Are You Talking to Me? Multi-Dimensional Language Analysis of Explanations during Reading

Laura K. Allen[1], Caitlin Mills[2], Cecile Perret[3], Danielle S. McNamara[3]

[1]Mississippi State University

[2]University of New Hampshire

[3]Arizona State University

Abstract

This study examines the extent to which instructions to selfexplain vs. other-explain a text lead readers to produce different forms of explanations. Natural language processing was used to examine the content and characteristics of the explanations produced as a function of instruction condition. Undergraduate students (n = 146) typed either self-explanations or otherexplanations while reading a science text. The linguistic properties of these explanations were calculated using three automated text analysis tools. Machine learning classifiers in combination with the features were used to predict instruction condition (i.e., self- or other-explanation). The best machine learning model performed at rates above chance (kappa = .247; accuracy = 63%). Follow-up analyses indicated that students in the self-explanation condition generated explanations that were more cohesive and that contained words that were more related to social order (e.g., ethics). Overall, the results suggest that natural language processing techniques can be used to detect subtle differences in students' processing of complex texts

Keywords: Computing methodologies~Natural language processing, Applied computing~Computer-assisted instruction, Applied computing~Psychology, Algorithms, Measurement, Performance, Languages, Theory

Are You Talking to Me? Multi-Dimensional Language Analysis of Explanations during Reading

Laura K. Allen
Mississippi State University
PO Box 872111
Mississippi State, MS, 39762
01+404-414-5200
lka22@msstate.edu

Caitlin Mills
University of New Hampshire
15 Academic Way
Durham, NH, 03824
01+603-862-2360
caitlin.mills@unh.edu

Cecile Perret
Arizona State University
PO Box 872111
Tempe, AZ, 85287
01+404-414-5200
cperret@asu.edu

Danielle S. McNamara
Arizona State University
PO Box 872111
Tempe, AZ, 85287
01+404-414-5200
dsmcnama@asu.edu

ABSTRACT
This study examines the extent to which instructions to *self*-explain vs. *other*-explain a text lead readers to produce different forms of explanations. Natural language processing was used to examine the content and characteristics of the explanations produced as a function of instruction condition. Undergraduate students (n = 146) typed either self-explanations or other-explanations while reading a science text. The linguistic properties of these explanations were calculated using three automated text analysis tools. Machine learning classifiers in combination with the features were used to predict instruction condition (i.e., self- or other-explanation). The best machine learning model performed at rates above chance (kappa = .247; accuracy = 63%). Follow-up analyses indicated that students in the self-explanation condition generated explanations that were more cohesive and that contained words that were more related to social order (e.g., ethics). Overall, the results suggest that natural language processing techniques can be used to detect subtle differences in students' processing of complex texts.

CCS CONCEPTS
• **Computing methodologies~Natural language processing** • *Applied computing~Computer-assisted instruction* • Applied computing~Psychology

General Terms
Algorithms, Measurement, Performance, Languages, Theory

LAK19, March, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303835

ACM Reference Format:
L. K. Allen, C. Mills, C. Perret & D.S. McNamara. 2019. Are You Talking to Me? Multi-Dimensional Language Analysis of Explanations during Reading. In *The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA.* ACM, New York, NY, USA. Article 4, 5 pages. DOI: https://doi.org/10.1145/3303772.3303835

KEYWORDS
Intelligent Tutoring Systems, Natural Language Processing, corpus linguistics, reading, comprehension

1 INTRODUCTION
Learning from text requires a complex set of processes that include understanding the surface-level words and content in a text, as well generating connections amongst these ideas and previously learned material related to the domain [18]. One way that students can enhance comprehension of a topic is by generating inferences to connect ideas within and outside of the text [19]. Prompting students to generate e*xplanation*s promotes inferences by encouraging students to establish connections between new information and their relevant prior knowledge [13]. Within the context of comprehending text, explanation processes help readers activate a stronger network of concepts, which leads to a more coherent mental representation and deeper comprehension of the text [18]. For instance, when readers self-explain, they establish more connections between ideas in the text and improve their understanding of complex topics [4][15]. Similar benefits have been observed when readers explain information to others, such as other students or peers. This process prompts readers to synthesize information from the text by relying on their prior domain knowledge [13].

Despite the similar benefits of these two forms of explanation, it is unclear whether self- and other-explanation are driven by the same cognitive processes. Evidence from prior research suggests that students rely less on their own prior knowledge when explaining to a peer and, instead, on the examples and definitions that are already present within the text. [20]. Additionally, students may adapt their explanations based on their perceptions of the audience's knowledge [13]. It may be the case, then, that self- and other-explanation instructions prompt readers to activate different sources of knowledge while reading. Consequently, these differences could influence the cognitive processes employed by readers, as well as their learning from the text.

In the current study, we examined whether instructions to self-explain or other-explain a text led readers to produce different forms of explanations. Natural language processing techniques were used to examine the extent to which these instructions related to the content and characteristics of the explanations that were produced. Linguistic properties of explanations have been shown to be strong proxies of cognitive processes [2]; thus, we hypothesized that potential differences in self- and other-explanation would be reflected in these analyses. In particular, we examined whether the characteristics of readers' explanations at multiple levels of representation differed based on the instructions they were given prior to reading.

In the following section, we provide a general overview of instructional technologies developed to teach comprehension strategies, such as self-explanation. We then describe the iSTART tutoring system, which provides instruction and feedback on self-explanation strategies for high school and college students.

1.1 Educational Technologies

One way that educators have worked to improve the quality of students' comprehension skills is through the development of educational technologies for literacy, which often prompt students to respond to the tutor using natural language. Developers of such systems are more frequently incorporating natural language processing techniques into these systems to increase the systems' adaptivity [9] [10]. In particular, automated analyses from these responses have been used to develop assessments of students' individual differences and performance [3] [8]. For example, McNamara and colleagues (2007) [16] found that natural language processing indices could be used to score the holistic quality of students' self-explanations. Previous research suggests that these interactions with language-based tutoring systems lead to significant learning gains compared to non-interactive learning tasks [21].

Although there have been important developments, these technologies have a number of weaknesses that remain unaddressed. One issue relates to the ability of these systems to measure the *on-line* cognitive processes of student users as they relate to different instructional demands. Educational tutoring systems typically prompt students to type explanations or other responses directly to the computer and, accordingly, it can be unclear how to determine whether students are engaging in the same processes across different situations. Thus, more research is needed to determine the factors that influence students' processes, as well as the ways in which these factors can be automatically detected in educational technologies.

1.2 iSTART

One of the primary aims of this study is to improve the adaptive capabilities of the Interactive Strategy Training for Active Reading and Thinking (iSTART) system, an intelligent tutoring system that teaches high school and college students self-explanation strategies to improve their understanding of texts [17]. iSTART is based on the Self-Explaining and Reading Training (SERT) intervention, which was created to teach students effective strategies for self-explaining a text followed by practice on how to use them as they read [15]. Previous research has demonstrated that both SERT and iSTART are effective at improving students' reading comprehension of texts [14].

iSTART focuses on self-explanation training through two modules: training and practice. Students are taught the strategies in the *training module* through the use of videos narrated by a pedagogical agent. The *practice module* is composed of multiple practice activities in two categories: identification and generative. Identification mini-games are games in which students are prompted to read a text with a self-explanation and then select the specific strategy that was used for that self-explanation. Generative practice is composed of game-based and non-game practice, which prompt students to generate their own self-explanations as they read. Students are given feedback on their response based on a complex algorithm that uses linguistic indices to determine the quality of the response.

The iSTART system relies on an algorithm that automatically scores students' self-explanations after each submission to the system. Analyses are based on the response, the specific target sentence prompting the self-explanation, and the previous sentences of the text. The algorithm uses word-based indices and Latent Semantic Analyses (LSA) to assess the quality of the self-explanation and determine the appropriate feedback. The algorithm produces a score using word-based indices and LSA-indices on a scale from 0 to 3. Research using the iSTART algorithm has shown that it scores as accurately as humans and that it can offer a summary of the cognitive processes used in reading comprehension [11].

Recent research with iSTART explanations has found that natural language processing techniques can be used to determine much more about the learning environment beyond holistic score. For example, studies have indicated that indices of local and global cohesion relate to students' comprehension and learning from the text [1] [3]. Thus, the purpose of these studies – and the current study – is to determine how these linguistic signatures can provide the iSTART system with information that can be used to enhance individualized training.

1.3 Current Study

The purpose of the current study is to investigate the degree to which the processes in which students engage during reading comprehension can be examined through multi-dimensional analyses of their natural language explanations of the text. In particular, we rely on natural language processing techniques to calculate multiple indices related to the properties of students' explanations at multiple dimensions of language (e.g., lexical, semantic). We then examine whether these automated indices are able to detect whether students were instructed to generate *self-explanations* or *other-explanations* (i.e., explain to a peer) while reading a science text.

2 METHODS

2.1 Participants

A total of 146 undergraduate students participated in this study located in the southwestern United States. The majority (63.7%) of students were in their first year of college, with the remainder in either their second (21.2%), third (11.6%), or fourth (0.03%) years. Of these students, 50% were female and 30.1% reported speaking English as a second language Additionally, 47.3% were Caucasian, 30.1% were Asian, 15.8% were Hispanic, 0.02% were African-American, and 0.05% reported other nationalities. One participant was removed from the analysis due to not following directions; therefore, analyses were conducted on 145 students.

2.2 Study Procedure

The data included in this study were collected during one laboratory session, wherein students read and explained two texts (one easy, one difficult), answered comprehension questions (for the second text), and completed a prior knowledge test. Students were randomly assigned to one of two conditions where they were instructed to type explanations of the text either to a peer (n=69; other-explanation) or to themselves (n=76; self-explanation). As they read, participants were prompted to generate explanations 9 times for an easier text on the topic of heart disease and 16 times for a challenging text on cell division. For the purposes of the current study, we only analyzed the data from the Demographics questionnaire and the explanations that students generated for the second text on cell division.

2.3 Self- and Other-Explanation Task

A self- and other-explanation and reading task was administered to analyze the on-line reading processes students employed during reading, as well as their comprehension of the text at the surface (text-based) and deep (bridging) levels. Students read and either self-explained or other-explained two science texts related to heart disease and cell division, respectively. The texts were presented one segment (i.e., two to three sentences) at a time, with each segment separated by a target sentence in bold. For each target sentence, students were instructed to write a self-explanation or an other-explanation of the information they had just read. Each student wrote 16 explanations for the second text that is analyzed here. Immediately following this procedure, the students were asked to answer eight comprehension questions.

2.4 Data Processing

To calculate the linguistic properties of students' explanations, their individual, sentence-level explanations (the 16 generated during reading) were aggregated. Hard returns were placed after each individual explanation to represent a 'paragraph break.' Linguistic indices were then calculated for each of these aggregated explanation files.

2.5 Model Building

2.5.1 Feature Engineering

To assess the linguistic properties of students' aggregated explanations, we utilized the Tool of the Automated Analyses of Lexical Sophistication (TAALES) [13], the Tool for the Automated Analysis of Cohesion (TAACO) [7], and the Sentiment Analysis and Cognition Engine (SEANCE) [6]. These are automated text analysis tools that compute linguistic indices for the lexical, cohesion, and semantic aspects of texts, respectively. A total of 79 features were computed based on students' explanation. The 79 features can be subdivided into three categories: Descriptive (21 features), Lexical (21 features), Cohesion (20 features), and Sentiment (20 features). Below we provide a brief overview of each type of feature.

Descriptive features. TAACO calculates a number of basic linguistic indices that provide simple counts of features in a text. This includes features such as the total number of words, unique words, and sentences in a given text. Additionally, TAACO calculates the average length of words and sentences and the lexical diversity (i.e., the degree to which the text contains unique words rather than repetitive language).

Lexical features. TAALES calculates indices related to the lexical properties of a given text. These indices describe the characteristics of the words that were found in a given text. Importantly, TAALES also calculates these indices separately for all words, content words, and function words. Examples of the indices calculated by TAALES include word frequency (i.e., the estimated frequency of a word in the English language) and concreteness (i.e., the degree to which the words represent concrete v. abstract concepts).

Cohesion Features. TAACO calculates measures of cohesion in a given text. Cohesion indices provide information about the type of connections that are made between ideas in a given text; relevant cohesion measures calculated by TAACO include: incidence of connectives (overall and specific types of connectives), word overlap for adjacent and all sentences). Importantly, these overlap indices are analyzed for all words, as well as for specific parts of speech.

Sentiment features. SÉANCE relies on a number of pre-existing sentiment, social positioning, and cognition dictionaries and can report on almost 3,000 indices. However, because such a large number of indices can be difficult to analyze, SEANCE reports on 20 components derived from these indices that describe sentiment categories, such as *social order, affect for friends and family, fear and disgust, politeness, positive verbs,* and *objects*. We focus on the scores from these components in the current analysis (see [6] for information about the SEANCE indices).

2.5.2 Supervised classification and validation

We relied on supervised machine learning techniques to build detectors to predict students' condition – i.e., whether they were asked to self-explain or explain to a peer. RapidMiner, a popular machine learning tool, was used to train binary classifiers to classify the experimental groups. In total, four binary classifiers provided by RapidMiner were used: Naïve Bayes, Neural Network, Random Forrest, and Support Vector Machine (SVM). The data contained 69 instances of other-explanation and 76 instances of self-explanation (baseline accuracy 54.5%).

All models were evaluated using leave-one-out cross-validation, in which *k*-1 instances are used in the training data set. The model was then tested on the instance not used in the training data. This process was repeated *k* times until every instance was used as the test set. Because each participant in this study only contributed one instance to the dataset, *leave one instance out* validation effectively served as *leave-one-subject out* validation as the participants could not be included in both the training and testing set. We can therefore be confident that models will be generalizable when applied to new participants because the testing and training sets are independent.

Classification accuracy was evaluated using Cohen's kappa, which indicates the degree to which the model is better than chance at correctly predicting whether students' were assigned to the *self-explanation* or *other-explanation* conditions. A kappa of 0 indicates that detection is at chance levels whereas a kappa of 1 indicates the detector performs perfectly. Below, we also report the percent of cases correctly classified (accuracy). However, it is important to note that this metric should be interpreted cautiously because our slight class imbalance tends to inflate accuracy (e.g., accuracy should be higher than 54.4% to indicate better than chance).

3 RESULTS

3.1 Classification Accuracy

The four classification algorithms (i.e., Naïve Bayes, a Neural Network, a Random Forrest, and a SVM) were applied to the dataset to classify students as belonging in either the self- or other-explanation conditions. The final models reported in this section were selected based on the highest kappa achieved after testing all four of these classification algorithms.

We first tested *combined* versions of our models, which contained the descriptive, lexical, cohesion, and semantic indices.

The best combined feature model was achieved with the Naïve Bayes classifier, which performed at rates above chance (kappa = .247; accuracy = 63%). Table 1 contains the confusion matrix for this final combined model. As can be seen in Table 1, this model had a relatively high rate of misses (.375), wherein actual self-explanations were predicted as other-explanations. In comparison, however, there were still more correct rejections (.632), such that self-explanations were accurately classified as not being other-explanations. This was complemented by a comparably lower rate of false alarms (.368).

Table 1. Confusion Matrix of Combined Model

	Pred. SE	Pred. OE	Priors
Actual SE	.625 (hit)	.375 (miss)	.524
Actual OE	.368 (false alarm)	.632 (correct rejection)	.476

Note. Pred.=Predicted; SE=Self-Explanation; OE=Other-Explanation

To build on this combined model, we examined how each of the four feature subtypes (i.e., descriptive, lexical, cohesion, and sentiment) successfully classified the self-explanation and other-explanation groups independently. Each feature group was tested independently using the same classification algorithms. A summary of the classification accuracies for the best performing models (selected based on highest AUC) is presented in Table 2. Although the subtype models all performed at levels that were above chance, none of the models performed as well as the combined model. This indicates that the self-explanation and other-explanation instructions prompted students to produce different forms of explanation across multiple dimensions. In other words, providing students with instructions to either self- or other-explain had an influence on the way in which students' respond across multiple levels of language.

Table 2. Confusion Matrix of Combined Model

Features in model	Kappa	Classifier
Combined Model	.247	Naïve Bayes
Cohesion Only	.174	Random Forest
Descriptives Only	.159	Random Forest
Lexical Only	.143	Naïve Bayes
Sentiment Only	.193	Naïve Bayes

3.2 Exploratory Feature Analysis

The classifiers reported above provide evidence that self-explanation and other-explanation instructions prompt students to produce different forms of language that can be automatically detected. To more closely examine the text properties driving these differences, we conducted exploratory analyses of the features in the models. In these analyses, we examined how the features related to the model's classifications of explanations in order to provide some initial insight into the primary ways in which the language of self- and other-explanations differ.

Independent-samples t-tests were conducted for each feature to assess mean differences across the two groups. Three of the features were significantly different across the two groups (no multiple-tests correction applied): Sentiment: Social Order component, Lexical: Concreteness (Content words), and Cohesion: Reason and Purpose connectives.

We additionally examined the effect sizes (Cohen's *d*) for the individual features. There were four features with $d > .3$, which is consistent with a medium effect size [5]. Table 3 presents a summary of these features using the combined model, as well as the standardized regression coefficient (β) for each feature.

Results of this exploratory analysis indicate that students in the self-explanation condition generated explanations that were more cohesive and contained words related to social order (e.g., ethics). They also generated words that were less concrete (more abstract) than students in the other-explanation condition.

These results are important because they suggest that other-explanation instructions may promote somewhat different cognitive processes than self-explanation instructions. In particular, self-explanation instructions may lead students to generate more connections between concepts, which in turn could lead them to generate more inferences about complex text concepts. Thus, although the exploratory analyses presented here should be interpreted with caution, they may be informative for future work as they highlight a few of the key dimensions along which explanations to oneself compared to a peer may diverge.

Table 3. Descriptives for Features with Effect Sizes > .3

Feature Type	Self-Explanation M (SD)	Other-Explanation M (SD)	*d*
Sentiment: Social Order	.331 (.099)	.289 (.114)	.393
Lexical: Concreteness (Content Words)	3.22 (.121)	3.26 (.123)	-.357
Cohesion: Reason and Purpose Connectives	.013 (.011)	.009 (.008)	.353
Cohesion: Sentence Overlap - Adjectives	.002 (.003)	.001 (002)	.320

4 DISCUSSION

In this study, we investigated the differential effects of explaining to oneself or to a peer on the linguistic properties of student explanations. We leveraged natural language processing techniques to capture the properties of students' language along multiple dimensions and to relate those properties to the self- or other-explanation instructions. Specifically, three automated text analysis tools were used to analyze students' aggregated explanations of a complex science text. These tools were able to provide linguistic information about the nature of students' explanations across multiple levels, including descriptive, lexical, cohesion, and semantic. Importantly, the features calculated by these tools successfully classified explanations to oneself or a peer. Therefore, by examining students' explanations across multiple dimensions, we were able to detect subtle differences in these two forms of explanation.

The results of the current study support our hypotheses that the multi-dimensional properties of students' natural language responses to text can provide important information about the type of processes in which students engage. In particular, our combined model performed significantly above chance in its classification of explanations as to oneself or a peer. These results are important because they suggest that other-explanation instructions may promote somewhat different cognitive processes than self-explanation instructions. If this is the case, it suggests that researchers need to place a greater emphasis on understanding the role of *audience* in research on the explanation process. This, and other questions, will be examined more thoroughly in future research. For example, we plan to conduct future research that examines how students flexibly adapt their language to different audiences based their presumed level of content knowledge.

The exploratory feature analysis was additionally able to provide important information about students' explanation processes. In particular, the results of this analysis indicated that four linguistic indices had a medium effect size in distinguishing between the two forms of explanation: Sentiment: Social Order component, Lexical: Concreteness (Content words), Cohesion: Reason and Purpose connectives, and Cohesion: Sentence Overlap - Adjectives. These analyses suggest that self-explanation instructions prompted students to generate explanations that were more cohesive but less concrete than their other-explanation condition peers. Further, they referred more frequently to themes related to social order. These analyses indicate that relatively simple instructions to explain a text to oneself compared to a peer can significant impact on the way in which students' explain text content. Educational technologies that examine natural language responses often fail to consider the *communicative* aspect of language production in their analyses and instruction. However, the results of the current study suggest that expanding these analyses to account for audience differences can provide critical information about students' learning processes.

Overall, our results suggest that natural language processing techniques can be used to provide important information about nuanced differences in students' comprehension processes. The ultimate goal of this research is to use these multi-dimensional indices to develop more nuanced assessments of explanations and drive formative feedback in iSTART. More broadly, this study suggests that multi-dimensional analyses of language can be used to develop more adaptive educational technologies that can respond to a number of educationally relevant variables, outside of simple holistic scores. This study is only a first step towards these forms of adaptive technologies; nonetheless, it provides a demonstration of how these indices can be used to model important aspects of the learning environment.

5 ACKNOWLEDGMENTS

This research was supported in part by IES Grants R305A120707 and R305A180261 as well as the Office of Naval Research (Grant No. N00014-16-1-2611). Opinions, conclusions, or recommendations do not necessarily reflect the view of the Department of Education, IES, or the Office of Naval Research.

6 REFERENCES

[1] Allen, L.K., Jacovina, M.E. and MNamara, D.S. 2016. Cohesive Features of Deep Text Comprehension Processes. *Grantee Submission*. (2016).

[2] Allen, L.K., McNamara, D.S. and McCrudden, M. 2015. Change your Mind: Investigating the Effects of Self-Explanation in the Resolution of Misconceptions. *CogSci* (2015).

[3] Allen, L.K., Snow, E.L. and McNamara, D.S. 2015. Are you reading my mind?: modeling students' reading comprehension skills with natural language processing techniques. *Proceedings of the fifth international conference on learning analytics and knowledge* (2015), 246–254.

[4] Chi, M.T., De Leeuw, N., Chiu, M.-H. and LaVancher, C. 1994. Eliciting self-explanations improves understanding. *Cognitive science*. 18, 3 (1994), 439–477.

[5] Cohen, J. 1992. A power primer. *Psychological bulletin*. 112, 1 (Jul. 1992), 155–159.

[6] Crossley, S.A., Kyle, K. and McNamara, D.S. 2017. Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*. 49, 3 (Jun. 2017), 803–821. DOI:https://doi.org/10.3758/s13428-016-0743-z.

[7] Crossley, S.A., Kyle, K. and McNamara, D.S. 2016. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*. 48, 4 (Dec. 2016), 1227–1237. DOI:https://doi.org/10.3758/s13428-015-0651-7.

[8] D'Mello, S.K., Dowell, N. and Graesser, A.C. 2009. Cohesion Relationships in Tutorial Dialogue as Predictors of Affective States. *AIED* (2009), 9–16.

[9] Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H.H., Ventura, M., Olney, A. and Louwerse, M.M. 2004. AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*. 36, 2 (2004), 180–192.

[10] Graesser, A.C., VanLehn, K., Rosé, C.P., Jordan, P.W. and Harter, D. 2001. Intelligent tutoring systems with conversational dialogue. *AI magazine*. 22, 4 (2001), 39.

[11] Jackson, G.T., Guess, R.H. and McNamara, D.S. 2010. Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*. 2, 1 (2010), 127–137.

[12] Kyle, K., Crossley, S. and Berger, C. 2018. The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods*. 50, 3 (Jun. 2018), 1030–1046. DOI:https://doi.org/10.3758/s13428-017-0924-4.

[13] Lombrozo, T. 2006. The structure and function of explanations. *Trends in cognitive sciences*. 10, 10 (2006), 464–470.

[14] Magliano, J.P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H.J. and McNamara, D.S. 2005. Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*. 32, 2 (2005), 185–208.

[15] McNamara, D.S. 2004. SERT: Self-explanation reading training. *Discourse processes*. 38, 1 (2004), 1–30.

[16] McNamara, D.S., Boonthum, C., Levinstein, I.B. and Millis, K. 2007. Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. *Handbook of latent semantic analysis*. (2007), 227–241.

[17] McNamara, D.S., Levinstein, I.B. and Boonthum, C. 2004. iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*. 36, 2 (2004), 222–233.

[18] McNamara, D.S. and Magliano, J. 2009. Toward a Comprehensive Model of Comprehension. *Psychology of Learning and Motivation*. B. Ross, ed. Elsevier Science. 297–384.

[19] Oakhill, J. and Yuill, N. 1996. Higher Order Factors in Comprehension Disability: Processes and Remediation. *Reading Comprehension Difficulties: Processes and Remediation*. C. Cornoldi and J. Oakhill, eds. Lawrence Erlbaum Associates, Inc.

[20] Roscoe, R.D. and Chi, M.T. 2008. Tutor learning: The role of explaining and responding to questions. *Instructional Science*. 36, 4 (2008), 321–350.

[21] 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 (2007), 3–62.