* This research is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A120370 to the University of Pittsburgh, and also by NFS 1122504.

Classroom Evaluation of a Scaffolding Intervention for Improving Peer Review Localization*

Huy V. Nguyen[2], Wenting Xiong[1,2], and Diane J. Litman[1,2]

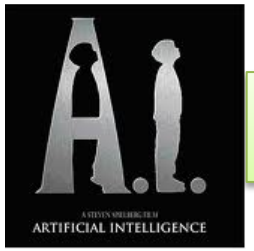
[1] LRDC | [2] CSD, University of Pittsburgh

ITS 2014
5 - 9 June

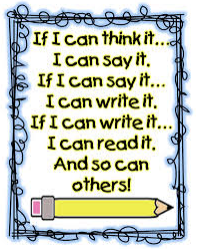
UNIVERSITY OF PITTSBURGH
LRDC
Learning Research & Development Center

DEPARTMENT OF COMPUTER SCIENCE

Context

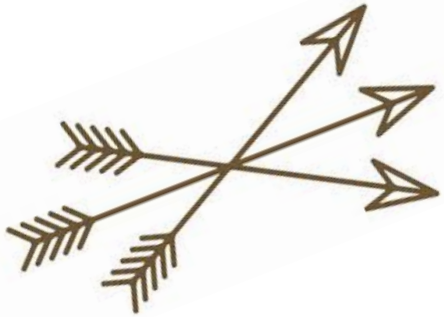


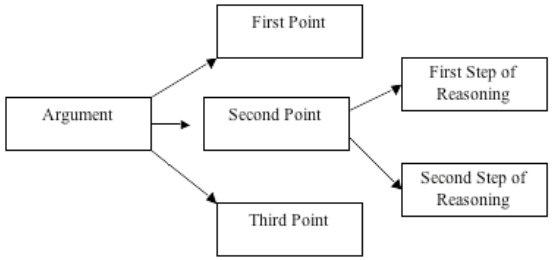
- For effective peer reviewing



SWoRD Peer Review

Panther Learning





Peer review helpfulness

Peer review

(Helpfulness as implementation likelihood)

Instructor guidance and review prompts have limited impact

A capable peer review system that actively help make helpful reviews.

- Natural Language Processing
- Machine Learning

Research goal

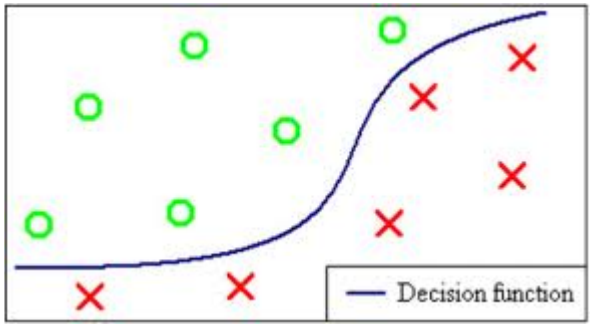
- How well does system predict review helpfulness/*its signal*?
  - Improve **system performance**
- How do student reviewers respond to system intervention?
  - Understand **student behavior**: agree vs. disagree
- How does scaffolding intervention impact reviewer revisions?
  - Analyze **student learning**, design interface

Feedback localization for more helpful review

Helpful peer review

Pinpoints where in the source the feedback applies (*Nelson & Schunn '09*)

> ***Study 17*** *doesn't have a connection to anything, which makes it unclear about it's purpose […].*
>
> *[…] need captions for* ***figure 1 and 2***



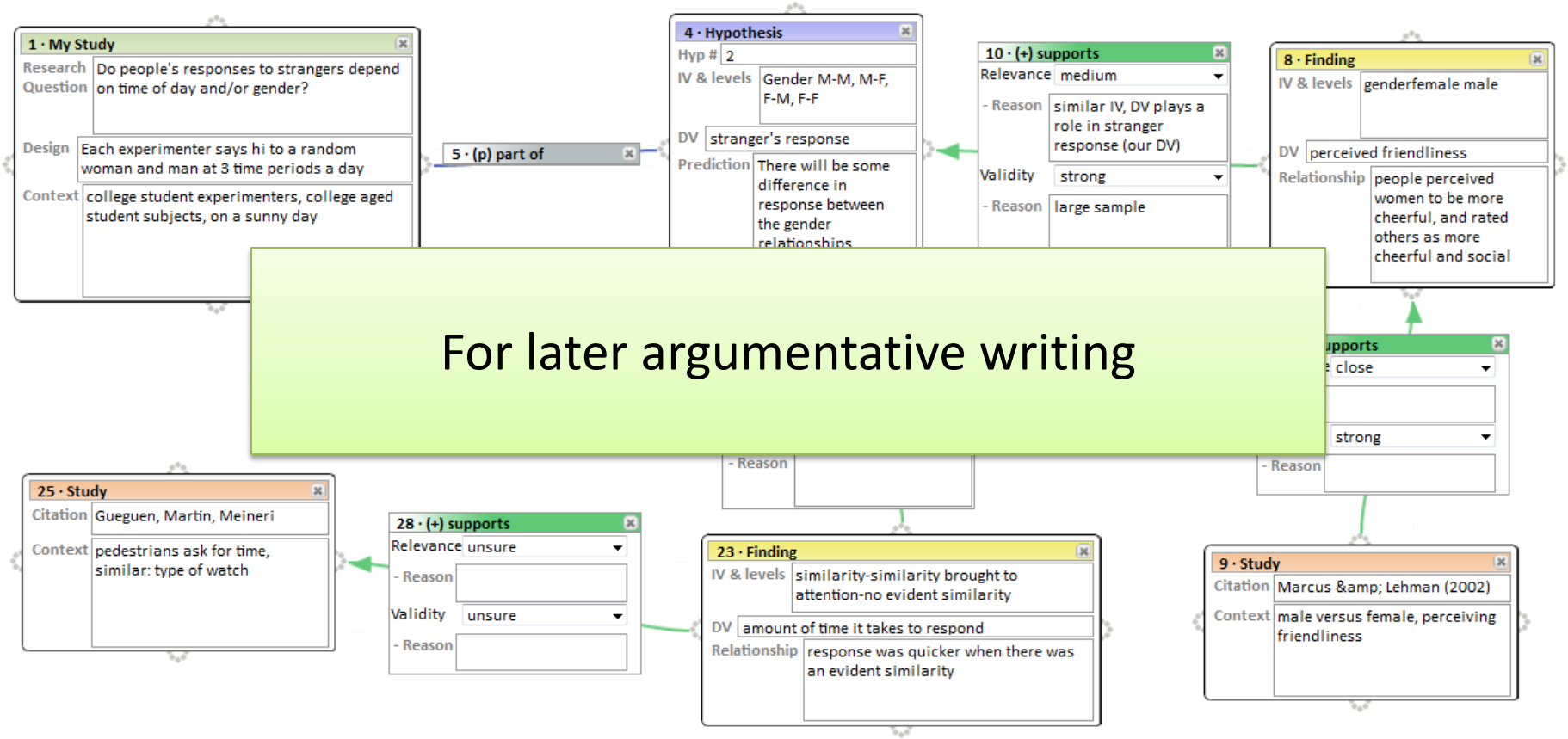
Predicting feedback localization (*Xiong & Litman '10, Nguyen & Litman '13*)

- Localized v. Not-localized
- Peer reviews of student argument diagrams v. papers

- Scaffolding intervention to help student reviewers
- Mining tool to help instructors

Argument diagram

(*edited in LASAD,* Scheuer et al. '10)



Scaffolding system feedback

(*an example of argument diagram review*)

Review prompt

#3. Are the relevance, validity, and reason fields in the supportive arcs complete and convincing? If not, indicate where the

System gives guidance

Make sure that for every comment below, you explain **where** in the diagram it applies. For example, you can indicate where your comments apply by:

(1) Specifying node(s) and/or arc(s) in the author's diagram to which your comment refers

- Your **conflicting/supporting [node-type]** is really solid!

(2) Quoting the excerpt from the author's textual content of node and/or arc to which your comment refers

- For your **[node-type] that talks about body chemistry and cortisol levels,** you should clarify how that is related to politeness specifically.

(3) Referring explicitly to the specific line of argumentation that your comment addresses

- Why does **claim [node-ID] support the idea that people will be more polite in the evening?**

Reviewer makes decision

I've revised my comments. Please check again.

REVISE

I don't know how to specify where in the diagram my comments apply. Could you show me some examples?

My comments don't have the issue that you describe. Please submit comments.

DISAGREE

Sample paper review

Make sure that for every comment below, you explain **where** in the paper it applies. For example, you can indicate where your comments apply by:

(1) Specifying page numbers and paragraph numbers in the author's text to which your comment refers

(2) Referring explicitly to the specific topic that your comment addresses

(3) Quoting the excerpt from the author's text to which your comment refers

#8. APA Style: Is APA style used correctly for the following? - *Numbers - Statistics - In-text citations - Paper header - Abbreviations - Section headings* Etc. Are the following elements formatted according to APA style? - *Abstract - Introduction - Method - Results - Discussion - References - Table/Figure*

Comment Entry 1: *(*Required)*

need captions for figure 1 and 2

Comment Entry 2:

go thru APA manual and make sure everything is formatted correctly

Related work

- Studies on effects/helpfulness of peer feedback (Gielen et al. ’10, Nelson & Schunn ’09)

- Automatic feedback feature prediction
  - Localization (Xiong & Litman ’10, Nguyen & Litman ’13)
  - Problem/solution (Xiong et al., ’12, Nguyen & Litman ’14)

- Feedback helpfulness prediction
  - Binary classification (Cho ’08)
  - Helpfulness rating (Xiong & Litman ’11)
  - Other measures (Ramachandran & Gehringer ’11)

- System scaffolding for feedback quality improvement (Kumar ’10, Razzaq & Heernan ’10)

Outline

- System setting and data
- Analyses (*regarding research goals*)
  - Prediction performance
  - Reviewer response (*to system intervention*)
  - Review revision
- Conclusions and future work

System setting

- Predicts review comments (of paper and argument diagram) for localization
  - Flags not-localized comments in red
  - Intervenes if ratio of localized comments < **0.5**
- Student reviewers response with one of two following options:
  - **REVISE**: Revise their reviews and resubmit
  - **DISAGREE**: Submit their reviews without revision

Peer review data

- Student reviews of argument diagrams and papers from Research Method 2013 course in psychology
- Diagrams → Peer reviews → Papers → Peer reviews

Table 1. Peer review data statistics

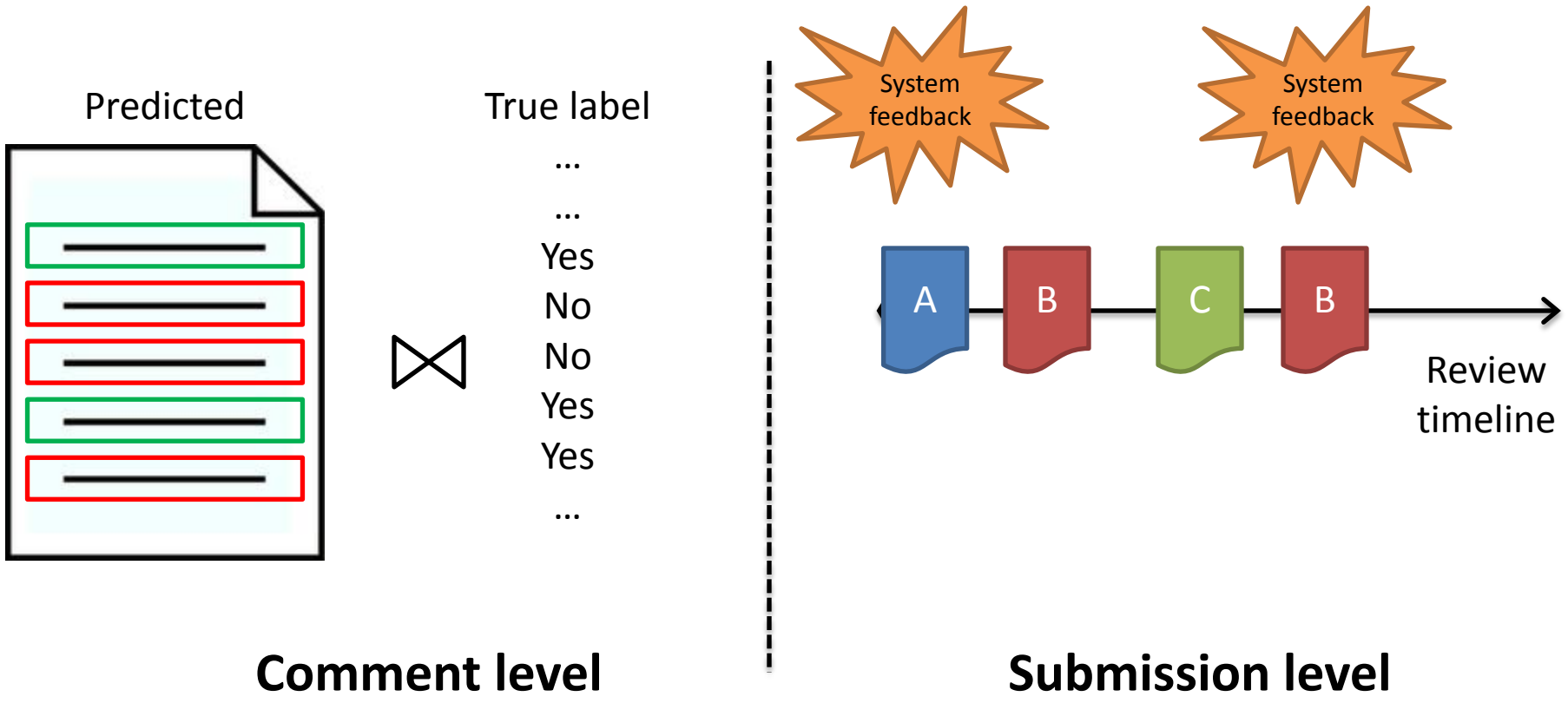
	Diagram review	Paper review
Reviewers/Authors	181/185	167/183
Submitted reviews	788	720
Intervened submissions	173	51
Localized comments	449	347
Not-localized cmnts.	718	336

> %

Outline

- System setting and data
- Analyses
  - Prediction performance
  - Reviewer response (*to system intervention*)
  - Reviewer revision
- Conclusions and future work

Prediction performance



Localization prediction

- Prediction at comment level: significantly outperform the majority baselines
- Close to with reported results (in experimental setting) of previous studies (*Xiong & Litman 2010, Nguyen & Litman 2013*)
  - Prediction models are robust even in not-identical training-testing

Table 2. Localization prediction performance

	Diagram review		**Paper review**	
	Accuracy	Kappa	Accuracy	Kappa
Baseline	61.5%	0	50.8%	0
Model	**81.7%**	**0.62**	**72.7%**	**0.46**

Intervention accuracy

- Student-perspective evaluation:
  - *Students do not know the localization threshold*
  - *An intervention is considered wrong by student if **all of its comments are localized***

- Intervention at submission level: only 1 incorrect intervention of diagram review

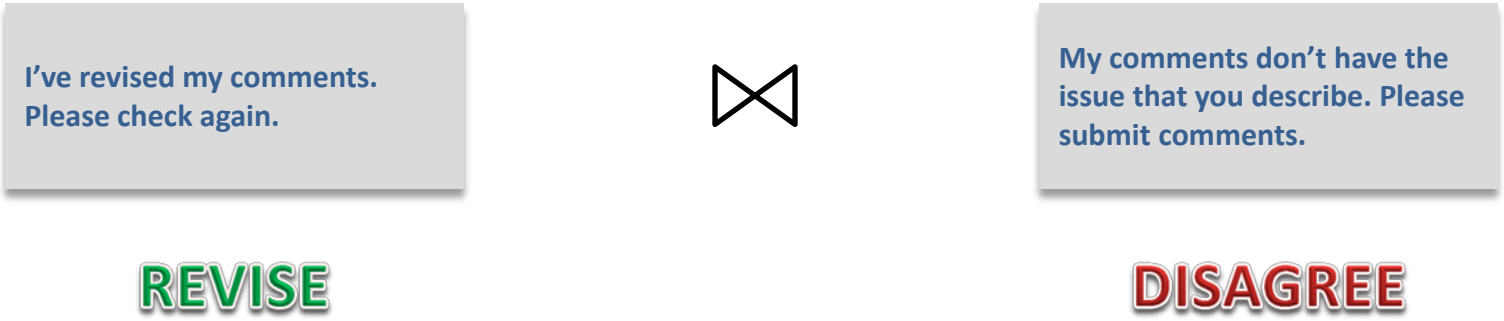
Table 3. Intervention accuracy

	Diagram review	**Paper review**
Total interventions	173	51
Incorrect interventions	1	0

Outline

- System setting and data
- Analyses
  - Prediction performance
  - Reviewer response (*to system intervention*)
  - Review revision
- Conclusions and future work

Student response analysis



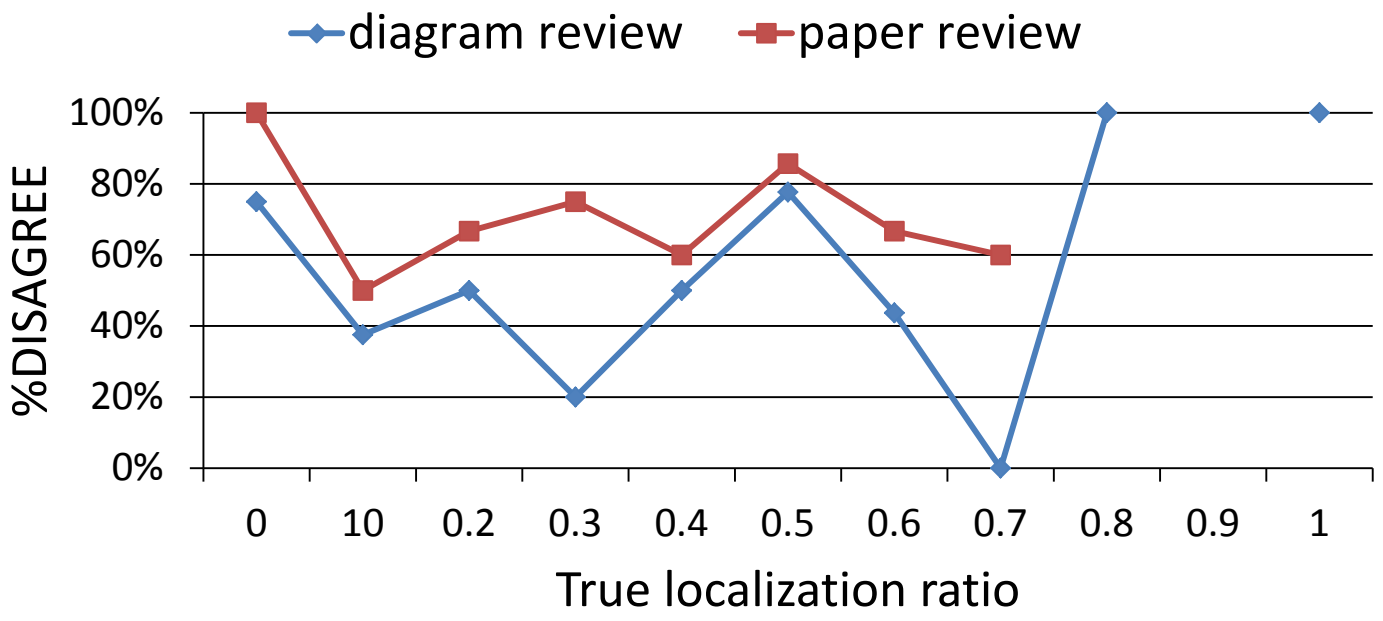
- Student reviewers disagreed more than they agreed with system intervention

Table 4. Student response percentage

	Diagram review	Paper review
REVISE	48%	30%
DISAGREE	52%	70%

Student response analysis

- Students' disagreement is not related to how well the original review were localized



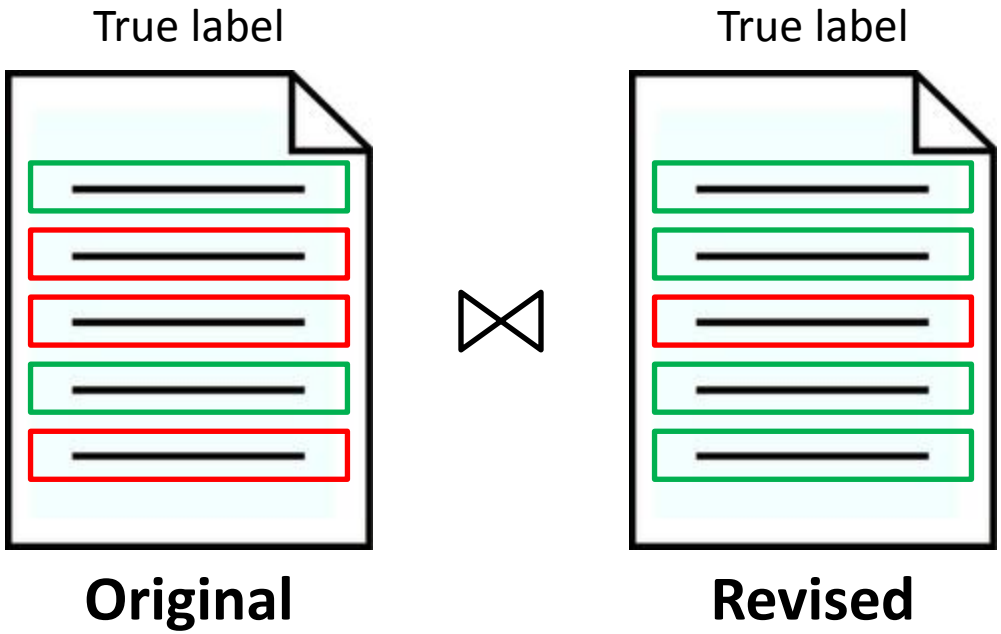
Outline

- System setting and data
- Analyses
  - Prediction performance
  - Reviewer response (*to system intervention*)
  - Review revision
- Conclusions and future work

Review revision: localization change

- Resubmissions from intervened and not-intervened reviews
  - Reviewers could resubmit review without being intervened

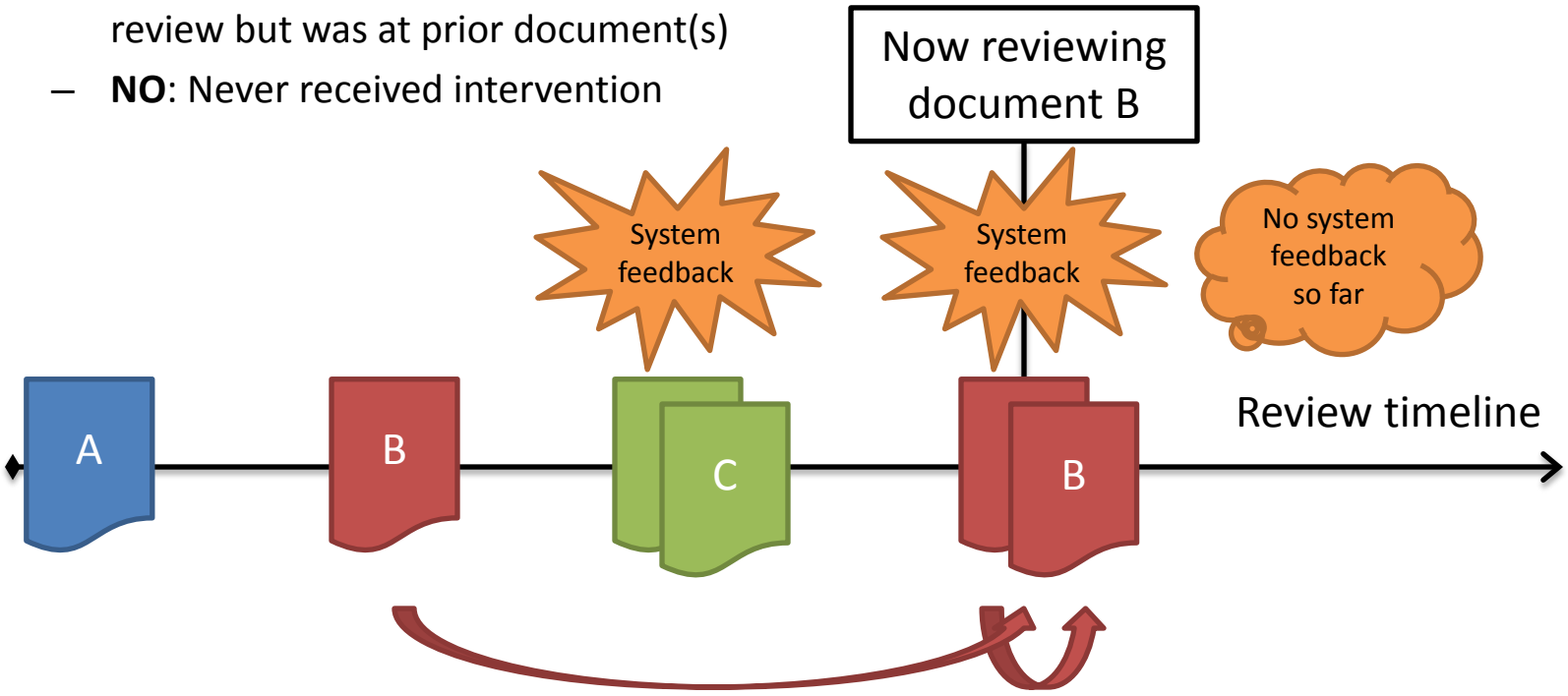
- Localization change patterns:



- Yes → No
- Yes → Yes
- No → No
- No → Yes

Review revision: intervention scope

- Intervention scopes
  - **IN**: Revision is in response to intervention of the current reviewing
  - **OUT**: Not intervened at the current review but was at prior document(s)
  - **NO**: Never received intervention



Review revision analysis

- Localization change between revisions: Yes (localized) v. No (not-localized)

Table 5. Comment change patterns by intervention scopes.

	Diagram review			Paper review		
	IN	OUT	NO	IN	OUT	NO
No → Yes	26	**7**	3	8	**2**	5
Yes → Yes	26	1	**16**	13	1	**29**
No → No	**33**	0	5	**19**	1	20
Yes → No	1	0	0	0	0	1

- Scope=IN: *potential improvement in system feedback*
- Scope=OUT: *impact of system feedback remains in later review sessions*
- Scope=NO: *revision might not be due to localization*

Outline

- System setting and data
- Analyses
  - Prediction performance
  - Reviewer response (*to system intervention*)
  - Reviewer revision
- Conclusions and future work

Conclusions

- Enhanced peer review system
  - Integrated two review localization models and implemented scaffolding intervention.
- High performance at both comment level and submission level
- Scaffolding intervention helped student localize their comments
  - … even in later not-intervened review sessions.

Future work

- Large number of student disagreements
  - *No relation with time-to-deadline*
  - Reviewing time
  - User study
- Large number of unsuccessful attempts
  - *Improve user interface, highlight localization text*
- Implement other feedback features
  - *Problem identification and solution providence*
- Adapt to other courses,
- *… also high school students*

***Study 17** doesn't have a connection to anything, which makes it unclear about it's purpose […].*

*[…] need captions for **figure 1 and 2***

ArgumentPeer and SWoRD projects

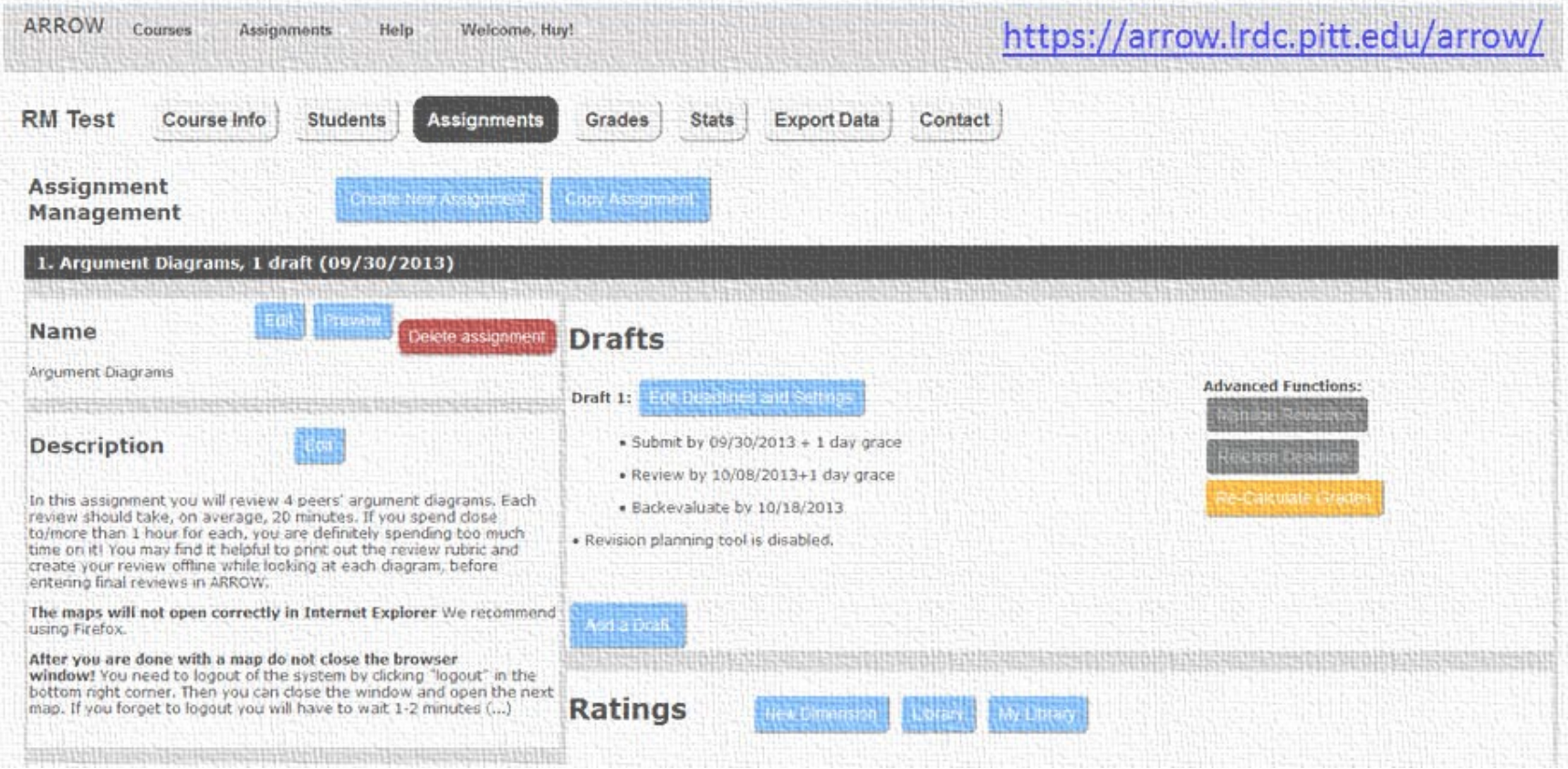
SWoRD Research

https://sites.google.com/site/swordlrdc/home

UNIVERSITY OF PITTSBURGH
LRDC
Learning Research & Development Center

DEPARTMENT OF COMPUTER SCIENCE

Thank you!



Prediction features

- Xiong and Litman 2011: studied syntactic features from the parsed dependency tree of sentence
- Domain word count (**dw_cnt**)
  - dictionary of domain word is learned automatically from set of argument diagrams
- **So_domain**: indicates whether domain word appear between subject and object of review
- **Det_count**: counts number of demonstrative determiners in comment
- Overlapping window features:
  - Compute the maximal overlapping window
  - Report window size (**wnd_size**) and number of common words (**overlap_num**)

Prediction features

- Location information must involve diagram component **keyword** surrounded by **supporting words**

- A diagram component keyword:
  - The words *node* or *arc*
  - Node/arc type from the ontology (*parsed automatically*)
- Supporting words are in proximity of a keyword which help locate the component

- Supporting words are selected from common words between review and node/arc content (*stemmed already*)
- Identified accordingly to 5 localization pattern (*applied to review sentences that have common words*)
- **Numbered ontology type**: supporting words are number/list of numbers right after keyword
- **Textual component content**:
  - Supporting words occur right before keyword
  - Or after keyword with distance less than 3
- **Unique component**: count number of node/arc of each type while parsing argument diagrams

- **Connected component**: extend node/arc text by the textual content of the other node/arc that it connects to
  - Supporting words must be in the extended content

- **Typical numerical expressions**: use held-out development data to learn regular expressions

ArgumentPeer and SWoRD projects

SWoRD Research

https://sites.google.com/site/swordlrdc/home

UNIVERSITY OF PITTSBURGH
LRDC
Learning Research & Development Center

DEPARTMENT OF COMPUTER SCIENCE