



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

Huy V. Nguyen<sup>2</sup>, Wenting Xiong<sup>1,2</sup>, and Diane J. Litman<sup>1,2</sup>

<sup>1</sup> Learning Research and Development Center | <sup>2</sup> Department of Computer Science, University of Pittsburgh



\* 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.

## Introduction of system setting

- Peer review systems, e.g. SWoRD [1], need intelligence for *detecting* and *responding* to problems with students' reviewing performance
  - E.g. **problem localization** (pinpoints the location of the problem mentioned in the feedback), *providing solution*
- Peer review localization models [2, 3] has been deployed in SWoRD :
  - Predicts comments of paper and argument diagram reviews for **localization**
  - Flags not-localized comments in red
  - Intervenes if ratio of localized comments less than a threshold of 0.5
- Student reviewers response with one of two following options:
  - REVISE**: Revise their reviews and resubmit
  - DISAGREE**: Submit their reviews without revision
- Student can resubmit multiple times. All resubmissions go through the same localization prediction procedure.

## Peer review corpus and research goal

- Student reviews of argument diagrams and papers from Research Method course in psychology

Table 1. Peer review data statistics. All resubmissions are counted.

	Diagram review	Paper review
Reviewers/Authors	181/185	167/183
Submitted reviews	788	720
Intervened submissions	173	51

- Questions of interest:
  - Q1**: How well did system predict feedback localization?
  - Q2**: How did student reviewers respond to system intervention?
  - Q3**: How did system scaffolding impact reviewer revisions?

- Data of interest:
  - Intervened submissions and subsequent resubmissions (if any).
  - Resubmissions occurred after a non-scaffolded submissions.

Table 2. Localization annotation results. Inter-rater kappa = 0.8

	Diagram review	Paper review
Localized comments	449	347
Not-localized comments	718	336

**References**

- K. Cho and C. D. Schunn (2007). *Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system*. Computers and Education, 48(3), 409-426.
- H. Nguyen and D. Litman (2013). *Identifying Localization in Peer Reviews of Argument Diagrams*. In Proceedings of the 16th international conference on Artificial Intelligence in Education (AIED), pp. 91-100. Springer-Verlag.
- W. Xiong and D. J. Litman (2010). *Identifying Problem Localization in Peer-Review Feedback*. 10th International Conference on Intelligent Tutoring System, Proceedings, 429-431.

## Examples of student comments and system intervention

### Sample comments of an argument diagram review

#1. Are any parts of the diagram hard to understand because they are unclear? If so, describe any particularly confusing parts of the diagram and suggest ways to increase clarity.

Comment Entry 1: (\*Required)

Although the text is minimal, what is written is fairly clear.

Comment Entry 2:

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

### Sample comments of a paper review

#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

### System scaffolded feedback of argument diagram review

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?

### System scaffolded feedback of 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

I've revised my comments. Please check again.

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.

Student possible responses: REVISE (left) and DISAGREE (right)

## Q1: Localization model performance

- Comment level for prediction performance:** significantly outperform the majority baselines.

Table 3. Localization prediction performance at the comment level.

	Diagram review			Paper review		
	Accuracy	F-measure	Kappa	Accuracy	F-measure	Kappa
Baseline	61.5%	0.47	0	50.8%	0.34	0
Model	81.7%	0.82	0.62	72.7%	0.73	0.46

- Submission level for intervention accuracy** (w.r.t student reviewer perspective)
  - A reviewer thinks the system incorrectly intervened when all of his comments were localized.
  - Correct intervention when at least one comment is human-labeled Not-localized.
  - Results: *Only one incorrect intervention of diagram review.*

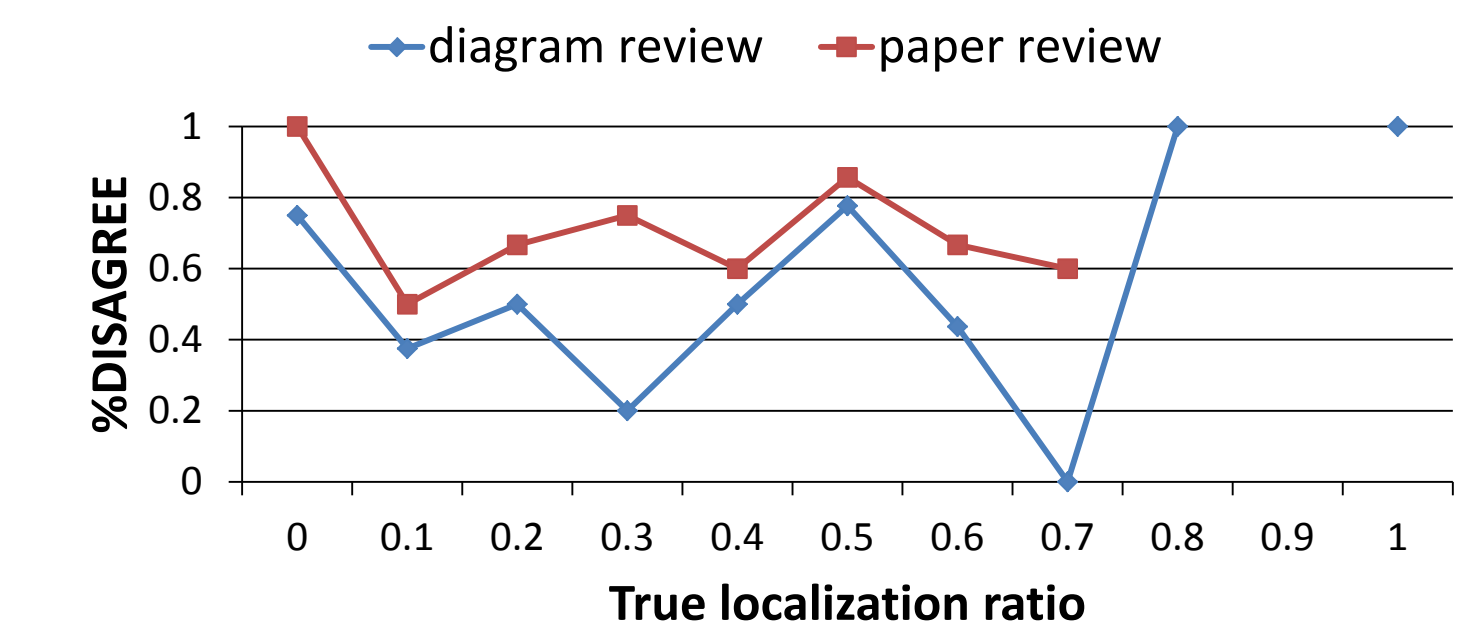
Table 4. Intervention accuracy.

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

## Q2: Student response analysis (to intervention of first submissions)

- Despite system's high intervention accuracy: *student reviewers disagreed more than they agreed with system scaffolded feedback.*
  - Diagram review:** 52% DISAGREE v. 48% REVISE (of 113 first interventions)
  - Paper review:** 70% DISAGREE v. 30% REVISE (of 43 first interventions)

- Student disagreement is not related to how well the original reviews were localized.*



## Q3: Review revision analysis (to intervention of first submissions)

- Classify intervention into scopes
  - Scope=IN:** intervened at current diagram/paper
  - Scope=OUT:** intervened at some prior diagram/paper but not current diagram/paper
  - Scope=NO:** never received intervention
- For resubmissions of each scope, collect edited comments and compare true localization labels {Yes/No} to that of prior comments.

Table 5. Comment change patterns by intervention scopes.

Scope	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, No→No** contributes the largest portion: *potential improvement in our scaffolding of review localization.*
- Scope=OUT**, relatively large numbers of *No→Yes* suggest that *impact of scaffolding feedback remains in later reviewing sessions.*
- Scope=NO**, large numbers of *Yes→Yes* may show that *those reviewers might revise their reviews for some reason other than localization.*

## Conclusions and future work

- Enhanced a peer review system by integrating two review localization models and implementing scaffolding intervention.
- Model performance was high at both comment level and submission level.
- Scaffolding feedback helped student localize their comments, even in later non-scaffolded review sessions.
- In future we aim to address current limitations:
  - Large number of unsuccessful attempts to localize comments: improve user interface, highlight localization text
  - Large number of student disagreements: user study