Open Web Annotation as Connected Conversation in CSCL

Abstract: Research has yet to explore how the social and technical affordances of open web annotation (OWA) can mediate connections between educators in service of their professional learning. This study examined educator participation in the Marginal Syllabus, a computer supported collaborative learning environment that encouraged connected conversation via OWA. Multiple quantitative methods, including text sentiment analysis and social network analysis, were used to discern key discursive characteristics among the nine conversations of the 2016-17 Marginal Syllabus (1,163 annotations authored by 67 participating educators). Key discursive characteristics of educators' connected conversations include: (a) generally positive sentiment; (b) educators who annotated most prolifically also authored the greatest percentage of annotations with neutral sentiment; and (c) conversations of at least four annotations tended to demonstrate a greater percentage of negative sentiment. The sentiment trends and study limitations are addressed in the final discussion.

Annotation as Connection

From handwritten scholia atop Homer's *Iliad* to layered Talmudic commentary, the act of annotation has notable historical roots that precede our digital era. Annotation, according to Sanderson and de Sompel (2011) is "a pervasive activity shared by all humanity across all walks of life" (par. 1). Inspired by the centuries-old practice of adding marginalia to printed texts, the need for digital forms of annotation appeared in Berners-Lee's (1989) proposal for an information system linked by hypertext – what we know as the World Wide Web – and was a feature of Mosaic, the first web browser. Today, web annotation allows a reader to comment upon, correct, highlight, and categorize online text. According to Udell (2017): "If we think of the web we've known as a kind of fabric woven together with links, the annotated web increases the thread count of that fabric. When we weave with pieces of URL-addressable documents, we can have conversations about those pieces, we can retrieve them, we can tag them, and we can interconnect them" (par. 11). Web annotation serves as a means of connection; more specifically, it can serve as a process mediating connected conversation among people, their ideas, and fine-grained linkages to online content like phrases, sentences, and data.

Today, connected conversations mediated by web annotation appear in journalism, legal education, scientific research, and scholarly publication. Within the field of education, a growing body of research has examined web annotation in relation to students' reading comprehension and critical thinking, the development of domain-specific knowledge, and as a form of collaboration (i.e. Johnson, Archibald, & Tenenbaum, 2010; McNutt, 2014; Su, Yang, Hwang, & Zhang, 2010). However, educational research has yet to robustly examine how the social and technical affordances of web annotation might mediate connections between educators in service of their interest-driven and professionally-relevant learning. Under what conditions might educators leverage web annotation to have connected conversations about their teaching practices and professional interests? What would be the discursive qualities of this conversation? And, ultimately, why would such conversation matter? Given the relevance of connected learning (Ito et al., 2013) to both educators' classroom practices (i.e. Garcia, 2014) and emerging models of connected teaching (Mirra, 2017), this study examines educators' connected conversation via open and collaborative web annotation as a form of professional learning.

Open Web Annotation Mediating Educator Connected Conversation via CSCL

Given growing interest in open educational movements, practices, and values (i.e. Cronin, 2017; Wiley, 2016), as well as the proliferation of open educational resources (OER) in both K-12 and higher education settings (i.e. Baker, Asino, Xiu, & Fulgencio, 2017; Hilton, 2016), we examine educators' connected conversation as mediated by *open web annotation* (OWA; Author, in review). OWA is defined by a standardized technical architecture (Web Annotation Working Group, 2017), interoperability, and open-source software. Moreover, OWA content may be publicly licensed (i.e. Creative Commons attribution) and, like blogging or online social networking, may encourage an "ethos of transparency" (Havemann, 2016) about open educational practices. While there is growing interest in and use of web annotation in educational contexts (i.e. Novak, Razzouk, & Johnson, 2012), the use of OWA remains scare (an exception is Chen, 2018) and, hence, is a motivator of this study.

This study leverages OWA to help architect a computer supported collaborative learning (CSCL) environment for educator learning. According to Kirschner and Erkens (2013), the three key elements of CSCL are *pedagogical*, *social*, and *technological*. Pedagogy, in CSCL, is intended to support cognitive and metacognitive tasks; by authoring open annotations, for example, an individual can make their thinking available to others, aiding the visibility of group cognition (Stahl, 2006) and public accessibility of knowledge resources. The social element of CSCL refers to collaborative interaction; while OWA systems are not, by definition, social networks, OWA-enabled learning environments can become a social medium, as evident by OWA use in academic communities of practice (Author, in review). The final element of CSCL is technological, or the computer supports that "facilitate the collaborative cognitive and social learning processes" (Kirschner & Erkens, 2013, p. 3). The technical affordances of OWA can create an "anchored environment" (Gao, 2013) for collaboration that supports multimodal expression, the additional of descriptive metadata (i.e. tags), and the curation of and cross-context linkages among distributed resources. More specifically, the OWA platform featured in this study, Hypothes.is, includes various technical affordances that facilitate collaboration, including replies, sharing individual annotations via social media, and public annotations attributed with a Creative Commons license to be reused by others. In consideration of these pedagogical, social, and technological elements, this study builds upon research about CSCL environments supporting educator learning (i.e. Lockhorst, 2004) and makes a novel contribution to the CSCL literature by investigating the role of OWA in mediating educator learning via connected conversation.

Study Context: The Marginal Syllabus as a CSCL Environment

The context of this study is a CSCL environment for educator learning that has been iteratively developed and implemented by a multi-stakeholder partnership among: University researchers; K-12 educators and administrators; authors and publishers of academic content; Hypothes.is, a non-profit organization building an open-source web annotation platform; and the National Writing Project, the nation's premier literacy education organization that provides professional development for educators. Launched during the 2016-17 academic school year, the Marginal Syllabus is a CSCL initiative that sparks and sustains conversations about educational equity via OWA. The project embraces a political and technical double entendre; the Marginal Syllabus partners with authors whose writing may be considered marginal – or contrary – to dominant education norms, and online conversations among educators and authors occur in the margins of digital texts using web annotation.

The 2016-17 syllabus featured nine conversations with ten partner authors about topics like critical literacy, curricular co-design, the business of educational technology, and culturally relevant pedagogy. Sixty-three educators participated in these public conversations by generating a corpus of 1,163 annotations. The 2017-18 syllabus, currently hosted by the National Writing Project, was organized around the theme "Writing Our Civic Futures;" it features eight conversations with 12 partner authors. Thirty-nine educators have, to date, participated in the first four public conversations of the 2017-18 syllabus by generating a corpus of 445 annotations. The Marginal Syllabus initiative includes: a project website; two sets of curated, open access digital texts contributed by multiple authors and scholarly publishers; public layers of Hypothes.is OWA dialogue; blogs authored by project partners; and webinars (most featuring text authors) hosted by the National Writing Project.

Research about web annotation in educational contexts has seldom focused on teacher learning. In response, this study suggests OWA may be a promising means of creating open, publicly-accessible CSCL environments within which educators might exercise political agency through connected conversation, question dominant schooling narratives, and critique inequitable educational practices (Blinded, 2018). The central research question guiding this study asks: *What are key discursive characteristics of educators' connected conversation as mediated by OWA in a CSCL environment?*

Methodology: Examining Connected Conversation

A persistent methodological challenge in CSCL studies is the definition of a unit of analysis to meaningfully describe participatory patterns. Stahl (2006) stated the problem simply: "I work and I learn in innumerable ways and modes," and added, "Working and learning with other people mixes these ways into yet more complex varieties. Technology multiplies the possibilities even more" (p. 3). In a CSCL environment like the Marginal Syllabus, learning not only occurs at the level of the individual educator (i.e. reading and annotating a text), it also occurs with and around a group of educators via their conversation (i.e. annotation replies). Puntambekar and colleagues (2011) suggest the methodological challenge of including individuals and groups in a CSCL study is akin to a sliding

scale: "Smaller segments in data provide finer grained analysis but little contextual information. On the other hand, larger units of analysis help create context but with the loss of detail" (p. xii).

We adopt multiple quantitative methods to examine *connected conversation* as the unit of analysis relevant to educator learning in the Marginal Syllabus as an open CSCL environment. We analyzed multiple discursive characteristics of connected conversation during all nine conversations of the 2016-17 Marginal Syllabus, a corpus of OWA data associated with 1,163 Hypothes.is annotations authored by 67 participants. To systematically investigate patterns of interaction and connected conversation in the Marginal Syllabus, we utilized an array of quantitative methods that included text sentiment analysis, social network analysis, time series analysis, as well as visual representations of data. We then applied these analytic methods at multiple levels of granularity including individual educators, annotation discussion threads, and discussion spanning all nine conversations. This study's quantitative analysis of the entire 2016-17 Marginal Syllabus complements a smaller-scale discourse analysis of a single conversation (August, 2016) that found the Marginal Syllabus helps to architect - across multiple sociopolitical texts and contexts - professionally-relevant learning opportunities for educators that amplify political dimensions of talk (Author & Co-Author, in review).

Sentiment Analysis

The first phase of our analysis measured the sentiment of educators' individual annotations. To perform this analysis, we used a tool called VADER (Valence Aware Dictionary for sEntiment Reasoning), a rule-based model for sentiment analysis, that returns a sentiment value of polarity and intensity. The sentiment value is returned as a continuous number between -1 and 1, with 0 indicating the text has neutral sentiment. VADER outperforms other sentiment analysis techniques in testing social media text (Hutto & Gilbert, 2014). In our assessment, VADER was an appropriate tool given that Hypothes.is annotations and replies resemble social media messages in size. Among the 1,163 Hypothes.is annotations analyzed, the mean word and sentence count per annotation was 37.7 and 2.4, respectively. VADER was also useful regarding the lexicon of the annotation corpus. For example, in measuring sentiment, VADER is capable of accounting for educators' use of emoticons (i.e. :)) and phrases like "+1" to indicate agreement.

Social Network Analysis

Unlike typical social media networks in which users define their own networks by selecting other users with whom they share a connection (Kane et al., 2012), Hypothes.is users form no explicit networks between other users. Rather, Hypothes.is users form networks implicitly through their annotations. In projects like the Marginal Syllabus, a reply to an original annotation becomes a discussion thread; such connections are central to educator interactions via OWA. In order to analyze the structure and patterns of the educators' connected conversations, we used a network analysis tool called NetworkX. Each annotation contains data that determines whether the user is referencing the source document, known as the anchor text, or whether the user is replying to an existing annotation. By using a directed graph - where each annotation is analyzed and represented as a node, and each reply is a directed edge to another node - we are able to analyze conversation threads as tree-like structures generated from Hypothes.is' underlying social network characteristics.

For this second phase of analysis, we focused on *threads* of connected conversation. Threads, in this study, were generated when a participating educator's original Hypothes.is annotation elicited at least one reply from another educator. To initiate this analysis, we discarded all annotations from the nine 2016-17 Marginal Syllabus conversations that did not generate any replies from other participants. As an example, consider a thread of connected conversation initiated by educator *bali* from the October, 2016 conversation; the annotation generated eight replies from five different educators (Table 1, below). Social network analysis of this conversation reveals different discussion patterns of connected conversation mediated by OWA. For example, Figure 1 (below) illustrates: (1) bali showed reciprocity of posts, replying to two educators who had replied to the initial post (left); (2) most annotations had positive sentiment scores (center); and (3) bali's reciprocity applied, in particular, to the first two subsequent posts made within 7.45 minutes. After the initial exchanges of this particular thread, bali did not return to the conversation (right).

Anchor text: *writers should write the books they wish to come upon.*
1 *bali*: Love this. I sense it's not enough but it's sthg i can gst behind [sic]. Should academics try to

write articles in voices and on topics they would wish to read? How do academia and publishing processes get in the way of that? [October 26, 2016 15:57 MDT]

> 1A *[blinded]*: That's a real tension. The so-called objectively of academic research certainly diverges from the motivation of some authors (such as those who write fiction). Paul Beatty, who just won the Booker prize, noted: "I'm just trying to create space for myself. And hopefully that can create space for others." [October 26, 2016 16:03 MDT]

>> 1A1 *bali*: Love that! And something I have thought of before. Persisting in dissent can make room for others to try [October 26, 2016 16:40 MDT]

> 1B *LauraRitchie*: I've just done this! Wrote my first non-fiction book - and I have no clue about the publishing... started sending letters to publishers. ...may be a bumpy road ahead! [October 26, 2016 16:04 MDT]

>> 1B1 *bali*: Wow! Good luck! And bravo [October 26, 2016 16:40 MDT]

>>> 1B1A *[blinded]*: +1 [October 26, 2016 16:44 MDT]

> 1C *BMBOD*: Such important questions. I wish academic publishing was more accessible in this way. [October 26, 2016 16:39]

>> 1C1 *ssimpson_01*: I think this is a beautiful sentiment. Everything is always better with passion; including writing and education. When you are truly passionate about your subject is when you can write something that is truly enjoyable *and* useful to others. #ED677 [February 19, 2017 11:51 MST]

> 1D *traceydean*: As an art teacher, I try to create lessons that I'd want to explore; media/technique/artists #ED677 [February 19, 2017 18:45 MST]

Table 1: Sample Marginal Syllabus thread of connected conversation featuring anchor text, original annotation, and subsequent replies.
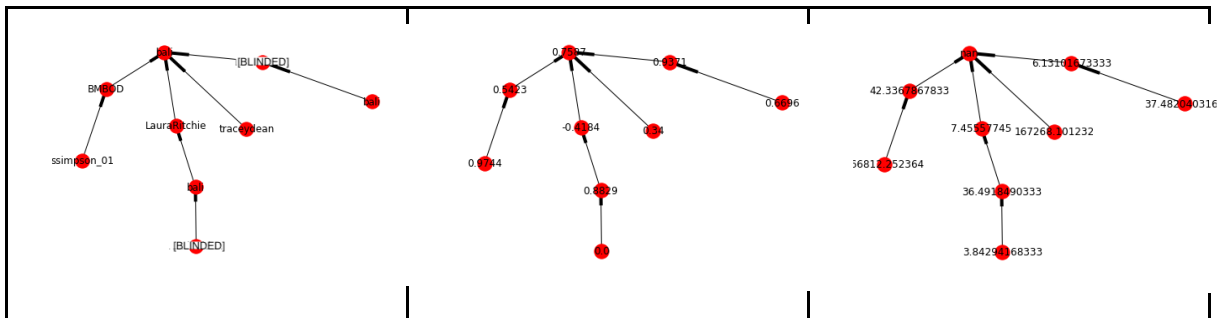

Figure 1: Network graphs of sample Marginal Syllabus thread with different labels, including usernames (L), sentiment score (C), and delay of reply in minutes (R).

In summary, data collection and analysis began at the lowest granularity - individual annotations collected via Hypothes.is. Through our analysis, patterns among connected conversations emerged to provide a perspective on educators' interaction in the Marginal Syllabus as an open CSCL environment. Social network methods provided the analytic tools necessary to better understand the social and collaborative elements of CSCL within the 1,163 Hypothes.is annotations comprising the entire 2016-17 Marginal Syllabus. Both sentiment analysis and social network methods measured patterns of educators' collaborative interaction, allowing us to move from the scale of individual activity in a learning environment to examine broader discursive characteristics of connected conversations that supported educators' professional learning.

Findings

Sentiment analysis of OWA authored by educators during Marginal Syllabus conversations indicates educators' connected conversations were generally positive (59%). Table 2 summarizes all 1,163

annotations of the 2016-17 Marginal Syllabus: The left column reports annotations organized according to three count ranges; "Educator Count" organizes the 67 educators within a range based upon their OWA production; and three columns report the mean sentiment of all annotations.

Annotation Count	Educator Count	Positive	Negative	Neutral
ALL	67	59.0%	9.0%	31.9%
1-4 annotations	26	63.1%	8.0%	28.8%
5-16 annotations	21	56.6%	9.8%	33.6%
16+ annotations	20	56.4%	9.5%	34.1%

Table 2: Sentiment analysis of 1,163 annotations in 2016-17 Marginal Syllabus

Just under one third of educators (20 of 67) authored more than 16 annotations during their Marginal Syllabus participation, about one third of others (21) authored between four and 16 annotations, and just over one third of educators (26) authored less than four, but at least one, annotations. The 20 educators who authored 16+ annotations produced content with sentiment that was, on average, less positive (56.4%) than annotation content authored by other educators. The group of most prolific annotators also authored the greatest percentage of content with neutral sentiment (34.1%), as illustrated by a sample annotation from the October conversation: "I wonder sometimes if some instructors don't make explicit the narrative of their course in part because they think it's obvious, because they put it together. When do we give instructors the time or permission or whatever to say to their students, 'This is why I am assigning these works, in this sequence'" (*dlanclos*).

Trends in the sentiment of educators' connected conversations also appeared in an analysis of thread length, or the number of replies to an original annotation (i.e. eight replies in Table 1 example). Table 3 summarizes all threads in the 2016-17 Marginal Syllabus: The left column reports the number of annotations within a thread; "Thread Count" organizes the 266 threads according to the number component annotations; and three columns report the mean sentiment of annotations within a thread.

Annotation Count	Thread Count	Positive	Negative	Neutral
All	266	58.8%	7.7%	33.5%
2 (1 annotation & 1 reply)	189	60.5%	6.2%	33.3%
3 (1 annotation & 2 replies)	31	50.0%	11.3%	38.7%
>4 (1 annotation & 3+ replies)	46	57.7%	11.6%	30.7%

Table 3: Sentiment analysis of 266 conversation threads in 2016-17 Marginal Syllabus

All threads were generally positive in sentiment (58.8%). Threads that featured three or more replies (17%; 46 of 266), while a minority of the overall dataset, evidenced the greatest percentage of negative sentiment (11.6%). In other words, connected conversations of at least four annotations - despite occurring in only about one fifth of all threads - tended to demonstrate a greater percentage of negative sentiment. The finding that threads with large annotation counts (>4) evidenced a higher percentage of negative sentiment may, in part, be explained by Figure 2 (below), which illustrates that the likelihood of a reply from another educator decreased in relation to more positive sentiment. An annotation with negative sentiment (VADER value -1) was replied to in 48% of all educator interactions, whereas the likelihood of a reply to annotations with both neutral sentiment (0) and positive sentiment (1) dropped to 39% and 37%, respectively. This trend in educator interaction might explain why the 46 conversation threads of greatest length evidenced the highest percentage of annotations with negative sentiment.
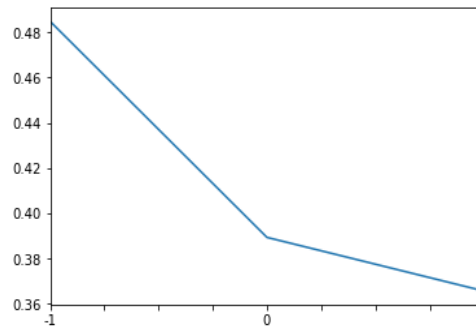
Figure 2: Likelihood of reply to annotation as measured by annotation sentiment.

Discussion

This study has described a number of discursive characteristics associated with educators' connected conversation as mediated by their use of OWA in the Marginal Syllabus, an open CSCL environment organized around conversations about educational equity. In this brief discussion we address two key characteristics educators' connected conversation: (1) sentiment trends in individual educators' annotations; and (2) trends in the sentiment of threads.

First, it is noteworthy that the group of most prolific annotators - nearly a third of Marginal Syllabus participants (20) - authored the greatest percentage of OWA content with neutral sentiment (34.1%). Perhaps these educators' sustained participation in Marginal Syllabus conversations led them, over time, to write more balanced commentary about important topics - such as pedagogical transparency, as indicated in the sample annotation by dlanclos. Alternatively, perhaps this neutral sentiment can be attributed to an awareness of social norms. Given these educators' more pronounced presence in Marginal Syllabus conversations, perhaps neutral annotations were more regularly contributed so as to keep the conversation "moving forward," or so as not to inadvertently offend another participant. Future study of connected conversation in the Marginal Syllabus as a CSCL environment should contextualize such trends in annotation sentiment by, for example, interviewing educators about their intent when authoring annotations, replying to others, and participating in such open learning.

A second notable discursive characteristic was that educators' lengthier connected conversations (4+ annotations) evidenced the largest percentage of annotations with negative sentiment. To better understand this dynamic, we consider Kadushin's (2005) assertion that sentiment affects interactions within networks and small groups and that, generally, "Positive sentiments lead to further interaction and negative sentiments lead to less interaction" (p. 2). Yet this was not the case for some connected conversations in the Marginal Syllabus. However, a more recent study by Backstrom and colleagues (2013) suggests that participation patterns in social network conversations (i.e. commenting on Facebook) are inversely patterned to those that are primarily task-oriented (i.e. editing Wikipedia). This may help us to understand why, as connected conversations in the Marginal Syllabus grew, annotations with negative sentiments made up a larger percentage of those OWA conversation, demonstrating an inverse effect of Kadushin's (2005) socially-focused findings. In other words, sentiment patterns associated with the Marginal Syllabus' longer connected conversations indicate that these interactions among educators may have been more task-oriented (and less social).

Finally, a limitation of this study concerns the methodological challenge of attributing social network characteristics to Hypothes.is, a system that affords social and collaborative interaction via annotation but is not expressly designed as a social network. The social network methods we applied allowed us to identify and pattern educators' participation in the Marginal Syllabus by providing a structure to our unit of analysis - connected conversations - based upon Hypothes.is' underlying data structure of annotations and replies. The social limitations of utilizing such an annotation system for connected conversation are also apparent from a participating educator's perspective; having authored a Hypothes.is annotation, an educator is only notified of a subsequent reply to their annotation via email. This may limit an educator's capacity to sustain or re-enter a conversation as they may be unaware that a social connection has occurred. Such challenges may suggest there may be benefits to facilitating synchronous OWA conversation atop a text, as well as the limitations associated with mediating educator professional learning via open - but primarily asynchronous - CSCL environments.

References

Author. (In review). Blinded for review.

Author & Co-Author. (In review). Blinded for review.

Backstrom, L., Kleinberg, J., Lee, L., & Danescu-Niculescu-Mizil, C. (2013). Characterizing and curating conversation threads: Expansion, focus, volume, re-entry. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* (pp. 13-22). ACM.

Baker, A., Asino, T.I., Xiu, Y., & Fulgencio, J. L. (2017). Logistical Issues With OER Initiative in a K-12 Environment. In M. Mills, & D. Wake (Eds.), *Empowering Learners With Mobile Open-Access Learning Initiatives* (pp. 125-146). Hershey, PA: IGI Global.

Berners-Lee, T. (1989, March). *Information Management: A Proposal*. Retrieved from https://www.w3.org/History/1989/proposal.html

Chen, B. (2018, January 11). Designing for networked collaborative discourse: An UnLMS approach. Retrieved from osf.io/ebjfp

Cronin, C. (2017). Open education, open questions. *EDUCAUSE Review*, 52(6), 11-20.

Garcia, A. (Ed.). (2014). *Teaching in the connected learning classroom*. Irvine, CA: Digital Media and Learning Research Hub.

Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education*, 17, 76–83. http://doi.org/10.1016/j.iheduc.2012.11.002

Havemann, L. (2016). Open educational resources. In M. A. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory*. Singapore: Springer Singapore.

Hilton, J. (2016). Open educational resources and college textbook choices: A review of research on efficacy and perceptions. *Educational Technology Research & Development*, 64(4), 573–590.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Ito, M., Gutierrez, K., Livingstone, S., Penuel, B., Rhodes, ...Watkins, C. (2013). *Connected learning: An agenda for research and design.* Irvine, CA: Digital Media and Learning Research Hub.

Johnson, T. E., Archibald, T. N., & Tenenbaum, G. (2010). Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior*, 26(6), 1496-1507.

Kadushin, C. (2005). Networks and small groups. *Structure and Dynamics*, *1*(1).

Kane, G. C., Alavi, M., Labianca, G. J., & Borgatti, S. (2012). What's different about social media networks? A framework and research agenda. *MIS Quarterly*, 1-69.

Kirschner, P. A., & Erkens, G. (2013). Toward a framework for CSCL research. *Educational Psychologist*, *48*(1), 1-8.

Lockhorst, D. (2004). *Design principles for a CSCL environment in teacher training*. IVLOS.

McNutt, M. (2014). More science in the classroom. *Science*, 346(6212), 895-895.

Mirra, N. (2017, July 31). From connected learning to connected teaching: A necessary step forward. Retrieved from https://goo.gl/jjsWiR

Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education*, *15*(1), 39-49.

Puntambekar, S., Erkens, G., & Hmelo-Silver, C. (Eds.). (2011). *Analyzing interactions in CSCL: Methods, approaches, and issues*. New York, NY: Springer.

Sanderson, R., & de Sompel, H. (2011, August 10). Open annotation: Beta data model guide. Retrieved from http://www.openannotation.org/spec/beta/

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge* (pp. 2-3). Cambridge, MA: MIT press.

Su, A. Y. S., Yang, S. J. H., Hwang, W.-Y., & Zhang, J. (2010). A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. *Computers & Education*, 55(2), 752–766. http://doi.org/10.1016/j.compedu.2010.03.008

Udell, J. (2017, May 5) Weaving the annotated web. Retrieved from https://goo.gl/yagWnn

Web Annotation Working Group. (2017). What are web annotations? In *Web annotation working group*. Retrieved from https://www.w3.org/annotation/

Wiley, D. (2016). Openness as a Value. In P. Blessinger & T.J. Bliss (Eds.), *Open education: International perspectives in higher education*. Cambridge, UK: OpenBook Publishers.

Acknowledgments