REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
30-12-2011	Final Technical Report	01-04-2008 - 30-12-2011

4. TITLE AND SUBTITLE
Investigating Relationships Among Macrocognitive Processes

5a. CONTRACT NUMBER
N00014-08-1-0676

5b. GRANT NUMBER
GRT00012190

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Patterson, Emily S., Woods, David D.

5d. PROJECT NUMBER
60016728

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Ohio State University, B030 Graves Hall, 333 W. 10th Ave., Columbus, OH 43210

8. PERFORMING ORGANIZATION REPORT NUMBER
Final

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research, Code 341 Rm 1051, 875 North Randolph St., Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR/CKI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The primary objectives for this project were: 1) Conduct an independent validation of the macrocognition model developed by researchers in the Collaborative and Knowledge Interoperability (CKI) program, and 2) Examine the reliability and validity of new macrocognition metrics for team collaboration processes. The research was conducted in two phases. Phase I examined the orthogonality of 20 macrocognitive processes with a card sort study with study participants with no prior knowledge vs. participants with expert knowledge. No differences were found for model components, but some differences in distinctions between components were found. Phase II replicated prior CKI findings about model stages with a new task, logistics planning for both face to face and virtual teams supported by audio SKYPE. In addition, the reliability and validity of three new macrocognition metrics are examined. One promising measure of analytic rigor, called the rigor metric, was found to have high inter-rater reliability and face validity. The rigor metric is now in operational use at the National Air and Space Intelligence Center.

15. SUBJECT TERMS
macrocognition, teamwork, collaboration, knowledge

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jeffrey Morrison
unclass	unclass	unclass	Same as Report	56	19b. TELEPHONE NUMBER: 703-696-4875

Standard Form 298

ABSTRACT

This project advances the current state of the science in team collaboration and knowledge interoperability by increasing our basic understanding of how macrocognition in teams is accomplished through a series of nested and interrelated processes. The primary objectives for this project were: 1) Conduct an independent validation of the macrocognition model developed by researchers in the Collaborative and Knowledge Interoperability (CKI) program, and 2) Examine the reliability and validity of new macrocognition metrics for team collaboration processes. The research was conducted in two phases. Phase I examined the orthogonality of 20 macrocognitive processes with a card sort study with study participants with no prior knowledge vs. participants with expert knowledge. No differences were found for model components, but some differences in distinctions between components were found. Phase II replicated prior CKI findings about model stages with a new task, logistics planning for both face to face and virtual teams supported by audio SKYPE. In addition, the reliability and validity of three new macrocognition metrics are examined. One promising measure of analytic rigor, called the rigor metric, was found to have high inter-rater reliability and face validity. The rigor metric is now in operational use at the National Air and Space Intelligence Center.

Investigating Relationships Among Macrocognitive Processes

This project independently investigates the validity of aspects of the macrocognition model previously developed by researchers in the Collaborative and Knowledge Interoperability (CKI) program. Overall, there are three main scientific and technical objectives:

1) an independent validation of aspects of the CKI macrocognition model,
2) identification of new components and distinctions to augment the model, and
3) exploring new macrocognition metrics for team collaboration processes and strategies.

CKI's macrocognition model (Warner, Letsky, and Cowen, 2005) conceptualizes macrocognition as the internalized and externalized high-level mental processes employed by teams to create new knowledge during complex, one-of-a-kind, collaborative problem solving. High-level is defined as the process of combining, visualizing, and aggregating information to resolve ambiguity in support of the discovery of new knowledge and relationships. The methodology is a series of three laboratory experiments examining the validity of different aspects of the macrocognition model.

A series of three laboratory experiments are used to accomplish the objectives.

In the first experiment, a repeated single criterion card sort methodology was employed (Rugg and McGeorge, 1997). Sixteen study participants with no prior knowledge of the CKI model sorted cards representing component processes of the model into related piles. Distinctions identified by study participants generally provide additional warrant for existing distinctions in the model. Additional distinctions were identified, suggesting possible new components and interactions to augment the model. In the second follow-on experiment, study participants with deep knowledge of the CKI model are using the same methodology to investigate how much prior knowledge of the model affects the groupings.

In the third experiment, 12 three-person teams of study participants conducted a challenging, face valid task of optimally moving troops and supporting materials to an attack location securely, economically, and within the least amount of time possible. Six teams were in a face-to-face condition and six were physically distributed with audio platform (SKYPE) support. Verbal transcripts were analyzed for evidence of macrocognition stages found in prior research: Knowledge Construction (KC), Collaborative Team Problem Solving (TPS), Team Consensus (TC), Outcome Evaluation and Revision (OER).

Findings from all three studies validated key aspects of the CKI model. In particular, definitions distinguishing these macrocognition concepts were validated: Individual vs. Team Processes, High vs. Low Dissension, Data vs. Knowledge, and Coordination vs. Collaboration. In Studies 1 and 2, there were no statistically significant differences in the categories of card sorts across naïve and knowledgeable participants. In addition, study 3 provided strong evidence for the existence of the macrocognition stages of the CKI model (knowledge construction, collaborative team problem solving, team consensus, and outcome evaluation and revision). Extensions to the CKI model might include going beyond the focus of knowledge building to incorporate tasks with analysis, planning, and executing a plan. Based on findings across all three studies, follow-on research might examine how macrocognition processes change for conditions with 1) high vs. low knowledge accuracy and 2) high vs. low knowledge specialization.

Card Sorting Macrocognitive Processes: Studies 1 & 2

Studies 1 & 2 Methodology:
The methodologies from studies 1 and 2 are presented together since they represent the same methodology applied to naïve study participants with no knowledge of the CKI model (study 1) and study participants with extensive working knowledge of the CKI model (study 2).

Study 1 Methods: 16 naïve study participants with no knowledge of the CKI model were recruited via an IRB-approved posting on listservs for undergraduate students specializing in homeland security and industrial engineering. A repeated single criterion card sort methodology was employed, where study participants were instructed to group related items together into multiple piles based upon a single overall sort criterion and labels for the criterion and each pile were elicited. Data collection and analysis was completed for all study participants with two text-based sorts conducted in individual 60-minute sessions. Prior to the session or at the end of the session, all the study participants completed an online learning styles questionnaire.

Study 2 Methods: Identical methods as Study 1 with the exception that an exhaustive sample of 5 study participants with extensive knowledge of the CKI model due to conducting research were recruited via professional connections.

Two card sorts were employed in both studies, representing two portions of the CKI model of macrocognition. Identical cards were used in both studies (Figure 1), which had the concept (with no associated label) and an example that was in the context of student teams doing a presentation on a group project in a class for a combined grade.

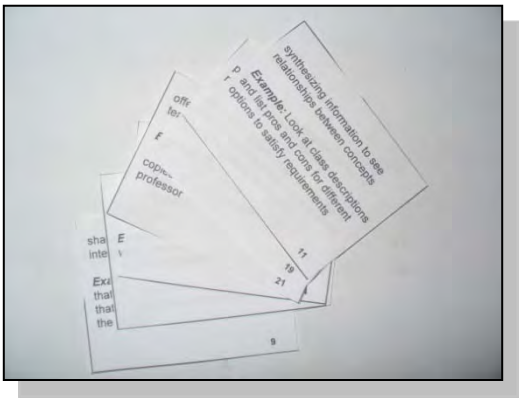

Figure 1. Cards used in Sort #1

In both studies 1 and 2, the following information was on the front of an individual card. The eight concepts in Table 1 and fourteen concepts in Table 2 comprise the entirety of semi-independent sections of the CKI model. These sections were identified by Dr. Mike Letsky and Dr. Norm Warner as most important to study. It was believed that Table 1 had the strongest theoretical foundation, and that there was a possibility that some of the concepts in Table 2 might be combined into a single category. In the pilot runs, examples that are face valid to students were found to be critical for having the naïve study participants understand the concept text.

Table 1. Text for card sort #1

Concept	Example
Acting to add to existing knowledge	Read a book, look at a map
Synthesizing information to see relationships between concepts	Look at class descriptions and list pros and cons for different options to satisfy requirements
Creating diagrams or table	Make a spreadsheet for which classes to take which quarter in order to graduate on time
Passing relevant information to the right person at the right time	A teammate points out that the room that they want to meet in will be locked on Sunday
Sharing explanations and interpretations with the team	A teammate tells the team that the professor emailed him back that they can have an extra day for the project
Offering potential solutions to the team	A teammate suggests going to Kinko's to make color copies of the presentation for the professor
Clarifying and discussing pros and cons of potential solutions	One solution is to go to Kinko's to make color copies of the presentation for the professor, but we have to pay. Another solution is to do it here in black and white, which is quicker and free.
Critiquing the team's process of solution after getting feedback	The team lost 10 points on the grade because they went 10 minutes longer than allotted for their presentation. Everyone agreed that they should have only had one presenter and then have the entire team answer questions.

Table 2. Text for card sort #2

Concept	Example
How much everyone understands their roles and the roles of the others on the team, and how much everyone understands the critical goals and locations of resources	Everyone knows what the homework assignment is, who is supposed to do what, and what the name of the Powerpoint file is for the presentation
How much everyone agrees on procedures and resources to do a team task	All five team members knew that they were going to leave on their cellphones so that they could coordinate while driving two cars to the science fair
How much everyone on a team knows their roles and how to interact with each other	Greta's teammates all knew she had an IPhone that she could use to look up a location on a map while they were driving by typing in the address.
How much everyone on a team agrees on the skill, knowledge, experience, dispositions and/or habits of the others	The team gave Bill the task of performing the statistical analysis for the project because he got an A in statistics.
How much everyone on a team is aware of moment-to-moment changes and agree on what the implications are	The team realized that they could not launch their rocket until the rain stopped
A team's collective understanding of resources and responsibilities associated with a task	Jill was the only one who knew that they had to keep original gas receipts to be reimbursed, but she didn't tell Joe when he filled the tank
Accurate knowledge held by team members that is useful for a task	Jim knew that only four students could fit in each car that the team had.
How much everyone has accurate knowledge of team roles, goals, responsibilities, access to information, constraints, and when to interact with other team members	The team expected Barb to tell Jodi when she was available to meet, so that Jodi could then schedule a room with the department secretary and then tell Tim, the leader, who would let everyone on the team know where and when to meet.
How much everyone has	John knew that Bill used to design websites and

an accurate knowledge of the expertise and behavioral habits of all their team members	is always five minutes late to meetings
How much an individual has an accurate awareness of moment-to-moment changes in the environment	Julia knew that it started raining ten minutes ago
Facts, relationships, and concepts that have been explicitly agreed upon by team members	Everyone on the team agrees that there is 68 miles to drive to the science fair because Joe mapped a route starting from their school to the fair using Google maps
How much everyone agrees on their task strategies and what events should change those strategies	The team agreed that if it rained they would have to wear rain ponchos to the test site.
How accurate patterns and trends identified by team members are	Jill remarked that there are 5 bullets on every slide in the presentation and no one pointed out that actually that was only true for two slides and that, in fact, 3 slides had 3 bullets on them.
How much everyone agrees on the status of a problem	Everyone agrees that heavy rain makes it impossible to launch the rocket

Study 1 & 2 Findings:

Study participants in both study 1 and 2 repeatedly sorted cards into categories which they personally generated until they chose to stop. There were no statistically significant differences in the sort categories used in study 1 and 2 (Table 3).

Table 3. Percentage of naïve and trained participants employing a sort category

Category label	% Naïve Participants	% Trained Participants
Exchanging thoughts and ideas	63	80
Teamwork activities/team working together	56	80
Team agreement on knowledge and information	56	100
Individual activities	50	80
Analysis of information	50	60
Awareness of patterns, trends and environment	50	40
Understanding of how team works	50	20
Passing information without added context	44	80
Making a decision	44	20
Knowing other team members' skill sets and role	38	80
Gaining knowledge	31	80

For each of the sorts, study participants in both study 1 and 2 labeled the distinctions across the categories (see Table 4). Six new distinctions were made by trained participants ($p<0.01$ using Fisher's exact test).

Table 4. Percentage of naïve and trained participants employing a distinction across sorts

Distinction in Relationships Between Piles	% Naïve Participants	% Trained Participants
Team vs. individual	75	80
High vs. low dissension	75	40
Analysis vs. synthesis	69	60
High vs. low knowledge specialization	69	20
Analysis vs. planning vs. acting	56	60
Sharing vs. working	56	40
High vs. low clarity in roles (who does what)	50	40
Early vs. late collaboration stages	50	20
Generating vs. evaluating	25	20
High vs. low information organization	25	0
High vs. low team unity	19	40
High vs. low information accuracy	19	60
Share vs. communicate*	0	60
Task vs. team vs. context*	0	60
Internal vs. external*	0	40
Measures of knowledge*	0	40
Reaction to types of changes*	0	40
Most risky vs. less risky if dropped*	0	40

* statistically significant

In summary, the findings from studies 1 and 2 are:

- No differences for study participants with and without prior knowledge of the CKI model for the labels of the piles (with the possible exception of "gaining knowledge" $p = 0.12$)
- Some differences for study participants with and without prior knowledge of the CKI model for distinctions across the piles:
    - Share vs. communicate
    - Task vs. team vs. context
    - Internal vs. external
    - Measures of knowledge
    - Reaction to types of changes
    - Most risky vs. less risky if dropped

Overall, the 'bottom line' implications from these studies are an independent validation of many of the key CKI model concepts, particularly for distinctions between:

- Individual vs. Team Processes
- High vs. Low Dissension

- Data vs. Knowledge
- Coordination vs. Collaboration

The findings also suggest that there is stronger evidence for a two-level distinction between data and information/knowledge than a three-level distinction between data, information, and knowledge. In particular, it is difficult to distinguish what elements are uniquely at the information level vs. what elements are at the knowledge level.

The findings also indicate that the model might benefit from additional research in order to clarify or extend the model for conditions of:

- High vs. low knowledge accuracy
- High vs. low knowledge specialization

Finally, the findings indicate that expanding the scope beyond knowledge building to include planning and action execution might be warranted based on difficulties uniquely distinguishing knowledge building from these other macrocognitive functions due to being highly interconnected.

Macrocognitive Phases during Logistics Task: Study 3

In summary, the methods for study 3 are:

- Between subjects design
- Study participant teams randomized to condition, teams assembled as first-come, first-assigned
- Condition: Face to face (6 teams) vs. Virtual supported by audioSkype (6 teams)
- Hidden profile task: Information distributed across specialized roles
- Single two-hour data collection session
- Logistics task: Move troops and vehicles from point A to point B
- Location: Georgia, Russia
- Constraints on task: Fastest and less than 2.5 hours, cheapest (least fuel), and acceptable security (avoiding known risks)
- Data collection: Digital audio and video, combined into integrated transcript
- Analysis: Performance, CKI stages, rigor metric

This laboratory study had twelve three-member ad hoc teams perform a logistics planning task in a single session. A between-subjects design randomly assigned six teams to a face-to-face condition and six teams to a distributed condition, where they had to communicate using an audio-only SKYPE program from three different rooms within a building. Study participants were recruited by an IRB-approved posting on a listserv for undergraduate students specializing in homeland security. The study participants were provided monetary compensation for their time. Study participants were assigned to teams based on the order in which they responded, with the exception that none of the team members were allowed to have worked together with our study participants previously in order to simulate an ad hoc team formulated with no prior working experience as a team.

Each team was tasked with the mission to transport troops and cargo to a desired location while optimally satisfying time, cost, and safety constraints. Each participant was given different information critical to task completion. The task was to transport 15,000 kilograms of cargo and 100 troops to the desired location in under 2.5 hours, while also minimizing cost and maximizing security. The team could choose the route and vehicles used in the mission. Each analyst had unique information about the safety, cost, and speed/distance of the vehicles/routes along with added intelligence information. Table 5 below outlines the vehicle information compiled from each analysts' information.

Table 5. Vehicle Information

Vehicle	Number Available	Fuel Consumption	Range	Speed	Security	Capacity
Oktokar Cobra	1	6.8 km/L	752 km	115 km/hr	Low Armor Protection	8 troops
BTR-80	15	1.2 km/L	600 km	80 km/hr	Armored, Difficult to Hide	7 troops or 1,000 kg
Kamaz 4308	6	7.1 km/L	320 km	100 km/hr	No Armor	8 troops or 3,000 kg
Tractor Trailer	2	2.7 km/L	400 km	60 km/hr	No Armor	2 troops and 5,000 kg
Train	1	.28 km/L	350 km	100 km/hr	No Armor	50 troops and 8,000 kg
Mi-8 Helicopter	1	.33 km/L	450 km	250 km/hr	Targetted by Saboteurs	24 troops and 3,000 kg

Table 6 outlines the route information.

Table 6. Route Information

Route	Distance	Security	Condition
A	190 km	No attacks	80
B	150 km	Minor attacks	40
C	150 km	Many attacks	60

The analysts were also given a map, displayed below in Figure 2, which helped the team to visualize the terrain, distance, and security threat differences between the routes to the desired location.

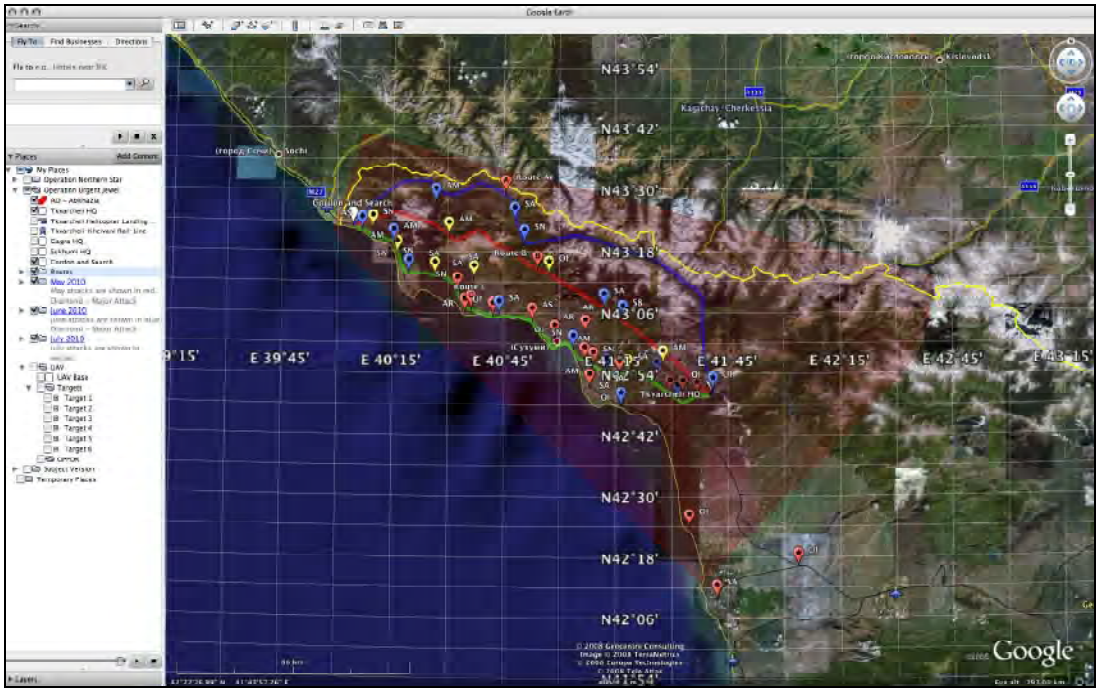

Figure 2. Map with Three Possible Routes

Team Performance Scoring

Each of the twelve teams had 90 minutes to come up with their best solution to the logistical task. After providing IRB-approved consent, the teams were video and audio-taped while working through the task together in a laboratory setting. Written transcripts for each team were compiled into spreadsheets for data analysis.

The teams were scored based on their ability to satisfy the time, cost, and safety constraints outlined in the problem. The following score sheet was used to evaluate the final solution for each team, shown in Table 7. This was also used as the grade of team performance.

Table 7. Team Performance Score Sheet

Reductions	Explanation	Points Lost
Ignore Information Resulting In Impossible Solution	Ignore limitations of vehicles, mistake "or" for "and"	10
- Ignore Weather	Use vehicles off-road that lack those capabilities	2
- Ignore Unavailable Airfield	Use Mi-8 even though the airfield is not operational	5
Unrealistic Solution	Ignore current date and/or plan operation in the future or the past	10
Utilize Resources That Are Not Provided	Utilize weapons systems not provided in story	8
Security Receives Low Consideration	The security of the operation has low priority	7
- Don’t Consider Operations Security	Use train with enemy intelligence agents on board	2
- Don’t Consider Physical Security	Use Route C, the route with the most enemy activity	2
Time Receives Low Consideration	Don't arrive at the target objective within the 2.5 hour deadline	7
Cost Receives Low Consideration	The cost of the operation has low priority	7
Total		60

Each team started with a perfect score of 60, but when constraints were ignored or the objective of the mission not realized, points were deducted from their score.

Study 3 Findings: Overall, the findings (Tables 8-12) validate that the macrocognition stages of the CKI model found in prior research (knowledge construction, collaborative team problem solving, team consensus, and outcome evaluation and revision) apply to this new task and that other stages are not needed (<5% of utterances were coded as not falling into the four stages).

As detailed in Table 8, there were no statistically significant differences ($p = 0.36$ with two-tailed assumption) between solution scores for face to face (average 66.7%) vs. distributed with audio Skype support (average 80.3%) teams.

Table 8. Solution Scores for All Teams

Team	Face-to-Face (F) or Distributed (D)	Solution Score
1	F	44/60 (73%)
2	F	53/60 (88%)
3	D	60/60 (100%)
4	F	31/60 (52%)
5	F	33/60 (55%)
6	F	19/60 (32%)
7	D	24/60 (40%)
8	D	53/60 (88%)
9	D	60/60 (100%)
10	F	60/60 (100%)
11	D	53/60 (88%)
12	D	39/60 (65%)

As shown in Table 9, there were no statistically significant differences (p = 0.62 with two-tailed assumption) between the time to the first utterance for face to face (average 5:20) vs. distributed with audio Skype support (average 4:28) teams.

Table 9. Time until First Utterance

Team	Face-to-Face (F) or Distributed (D)	Time of first utterance
1	F	3:58
2	F	9:30
3	D	6:35
4	F	9:16
5	F	4:50
6	F	0:47
7	D	0:01
8	D	5:04
9	D	6:22
10	F	3:40
11	D	5:11
12	D	3:37

In Table 10, X is defined as the member first to speak following the pause when team members read the provided materials, Y second, and Z last. The probability of the first to speak having more utterances over the session than the last to speak approached statistical significance ($p(X>Z)= 0.08$ using a one-tailed assumption). Nevertheless, with the removal of team 7, which appears to be an outlier with team member Z speaking the least of all the teams, this statistic is not significant ($p(X>Z) = 0.23$ using a one-tailed assumption) and is similar to the statistic for comparing the first and second speakers, which is also not statistically significantly different ($p(X>Y) = 0.28$ using a one-tailed assumption).

Table 10. Distribution of Team Member Utterances

Team	X%	Y%	Z%
1	30.9%	27.6%	41.6%
2	31.8%	26.5%	41.7%
3	19.9%	42.4%	37.7%
4	39.7%	36.7%	23.6%
5	44.4%	23.8%	31.8%
6	37.0%	32.8%	30.1%
7	44.6%	42.8%	12.6%
8	34.7%	37.6%	27.7%
9	34.5%	42.4%	23.0%
10	33.6%	30.6%	35.9%
11	40.8%	23.4%	35.8%
12	33.3%	38.1%	28.6%
Average	35.4%	33.7%	30.8%

The findings detailed in Table 11 are similar to prior studies, and provide additional support that for this new domain, the stages of collaboration of knowledge construction, collaborative team problem solving, team consensus, and outcome evaluation and revision in the macrocognitive model apply. The other category was minimal and the judgment of the coder was that no additional categories were needed beyond these four.

Table 11. Distribution of Collaboration Stages in Macrocognitive Model

Team	KC%	CTPS%	TC%	OER%	Other
1	31.9%	55.9%	9.7%	2.5%	0.0%
2	0.4%	74.9%	17.5%	0.9%	6.3%
3	36.8%	49.3%	4.9%	2.1%	6.9%
4	21.1%	58.8%	4.5%	7.0%	8.5%
5	0.4%	63.9%	27.1%	1.8%	6.9%
6	0.5%	70.1%	24.9%	0.7%	3.7%
7	1.9%	62.5%	24.9%	4.1%	6.7%
8	9.1%	59.1%	26.4%	2.1%	3.3%
9	21.6%	65.5%	2.2%	6.5%	4.3%
10	33.6%	56.8%	1.0%	6.0%	2.7%
11	19.4%	65.7%	4.5%	4.5%	6.0%
12	8.3%	84.4%	1.6%	1.6%	1.6%
Average	15.4%	63.9%	12.4%	3.3%	4.7%

As shown in Table 12, there were no detectable differences between the distribution of stages across the two conditions.

Table 12. Distribution of Collaboration Stages in Macrocognitive Model: Face to face vs. Distributed

Face to Face				
Team	**KC%**	**CTPS%**	**TC%**	**OER%**
1	31.9%	55.9%	9.7%	2.5%
2	0.4%	74.9%	17.5%	0.9%
4	21.1%	58.8%	4.5%	7.0%
5	0.4%	63.9%	27.1%	1.8%
6	0.5%	70.1%	24.9%	0.7%
10	33.6%	56.8%	1.0%	6.0%
Average	14.6%	63.4%	14.1%	3.2%
Distributed				
3	36.8%	49.3%	4.9%	2.1%
7	1.9%	62.5%	24.9%	4.1%
8	9.1%	59.1%	26.4%	2.1%
9	21.6%	65.5%	2.2%	6.5%
11	19.4%	65.7%	4.5%	4.5%
12	8.3%	84.4%	1.6%	1.6%
Average	16.2%	64.4%	10.7%	3.5%

Exploration of New Macrocognition Metrics

One of the objectives of this work is to explore new macrocognition metrics that might be more sensitive, easier to obtain, have a higher inter-rater reliability and/or shed more insight into how to improve process than the current state-of-the-art in manually coding the verbal transcripts with multiple coders. To this end, the following metrics were explored:

1) Process tracing analysis on all macrocognitive events
2) Process tracing analysis on the macrocognition function of 'deciding'
3) Rigor metric for the macrocognition function of 'sensemaking'

1) Process tracing analysis on all macrocognitive events

First, a process tracing analysis (Woods, 1993) was employed to map sequences of macrocognitive events. The hope was that sequenced event maps might be sensitive measures that correlate with team performance scores.

In order to pilot this approach, one integrated theoretical framework for macrocognition was used to provide top-down input on the search for events to include in a process tracing analysis. In particular, five macrocognition functions were previously identified across a wide variety of settings (Klein et al., 2003; Patterson et al., 2010, Patterson et al., 2011, Patterson and Hoffman, 2012):

1) *Detecting:* This is noticing that events may be taking an unexpected (positive or negative) direction. This change requires explanation and might signal a need or opportunity to reframe how a situation is conceptualized (sensemaking) and/or revise ongoing plans (planning) in progress.
2) *Sensemaking:* This is collecting, corroborating, and assembling information and assessing how the information maps onto potential explanations. This includes generating new potential hypotheses to consider and revisiting previously discarded hypotheses in the face of new evidence.
3) *Planning:* This is adaptively responding to changes in objectives from supervisors and peers, obstacles, opportunities, events, or changes in predicted future trajectories. When ready-to-hand default plans are applicable, there is still a need to adapt a prespecified plan into actions within a window of opportunity. When ready-to-hand default plans are not applicable to the situation, this can include creating a new strategy for achieving one or more goals or desired end states. This function includes adapting procedures, based on possibly incomplete guidance, to an evolving situation where multiple procedures need to be coordinated, procedures which have been started may not always be completed, or when steps in a procedure may occur out of sequence or interact with other actions. Executing a plan is never distinguished from replanning, even when the individual or team that generates a plan is different from the individuals or teams who perform the actions to execute it (Klein 2007a and 2007b).

4) *Deciding:* This is committing to one or more course of action options. The commitment may constrain the ability to reverse courses of action. This function is inherently a continuous process conducted under time pressure. It involves re-examining embedded default decisions in ongoing plan trajectories for the predicted impact on meeting objectives, including whether to sacrifice decisions to which agents were previously committed based on considering trade-offs. This function may involve a single individual or might require consensus across distributed actors with different stances towards decisions. This function is far more complex than classical discussions of decision-making, including increased uncertainty about when a decision can be modified, the level of commitment to a future planned action, distributed perspectives with associated goal trade-off tendencies negotiating an agreement, and temporal dynamics, including rallying points, changes in the ability to modify an action, and impacts of changes to plans of other stakeholders (see Hoffman and Yates, 2005).
5) *Coordinating:* This is managing interdependencies of activity and communication across individuals acting in roles that have common, overlapping or interacting (and possibly conflicting) goals.

In Figure 3, the relationship of the macrocognition functions is illustrated.

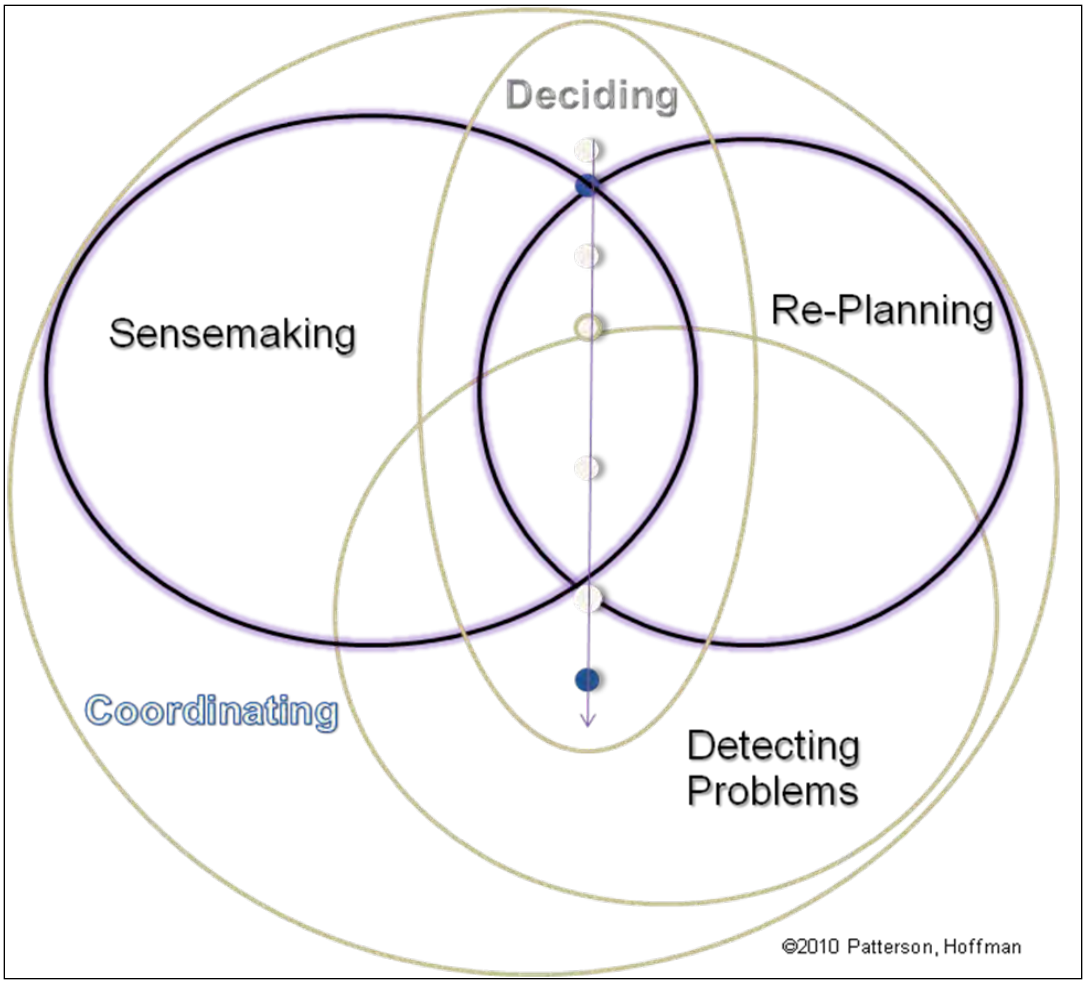

Figure 3. Integrated theoretical framework of macrocognition functions

In parallel, one coder tagged all events which he judged to be “non-routine” bottom-up next to the relevant portion on every transcribed session. This approach was inspired by the use of non-routine events as a more sensitive measure than human errors. In the research by Weinger et al. (2002) in the medical field, a Non-Routine Event (NRE) was defined as “any aspect of care perceived by clinicians or observers as a deviation from optimal care based on the context of the clinical situation.” Figure 4 shows a portion of the figure that was used to illustrate this definition of a NRE. A NRE needed an intervention to realign with the optimal care path. A NRE only led to an adverse event if an intervention was not made. By tracking the more frequent NREs, a more robust systems understanding of failure modes could be developed to drive quality improvements for patient experience. In addition, it was developed into a predictive measure for the patient risk during a procedure.

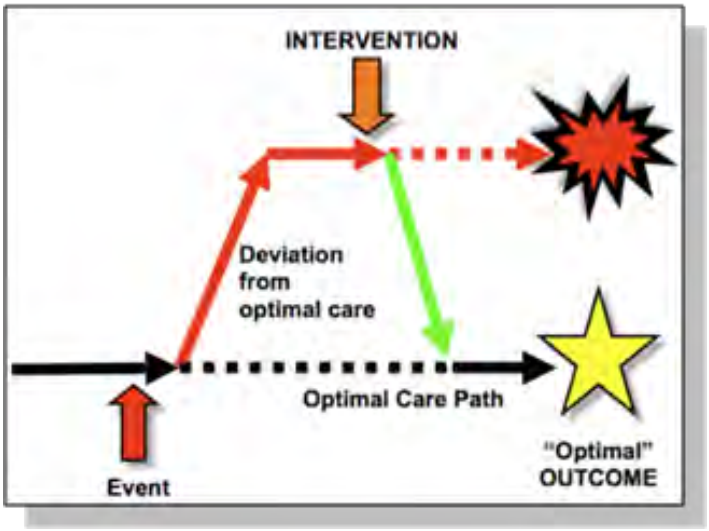

Figure 4. Non-Routine Event (adapted from Weinger et al., 2002)

The macrocognition function of Detecting was judged not to have played an important role in the laboratory study, since the team members were not provided with any real-time updates to provided data, such as would occur with telemetry or sensor data or satellite images. Therefore, only four of the five macrocognition functions were selected for inclusion, and the leading pattern identified from the non-routine event tagging that mapped for the macrocognitive function was used in the event mapping. The definitions for the macrocognitive events are provided in Table 13.

Table 13. Definition of Coded Macrocognitive Events

Macrocognitive Event	Macrocognitive Function	Definition
Assuming	Sensemaking	Adding constraints not explicitly outlined in the problem
Eliminating	(Re)planning	Eliminating a potential option to simplify the final decision
Delaying Commitment	Deciding	Delaying final commitment in favor of further analysis
Dismissing	Coordinating	"In-group vs. Out-group" events

A more detailed explanation of each macrocognitive event, along with an example from the transcripts is provided in Figures 5-8.

Assuming: (A)

Macrocognitive Function: Sensemaking

Description: The Assuming event occurs when a teammate uses creativity or intuition to add a characteristic or complexity to the problem that is not explicitly stated in the problem description given. Any mention of the following items is considered assuming:

- Health insurance
- Battles/Fighting
- Protection of vehicles or troops
- Disguising or splitting up troops or vehicles
- Rerouting vehicles
- Adjusted speeds of vehicles
- Delaying time to start task

Example: There is no information given about the ability of vehicles or troops to fight. This is simply a logistical task where the assumption of a battle is out of scope.

Analyst	Statement
Z	Is that going to be enough man power to fight. Should they fall under attack?
X	That's just the initial... I would propose mixing the otocar cobras and the kamaz.

Figure 5. Definition and example of assuming

Eliminating: (E)

Macrocognitive Function: (Re)planning

Description: The Eliminating event occurs when the team has agreed to remove one of the potential solutions (vehicle and/or route combination) from the problem because of its inability to satisfy one of the objectives of lowest cost, most safety, and least time.

Example: The train is eliminated from consideration because of both high cost and low amount of safety because of enemy agents on rail road.

Analyst	Statement
X	I assume we don't want to use the train because we know that there are intelligence agents working on the rail road.
Y	And the train will also cost. We also have to take in consideration the fuel. That's going to cost a lot for fuel because it is only 0.28 km/l.
X	Right, but I think we should consider fuel last as a consideration considering that our objective is to get the mission accomplished if need we'll have to pay extra for fuel. They have to dig deeper in their pockets. We need to get our people there safely and all of the equipment to support them. But yeah, the train gives horrible gas millage too. Although it's fast and can carry a lot of troops and cargo. Any significant amount of use on that will definitely rise my eyebrows.
X	So I think we should cross the train out of the list at this point.

Figure 6. Definition and example of eliminating

Delaying Commitment: (DC)

Macrocognitive Function: Deciding

Description: The Delaying Commitment event occurs when one or more of the teammates attempts to stop rushed decisions or guesses of potential solutions, and encourages further unbiased analysis.

Example: Analyst Z is attempting to encourage the team to halt making decisions until all the intelligence of each analyst has been revealed.

Analyst	Statement
Z	Yeah, let's hold on. We need to keep combining our intelligence, because you have something different than I do. I have fuel consumption.

Figure 7. Definition and example of delaying commitment

Dismissing: (D)

Macrocognitive Function: Coordinating

Description: The Dismissing event occurs when one of the teammates cuts off another teammate mid sentence (expressed in transcript as "…") or rudely dismisses their input/contribution to the team.

Example: The questions asked by analyst Z are regularly dismissed and never answered by X and Y who are having their own discussion.

Analyst	Statement
Z	Is the train out? Because it has to go on C, which is dangerous? Can it handle security?
Y	We only have 1 COBRA.
X	Since we have 1 COBRA, might as well use it for 8 troops.
Z	How much will that cost us?
Y	What route are we going to send that on?
Z	Not for 2 more days. Can't go off-road in wet weather.
Y	It can go A or C.
Y	I thought A's the longest but still safe.

Figure 8. Definition and example of dismissing

Macrocognitive Events

The results of the number of macrocognitive events identified for each team are displayed in Table 14.

Table 14. Frequency of Macrocognitive Events By Team

Team	Assuming	Eliminating	Delaying Commitment	Dismissing	Total
1	3	2	0	0	5
2	1	2	1	1	5
3	2	9	3	0	14
4	9	3	0	0	12
5	4	5	2	1	12
6	2	4	1	1	8
7	1	7	0	0	8
8	6	3	0	0	9
9	0	7	1	2	10
10	0	3	0	2	5
11	0	5	1	0	6
12	1	3	2	1	7
Total	29	53	11	8	101

In Figure 9, the macrocognitive event maps for each team are visually represented. The events are labeled: Assuming (A), Dismissing (D), Delaying Commitment (DC), and Eliminating (E). The length of the horizontal line indicates the time the team spent on the task, and the team's performance score is displayed at the far right of the map.

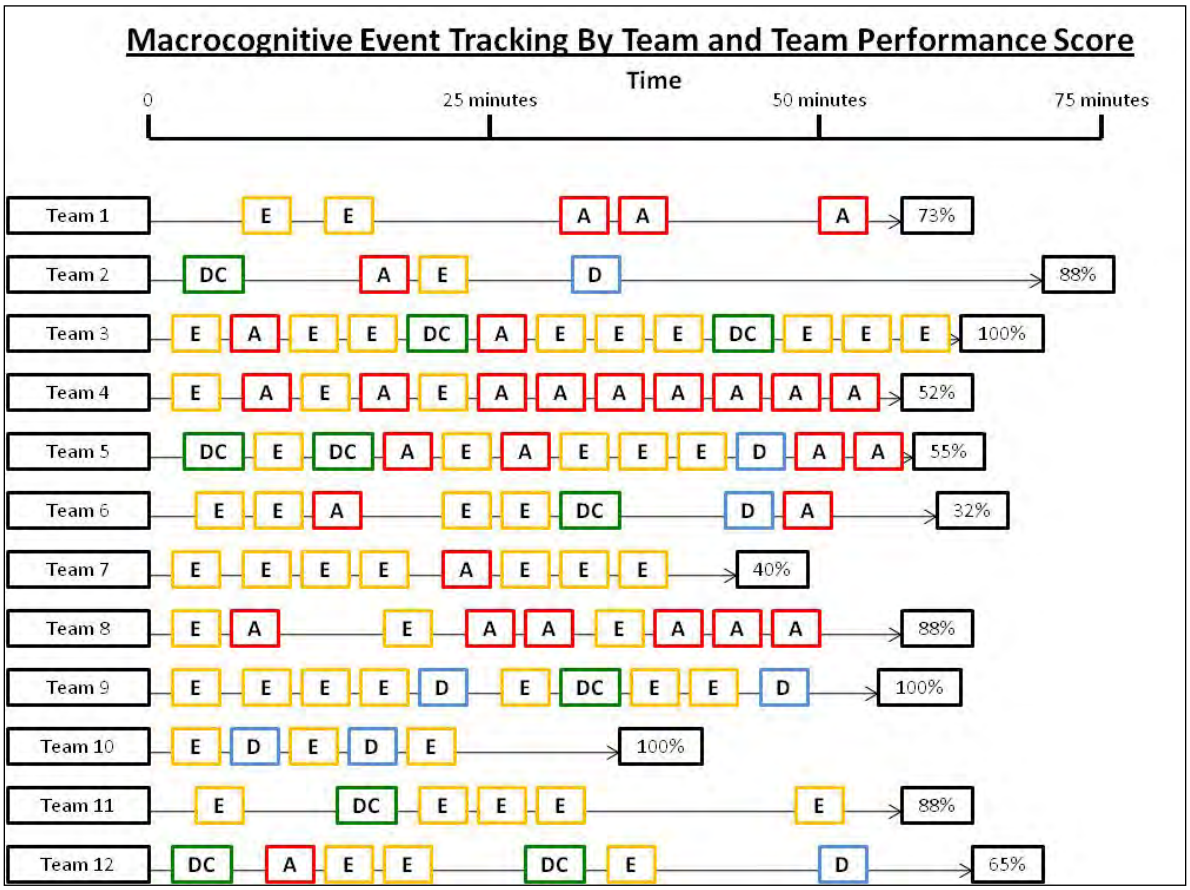

Figure 9: Macrocognitive Event Maps for Each Team

Inter-rater reliability

Figure 9 displays the codes determined by the first coder, who defined each code. Inter-rater reliability across two coders suggested only moderate agreement (Landis & Koch, 1977) on the codes ($\kappa_w = 0.57$). Analysis revealed that 40% of the differences in coding were accounted for by disagreement between the Assuming and Eliminating events. Without recoding, combining the Assuming and Eliminating codes increased the inter-rater reliability score to an acceptable level ($\kappa_w = 0.68$). In future work, it is suggested that the parsing strategy be done differently for Sensemaking and (Re)planning functions. There are also possibly theoretical implications stemming from the challenges in parsing reliably for these functions. In particular, it might not be warrnated to

define Sensemaking and Replanning as semi-independent constructs. In the transcript data, teams were often gathering information in parallel with formulating potential options, providing evidence that these functions are harder to separate than the other functions.

Correlating Macrocognitive Events with Team Performance

A primary purpose of generating macrocognitive event maps was to explore how well these maps correlated with team performance outcome data. To this end, linear regression analysis was performed on the macrocognitive events and performance scores.

The inputs for the most accurate linear regression model were:

- Assuming event frequency
- Dismissing event frequency
- Additional factor: Assuming * Dismissing

The output analyzed was team performance score. The complete model is shown below in Figure 10.

Regression Analysis: Team Performance versus Assuming, Dismissing, ...

The regression equation is
Team Performance = 0.764 - 0.0113 Assuming + 0.102 Dismissing
- 0.105 Assuming*Dismissing

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.7636	0.1333	5.73	0.000	
Assuming	-0.01135	0.02908	-0.39	0.707	1.374
Dismissing	0.1016	0.1057	0.96	0.365	1.461
Assuming*Dismissing	-0.10464	0.05875	-1.78	0.113	1.127

S = 0.225887 R-Sq = 36.8% R-Sq(adj) = 13.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.23769	0.07923	1.55	0.275
Residual Error	8	0.40820	0.05102		
Total	11	0.64589			

Figure 10. Linear regression model

The regression equation highlighted at the top of the figure with a rectangle was used to calculate the predicted team performance score based on the values of the inputs. The three red squares highlight important indicators of model accuracy. A trustworthy model has Variance Inflation Factors (VIF) less than 10 and an adjusted R-Squared > 0.7. The best possible model for this research had

VIFs less than 10, but only an adjusted R-Squared of 13.1%. Therefore, this correlation is not judged to be strong enough to be a trusted model.

In summary, using macrocognitive events for measurement was not a promising direction. An acceptable level of inter-rater reliability was eventually achieved. Nevertheless, parsing sensemaking and replanning macrocognitive functions was identified to be a challenge and the event were not strongly correlated with team performance, either negatively or positively. Interestingly, macrocognitive events such as dismissing that were originally assumed to be negative events were sometimes correlated with high performance scores. Therefore, an area for future exploration is whether these events could be either positively or negatively correlated with performance based on context. For example, dismissing might be correlated with better performance when there is a team member with limited cognitive abilities (see Team 10 in Figure 9). Alternatively, assuming might be negatively correlated with performance if there are multiple instances, but not if it is an infrequent event (Teams 4 and 8 vs. Teams 3, 6, and 12). Finally, the study was a between subjects design, which can increase variation. Either a within subjects design or having more teams in the study might have allowed more factors to be included in the model, and thus resulting in a better correlation.

2) Process tracing analysis on the macrocognition function of 'deciding'

Since the macrocognitive event maps did not yield promising findings, we re-conceptualized process tracing efforts based upon focusing solely on the macrocognition function of deciding. The study task was purposely simplified to eliminate the macrocognition function of detecting events and coordinating outside the 3-person team. Therefore, the primary aspects of decision making, or specifically the macrocognition function of deciding, related to the problem space for optimizing the movement of troops and supplies. Three elements emerged from the iterative analysis as important elements of decision-making in this task:

- R = Rule-out: Eliminate a portion of the solution space from consideration. Example: "I think the train should be out."
- A = Add factor: Introduce a consideration to factor into the problem solving process that was not in the original framing of the problem (which was primarily information about routes and vehicles that had implications for safety, cost, and fuel efficiency). Example: "That would be an excuse for Russian intervention."
- D = Delay commitment: Delay committing to a solution or locking in a portion of the solution at that moment. Example: "Yeah, let's hold on. We need to keep combining our intelligence, because you have something different than I do."

106 elements were uniquely coded as rule-out, add factor, or delay commitment independently by two investigators with high inter-rater reliability ($\kappa_w = 0.854$), as detailed in Table 15.

Table 15. Deciding macrocogitive events: Rule-out, Add factor, Delay Commitment

No.	Code	Time	Team	Transcript
1	Rule-out	14:15	1	Because of the reason that it has been wet weather BTR80 it can be the only one who can actually complete the course, because of its off road capabilities.
2	Rule-out	23:52	1	If you want to go A, you cannot use the tractor trailer at all because it is only 60 km/h it cannot make it on time.
3	Add factor	39:55	1	Health insurance is more expensive.
4	Delay	15:24	1	I think we should summarize what's going so we're all on the same page and know what's going on.
5	Add factor	41:03	1	There is a battle going on, it would take more time to reach…this is the slowest tractor trailer at 60 km/h and it will only be on time in 2.5 hours. But what if there is some attacks from terrorists? We have to stop there for a while. It is not like 10 seconds or 15 seconds and miss several hours.
6	Add factor	55:22	1	Depending on how BTR can protect kamaz and tractor trailer. It all depends on that.
7	Add factor	22:06	2	As far as strategic methods go, it would be beneficial to split them up as your chances don't become all or nothing. I think this is a limiting factor.
8	Rule-out	25:07	2	Capacity wise it holds 24 troops and 3000 kg of cargo and has a speed of 250 kms/hr but that's out. We can't use it.
9	Rule-out	53:20	2	If the KAMAZ cannot off road we can't send it on B, which means we can't use it.
10	Rule-out	13:40	3	So, when it comes to safety, I think that Route C, even though it's the fastest, I think that that's just right out.
11	Add factor	13:56	3	If we can somehow disguise all of our military things with a civilian train or something, maybe... I mean it's a risk, we might be able to make it seem like it's a normal train ride, but in reality it's some of our equipment.
12	Rule-out	17:05	3	The repairs are expected to be completed on August 6, which is two days, so we can't use helicopters in two days.
13	Delay	21:29	3	Yeah, let's hold on. We need to keep combining our intelligence, because you have something different than i do. I have fuel consumption.
14	Add factor	30:11	3	That would be an excuse for Russian intervention too.
15	Rule-out	30:13	3	Let's abort going into Russia even though is still in the Area of Operations.
16	Rule-out	33:38	3	Yeah, I think the train should be out as well.
17	Rule-out	34:39	3	Route C is probably not a good way to go. If we use anything on route C it should be really fast. Not a tractor trailer because that is so easy to hit.
18	Rule-out	45:30	3	So we need to rule out the tractor trailer for route A. It would be able to make it on time on other routes but not on route A unfortunately.
19	Rule-out	46:34	3	I prefer not to use route C because of the high enemy activity. Basically if we get hit it's all over, right?
20	Rule-out	47:15	3	Can we rule the otokar cobra out? Just because it does not really help us?
21	Delay	48:05	3	I think so. Can we go through all of our options? What about the Kamaz? What's the downside of the kamaz?
22	Delay	1:00:02	3	So here is the thing, so basically what we are doing here is that we are working on time, cost and security. We are doing everything in time but they seem to be putting more weight on security than the cost? Do you think that's a good thing to do?
23	Add factor	26:02	4	I think we need to give or take a little bit because of the terrain and because of the weather.

Table 15 (Cont.) Deciding macrocogitive events: Rule-out, Add factor, Delay Commitment

24	Rule-out	28:24	4	So I think we should cross the train out of the list at this point.
25	Rule-out	28:30	4	And I would say because of the situation with the helicopters. I would say that we could just ignore those for the time being, if we have to make adjustments, we could do so accordingly.
26	Add factor	30:25	4	I don't know how many people I want to really send down route B just because its proximity to the mountain ranges as well, and the fact that it is off road, the chance that they may hit a mud hole that blocks the entire road.
27	Add factor	31:14	4	It looks like we have to send the BTR along route C if we do this because we could send a couple down route C but it is the most, correct me if I'm wrong, but it's the heaviest armored of the vehicles, which could be the one that could withstand.
28	Add factor	32:10	4	Since the BTR is faster than the tractor trailer we could send one down the road first just to scout, that way we would only be risking 1 vehicle, send the tractor trailer behind that, follow it with more BTR's and send a portion of those BTR down route B too, and just load it with troops and equipment
29	Add factor	33:19	4	Is that going to be enough man power to fight, should they fall under attack?
30	Add factor	33:56	4	Plus we also have this helicopters that they have been targeted by saboteurs, but if we are rolling down route C we can fly out over the ocean and run support there and since they are moving at 250 km/h and they got a range of 450 km they can support us twice over on both route B and C.
31	Add factor	39:20	4	I just think that we don't need to send out unnecessary vehicles. We also have to think about gas. We can't just throw helicopters just to get us covered if they are not really doing anything. You know our vehicles have their armor and we have the people with then. So the last thing we need is spending unnecessary fuel.
32	Rule-out	45:38	4	The helicopter can carry 24 troops a piece, but they have also been targeted by saboteurs so…we are ruling those out to get troops or cargo because they have been targeted so they may crash, they may not but if they are going to crash I rather not have troops and cargo for the search coordinates in there.
33	Add factor	46:34	4	We don't know, we don't have information on what weapons it has, as long as we know, they don't have any weapons.
34	Add factor	47:06	4	Can we split it in 2 helicopters in case we lose 1? Because that way we would at least have one helicopter to fly over each group.
35	Delay	9:44	5	Should we rank them, as far as what's best with security, fuel?
36	Delay	12:40	5	We have to get there in less than 2.5 hours. Maybe it'll help if we go through and list speed and security for all vehicles.
37	Add factor	15:22	5	Can we send out troops disguised?...like civilians?
38	Rule-out	16:48	5	We can't take the tractor on A because it wouldn't get there in time.
39	Add factor	27:48	5	I think it'd be more conspicuous if you had both, or would it be more conspicuous with cargo and civilians?
40	Rule-out	28:10	5	I don't like C because it's going through all these cities.
41	Rule-out	29:58	5	I feel like we can't use the train at all, because it says that any train activity running N-S will be under enemy observation.
42	Rule-out	30:34	5	We can't take F because a VBIED was detonated on the landing pad and they have to use an airfield outside of the AO until August 6th.
43	Add factor	45:04	5	Can they walk?
44	Rule-out	05:44	6	That takes out helos.
45	Add factor	07:00	6	The thing with the weather: would that not help us? If it's raining both parties are going to be at a disadvantage.
46	Rule-out	18:08	6	We wouldn't be able to use a helo today or tomorrow anyway, right?

Table 15 (Cont.) Deciding macrocogitive events: Rule-out, Add factor, Delay Commitment

47	Add factor	42:47	6	The weather in the AO has been very wet, raining 3 of the last 5 days. Rain is forecast 24-48 hours. If we wait for the 6, we're taking a guess.
47.5	Delay	06:00	7	[I'm thinking that maybe Route A with the train would be a good idea….] Let's think: what else are possibilities route wise?
48	Rule-out	08:15	7	We couldn't use the KAMAZ, the tractor trailer, or the train, because they have no armor.
49	Rule-out	09:06	7	So I say right now that Route C wouldn't be a good idea anyway.
50	Add factor	21:38	7	[rainy day…] That wouldn't necessarily be a problem..could kinda be an advantage for us since it's not going to affect our travel but could affect the enemies' ability to find us.
51	Rule-out	25:54	7	The only thing that can move on B are the BTR-80s, so B's done except for the BTR-80s.
52	Rule-out	26:20	7	The KAMAZ, trailer, and train have no armor. We don't want to throw any of those on Route C.
53	Ru9le-out	07:11	8	Because of the COBRA's capacity, I don't think it would be a viable option.
54	Add factor	08:01	8	That might be a possibility then. If you can load it from the side, to have the troops positioned on the sides of the helo.
55	Add factor	25:40	8	I don't know how we intend to hide 100 people and all their equipment moving in a convoy.
56	Add factor	30:34	8	Question. It's rainy and we're going through rugged, unpaved terrain in mountains. Can we assume it's going to go at its top speed?
57	Rule-out	41:28	8	So I guess A for the train line is out of the question.
58	Add factor	47:15	8	My only concern with putting it all in the most heavily armored vehicles available, if anybody sees us at any point, they could call their leaders and they could bolt.
59	Add factor	48:19	8	If we did use the helos, that could be done at night. That'd give us cover in darkness.
60	Add factor	51:10	8	I'd like to have a healthy mix of troops and supplies in whatever convoy. I like A for its hybrid virtues.
61	Rule-out	8:24	9	I'd rule out the green out
62	Add factor	8:32	9	It has way too many towns so it means the level of security is going to be low. They may see troops coming.
63	Rule-out	8:38	9	So we can't use helicopters then.
64	Rule-out	9:49	9	So that rules out route C, it's the most dangerous.
65	Rule-out	9:51	9	Route A, I think we have to avoid using the train system.
66	Rule-out	9:53	9	Route B, we can't use the helicopters.
67	Rule-out	10:04	9	And route B we can only use vehicle BTR80
68	Add factor	10:11	9	[BTR80] Which is armored and difficult to hide.
69	Rule-out	23:06	9	I think we should eliminate route B
70	Add factor	23:12	9	Route B is not good in this weather
71	Rule-out	24:09	9	[helicopter traffic has been routed out of our AO until August 6] Screws that plan then
72	Rule-out	24:12	9	We can't use the train for route A, E or D on route A, so that leaves us with the first 4 vehicles.

Table 15 (Cont.) Deciding macrocogitive events: Rule-out, Add factor, Delay Commitment

73	Rule-out	28:08	9	That won't work. If we send everything on C, they'll get there but not in one piece.
74	Add factor	30:05	9	I would say speed could overcome it [light armor protection]
75	Add factor	30:50	9	There's some spies, but if they don't see troops, I don't see the problem, they will only see boxes.
76	Delay	35:00	9	Let's keep that on the table
77	Add factor	38:40	9	You want to armor the troops as much as possible, the cargo you don't need to armor as much.
78	Rule-out	40:30	9	If we can avoid route C at all costs, it is a good idea to me.
79	Rule-out	0:30	10	So we are crossing out C.
80	Rule-out	10:27	10	[we have 2.5 hours] No, but that's 3 hours for 180 right? 60 and 60 and 60 is 180. Tractor trailer moves…60 km/h
81	Rule-out	11:44	10	We can't get that one on route B though,
82	Rule-out	12:06	10	I think that we should leave the cargo on the BTR 80 for sure, because it's not efficient to put the troops on the BTR 80
83	Delay	13:02	10	Let's experiment with people and see what happens.
84	Delay	15:26	10	So at this point we can solve the time issue, the people and cargo issue but probably not the gas issue.
85	Rule-out	33:12	11	I think C should be out
86	Add factor	34:39	11	The train is likely to be seen and shot down by missiles.
87	Add factor	34:59	11	And all of them are planning on attacking the helicopter anyway
88	Rule-out	36:16	11	So train and helicopter is out
89	Rule-out	38:02	11	Yeah, so it has to be route A
90	Add factor	44:20	11	They won't see troops, but still if they see a bunch of boxes.
91	Rule-out	53:20	11	Let's just x the train out.
92	Delay	7:41	12	Maybe we should go through like just say what we have so that we can write it down. I'll start
93	Add factor	18:20	12	If we did decide to go with the train, one of the major concerns is all the insurgents have to do is bomb the train and we would be done with. We don't have time to rebuild railroads.
94	Rule-out	19:00	12	Let's stay clear of the train
95	Delay	19:20	12	I think we could b able to allocate our resources in different. We don't want to use all of the same kind. But we want to use the best, so not to necessarily rule it out, just to be aware of it.
96	Add factor	26:10	12	So I say we might want to split up the routes because if one route gets attacked a lot then we should...
97	Delay	26:25	12	So I guess our best bet is to look at each route, start with route A and see which vehicle we can use for it and which you can't
98	Rule-out	29:30	12	route A no train
99	Rule-out	29:33	12	So do you think the tractor trailer is too big for route A. probably?
100	Rule-out	30:36	12	We definitely want to use the BTR for C
101	Delay	30:50	12	So, **IF** we want to use C use the BTR 80.

Table 15 (Cont.) Deciding macrocogitive events: Rule-out, Add factor, Delay Commitment

102	Rule-out	30:56	12	But maybe just the BTR. if there is a lot of attacks, anything that is not armored is going to be extremely vulnerable.
103	Rule-out	31:36	12	I'm gonna say the train is probably not the best idea
104	Rule-out	31:48	12	I just don't like route C, because it says that the attacks have been increasing in severity and there are also 30 friendly troops that were killed, so if they see this kind of activity on route C. of course we are going to be attacked. I just don't think that would be a safe route to choose.
105	Delay	33:30	12	X: Actually, I don't think there is any way you can do this without using the train. There is no way you can do it without using the train. Y: Well, let's see. If you use all of the BTR can carry 105 troops total.
106	Add factor	46:45	12	Another thing you want to think about is that you kind of want all your troops arriving at the same time.

Process trace maps were constructed and are displayed in Figures 11 and 12. The only marginally statistically significant difference between the teams was that rule-out was conducted more in the virtual, distributed teams (p = 0.065).

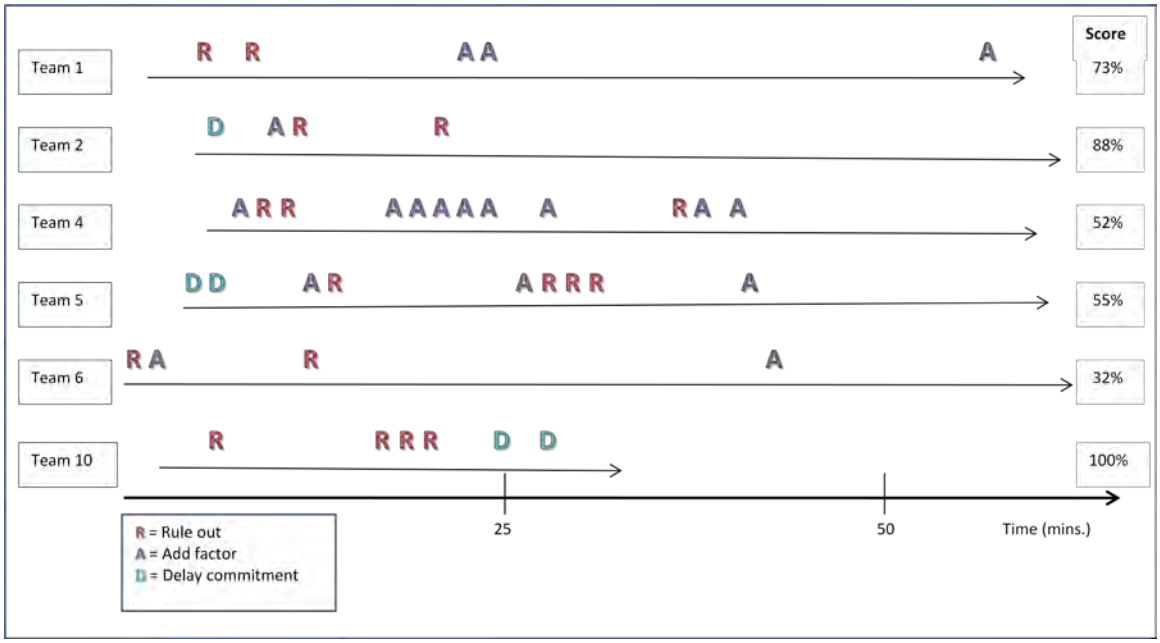

Figure 11. Process Trace for Deciding Macrocognition Function for Face to Face Teams

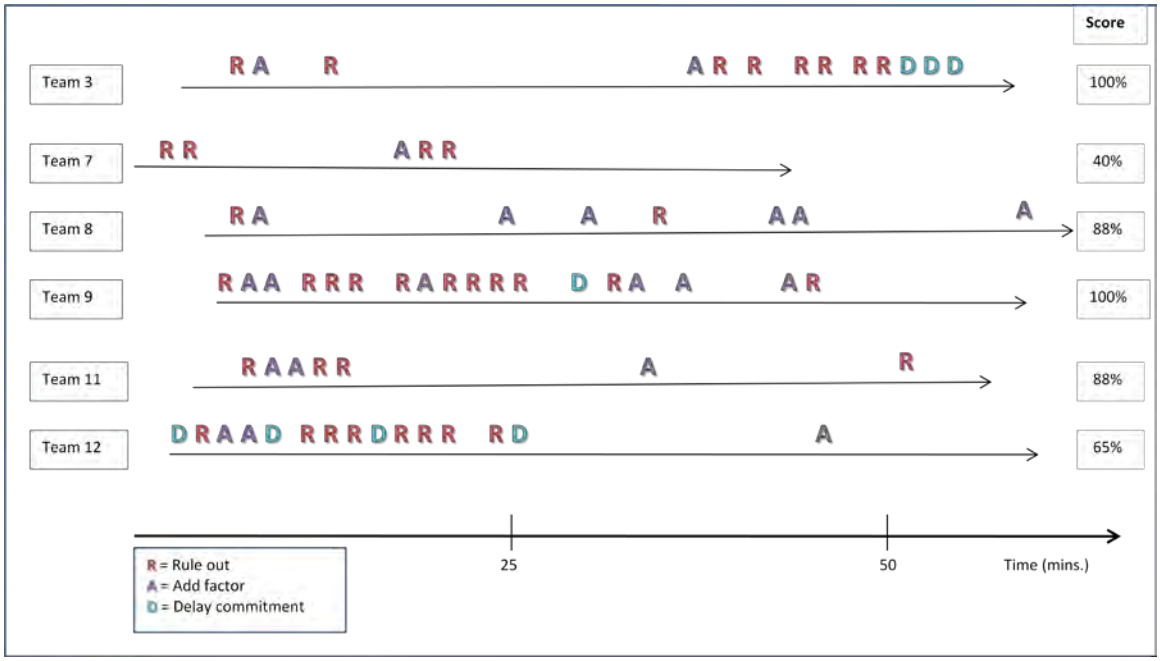

Figure 12. Process Trace for Deciding Macrocognition Function for Virtual Teams

3) Rigor metric for the macrocognition function of 'sensemaking'

The final macrocognitive metric to be explored was a measure of analytic rigor, the so-called 'rigor metric' (Zelik, Patterson, and Woods 2010). As previously stated, one of the objectives of this work was to explore new macrocognition metrics that might be more sensitive, easier to obtain, have a higher inter-rater reliability and/or shed more insight into how to improve process than the current state-of-the-art in manually coding the verbal transcripts with multiple coders.

The rigor metric was designed to assess the macrocognition function of 'sensemaking'. This measurement approach is contrasted with conventional perspectives for measuring analytical rigor which tend to identify defects in analysis as compared to a prescribed methodlogy. For example, Crippen et al (2005, p. 188) defines rigor as "scrupulous adherence to established standards", the Military Operations Research society (2006, p.4) as "application of precise and exacting standards" and Morse (2004, p.501) as "methodological standards for qualitative inquiry."

Unfortunately, such definitions suggest a conceptualization of analytical activity that is neither particularly likely to reflect rigorous analysis work as practiced in a team setting based upon macrocognition (Dekker, 2005; Sandelowski, 1993, 1986). Consequently, rather than on a standards-based notion of rigor, our measurement approach focuses on how the risk of shallow analysis is reduced via analyst-initiated strategies that are opportunistically employed throughout the analysis process. These strategies are alternatively conceptualized as "broadening" checks (Elm et al., 2005) insofar as they tend to slow the production of analytic product and make explicit the sacrifice of efficiency in pursuit of accuracy, a central tenet of the framework.

The rigor metric is therefore oriented around detecting generic strategies employed to increase warrant in an analytic conclusion opportunistically at any point during a free-flowing process of 'making sense' of the interactions of agents in a complex environment. The eight attributes of the rigor metric are organized around eight inter-related risks from having a shallow analysis process that:

1) Is structured centrally around an inaccurate, incomplete, or otherwise weak primary hypothesis, which analysts sometimes described as favoring a "pet hypothesis" or as a "fixation" on an initial explanation for available data.
2) Is based on an unrepresentative sample of source material, e.g. due to a "shallow search", or completed with a poor understanding of how the sampled information relates to the larger scope of potentially relevant data, e.g. described as a "stab in the dark".
3) Relies on inaccurate source material, as a result of "poor vetting" for example, or treats information stemming from the same original source as if it stems from independent sources, labeled variously as "circular reporting", "creeping validity", or as the "echo chamber" effect.

4) Relies heavily on sources that have only a partial or, in the extreme, an intentionally deceptive stance toward an issue or recommended action, often characterized by analysts in terms of “biased”, “slanted”, “polarized”, or “politicized” source material.
5) Depends critically on a small number of individual pieces of often highly uncertain supporting evidence proving accurate, identified by some individuals as an analysis heavily dependent upon “hinge evidence” or, more generically, as a “house of cards” analysis.
6) Contains portions that contradict or are otherwise incompatible with other portions, e.g. via the inclusion of lists or excerpts directly “cut and paste” from other documents or via an assessment that breaks an issue into parts without effectively re-integrating those parts.
7) Does not incorporate relevant specialized expertise, e.g. an analyst who “goes it alone”, or, in the other extreme, one who over relies on the perspectives of domain experts.
8) Contains weaknesses or logical fallacies in reasoning from data to conclusion, alternatively described as having a “thin argument”, a “poor logic chain”, or as involving “cherry picking” of evidence.

In addressing each of these sources of risk, eight corresponding attributes of analytical rigor comprise the metric (see Figure 13), with each attribute categorized into low, moderate, and high levels.

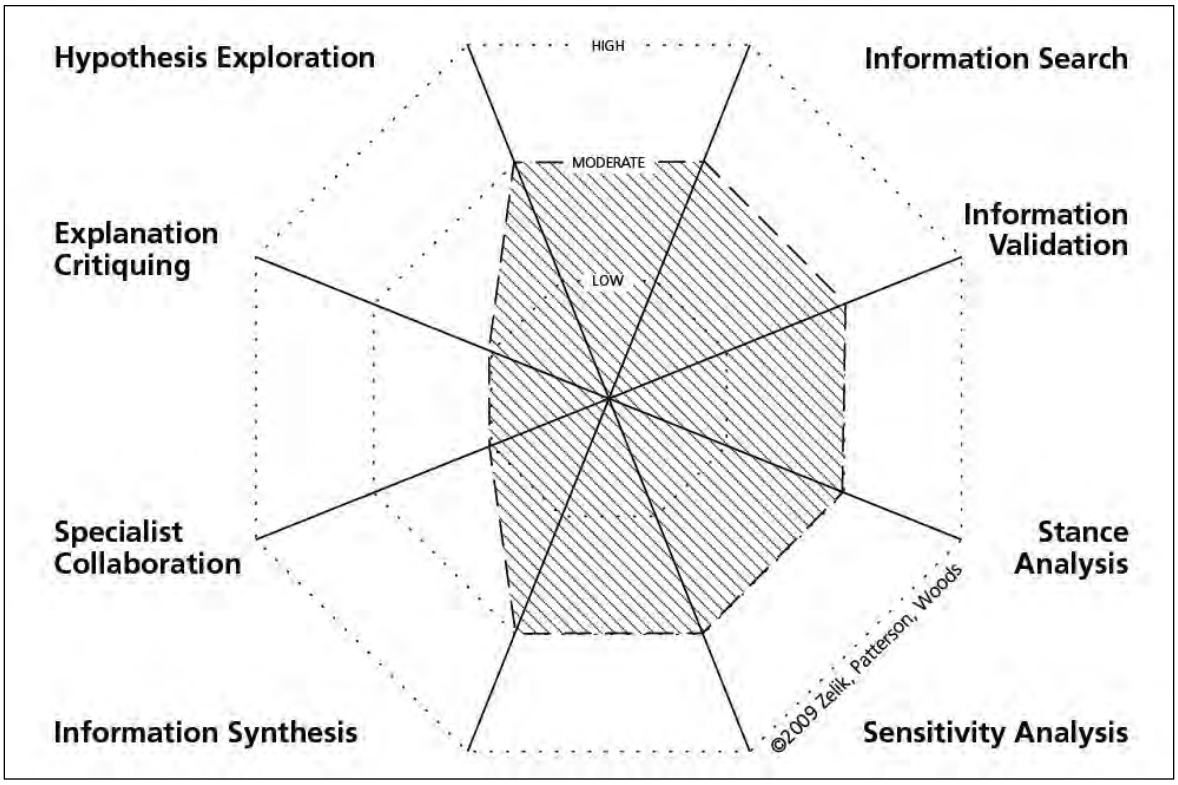

Figure 13. Rigor Metric

The eight attributes of the rigor metric are:

1) *Hypothesis Exploration*. Hypotheses are among the most basic building blocks of analytical work, representing candidate explanations for available data. For this attribute, rigorous analysis is identified by the depth and breadth of the generation and consideration of alternatives, by the incorporation of diverse perspectives in brainstorming hypotheses, by the evolution of thinking beyond an initial problem framing, and by the ongoing openness to the potential for revision.
2) *Information search*. Similarly viewed as a fundamental component of analysis work, this attribute h encompasses all activities performed to gather task-relevant evidence—including those to broaden as well as deepen, those that are active as well as passive, and those hypothesis-driven as well as data-driven. Note that this framing of information search reflects the diverse nature of analytical activity and emphasizes the fact that, for the professional analyst, supporting evidence comes in many forms, and not simply as raw data. Information search is primarily concerned with where and how analysts look for supporting information. A strong information search process is characterized by the extensive exploration of relevant data, by the collection of data from multiple source types, and, most critically, by an active approach to information collection. A weak information search in contrast is identified by failure to go beyond routine and readily available data sources, by reliance on a single source type or on “distant” data that is removed from original source material, and by passive dependence upon “pushed” rather than actively collected data.
3) *Information Validation*. This attribute is concerned with the critical evaluation of data and with determining the level of agreement and disagreement among sources. In rigorous analysis, analysts make an explicit effort to distinguish fact from judgment and are concerned with consistency and credibility among, as well as within, sources. Thus, a strong validation process involves assessing the reliability of sources, assessing the appropriateness of sources relative to the task question, and the use of proximate sources whenever possible. It also involves an explicit effort to seek out multiple, independent sources of converging evidence for key findings. In contrast, weak information validation is reflected in the uncritical acceptance of data at face value, little or no clear effort to establish underlying veracity, and a failure to collect independent supporting evidence. Poor tracking and citation of original sources also identify such analyses. Between strong and weak characterizations, a moderate validation process involves the recognition of inconsistencies among sources and, often times, involves the use of heuristics to support judgments of source integrity—such as deference to sources that have previously proven highly reliable and avoidance of those that have not. On the aggregate, then, information validation can be described as an intense concern with issues of agreement, consistency, and reliability with respect to the set of collected data.

4) *Stance analysis*. Stance" refers to the perspective of a source on a given issue and often it is characterized informally in terms of slant, bias, or predisposition. Stance analysis refers to the evaluation of information with the goal of identifying the positions of sources with respect to a broader contextual understanding and in relation to alternative perspectives on an issue. A process in which little attention is paid to issues of stance reflects weak analysis. In such instances, the analysis may identify heavily slanted sources or sources that support a well-defined position on an issue but yet reflect little in the way of a nuanced understanding. A somewhat better stance analysis would incorporate basic strategies for considering the perspectives of different sources. For example, dividing evidence into camps that are "for" or "against" an issue represents a simplifying heuristic for organizing and making sense of various stances on that issue. A significantly stronger stance analysis involves research into, or leverages a preexisting knowledge of, the backgrounds and views of key individuals, groups, and thought leaders. Where appropriate, it may also include a more formal assessment that employs structured methods to identify critical relationships, to predict how the general worldview of a source is likely to influence his or her stance toward specific issues, or to detect the intentional manipulation of information.
5) *Sensitivity analysis*. The term "sensitivity", as it is used here, has a meaning most similar to its usage in the statistical analysis of quantitative variables, wherein it describes the extent to which changes in input parameters affect the output solution of a model. However, rather than with the relationship between output and input variance, our concern is with the strength of an analytical assessment given the potential for low reliability and high uncertainty in supporting evidence and explanations. Phrased differently, sensitivity analysis describes the process of discovering the underlying assumptions, limitations, and scope of an analysis as a whole, rather than those of the supporting data in particular, as with the related attribute of information validation. Many in the intelligence community emphasize the importance of examining analytical assumptions. To that end, a strong sensitivity analysis goes beyond simple identification, meticulously considering the strength of explanations and assessments in the event that individual supporting evidence or hypotheses were to prove invalid. It also specifies the boundaries of applicability for the analysis. With weak sensitivity analysis, in contrast, explanations seem appropriate or valid at surface level, with little consideration of critical "what if" questions—e.g., "What if a key data source misidentified a person of interest?" Likewise, the overall scope of a weak analysis process may be unclear or undefined.
6) *Information synthesis*. Often emphasized by experts more than casual analysts is that rigorous analytical work is as much about putting concepts together as it is about breaking an issue apart. That is to say, rigorous analysis demands not only "analytic" activity in the definitional sense, but "synthetic" activity as well. Thus, information synthesis is a reflection of the

extent to which an analysis goes beyond simply collecting and listing data to provide insights not directly available in individual source data. Weak information synthesis is reflected in analyses that succeed in compiling relevant and “on topic” information, but that do little in the way of identifying changes from historical trends or providing guidance for broader or more long-term concerns. Indicators of weak synthesis include extensive use of lists, copying material from other sources with little reinterpretation, and a lack of selectivity in what is emphasized by the analysis. A stronger synthesis is reflected by explicit efforts to develop an analysis within a broader framework of understanding. The depiction of events in relation to historical or theoretical context and the framing of key issues in terms of tradeoff dimensions and interactions also identify such analysis. Stronger still is synthesis that has integrated information in terms of relationships rather than components, with a thorough consideration of diverse interpretations of relevant data. In addition, such synthesis is performed by reflexive analysts who are attentive to ways in which their particular analytical processes may hinder effective synthesis and who are attuned to the many potential “cognitive biases” that manifest in analytical work.

7) *Specialist collaboration*. Inevitably, analysts encounter topics on which they are not expert or that require multiple areas of expertise to fully make sense of. Even in instances where an analyst has expertise in pertinent topics, success for the modern analyst still demands the incorporation of multiple perspectives on an issue. Accordingly, analytical rigor is enhanced when substantive expertise is brought to bear on an issue. The level of effort expended to incorporate relevant expertise defines effective specialist collaboration. In a process with little collaboration, minimal outside expertise is sought out directly. A moderately collaborative analysis process involves some interaction with experts, though at this level such expertise is often drawn from existing personal or professional networks, rather than from organizationally external sources. In a high-rigor process, independent experts in key content areas are identified and consulted. Thus, a strong specialist collaboration process is defined by efforts to go beyond a “core network” of contacts in seeking out domain-relevant expertise. In many cases, additional resources and “political capital” are expended to gain access to such specialized knowledge.
8) *Explanation critiquing*. Specialist collaboration and explanation critiquing are related in that both are forms of collaborative analytical activity that reflect the influence of diverse perspectives. However, whereas specialist collaboration primarily relates to the integration of perspectives relative to information search and validation, explanation critiquing relates to the integration of perspectives relative to hypothesis exploration and information synthesis. More succinctly, explanation critiquing is concerned with the evaluation of the overall analytical reasoning process, rather than with the evaluation of content specifically. Similar to specialist collaboration however, this attribute is largely defined by the extent to

which analysts reach beyond immediate contacts in collecting and integrating alternative critiques. A low quality explanation critiquing process has limited instances of such integration, while a more moderate process leverages personal and professional contacts to examine analytical reasoning. In the latter case, it is often peer analysts, supervisors, or managers who serve as the primary source of these alternative critiques. In a still stronger analysis process, independent as well as familiar reviewers have examined the chain of analytical reasoning and explicitly identified which inferences are stronger and which are weaker.

Using the verbal transcripts and a description of the eight rigor metric attributes, two coders independently holistically coded the entire process employed by the team to come up with the final plan. The results are detailed in Table 16, only displaying the final codes after the coders resolved all disagreements. Note that one of the dimensions did not apply to this study, specialist collaboration, which might suggest a limitation regarding the face validity of the task. Note that both coders were provided the solution scores of the teams prior to coding, which likely influenced the ratings, but also likely reduced variation since both of the coders were provided the same information.

Of the 84 items scored (12 teams x 7 scored attributes), 76 were judged by both raters to fall into the same low, moderate, or high category, implying strong agreement between raters ($\kappa_w = 0.86$). Overall there was general consistency in how the coders applied the framework to assess the analytical rigor of the processes employed by the teams. Disagreements for the initial codes, which were subsequently resolved by discussion, were on:

- Information search (3), because one coder included sharing information among the team and deeply processing information under this whereas the other coder did not agree
- Hypothesis exploration (2), because one coder based it more on process and the other coder based it more on how well the team did on the task,
- Information validation (1), because one coder gave a higher score for the team taking notes whereas the other disagreed that this increased this attribute, and
- Stance analysis (1), because one coder defined stance as including revisiting things that were previously closed and the other coder disagreed that this was related to this attribute

Table 16. Detailed Justification for Team Process Rigor Attribute Ratings

Team	Hypothesis Exploration	Information Search	Information Validation	Stance Analysis	Sensitivity Analysis	Information Synthesis	Explanation Critiquing
1	Low: Did not consider security (route C) or cost constraints	Medium: Shared info	Low: Small checking of calculations	Low: No evidence of consideration	Low: Asked about % chance to rain and accepted certainty in scenario	Low: No evidence of consideration	Low: Some embedded checking remarks, but little explicit or overall
2	Medium: Cost had low priority	Medium: Figured out who had what kind of info but didn't give everything to everyone immediately	Low: Comment to assume raining	Medium: Decision to choose route B was done without strong preconceptions	Low: No evidence of consideration	Low: No evidence of consideration	Medium: Invited input from all: "Everybody just give a minute of what they think we should do."
3	High	High: Slow to share detailed info with each other, but when did, then "word for word" and verifications	Medium: Sensitive to risk: "Our intelligence might be an instance of just us have incomplete intelligence which it always happens. because the train track branches two ways"	Medium: They didn't seem to push an agenda, but they settled on B as important and ruled C out fairly quickly	Medium: Commented that calculations were suspect	Medium: Talked about high-level goals before diving into the weeds; generally high insight during discussions	High: Error checking as a group along the way

Table 16 (Continued). Detailed Justification for Team Process Rigor Attribute Ratings

Team	Hypothesis Exploration	Information Search	Information Validation	Stance Analysis	Sensitivity Analysis	Information Synthesis	Explanation Critiquing
4	Medium: Cost had low priority	High: Once they realized that they all had partial intel, which was immediate, they made sure in a systematic way that all the data was shared	Medium: Checked that everyone had the same cost information and there were a few confirmatory “did you say X?”, said to “give or take a little” on calculations due to weather	Low: Route C pushed too hard from the beginning; generally “positions” given strongly early on without inviting comment and flexibility; tone was “I’m in charge and we have to make decisions quickly”	Low: No checks to see if wrong and leader dismissed requests to check when made by others	Medium: Thinks to “give or take a little” on calculations due to weather and tries to get the “gist” of what is going on rather than do math calculations	Low: When Analyst Z was involved in the conversation, her contributions were ignored; the leader was dominant and didn’t like to admit mistakes
5	Medium: Cost had low priority	Medium: Ignored limitations of vehicles, and there was more information pull than push	Medium: Started off verifying objectives are the same on sheets, Date wrong	Medium: There was no one strong stance going in, but they talked all around all of the constraints without really thinking through considerations	Low: No evidence of consideration	Low: Wanted to stay within the rules like <2.5 hours and if the instructions did not explicitly say cannot do it, then did it	Low: Ignored limitations of vehicles, mistake "or" for "and", Ignored current date and plan operation in future or past, rushed to find an answer that didn’t break the rules and then stop
6	Medium: Didn’t consider security (route C) or cost constraints, but everyone did their own plan first	Medium: Mix of push and pull, non-systematic	Low: Enemy agents known, but ignored, Date wrong	Medium: Split up for 5 mins to devise own plans; didn’t seem to be paying much attention to multiple constraints or biases	Low: No evidence of consideration	Low: Didn’t meet deadline of <2.5 hours	Low: Ignored limitations of vehicles, mistake "or" for "and", Ignored current date and planned operation in future or past

Table 16 (Continued). Detailed Justification for Team Process Rigor Attribute Ratings

Team	Hypothesis Exploration	Information Search	Information Validation	Stance Analysis	Sensitivity Analysis	Information Synthesis	Explanation Critiquing
7	Medium: Did not consider time or cost constraints and missed enemy agents, but didn't jump on first solution	Medium: Seemed like one person did not contribute much, perhaps due to poor English; person finally says: what do you have?	Low: Asked to repeat what was just said but seemed to miss constraints	Medium: Compared multiple options	Low: "I think we're doing the best in terms of cost and mileage." was said without any justification provided	Low: Did not meet deadline of <2.5 hours	Medium: "That's the safest bet, but using the train and BTR-80s has to cost a lot of money."
8	Medium: Time had low priority	High: Shared information early and well, even though much was "pulled," it was persistent to get a lot	Low: "I think I might've done the calcs wrong for C." was said but no one encouraged checking it or trying themselves	Medium: Didn't seem to really explicitly compare options, but sort of talking through a wide turf of thoughts	Low: little interest in this kind of thing	High: Knowledge beyond what was written brought into the thinking patterns as "intent" not just meeting constraints (two kinds of security to think about here); Our mission, as a team, is to capture the insrugent leaders.	Medium: Some; Why not just one convoy? This is an all or nothing mission.
9	High: seemed to work through everything pretty well	Medium: Noticed different information quickly, nothing particularly systematic in sharing but not bad either	Low: no obvious consideration	Low: Tried to eliminate constraints with poorly supported justifications	Low: No evidence of consideration	Low: Just "get it done" as ordered mindset	Medium: 16 rows of just to be sure critiquing after said that they were pretty much done and in agreement and had yes yes yes approach

Table 16 (Continued). Detailed Justification for Team Process Rigor Attribute Ratings

Team	Hypothesis Exploration	Information Search	Information Validation	Stance Analysis	Sensitivity Analysis	Information Synthesis	Explanation Critiquing
10	Medium: quickly "ruled out" things that weren't ideal	Medium: People did not volunteer info quickly but seemed to get there eventually	Medium: Sensitivity that calculations could be wrong;	Medium: Some confusion from math uncertainty but no strong stances noted	Low: Only worried about math calculations	Low: Just "get it done" as ordered; "wish we had a budget"	Low: Only supportive comments without basis given "I think that's probably the best way to go'
11	Medium: Cost low consideration	Medium: Shared reasonably well	Low: No evidence of consideration	Medium: Discussion of use of train highly influenced by prior decision to use other routes	Low: No evidence of consideration	Low: No evidence of consideration	Medium: Realized enemy agents were the wrong plan; started caring about cost and then dropped it
12	Low: Did not consider time or cost constraints	Medium: "Sorry I have that, it is part of my information, I keep forgetting that we don't have the same thing."	Medium: Tried to clear up confusion on "and/or"	Low: Seemed to accept statements like "so we can try the helicopter" and then accept something that contradicted it just as easily	Low: No evidence of consideration	Low: Didn't meet deadline of <2.5 hours	Low: No one pointed out that statements were contradictory, blanket agreement with leader on whatever he said

A verbal summary of where the two raters disagreed on the ratings is provided in Figure 14.

Attribute	Team 1	2	3	4	5	6	7	8	9	10	11	12
Hypothesis Exploration	Low	Moderate	**High**	Moderate / Low	Moderate	Moderate	Moderate / Low	Moderate	**High**	Moderate	Moderate	Low
Information Search	Moderate / Low	Moderate	**High**	**High**	Moderate	Moderate	Moderate	**High**	Moderate / **High**	Moderate	Moderate / **High**	Moderate
Information Validation	Low	Low	Moderate / **High**	Moderate / **High**	Moderate	Low	Low	Low	Low	Moderate	Low	Moderate
Stance Analysis	Low	Moderate	Moderate / **High**	Low	Moderate	Moderate	Moderate	Moderate	Low	Moderate	Moderate	Low
Sensitivity Analysis	Low	Low	Moderate	Low	Low	Low	Low	Low	Low	Low	Low	Low
Information Synthesis	Low	Low	Moderate	Moderate	Low	Low	Low	**High**	Low	Low	Low	Low
Specialist Collaboration	—	—	—	—	—	—	—	—	—	—	—	—
Explanation Critiquing	Low	Moderate	**High**	Low	Low	Low	Moderate	Moderate	Moderate	Low	Moderate	Low
Solution Score (out of 60)	44/60	53/60	60/60	31/60	33/60	19/60	24/60	53/60	60/60	60/60	53/60	39/60

Figure 14. Differences in rigor measure ratings for two independent reviewers

No statistically significant differences were found for face to face vs. distributed teams on the rigor measures, as displayed in Figures 15 and 16.

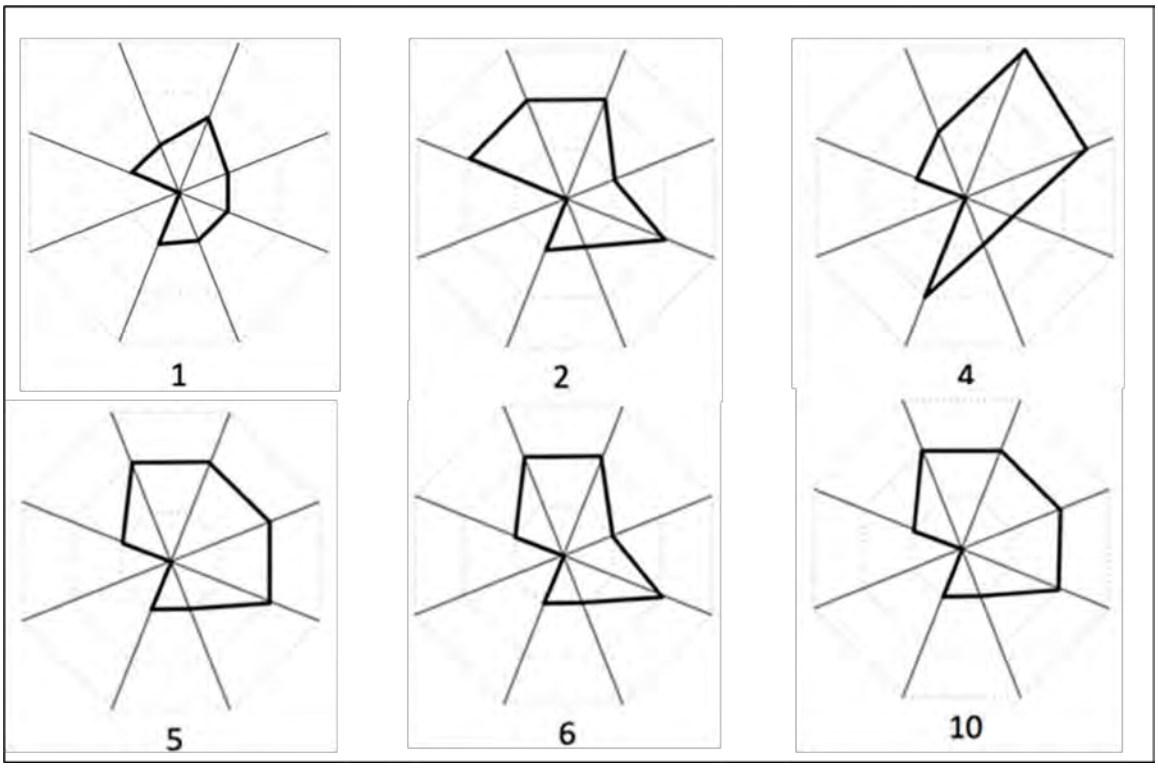

Figure 15. Rigor scores for face to face teams

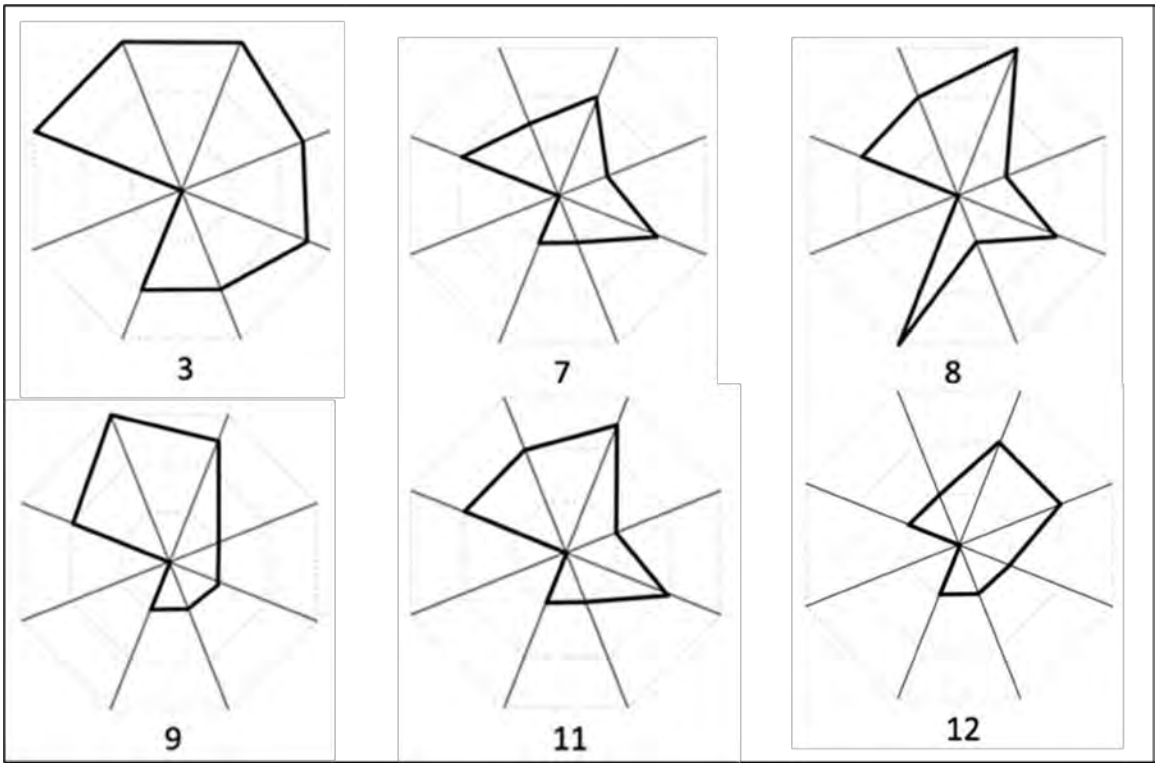

Figure 16. Rigor scores for virtual teams

In Figure 17, the rigor scores from Figures 15 and 16 are combined into a composite graph where the higher the graphed point is, the more teams got scores on the peak. It is anticipated that composite patterns across teams can begin to show weaknesses in populations or across types of tasks that are difficult to quickly visualize with individual rigor figures. For example, information search had the most “high” ratings across all of the teams, with sensitivity analysis having nearly all “low” ratings, and only one “moderate” rating.

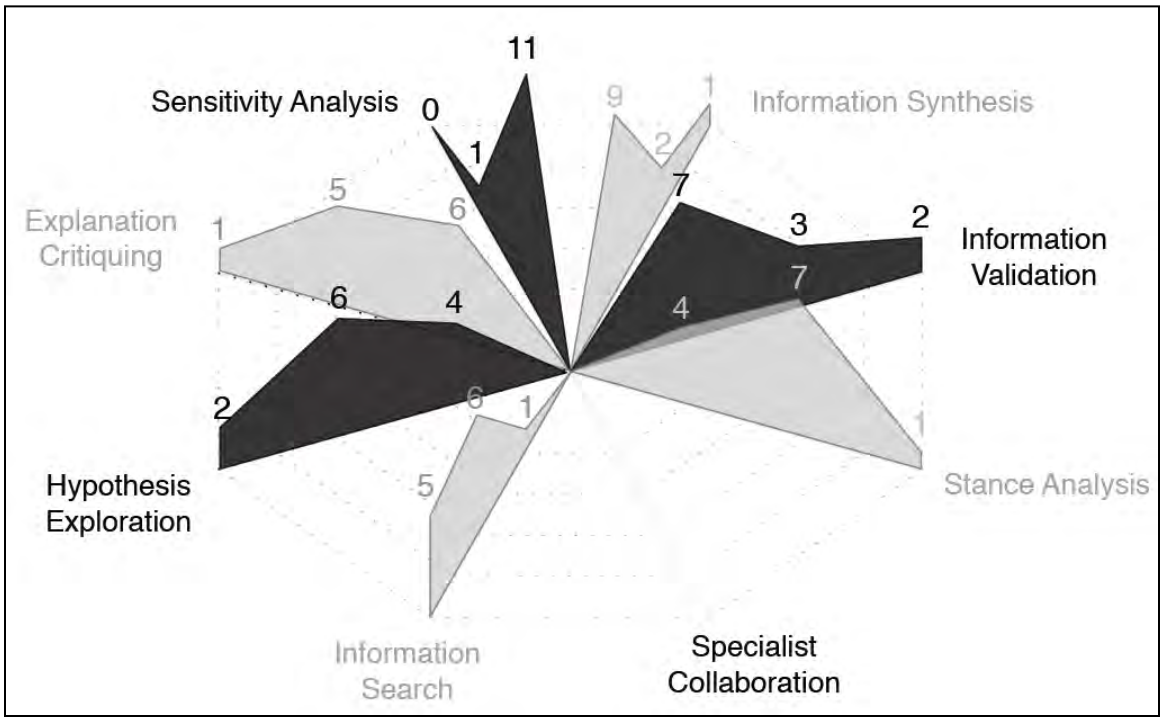

Figure 17. Composite rigor display for all teams

In summary, this effort laid a firmer foundation for macrocognitive research. This project conducted an independent validation of distinctions embedded in the macrocognition model developed by researchers in the Collaborative and Knowledge Interoperability (CKI) program, provided additional verification of macrocognitive stages in the CKI model for a new, face valid task of logistics planning in both face to face and virtual configurations, and examined the reliability and validity of three new approaches to measurement of macrocognition. A rigor metric was found to have high inter-rater reliability and face validity. The rigor metric has already transitioned into practice. It is currently in operational use at the National Air and Space Intelligence Center for annual performance evaluations for all intelligence analysts.

References

Crippen, D. L., Daniel, C. C., Donahue, A. K., Helms, S. J., Livingstone, S. M., O'Leary, R., & Wegner, W. (2005). Annex A.2: Observations by Dr. Dan L. Crippen, Dr. Charles C. Daniel, Dr. Amy K. Donahue, Col. Susan J. Helms, Ms. Susan Morrisey Livingstone, Dr. Rosemary O'Leary, and Mr. William Wegner. In Return to Flight Task Group (Eds.), Return to Flight Task Group final report (pp. 188–207). Washington, DC: U.S. Government Printing Office.

Cummins D.D., Dominance Hierarchies and the Evolution of Human Reasoning. Minds and Machines, Volume 6, Number 4, November 1996, pp. 463-480(18).

Dekker, S. (2005). Ten questions about human error: A new view of human factors and system safety. Mahwah, NJ: Lawrence Erlbaum.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., & Patterson, E. (2005). Finding decision support requirements for effective intelligence analysis tools. In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (pp. 297–301). Santa Monica, CA: Human Factors and Ergonomics Society.

Fiore, S. M., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky, M. (2010). Toward an Understanding of Macrocognition in Teams: Developing and Defining Complex Collaborative Processes and Products. Theoretical Issues in Ergonomic Science , 11(4):250-271.

Fiore, S. M., Rosen, M., Salas, E., Burke, S., & Jentsch, F. (2008). Processes in Complex Team Problem Solving: Parsing and Defining the Theoretical Problem Space. In M. Letsky, N. Warner, S. M. Fiore, & C. Smith (Eds.), Macrocognition in Teams: Theories and Methodologies (pp. 143 – 163). London: Ashgate Publishers.

Hollnagel, E. & Woods, D.D. (2005). Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. Taylor & Francis.

Hoffman, R. R. & Yates, J. F. (July/August, 2005). Decision(?)-Making(?). IEEE: Intelligent Systems, pp. 22-29.

Klein, G. A., Ross, K. G., Moon, B. M., Klein D. E., Hoffman, R. R., & Hollnagel E. (2003). Macrocognition, IEEE Intelligent Systems, 18, 81-85.

Klein, G. (2007a). Flexecution as a paradigm for replanning, Part 1: IEEE Intelligent Systems, 22, 79-83.

Klein, G. (2007b). Flexecution, Part 2: Understanding and Supporting Flexible Execution: IEEE Intelligent Systems, 22, 108-112.

Landis, J.R. and Koch, G. G. (1977) "The measurement of observer agreement for categorical data" in Biometrics. Vol. 33, 159-174.

Letsky, M. P., Warner, N. W., Fiore, S. M., & Smith, C. A. P. (Eds.). (2009). Macrocognition in teams: Theories and methodologies. Burlington, VT: Ashgate Publishing Company.

Military Operations Research Society. (2006, October 3). MORS special meeting terms of reference. Paper presented at Bringing Analytical Rigor to Joint Warfighting Experimentation, Norfolk, VA. Retrieved December 11, 2008, from http://www.mors.org/meetings/bar/tor.htm.

Morse, Janice M. (2004). Preparing and Evaluating Qualitative Research Proposals. In C. Seale, G. Gobo, J. Giubrium, & D. Silverman, (Eds.) Qualitative Research Practice (pp. 493–503). London: SAGE.

Patterson ES, Hoffman R. (2012, forthcoming). Visualization Framework of Macrocognition Functions. Cognition, Technology, and Work: Special issue on Adapting to Change and Uncertainty.

Patterson ES, Miller JE, Roth EM, Woods DD (2010). Macrocognition: Where Do We Stand? In E.S. Patterson and J. Miller (eds.) Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams. Ashgate Publishing. ISBN 978-0-7546-7578-5.

Patterson ES, Stephens R. (2011) A Cognitive Systems Engineering Perspective on Shared Cognition: Coping with Complexity. In Salas E, Fiore S, Letsky M (Eds). Theories of Team Cognition: Cross-Disciplinary Perspectives. Taylor & Francis.

Rugg, G., McGeorge, P. (1997). The sorting techniques: Card sorts, picture sorts and item sorts. Expert Systems, 14(2):80-93.

Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. Advances in Nursing Science, 16(2), 1–8.

Warner, N., Letsky, M., & Cowen, M. (2005). Cognitive model of team collaboration: Macro-cognitive focus. Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomic Society. Santa Monica, CA: Human Factors and Ergonomics Society.

Weinger MB, Slagle J. Human Factors Research in Anesthesia Patient Safety: Techniques to Elucidate Factors Affecting Clinical Task Performance and Decision Making. JAMIA 2002; 9 (suppl 6):S58-63.

Wood, D.D. (1993) “Process-Tracing methods for the study of cognition outside the experimental psychology laboratory”, in Klien, G., Orasanu, J., Colderwood, R. and Zsambok, E. (eds.), Decision Making in Action: Models and Methods. Norwood, pp. 228-251.

Zelik D, Patterson ES, Woods DD. 2010. Measuring Attributes of Rigor in Information Analysis. In Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams. Edited by Patterson ES, Miller J. Hampshire, UK: Ashgate Publishing. 65-84.