Post-System Debriefing Interview Schedule

## ARDA Challenge Workshop

**POST-SYSTEM DEBRIEFING INTERVIEW** (Note: Space between questions deleted in this version)

Date: \_\_\_\_\_  
Start Time: \_\_\_\_\_  
End Time: \_\_\_\_\_

Observer: \_\_\_\_\_  
Analyst: \_\_\_\_\_  
System: \_\_\_\_\_

At the beginning of the Glass Box recording state the information that is at the top of this form, omitting the end time.

1. Do you feel that you've gotten the hang of it?
  
2. Did you think the system was particularly better suited to one of the scenarios, rather than the other? Which one? Why?
  
3. Looking over these answers (from online instrument) I see that your most negative thoughts are about \_\_\_\_\_. Can you tell me a little more about this? [WRITE DOWN QUESTION NUMBER]  
  
(repeat if there is another negative outlier – use back of this paper if necessary).
  
4. I see that you gave a pretty high score here on \_\_\_\_\_. Can you tell me more about that? [WRITE DOWN QUESTION NUMBER]  
  
(repeat if there is another positive outlier – use back of this paper if necessary).
  
5. Is there anything else I should have asked about this system?
  
6. Do you have any other comments about the system?

**THANK THE ANALYST :O)**

## Appendix S. Cross-Evaluation Interface

This is a sample of the web page that the analysts used to perform cross-evaluation.

### Evaluating reports on scenario H

We are at the final stage in judging the results of several searches on the same topic. You will examine each of the draft reports that has been prepared for this specific topic. Selecting the report you are reading from the list in the table below. To record your judgment, you must click [SUBMIT] for each report. **Repeat this process until you have judged all user reports.** The judgments you made will show up in the lower part of the window. When you have completed them all, your work is done.

Please judge each summary on the following characteristics. Use scores of 0 (worthless!) to 5 (marvelous!). You can make each choice by clicking on a "radio button".

<b>Please select the report you are evaluating:</b>							Rep12 ▾
	0	1	2	3	4	5(Best)	
<b>Covers the important ground:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Avoids the irrelevant materials:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Avoids redundant information:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Includes selective information:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Is well organized:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Reads clearly and easily:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Overall rating this summary:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Other Comments:</b>							
<div style="border: 1px solid black; height: 80px; width: 100%;"></div>							
<input type="button" value="SUBMIT"/>							

How you have evaluated reports so far:								
Report name	Covers the important ground	Avoids the irrelevant materials	Avoids redundant information	Includes selective information	Is well organized	Reads clearly and easily	Overall rating this summary	Comment

You have finished 0 reports, there are 8 reports to judge.

You can change your mind by re-submitting your judgment. When you have completed them all, click [here](#) and your work is done.

## **Appendix T. Scenario Difficulty Assessment Instructions**

Please help us in our research on intelligence analysis tools. One challenge that we have in conducting this research is to be able to compare the difficulty of intelligence analysis tasks and to understand the factors that contribute to the difficulty or complexity of intelligence analysis tasks. To help us gain an understanding of these factors, we have designed a survey/questionnaire that we'd like you to fill out. The survey asks you to rate a number of intelligence analysis tasks on the basis of a series of potential difficulty factors. You have worked with some of these tasks as part of the AQUAINT METRICS CHALLENGE study, but you have not worked on all of the tasks. We are interested in your assessments about each task, even if you did not actually work on a given task.



Did the amount of data available to address this scenario make it more or less difficult to produce the required two-pager? [Only scenarios A-H]	more	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	less
Reliability of the sources [Only scenarios A-H]	low	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	high
Rate this tasking on the degree to which there is reason to expect deceit in the data.	low potential	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	high potential

# Appendix U. Focus Group on Systems Schedule

## ARDA Challenge Workshop: Analysts Focus Group Guide

Date: \_\_\_\_\_  
Location: \_\_\_\_\_  
Start Time: \_\_\_\_\_  
End Time: \_\_\_\_\_

### **PARTICIPANTS**

Moderator:

Analysts:

Others Present:

### **RECORDING INSTRUCTIONS**

Label the cassette tape with the information at the top of this guide, excluding analysts and others present. Check the tape player by recording this information at the beginning of the tape. Rewind and verify that things are working properly. **BE SURE** to start the recording again. Make sure that the tape player is recording (wheels spinning, red light on).

### **ENVIRONMENT**

In the space below, please take a moment to draw the configuration of the group. Please be sure to indicate the positions of all participants.

## DISCUSSION

“We’re going to talk now about the systems, rather than the specific scenarios. We have heard a lot from each of you, but we want to take another look at some key questions. Of course these questions don’t have right or wrong answers, but they will help to keep us focused.”

“I am going to begin by sharing a story with you.” [SHOW SLIDE]

### The TACIT CLARINET

The acoustical physicist Arthur Benade once did a careful study of the acoustical wave equations, and made a clarinet with the holes drilled in all the wrong places. No matter which combination of keys he pressed, it would not make a sound. Among his many friends was the principal clarinetist of the Cleveland Orchestra, whom he invited over for dinner. After dinner, he handed his friend the tacit clarinet and said, "What do you think of this." The musician blew into it, and it made no sound. He stopped and thought for a few moments, and then began to play it. He said, "It's a little hard to play, but I've seen worse."

*Contributed by Paul Kantor*

### Questions for Guiding the Discussion

1. Do you think any of these systems would be helpful, as they are right now, for your real work?
2. Could they be improved a little? How?
3. If you could afford to buy just one of them, which one would that be? Why?
4. If you could afford to buy just *two* of them, would you still only pick the best one? Would it make more sense to have two of them? Why -- or why not? Which ones?
5. You did some online judgments about the systems. How confident do you feel about those judgments? Why?
6. Do you think that a person's judgments about these systems would be stable over time? Or would they change with more use? Why?
7. Is there anything else I should have asked you?

**THANK THE PARTICIPANTS :O)  
STOP THE RECORDING.**

# Appendix V. Focus Group of Entire Workshop Schedule

## ARDA Challenge Workshop: Analysts + Observers Focus Group Guide

Date: \_\_\_\_\_  
Location: \_\_\_\_\_  
Start Time: \_\_\_\_\_  
End Time: \_\_\_\_\_

### PARTICIPANTS

Moderator:

Analysts:

Observers:

Others Present:

### RECORDING INSTRUCTIONS

Label the cassette tape with the information at the top of this guide, excluding analysts, observers, and others present. Check the tape player by recording this information at the beginning of the tape. Rewind and verify that things are working properly. BE SURE to start the recording again. Make sure that the tape player is recording (wheels spinning, red light on).

### ENVIRONMENT

In the space below, please take a moment to draw the configuration of the group. Please be sure to indicate the positions of all participants (use the back of this sheet if necessary).

## DISCUSSION

"Now we are going to completely switch gears, and chat about the whole workshop -- which is almost, but not quite, over. Of course these questions also don't have right or wrong answers, but they will help to keep us focused. Some questions are for observers, some are for analysts, and some are for everybody."

1. For **everybody** -- Was there any particular thing about this whole process of observing and being observed that made you feel a little embarrassed?
2. For the **observers** -- you've only been asking questions now, so we'll give you a chance to talk for a while. Did you feel comfortable about having to interrupt the analysts, or having to ask them the *same* questions at each follow up interview?
3. For the **analysts** -- was it a problem that observers interrupted your work? Or that they were asking you the same questions each time?
4. For the **analysts** -- the cross-evaluation stuff. Were you pretty confident in the judgments that you made about the reports? [Actually, we have some statistics here that shows you were in \_\_\_\_\_ agreement most of the time]
5. For the **analysts** -- the business of talking into the lapel mikes. Do you think we got most of the important points, or was it something that you only could do when you happened to remember it?
6. For the **observers** -- did you see some important things that you think we would completely miss, if we go home with nothing but the data, and analyze it without asking you about them?
7. For the **analysts** -- with all the data we have -- screen logs, and tape recordings, and online questionnaires, and cross-evaluations -- we still worry that we might be missing some essential kinds of information, because we can never "see" what you're thinking. What can you say about that?
8. Is there anything else I should have asked you about?

**THANK THE PARTICIPANTS :O)  
STOP THE RECORDING.**

## Appendix W. Post-Session Questionnaire Results

### Results of the Post-Session Questionnaire

This questionnaire was administered at the end of each session (twice during a single block). Each of the seven analysts completed two of these questionnaires (one for each scenario) for each of the systems. Each cell displays the mean (standard deviation), which is composed from 14 data points. Questionnaire items (Q1, Q2, etc.) are organized according to hypotheses and level of significance (p-value).

In most cases, questionnaire items were assessed on a 5-point scale where 1=strongly disagree and 5=strongly agree. However, in some cases, the scale labels differ from this, in which case, these are noted next to the question. Many questions refer to "System [X];" for these questions, the appropriate system name replaced the "X." A "T" appears in the cell when there was a tie between systems.

The rankings below reflect the highest, or best, score for each question. The best value for the question is dependent on the wording of the question. In most cases, the best value is the high score, but in some cases the best value is the low score (this is scale dependent). It should be noted that a column does not represent all scores for a single system. Instead, scores are presented according to rank, regardless of which system they were associated. The score for the baseline system is shaded. When significant, the F-value and degrees of freedom from the General Linear Model is provided along with the probability value (p-value). The values in the column labeled "Diff" are the results of the Scheffe's pair-wise comparison tests.

<b>H1: QA system should result in gathering the same type of information with a lower cognitive workload.</b>							
None							
<b>H2: QA system should assist the user in exploring more paths/hypotheses.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q12: If you had to perform a task like the one described in the scenario at work, do you think that having access to the [X] system would help you find information that you usually have trouble finding? [not at all ... a lot]	4.00 (.68)	3.07 (1.28)	2.79 (1.25)	2.47 (.64)	7.62	.000	1 > 2, 3, 4
<b>H3: QA systems should enable analysts to produce higher quality reports.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q4: If you had to perform a task like the one described in the scenario at work, do you think that having access to the [X] system would help answer specific questions that you are currently having trouble answering? [not at all ... a lot]	4.14 (.66)	3.13 (1.13)	2.87 (.64)	2.56 (1.28)	7.66	.000	1 > 2, 3, 4
Q8: In general, were the answers that the system provided helpful in meeting the goals set forth in the scenario? [unhelpful ... helpful]	4.14 (.53)	3.13 (1.06)	3.00 (1.24)	2.67 (1.18)	6.17	.001	1 > 2, 3, 4
<b>H4: QA systems should provide useful suggestions to analysts.</b>							

None							
<b>H5: QA systems should provide more good surprises than bad.</b>							
None							
<b>H6: QA systems should allow the analyst to focus more on analysis, higher level, than to focus on data collections efforts.</b>							
None							
<b>H7: QA systems should allow the analyst to collect more data in less time.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q2: In comparison to other systems that you normally use for work tasks, how would you assess the length of time that it took to perform this task using the [X] system? [less time ... about the same ... more time]	2.21 (.97)	3.43 (1.28)	3.73 (.70)	3.73 (1.22)	9.21	.000	1 < 2, 3, 4
Q13: If you had to perform a task like the one described in the scenario at work, do you think that having access to the [X] system would help increase the speed with which you find information? [not at all ... a lot]	4.07 (.92)	3.00 (1.20)	2.50 (1.29)	2.27 (.70)	10.827	.000	1 > 2, 3, 4
<b>H8: QA systems should allow the analyst to reduce reading time.</b>							
None							
<b>H9: QA systems should provide the analyst with identification of gaps in the knowledge base.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q10: After using the [X] system, it is clear to me that there is more information that I need to know about the scenario that I cannot find in the collection.	2.93 (.92)	3.36 (1.01)	3.47 (.99)	3.60 (.99)	-	-	-
<b>H10: QA systems should help the analyst recognize gaps in their thinking.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q3: The [X] system helped me to better understand the scenario by the information that it provided.	3.79 (.80)	3.13 (1.06)	2.79 (.89)	2.53 (.83)	5.54	.002	1 > 2, 3, 4
Q11: The [X] system stimulated my thinking about the scenario.	3.79 (.70)	3.20 (.77)	3.00 (1.24)	2.47 (.83)	5.86	.002	1 > 3, 4
Q15: The [X] system expanded my understanding of the scenario.	3.71 (.83)	3.13 (.83)	3.07 (1.21)	2.60 (.51)	4.28	.009	1 > 4
Q7: The [X] system helped me to think about the scenario in new ways.	3.64 (.93)	3.13 (.99)	2.79 (.97)	2.53 (.83)	3.96	.013	1 > 3, 4
<b>H11: QA systems should provide relevant context for information.</b>							
None							

<b>H12: QA systems should provide coherence and continuity for the user.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q6: In general, how appropriate were the answers that the system provided for the questions that you asked? [inappropriate ... appropriate]	4.00 (.78)	3.33 (1.11)	2.86 (1.41)	2.73 (.96)	4.82	.005	1 > 3, 4
<b>H13: QA systems should allow analyst to locate a specific document or piece of information previously seen.</b>							
None							
<b>H14: QA systems should be easy to learn and use.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q5: In general, how did you find formulating questions that resulted in useful responses from the system? [easy ... difficult]	2.00 (.96)	2.73 (.88)	3.53 (.64)	3.71 (1.27)	13.90	.000	1 < 2, 3, 4
Q14: It was easy to gather relevant information about this scenario with the [X] system.	4.14 (.66)	3.13 (1.30)	2.79 (1.37)	2.53 (.92)	6.44	.001	1 > 2, 3, 4
<b>H15: QA systems should make its user confident in the exploration of the available data, and in the report produced as a result.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,58)</b>	<b>p-value</b>	<b>Diff.</b>
Q9: I felt that the system showed me all of the available relevant information about this scenario.	3.57 (.85)	2.87 (1.13)	2.57 (1.02)	2.20 (.68)	7.15	.000	1 > 2, 3, 4
Q1: How confident were you of your ability to make the [X] system work to accomplish the assigned tasks? [unconfident ... confident]	4.43 (.65)	3.80 (1.08)	3.47 (1.19)	2.86 (1.35)	5.00	.004	1 > 3, 4

# Appendix X. Post-System Questionnaire Results

## Results of the Post-System Questionnaire

This questionnaire was administered at the end of each block. Each of the seven analysts completed one of these questionnaires for each of the systems. Each cell displays a mean (standard deviation), which is composed from seven data points. Questionnaire items (Q1, Q2, etc.) are organized according to hypotheses and level of significance (p-value).

All questionnaire items were assessed on a 5-point scale where 1=strongly disagree and 5=strongly agree. A “T” appears in the cell when there was a tie between systems.

The rankings below reflect the highest score for each question. The best value for the question is dependent on the wording of the question. In most cases, the best value is the high score, but in some cases the best value is the low score (this is scale dependent). It should be noted that a column does not represent all scores for a single system. Instead, scores are presented according to rank, regardless of which system they were associated. The score for the baseline system is shaded. When significant, the F-value and degrees of freedom from the General Linear Model is provided along with the probability value (p-value). The values in the column labeled “Diff” are the results of the Scheffe’s pair-wise comparison tests.

<b>H1: QA system should result in gathering the same type of information with a lower cognitive workload.</b>							
-							
<b>H2: QA system should assist the user in exploring more paths/hypotheses.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q10: The system helped me think of new ways to search for information.	3.86 (.69)	3.14 (.90)	3.14 (1.07)	2.29 (.49)	3.90	.030	1 > 4
Q1: The system allowed me to easily change my line of questioning.	4.00 (.58)	3.43 (.98)	2.71 (1.11)	2.29 (1.25)	3.51	.042	1 > 3, 4
Q22: The system often presented me with novel information.	3.49 (.79)	3.00 (.82)	2.86 (.69)	2.29 (.49)	3.11	.058	1 > 4
Q15: The system allowed me to easily change my search strategy.	3.57 (.96)	3.29 (1.38)	3.14 (1.21)	3.14 (1.21)	-	-	-
Q27: Having the system at work would help me find information that I cannot currently find.	3.29 (1.25)	2.86 (1.07)	2.86 (1.07)	2.43 (.53)	-	-	-
<b>H3: QA systems should enable analysts to produce higher quality reports.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q9: The system helps me find important information.	4.29 (.49)	3.43 (.79)	2.86 (.69)	2.86 (1.25)	5.78	.008	1 > 3,4
Q32: The system often presented me with redundant information.	2.71 (1.11)	3.57 (1.40)	3.71 (1.11)	4.14 (.90)	3.91	.030	1 < 4 2 < 4
Q18: I could not find enough documents with relevant information.	2.00 (.58)	2.86 (1.07)	3.43 (.98)	3.71 (1.50)	3.57	.040	1 < 3,4
<b>H4: QA systems should provide useful suggestions to analysts.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>

Q11: The system provided me with suggestions about which I had not thought.	3.57 (.98)	3.29 (1.50)	2.57 (.53)	1.86 (.69)	3.68	.036	1 > 4 2 > 4
<b>H5: QA systems should provide more good surprises than bad.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q24: The system often surprised me by doing bad things.	1.86 (.69)	2.71 (.76)	3.00 (.58)	3.29 (1.50)	3.45	.044	1 < 3, 4
Q7: The system often surprised me by doing good things.	3.14 (.69)	3.00 (.58)	2.71 (1.25)	2.57 (.53)	-	-	-
<b>H6: QA systems should allow the analyst to focus more on analysis, higher level, than to focus on data collections efforts.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q13: I spent more time reading documents than I did analyzing the information that the system provided.	3.14 (.69)	3.43 (.79)	3.43 (1.51)	4.14 (.69)	-	-	-
<b>H7: QA systems should allow the analyst to collect more data in less time.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q23: Having the system at work would help me find information faster than I can currently find it.	4.00 (.82)	2.57 (.96)	2.57 (.96)	2.43 (1.13)	5.95	.007	1 > 2, 3, 4
Q6: The system slows down my process of finding information.	2.43 (1.27)	3.00 (1.63)	3.14 (.69)	3.42 (1.40)	-	-	-
<b>H8: QA systems should allow the analyst to reduce reading time.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q3: I spent more time reading documents than I did searching for information.	3.43 (1.40)	3.86 (.38)	4.14 (.90)	4.71 (.49)	-	-	-
<b>H9: QA systems should provide the analyst with identification of gaps in the knowledge base.</b>							
-							
<b>H10: QA systems should help the analyst recognize gaps in their thinking.</b>							
-							
<b>H11: QA systems should provide relevant context for information.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q5: Most of the time, I understood the relationship between the question that I asked and the answer that the system provided.	4.00 (1.15)	3.57 (.98)	3.00 (.82)	3.00 (1.00)	3.57	.040	1 > 3,4
<b>H12: QA systems should provide coherence and continuity for the user.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q25: The system provided coherent answers.	4.00 (.82)	3.00 (.58)	3.00 (1.00)	2.86 (.69)	4.23	.024	1 > 2, 3,4
Q8: The system allowed me to navigate easily between searching activities.	4.57 (.53)	3.57 (.98)	3.00 (1.15)	3.00 (1.73)	4.08	.026	1 > 3,4

Q20: The system did a good job of helping me to integrate the materials that I found while searching.	3.86 (.69)	3.57 (1.13)	2.86 (1.21)	2.29 (.76)	-	-	-
<b>H13: QA systems should allow analyst to locate a specific document or piece of information previously seen.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q29: It was easy to re-examine my previous searching activities with the system.	4.43 (.79)	3.57 (1.13)	3.57 (1.13)	3.00 (.58)	-	-	-
<b>H14: QA systems should be easy to learn and use.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q19: The system is hard to use.	1.29 (.49)	2.14 (.90)	2.43 (1.27)	3.00 (.82)	4.44	.020	1 > 3,4
Q28: The system was inflexible.	2.00 (1.00)	2.86 (1.25)	3.00 (.82)	3.57 (1.27)	4.31	.022	1 > 3,4
Q2: It was difficult to get the system to do what I wanted it to do.	2.14 (.69)	2.86 (.90)	3.14 (.90)	3.29 (1.11)	-	-	-
Q14: My skill at using the system improved over the course of the workshop.	4.29 (.49)	4.00 (.58)	4.00 (1.00)	3.57 (1.27)	-	-	-
Q21: I feel that I have become pretty proficient at using the system.	4.29 (.76)	3.71 (1.11)	3.43 (1.27)	3.00 (.82)	-	-	-
Q33: In general, I like using the system.	4.43 (.53)	3.14 (1.35)	3.14 (1.46)	3.00 (.82)	-	-	-
<b>H15: QA systems should make its user confident in the exploration of the available data, and in the report produced as a result.</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q4: I often wanted to verify the system's answers.	2.43 (.96)	3.00 (.82)	3.43 (1.27)	3.43 (1.27)	-	-	-
<b>System Readiness</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q12: The system is not ready yet to be used in my regular work environment.	2.14 (1.07)	2.71 (.95)	3.43 (1.13)	4.14 (.90)	6.15	.006	1 > 3,4
Q16: The system would be a useful addition to the tools that I already have at work.	4.43 (.53)	3.29 (.49)	3.00 (1.41)	2.86 (1.21)	4.40	.021	1 > 2, 3, 4
Q17: The system would let me stop using some of the tools that I currently use at work.	2.71 (1.12)	2.29 (.76)	2.14 (1.07)	1.86 (.90)	-	-	-
<b>Training</b>							
<b>RANK:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>F(3,27)</b>	<b>p-value</b>	<b>Diff.</b>
Q30: The training materials were hard to understand.	1.57 (.79)	1.57 (.98)	2.43 (1.13)	3.00 (.58)	5.09	.013	1 > 4 2 > 4

Q26: The training on the first day gave me the skills needed to use the system successfully.	4.57 (.53)	4.29 (.95)	3.57 (1.40)	3.00 (1.15)	3.21	.053	1 > 4 2 > 4
Q31: The training materials contained most of the information I needed to learn to use the system.	4.14 (.69)	4.14 (.90)	3.71 (.95)	3.14 (1.06)	-	-	-

## Appendix Z. Post-Scenario Questionnaire Results

### Results of Post-Scenario Questionnaire

Scenario Assessment [5-point scales]								
Q1: How similar was this scenario to tasks you typically perform at work? [dissimilar ... similar]								
B	A	D	C	E	F	H	G	Sig.
3.00 (1.15)	3.14 (1.46)	3.29 (1.50)	3.29 (1.60)	3.71 (.95)	3.71 (.95)	3.86 (.90)	3.86 (.90)	p=.041 B < E, F, G, H; A < G, H
Q2: How would you assess the amount of time that you were given to perform this task? [too little ... just right ... too much]								
A	D	G	H	F	E	B	C	Sig.
2.57 (1.51)	2.43 (.98)	2.71 (.49)	2.71 (.49)	2.86 (.69)	3.00 (1.00)	3.14 (1.07)	3.29 (.76)	-
Q3: How did the scenario compare in difficulty to tasks that you normally perform at work? [less difficult ... more difficult]								
A	F	G	H	E	B	C	D	Sig.
3.00 (1.00)	3.00 (.58)	3.00 (.00)	3.00 (.00)	3.14 (.38)	3.14 (.90)	3.29 (.49)	3.43 (.79)	-
Q5: How realistic was the scenario? [unrealistic ... realistic]								
D	E	H	B	G	A	C	F	Sig.
3.29 (.49)	3.43 (.79)	3.43 (.79)	3.57 (.98)	3.86 (.90)	4.00 (1.29)	4.00 (.82)	4.00 (1.00)	-

Baseline Shaded.

Systems ordered according to rank for each Scenario.

H3: QA systems should enable analysts to produce higher quality reports.								
Q6: Given that you were performing this scenario outside of your standard work environment, with access to a restricted set of documents, how satisfied were you with the quality of the deliverable that you produced? [unsatisfied ... satisfied]								
	A	B	C	D	E	F	G	H
	4.50 (.70)	4.00 (.00)	5.00 (.00)	5.00 (.00)	4.00 (1.41)	4.50 (.71)	5.00 (.00)	5.00 (.00)
	4.00 (.00)	4.00 (1.41)	4.00 (1.41)	2.50 (.71) T	3.00 (2.83)	4.00 (1.41)	3.00 (.00)	3.00 (.00)
	3.00 (1.41)	3.00 (2.83)	2.50 (.71) T	2.50 (.71) T	1.00 (.00) T	3.00 (.00)	3.00 (1.41)	2.50 (.71)
	2.00 (.00)	2.00 (.00)	2.50 (.71) T	2.50 (2.12)	1.00 (.00) T	3.00 (2.83)	2.00 (1.41)	1.50 (.71)
Total	3.57 (1.13)	3.43 (1.51)	2.86 (1.07)	3.43 (1.27)	2.43 (1.90)	3.71 (1.50)	3.00 (1.29)	2.71 (1.25)

**H15: QA systems should make its user confident in the exploration of the available data, and in the report produced as a result.**

Q4: Given you were performing this scenario outside of your standard work environment, without many of your standard resources, how comfortable were you with the process of preparing your deliverable?  
[uncomfortable ... comfortable]

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
	5.00 (.00)	4.00 (.00)	4.50 (.71)	4.00 (.00)	4.00 (.00)	4.00 (.00)	5.00 (.00)	5.00 (.00)
	2.50 (.71) T	3.50 (.71)	4.00 (.00)	3.50 (.71) T	1.00 (.00) T	3.50 (2.12)	3.50 (.71)	3.50 (.71)
	2.50 (.71) T	3.00 (.00)	3.00 (.00) T	3.50 (.71) T	1.00 (.00) T	3.50 (.71)	3.00 (1.41)	3.50 (2.12)
	2.00 (.00)	2.50 (.71)	3.00 (.00) T	2.00 (.00)	2.50 (2.12)	3.00 (.00)	2.50 (.71)	1.50 (.71)
<b>Total</b>	3.14 (1.25)	3.29 (.76)	3.57 (.79)	3.14 (.90)	2.29 (1.60)	3.57 (1.27)	3.29 (1.11)	3.14 (1.57)