Modeling Programming Learning in Online Discussion Forums

I-Han Hsiao
School of Computing, Informatics & Decision Systems Engineering,
Arizona State University, 699 S. Mill Ave., Tempe AZ, U.S.A.

Keywords: Learning Activity, Engagement Activity, Discourse Analysis, Constructive Learning, Discussion Forum, Programming, Educational Data Mining, Semantic Modelling, Topical Focus, Learning Assessment.

Abstract: In this paper, we modelled constructive engagement activities in an online programming discussion. We built a logistic regression model based on the underlined cognitive processes in constructive learning activities. The findings supported that there is *passive-proactive* behaviour and suggested that detecting constructive content can be a helpful classifier in discerning relevant information to the users and in turn creating opportunities to optimize learning. The results also confirmed the value of discussion forum content, disregarding the crowd approves or not.

1 INTRODUCTION

With the rapid growth of free, open, and large user-based online discussion forums, it is essential for education researchers to pay more attention to emerging technologies that facilitate learning in cyberspace. In programming, these free online discussion sites (i.e. stackoverflow: http://stackoverflow.com, Dream.In.Code: http://www.dreamincode.net, etc.) are popular trouble-shooting/problem-solving technologies for online courses. They allow programmers and learners to reach out for help so that they can freely discuss programming problems, ranging from general to specific and simple to complex topics. These sites therefore not only throw open unbounded topics in the form of questions and answers, but are especially attractive for open-ended problem discussions. Over the decades, discourse analysis on discussion forums has been carried out through various formats, such as network analyses, topical analyses, interactive explorers, knowledge extraction, semantic connections etc. (Dave, Wattenberg, and Muller, 2004; Gretarsson et al., 2012; Indratmo, Vassileva, and Gutwin, 2008; Lee, Kim, Cho, and Woo, 2013; Shum, 2008; Wei et al., 2010). However, the scale and types of posts are often very diverse in terms of user background, coverage of topics, post volumes, post-response turnaround rates, etc. It is a typical "open corpus" challenge, where content sources are diverse and usually unbounded; therefore it is challenging to estimate student's knowledge and further provide personalized support. In addition, these platforms are usually not moderated or guided by teachers or teaching assistants, but are essentially governed by the community. There has been considerable research on strategies to filter the quality of content and encourage participation of online communities via crowdsourcing, rating, tagging, badges, etc. (Hsiao and Brusilovsky, 2011; Jeon, Croft, and Lee, 2005; Kittur, Chi, and Suh, 2008; Snow, O'Connor, Jurafsky, and Ng, 2008). Such social mechanisms tend to filter and point out the most possible correct solutions. However, in the context of online learning, the correct solutions may not necessarily be the best next steps for all learners (Graesser, VanLehn, Rose, Jordan, and Harter, 2001; van de Sande and Leinhard, 2007). The majority of the online large-scale discussion forums investigate in content quality and management; this work aims to centre on understanding how people learn from these online discussion forums.

The juncture of Intelligent Tutoring Systems/Artificial Intelligence in Education (ITS/AIED) and Learning Science/Computer Supported Collaborative Learning (LS/CSCL) literature has successfully demonstrated that students can learn from a wide range of dialogue-based instructional settings, such as dialogic-based tutor, asynchronous discussion forums, etc. (Aleven, McLaren, Roll, and Koedinger, 2006; Aleven, Ogan, Popescu, Torrey, and Koedinger, 2004; Boyer et al.,

Hsiao I..
Modeling Programming Learning in Online Discussion Forums.
DOI: 10.5220/0005452002530259
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 253-259
ISBN: 978-989-758-108-3

2011; Chi, 2009; Chi, Roy, and Hausmann, 2008; Muldner, Lam, and Chi, 2014; VanLehn et al., 2007). Recently, studies show an alternative instructional context by *learning from observing others learn* (Chi et al., 2008) and is considered as a promising learning paradigm (Muldner et al., 2014). It suggests passive participants (such as lurkers who consume content without contributions) can still learn by reading the postings-and-replies exchanges from others due to the constructive responses in the content (Chi and Wylie, 2014). Such *learning-from-observing* paradigm addresses a major limitation on development time in ITSs and liberated the domains from procedural skills to less structured fields. However, to what extend can we capitalize such learning activity: reading others' constructive dialogues voluntarily and engage in some sort of learning activity after that? In the context of programming learning, can we successfully model users' learning activities in such large-scaled open corpus environment? In this paper, we focus on modelling such behaviour and exploring the associated learning activities in an online programming discussion forum.

2 LITERATURE REVIEW

In modelling learning activities, Wise, Speer, Marbouti, and Hsiao (2013) studied an invisible behaviour - listening behaviour in online discussions, where the participants are students in a classroom instructed to discuss tasks on the platform. (Sande, 2010; van de Sande and Leinhard, 2007) investigated online tutoring forums for homework help, making observations on the participation patterns and the pedagogical quality of the content. (Hanrahan, Convertino, and Nelson, 2012; Posnett, Warburg, Devanbu, and Filkov, 2012) studied expertise modelling in similar sort of discussion environment. (Goda and Mine, 2011) quantify online forum comments by time series (*Previous*, *Current* and *Next*) to infer the corresponding learning behaviours. The ICAP (Interactive, Constructive, Active, Passive) learning activity framework defines "learning activities" as a broader and larger collection of instructional or learning tasks, which allows educational researchers to explain subtle engagement activities (invisible learning behaviours) (Chi, 2009; Chi and Wylie, 2014; Muldner et al., 2014). The framework examines comparable learning involvement, where *Interactive* modes of engagement achieve the greatest level of learning, then the *Constructive* mode, then the *Active* mode, and finally, at the lowest level of learning, the *Passive* mode. This allows prediction of learning outcome and estimation of knowledge transformation. However, effective evaluation and harnessing of students' learning activities usually relies on qualitative human-coded methods (i.e. domain expert judges), which is typically difficult to scale and challenging to keep persistent traces of for current knowledge prediction (Blikstein, 2011). In addition, crucial learning moments can be easily missed and difficult to reuse. We are beginning to see more data driven approaches attempting to address these problems (Hsiao, Han, Malhotra, Chae, and Natriello, 2014; Rivers and Koedinger, 2013).

3 METHODOLOGY

To model learning activity in an online programming discussion forum, we have to firstly analyse forum content by extracting features in presenting content corresponding constructive engagement activities. We consider two dimensions of features 1) Social aspects features, including posting votes, poster reputation, poster status (Stack Overflow utilizes gamification mechanism, which allows community members to vote and gain badges in reflecting community status (i.e. gold, silver, bronze, etc.)) and the number of favourites bookmarked by users; 2) Content related features, including code snippets, content syntactic (length, average sentence per thread, novelty terms), content semantics (sentiment polarity, topic entropy, topic coherence, topic complexity, concept entropy) and most importantly, the constructive lexicons. We define the value to provoke learning as *constructiveness* based on the constructive lexicon. According to ICAP learning activity framework (Chi and Wylie, 2014), a constructive learning activities include the following possible underlying cognitive processes, *inferring, creating, integrating new with prior knowledge, elaborating, comparing, contrasting, analogizing, generalizing, including, reflecting on conditions, explaining why something works*. Based on these cognitive processes, we build a constructive lexicon library to capture comparing and contrasting words, explanation, and justification and elaboration words. We extract comparing and contrasting keywords from a comparative sentence dataset, which was originally used in sentiment analysis for detecting and comparing product features in reviews (Ganapathibhotla and Liu, 2008). For example, comparative or superlative adjective

Table 1: Overview of Features.

Feature	Description
	Social Features (SF)
Vote	Community democracy to evaluate content quality based on up or down votes
Reputation	Community trust measurement based on user's previous activities on the site, including up-voted questions and answers, answer acceptance
Status	The accumulated scores on user profile to symbolize the amount of work done in the community. i.e. Gold indicates important contributions; silver indicates strategic questions or answers; bronze shows rewards for participation
Favourite	Number of saved bookmarks by the community
	Content Features (CF)
Code length	Number of code lines
Concept count	Number of code concepts parsed by programming language parser (Hosseini and Brusilovsky, 2013)
Code Concept entropy	Code topic distribution among all codes to measure community code topic focus
Post length	Number of words of the post
Post entropy	Post topic distribution to measure community topic focus, where post topics are generated by TFM (Hsiao and Awasthi, 2015)
Sentiment polarity	Positive and negative sentiments of the content based on a list of positive and negative sentiment words in English (Hu and Liu, 2004) Polarity = #(PosTerm) + #(NegTerm)
Topic coherence	UMass score is measured as pairwise score to represent how much a word in a post triggers the corresponding concept. (Mimno, Wallach, Talley, Leenders, and McCallum, 2011) $\mathrm{UMass}(w_t, w_p) = log\frac{P(w_t, w_p)}{P(w_p)}$
Novelty	Novelty words (w) of a post (p) compared to other post of the same question. Informativeness is calculated by $\Sigma_{w \in p} tfidf(w, p)$
	Dependent Variable
Constructiveness	The number of constructive word counts based on the constructive lexicon described above

and adverb words, such as versus, unlike, most etc. We then modify an arguing lexicon to extract explanation, justification and elaboration words (Somasundaran, Ruppenhofer, and Wiebe, 2007). We focus on the *assessment, emphasis, causation, generalization,* and *conditionals* sentence patterns and include WH-type and punctuation features in generating associated constructive lexicons. For instance, "in my understanding…", "all I'm saying is…" (*assessment*), "…this is why…(*emphasis*)", "…as a result…(*causation*)", "…everything…(*generalization*)" and "…it would be…(*conditionals*)". Table 1 presents an overview of all features.

4 EVALUATION

According to the engagement activity framework reviewed above, we construct the learning activity model based on the features identified. We then further analysed the forum content semantics in examining the validity of the findings from the results discovered from the model.

4.1 Data Collection

We sampled one year (year 2013) of forum posts in topic Java from StackOverflow site through StackExchange API. Stack Exchange (http://stackexchange.com) is a question and answer website network for various fields. The data pool was selected from the top 10 frequent tagged questions due to most the posts in this section contained at least one accepted answer. It will allow us to build a baseline on the answer quality according to crowdsourced votes. There are total 16,739 posts, including 3,725 questions, 13,014 answers, with 3,718 accepted answers.

4.2 Model Learning Activity Analysis

To capture whether the observed assumptions on the features would account for the variation in user engagement prediction, we performed logistic regression analysis. The full model was able to successfully predict *constructiveness* at 0.001 level, *adjusted-*R^2= 0.6496. We tested the goodness of the models reserving 20% of the observations for testing with 10-fold cross validation (MAE_{10FOLD}= 7.08) and selected a final model.

We found that there are significant more constructive words within Accepted Answer (M= 0.827, SE= 1.334) than Answers (M=0.583, SE= 1.005), p< 0.01 (Table 3). The result confirmed that the answers accepted by the crowd not only agreed as correct solutions among the best available answers, but also contained higher constructive

Table 2: The logistic regression model on Constructiveness.

Feature	Coefficient
SF-vote	6.900
SF-reputation	9.587*
SF-gold	-3.866
SF-silver	-4.269
SF-bronze	3.527
SF-favourite	1.028
CF-code_length	9.761
CF-concept	-1.555***
CF-code_entropy	2.841**
CF-post_length	4.154***
CF-post_entropy	2.897
CF-polarity	1.205***
CF-coherence	-1.895**
CF-novelty	7.852***
constant	-2.255(.)

Significance codes: 0****, 0.001**, 0.01*, 0.05(.)

information. Accepted Answers also showed a positive correlation between user favourites and the amount of constructive words (r= 0.0781, p< 0.01), but we did not see such correlation between Questions/Answers and the amount of constrictive words. This result is not surprising. It indicates the community tends to bookmark useful Accepted Answers, but not Questions nor Answers. However, we found the community provided as many votes to Answers and Accepted Answers, no matter how constructive the content were. This observation was very interesting and revealed that the community may not bookmark the Answers as frequent as they do to Accepted Answers, but it did show the effort to screen the Answers and provide votes to them.

We further divided the content into two categories, Easy and Difficult (based on the topics covered in CS1 or CS2 courses). Easy topics include *Classes, Objects, Loops, ArrayLists* etc.; difficult topics contain *Inheritance, Recursion, Multithreading, User Interfaces* etc. We found that easier content had slightly higher constructive words than difficult content, but it was not significant. It was understandable that simpler problems may be easier to provide examples and tougher problems may require more efforts to justify the answers. However, we found that among Answers, users bookmarked more and up voted more in difficult content when the content had also more constructive words. But we saw no such pattern in Accepted Answers or in Questions. This again showed important evidence that the users in the community spending efforts in locating relevant information to themselves, even the answers are not accepted by the crowd. These results suggested that there was a *passive-proactive* learning behaviour, which users did not just read the Accepted Answers, but also Answers, and further provided some sort of actions (up voted, bookmarked etc.) The findings also suggested that detecting constructive content could be a helpful classifier in discerning relevant information to the users, and in turn providing learning opportunities.

Table 3: Constructive word counts by content types and difficulties.

Topic/Type	Question	Accepted Answer	Answer
Easy	0.956±1.253	0.959±1.385	0.646±1.035
Difficult	0.984±1.355	0.827±1.294	0.583±0.981
Average	0.971±1.309	0.827±1.334	0.583±1.005

4.3 Semantic Content Analysis

From learning activity model analysis we learn that there are learning opportunities in utilizing discussion forum content and not limited to the crowd accepted content only. To further understand why and how people can benefit from the content (not just the Accepted Answers, but also the Answers), we analysed the forum content semantics.

We recognize that programming discussion forums are places for users to solve or to search for code solutions. The forum posts consist of combination of natural language posts and programming codes. Therefore, to extract content semantics, it requires two different semantic parsers. For natural language forum post texts, we applied Topic Facet Modelling (TFM) algorithm to extract concepts from forum texts into corresponding sets of topics (Hsiao and Awasthi, 2015). For programming codes, we used the program code parser (Hosseini and Brusilovsky, 2013) to obtain the code semantics. TFM is a modified Latent Dirichlet Allocation (LDA) probabilistic topic model, which automatically detects content semantics in conversational and relatively short texts. It is fully explained and reported in (Hsiao and Awasthi, 2015).

After extracting all the content semantics, we applied Shannon entropy (1) to gauge the content topical focus (Momeni, Tao, Haslhofer, and Houben, 2013; Wagner, Rowe, Strohmaier, and Alani, 2012). We calculated the distance topic distribution of each post (text and codes separately). We define entropy of topic distribution of the forum post authored by the user, u. Where t is a topic and n is #topics. Low topic entropy indicates high focus. We assume the topical focus of posts has influence on the usefulness of content for learning.

We found that post texts had consistent topical focus across three different content categories, and program codes yielded higher topical focus than post texts. This is understandable due to the reason that people often come to the programming discussion forums to look for code solutions. Most importantly, we found the codes in Answers generated the highest topical focus than any other content type and content categories. It demonstrated the value of massive Answers in the discussion forum, even the content are not *approved* by the crowd as the Accepted Answers. Possible explanations could be, while the Answers may not be the best solutions to the questions, they can still be the most appropriate resource for the viewer. Because the person who ends up browsing the content can have his/her questions in mind, which are not exactly the same or fully expressed as the questions presented in the forum. Such findings again demonstrated the value of the forum content, which can be resourceful learning objects even they are not crowd approved.

$$Entropy(\hat{u}) = -\sum_{j=1}^{n} p(t_{i,j}) \log_e p(t_{i,j}) \quad (1)$$

Table 4: Text and code entropy by content types.

Content Category/Type	Text	Code
Question	4.302±0.251	2.316±2.165
Accepted Answer	4.179±0.554	2.455±2.110
Answer	4.108±0.711	**1.758**±2.085

5 DISCUSSION

In this paper, we modelled constructive engagement activities in an online programming discussion. We built a constructive word lexicon based on constructive learning activities underlined cognitive processes described in the ICAP learning activity framework. We then performed logistic analysis and selected a model, which was able to explain 64.96% of users' engagement activities. Deeper analysis confirmed that the crowd perceived Accepted Answers were likely to contain more constructive words. Moreover, users had more up votes interactions with Answers and Accepted Answers disregard the quantity of constructive words. Besides, they especially bookmarked more and up voted more in difficult Answers when the content had also more constructive words. In addition, in the semantic content analysis, we found higher topical focus of the program codes in Answers in the discussion forum. This again demonstrated the value of discussion forum content, no matter the crowd approves the content or not.

All these findings combined together suggested the existence of *passive-proactive* in large-scaled online discussion forum and the content of the discussion forum are valuable assets for learning, disregarding the acceptance by the crowd or not. It also suggested that detecting constructive content could be a helpful classifier in discerning relevant information to the users, and in turn providing learning opportunities. For instance, we can optimize learning opportunities in the open corpus large-scaled discussion forum by identifying and ordering content based on the quality and constructiveness, which may result in better efficiency for mass *passive-proactive* users. (As oppose to traditional layout of the content, which is ordered by the content quality and reversed chronological order.) Similarly, the value of the Answers in the massive amount of discussion forums should be harnessed and better utilized. For example, recommend relevant Answers to learners, instead of Accepted Answers.

6 LIMITATION AND FUTURE WORK

We recognized two major limitations during the exploratory modelling process. 1) We currently only considered the constructive learning activity, and neglected other activities, such as *Interactive* learning activity. Learning is complex. All sorts of learning activities can be intertwined among the same context. 2) Current model considered limited social features to capture users' profiles. We believe that a learning-inductive post should also take into account the content poster's expertise, rather than just the amount activities in the community. Therefore, in the future, we plan to integrate other learning activities associated with constructive ones and conduct more rigorous evaluation in modelling forum posters' expertise. Moreover, we are currently testing innovative learning analytics interfaces, which present personalized views, sequencing, and summaries in assisting users to better use of the massive content from discussion forums. More exhausted user studies are planned to evaluate predictive model effectiveness.

REFERENCES

Aleven, V., McLaren, B., Roll, I., and Koedinger, K. (2006). Toward Meta-cognitive Tutoring: A Model of

Help Seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16(2), 101-128.

Aleven, V., Ogan, A., Popescu, O., Torrey, C., and Koedinger, K. (2004). Evaluating the Effectiveness of a Tutorial Dialogue System for Self-Explanation. In J. Lester, R. Vicari and F. Paraguaçu (Eds.), Intelligent Tutoring Systems (Vol. 3220, pp. 443-454): Springer Berlin Heidelberg.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. Paper presented at the Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, Canada. http://dl.acm.org/citation.cfm?doid=2090116.2090132.

Boyer, K. E., Phillips, R., Ingram, A., Ha, E. Y., Wallis, M., Vouk, M., and Lester, J. (2011). Investigating the Relationship Between Dialogue Structure and Tutoring Effectiveness: A Hidden Markov Modeling Approach. International Journal of Artificial Intelligence in Education, 21(1), 65-81. doi: 10.3233/JAI-2011-018.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. Topics in Cognitive Science, 1(1), 73-105. doi: 10.1111/j.1756-8765.2008.01005.x.

Chi, M. T. H., Roy, M., and Hausmann, R. G. M. (2008). Observing Tutorial Dialogues Collaboratively: Insights About Human Tutoring Effectiveness From Vicarious Learning. Cognitive Science, 32(2), 301-341. doi: 10.1080/03640210701863396.

Chi, M. T. H., and Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educational Psychologist, 49(4), 219-243. doi: 10.1080/00461520.2014.965823.

Dave, K., Wattenberg, M., and Muller, M. (2004). Flash forums and forumReader: navigating a new kind of large-scale online discussion. Paper presented at the Proceedings of the 2004 ACM conference on Computer supported cooperative work, Chicago, Illinois, USA. http://dl.acm.org/citation.cfm?doid=1031607.1031644.

Ganapathibhotla, M., and Liu, B. (2008). Mining opinions in comparative sentences. Paper presented at the Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1.

Goda, K., and Mine, T. (2011). Analysis of students' learning activities through quantifying time-series comments. Paper presented at the Proceedings of the 15th international conference on Knowledge-based and intelligent information and engineering systems - Volume Part II, Kaiserslautern, Germany.

Graesser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., and Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. AI magazine, 22(4), 39. doi: http://dx.doi.org/10.1609/aimag.v22i4.1591.

Gretarsson, B., O, J., Donovan, Bostandjiev, S., H, T., #246, Smyth, P. (2012). TopicNets: Visual Analysis of Large Text Corpora with Topic Modeling. ACM Trans. Intell. Syst. Technol., 3(2), 1-26. doi: 10.1145/2089094.2089099.

Hanrahan, B. V., Convertino, G., and Nelson, L. (2012). Modeling problem difficulty and expertise in stackoverflow. Paper presented at the Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion, Seattle, Washington, USA.

Hosseini, R., and Brusilovsky, P. (2013). JavaParser: A Fine-Grained Concept Indexing Tool for Java Problems. Paper presented at the AIEDCS workshop Memphis, USA. .

Hsiao, I.-H., and Awasthi, P. (2015) Topic Facet Modeling: Visual Analytics for Online Discussion Forums. Paper presented at the The 5th international Learning Analytics and Knowledge Conference, Marist College, Poughkeepsie, NY, USA.

Hsiao, I.-H., and Brusilovsky, P. (2011). The Role of Community Feedback in the Student Example Authoring Process: an Evaluation of AnnotEx. British Journal of Educational Technology, 42(3), 482-499. doi: http://dx.doi.org/10.1111/j.1467-8535.2009.01030.x.

Hsiao, I.-H., Han, S., Malhotra, M., Chae, H., and Natriello, G. (2014). Survey Sidekick: Structuring Scientifically Sound Surveys. In S. Trausan-Matu, K. Boyer, M. Crosby and K. Panourgia (Eds.), Intelligent Tutoring Systems (Vol. 8474, pp. 516-522): Springer International Publishing.

Hu, M., and Liu, B. (2004). Mining and summarizing customer reviews. Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

Indratmo, Vassileva, J., and Gutwin, C. (2008). Exploring blog archives with interactive visualization. Paper presented at the Proceedings of the working conference on Advanced visual interfaces, Napoli, Italy. http://dl.acm.org/citation.cfm?doid=1385569.1385578.

Jeon, J., Croft, W. B., and Lee, J. H. (2005). Finding similar questions in large question and answer archives. Paper presented at the Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany. http://dl.acm.org/citation.cfm?doid=1099554.1099572.

Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy. http://dl.acm.org/citation.cfm?doid=1357054.1357127.

Lee, Y.-J., Kim, E.-K., Cho, H.-G., and Woo, G. (2013). Detecting and visualizing online dispute dynamics in replying comments. Software: Practice and Experience, 43(12), 1395-1413. doi: 10.1002/spe.2153.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, United Kingdom.

Momeni, E., Tao, K., Haslhofer, B., and Houben, G.-J.

(2013). Identification of useful user comments in social media: a case study on flickr commons. Paper presented at the Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, Indianapolis, Indiana, USA.

Muldner, K., Lam, R., and Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. Journal of Educational Psychology, 106(1), 69-85. doi: 10.1037/a0034448.

Posnett, D., Warburg, E., Devanbu, P., and Filkov, V. (2012). Mining Stack Exchange: Expertise Is Evident from Initial Contributions. Paper presented at the Social Informatics (SocialInformatics), 2012 International Conference on, Lausanne.

Rivers, K., and Koedinger, K. (2013). Automatic Generation of Programming Feedback: A Data-Driven Approach. Paper presented at the Workshops at the 16th International Conference on Artificial Intelligence in Education AIED.

Sande, C. v. d. (2010). Free, open, online, mathematics help forums: the good, the bad, and the ugly. Paper presented at the Proceedings of the 9th International Conference of the Learning Sciences - Volume 1, Chicago, Illinois.

Shum, S. B. (2008). Cohere: Towards web 2.0 argumentation. COMMA, 8, 97-108.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii.

Somasundaran, S., Ruppenhofer, J., and Wiebe, J. (2007). Detecting arguing and sentiment in meetings. Paper presented at the Proceedings of the SIGdial Workshop on Discourse and Dialogue.

van de Sande, C., and Leinhard, G. (2007). Online tutoring in the Calculus: Beyond the limit of the limit. education, 1(2), 117-160.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., and Rosé, C. P. (2007). When Are Tutorial Dialogues More Effective Than Reading? Cognitive Science, 31(1), 3-62. doi: 10.1080/03640210709336984.

Wagner, C., Rowe, M., Strohmaier, M., and Alani, H. (2012). What Catches Your Attention? An Empirical Study of Attention Patterns in Community Forums. Paper presented at the ICWSM.

Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M. X., Qian, W., Zhang, Q. (2010). TIARA: a visual exploratory text analytic system. Paper presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA. http://dl.acm.org/citation.cfm?doid=1835804.1835827.

Wise, A., Speer, J., Marbouti, F., and Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. Instructional Science, 41(2), 323-343. doi: 10.1007/s11251-012-9230-9.