

Understanding Students' Revisions in Writing: From Word Counts to the Revision Graph

Antonette Shibani, Simon Knight, and Simon Buckingham Shum

Technical Report CIC-TR-2018-01

Connected Intelligence Centre, University of Technology Sydney, Australia

To cite this report,

Shibani A, Knight S and Buckingham Shum S: Understanding Students' Revisions in Writing: From Word Counts to the Revision Graph (2018). Technical Report CIC-TR-2018-01, Connected Intelligence Centre, University of Technology Sydney, Australia, <https://utscic.edu.au/research/publications/?target=reports>

Parts of this report were originally published as:

Shibani A, Knight S and Buckingham Shum S. Understanding Revisions in Student Writing through Revision Graphs. In 19th International Conference on Artificial Intelligence in Education, London, UK.(2018)

That final publication is available at Springer via [http://dx.doi.org/\[insert DOI\]](http://dx.doi.org/[insert DOI])

Understanding Students' Revisions in Writing: From Word Counts to the Revision Graph

Antonette Shibani, Simon Knight and Simon Buckingham Shum

University of Technology Sydney, PO Box 123, Broadway, Ultimo, NSW 2007, Australia
antonette.aileenshibani@student.uts.edu.au

Abstract. Text revision is regarded as an important process in improving written products. To study the process and products of this revision activity from authentic classroom contexts, this report provides an exemplification of methods involving text analysis and visualization. These analyses open up the possibility of exploring the features of text to understand writing revision and the stages in students' revision of drafts, which can lead to further automation of revision analysis for researchers, and formative feedback to students on their writing. A novel *Revision Graph* is introduced to aid detailed analysis of the writing process. This could be automated and applied to study the direct impact of automated feedback on students' revisions and written outputs, thus evaluating its effectiveness in pedagogic contexts.

Keywords: Learning Analytics, Writing Analytics, Revision Analysis, Writing, Revision process, Visualizations, Revision Graph

1 Introduction¹

Text revision is considered an important process in writing to support the reworking of writer's thoughts and ideas, playing a major role in the outcome of the writing [1]. The cognitive process theory of writing defines revision as a recursive process that can be called any time during writing [2]. As per this model, writers engage in processes like task definition, evaluation, goal-setting and strategy selection to make revisions, thus leading to improvements in a text. To teach students revision skills to improve their writing, it is essential for researchers and educators to understand what contributes to good revision and how it occurs. This can be supported by Writing Analytics, which could be thought of as a sub field of Learning Analytics that involves "the measurement and analysis of written texts for the purpose of understanding writing processes and products, in their educational contexts" [3]. Such analytics might be deployed both to provide feedback to students on their revisions, and in research to understand the revision process using textual features.

Analysis of revisions has typically relied on resource-intensive manual observation and coding of writing behaviour, and/or on analyzing and/or grading the textual products of revision. For example, revisions made in writing have been studied using manual coding to categorize the types of revisions like unit (e.g., word, phrase, sentence), type (e.g., addition, substitution, spelling), and quality [4]. There have also been recent efforts to automate classification of revisions based on the content of the changes made [5]. In addition, studying the process of revision in a detailed manner helps researchers and educators gain insights into the processes involved in the creation of a written document.

¹ A shorter version of this report on the process-centric analysis can be found at [16].

The process of revision has also been studied using personal testimonies of participants regarding their cognitive process in revising, or by process tracing and participant-observer methods that observe the behaviors involved in revision [1]. Such resource intensive manual observation and coding are improved with advanced data collection and analysis techniques in more recent work. With the right unit of analysis and visualization techniques, the revision process can be easily understood and explained to educators and other non-technical audience. However, because analysis of revisions is generally conducted on multiple drafts of a produced text, it is challenging to interpret differences between texts. Analysis of revisions on a provided ‘base text’ may provide an innovative approach to addressing this gap, an approach detailed in Section 2.

While automatic classification of writing is useful to develop large scale solutions such as automatic graders, the objective is to assign a grade that is sufficiently close to a human grader. Such approaches do not always provide data that can serve either as formative feedback to students, or assist human sensemaking in research contexts. Further work is required to understand the nature of revised texts and the revision process, in order to develop models for meaningful formative feedback to students. For instance, a machine learning black box that classifies students’ essays as a good or bad essay based on a corpus of input essays may not be useful to understand and give feedback to students on what changes can be made to improve the essay. When it comes to providing good feedback, whether automated or human-based, it should be actionable and meaningful for a learner in order for them to learn from it. We argue that this requires a deeper understanding of the features of revised texts and the processes that led to them. In this way, we can bridge the gap between computational outputs and behavior to enable the learner. The product and processes involved in the revision process can be studied with text analysis and visualization techniques; this paper provides exemplifications of some of these methods, illustrating their potential use for learning analytics.

2 Research Context

The research context from which the data for this paper is drawn is a pedagogic intervention that made use of a web-based tool integrated with multiple tasks to help students write better essays for their subject [6]. In the main revision task, students worked on revising a short essay that was provided to them (original essay), to produce an improved version, with a focus on improving the rhetorical structure of the text, in study conditions with and without using automated writing feedback. In this paper, we introduce methods to exemplify analysis of the revised essays and revision process produced in this task, which could be applied to similar revision tasks in writing instruction. To study the features of revision using this analysis, the revised essays were marked by tutors on a scale of 0-3 (0- degraded, 1-no change, 2- minor improvement, 3-major improvements) based on the subject’s rubric elements as indicators of quality. Sample students’ revised essays which were used to characterize good and bad essays with these techniques included the most degraded (16) and the most improved (15) essays to distinguish the extremes. The original essay given to students for revision was on the legal topic of using video conferencing in civil trials, consisting of 550 words with four small paragraphs and a reference list. Drafts from students’ revisions were captured every one minute using the AWA-Tutor tool which scaffolds the tasks in the intervention [7]. Students’ usage of automated feedback and its output every time the student requested for feedback were captured to study the impact of automated feedback as well. This data will be used in the second part of analysis that studies the process of revising drafts in section 4.

3 Studying the Revision Products

In this section, we will describe methods that can be used to analyze the *products* of revision, which are the revised essays of students in this case. The first two are simple quantitative measures which can be used to

quantify the revision in terms of textual features. These measures assist in providing an overview of the revisions made and are usually good first-hand indicators of revision quality. Graded revised essays of students mentioned in the research context section were studied using these measures. The next two are visualization methods to dig deeper into the quality of actual revisions made. A good visualization should help the user make sense of complex information in order to support particular tasks, in this case, revealing the feature order (e.g. word/ sentence order) in a text. To illustrate the use of these methods, we will introduce individual essay revisions of varying quality from two students. These two examples are from opposite ends of the improvement/degradation spectrum, to study their unique characteristics in detail. The methods described here are particularly useful to analyze essays which are very similar (revised from the same base text). Other linguistic features that are used to quantify texts like cohesion [8] are not included since they do not detect the nuanced differences made in the small texts in this context (i.e. original and improved/degraded essays as a whole do not give significantly different cohesion scores).

3.1 Word counts

A common measure to characterize text differences is a simple word count. Although it is a simple measure, word counts can be an effective indicator to get a general sense of the quantity of revisions made by a student at the word level [1]. To further characterize revision, the relative change in word count in comparison to the original essay can be calculated as the difference between the word counts in the given and the revised essays. Even though this measure of the number of words changed cannot identify individual additions and deletions, it can still be an indicator for good or bad performance, depending on the context.

In the examples used for our analysis, the mean word count change was -24.5 for degraded essays (SD=48) and +11.73 for improved essays (SD=70). The actual values ranged from -103 to 33 for degraded essays and -110 to 99 for improved essays. This shows that improvement need not involve addition of words; students could make improvements by deleting words or rephrasing them to be more succinct. However, these simple measures do not capture other information including text-reorganizations, which may have significant impact on text quality. Simple variants on this method could capture sentence or paragraph counts, which represent some structural information. In addition, distinct words (i.e. words in the new text but not the original) could be represented to give some measure of textual change. Other text similarity measures also exist to capture this information, detailed next.

3.2 Text similarity measures

Text similarity measures are commonly used in information retrieval and document clustering to compute the similarity or dissimilarity between two documents [9]. In the current context, we are interested in studying the dissimilarity of revised essays in comparison to the original essay to quantify the changes made. This is measured using Cosine Distance, which is a measure that quantifies the distance between two documents in a vector space model. To calculate this measure, text documents to be compared are represented as vectors and the cosine of the angle between them are measured in an n-dimensional vector space, where n is the number of words. For instance, consider a reference document containing the terms “essay” and “text” once each, that is compared with a comparison document (1) containing “essay” 5 times and “text” 0 times, and comparison document (2) containing “essay” and “text” once. Here, the reference document is more similar to document (2) than document (1). This is because the direction of vector contributes to the angle and not its magnitude (frequency). The measure was calculated using the ‘stringdist’ package in R.

From our examples, the cosine distance serves as a good proxy to study the performance of students in revision. The mean cosine distance (scaled) was 5.2 for degraded essays (SD=4.8) and 9.77 for improved essays (SD=7.6). The cosine distance measure positively correlated to the revision score ($r(29) = 0.41$, $p <$

0.05). So in general, the higher the distance between the original and the improved essay, the higher the performance. This means that students who made more changes to the essay had better performance than students who did not make many changes to the text. However, this measure is based only on the angle between the document vectors, and does not take into account the frequency of words, potentially losing some information about the revisions made. In cases where a semantic similarity measure would be helpful to analyze the content of revisions, more advanced word2vec word embeddings can be used [10].

3.3 N-gram graph

Words and their occurrences in the revised essays can be analysed to study the revisions made to the given text. This can be visualized using a n-gram graph that captures all important words in the essay along with the position of co-occurring words. In this graph, nodes represent words and edges represent their occurrence next to each other in terms of the ‘n’ number of words selected. An example trigram graph constructed on our example revised essays using the ‘igraph’ package in R is shown in Fig. 1.

The graph shows the density of occurrences and connectivity among the words used in the essays. The improved essays contain additional words like connectives (e.g. *however*) that connect the given words, producing a denser graph than the degraded and the original essay. This is visible from the change in clusters created in the graphs. This could be a useful way to represent essays visually to study their revisions, going beyond word-count or broad-similarity measures. Graph metrics such as distance, centrality and connectivity metrics could be utilized to further study these graphs quantitatively. Graphs could also be used to develop word clouds, or concept maps, to visualize the key concepts and their connectedness in a document. This visualization may be particularly useful in studying revisions made by adding many new concepts and ideas to the given base text which can be observed by changing clusters of words.

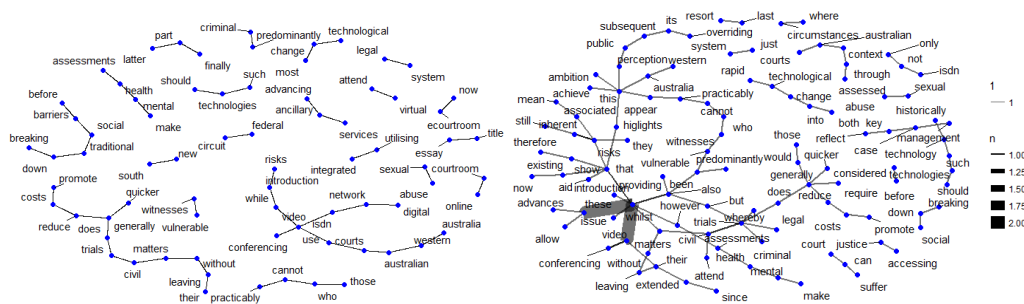


Fig.1. Trigram graphs from a degraded (left) and an improved essay (right)

3.4 Rhetorical moves graph

The next analysis on the revised texts is based on rhetorical moves, since the core aim of the task given to students was to understand the use of rhetorical moves and discourse markers to improve an essay. To illustrate this analysis, the improved and degraded essays are mapped as a rhetorical move graph to study their sequences. Writing analytics tools like Academic Writing Analytics (AWA) can automatically detect rhetorical moves in an essay using natural language processing techniques, simplifying analysis of this kind [11]. The rhetorical move graph is based on the method of visualizing rhetorical moves as proposed by the sequence mining approach to study the patterns of rhetorical moves in student texts [12]. The differences in the moves graph between the degraded and the improved essays can be seen in Fig. 2.

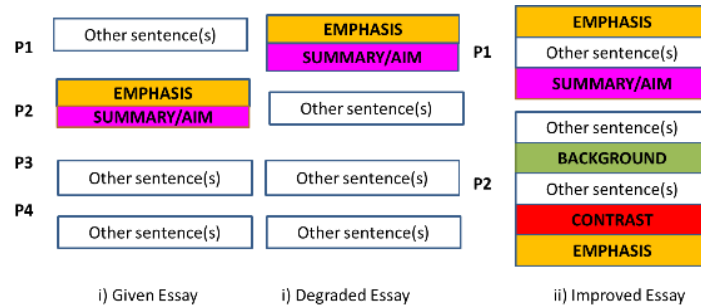


Fig. 2. Sequence of rhetorical moves from a degraded and an improved essay

As detailed in [9], *Emphasis* is a move that usually highlights the key statement the essay's argument and *Summary* signals to the reader the author's intent and textual organisation. These were provided in the original essay along with some supporting statements. In our example shown in Fig 2, the degraded essay did not add new rhetorical moves and only reorganized the given sentences. The other paragraphs contained no rhetorical moves to explicitly guide the reader with the flow of the text. On the other hand, the improved essay included discourse markers that indicate rhetorical moves such as explaining *Background* work, *Contrasting* critical statements, *Emphasizing* in the combined second paragraph, and removing inappropriate paragraphs. Although the examples described above show only a type of revision made using rhetorical moves, they illustrate the possibility of using such patterns to characterize other revisions as good and poor.

4 Studying the revision process

Analysis of the *process* of writing revision predominantly involves developing visualizations to represent the process of drafting and revision. To visualize modification patterns in an online document, Caporossi and Leblay [13] developed a graph theory approach to represent the movement of text through a document. In that work, log data of keystrokes and cursor movements from the document editing process was used to generate a graph. However, there is no evidence that educators would find keystroke-level data insightful for understanding revision patterns, nor that students would find this meaningful feedback to improve their writing. In other work that investigated collaborative writing processes, a revision map was created to represent the joint development of ideas by a group of authors [14]. These maps helped researchers gain insights into the location and time of text edits made, and the collaboration of students to develop the document, but no effort was made to test this as feedback to the students themselves.

Recent work introduced the use of Sequence Homology Analysis (SHA) to study the evolution of public speech drafts by comparing the changes in characters [15]. SHA is a method from molecular biology to study differences in DNA strands. It was applied to identify the character differences between text drafts, and a draft network was proposed. This network maps sentences in students' drafts based on the strength of revisions made across the drafts. While computational analysis techniques which are previously available can be used to study learning processes, new ways of looking at data also helps uncover new patterns of the current pedagogical contexts and learning scenarios. A shorter version of this process analysis can be found in [16].

4.1 A novel approach to revision analysis

Finally, building on previous work in studying the process of revision, we provide a novel analysis of revisions over multiple drafts created through the text-revision exercise using a 'Revision Graph'. This draft level analysis can aid to uncover the previously unknown processes involved in the editing of the final revised essay. This new manual analysis focuses on the ordering of sentences and revision actions, which could be potentially automated. In this revision graph, the nodes represent sentences from the drafts and the edges represent changes in the organization of sentences across multiple drafts. The sentences are represented in the sequence of occurrence across the paragraphs. The colors of the nodes (Fig. 3) indicate the type of revision action made at the sentence level: i) minor revisions are when students predominantly use the given text, but add or substitute few words, ii) major revisions are when students add a substantial number of words and explanations to the given text with the inclusion of their own writing, iii) no changes made and iv) no change in the current stage, but deleted in the next stage. Red inverted triangles represent the request of automated feedback during the revision process. Dotted edges are used to represent the repetition of similar concepts across multiple sentences inside a draft. This could be a good indicator of word repetition/ overlap leading to high cohesion in the document.

Fig. 3 shows the revision graph constructed manually from the sample improved essay's drafts to show the evolution of a high-scoring revised essay. The drafts were selected from certain intervals (every 6 minutes in this analysis) using the time spent on revision. The graph shows the stages in the revision of the given text containing four paragraphs and 15 sentences to the final product containing two paragraphs and 10 sentences. In the first draft stage, the student has deleted some broad introductory sentences from the original essay. The first paragraph of the draft has been shaped up by making minor and major revisions to the given sentences and reordering them, while the other paragraphs remain untouched. In the second stage of drafting, the student has deleted the previous second paragraph and mainly worked on the revision of this paragraph from the other paragraph sentences. Here the text has been reduced to three paragraphs.

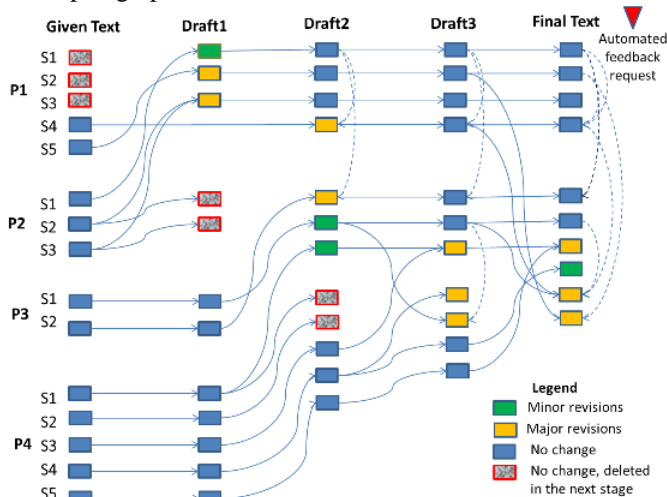


Fig. 3. Revision graph of a sample improved essay

From the third draft, the first paragraph remains stable. The student has made some extensive changes to the sentences by revising and consolidating them to produce a final text consisting of only two paragraphs. The number of references to the previously written words increases in each stage of the draft as shown by

the dotted edges. The final text has many such cross references made to the previous sentences, which has improved the cohesion of the text. This student requested automated feedback (red triangle) after completing the final text and made no more changes after that. This information is made visible by matching the timestamp of feedback request with those of the drafts. It informs that the changes made to the text by the student were not an effect of the feedback received. In cases where we do not have such process information to study writing, it is feasible that the revision effect is attributed to feedback, but they are in fact not related. This revision graph is thus serving its purpose of making visible, at an appropriate granularity, the nature of the revisions, and whether the automated feedback component impacted subsequent revisions.

In the revision graph of a sample degraded essay shown in Fig. 4, there are no revisions made by the student to the given sentences. The introductory sentences have been removed in the first draft, and sentences have been reorganized in the second draft. No further changes have been made from the second draft to create the final revised essay, leading to a degraded version of the given text. The last three drafts have remained stable, meaning the student has stopped working in the last few minutes of the revision task.

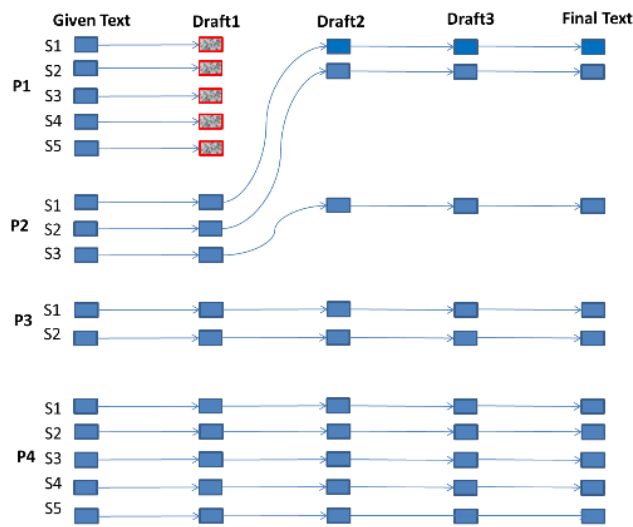


Fig. 4. Revision graph of a sample degraded essay

The above manually constructed revision graphs could potentially be automated for a large scale analysis of revision process. Sample pseudo code is shown in Fig.5, which would operate on writing process data collected automatically (students' drafts at specific intervals) by the platform described in [7]:

```

Create the corpus of n ordered drafts based on timestamp
For each draft document:
  Parse document into sentences
  Create a column of nodes in the graph for each sentence ordered by paragraph
  If n > 1: (Start from the second draft to add edges)
    For each sentence:
      Calculate similarity score with each sentence from the previous draft
      If similarity score > similaritythreshold:
        Add edge in graph between the current sentence and previous
        draft sentence
        Use similarity score to colour nodes for minor, major revisions
        and no change
      Calculate word overlap with all other sentences in the same draft
      If word overlap > overlapthreshold:
        Add dotted edge in graph between the sentences
  If no edge originates from a sentence in draft (excluding last draft):
    Mark sentence node as red deleted node
  If automated feedback request falls in the timestamp between two drafts:
    Add red inverted triangle entry for automated feedback request

```

Fig. 5. Sample pseudo code for automatically constructing a revision graph

In the current context, the revision graph was used to study the revisions made across drafts in terms of the strength of revision actions like insertions, deletions, substitutions etc. It can also be applied to other specific changes we would like to study like the types of revisions (e.g. concepts, rhetorical moves, surface errors, etc.). Another application of this revision graph, as mentioned previously in the revision process analysis of a good revised essay, is to study the effect of automated writing feedback using actual revisions made by students. This way of evaluating the effectiveness of Learning Analytics applications (automated writing feedback in this case) is thus made possible using Learning Analytics itself (tracking the revision process in student drafts for detailed study).

5 Discussion and Conclusion

The paper illustrated methods to study revision with the use of text analysis and visualization techniques, exemplifying its potential for learning analytics applications by using real examples of students' writing from a revision task. The combination of product and process centric methods to study revision helps to uncover deeper aspects of writing. The methods demonstrated different granularities of analysis of revisions starting from a set of essays, extending to individual essays and internal processes.

In sections 3.1 and 3.2, word count change and a text similarity measure – cosine distance – were used to quantify the changes made in the final product of revision in a set of degraded and improved essays. The cosine distance was a good measure of the performance since students who make more changes to the given text generally score higher in the task. Visualization techniques like the n-grams graph based on word occurrences in text and the rhetorical moves graph constructed from the sequence of rhetorical moves in sections 3.3 and 3.4, were useful to study further the revisions made in individual texts. Such visualizations of texts are an effective way to observe the characteristics of these revised texts.

The construction of a revision graph to visualize the revision process of essays introduced an effective technique to study the evolution of writing in terms of the actions that led to the final product. This novel

visualization (section 4.1) revealed a pattern of activities like addition, deletion and re-organization of sentences in the generation of the improved essay showing the importance of understanding textual restructuring and the revision process in writing. It demonstrated the opportunity to study the different ways that good or poor writing may evolve in its revision stages. It could be particularly useful to study the effects of writing feedback on students' revisions at multiple stages, thus helping to find effective forms of feedback leading to revisions. This could be the first step towards studying the contexts in which automated feedback can work better, and other contexts in which other forms of feedback like human feedback are well suited. Further cognitive processes can be studied using think aloud techniques to capture the mental models while adopting/ rejecting the feedback. Fully automated approaches to do these analyses would reduce human effort, but they are yet to evolve. The current approach where the researcher is in the loop of the analysis is still useful, as it brings together contextual and cognitive constructs to understand revisions better.

Learning analytics methods to collect and analyze such process data can help researchers to study the writing processes in great detail. It can inform them of the additional processes that happen in between the submission of drafts and provide feedback on the revisions made; the web-based ArgRewrite tool [5] provides one example, classifying revisions as either 'surface' or 'content' based changes in students' drafts. This can provide a lens to study the writing style of different authors and find synergies among certain set of writers that attribute to better quality. We do not yet know if these techniques can be used to differentiate texts that are not extreme cases of performance; thus, having demonstrated the utility of the revision graph in principle, to test its performance on text corpora at scale requires software implementation. Finally, to extend their usage in educational contexts, further work has to be done to characterize essays based on the discussed features to provide meaningful feedback to educators and students. The feedback might be based on writing patterns that emerge or revision types, e.g. to draw attention to the fact that there have been no substantive changes in graphs after 2 drafts or within a defined time interval, or changes that only involve surface level error corrections. The methods described in the paper can thus be applied in writing research to analyze the revision process and product using different lenses.

References

1. Fitzgerald J: Research on revision in writing. *Review of educational research*, 57(4): 481-506 (1987).
2. Flower L and Hayes J R: A cognitive process theory of writing. *College composition and communication*, 32(4): 365-387 (1981).
3. Buckingham Shum S, Knight S, McNamara D, Allen L, Bektik D and Crossley S: Critical perspectives on writing analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, pp. 481-483, NewYork (2016).
4. Crawford L, Lloyd S and Knoth K: Analysis of student revisions on a state writing test. *Assessment for Effective Intervention*, 33(2): 108-119 (2008).
5. Zhang F, Hwa R, Litman D J and Hashemi H B: ArgRewrite: A Web-based Revision Assistant for Argumentative Writings. In *NAACL-HLT 2016 (Demonstrations)*, pp. 37-41, San Diego, California (2016).
6. Shibani A, Knight S, Buckingham Shum S and Ryan P: Design and Implementation of a Pedagogic Intervention Using Writing Analytics. In *25th International Conference on Computers in Education*, Asia-Pacific Society for Computers in Education, New Zealand (2017).
7. Shibani A: AWA-Tutor: A Platform to Ground Automated Writing Feedback in Robust Learning Design. In *8th International Conference on Learning Analytics & Knowledge (LAK18)*, Sydney (2018).
8. McNamara D S, Graesser A C, McCarthy P M and Cai Z. *Automated evaluation of text and discourse with Coh-Metrix*. 2014.
9. Huang A: Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand, pp. 49-56 (2008).

10. Mikolov T, Sutskever I, Chen K, Corrado G S and Dean J: Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111-3119 (2013).
11. Knight S, Buckingham Shum S, Ryan P, Sándor Á and Wang X: Designing Academic Writing Analytics for Civil Law Student Self-Assessment. *International Journal of Artificial Intelligence in Education*(Special Issue on Multidisciplinary Approaches to Reading and Writing Integrated with Disciplinary Education): 1-28 (2016).
12. Knight S, Martinez-Maldonado R, Gibson A and Buckingham Shum S: Towards mining sequences and dispersion of rhetorical moves in student written texts. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ACM, pp. 228-232 (2017).
13. Caporossi G and Leblay C: Online writing data representation: a graph theory approach. In *International Symposium on Intelligent Data Analysis*, Springer, pp. 80-89 (2011).
14. Southavilay V, Yacef K, Reimann P and Calvo R A: Analysis of collaborative writing processes using revision maps and probabilistic topic models. In *Proceedings of the third international conference on learning analytics and knowledge*, ACM, pp. 38-47 (2013).
15. Wininger M: Measuring the evolution of a revised document. *Journal of Writing Research*, 6(1) (2014).
16. Shibani A, Knight S and Buckingham Shum S. Understanding Revisions in Student Writing through Revision Graphs. In *19th International Conference on Artificial Intelligence in Education*, London, UK.(2018)