Assessing CPS in Action: Practical Considerations and Preliminary Results

Jiangang Hao
Center for Advanced Psychometrics
Educational Testing Service
Goal: explore the possibility of measuring cognitive skills and CPS skills in the domain of science

Team members:

- Jiangang Hao, Alina von Davier [Center for Advanced Psychometrics]
- Lei Liu, Diego Zapata, Tanner Jackson [Cognitive and Learning Sciences]
- Patrick Kyllonen [Center for Academic and Workforce Readiness and Success]
- Heather Walters, Chris Kitchen, Lauren Carney [Research Coordination & Support]

Others worked on the project: Su-Youn Yoon, Lei Chen [NLP], Zuowei Wang [Univ. of Michigan, 2014 summer intern]

Initiated in June 2013

Develop simulation tasks to assess CPS

Aim at collecting 1500 responses from Amazon Mechanical Turk

75% of the data collection is done by now
Challenges of assessing CPS

- Construct definition
- Entangled cognitive skills and social skills
- How to determine the individual contributions to the team performance
- Many confounding factors
  - Personality
  - Demographic information
  - General science knowledge
- ...

3
Key research questions

- Identify the CPS construct for psychometrically rigorous assessment
- Prototype game-like environment for assessing CPS
- Explore the relation between cognitive and social skills in a CPS task
- Explore the impact of individual cognitive skills on the team performance
- Explore the impact of individual social skills on the team performance
- Explore the impact on the collaboration by personality, demographic information, general science knowledge
- Developing automated facilitating mechanism (intelligent facilitator)
Simulation task

- Modified Volcano Triologue as web-based version
- Added additional layer for collaboration

Single player version

CPS version
Collaborative Layer

System Prompt Area to facilitate:
- Flow of task
- Collaboration

Chat Box Area for team members to:
- Share ideas
- Negotiate & develop team response
- Track history of discussion

Slide from Lei Liu
Collaboration Facilitation

For each item in the simulation

1. Two participants are prompted to respond separately first.
2. Two participants are prompted to:
   - Exchange their answers
   - Discuss to get the best answer
3. Each participant is given an opportunity to revise his/her initial answers after discussion
4. One participant is randomly chosen to submit a team answer.
Collaborative Problem Solving Skills Framework

CPS Skills

- Sharing Resources / Ideas
- Assimilating & Accommodating Knowledge / Perspective Taking
- Regulating Problem Solving Activities
- Maintaining Positive Communication

Note: CPS framework was developed based on review of CSCL research, the PISA 2015 CPS framework, and ECD principles.

Liu et al, 2014
Experiment Design

- Four main components
  a) Two Simulation tasks about volcano science
     i. Single player version
     ii. CPS version (dyads) [the only place that CPS happens]
  b) General science knowledge test consist of 40 multiple choice items
  c) Questionnaires for demographic information
  d) Questionnaires for personality (TIPI)

- Data Collection: amazon turk
  - 1500 total participants, all take b), c) and d)
  - 500 participants take Single player version
  - 1000 participants take CPS version
# Project timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>Milestone</th>
</tr>
</thead>
</table>
| June 2013 - Nov. 2013 | • Single player version development  
                        | • Usability study                                                 |
| Dec. 2013             | • Data collection for single player version  
                        | • 500 responses                                                   |
| Jan. 2014 – May 2014  | • CPS version development  
                        | • Usability study                                                 |
| May 2014 - Now        | • Data collection for CPS version  
                        | • Responses from 320 dyads                                        |
| June 2014 - Now       | • Scoring rubrics development  
                        | • Data reduction pipeline                                         |
|                       | • Preliminary data analysis                                       |
| 2015                  | Data analysis and report findings                                 |
Preliminary results

- Collaboration has effects

![Graph showing sum scores of the first 7 SR items for different groups. The graph indicates a significant effect with a confidence level of greater than 4.5 σ.](image)
Collaboration heat: number of words in the conversation

"heat" of communications based on 318 dyads
An interesting conversation

hi how are you?
I am fine. How about yourself?
good glad to be able to work on this with you
I feel the same way.
I wonder what we will be doing.
I was just about to ask what should we do now
click next?
I would wait a bit.
some thing dealing with volcanos, pretty sure.
tracking the I think
Yes.
Have you ever seen a real volcano?
nope, but that would be fun. You?
I saw Mt. St. Helens when I was younger.
Nice!
It was real pretty.
I bet there a little scary to I think too. A good mix
of fun too!
Clustering based on the heat vector
Process data: positive or negative outcomes

Response change for total score 0

total # of teams: 103
Process data: positive or negative outcomes

Response change for total score 0

total # of teams: 103
The special team
Ongoing

- Compare the performance for CPS version and single player version and model the change
- Establish the relations between the performance in the task and the social skills from the chat
- Examine relations between the social skills and the demographic information and personality
- Conduct data mining of the log files, NLP features of the chat
- Develop algorithm to automatically classify the conversation, pave the road for an intelligent facilitator.
Summary

- Task development and data collection has been going successfully.
- The participants are serious about our task.
- Preliminary analyses show that there is a significant performance improvement due to collaboration.
- Different topics in the task elicit different amount of communications.
- More results are coming.
Thank You!
Back up slides
Task deployment

Multiple choice science test
Demographic survey
Personality questionnaire

Single player version simulation
CPS version simulation

Amazon Mechanical Turk

Websuveyor.com

Bluehost server

ETS server

Dialogue analysis engine
Log file management

- Structured log file in xml
- Time-stamped actions
- Distributed file uploading:
  - Each participant upload his/her own log
  - Logs updated after each action/event
  - Standardized tags

\[
\begin{align*}
\text{<action name>} \\
\text{<action time>} \\
\text{<action to>} \\
\text{<action by>} \\
\text{<action results>} \\
\text{<slide name>}
\end{align*}
\]

Hao et al, in preparation, 2014
Data reduction pipeline

- A set of python modules to:
  - Extract participant IDs
  - Matching participants across tasks
  - Extract and aggregate information
  - Implement scoring rules
  - Identify new features
  - NLP feature extraction
  - ....

- Can be applied to Trialogue tasks
Preliminary results

- Turks are serious about our task
- Based on MC science items
  Cronbach’s Alpha ~ 0.89
- Other CTT parameters are also reasonable

<table>
<thead>
<tr>
<th>Sample</th>
<th>SD</th>
<th>Item.total</th>
<th>Item.Item</th>
<th>Difficult</th>
<th>Item.Disc</th>
<th>Item.Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0</td>
<td>0.36</td>
<td>0.35</td>
<td>0.35</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>X1</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>X2</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>X3</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>X4</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>X5</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>X6</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>X7</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>X8</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>X9</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>X10</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>X11</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>X12</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>X13</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>X14</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>X15</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>X16</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>X17</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>X18</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>X19</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>X20</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>X21</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>X22</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>X23</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>X24</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>X25</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>X26</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>X27</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>X28</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>X29</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>X30</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>X31</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>X32</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>X33</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>X34</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>X35</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>X36</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>X37</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>X38</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>X39</td>
<td>1.16</td>
<td>1.16</td>
<td>1.16</td>
<td>1.16</td>
<td>1.16</td>
<td>1.16</td>
</tr>
</tbody>
</table>