No more ‘soft’ science quantitative evaluation of collaboration effectiveness

Nathan Bos
Johns Hopkins Applied Physics Lab
Nathan.bos@jhuapl.edu
nathanbos@gmail.com

The 2009 International Symposium on Collaborative Technologies and Systems (CTS 2009)

‘Soft’ science doesn’t mean soft methods

Use established social science methods to test subjective findings

- Observational
- Quasi-experimental
- Experimental

- Opinion
- Aggregated opinions
- Multiple data sources triangulated

Randomness Certainty
Outline

- Macro-level methods for organizations and larger groups
  - Interview research and grounded theory
  - Survey research including scale development
  - Social network analysis of collaboration and coauthorship
  - Emerging methods: large scale text analysis with LIWC, LSA
- Group level experiments.
  - Game theory tasks for measuring trust
  - ‘Hidden profile’ tasks for measuring information sharing
  - Coordination/ cooperation tasks
  - Research tasks for measuring influence, negotiation, effectiveness, and cultural contrasts
  - New experimental methods: use of software agents as confederates; virtual worlds and MMOG’s for large scale distributed information sharing
- Micro-level methods of individuals in collaboration
  - Survey measures of workload, stress, and situational awareness
  - Brain science, social cognition and the ‘theory of mind’
  - Biological markers of trust

Survey research

- Observation
  - Random
  - Aggregated opinions
  - Multiple data sources triangulated
- Quasi-experimental
- Experimental

Certainty
Survey research

- Use grounded theory unless you are sure you are asking the right questions
- Start with open-ended interviews
- Keep a list of issues, opinions, examples
- Adapt these into a closed-ended survey

‘Grounded theory’ approach

<table>
<thead>
<tr>
<th>Collect:</th>
<th>Construct:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>Survey items</td>
</tr>
<tr>
<td>Stories</td>
<td></td>
</tr>
<tr>
<td>Lessons learned</td>
<td></td>
</tr>
</tbody>
</table>

“I’ve had success with phone conferences when everyone is online, but hard when most of the people are collocated and only a few are joining remotely”

I have had success with phone conferencing in the past

It is difficult to be the only person remote in a phone conference

Moving from data to theory

*Use interviews and observations to capture the range of opinions and ideas*
E.g. These are all the things people said when we asked about barriers to collaboration

*Use surveys to understand frequency and importance*
♣ 40% of people agreed that Idea 41 is a problem
♣ Only 7% of people agreed that Idea 33 is a problem

Scale construction

♣ Single items may be misinterpreted, may be unreliable
♣ Asking about specific behaviors or experiences is more valid than asking for general observations

Joining a meeting by phone is as good as being there in person
Strongly agree ———- Strongly disagree

I have been in meetings where people on the phone were ‘left out’
Strongly agree ———- Strongly disagree

It is difficult to be the only person remote in a phone conference
Strongly agree ———- Strongly disagree

People in this company do a good job of making sure remote collaborators are kept ‘in the loop’
Strongly agree ———- Strongly disagree
Scales re-used from another source


- **Affective subscale**
  - I would have preferred to be in a different group that the one I was in.
  - Members of this group like one another.
  - I enjoyed interacting with the members of this group.
  - I don’t like many of the other people in this group.

- **Behavioral subscale**
  - In this group, members relied on one another.
  - All members need to contribute to achieve the group’s goals.
  - This group accomplished things together that no single member could achieve.
  - In this group, members did not cooperate to complete tasks.

- **Cognitive subscale**
  - I see myself as quite similar to other members of this group.
  - I see myself as quite different from other members of this group.

---

Example scale:
SUMI helpfulness

3. The instructions and prompts are helpful.
8. I find that the help information given by this software is not very useful.
13. The way that system information is presented is clear and understandable.
18 There is never enough information on the screen when it’s needed.
23 I can understand and act on the information provided by this software.
28 The software has helped me overcome any problems I have had in using it.
33. The organization of the menus or information lists seems quite logical.
38 Error prevention messages are not adequate.
43 Either the amount or quality of the help information varies across the system.
48 It is easy to see at a glance what the options are at each stage.
Cronbach’s alpha example

In JMP software: Analyze › Multivariate methods › Multivariate › Item reliability › Cronbach’s alpha

SUMI output
Online services

- SurveyGizmo
- SurveyMonkey
- Zoomerang
- Checkbox

What can't the major online services do?

See Survey of Survey tools http://www.wac.ohio-state.edu/workshops/survey_of_surveys/

Social network analysis
Social network analysis

Q1. With whom do you exchange information, documents, schedules, and other resources to get your job done?
Q2. With whom do you discuss what is going on in the company, and who is doing what?
Q3. From whom do you seek inputs and opinions before making a key decision?
Q4. With whom do you discuss new ideas and innovations in financial products and services?
Q5. With whom do you discuss customer needs, requests, and feedback?

Bob
Carol
Ted
Alice

From Valdes Krebs’ InFlow

Network structure

A PL
Network structure

Two views of the same network

Other data sources
e.g. charging data

Use enterprise charging data as a proxy for collaboration

Look at collaboration patterns between specific groups, departments
Quasi-experimental methods

- Methods that borrow elements of experimental design but are not true experiments
  - Comparisons to expert performance (expert-novice comparison)
  - Comparison to theoretical 'optimal' performance
  - Comparison to doctrine or policy
  - Comparison to simulation result
  - Compare across non-equivalent conditions with statistical control of mediating variables
  - Post test compared with estimated pre-test
ADVISE timed tests

• Compare performance on small subtask to available expert
• In this case: expert was PI, self-timing on a set of analytical questions

Re-using experimental tasks

❖ Examples:
   ❖ Stasser’s Hidden Profile experiment
   ❖ Game theory tasks
Stasser’s Hidden Profile murder mystery

❖ Who killed Mr Guion?

❖ Group’s solution to a murder mystery is strongly influenced by how the information is distributed.

❖ Information is distributed so that each person has a deceptive first impression.

❖ ‘Red herring’ clues are widely distributed; critical clues are given to only one person.


Hidden Profile task

❖ Factors that make a hidden profile more difficult:
  ➤ Tendency of groups to share common knowledge and ignore unique knowledge.
  ➤ Time and resource pressures that prevent consideration of evidence.
  ➤ Poor knowledge-sharing technology.
  ➤ Poor knowledge sharing practices.
Strike planning task

- Choose CVOA: which of three seas should the task force move toward
- Identify suitability for strike
- Identify risks to forces in that sea
- Submit RFI's
- Submit requests for additional resources

Information distribution

Common information
- Warning Order
- Commander’s Intent
- Enemy Order of Battle
- including missile, sub, air capabilities
- Blue Force capabilities
- Weather updates
- Maps, including ranges

Strike only
- Strike target info
- SOF capabilities
- Bunker-busting bomb brief
- Updates on strike target situation

AW only
- Brief on civilian air traffic
- Brief on red missiles
- Updates on red missile types and locations
- Liaison to COCOM
- Request for tankers
- Request for P3

ASW only
- Detection ranges and region maps
- Brief on small boat threat
- Brief on red sub tactics
- Updates on blue, green, red sub positions
## Roles’ initial preferences for CVOA

<table>
<thead>
<tr>
<th>Strike</th>
<th>Strike</th>
<th>AW</th>
<th>AW</th>
<th>SCC</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strike distance</td>
<td>SOF insertion &amp; support</td>
<td>Air threat</td>
<td>Missile threat</td>
<td>Sub threat</td>
<td>Surface threat</td>
</tr>
<tr>
<td>Java Sea</td>
<td>Best (1)</td>
<td>Worst (3)</td>
<td>Worst (3)</td>
<td>Worst (3)</td>
<td>Worst (3)</td>
</tr>
<tr>
<td>Banda Sea</td>
<td>Med (2)</td>
<td>Best (1)</td>
<td>Med (2)</td>
<td>Best (1)</td>
<td>Best (1)</td>
</tr>
</tbody>
</table>

### Phase I Scores

<table>
<thead>
<tr>
<th></th>
<th>possible pts</th>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUM</td>
<td>221</td>
<td>77</td>
<td>40</td>
<td>53</td>
</tr>
<tr>
<td>Banda sea assessment</td>
<td>54</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Suitability for Strike</td>
<td>21</td>
<td>1</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Known risks to forces</td>
<td>33</td>
<td>10</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Celebes sea assessment</td>
<td>60</td>
<td>18</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Suitability for Strike</td>
<td>24</td>
<td>7</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Known risks to forces</td>
<td>36</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Java sea assessment</td>
<td>57</td>
<td>12</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Suitability for Strike</td>
<td>21</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Known risks to forces</td>
<td>30</td>
<td>9</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Ranking</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>RFTs assets</td>
<td>21</td>
<td>9</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Another experiment type: Social Dilemmas

- Everyone can put any amount, or nothing, into your envelope
- Total amount in envelope will be increased by 50%, and divided back evenly
- No talking while you decide!

Envelope game

- What is the group maximum?
- For every dollar you put into the envelope, how much extra pay do you receive?
- What would a game theorist say is the optimum strategy for any one player?
- Envelopes game is a simple ‘Social Dilemma’: a situation where the rational action taken by every individual leads to a sub-optimal solution for the group
- What is your prediction for most random groups of people?
Trust in four media conditions

Other experimental tasks

- Social dilemma
- Trust game
- Ultimatum game
- Negotiation
- Arctic survival
- Iowa gambling task
Example: Use of multiple methods in software evaluation

<table>
<thead>
<tr>
<th></th>
<th>Overall Score</th>
<th>SUMI</th>
<th>Daily</th>
<th>Wrap-up</th>
<th>Timed</th>
<th>Reports</th>
<th>Event Tags</th>
<th>Debrief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>4</td>
<td>4</td>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>2</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learnability</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helpfulness</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>4</td>
<td></td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stability</td>
<td>3</td>
<td></td>
<td>4</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suitability</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

Future ‘micro’ methods

- Language analysis
- Physiological methods
- Brain science methods
Language analysis - LIWC

- Pennebaker and colleagues aggregated keyword-based language studies into a single tool
- Useful for characterizing large corpuses of text
- Most successful so far:
  - Positive and negatively valenced words
  - 'I' versus 'we' analysis
  - National online 'conversation' shifted for two weeks after 9-11 to use more first person plural

- http://www.liwc.net/


Language analysis - LSA

- LSA treats texts as a 'Bag of words'- it has no explicit knowledge of grammar and discards information about word order
- LSA models train on a large corpus of text, then perform multiple choice or classification tasks
- Uses singular value decomposition
- Similar to a neural network

Proximity of 'airplane' to other terms across a large text corpus

Airplane
Plane
Jet
Rocket
Boeing
Wing
stewardess

Physiological and facial data

Physiological monitoring to understand emotional states, workload, arousal

Physiological measures can be gathered more easily

Facial action coding can be done from video!


Brain Sciences

<table>
<thead>
<tr>
<th>Spatial resolution</th>
<th>Temporal resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 cm</td>
<td>1 ms</td>
</tr>
<tr>
<td>10 cm</td>
<td>10 ms</td>
</tr>
<tr>
<td>1 cm</td>
<td>2s</td>
</tr>
<tr>
<td>5 mm</td>
<td>2 min</td>
</tr>
<tr>
<td>.1 mm</td>
<td>Ideal</td>
</tr>
</tbody>
</table>

EEG

fMRI

 Basic Research

Applied Research

MEG

NIRS

PET

APL
Chemical data?

- Oxytocin produces more trusting behavior in a well known social dilemma
- Oxytocin is a hormone that is internally released in social situations
- Oxytocin can be nasally administered

Increased trusting behavior toward other humans

(No general increase in risk tolerance)

Example problem

- Collaboration in a large, distributed organization
- How do you measure collaboration?
- How do you measure attitudes?
- How do you measure change?