Assessment of Learning in Collaborative Tasks

Carolyn Penstein Rosé
Language Technologies Institute
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon

With funding from the National Science Foundation and the Office of Naval Research
Outline

- Context of the Work: Dynamic Support for Collaborative Learning

- Overview of Souflé framework
  - Transactivity
  - Engagement
  - Authoritativeness

- Example Application: MathTalk
  - Assessing social climate effects on Authoritativeness, help exchange, and learning
Students learn 1.24 s.d. more when working with a partner and automated support than students working alone (Kumar et al., 2007)
Empirical Support for Design Principles

- **Personalized agents** increase supportiveness and help exchange between students (Kumar et al., 2007)

- Agents are more effective when **students have control** over timing of the interaction (Chaudhuri et al., 2008; Chaudhuri et al., 2009)

- Agents that employ **Balesian social strategies** are more effective than those that do not (Kumar et al., 2010; Ai et al., 2010)

- Students are sensitive to agent **rhetorical strategies** such as displayed bias (Ai et al., 2010), **displayed openness** to alternative perspectives (Kumar et al., 2011), and targeted elicitation (Howley et al., 2012)

- **Accountable talk agents** (Dyke et al., 2013; Adamson et al., 2014)
Low level features we used to believe in

- General indicators of interactivity
- Turn length
- Conversation length
- Number of student questions
- Student to tutor word ratio
- Student initiative


Souflé Framework

(Howley et al., 2013)

Authority

Transactive Knowledge Integration

Engagement

Person

Engagement

Authority
Souflé Framework
(Howley et al., 2013)

3 Dimensions:
- Transactivity
- Engagement
- Authoritativeness
Souflé Framework
(Howley et al., 2013)

Transactive Knowledge Integration

Person ↔ Person
• Definition of Transactivity
  • building on an idea expressed earlier in a conversation
  • using a reasoning statement

I think the tube will get heavier because water is going in

That’s true, but the important point is that water can flow in, but starch can’t flow out.
Transactivity (Berkowitz & Gibbs, 1983)

- **Findings**
  - Moderating effect on learning (Joshi & Rosé, 2007; Russell, 2005; Kruger & Tomasello, 1986; Teasley, 1995)
  - Moderating effect on knowledge sharing in working groups (Gweon et al., 2011)

- **Computational Work**
  - Can be automatically detected in:
    - Threaded group discussions (Kappa .69) (Rosé et al., 2008)
    - Transcribed classroom discussions (Kappa .69) (Ai et al., 2010)
    - Speech from dyadic discussions (R = .37) (Gweon et al., 2012)
    - Predictable from a measure of speech style accommodation computed by an unsupervised Dynamic Bayesian Network (Jain et al., 2012)
Souflé Framework
(Howley et al., 2013)

Transactive Knowledge Integration

Engagement → Person ↔ Person → Engagement
Engagement (Martin & White, 2005)

- **Findings**

  - *Correlational analysis*: Strong correlation between displayed openness of group members and articulation of reasoning (R = .72) (Dyke et al., in press)

  - *Intervention study*: Causal effect on propensity to articulate ideas in group chats (effect size .6 standard deviations) (Kumar et al., 2011)

  - Mediating effect of idea contribution on learning in scientific inquiry (Wang et al., 2011)
Souflé Framework
(Howley et al., 2013)

Authority

Transactive Knowledge Integration

Person

Engagement

Authority

Engagement
Authoritativeness (Martin & Rose, 2003)

Findings

- Authoritativeness measures display how students respond to aggressive behavior in groups (Howley et al., in press)
- Authoritativeness predicts learning ($R = .64$) and self-efficacy ($R = .35$) (Howley et al., 2011)
- Authoritativeness predicts trust in doctor-patient interactions ($R$ values between .25 and .35) (Mayfield et al., under review)

Computational Work

- Detectable in collaborative learning chat logs ($R = .86$)
- Detectable in transcribed dyadic discussions in a knowledge sharing task ($R = .95$) (Mayfield & Rosé, 2011)
- Detectable in transcribed doctor-patient interactions ($R = .96$) (Mayfield et al., under review)
Example: MathTalk

- Manipulation of Social Climate
- Social Positioning Risk Taking
- Engagement in Problem Solving and Learning
Example: MathTalk

- Personalized Agent condition vs Control condition
- 30 6th graders
  - Randomly assigned to pairs, conditions
- Procedure

Social Dialogue Agent Study (Kumar et al, 2007)
### Example: MathTalk

<table>
<thead>
<tr>
<th>Tutor:</th>
<th>Student1, if you had to choose between a long flight or a long car ride, which seems less uncomfortable?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student1:</td>
<td>I'll take the car ride</td>
</tr>
<tr>
<td>Tutor:</td>
<td>Ok, Student1.</td>
</tr>
<tr>
<td>Tutor:</td>
<td>Student2, which are more entertaining- books or movies?</td>
</tr>
<tr>
<td>Student2:</td>
<td>definitely books!</td>
</tr>
<tr>
<td>Tutor:</td>
<td>Ok, Student2.</td>
</tr>
</tbody>
</table>

Jan packed several books to amuse herself on a long car ride to visit her grandma. After 1/5 of the trip, she had already finished 6/8 of the books she brought. How many times more books should she have brought than what she packed?
Main Results, Advantage for Social Condition

[Kumar et al, 2007]

- Significant increase in perception of amount of help given and received
  - Significant increase in amount of help given per problem (Gweon et al., 2007)
  - Students marginally more likely to complete a step on their own after receiving help (Cui et al., 2009)
- Marginally higher learning gains

- But why?
Understanding the Effect of Social Climate on Positioning and Risk Taking

- Coded chat logs for instances of aggressive behavior
  - Pushy behavior
  - Insults

- Coded for Negotiation (especially K1 and K2)
  - Based on counts of K1 and K2, computed an authoritativeness score for each student per lab day
    - $\frac{K1}{K1 + K2}$
  - Computed a Shift score per student
    - Residual from linear regression predicting Day 2 authoritativeness from Day 1 authoritativeness
  - Binary Shift variable (within pair, which student shifted up to a more authoritative stance versus shifted down)
Significantly more aggressive behavior in Control condition
- $F(1,56) = 8.93, p < .005 \, **$, effect size $0.63\sigma$

Significantly more aggressive behavior on Day 2
- $F(1,56) = 15.61, p < .0005 \, **$, effect size $0.87\sigma$
- Significant interaction with Condition
  - $F(1,56) = 6.06, p < .05 \, **$
  - Only significant increase in aggressive behavior on Day 2 in the Control condition

In each pair, identified student with higher amount of aggressive behavior on Day 2 as “the bully” for further analysis
Authoritateness and Shift

- Significant difference in Authoritateness of Bullies and Non-Bullies in Control condition on Day 2
  - $F(1,23) = 5.92, \ p < .05^{**}$

- Visible Shift only in Control Condition
  - $F(1,23) = 5.28, \ p < .05^{**}$, effect size $0.15\sigma$
    - Bullies in Control condition shifted to more authoritative stance
    - Non-bullies in Control condition shifted to less authoritative stance
Learning

- No significant main effect of Aggressive behavior on learning
  - Bullied students in Control condition learned significantly less than Social Condition students
  - Recall that students respond differently to help in Control condition

- Significant interaction between Shift and Condition on Learning: $F(1,20) = 7.91, p = .01^{**}$
  - Opposite trend in Social Condition
  - Significant correlation between amount of shift and learning only within Control condition
  - Shifting down was associated with less learning
Conclusions

- 3 Dimensional Souflé framework captures the interplay between social and cognitive factors in discussion based learning

- Learning is explained by cognitive factors

- Social factors inhibit or enhance cognitive processes related to learning
Thank You!

Questions?