Title	アカデミックライティングにおける修正スキル向上のためのコーパスを利用した適応的ツール：TRONAの場合
Author(s)	Ocharo, Harriet Nyanchama
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15525
Rights	
Description	Supervisor:長谷川 忍，情報科学研究科，博士

A Corpus-based Adaptive Tool to Improve Revision Skill in Academic Writing: The Case of TRONA (Topic-Related revisiON Assistant)

Harriet Nyanchama Ocharo

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

A Corpus-based Adaptive Tool to Improve Revision Skill in Academic Writing: The Case of TRONA (Topic-Related revisiON Assistant)

Harriet Nyanchama Ocharo

Supervisor: Shinobu Hasegawa

School of Information Science
Japan Advanced Institute of Science and Technology

September, 2018

Abstract

Academic writing occurs in the final phase of a research activity cycle, where students present the results of their research. Students and researchers in higher education measure their achievements through the number of quality research articles that they publish. Therefore, studying how to improve the quality of writing is an important area of research. Quality writing is required in research in order to convey ideas clearly, especially when students write research articles or dissertations.

Writing is a cognitive activity consisting of generating, translating and revising. Much of the research in the area of academic writing tools is focused on equipping students with grammatical skills (especially in the case of English as a second language) and technical writing skills. Some of the tools also assist students with the generating stage (idea generation, planning etc.). However, there is not much research on the use of software tools to assist students during the revision process of academic articles. This is the motivation for this dissertation to focus on the revision aspect of the writing process.

This research designed and implemented a corpus-based adaptive tool, TRONA (**T**opic **R**elated revisi**ON A**ssistant), to support the improvement of revision skill in academic writing. The revision corpus consisted of articles written by former students in one laboratory and it included the raw drafts as well as the final articles, and the feedback from the laboratory supervisor in the form of comments that helped those students improve their drafts. Natural language processing and machine learning techniques were applied to reliably predict the most important comments. These comments were used to provide adaptive support in the form of hints to help students resolve reviewer comments in their own article drafts.

The type of hints provided depend on the student’s skill level. The type of adaptive support given is based on the teaching methods of the cognitive apprenticeship theory: specifically, modeling, coaching, scaffolding and fading. The cognitive apprenticeship theory is a widely accepted pedagogical theory of teaching cognitive skills in an explicit way. Through the adaptive interface, novices are provided with modeling support, intermediate students with coaching, while the support for advanced students fades so that they can become more independent.

The Item Response Theory (IRT) was applied to estimate the student's revision skill and the comment difficulty. The estimated student revision skill score for 7 students and the comment difficulty measure for 20 comments by IRT was compared with a manual evaluation by a supervisor of the laboratory. The Pearson's correlation analysis results showed a significant correlation between the student scores by IRT and supervisor estimations.

Furthermore, a machine learning algorithm (SVM) was applied to classify the comments in the article drafts in the corpus as *content-related* (comments that encourage global revision) or *not content-related* (comments on simple spelling and grammatical errors). With performance measures of 89% that were achieved for both recall and precision, it was demonstrated that machine learning can be applied to automatically and reliably predict whether a reviewer comment in an academic article is content-related or not. Once a student uploads their document to TRONA, about 90% of the *non content-related* comments can be filtered out. The student can therefore first focus on revising the comments that encourage global revision. The classification method was also incorporated into TRONA to select the content-related comments that were applied in the Item Response Theory to estimate the students' revision skill level.

The contribution of this research is in the area of writing tools that use artificial intelligence to support the revision process of students in higher education. This study presented a way to construct a revision corpus of raw article drafts from previous students in one laboratory, as well as a way of using machine learning, to make the reviewer comments in the drafts more meaningful to the students' revision process. The Item Response Theory was proposed as a suitable method to estimate students' revision skill. In addition, this study demonstrated how to achieve adaptation in a revision support tool through the cognitive apprenticeship methods of modeling, coaching and fading. Acquisition of revision skill is highly dependent on the laboratory style of writing; therefore this research could have an impact on laboratory education.

Keywords: revision support system, academic writing skill, revision skill, comments classification, laboratory education

Publications

Some ideas, figures and tables of this dissertation may have previously appeared in the following five publications:

International Journals:

1. Harriet Nyanchama Ocharo & Shinobu Hasegawa: *Using Machine Learning to Classify Reviewer Comments in Research Article Drafts to Enable Students to Focus on Global Revision*. Education and Information Technologies, Volume 23, Issue 5, pp 2093-2110 (2018).

International Conferences:

1. Harriet Nyanchama Ocharo & Shinobu Hasegawa: *Adaptive Interface that Provides Coaching, Modeling and Fading to Improve Revision Skill in Academic Writing*. Human Interface and the Management of Information: Applications and Services. Proceedings of the 20th International Conference on Human and Computer Interaction (HCII2018), pp.300-312 (2018).

2. Harriet Nyanchama Ocharo & Shinobu Hasegawa: *Resource Description Framework Models for Representing the Revision Process in Research Support Systems*. Proceedings of the 25th International Conference on Computers in Edcation (ICCE2017), pp. 252-260 (2017)

3. Harriet Nyanchama Ocharo, Shinobu Hasegawa and Kiyoaki Shirai: *Topic-based Revision Tool to Support Academic Writing Skill for Research Students*, Proc. of The Tenth International Conference on Advances in Computer-Human Interactions (ACHI2017), pp.102-107, (2017)

4. Harriet Nyanchama Ocharo & Shinobu Hasegawa: *An Adaptive Research Support System for Students in Higher Education: Beyond Logging and Tracking*. Human Interface and the Management of Information: Applications and Services, Volume 9735 of the series Lecture Notes in Computer Science, pp. 178-186, (2016)

Dedication

For Jeremy, and his namesake, Jeremiah.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Shinobu Hasegawa, for his invaluable support during the duration of this research. I believe that under his able guidance, I have become a competent researcher capable of carrying out independent scientific inquiry. More than just academic guidance, he provided a friendly and flexible environment under which I was able to carry out productive research while taking care of my young son.

I am grateful too, to my second supervisor, Prof. Hiroyuki Iida for his insightful feedback and thought provoking questions. Special thanks to my minor research supervisor, Dr. Kiyoaki Shirai. His openness and willingness to answer my questions relating to NLP and Machine Learning allowed me to push my boundaries and made me interested in becoming an ML application researcher.

I would also like to thank the Japanese government for awarding me the prestigious MEXT scholarship and thus the opportunity to carry out my research in Japan. I am also grateful to my university, JAIST and the lab I worked under, Hasegawa's Lab, for the funding to attend many international conferences, where I got to interact and share ideas with researchers from all over the world.

I cannot forget to thank my colleagues in Hasegawa lab for their encouraging feedback throughout my research. Many thanks to my friends in JAIST for their support, taking turns babysitting Jeremy whenever I had urgent meetings, etc. We made a lifetime of unforgettable memories! I also had a lot of community support from friends in Hope House, and from many Japanese friends, including Harumi and the Nishikawas, whom I consider my "Japanese parents".

I am grateful to my friends in Kenya, for believing in me and giving me that much needed confidence boost whenever times were tough.

And last, but most certainly not least, my parents and brothers, who provided me with all the support and encouragement I needed to leave Kenya and pursue my PhD in Japan.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CAT	Cognitive Apprenticeship Theory
DT	Decision Trees
EFL	English as Foreign Language
ESL	English as a Second Language
L2	Second Language
IRT	Item Response Theory
NLP	Natural Language Processing
SVM	Support Vector Machine
TRONA	Topic-Related revisiON Assistant
VBA	Visual Basic for Applications
XML	Extended Markup Language

1. Introduction

1.1 Background: Education in the 21st Century

In the current period, technology is changing rapidly, bringing about globalization and other changes. Job categories are disappearing as new ones are created. Education is the key tool to manage the challenges ahead. The economies of the future will be knowledge-driven. That's why we must use education to help people ride this wave of change and give them the skills they need for the new jobs of the 21st century. Figure 1 shows the top 10 skills necessary in the 21st century, according to the Future of Jobs Report (World Economic Forum, 2016).

However, most education systems were built for the needs of the 20th century. For example in higher education, many courses are designed to transfer knowledge but have not been designed to deliver the skills needed for the changes ahead. Institutions of higher education must provide students with the ability to continually learn, to think critically and theoretically, to be reflective and reflexive, to innovate and break the status quo, and to navigate in the unstable waters of the global economy.

	in 2020		in 2015
1.	Complex Problem Solving	1.	Complex Problem Solving
2.	Critical Thinking	2.	Coordinating with Others
3.	Creativity	3.	People Management
4.	People Management	4.	Critical Thinking
5.	Coordinating with Others	5.	Negotiation
6.	Emotional Intelligence	6.	Quality Control
7.	Judgment and Decision Making	7.	Service Orientation
8.	Service Orientation	8.	Judgment and Decision Making
9.	Negotiation	9.	Active Listening
10.	Cognitive Flexibility	10.	Creativity

Figure 1: The top 10 skills needed in the 21st century (Future of Jobs Report, World Economic Forum, 2016)

Teaching and research are two core activities in institutions of higher education. Therefore students need to acquire these skills in the course of attending lectures or when carrying out research. The broader research question is how can we ensure that the acquisition of these skills during course design and delivery or when carrying out research activities? Can digital technology enhance the acquisition of knowledge and skills?

We are living in the big data and artificial intelligence era, where researchers and practitioners try to leverage the power of computers to improve and enhance learning. In recent years, there has been a growing uptake of technology in support of education. This support includes new platforms for learning and new ways of owning knowledge. Computers enable efficient storage and easy retrieval of information, quick data processing, provide audio-visual aids in teaching, better presentation of information, and access to the Internet, efficient communication, etc. Computers and the Internet have also given an impetus for the uptake of distance learning, which means the methods of teaching in a classroom need to be adapted for online/distance learning.

1.2 Research Activity Cycle and Research Support Systems

Research being one of the core activities of institutions of higher education, it follows that ways of improving research skills of the students is an important area of study. Teaching students research skills provides them with information and facility for improving their research capacity, quality and productivity with the aim of better quality, more effective and efficient research output from institutions of higher education.

In institutions of higher education, where learning as well as research takes place, several software systems have been developed to assist students to learn and to also carry out research activities. This study is situated in the area of research support systems - various tools and systems that are designed to lend support to students in their daily research activities.

The daily research activities of students include, for example, the logging and tracking of the progress of experiments, communication with supervisors, managing the research schedule, etc. Students need research support systems that offer such functionality. This is because tracking of research can also enable students to complete their research in time (Suhaily , Rozainun, & Azmi, 2015). A research support system can also help students to find relevant information, to choose the right tools and to produce effective presentation of research results (Yao, 2003).

In order to provide the required support for students' research activities, it is important to understand the research activity cycle. This is because there is need to identify the requirements and the type of support that is necessary to enable students to carry out their research effectively. There is need to learn as much as possible about the research activity process in order to understand the research context and the goals that the student is trying to achieve. After requirements have been understood, then it will be possible to support students in achieving their research goals.

In a typical research process (Fankfort-Nachmias & Nachmias, 1992), the student goes through the following phases, as illustrated in Figure 2: setting a research theme, literature review, design of model or experiment, development or experimentation, testing and evaluation, and finally presentation of results. This research activity process is more or less similar for every discipline, from "hard" sciences to the social sciences (Lynch, 2013). The key difference across disciplines is in the subject matter, and therefore, the type of data used and the methods for gathering it.

In setting a research theme, students identify the area of research they would like to investigate by reading relevant literature or by discussing with their supervisor or colleagues about a theory or a problem they are interested in investigating. During the literature review phase, students carry out a critical analysis of existing literature in the field in order to identify any gaps in knowledge. At the end of this phase, students have a formal problem definition and objectives of the study. In the design phase, students either come up with a model or experiment design to realize the objectives. The development or experimentation phase is where the model or design is implemented. Then design is tested and evaluated, and the results analyzed in the testing and evaluation phase. The final stage is the presentation of research results, which is a culmination of all the efforts put into a research study.

There are tools to help students at various phases of the research process; for instance (Anzai, et al., 2012) present a system that provides a platform for the polishing and refinement of presentation slides.

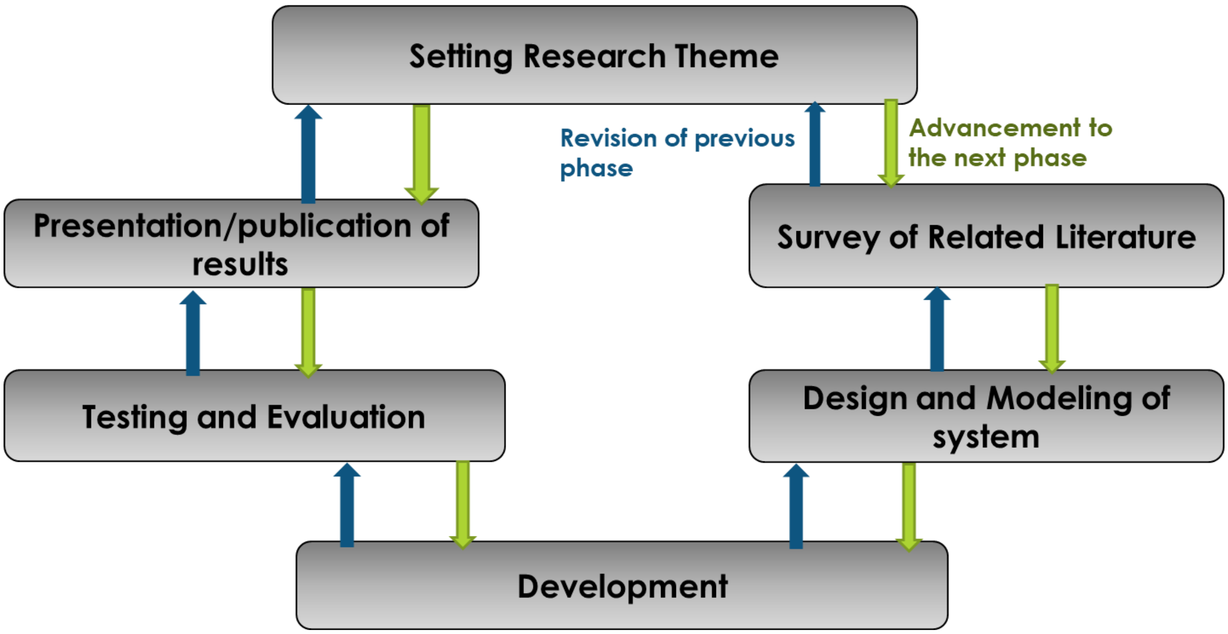


Figure 2: Research Activity Cycle

Table 1 shows some examples of research tools available for students in the Information Science field to support their research activities throughout the various phases. When students set a research theme, they have to come up with creative or innovative ideas or hypothesis. In this case, the students need tools that support brainstorming or coming up with ideas; such as mind mapping tools. A mind map is a diagram used to visually organize information. A mind map may be hierarchical and shows relationships among pieces of information or ideas.

There are also tools that support the literature review phase. During the literature review, students carry out a critical analysis of existing literature in the field in order to identify any gaps in knowledge. They therefore need software that supports the easy storage and retrieval of articles, tagging, annotation and citation. Examples of such tools include Mendeley and Readcube. For the design and modeling phase of the research activity cycle, students need tools that support the visualization of their designs. Such tools may include graphics or simulation software. The development phase in the case in information science students may involve coding of the designed system. In such a case, students need an integrated development environment that provides support for writing, debugging, and interpreting code in the target programming language. An example of such is IDLE for Python. In the testing and evaluation phase, students need functionality and support for coding and analyzing the results obtained, an example of such a tool is SPSS. Finally, in the presentation phase, students need support to write up and present their research results and conclusions.

Table 1: Research support tools, adapted from (Ocharo & Hasegawa, 2017)

Research Phase	Examples of Research Tools
Setting a research theme	Mind mapping tools, brainstorming tools, e.g.Mindmeister, MindMup, Coggle etc.
Survey of related literature	Reading and citation managers e.g. Mendeley, Readcube, etc.
Design and modeling of system	Graphics applications e.g. Microsoft Visio
Development	Integrated Development Environments (IDEs) e.g. IDLE for python
Testing and Evaluation	Statistical software packages e.g. SPSS
Presentation/publication of results	Word processors e.g. Microsoft Word, Presentation software

Academic writing is at the core of the presentation of research results. Academic writing constitutes writing logically organized research papers, essays or reports that are presented in a well-structured, concise format using the academic style of writing in third-person style, passive writing, proper citations, etc. The importance of quality academic writing cannot be overstated. This is because it is an opportunity to share the research accomplishments with not only the fellow researchers but also with the public at large. Indeed, achievements in the academic world are measured by the number of quality research articles published. Students write essays, reports, conference papers, journal articles and finally, theses or dissertations. To produce quality writing, students need to have sufficient academic writing skill. Academic writing skill is the ability to write logically organized papers, essays or reports in a well-structured, concise format (Ocharo, Hasegawa, & Shirai, 2017). The student has to be able to present complex ideas simply and objectively.

Writing is a cognitive activity consisting of three stages: generating, translating and reviewing (Flower & Hayes, 1981). Generating involves generating ideas and information that might be included in the article, setting goals and organizing the retrieval of information from memory. Translating is the process of converting ideas into textual output. Reviewing involves two processes: evaluating and revising the text.

Much of the research in the area of academic writing tools is focused on equipping students with grammatical skills (especially in the case of English as a second language) and technical or scientific writing skills. Some of the tools also assist students with the generating stage. However, there is not much research on the use of software tools to assist students during the reviewing process of academic articles. This is the motivation for focusing on the revision aspect of the reviewing process, with the aim of improving the students' revision skill.

Revision skill is important in order to improve the quality of an academic article. Revision is more than just editing or proof reading a document to fix spelling or punctuation. It might involve restructuring the arguments, reviewing the evidence, refining or even reorganization of the entire article. For the purpose of this study, the latter definition of revision applied i.e. changes to the text to improve the strength of an argument, the overall structure and content.

The challenge is that it is a difficult skill to learn because like any cognitive skill, learning it is an implicit process. It may be learned directly from language teachers or by co-authoring papers with supervisors and other students in the same learning environment such as a common laboratory (Hyland, 2000). However, research students are often pressed for time and may end up copying the writing style from bibliography. The problem is that these published articles they learn from are in their final form, so the students have no way of learning from the revision process that led to the final articles.

In addition to providing a way for students to learn from the revision process itself, there is need to consider the differences in the skill levels of the students. Research has shown that in the writing and subsequent revision process, there are important differences between novices and experts (Hayes, Linda, Schriver, Stratman, & Carey, 1987). Novices may be those whose writing is judged to be of poorer quality, such as first year college students. Experts may be those whose writing is judged to be of higher quality, such as professors or advanced college students (Kozma, 1991).

Another important reason why it is difficult for educators to teach students revision as there is no standard revision curriculum. While there are many books on academic writing, it is quite difficult to find a general guide suitable for everyone because in academic writing, different fields have different writing styles. Much of scientific research is organized around laboratories and each laboratory has its own specific way or style of writing and presenting research results. Therefore, a revision corpus should be gathered that is unique to each laboratory. However, the amount of literature produced by one laboratory each year may be quite limited. The limited amount of data means it is difficult to apply usual machine learning approaches to extract laboratory knowledge from such a corpus to support current and future students.

There is a lot of research into the use of a corpus to support students in the process of academic writing. However, in most cases the articles in the corpus are in their final form. This means that students have no way of learning from the actual revision process that led to the final documents. The revision corpus for a laboratory should also contain the drafts that led to the final copy, as well as any feedback from reviewers or supervisor that was useful in the revision process. This feedback is usually in the form of comments embedded in the drafts, and can include trivial comments to correct fix spelling and grammar or more important comments to change the structure of the arguments. How to identify these important comments by using machine learning techniques is one of the interesting problems in this research area.

1.3 Problem Statement

In the course of this study, four problems were identified in the research area of supporting students during the revision process.

1. First, a lot of research into software tools focuses on other aspects of writing such as grammar but there is not much on technology-based tools that support the explicit instruction of revision skill. Revision skill is an implicit cognitive process therefore students may find it hard to learn. They may learn this skill directly from a traditional classroom, but research students are often pressed

for time therefore they don't have the chance to be in a traditional classroom. They may also learn revision skill by working with senior researchers in the same laboratory; however this can be a limitation in cases where there is no frequent interaction with other researchers such as in the case of distance or online learning.

2. Secondly, existing tools in this area do not take into account the implicit cognitive processes involved during writing and revision by novice, intermediate or advanced students. They are not adaptive to the skill level of the students. Previous research in the cognitive processes involved in writing have shown that novice students have different needs than more skilled students.
3. Thirdly, while many tools make use of a corpus to support the revision process of students, in most cases the articles in the corpus are in their final form. This means that students have no way of learning from the actual revision process that led to the final documents. Moreover, when students graduate from institutions, the knowledge and skills that they possess, such as revision skill, are lost unless their documents are archived, the knowledge extracted (for example, through machine learning and data mining methods) and passed on to incoming students.
4. Students usually receive feedback from reviewers or supervisor to improve their drafts during the revision process. This feedback is usually in the form of comments embedded in the drafts, and can include trivial comments to fix spelling and grammar or more important comments to change the structure of the arguments (content-related comments). How to identify these important comments is one of the interesting problems in this research area. Because the focus of this dissertation is how to improve revision skill, it is important to identify the content-related comments as they are most related to revision skill.

1.4 Research Objectives

The following are the research objectives considered in line with the problem statement identified in this dissertation:

1. To investigate the theories of learning cognitive revision skill and how to apply them in the absence of a traditional classroom

2. To design input, content and output necessary for the development of a tool to support the revision process in academic writing
3. To propose a way for the tool to achieve adaptation to the revision skill of students
4. To apply machine learning techniques to classify the comments in revision article drafts as either content-related or not

To realize the research objectives, this dissertation designed and developed a software tool (TRONA – ***T****opic* ***R****elated Revisi****ON*** ***A****ssistant*) that incorporates the Cognitive Apprenticeship Theory (CAT) (Collins, Brown, & Newman, 1989), one of the most widely accepted theories for teaching cognitive skills in an explicit way. Incorporating the CAT in software tools to support the learning of cognitive skills has been implemented fairly successfully in various areas such as in learning metacognitive skill (Kashihara, Shinya, Sawazaki, & Taira, 2008) and in learning programming (Chee, 1995). The CAT is a prominent model of instruction that proposes teaching of cognitive skills through the methods of modeling, coaching, fading, scaffolding, articulation, reflection and exploration. TRONA incorporates some of the CAT methods of instruction to provide adaptive hints to students revising their own texts. The hints provided are different, depending on the revision skill level of the student. If current students have a problem resolving the comments during their own revision, they can upload their drafts with comments into TRONA and it will show them hints by displaying how previous students resolved similar comments. The hints in this case means that the interface displays information in such a way that clearly shows how the revised text changed in response to the comments. TRONA is able to do this because it utilizes a revision corpus of academic articles collected from students in one laboratory. The articles include the initial drafts to the final drafts, and the corresponding comments from the supervisor that led to the revision. In this way, TRONA enables current students to observe and learn from the revision process itself. TRONA also makes use of machine learning techniques to classify the comments in the drafts. All the comments in the drafts that the students upload are classified as content-related or not; and thus TRONA enables students to first focus on resolving the content-related comments.

1.5 Research Questions

To address the research objectives identified in this study and to realize TRONA, the following four research questions were posed and addressed:

1. How can students improve their revision skill?
2. How can a tool be designed and developed to support the improvement of revision skill?
3. How can the tool be adaptive to the various needs of students depending on their revision skill level?
4. How can machine learning techniques be applied to classify the comments in revision article drafts as either content-related or not?

1.6 Dissertation Outline

Chapter One provides the background to this dissertation and discusses the basis for research support systems in general, and tools that support academic writing skill in particular. It places the context of this research as a tool to support revision skill in academic writing. The problem statement, research objectives and research questions are stated.

Chapter Two addresses the first research question. It follows up on the introduction by critically reviewing previous research literature in the area of tools to support academic writing, especially the revision aspect. This chapter elaborates the originality of this research by comparing and contrasting related literature regarding tools that support the revision process. The resulting knowledge gap is discussed as well as how this research can contribute to bridge that gap.

Chapter Three addresses the second research question. The revision process model that students go through is presented. This is followed by a discussion of the functional requirements and the content (a revision corpus of article drafts and corresponding reviewer comments) that is necessary to support students in the learning of revision skill. Finally, this chapter presents the design, architecture and development of a tool that supports revision skill (TRONA).

Chapter Four addresses the third research question. In this chapter, the need for adaptation is discussed as well as how to apply the Cognitive Apprenticeship Theory (CAT) to provide adaptation to the revision skill level of students. This is illustrated with some examples in TRONA. This chapter also discusses how to use content-related comments as item models in the Item Response Theory (IRT), which is used to estimate the revision skill of students. The evaluation of IRT is also discussed.

Chapter Five addresses the fourth research question and presents the rationale for using machine learning to classify reviewer comments in research article drafts as content-related or not. It is important to classify the comments because content-related comments are at the core of the IRT item models; further, classification will enable students to be aware of, and to focus on, global revision which may improve their revision skill.

Chapter Six is the conclusion and future work. This chapter begins by restating the research questions and providing a summary of how the empirical findings addressed the research questions. Thereafter, the chapter discusses the implications of the study. Finally, the chapter discusses the limitations of the study and ideas for future research.

2. Improvement of Revision Skill

2.1 Introduction

This research combines knowledge from several areas of study including pedagogy (education), psychology/cognitive studies and the use of computers in education. Therefore, this chapter follows up on the introduction chapter by critically reviewing previous research literature in the area of tools to support academic writing. It then explores the studies implementing cognitive learning theories in software tools that support learning of various skill, especially revision skill. Thereafter, the originality of this research is elaborated by comparing and contrasting related literature. The resulting knowledge gap is discussed and identified as an opportunity where this research can contribute to bridge that gap.

Section 2.2 is a discussion of the cognitive processes involved during writing and revision. Through this discussion, the need for adaptation to revision skill level in a tool that supports revision becomes apparent. Tools that support the writing process are reviewed in detail in section 2.3 and their merits and shortcomings critiqued as to whether the tools support the cognitive processes of students, whether they are adaptive or not, etc. Finally, the need for a corpus is discussed in section 2.4 and section 2.5 is a summary of the knowledge gap discovered during this review of related literature.

2.2 Cognitive processes during writing and revision

It is a well-established fact that quality academic writing is important in research. Academic writing is the kind of writing students in higher education do when writing reports, research papers, or dissertations. It is different from everyday writing or formal writing in three ways:

- Academic writing involves writing technically – this means adopting a formal, impersonal style of writing by avoiding casual language such as contractions or informal vocabulary. There is also need to develop and continuously learn the technical vocabulary of the concepts and objects specific to the student's discipline e.g. physics or chemistry.

- Academic writing in each discipline not only differs in the technical vocabulary, but also the style. For example, in social sciences, the writing tends to be quite long as the writers try to build their arguments and may also involve some subjectivity. However, in physical sciences, short, factual paragraphs are expected and the use of personal pronouns/objective views (E.g. "My view is…") may not be acceptable. The emphasis in the latter case is on facts.
- The structure – different disciplines have different structures for each type of academic document e.g. report, case study. Students need to learn the right or acceptable structure for the type of academic document they are writing.

What cognitive process do students go through during the writing process (for academic wring or otherwise)? Writing is a complex task and before exploring the tools that support it, it is important to understand the cognitive processes involved in it.

The writing process is constrained by information in long-term memory such as topic-relevant information, knowledge and expectation of audience, and grammatical rules and persuasive strategies (Kellogg, 1988). Writing is centered around the cognitive processes of generating, translating and reviewing (Flower & Hayes, 1981). Please refer to Figure 3. Generating involves planning and coming up with ideas and information that might be included in the composition, setting goals and organizing the retrieval of information from memory. Translating is the process of converting ideas into textual output. Reviewing involves two subprocesses - evaluating and revising the text. These are discussed in detail below.

1. In the generating phase, the author takes the writing assignment and long-term memory as input, which then produces a conceptual plan for the document as output. This phase includes the sub-activities of planning (coming up with ideas), organizing (arranging those ideas logically in one's head), and goal setting (determining what effects one wants to achieve and modifying one's generating and organizing activities to achieve local or global goals). Students who are not experts in academic writing might struggle with the organizing or goal setting tasks – and for this reason they usually work together with their supervisors or senior colleagues in the laboratory.

2. Translating takes the conceptual plan for the document and produces text expressing the planned content. Students who are writing in a language that they are not fluent in often struggle with this phase of writing.
3. In reviewing, the text produced so far is read, with modifications to improve it (revise) or correct errors (proofread). In this dissertation, the term revision may be used to refer to the overall reviewing process or the specific revision process itself. Students often struggle with revision when writing academic articles because they not only need to consider their own writing goals, but must also revise their texts to persuade readers of their arguments, and thus must pay careful attention to the demands of the audience if they are to be successful (Fahnestock & Secor, 1988).

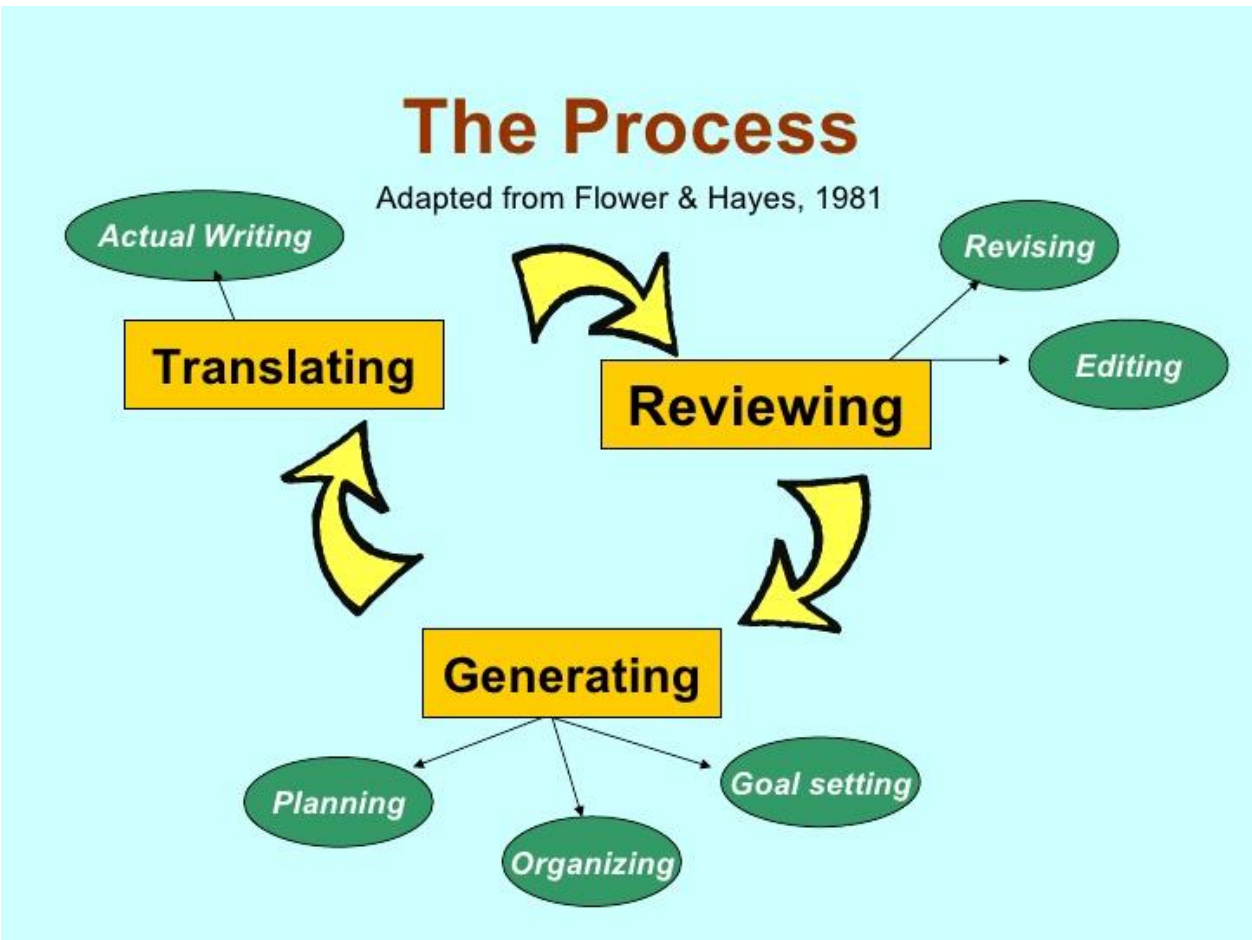


Figure 3: The cognitive processes involved in writing (Flower & Hayes, 1981)

When it comes to reviewing, more than proofing skills, revision skill is needed in order to improve the quality of written articles. Revision is more than just proofreading or editing an article. Revision involves restructuring the arguments, reviewing the evidence, refining or reorganization of the entire article (Ocharo, Hasegawa, & Shirai, 2017). In other words, skilled revising is the kind of revision that leads to meaning-level changes and requires additional reading strategies. Skilled revising involves understand the implicit and explicit purposes of revising, activating the relevant background knowledge, allocating attention to major content, evaluating the content for internal consistency, monitoring ongoing comprehension, and drawing and testing inferences (Palinscar & Brown, 1984).

This research makes a distinction between revision knowledge and revision skill. Revision knowledge is the theoretical or practical understanding of what is expected during the revision process – for example, rewriting a draft in order to improve its quality. So a student might have knowledge what is expected in the revision process, such as reorganizing the structure to improve logical coherence or proofreading to improve grammar and clarity. This doesn't mean that the student will actually be able to carry out the revision successfully. For that, they need revision skill.

Revision skill is the proficiency developed through training or experience. Revision skill is usually something that has been learned during experience of revising previous drafts. Therefore, students can develop skills through the transfer of knowledge during practice.

In their research (Hayes, Linda, Schriver, Stratman, & Carey, 1987), Linda et.al. attempted to map further the cognitive processes involved during revision as shown in Figure 4. Acquiring implicit cognitive proficiency in these processes is what it means to acquire revision skill. These are task definition, evaluation, goal setting and modification of text or plan. The task definition for revision specifies the goals of the reviser and the features of the text that should be examined e.g. local, global, or both. It also specifies how the revision process should be carried out. The evaluation process applies the goals and criteria of the task definition to the texts and plans.

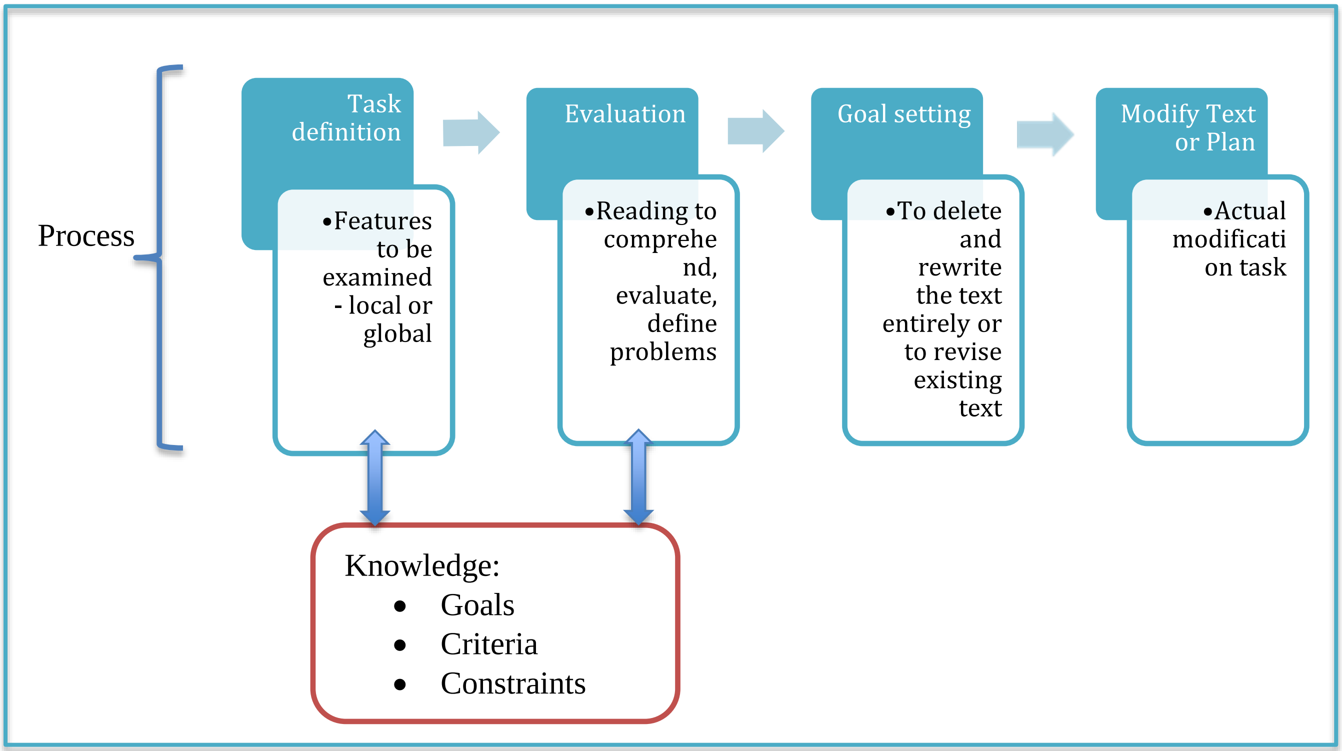


Figure 4: The process model of revision. Adapted from (Hayes, Linda, Schriver, Stratman, & Carey, 1987)

An important task during task definition is the selection of features to be examined i.e. local or global features. Global revision focuses on organization, development of ideas, and overall consistency. Global revision leads to larger conceptual, rhetorical, and structural revisions that would most significantly improve the quality of a paper. Local revision focuses on editing for word choice, sentence fluency, grammar, spelling and punctuation. The researchers (Hayes, Linda, Schriver, Stratman, & Carey, 1987) noted that there are important differences between novices and experts. Experts make more revisions than do novices, and experts revisers also attend to more global revising problems than do novices. Novices may be those whose writing is judged to be of poorer quality, such as first year college students. Experts may be those whose writing is judged to be of higher quality, such as professors or advanced college students (Kozma, 1991).

What consists “quality writing”? While the definition of quality in academic writing might be different depending on the discipline, most research papers require careful attention to the following stylistic elements (Hartley, 2008):

- **Formal and logical structure -** the writing must be cohesive and possess a logically organized flow of ideas; this means that the various parts are connected to form a unified whole. The introduction should include a description of how the rest of the paper is organized.
- **Neutral tone -** the paper should state the strengths of the arguments confidently, using language that is neutral, not confrontational or dismissive.
- **Vocabulary choice** – the author should use concrete words that convey a specific meaning. If this cannot be done without confusing the reader, then the author must explain the meaning of the word within the context of how that word is used within a discipline.
- **Grammar and spelling** – there should be no spelling or grammatical mistakes. The language used should be precise and formal.
- **Academic conventions** – Sources of information must be cited in the body of the paper and a list of references as either footnotes or endnotes should be provided. It is essential to always acknowledge the source of any ideas, research findings, data, or quoted text that have been used in a paper as a defense against allegations of plagiarism
- **Thesis-Driven** – academic writing is “thesis-driven,” meaning that the starting point is a particular perspective, idea, or position applied to the chosen research problem, such as, establishing, proving, or disproving solutions to the questions posed for the topic.
- **Complexity and Higher-Order Thinking -** Academic writing addresses complex issues that require high-order thinking skills to comprehend (e.g., critical, reflective, logical, and creative thinking). This implies the ability to explain complex ideas in a way that is understandable and relatable to the topic being presented. This includes cognitive processes that are used to comprehend, solve problems, and express concepts or that describe abstract ideas that cannot be easily acted out, pointed to, or shown with images. The writing should be a summarization of complex information into a well-organized synthesis of ideas, concepts, and recommendations that contribute to a better understanding of the research problem.

2.3 Analysis of Writing Support Literature and Tools

There has been a lot of research on writing tools to assist in the process of academic writing. Early research into important linguistic aspects of a good writing style such as readability, sentence and word length, sentence type, word usage and sentence openers (Cherry, 1982), enhanced the capability of word processors beyond mere spell-checking. Examples of word processors include *Microsoft Word*©, *OpenOffice*©, etc. In addition to word processors, grammar checking tools are available that can automatically recognize and clean up grammatical errors in writing (Blake, 2011). An example of an advanced grammar checker is *Grammarly*©. *LaTex*© is an example of a tool that goes beyond word processors and makes it easier to create complex technical documents suitable for scientific writing i.e. support for equations, figures, tables, citations, etc. (Wright, 2010). However, the quality of writing cannot be evaluated by structural or grammatical accuracy alone (Narita, 2012).

Word processors and grammar checking tools offer little support to the planning, retrieval and organization of knowledge or the evaluation and revision of ideas (Kozma, 1991). Therefore, there is need for tools that are designed to support idea generation, planning and revision; in other words, tools that support the cognitive processes involved in writing. In recent years, with the understanding of the limitations of word processors, the focus of many researchers has been tools that help to improve the students' overall competency in academic writing. This section examines some examples of such tools.

Much of the research in the area of academic writing tools is focused on equipping students with grammatical skills (especially in the case of English as a second language) and technical writing skills. Some of the tools also assist students with the planning or idea generation phase. These examples are summarized in Table 2 and analyzed further in subsequent subsections.

Examples of computer tools to support planning include idea prompters and organizers, structure or outline organizers, etc. Examples of tools to support the translation from ideas into text include English-as-a-Second-Language tools, abstract summarizers, etc. To support the review process, there are several tools that provide spelling and grammar checking functions, text analysis, etc. These tools are described further in the respective subsections i.e. subsection 2.3.1, subsection 2.3.2 and subsection 2.3.3. The target of this research is a tool that supports the overall revision of ideas, not just proofreading.

Table 2: Examples of Writing Support Tools

Cognitive process in Writing	Writing Support Tools
Generating	Idea organizers, prompters, outliners, and tools to help with the structure, concept mapping or drafting, etc.
Translating from ideas to text	Word processors, English-as-a-second language tools, abstract summarizers, etc.
Review	Evaluating - spelling and grammar checkers, text analyzers, collaboration tools, etc.
	Revision - Tools to support the overall revision of ideas (*this paper is a contribution in this subsection*)

2.3.1 Planning and Idea Generation Support Tools

Previously, some researchers (De-Smet, Broekkamp, & Brand-Gruwel, 2011) developed an outline tool that helped students to engage in the planning aspect of writing. They found out that electronic outlining improved the quality of students' argumentative texts and helped students to engage in the planning aspect, which they would not otherwise do unless explicitly instructed. The outline tool is clearly an example of a tool that explicitly targets one of the cognitive processes involved in writing. However, this tool does not take into account the differences between novices and experts.

Another tool that help students in knowledge construction, which is necessary for planning, was developed by (Chang & Kuo, 2011). It combined a corpus with genre analysis to develop research-based online teaching materials for English as a Second Language (ESL) graduate students in computer science. Similar to the planning tool reviewed prior, this tool also does not consider the differences between novices and experts.

Additionally, (Neuwirth & Kaufer, 1989) investigated the role of hypertext-based interfaces in supporting the building of knowledge structures by using graph/tree representations of the written statements. They note that knowledge structures are necessary in the planning stage of writing and that there are several differences between novices and experts. Their results suggest that it is important to take note of the cognitive skill of the students when designing a tool to support the writing process.

2.3.2 Tools that Support the Translation of Ideas into Text

Other researchers have developed tools to help students translate their writing plans and ideas into text. Many of these tools help students writing in English as a Second Language (ESL). Some researchers such as (Aluisio & Gantenbein, 1997), applied SFL - Systemic Functional Linguistic model of language to help students with grammatical revisions. In a related research, (Grami & Alkazemi, 2016) provide examples of the correct usage of sentences accompanied by statistical feedback from web-based applications for learners to emulate. They found that results improved for those who used the software. However, their focus is on sentence-level word patterns. This implies that there is no consideration for the underlying cognitive processes. In addition, all students with different abilities are treated in the same way.

Similarly, WriteAhead is an abstract writing assistant developed by (Tzu-Hsi, Wu, Chang, Boisson, & Chang, 2015). It provides ESL learners with suggestions such as collocations or transitional words - using domain-specific corpora of abstracts from the Web. Findings showed that the experimental group wrote better, and most students were satisfied with the system concerning most suggestion types, as they could effectively compose quality abstracts through provided language supports. The limitation in this case is that the evaluation was qualitative.

Additionally, there are several tools, such as Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) and “The tool for the automatic analysis of text cohesion” - TAACO (Crossley, Kyle, & McNamara, 2016), that utilize computational linguistics and NLP to analyze text for local and global cohesion, language and readability. These tools can allow students of different levels, to analyze their own texts cohesion and readability, thus the quality of their writing. They also enable researchers to collect information about corpora with less effort. Furthermore, the researchers of (TAACO) proved that expert judgments of text coherence and quality are positively correlated with global cohesion.

2.3.3 Revision Support Literature and Tools

Revision continues to be a difficult process to model and to teach. Revision is a complex cognitive task where the writer must reflect over time the changes needed to make the draft congruent with a writer's changing intentions. Various researchers have undertaken ways to help students of all ages – from elementary to college students – to improve their revision skills. All the studies indicate that less skilled writers attend mainly to surface level features, whereas skilled writers show more concern for content and larger segments of discourse, revising on both local and global levels.

Teachers can provide direct instruction to students to help them with their revision. However, (Witte, 2013) found out that strategies the teachers used when revising their own texts were not the strategies they used with students. There is also a lack of revision curriculum. Another researcher (Sommers, 1980) already established that teacher comments on college students' writing were usually text-specific and therefore not very helpful. They often took students' attention away from their own purposes and focused it on those of the teacher. After the teacher participants in the research by (Witte, 2013) underwent training, the strategies they used were more aligned when teaching writing revision to students in their own classrooms. Teachers need to help students realize that revising is not just about fixing grammatical and surface errors, but also refers to improving the strength of an argument, overall structure and content. Teachers can do this by providing more specific comments that emphasize the whole text over its parts (Lehr, 1995) and also by providing enough time for students to revise.

Struggling writers can benefit from designing instruction that teaches them to learn to evaluate writing and revising effectively (MacArthur, 2009). Further, research by (Kobayashi & Rinnert, 2001) indicate that explicit instruction plays an important role in students' essay level revisions and use of correction strategies. Second language (L2) writers can learn to improve essay level coherence problems through instruction combined with the experience of writing and revising in an instructional setting. It was also demonstrated by (Kakh & Wan Mansor, 2014) that a series of disciplinary-based writing tasks to facilitate revision helped participants to develop cohesion and coherence in their texts. With explicit instruction, even reviewing one's own texts is sufficient to develop writing skills (Wakabayashi, 2013).

The challenge with direct instruction by teachers is that it may not be suitable for graduate students who are focused on research. The time limitations mean that they may not have the opportunity to learn revision skill in the traditional classroom. Researchers have explored other ways of helping students to improve their revision skill, such as through peer review or collaboration. Through reviewing, students too can learn overall writing skills. This is because reviewing is a problem solving activity that engages problem detection, diagnosis and solution generation. The results of (Cho, Schunn, & Kyungbin, 2007) showed that reciprocal peer review improved the students' own writing skills.

Utilizing a collaboration wiki can enlarge young writers' experience of the process of composition and revision, both through their own efforts and observing the process in others (Pifarré & Fisher, 2011). A wiki is a collaborative website whose content can be edited by visitors to the site, allowing users to create, edit and revise collaboratively. Using wikis to encourage collaborative writing and revision is an example of the application of the power of technology to support the teaching and learning of revision skills. Utilizing dictionaries or an online corpus can also enhance revision skills. In a 2015 study by (Mueller & Jacobsen, 2016), they established that online corpus consultation, even for learners at fairly low proficiency levels, appear to have practical benefit in enhancing learners' ability to solve language issues. Furthermore, (Kakh & Wan Mansor, 2014) also illustrated that corpus-based tasks and critical-analytical reading helped students to find proper samples for writing in the disciplines. The need for a corpus is discussed further in section 2.4.

A search for more technology-based tools to support the revision process revealed that there is not much research on the use of software tools to assist students in the revision process of academic articles. Modern tools have emerged to help teachers in the teaching of revision, but the new tools are lacking in explicit revision instruction (Witte, 2013). Technology can serve as a motivator for participating in revisions, and can be especially useful in the case of research students who do not have time to learn revision skills through the traditional classroom or through peer review or collaboration.

Therefore, the question is: how can technology-based tools improve revision skills for research students? How can such tools provide explicit revision instruction? From the literature review, this can summarized (although not exhaustively) as:

- The tool should provide a way for the students to focus on revision of the whole text over its parts.
- The tool should be adaptive to the revision skill level of students – novices, intermediate or advanced
- The tool should provide a corpus of articles in the discipline for students to learn from, because they may not have time to learn from teachers in a traditional classroom

2.4 The need for a corpus

One important question to consider when designing a tool to support the writing process of students is what constitutes the content to be used for learning. Several studies have highlighted the need for corpora, which is a collection of text material assembled for the purpose of linguistic research. Studies such as (Friginal, 2013) and (Yoon & Hirvela, 2004) have highlighted the potential of corpora in ESL writing.

A target corpus can provide resources for both teachers and students, facilitating the development of learning materials and increasing the motivation of learners to study the target genre (Chang & Kuo, 2011). The message of this collection is that language use is purposeful and culture specific and that small corpus analysis is an effective method of linguistic investigation (Ghadessy, Henry, & Roseberry, 2001).

Another researcher (Charles, 2014) demonstrated that a specialized and personal corpus is useful to students who check their grammar and lexis while composing and revising. In the experiment, 93% of the students considered that corpus use had improved their academic writing. However, the evaluation was qualitative. In addition, the articles are in the final form, and no adaptation is provided to students beyond their personal preferences, since they are in charge of creating their own corpora.

Other studies, for example, (Gaskell & Cobb, 2004), (Gilmore, 2008), (Todd, 2001) – report positive outcomes for the success rate of error correction/self-correction when using a corpora. (Narita, 2012) also concluded that a corpus-based tool of previous students' work can be vital for improving second language learners' grammatical knowledge.

In summary:

- A corpus can be effective in improving the students' grammatical knowledge, as well as improving the success rate of error correction
- Students can also improve their overall academic writing skills by learning from articles in the corpus
- A specialized and personalized corpus is more effective

The challenges it that the articles in these corpora are all in the final version. Therefore, students do not have a chance to learn from the revision process that leads to the final copy. If they face a problem during the revision of their own articles, they may be stuck.

2.5 Summary of Gap and Opportunities

The review of related work highlighted some gaps that motivated this study. These are summarized below:

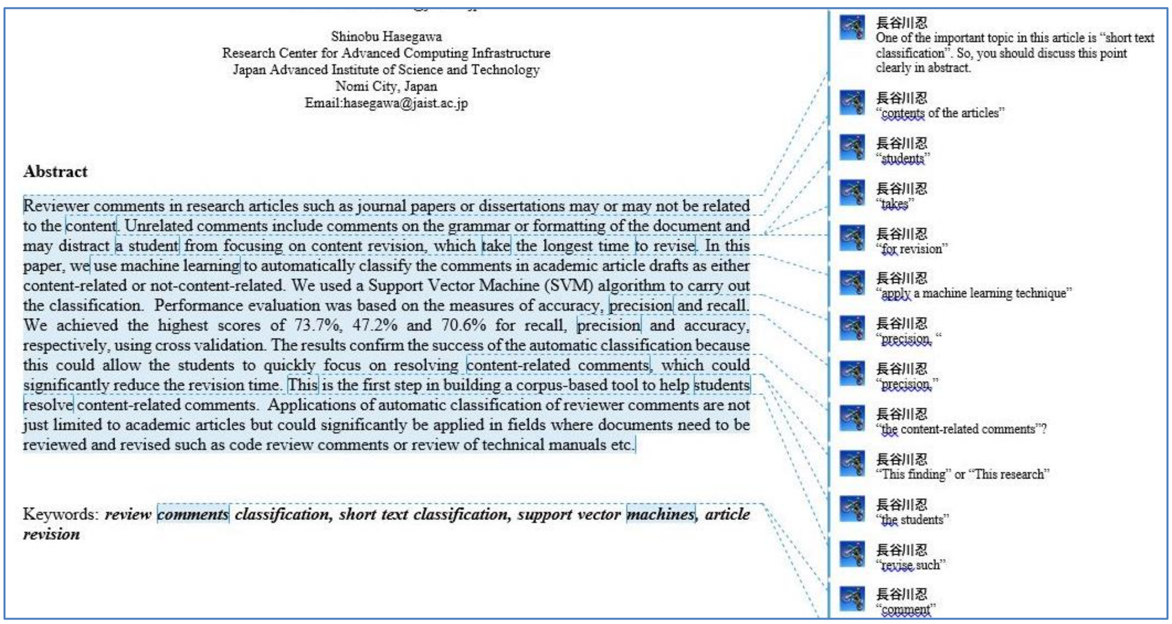
- While there is quite a number of writing tools to assist students, there is not much focus on the use of software tools to assist students in the overall revision process of academic articles.
- A detailed analysis of the existing tools further revealed that many of the tools do not take into account how to support the underlying cognitive processes involved in writing; and that they are also not adaptive to the needs of novice, intermediate or advanced students.
- A study of the use of corpora revealed that many of the published articles they contain are in their final form, so the students have no way of learning from the revision process that led to the final articles.

3. Design and Development of TRONA

3.1 Introduction: Modeling the writing and revision process

This chapter begins by first modeling the revision process. When it comes to revision in the academic setting, the supervisor acts as the prompter, sets the goals of the revision and gives comments to that end.

In many cases, the software often used for creating and modifying the drafts is a word processor application such as Microsoft Word. Microsoft Word has a comment feature that allows the supervisor to give feedback to the student to improve the draft. An example is shown in Figure 5.



Shinobu Hasegawa
Research Center for Advanced Computing Infrastructure
Japan Advanced Institute of Science and Technology
Nomi City, Japan
Email:hasegawa@jaist.ac.jp

Abstract

Reviewer comments in research articles such as journal papers or dissertations may or may not be related to the content. Unrelated comments include comments on the grammar or formatting of the document and may distract a student from focusing on content revision, which take the longest time to revise. In this paper, we use machine learning to automatically classify the comments in academic article drafts as either content-related or not-content-related. We used a Support Vector Machine (SVM) algorithm to carry out the classification. Performance evaluation was based on the measures of accuracy, precision and recall. We achieved the highest scores of 73.7%, 47.2% and 70.6% for recall, precision and accuracy, respectively, using cross validation. The results confirm the success of the automatic classification because this could allow the students to quickly focus on resolving content-related comments, which could significantly reduce the revision time. This is the first step in building a corpus-based tool to help students resolve content-related comments. Applications of automatic classification of reviewer comments are not just limited to academic articles but could significantly be applied in fields where documents need to be reviewed and revised such as code review comments or review of technical manuals etc.

Keywords: ***review comments classification, short text classification, support vector machines, article revision***

長谷川忍
One of the important topic in this article is "short text classification". So, you should discuss this point clearly in abstract.

長谷川忍
"contents of the articles"

長谷川忍
"students"

長谷川忍
"takes"

長谷川忍
"for revision"

長谷川忍
"apply a machine learning technique"

長谷川忍
"precision, "

長谷川忍
"precision,"

長谷川忍
"the content-related comments"?

長谷川忍
"This finding" or "This research"

長谷川忍
"the students"

長谷川忍
"revise such"

長谷川忍
"comment"

Figure 5: An example of the typical comments found in research article drafts

After a student writes the first draft of a research article, he/she sends it to the supervisor for feedback. The supervisor then inserts comments to help the student improve the draft and sends it back to the student. The student thereafter revises the draft based on the comments and sends it back to the supervisor for feedback... and so on until the final draft is approved. This can be modelled as shown in Figure 6.

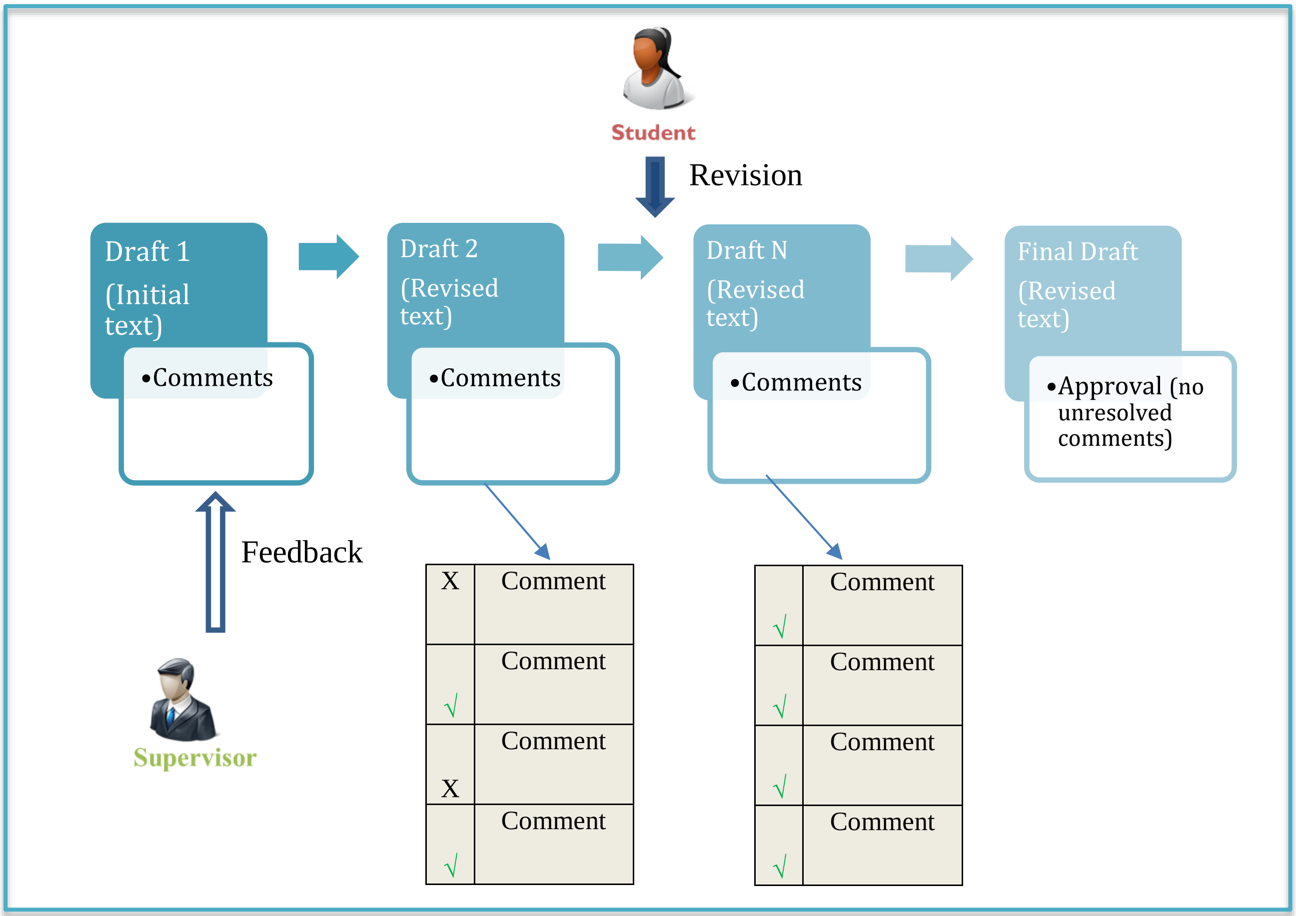


Figure 6: The revision process in academic writing

In the revision process model expressed in Figure 6, the four processes that entail revision skill as defined in section 2.2 would be mapped as followed:

i.) Task definition – a highly-skilled student would read the comments, make a decision on the comments that are related to global revision, and focus on revising those comments first. The student is aware of the need for global revision. The student has knowledge of the goals, criteria and constraints such as the time needed to carry out the revision.

ii.) Evaluation – the highly-skilled student re-reads their texts and the corresponding comments to understand the extent of the required revision, evaluates and defines what needs to be changed

iii.) Goal Setting – making a decision on whether to rewrite entire sections or just modify existing text

iv.) Modification of text or plan – actual modification of the text or of the revision plan

However, when the student is revising, he or she cannot query Microsoft Word to obtain an overview, at a glance, of information such as:

- The total number of comments, the resolved comments, the open comments, persistent comments in subsequent drafts, etc.
- Which comments are more important than others i.e. some comments are about simple spelling mistakes (local revision) while others may be suggestions to revise and reorganize the entire draft (global revision).
- The total time spent on revision so far, because there is no direct link between two distinct files that are separate drafts of the same document. Such functionality can help the student evaluate the constraints such as time.
- Commonly occurring comments in the drafts of other students so the student can know which common mistakes to look out for.
- Other metadata about the research phase such as due date, document and comment authors, etc.
- Sometimes they are not able to resolve these comments or they take too long to resolve them. However, word processors cannot provide them with hints to resolve such comments.

There is need for a complementary tool in the revision process that offers this kind of support. In addition, the tool should utilize a specialized revision corpus that will enable students to learn from the revision process of previous students by, for example, looking up similar comments and observing how those comments were resolved.

Following the challenges of using a word processor that were identified previously, the following are some of the requirements that were identified to support the revision process:

- A corpus of previous article drafts and the corresponding comments from the supervisor to be used to provide hints to students to resolve their own comments
- A checklist of commonly occurring comments in the previous drafts for the student to do a self-check

- Ability to track the revision process e.g. duration spent so far on revision, time remaining to the deadline, number of unresolved comments, etc.
- Ability to highlight the most important comments i.e. comments that trigger global revision of the contents
- Ability to be adaptive to the revision skill level of the student

The objective of TRONA is not to take over the functions of word processors such as Microsoft Word but to act as a complementary tool in the revision process by utilizing research output file metadata and a corpus of revision drafts.

3.2 Building the Revision Corpus

One of the objectives of this research was to design input, output and the content necessary to help students learn revision skill. From the literature review, the need for a specialized corpus was established. The main reason for the corpus is to enable students to learn revision skill by searching an archive of previous students' drafts, to observe the comments and the revision history. The revision process would be presented to the student in a way that clearly showed the changes to the text that corresponded to the comments. If current students have problems resolving their own comments, then they can search the corpus for similar comments and thereby learn how they can resolve their own comments in a similar way.

The content required for a revision corpus that fulfils the requirements of supporting the revision process includes:

- A collection of all the academic articles (reports, dissertations, conference papers, journal articles, etc.) in one laboratory. This includes not just the final articles but also the subsequent drafts from the initial to the final one.
- The comments extracted from the drafts – they are instrumental in the revision process.
- The text in the drafts that is covered by the comments. This is the highlighted part of the main body of the text that corresponds to the comment. Subsequently referred to as the comment range.

- Article drafts metadata as well as comments metadata. Article drafts metadata includes information such as author, creation date, modification date, version number, due date, purpose of article (e.g. conference or report) and other relevant metadata. For the comments, the metadata includes comment author, comment date, comment status, parent or child of comment (whether a reply or not), whether it was present in a previous draft or not… etc.

This content should be uploaded to a database or some other form of suitable storage that enables convenient access. For example, when the building the revision corpus in TRONA, articles written in English by previous students in our laboratory were collected. However, not just the final articles but also the subsequent drafts throughout the revision process. There were 42 articles in total which included dissertations, journal articles, conference papers, research proposals etc. There were 287 drafts in total, so each article had an average of 6.8 drafts. The articles were mostly in Microsoft Word format but some were in pdf. Table 3 shows all the manuscripts and the corresponding number of drafts contained in the corpus.

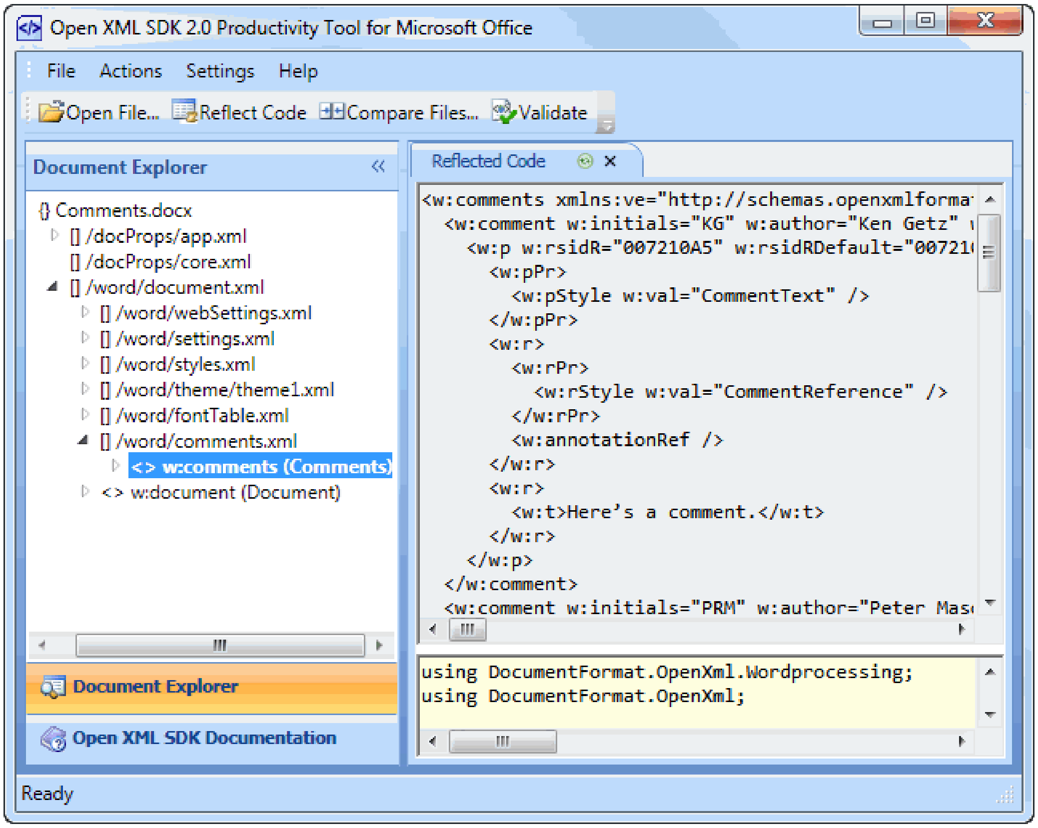


Figure 7: Exploring comments in a Word Document using Open XML SDK.

The comments in each draft were also extracted using Excel VBA (Visual Basic for Applications). A total of 7,786 comments was extracted from the drafts written in English by the previous students in our laboratory. Each draft had an average of 33 comments. The metadata of the drafts and the comments were uploaded to a MySQL Database which was used as the backend of TRONA to enable students to access and search the corpus. It is also possible to extract comments and the corresponding comment ranges by using XML (Extended Markup Language) parsers for Microsoft Word documents because they are saved in Office Open XML format. An example of such a parser is the Open XML SDK (Standard Development Kit) for Office (see Figure 7).

Table 3: The article drafts from our laboratory that were used to construct the revision corpus

Row Labels	Number of drafts
Conference	**168**
Ball_SIG-ALST	4
BallCCE2013	5
Carson_IARIA	3
Dandhi_HCII_Short	3
Dandhi_HCII2016Abst	5
Dandhi_TVET	4
Dandhi_TVET_Abst	2
Didin_ECGBL2013	14
Didin_GLS	4
Didin_ICCE2013	19
Didin_ICCE2014	7
Didin_ISAGA2014	6
Didin_SIG-ALST	4
Harriet_HCII2016	9
Harriet_ICCE2017	5
Harriet_IRAIA	6
Li_HCII2014	12
Li_ICCE2010	8
Li_ICCE2012	11
Li_ICCE2013	11
Li_ICCE2014	5
Dissertation	**41**
AsheryMastersThesis	4
CarsonMastersThesis	3
Dandhi_PreliminaryDefense	6
Harriet_ResearchProposal	7
Li_MasterThesis	5
Li_ResearchProposal	3
Mohamed_DissertationOutline	8
Mohamed_PreliminaryDefense	5
Examination	**9**
Harriet_PhDProposal	5
Li_DoctoralResearchPlan	4
Journal	**69**
Dandhi_IJECE	5
Dandhi_JCAL	5
Dandhi_JECR	6
Dandhi_JSiSE	3
Dandhi_JSiSE_2nd	7
Didin_RPTEL	6
Li_RPTEL	9
Li_RPTEL_Revision	7
Mohamed_BJET	4
Mohamed_Extention	6
Mohamed_JSISE	11
Grand Total	**287**

In the case of TRONA, the articles metadata, comments, comment range and comments metadata was uploaded to a MySQL Database. Part of the database schema for the MySQL Database is shown in Figure 8. There is a table to store the student's data (*studentprofile*) and another table to store metadata about the manuscripts (*papertitles*). The *drafts* table has the metadata about each draft, with a separate table for each comment contained in each draft. There are various relations between these core tables that are necessary to maintain the integrity of the database.

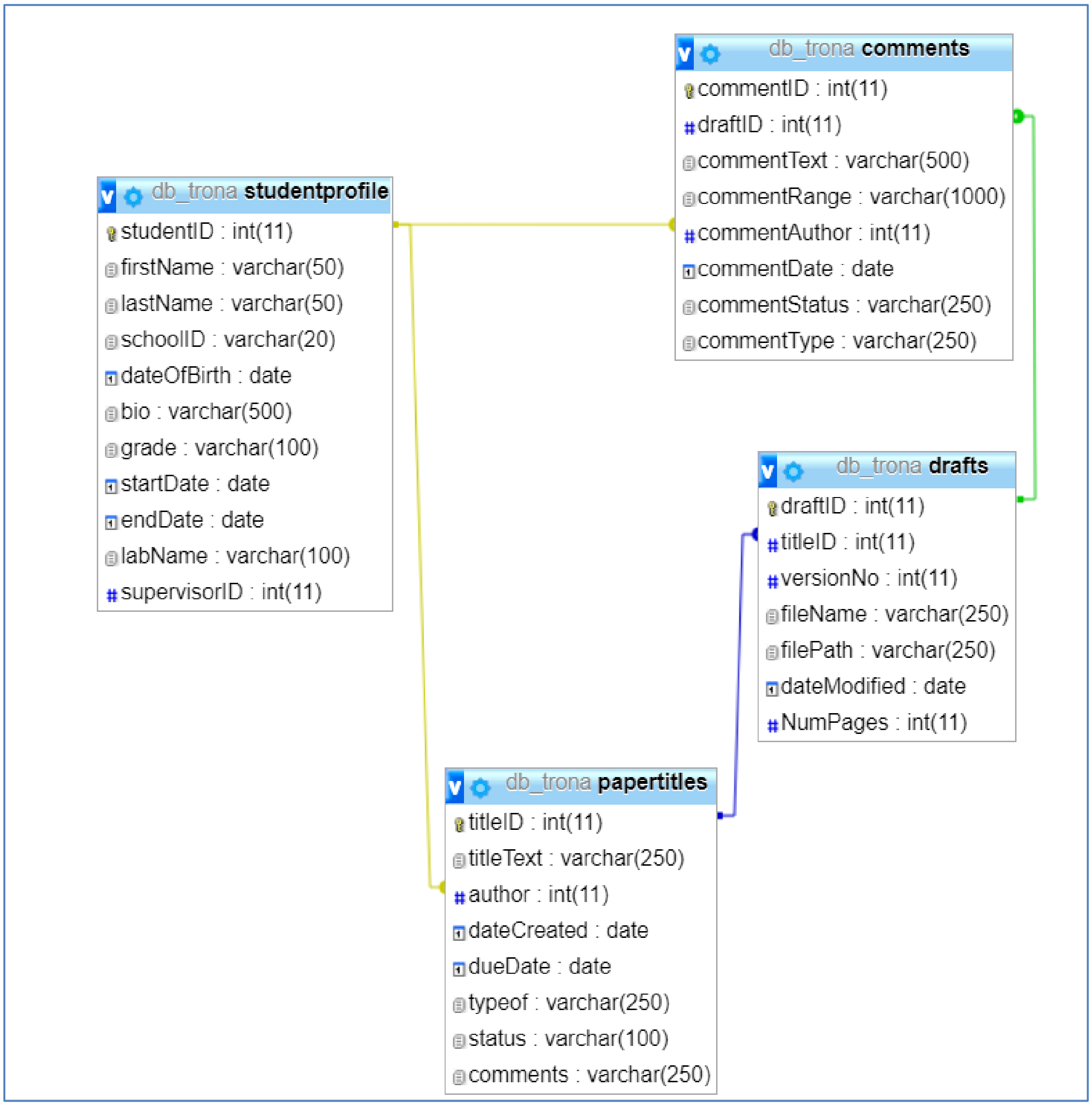


Figure 8: Part of the MySQL Database schema for TRONA

3.3 TRONA Design and Architecture

To fulfil the requirements identified, the tool TRONA contains the following modules presented through a web interface:

- **Profile manager** – provides the function to create, modify and delete a student user account.
- **Resource manager** – the student can upload their drafts and these are saved in the database. Therefore the student is able to reflect on their revision history.
- **Self-check List** of the most common comments.
- **Schedule Manager** – the student is able to view their own duration of revision, upcoming deadlines, etc.
- **Search function** – the student can look up previous drafts in the laboratory, such as reports, research proposals, conference papers, journal articles, etc. They are also able to manually search the comments' database to look up how those comments were resolved.
- **Adaptive Hints Provider** – provides adaptive hints to students to resolve the content-related comments by showing them similar comments in the database and observing how those comments were resolved.
- **Comment Extraction and Classification Module** – once the draft is uploaded, the comments are extracted and automatically classified. The important, content-related comments are highlighted.

The profile manager, resource manager, search and self-check list are straightforward implementations provided by the web framework, and this is discussed in section 3.4. However, providing adaptation, and extracting and classifying comments required further discussion and research effort. Therefore, they are discussed as follows:

- Providing adaptation – chapter 4
- Comments classification – chapter 5

The system architecture of TRONA is illustrated in Figure 9, which includes the web interface providing applications to the students, a web services layer, and a database layer.

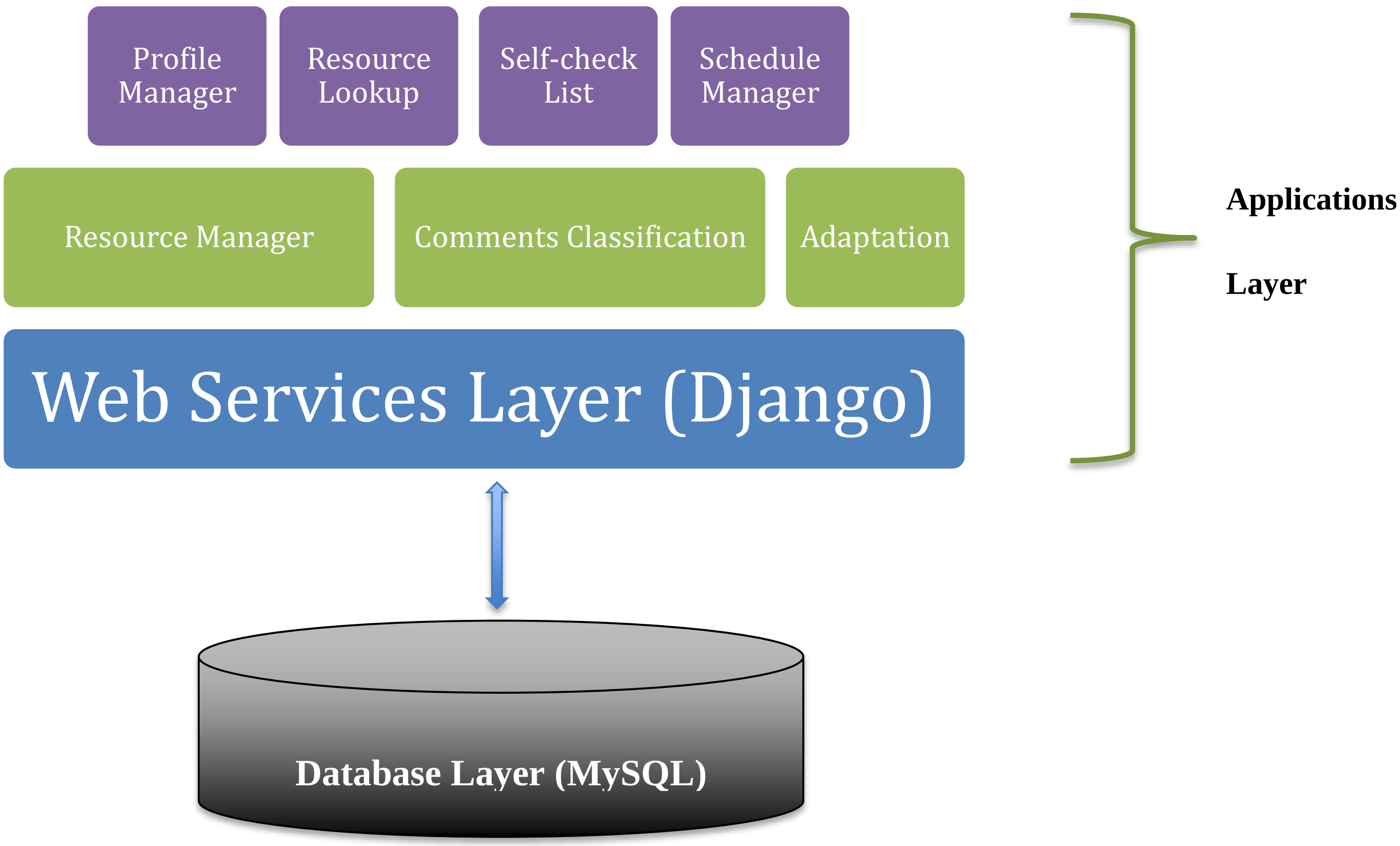


Figure 9: TRONA Architecture

The application layer contains all the modules and functionalities provided to the students by TRONA. The web services layer contains the services provided by the python framework, Django, that enables the website to run; such as session management, and it also provides communication with the database.

3.4 System Development and Overview

Figure 10 illustrates the web interface of the TRONA modules. The interface was developed using Django, a python web framework. To provide the above functionality, TRONA relies on a MySQL Database containing a revision corpus of draft articles, comments and other metadata.

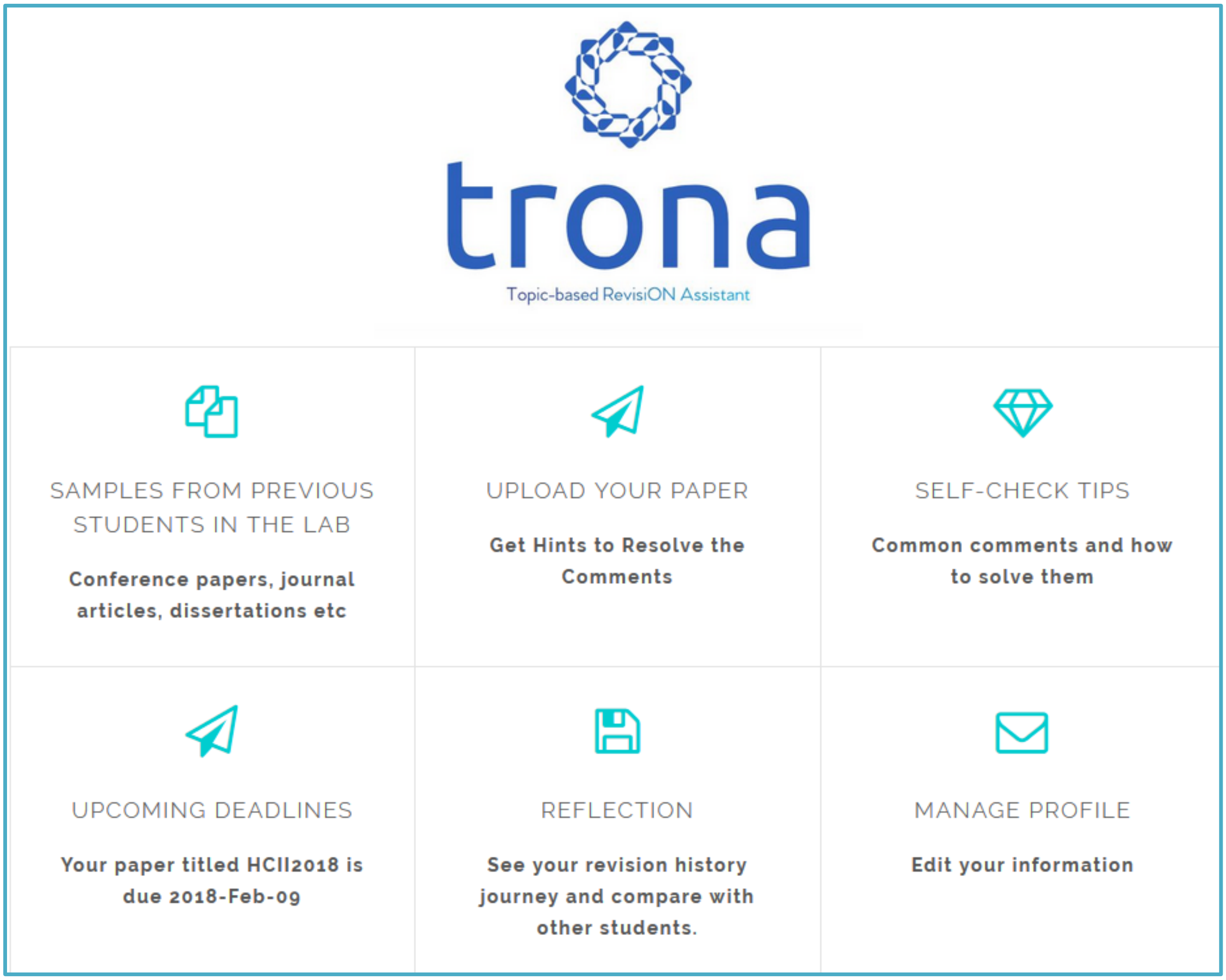


Figure 10: TRONA's web interface showing the modules provided

The functionalities provided in Figure 10 are explained below

- Samples – students can simply view all the previous articles in the revision corpus by category e.g. conference papers, journal articles, dissertations, etc. The interface that the student is presented with is illustrated in Figure 11.

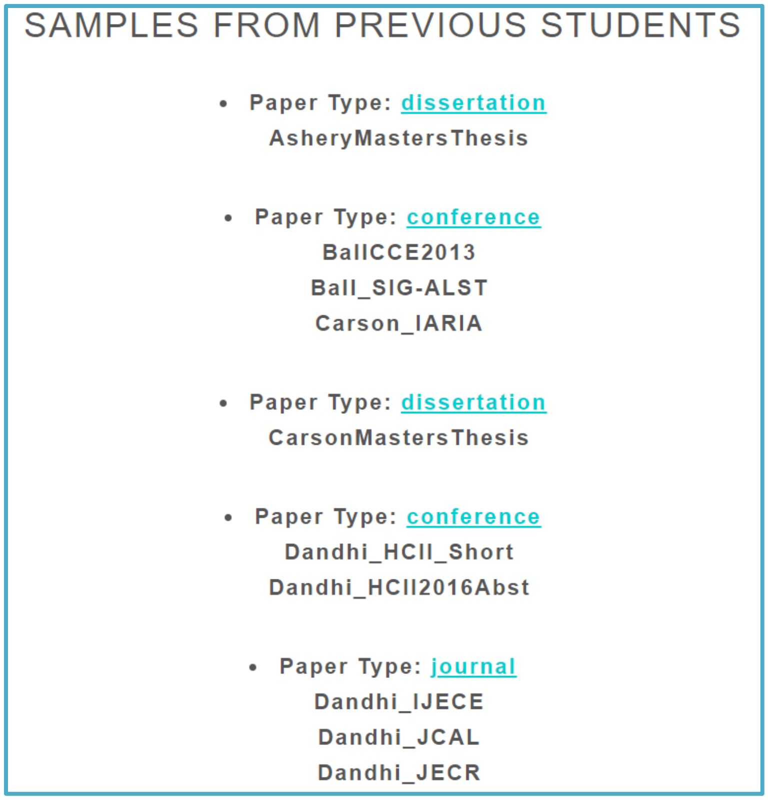


Figure 11: Samples of articles from previous students

- Upload your paper – students can upload their drafts and get hints to resolve the comments. The hints provided are discussed in detail in chapter 4.
- Self-check tips – a list of the common comments that students in the laboratory get in their drafts. Students can also see how the previous students resolved these comments. Please see Figure 12.

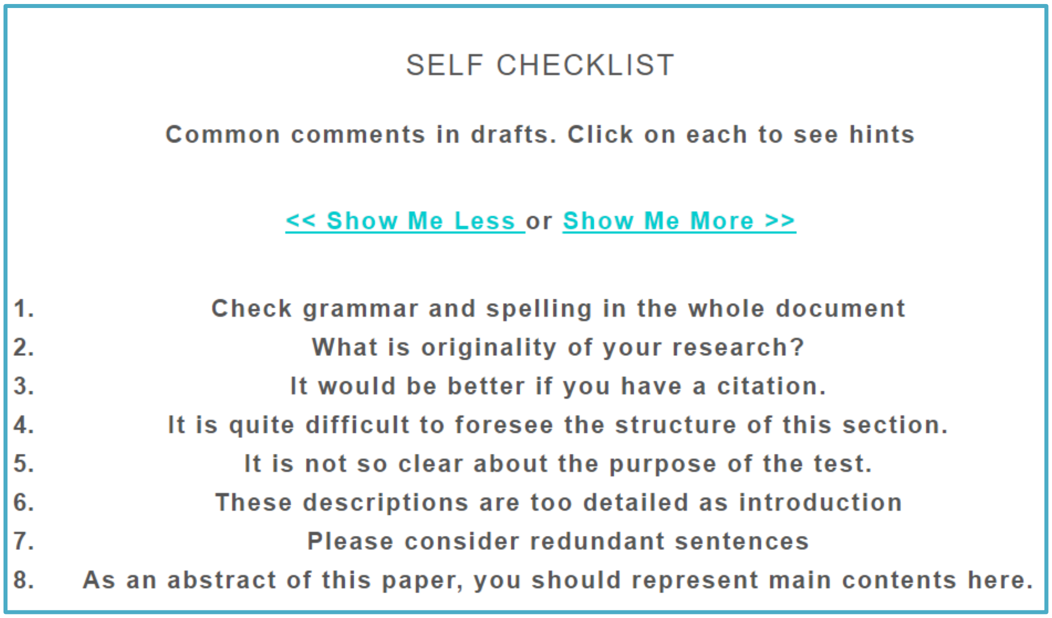


Figure 12: Self checklist

- Upcoming deadlines – this is provided by a scheduling module that checks any uploaded drafts against their due date and lets the students know of any upcoming deadlines. An example is illustrated by Figure 13.

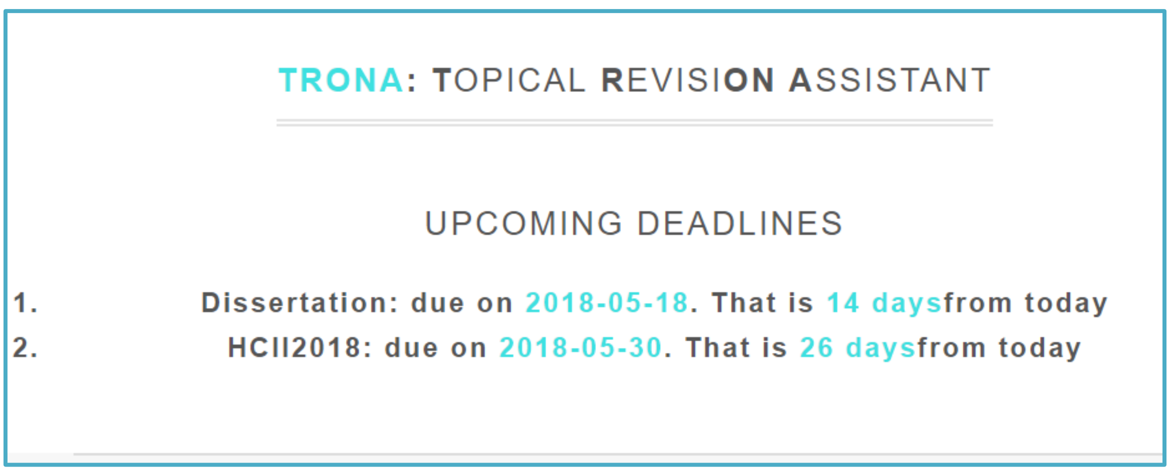
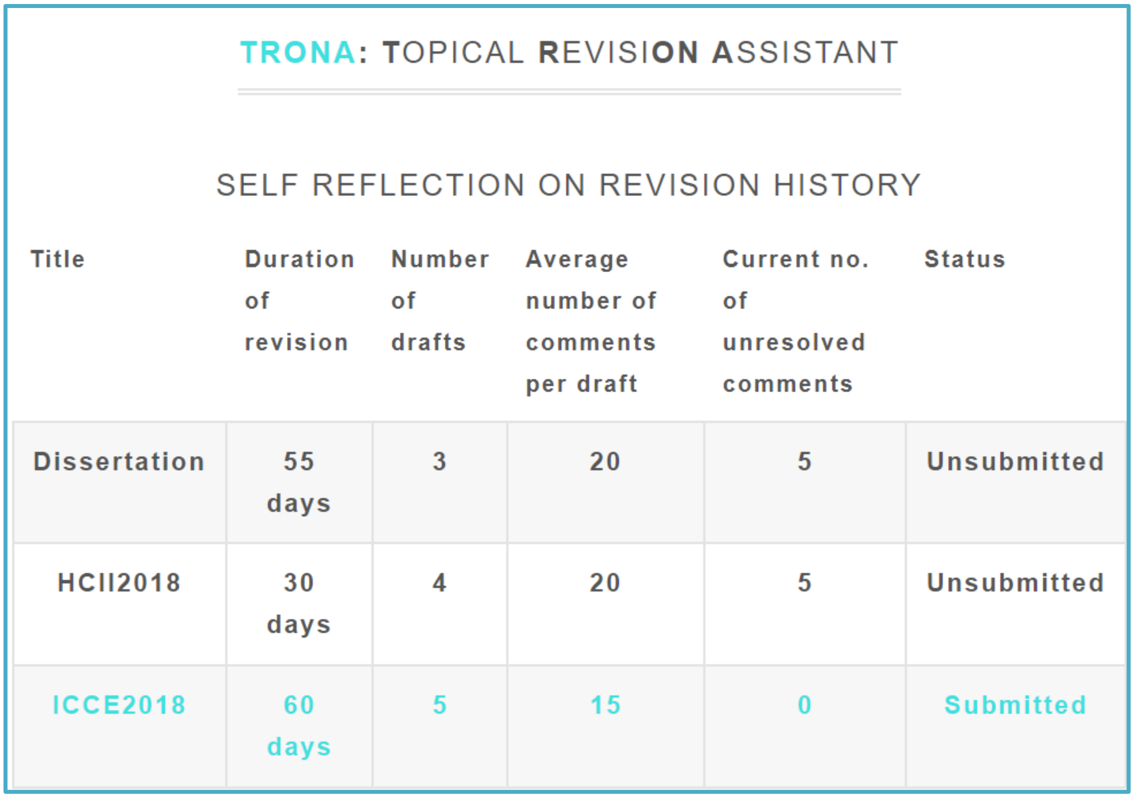


Figure 13: Upcoming deadlines for the logged in student

- Reflection module – the student can reflect on information such as the total duration taken during the revision of an article, the total number of comments, the total number of versions before arriving at the final draft, etc. For an example, please refer to Figure 14.



TRONA: TOPICAL REVISION ASSISTANT

SELF REFLECTION ON REVISION HISTORY

Title	Duration of revision	Number of drafts	Average number of comments per draft	Current no. of unresolved comments	Status
Dissertation	55 days	3	20	5	Unsubmitted
HCII2018	30 days	4	20	5	Unsubmitted
ICCE2018	60 days	5	15	0	Submitted

Figure 14: Reflection module

- Manage profile – the student can update or delete their profile, change password, profile photo, etc. Please see an example in Figure 15.

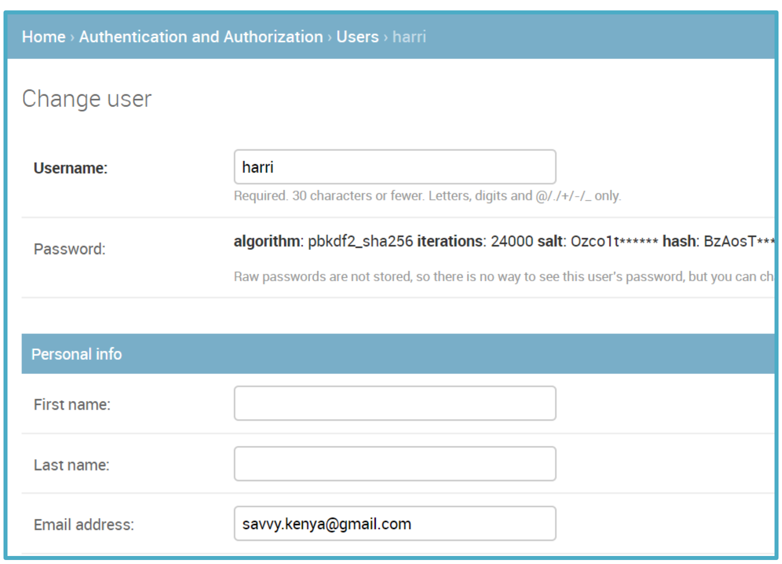


Figure 15: Profile manager

3.5 Chapter Conclusion

In this chapter, the design and development of a tool to support the revision process of students was discussed as well as an example of such an implementation i.e. TRONA. The chapter began by modeling the writing and reviewing process to identify the requirements and the challenges of using existing tools. One of the requirements is the need for a revision corpus, and this was discussed in section 3.2 and the process of building the actual corpus was demonstrated. The design and architecture of TRONA was then implemented with the modules that would fulfil the requirements identified. The next chapter will discuss in detail the provision of an adaptive interface depending on the revision skill level of the student.

4. Adaptation to Revision Skill Level

4.1 Introduction

In the previous chapter, this dissertation discussed a way to construct a revision corpus to support the revision process of students in one laboratory. The articles to be included in such a corpus are the raw drafts and the feedback from the supervisors or reviewers in the form of comments. Students can then learn about the revision process from this corpus by observing how previous students revised their articles. However, it is important to recognize that the needs of novice students are different from those of more experienced students.

For the improvement of revision skill, this research applied the cognitive apprenticeship theory which proposes several methods to teach cognitive skills to novices, such as modeling, coaching and scaffolding/fading. In modeling, the novice student is presented with a conceptual model of the processes required to accomplish a task, and the revision of an article is an example of such a task. In coaching, a student is provided with hints to help them accomplish a task. In scaffolding/fading, the student is provided with less support as their skill level increases.

In the case of using TRONA, the student uploads his/her own article and is provided with hints for improving that article. In fading, the hints provided become fewer and are presented in less detail as the student's revision skill improves. In this chapter, the adaptive design for TRONA's interface is discussed. It shows the revision process in academic writing from the initial drafts to the final drafts so that the students can learn revision skill from practical observation. In addition, the design presented is adaptive to the cognitive needs of the students. In order to provide adaptation, there is need to estimate the revision skill level of the student. Therefore, the application of the Item Response Theory (IRT) to estimate the revision skill level is also discussed.

4.2 The Cognitive Apprenticeship Theory

As was discussed in chapter two, in the revision process, there are important differences between novice writers and expert writers. Novices tend to focus on local and sentence-level aspects of revision such as grammar, spelling, and punctuation. Experienced writers, on the other hand, revise more on a global scale - their primary goal is to shape the argument (Sommers, 1980). Another related difference it that novices fail to detect problems in the text that need revision while experts easily detect both local and global problems in the text (Hayes, Linda, Schriver, Stratman, & Carey, 1987).

Therefore, the needs of both novices and experts should be considered when designing a revision support system. Furthermore, ways to help novice students improve their revision skill should also be considered. Since revision is an implicit cognitive process, there is need to make it explicit so that novices can learn from it. The Cognitive Apprenticeship Theory (CAT) (Collins, Brown, & Newman, 1989) in education theory proposes a way to make "thinking visible". Novices can be taught cognitive skills through the methods of modeling, coaching, scaffolding, articulation, reflection and exploration. An important aspect of the CAT is fading, were the support given to the novice students fades as their skill level rises.

The following items are the core teaching methods in CAT:

- **Modeling** - In modeling, an expert performs a task so that students can observe his actions and build a conceptual model of the processes required for task accomplishment.
- **Coaching** - students are engaged in problem-solving activities that require them to appropriately apply and actively integrate subskills and conceptual knowledge. The expert coaches students by providing hints, feedback, and reminders to assist students to perform closer to their level of accomplishment. The content of coaching interaction is related to specific problems that students face while carrying out a task.
- **Scaffolding** - Scaffolding is coupled with fading, the gradual removal of the expert's support as students learn to manage more of the task on their own.

- **Articulation -** an expert encourages students to explicate their knowledge, reasoning, and problem solving strategies. Such activities provide the impetus for students to engage in the refinement and reorganization of knowledge.
- **Reflection** - the expert provokes students to compare their problem solving processes with his own, with that of other students, and with an internal cognitive model of the relevant expertise.
- **Exploration** - the expert pushes students to be independent learners. At the same time, they are encouraged to identify personal interests and pursue personal goals.

The CAT has been implemented in various online tools and interfaces to adapt to the cognitive skill level of students. CAT methods were implemented in the domain of learning Smalltalk, a programming language (Chee, 1995). Another researcher (Kashihara, Shinya, Sawazaki, & Taira, 2008) applied CAT methods in web-based navigational learning in which learners can adjust the scaffold level in accordance with their metacognitive skill. The general conclusion is that if such systems are well designed and used judiciously, they can make a positive contribution towards achieving learning goals.

In the case of revision skill, it is important to support students of various levels.

- Novice students would benefit most from a revision process **model** - a conceptual model of the processes required to accomplish the revision of an article.
- Intermediate students require **coaching and scaffolding**.
- The expert (in this case the system) coaches the students by providing hints, feedback, and reminders to assist students to perform the revision process. As the students improve their revision skill, they will benefit from the **fading** aspect of scaffolding. Fading is the gradual removal of the expert’s support as students learn to manage more of the task on their own.

4.3 Design of the Adaptive System

The adaptive design overview of TRONA is illustrated in Figure 16. The corpus contains not only the final copies of the articles (42**)** but also the initial drafts (287 in total) and the corresponding feedback in the form of comments (7,809) by the supervisor. The comments by the supervisor prompt and suggest the revisions to be undertaken by the students. Some of the comments such as those on the grammar or formatting of the document may distract students, especially novices, from focusing on the overall revision of the content (Sommers, 1980). Aside from grammatical comments, the drafts also contain content-related comments. These are comments that require time and effort to revise as they involve clarifying, explaining or refining the idea or topic that the article is based on. In chapter 5, the application of a machine learning approach to automatically classify the comments in the revision corpus as *content-related* or not, is discussed in detail. The content-related comments and the corpus of articles will be used as hints and feedback to help current students during their own revision process.

To initiate the process of receiving adaptive hints, a student uploads their draft with comments from the supervisor into the system, as illustrated in Figure 16. The comments are then automatically extracted and classified as content-related or not by a component of the adaptation engine. After that, the adaptation engine estimates the student's revision skill level by evaluating the number of content-related comments in the draft via the Item Response Theory Models that are stored in a database as shown in Figure 16 (which are made of a selected number of content-related comments). Estimation of the revision skill level using the Item Response Theory is discussed in further detail in section 4.5 and subsequent sections. The newly uploaded draft and its comments are then added to the database of the revision corpus of articles and comments. The student model in the database is updated with the current estimated user level. Depending on the level of the student, appropriate hints are provided to help the student resolve the comments and thus improve their draft.

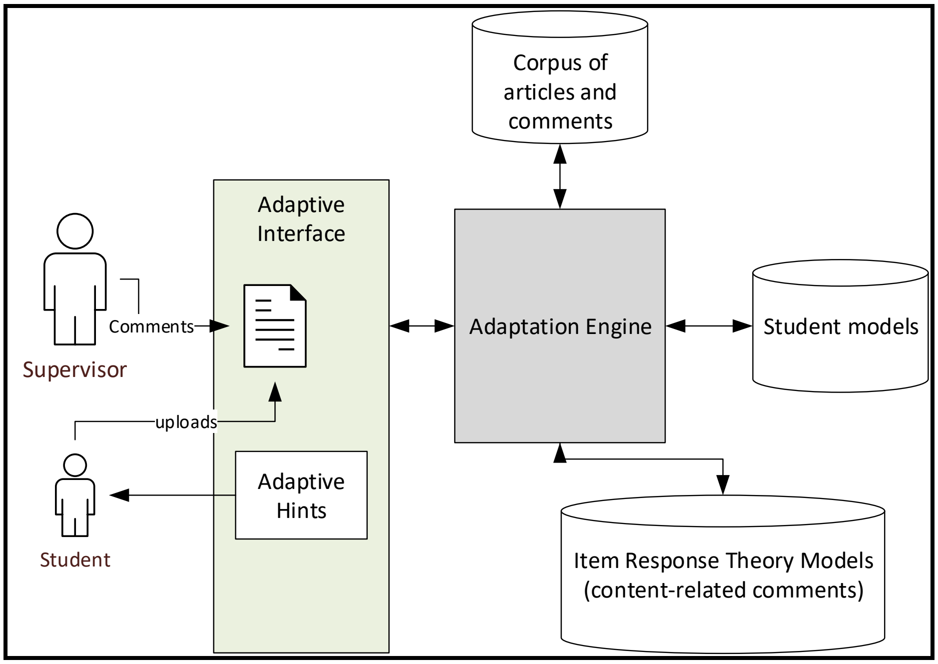


Figure 16: Design and architecture of the adaptive system in TRONA

To initiate the process of receiving adaptive hints in TRONA, the student uploads their draft into the system (see Figure 17). The comments are then automatically extracted and classified as content-related or not.

After the student uploads their draft, the content-related comments are highlighted and the students is able to see hints to resolve each content-related comment, as illustrated in Figure 18.

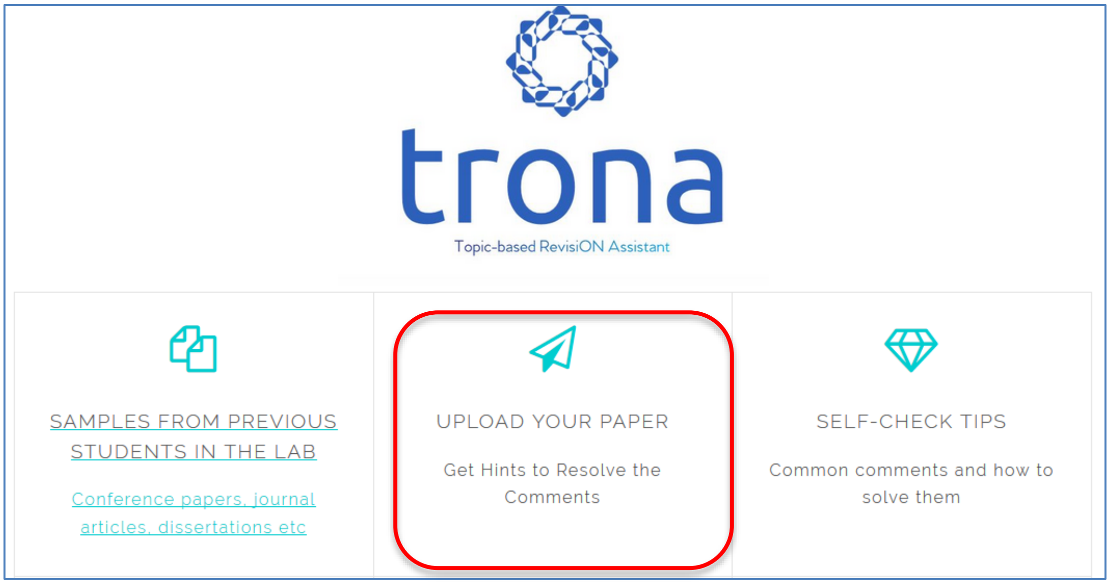


Figure 17: Students can upload their drafts containing comments into the system

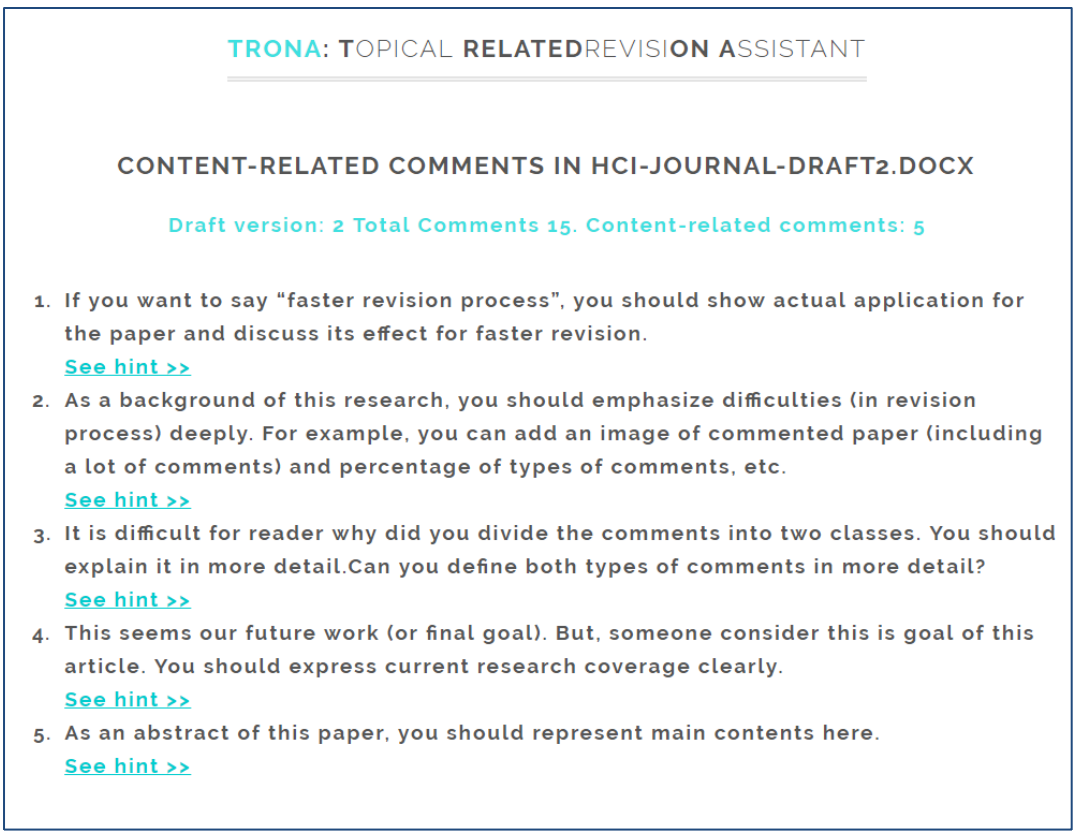


Figure 18: After the student uploads their draft, the content-related comments are selected and the student is able to see hints to resolve the comments.

4.4 Using Coaching, Modeling and Fading to Provide Adaptation

There are several approaches to realize adaptation. Adaptation can be interface-based, learning-flow based or content-based (Burgos, Tattersall, & Koper, 2007). The design discussed here applies content-based adaptation, where the content presented to users is different depending on their cognitive skill level. As illustrated in Figure 19, skill level estimation is carried out by the adaptation engine. The student is then classified as a novice, intermediate or advanced. For novice students, the interface provides modeling, for intermediate, coaching and for the advanced, fading as illustrated in Figure 19.

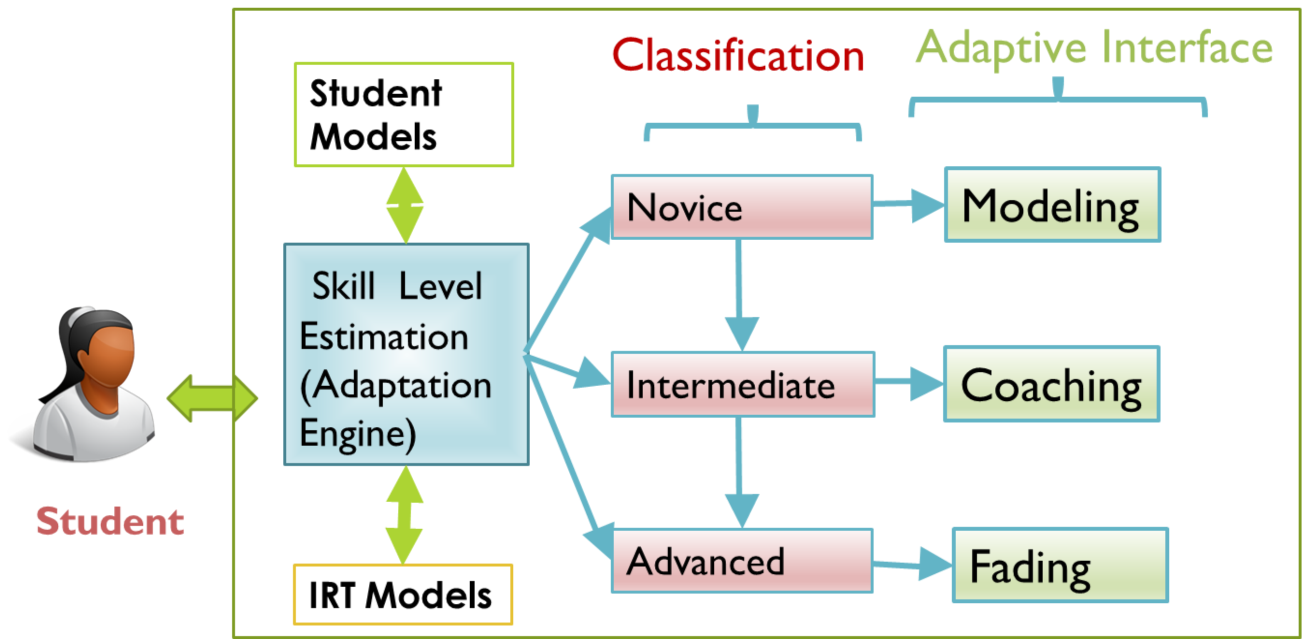


Figure 19: Providing adaptation through modeling, coaching and fading

Mapping of the CAT methods to the actual content presented by the interface is implemented as presented in Table 4.

Table 4: Mapping CAT Methods to the content of the interface

CAT Method	Content in the Interface Implementation
Modeling for novices	• Revision process overview • Checklist of common comments for novices
Coaching for intermediate	• Hints: similar comments and the corresponding comment ranges are shown from first draft to the final draft. • Checklist of common comments for intermediate students
Fading as intermediate skill level rises towards advanced	• Hints: Examples of similar comments but without the corresponding comment ranges • Checklist of common comments for advanced students

First time users of the system are assumed to be novices until they upload their article. For novice students, **a conceptual model** of the processes required to accomplish the revision of an article is presented. In terms of the actual content, the student chooses the type of article they need to work on, such as a conference paper, journal article, research proposal, etc. If the student chooses a conference paper, then previous conference papers and an overview of those papers' revision processes are presented to the student. Information such as the number of drafts, number of comments and duration of revision is displayed (please refer to Figure 20). The student can then can click on a specific conference paper to see it in detail, and can read each draft version from the first one to the last.

The system also presents a checklist of the most common comments found in novice drafts so that the novice students can keep these in mind when preparing their own articles (please refer to Figure 21).

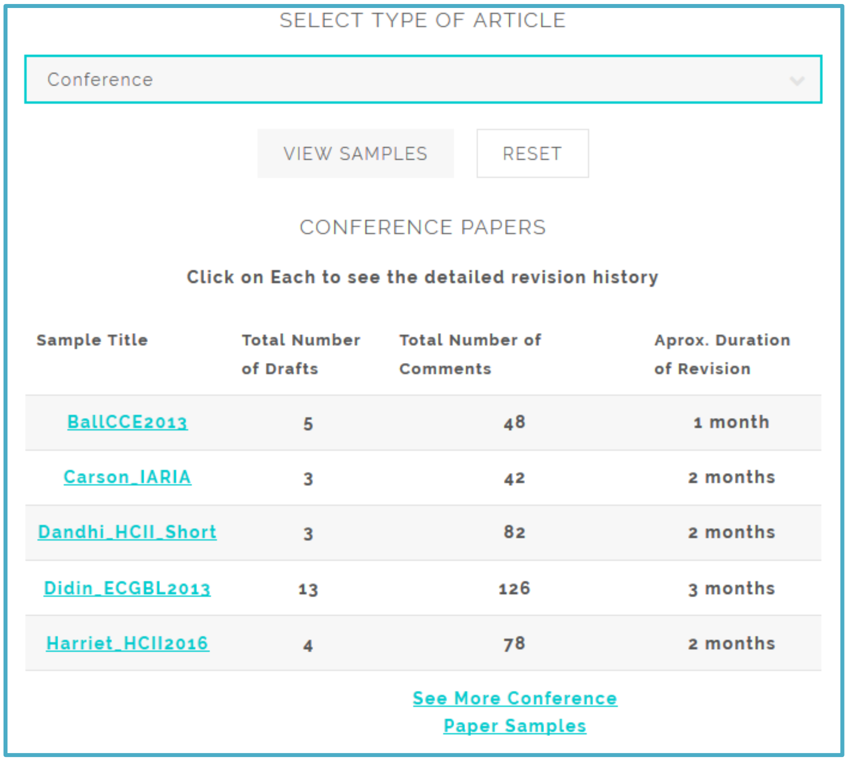


Figure 20: Providing a conceptual model of the revision process for the novice student

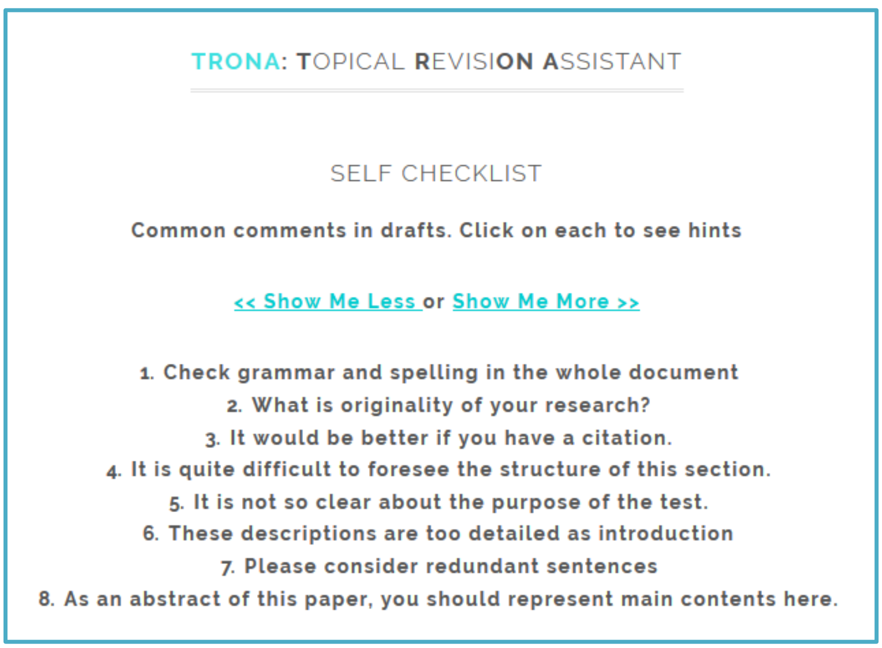


Figure 21: Showing a checklist of the most common comments for novices

Once the student has written their own article, they upload it to the system where the supervisor gives feedback in the form of comments. Based on the number and type of comments, the student is evaluated as either remaining a **novice**, or moved to **intermediate** or **advanced** categories.

The interface then displays hints on how to solve them (**coaching**) by searching the corpus for similar comments in drafts by the previous students. For each comment, the corresponding texts (comment range) in the drafts are displayed that shows the changes made to the comment range from the initial draft to the final article. For the novice student, they need more specific examples to follow to resolve their comments. This includes several examples of closely matching comments with links to the full drafts so that novices can examine them in detail.

However, for the intermediate students, the hints provided to help the student resolve a particular comment differ in the level of detail depending on the revision skill of the student. If a student has lower skill, then they are given more hints; the level of detail increases, e.g. they are shown more examples. The student can choose to see more or less detail, and this **interaction data** is saved and used to update the student model. Figure 22 is an example of a typical screen presented to an intermediate student to help resolve the comment, "***What's the originality of your research***?"

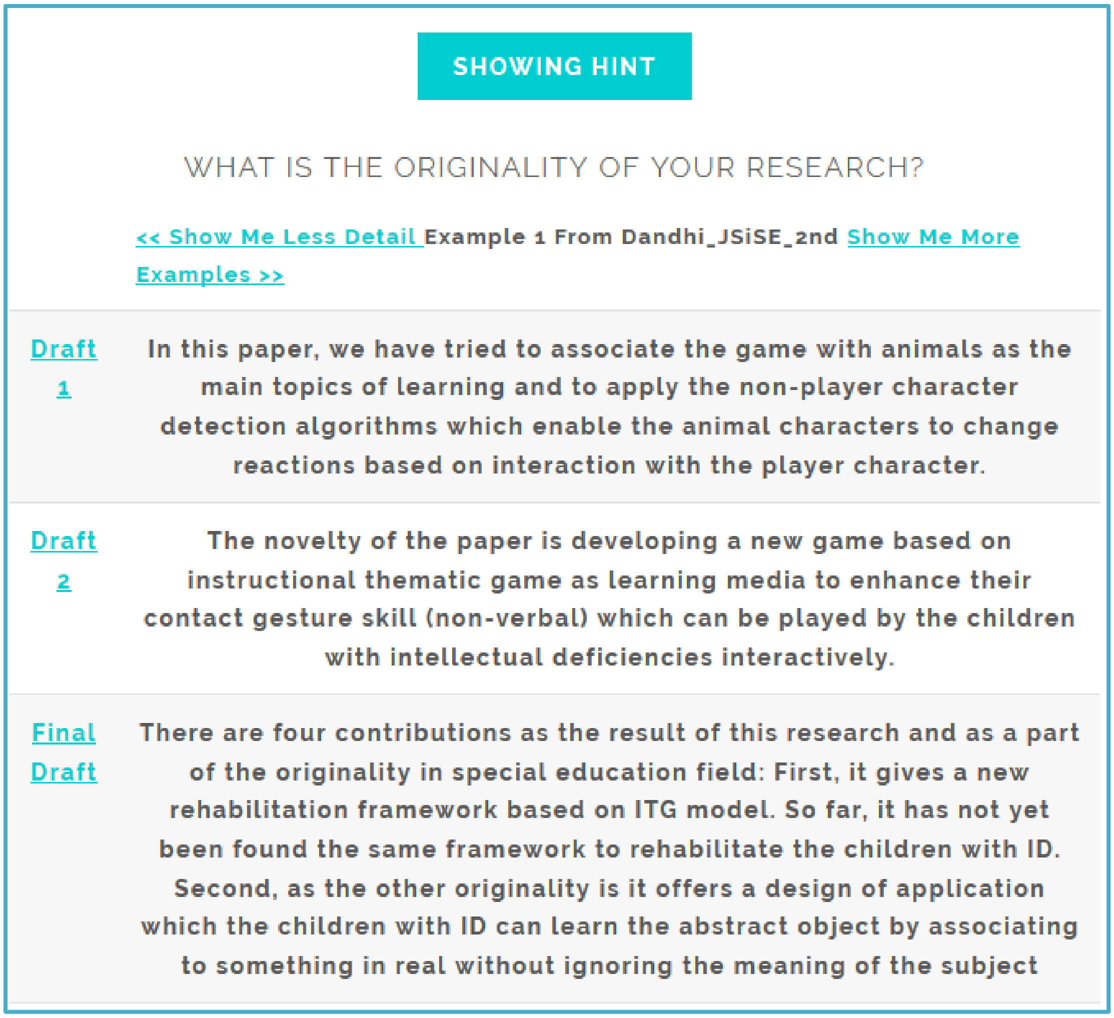


Figure 22: Intermediate Screen Example showing hints to resolve "What's the originality of your research."

The more advanced students with the highest skills are presented with the least details. As the student's revision skill level increases, the amount of detail presented to the student decreases so as to avoid over-reliance on the tool (**fading**). In the case of the same comment "***What's the originality of your research***?", the advanced student is presented with just examples of similar comments but without the corresponding comment ranges as shown in Figure 23.

Students can choose to request more or less hints at any point. They can also choose to see "models" of previous examples, as well as self-check tips of the most common comments. This information is used update the user model in order to provide a better estimation of the student's skill level. As the student's skill level changes, the type of hints they get also changes.

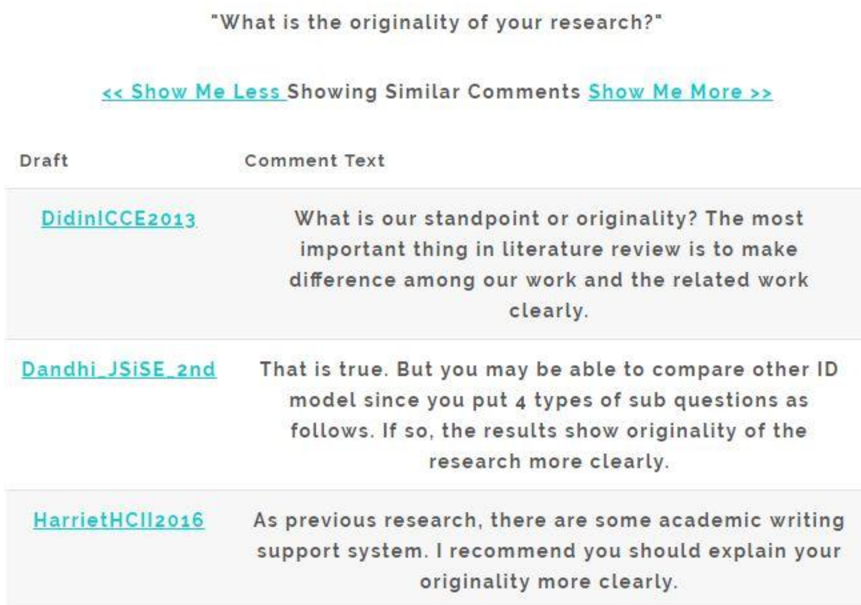


Figure 23: Advanced Screen showing similar comments to "What's the originality of your research."

4.5 Application of the Item Response Theory to Estimate Revision Skill

Estimating revision skill level is difficult because it is a cognitive skill that involves many complex processes. In addition, it takes many years to acquire. The outcome of revision can be measured, but there are many different types of academic articles with different styles in different fields. Therefore, it is hard to come up with objective standards of measurement.

The Item Response Theory (IRT) offers a way to evaluate the cognitive ability of a student in a certain subject by taking into account the student's ability and the difficulty of the questions in the test (Baker, 2001). This research proposed using the IRT to estimate the revision skill of the student. The IRT is a psychometric instrument for measuring abilities, attitudes, or other variables. There are several advantages of using IRT, such as it allows people to be compared to one another, even though they may have completed different items, allowing for computer-adapted testing such as in the case for online tests -

(Merrouch, Hnida, Idrissi, & Bennani, 2014). It is also simple enough that it can be used by many people without formal training in psychometrics. This makes it especially suitable for use in TRONA's adaptation engine to evaluate revision skill based on the number of content-related comments (items) that a student has in their draft.

The comments to be used in the IRT Model are the most common or most critical content-related questions. A number of the most common comments were selected as the items. If a student has such a comment in their document, they are deemed to have scored "incorrectly" according to the IRT model. If such a comment is absent in their document, then they have done "well" as it indicates their higher level of revision skill. Therefore, it is deemed a "correct score."

The IRT incorporates not only the student model but also the comment model. This makes IRT suitable as the type of support given to the student depends not only on their skill level but also on the type of comment as well. To realize adaptation, there is need to estimate the skill level of the student while considering the difficulty of the comments.

Suppose we obtain comments from students' drafts that are coded 1 for a correct score (the absence of a particular comment) and 0 for an incorrect score (presence of a comment). In Table 5, adapted from (De Boeck & Wilson, 2004), the results from five students are listed. C denotes comment, S denotes student.

Table 5: An example of students' score result

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**
S1	1	1	1	0	0	0	0	1	0
S2	0	0	1	0	0	0	0	1	1
S3	0	0	0	1	0	0	1	0	0
S4	0	0	1	0	0	0	0	0	1
S5	0	1	1	0	0	0	0	1	0

The goal is to test the students' revision skill and classify the students into novice, intermediate or advanced groups. We can test the students' skill by looking at the total score, but the problem is that it depends on the comments considered. If the comments are *easy to revise*, many students will appear advanced, and if all the comments are *hard to revise*, even advanced students will be considered novices.

The relationship between the probability of correctly answering an item P(θ), and the estimated ability of the student (θ), is expressed by a function called the Item Information Function (refer to equation 1) and plotted as the item characteristic curve (ICC). The probability of a student answering an item correctly P(θ), is expressed by equation 1:

$$\mathrm{P}(\theta) = \frac{1}{1 - e^{-a(\theta - b)}} \qquad (1)$$

Where:

- *e*: 2.718
- a: the discrimination parameter, an index of how well the item differentiates low from top ability students; typically ranges from 0 to 2, where higher is better
- b: the difficulty parameter, an index of what level of examinees for which the item is appropriate; typically ranges from -3 to +3, with 0 being an average examinee level
- θ is an ability level.

By utilizing available IRT evaluation software, a model can be created to estimate the parameters *a* and *b* for each comment. The ICCs of the comments in Table 5 can be visualized as in Figure 24, which also shows the estimated difficulty of the comments. A student with a higher ability (θ) has a higher probability of answering a question correctly.

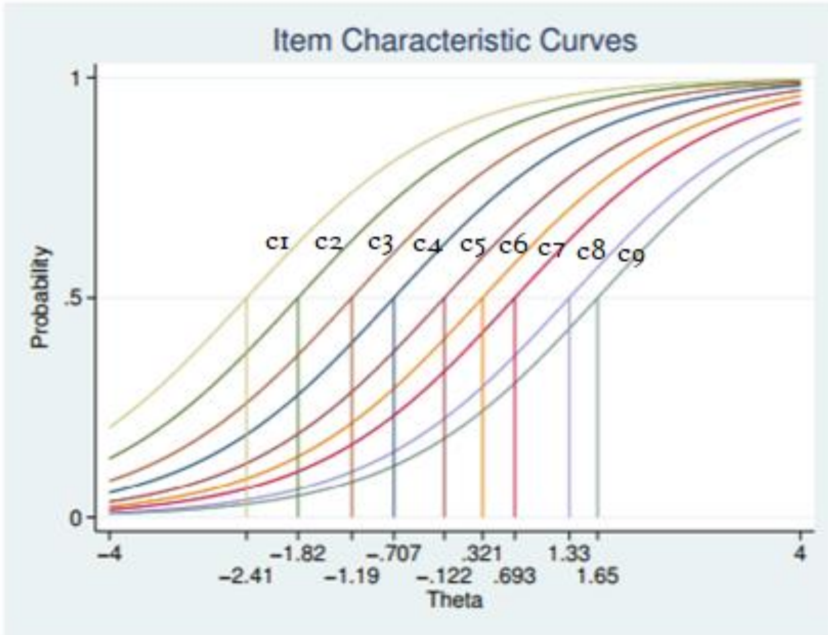


Figure 24: Item Characteristic Curves showing the difficulties of the comments (1) to (9): adapted from (De Boeck & Wilson, 2004)

The probabilities in Figure 24 represent the expected scores for each item along the ability continuum. For this particular model, the midpoint probability (0.5) for each item corresponds to the estimated difficulty parameter.

Once the ICCs of the items have been calculated, the students' ability can then be estimated. The following equation (equation 2) is used to calculate the student's ability, θ (Baker, 2001).

$$\theta_{s+1} = \theta_s + \frac{\sum_{i=1}^{N} -a_i[u_i - P(\theta_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\theta_s) * Q_i(\theta_s)} \quad (2)$$

Where:

- θs: Learner ability within iteration S, the value of θ is theoretically between $-\infty$ and $+\infty$ but in reality is limited between -3 and 3.
- i: Current asked item
- N: Number of items
- u_i: The learner response to item i
    - $u_i = 1$ for a correct answer
    - $u_i = 0$ for a wrong one
- $P_i(\theta s)$: The probability to give a correct answer to question i in iteration s
- $Q_i(\theta s)$: The probability to give an incorrect answer to question i in iteration s

Initially, the θs on the right side of the equal sign is set to some arbitrary value, such as 1. After each iteration, the estimation gets more precise until a stable competence value, and a low error value, are obtained. Using available statistical software, we can estimate the ability of a student (θ) by supplying pre-calculated parameters (*a* & *b*) and the student's comments.

4.6 Example of Estimating the Revision Skill Level of a Student

To illustrate the estimation of revision skill level of a student, this research utilized a spreadsheet obtained from an assessment website (Assessment Systems, 2018). For example, 20 content-related comments are being used in the estimation. First, the *a*, *b* and *c* parameters are estimated for each of the 20 content-related comment as shown in Table 6.

Table 6: The a, b and c parameters for the 20 content-related comments

Item	a	b	c
1	0.83	-3.73	0.27
2	0.89	-0.98	0.25
3	0.78	0.53	0.25
4	0.53	-0.80	0.26
5	1.15	0.69	0.41
6	0.42	0.08	0.24
7	0.29	0.51	0.25
8	0.26	-3.55	0.29
9	0.30	-3.73	0.44
10	0.50	-0.49	0.24
11	0.38	0.74	0.21
12	0.17	-3.73	0.55
13	0.40	1.15	0.23
14	0.25	-0.78	0.25
15	0.21	2.24	0.23
16	0.22	-2.41	0.26
17	0.95	-1.91	0.24
18	1.00	-1.07	0.24
19	0.63	0.11	0.20
20	1.02	-0.48	0.23

Having had the item parameters estimated, for each student it is possible to calculate the revision skill, which is given by theta, again using the same spreadsheet provided by (Assessment Systems, 2018).

If a student gets all responses "correct" – in this case, there are no content-related comments in their draft, then they get a score of 3, which is the maximum as illustrated in Figure 25.

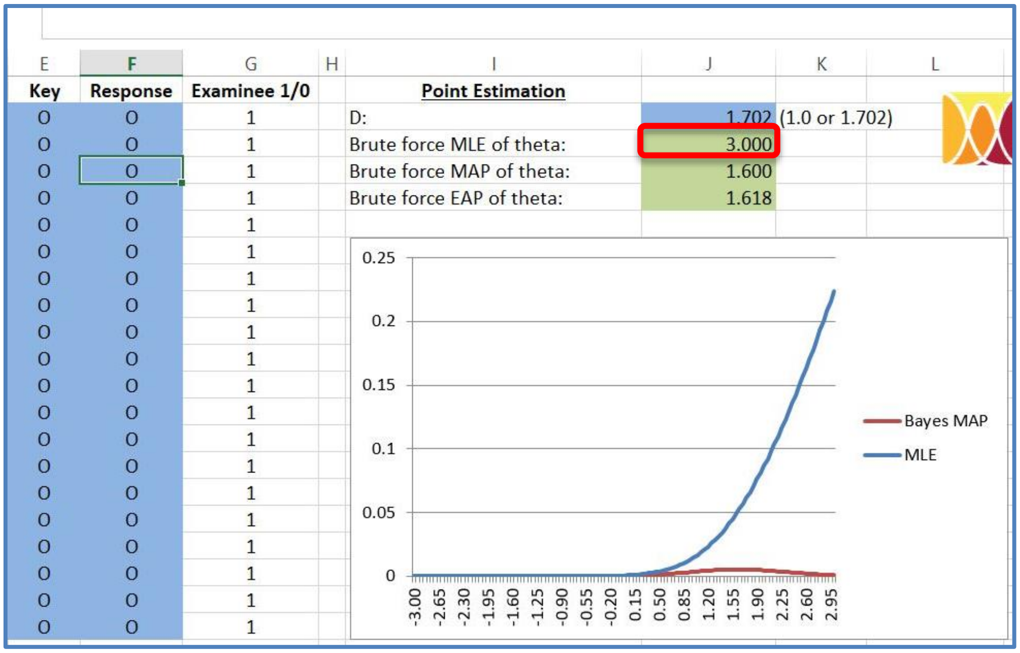


Figure 25: IRT Scoring with a maximum theta of 3

If a student’s draft has all the comments considered in the estimation, then they get the minimum value of theta, which is -3 as shown in Figure 26.

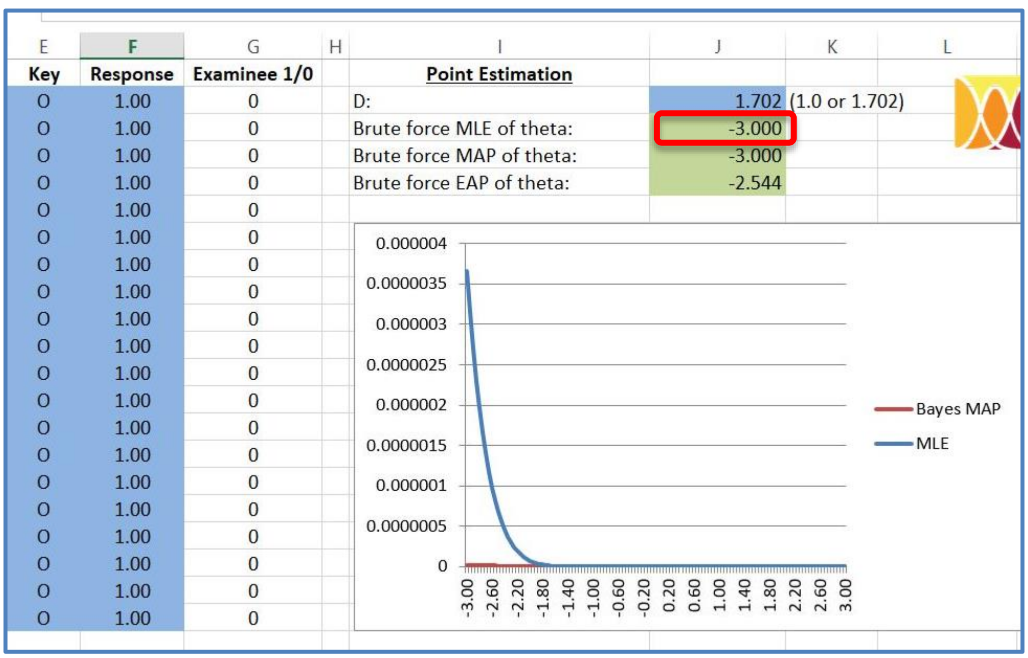


Figure 26: IRT Scoring with a maximum theta of 3

For example, on a scale of -3 to 3, a student who gets a score of x where

- -3 <=x < -1, may be said to be a novice,
- -1<= x <1, may be said to be an intermediate student, while
- 1 <= x < =3, may be said to be an advanced student

Let's take the case of a student who has a draft with 4 of the content-related comments considered in the evaluation. In this case, the evaluation on the 4 comments is "0" while they were scored "1" for the comments that were absent from their draft as shown in Figure 27. The student's revision skill is estimated to be 1.4 and based on the scale above, the student is estimated to have advanced revision skill.

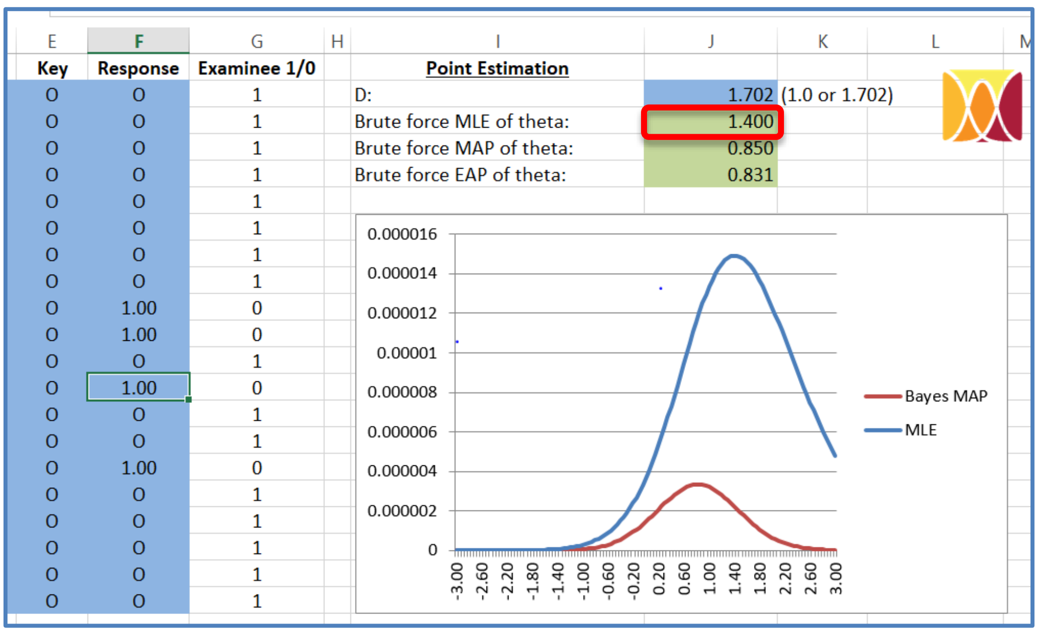


Figure 27: estimating the revision skill of the student in the example

4.7 Evaluation of the Item Response Theory in Estimating Student Skill

The goodness of fit of a statistical model such as an IRT model, describes how well the model matches a set of observations. Assessing the absolute fit of a model (i.e. the discrepancy between a model and the data) is critical in applications, as inferences drawn upon poorly fitting models may be misleading.

Assessing the absolute fit of a statistical model involves determining whether the model could have generated the observed data. In IRT applications, however, degrees of freedom are most often so large that no model can be expected to fit the data perfectly (Maydeu-Olivares, 2015). In models with so many degrees of freedom, it is instead recommended instead to assess whether the model approximately fits the data.

In this evaluation, the main concern was whether the IRT model constructed would match expected observations carried out by other means of estimating the revision skill of the students. Evaluating the goodness of fit is useful when refining the comments to be used in the IRT estimation model. Therefore in this evaluation, the estimated revision skill level of the student was compared with the result with a manual evaluation by the students' supervisor. At the same time, the result of 'comment difficulty' by the IRT evaluation was compared with the values of 'comment importance' provided by the supervisor.

4.7.1 IRT Estimation Vs Independent Supervisor Evaluation of Student Skill Level and Comment Difficulty

20 of the most common content-related comments were selected to be used in the IRT estimation model. These comments are shown in Appendix 3. Following that, 7 students were evaluated based on their final drafts of conference or journal articles. The presence of a comment in the draft is marked by Y and absence is marked by N, as shown in Table 7.

Table 7: Student drafts and the presence or absence of comments in the drafts

Student/comment	c1	c2	c3	c4	c5	c6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C8	C19	c20
s1	N	Y	N	N	N	N	Y	N	N	N	N	N	N	N	Y	N	N	N	Y	N
s2	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N
s3	N	Y	N	Y	N	N	N	N	Y	Y	Y	N	N	N	N	N	Y	N	N	N
s4	N	N	Y	N	Y	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N
s5	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
s6	N	Y	Y	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	Y	N	N
s7	N	Y	Y	N	N	N	N	N	N	N	N	N	Y	Y	Y	Y	N	N	N	N

This information in Table 7 was used to construct a Rasch IRT model using Ministep [Linacre, 2018], a free IRT analysis software. The students' revision skill level was estimated by IRT and also evaluated independently by the supervisor, as shown in Table 8.

Table 8: IRT estimation vs independent supervisor's evaluation of students' revision skill level

Student	Ability Estimation by IRT	Supervisor Evaluation
S1	-1.21	-2
S2	2.3	0
S3	0.42	1
S4	-1.21	-1
S5	3.2	2
S6	0.79	0
S7	-0.42	-2

Using SPSS© software, a Pearson correlation analysis was done to determine the relationship between the ability measure by IRT and the supervisor's evaluation of the students' ability. The results are shown in Figure 28.

Correlations

		IRT	Sup
IRT	Pearson Correlation	1	.813*
	Sig. (2-tailed)		.026
	N	7	7
Sup	Pearson Correlation	.813*	1
	Sig. (2-tailed)	.026	
	N	7	7

*. Correlation is significant at the 0.05 level (2-tailed).

IRT- value of IRT estimation

Sup: value of supervisor evaluation

Figure 28: Pearson correlation result of students' revision skill level estimation by IRT and the independent evaluation by the supervisor.

Based on the results, the IRT estimation and supervisor evaluation values appear to have a strong, positive correlation at $p = 0.026$, which is statistically significant. However, due to the small sample size (n=7), the estimate of the correlation may not be stable. A much larger sample size is needed.

In the same way, the comment difficulty was estimated by IRT and also evaluated independently by the students' supervisor as shown in Table 9. Using SPSS© software, a Pearson correlation analysis was run to determine the relationship between the difficulty measure by IRT and the supervisor's evaluation of the comment's "importance". The results are shown in Figure 29.

Table 9: IRT estimation and supervisor evaluation of comment difficulty

Comment	Difficulty Measure by IRT	Importance evaluation by supervisor
C1	2.02	3
C2	-1.65	-1
C3	-1.65	-2
C4	0.7	2
C5	0.7	0
C6	2.02	3
C7	0.7	2
C8	2.02	2
C9	0.7	1
C10	-0.97	-3
C11	-0.26	1
C12	2.02	-1
C13	0.7	-1
C14	0.7	1
C15	-2.42	-2
C16	0.7	-1
C17	0.7	-1
C18	0.7	0
C19	0.7	1
C20	2.02	0

Correlations

		IRT	Sup
IRT	Pearson Correlation	1	.646**
	Sig. (2-tailed)		.002
	N	20	20
Sup	Pearson Correlation	.646**	1
	Sig. (2-tailed)	.002	
	N	20	20

**. Correlation is significant at the 0.01 level (2-tailed).

IRT- value of IRT estimation

Sup: value of supervisor evaluation

Figure 29: Pearson correlation result comment difficulty estimation by IRT and comment “importance” evaluation by the supervisor.

Based on the results, the following can be stated:

- The IRT estimation and supervisor evaluation values have a strong and positive correlation.
- At $p = 0.002$, the correlation is statistically significant.

4.7.2 Discussion

To assess the suitability of IRT in estimation students' revision skill, the estimated values by IRT were compared with an independent manual evaluation by the students' supervisor. The comment difficulty estimation by IRT was also compared with the values of "comment importance" by the supervisor. There was a strong and positive, statistically significant correlation between IRT estimation and manual supervisor evaluation. However, for the students' ability estimation, the sample size was small size (n=7). The estimate of the correlation may not be stable. A much larger sample size is needed. The significant correlation implies that IRT may indeed be a suitable way to estimate revision skill.

4.8 Chapter Conclusion

In this chapter, the design for an adaptive interface that implements the cognitive apprenticeship theory techniques of modeling, coaching and fading to improve the revision skill of students writing academic articles was discussed. The interface provides novices with a model of the revision process and a self-check list of the most common comments in the drafts. For intermediate students, it provides hints in form of revision histories of similar comments by previous students in the lab (coaching). For advanced students, it also gives hints but they are less detailed (fading) than those presented to intermediate students. The application of the Item Response Theory to estimate the revision skill level of students was also evaluated and discussed. At the core of the IRT are the content-related comments to be used as item models, therefore, the classification of comments in research article drafts as content-related or not is discussed in detail in chapter 5.

5. Using Machine Learning to Classify Comments in Research Article Drafts as Content-Related or Not

5.1 The Need for Classification

In this research, one of the most important aspects is the classification of comments in the research article drafts as *content-related* or not. There are two reasons why classification of comments is important.

i.) Reviewer comments by the supervisor can trigger local or global revision. To enable students to increase their awareness of and focus on global revision, there is need to classify the comments and highlight those that are content-related i.e. comments that trigger the global revision of comments. If the comments are automatically classified as *content-related* or not, the students are forced to reflect and revise on a global scale. The concept of automatic filtering comments has been applied by other researchers for websites' comments. Automatic classification can give the reader an overview of comments, because they may be too many for the reader to read through all of them (Mukherjee & Bing, 2012). Comments can be classified as negative or positive (Kaszuba, Albert, & Adam, 2009) or in our case they can be classified according to the type of revision they will trigger. Thus, the student can gain perspective into the overall global revision needed to improve the quality of the draft.

ii.) To realize adaptation there is need to first evaluate which comments are important (content-related). The adaptive hints given to students are based on their skill level and the content-related comments in their drafts. The IRT is used to estimate revision skill and at its core are the content-related comments to be used as item models, therefore, this research focused on evaluating the comments classification. In this research, there is an assumption that a student with low revision skill will have more content-related comments in their drafts. On the other hand, if a student has very few content-related comments in their draft, it may imply that the draft is already of good quality and therefore the skill level of the student is high. This is the second reason that it is important to obtain the content-related comments from the drafts –a number of selected content-related comments will be used in the IRT as item models.

5.2 Examples of Comments in a Research Article Draft

In the article drafts obtained when building the revision corpus, the average number of comments per draft was 33. A majority (64%) of the comments in the drafts were concerning the spelling, grammar or formatting of the document. These comments do not require much effort to revise as they are straightforward instructions on improving the relevant section. However, some of the comments did require effort into revising and improving the actual content of the text in the comment range. This may involve instructions to provide more explanation about an idea, adding information to justify a concept or other comments that require a considerable amount of revision effort and time. This former type of revision is global revision and research has shown that expert writers carry out global revision while novices focus on local, sentence level revisions (Hayes, Linda, Schriver, Stratman, & Carey, 1987).

This is illustrated with an example: During the writing of a particular draft which was prepared using Microsoft Word, various comments were received from the supervisor as shown in Figure 30. The first two comments may require considerable effort in revision. However, the rest of the comments are typos that can be quickly revised. Typos and other such comments can become a distraction if the student focuses too much on them, instead of giving priority to the comments that are concerned with global revision.

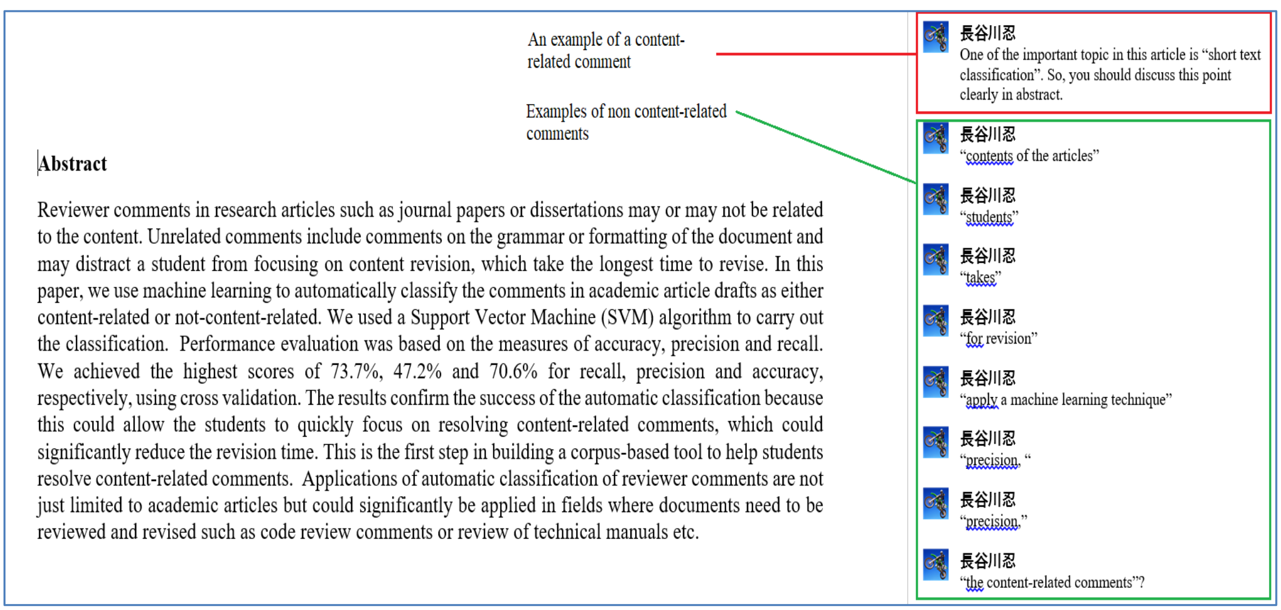


Figure 30: An example of the typical comments found in research article drafts

Comments in the drafts are generally short in length, sometimes as short as one word. Each comment is also specific to a certain section of the draft – the comment range. Depending on the type of revision they trigger, the comments were classified into two classes:

- ***Content-related* comments**: These are comments that trigger a global revision e.g. restructuring of ideas, reorganization of the document, etc. In addition, they require a lot of time and effort to revise. They are about clarifying, explaining, expounding etc. the actual idea or topic of the article. Example from Figure 30*: If you want to say "faster revision process", you should show actual application for the paper and discuss its effect for faster revision.*

- ***Non content-related* comments:** These are comments that trigger local revision of words or sentences**.** These comments are easy to revise**.** The *non content-related* comments are not directly related to the content or topic of discussion, but may be about the correction of grammatical or spelling mistakes, punctuation, or formatting instructions such as adjusting font type and size, positioning of figures and tables, page limitations, etc. Example from Figure 30*: "takes"* which is a typo and should be *"take".*

5.3 Related Literature - Short Text Classification

In this research, a machine learning technique was applied to automatically classify the comments as *content-related* or not. If the classification were to be done manually, it would take too much time which could be used to do the actual revision. The average number of comments per draft was 33 and one of the drafts actually had a maximum of 500 comments.

This section constitutes a review of literature in the area of classification of short length texts or sentence-level classification. This is a more difficult task because there is not much text information, unlike in document-level classification. An important factor to consider is feature selection, which can have a significant impact on the machine learning method's performance (Bird, Loper, & Klein, 2009). This has motivated various researchers to investigate the type of features necessary in each particular case in order to improve classification accuracy. Much of prior work has been on internet comments on social media sites, auction sites, etc. The closest research to this classification of reviewer comments in written text is code review comments carried out at Microsoft (Amiangshu, Greiler, & Bird, 2015).

From previous research (Iwai, Hijikata, Ikeda, & Nishida, 2014), (Mukherjee & Bing, 2012) , (Joty, et al., 2015), (Kaszuba, Albert, & Adam, 2009), (Yue, Zhai, & Sundaresan, 2009), it can be concluded that for short texts, other features apart from text information are needed to improve classification accuracy. This is because short texts do not possess enough textual information. It can also be concluded that in each case, these features are unique to the application area. This motivated this research to investigate the features that can give the best performance in the case of reviewer comments in academic texts. There are characteristics of the reviewer comments in academic drafts that make them different from those in social media comments, such as the pairwise relationship with the comment range. Reviewer comments are also typically short in length and they are intended for the author to improve the main document.

Another factor is the type of machine learning algorithm chosen (Iwai, Hijikata, Ikeda, & Nishida, 2014). Among the most common algorithms considered for short text classification are Naive Bayes, SVM and Decision Trees. For classifying comments from the social network Facebook, as whether weather-related or food-related, SVM performed the best (Mosab, Abdulla, Al-Ayyoub, Jararweh, & Quwaider, 2014). In preliminary evaluation, this research applied a Decision Trees (DT) algorithm as well as SVM. SVM outperformed as expected because it generally performs better for short texts (Atreya, Walters, & Shepherd, 2003). The classification task was implemented using LIBSVM (Chih-Chung & Chih-Jen, 2011) which is a fast, easy to use library implementation of SVM.

5.4 Data Preprocessing

7,786 comments and the corresponding comment ranges (text covered by the comment) were extracted from the drafts of various academic articles by current and former students in our laboratory. These articles included as research papers, dissertations, reports, etc. The data was preprocessed by removing duplicates and blanks. In supervised machine learning, the desired classes of the data is provided, in this case the classes were *content-related* or not. To train the algorithm to predict correctly, the test data is labeled with the correct classes. This labeling was done manually. There was a total of 4,551 comments, of which 1,565

(36%) were *content-related* and 2,786 (64%) were *non content-related*. The comments were then randomized so that comments from the same article draft were spread out in a random order in the master list.

Each comment was analyzed and annotated using the Stanford CoreNLP (Manning, et al., 2014). The annotation was tokenization, sentence splitting, and lemmatization. The main interest of this research was the lemmas of content words. A lemma is the dictionary form of a word. The content words included were nouns, verbs, adjectives, and adverbs. Pronouns, articles and other parts of speech were not considered relevant features for the classification algorithm. The preprocessing for a content-related and non content-related comment is shown in the following examples:

Content-related example

- **Before preprocessing**: In research paper, we have to claim the novelty of the research through comparison of related work. You don't follow the way of academic writing.
- **After preprocessing**: (*content-related*): research, paper, claim, novelty, research, comparison, relate, work, follow, way, academic, writing

Non content-related Example

- **Before preprocessing**: You should check capitalized letter in the whole document
- **After preprocessing**: (*non content-related*): check, capitalize, letter, whole, document

In the examples, the stop words such as in, we, to, etc. have been removed. Stop words are commonly occurring grammatical words that tell us nothing about a document’s content. The content words such as research, paper, check, capitalize etc. are retained. However, the tenses and plurals are stripped so that only the lemmas of the content words remain.

In summary, in the preprocessing stage, the text of each comment was reduced to the lemmas of the content words. The lemmas are considered features of the comment. After that, labels were manually assigned by the researcher. Then during the evaluation phase, the classification result from the machine was compared with the result by the human researcher. The measures of performance were derived from this comparison. Accuracy is a measure of the similarity between the classification result from the machine and the classification result by the human researcher.

5.5 Evaluation Method

When testing the reliability of a machine learning model to predict the classification of data, the model is usually given a dataset of known or labeled data on which training is run (training dataset), and a dataset of unknown data against which the model is tested (called the validation dataset or testing set). In conventional testing of prediction models for machine learning, also known as hold-out validation, it is common for the dataset to be partitioned into two sets, one set for training and the other for testing. One reason for not using conventional cross validation is that in some cases, there is not enough data available to partition it into separate training and test sets without losing significant modelling or testing capability. The downside is that the results are highly dependent on the choice for the training/test split. The instances chosen for inclusion in the test set may be too easy or too difficult to classify and this can skew the results.

In such cases, cross-validation is a more suitable way to properly estimate model prediction performance (Refaeilzadeh, Lei, & Huan, 2009). Cross-validation combines (averages) measures of fit (prediction error) to derive a more accurate estimate of model prediction performance.

In cross validation (k-fold cross validation), the data is first partitioned into k equal (or nearly equal) sized segments or folds. For example, the data may be partitioned into 10 equal segments (k=10). One fold (k) is held for validation while the remaining 9 folds (k-1 folds) are used for training. In this case, 90% of the data is used for training, and 10% for testing. Subsequently, k iterations of training and validation are performed such that within each iteration, a different fold of the data is held-out for validation while the remaining $k - 1$ folds are used for training.

The confusion matrix (Table 10) was used to calculate the performance measures during each iteration of cross validation:

Table 10: Confusion matrix

(Total N =Actual No + Actual Yes, as labeled by human researcher)	**Predicted *Non content-related***	**Predicted *Content related***
Actual No	True Negatives (TN)	False Positives (FP)
Actual Yes	False Negatives (FN)	True Positives (TP)

Performance measures:

- **Accuracy**: the number of correctly classified comments over the total number of comments. In other words, the percentage of correct results that was achieved. It is calculated by (TP+TN)/N
- **Recall**: the number of *non content-related* comments that was predicted as such from the total of all actual *non content-related* comments. In this case, it is calculated by TN/Actual No
- **Precision**: the number of actual *non content-related* comments from those that were predicted as being *non content-related*. In this case, it is calculated by TN/(TN+FN)

5.6 Baseline Model

The baseline model was selected by using the simplest features when it comes to text classification. In this case, the feature used was only the lemmas of the content words contained in the comment text. The content words included were nouns, verbs, adjectives, and adverbs. Pronouns, articles and other parts of speech are not very informative in classification tasks.

Using the lemmas of the content words (of the comment text and comment range), the baseline model was created to predict whether a comment was content-related or not. The number of features (lemmas of content words) was 902. By applying supervised learning using LIBSVM, an average accuracy rate of 0.562 was obtained. For the content-related comments, the recall and precision rates were 0.56 and 0.64 respectively.

Utilizing the text data meant that there was a high number of features, 902 in this case. This led to the curse of dimensionality. However, it provided a baseline to use when considering what features other than actual text words could be used to improve the performance of the classification algorithm.

5.7 Feature Selection

One of the main tasks in this classification was a selection of features that would improve the baseline result. Feature selection is important for effective data mining, especially with high-dimensional data. As supervised learning was applied, it was possible to infer some characteristics of *content-related* comments vs. the *non content-related* comments during the manual labeling process. A combination of several different features was tried, including the *content-related* keywords in the comments.

In Table 11 is the summary of some of the feature combinations that were tested, at each time adding more features to improve the performance metrics. A Decision Trees algorithm was used for testing the features. This is because with Decision Trees, it is possible to carry out analysis to examine how the branching (categorization) occurred.

While the recall was quite high, the precision rate wasn't as impressive. However, precision is also an important measure to keep ensure the content-related comments were not mislabeled. A higher precision rate is therefore desired.

Table 11: The results of feature selection using Decision Trees

Feature List	Accuracy %	Recall %	Precision %
Comment length (actual no. of words)	64	100* all were predicted as *non content-related*	64
Comment length (symbolic – short, medium, long)	70.4	98.7	68.8
Comment length + comment-range length	71.6	98.7	69.6
Comment length + comment-range length + edit distance*	72.1	98.7	70.0
Comment length + comment-range length + edit distance + grammatical keywords	76.2	98.7	73.3
Comment length + comment-range length + edit distance + grammatical keywords + content keywords	81.4	85.8	85.3

* A measure of similarity between comment text and comment-range text

Judging from the results in Table 11, the following features were selected for the classification process - length of comment, length of comment-range, similarity between comment text and comment-range text, grammatical keywords and content keywords. Each feature is explained in detail below:

a) **Comment length**: The *content-related* comments tended to be longer than the *non content-related* comments. Figure 31 and Figure 32 show the comment length distribution for the *content-related* comments and the *non content-related* comments respectively. The comment length was normalized between 0 and 1 using the formula: $zi=(xi-min)/max-min$, where zi is the normalized value, xi is the original value, min is the minimum value of length in the sample and max is the maximum value of length in the sample.

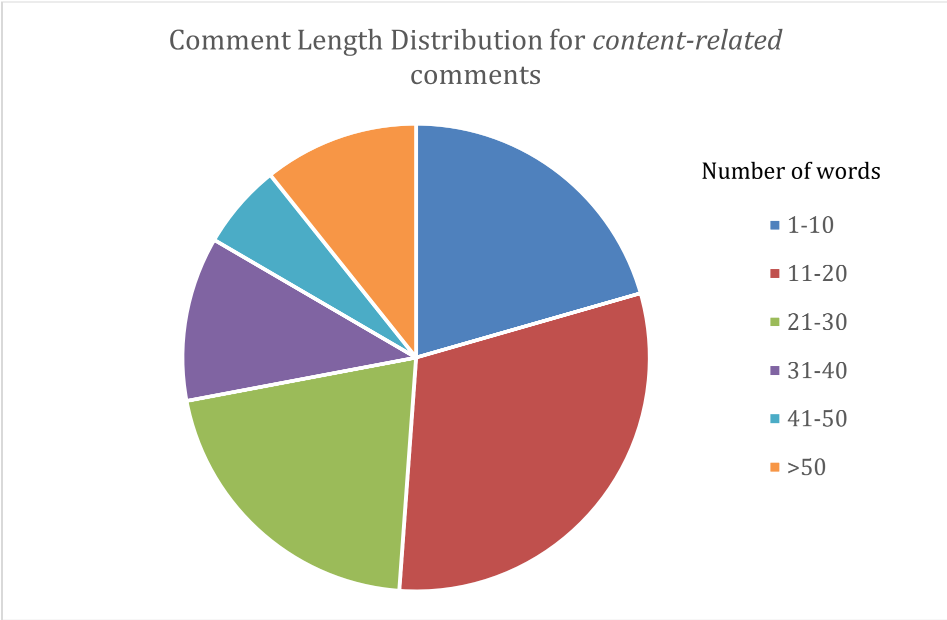


Figure 31: Chart showing the comment length distribution for the *content-related* comments

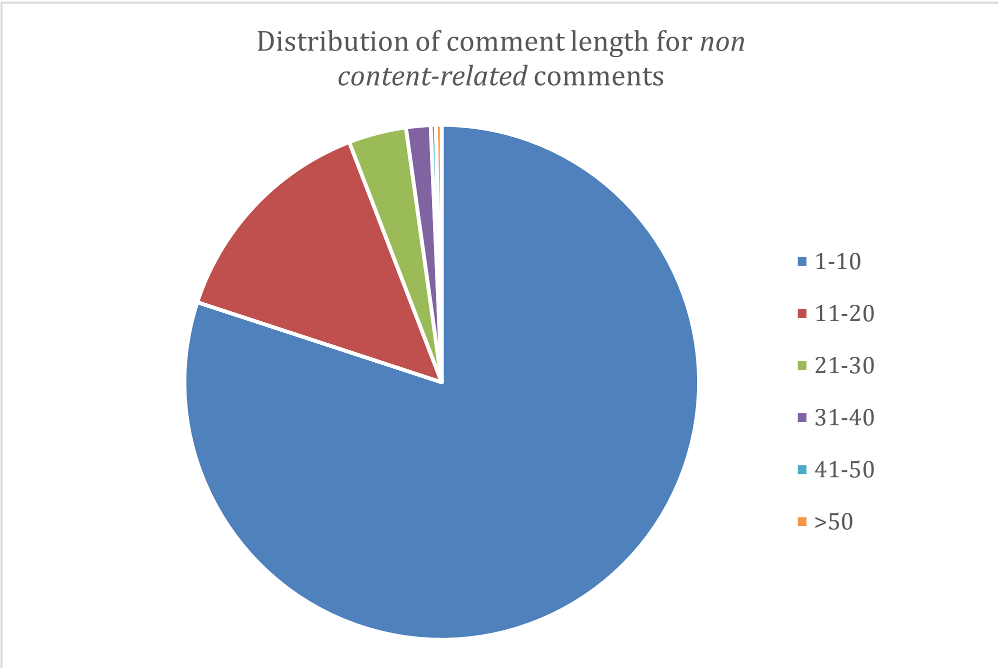


Figure 32: Chart showing the comment length distribution for the *non content-related* comments

b) **Comment-range length**: Similarly, the comment-range length tended to be longer for the *content-related* comments than for the *non content-related* comments. As was the case for comment length, the comment-range length was also normalized between 0 and 1 using the formula: $zi=(xi-min)/max-min$, where zi is the normalized value, xi is the original value, min is the minimum value of length in the sample and max is the maximum value of length in the sample.

c) **Edit distance:** This is a measure of similarity between the comment and the comment-range. If a comment was a correction of grammar or spelling, then the comment and comment-range tended to be similar. There are several measures of similarity (Sun, Ma, & Wang, 2015), but for simplicity reasons, edit distance (also known as Levenshtein Distance) was considered because the interest is in how different the comment was from the comment range. Comments to correct the grammar tended to have a higher similarity measure than *content-related* comments. The value of this feature was a real number between 0 and 1.

d) **Grammatical keywords:** The comments that were *non content-related* commonly contained words or phrases related to the grammar or formatting of the comment range text, such as 'redundant', 'sentences', 'misspell' 'font', 'Calibri', 'style', 'move figure', 'change order', etc. 51% of the comments that contained any of these keywords were *non content-related*. In other words, if a comment contained any of these keywords, there was a 51% chance that it was a *non content-related* comment. Although it is not in itself a very discriminatory feature, when combined with the previous features, the results are better than when not included at all. The list of the words used can be found in Appendix 1. The values of this feature were Yes (it contained the keywords) or No (it did not contain keywords). The reason for not considering each individual keyword as a feature was to keep the number of features small. From several tests, considering each keyword as an individual feature did not improve the results of the classification algorithm.

e) **Content keywords:** *content-related* comments generally had words or phrases such as 'describe', 'recommend', 'expound', etc. 68.5% of the comments that contained these keywords were *content-related*. In other words, if a comment contained any of these keywords, there was a 68.5% chance that it was a *content-related* comment. The list of *content-related* keywords was compiled using the frequency count as well (refer to Appendix 2). These words may imply that the author has to put in considerable effort in revising the relevant section. Similar to the grammatical keywords, the values of this feature were Yes (it contained the keywords) or No (it did not contain keywords).

5.8 Results and Discussion

In the testing phase, stratified k-fold cross validation was applied, with k=10. Stratification is the process of rearranging the data so as to ensure that each fold is a good representative of the whole. Therefore, in each fold, the proportion of comments (36% *content-related* and 64% *non content-related*) was maintained. Since there were 4,350 total comments used in the cross validation, in terms of the actual numbers, each fold had 435 items, of which 157 (36%) were *content-related* and 278 (64%) were *non content-related*. In each iteration there were 3,915 training data items and 435 validation data items.

The results obtained using LIBSVM (Chih-Chung & Chih-Jen, 2011) during classification are shown in Table 12. When using LIBSVM, it is recommended to first discover the best parameters for the data. Using the script provided, the parameters C= 512 and *g*=2 were recommended for a Radio Basis Function (RBF) kernel, which is a non-linear implementation of SVM. The RBF kernel nonlinearly maps samples into a higher dimension. The parameter C is for the cost of computation, which trades off misclassification of training examples against the simplicity of the decision surface. The parameter gamma, *g*, defines how far the influence of a training example reaches.

Performance criteria was accuracy, recall and precision of the *non content-related* comments as defined in Section 5.5. The average accuracy from cross validation was 86.3%. The average recall rate for the *non content-related* comments was 89.3%, and the precision for the same was 88.9%.

Table 12: Classification Results using SVM

k	***accuracy***	***recall***	***precision***
1	86.7	89.4	89.4
2	86.0	88.2	89.8
3	91.0	92.1	89.7
4	87.4	90.9	90.0
5	86.2	89.2	89.8
6	84.6	89.5	86.0
7	84.6	88.9	87.4
8	88.7	92.7	90.4
9	85.5	86.0	90.3
10	82.6	85.6	86.0
Average	**86.33**	**89.25**	**88.88**

There were five features selected to be used in the classification model: *comment length, the length of the comment range* (the text covered by the comment), *the edit distance* (similarity between comment text and comment range), *grammatical keywords* and *content-related keywords*. The algorithms used during the classification process were decision trees (DT) and Support Vector Machines (SVM). Better measures of performance were obtained using SVM.

- The **accuracy** result means that this approach correctly classified each comment 86.3% of the time. However, this does not tell us how the comments were actually classified. This is what the other measures of precision and recall are for.
- The average **recall rate** for *non content-related* comments was 89.3%. This means that using this approach, 89.3% of the *non content-related* comments were correctly identified as such.
- The **precision rate** was 88.9%, and it indicates the percentage of actual *non content-related* comments from among those predicted as "*non content-related*" by the algorithm. This means that if a comment was classified as *non content-related*, there was a high chance that it was actually so and would not require much revision effort.

As evident from the above results (over 89% in all the performance measures of recall and prevision), this research has proved that it is possible to use machine learning to automatically classify, to a reliable extent, the reviewer comments in academic article drafts. The application of comment classification in a revision tool could prove useful in filtering out the *non content-related* comments so that students can focus on global revision, which results in better quality of the revised articles. The comments classified as *content-related* can be applied in the IRT to estimate the revision skill of the students.

One limitation of this study is that it only classified the comments in the academic drafts from one laboratory. Different laboratories may have different writing styles, and the reviewers may also have a different way of writing their comments. However, it is believed that with minimum modifications of the features discussed here, it is possible to extend the classification model to reliably classify the comments of reviewers in other fields.

5.9 Error Analysis

The misclassified comments were analyzed. Table 13 shows some examples of the *content-related* comments that were misclassified as *non content-related*. The *content-related* comments that were misclassified as *non content-related* (Table 13) tended to have a higher edit distance (this means a high similarity between the comment text and comment range). This is the case for comments 6, 45, 58 and 59. Another possible reason for misclassification of this nature is that the comments had grammatical keywords, but not content keywords (comment 17). This implies that the keywords may not be optimal for all the cases. The length of comment 4 and corresponding comment range length were also short. Short comments generally tended to be *non content-related*. The impact of this type of error is that significant comments (*content-related*) may be filtered out. This means the student will be unable to attend to this type of comments first. Therefore it is important to find ways to reduce this type of misclassification errors.

Table 13: Examples of *content-related* comments that were misclassified as *non content-related.* The normalized values are in brackets ()

6	**Comment**	“Attaining the knowledge regarding designing and producing OVLs,” # Which type of knowledge?			
	Comment range	The participant’s attainment knowledge of designing and producing online virtual labs.			
	Comment length	**Comment Range length**	**Edit Distance**	**Content Keywords**	**Grammatical keywords**
	12 (0.041)	11 (0.013)	0.55	Yes	No
17	**Comment**	Local proxy should be added in this figure and this section.			
	Comment Range	Figure 9 System Architecture			
	Comment Length	**Comment Range Length**	**Edit Distance**	**Content Keywords**	**Grammatical keywords**
	11 (0.037)	4 (0.004)	0.30	No	Yes
45	**Comment**	Why PC’s view angle?			
	Comment Range	PC’s view angle			
	Comment Length	**Comment Range Length**	**Edit Distance**	**Content Keywords**	**Grammatical keywords**
	4 (0.008)	3 (0.003)	0.86	Yes	No
58	**Comment**	How about victim? The other agents means NPCs?			
	Comment Range	The other agents			
	Comment Length	**Comment Range Length**	**Edit Distance**	**Content Keywords**	**Grammatical keywords**
	7 (0.020)	3 (0.003)	0.52	Yes	No
59	**Comment**	In research paper, we have to claimed the novelty of the research through comparison of related work. You don't follow the way of academic writing.			
	Comment Range	Therefore, the novelty of the paper is developing a new game based on instructional thematic game for the children with ID.			
	Comment Length	**Comment Range Length**	**Edit Distance**	**Content Keywords**	**Grammatical keywords**
	25 (0.090)	21 (0.026)	0.48	No	No

Table 14 has examples of the *non content-related* comments that were misclassified as *content-related*. The *non content-related* comments that were misclassified as *content-related* (Table 14) had a low edit distance value (such as the case in comment 19), and also contained *content-related* keywords such as describe, define, content (comments with index 63, 63, 75, 80). However, this type of misclassification may be tolerated since it is better than we keep all *content-related* comments and therefore a few *non content-related* comments may be acceptable.

Apart from some *content-related* comments whose length was short, other sources of misclassification errors included lack of a clear distinction between some *content-related* comments and the *non content-related* comments. This occurred in cases where the comment contained both instructions to check the grammar and also to improve the content, or in cases where the comment required that the structure of a sentence or paragraph undergo considerable revision. In such cases, it would be better to classify the sentence as *content-related*. However, an additional class can be created to handle such comments, and this could lead to more satisfying results for the student who is revising the article.

Table 14: Examples of *non content-related* comments that were misclassified as *content-related.* The normalized values are in brackets ()

<table>
<tr><td rowspan="3">19</td><td>Comment</td><td colspan="4">You should check capitalized letter in the whole document. In this case, maybe “educational technology, and also…</td></tr>
<tr><td>Comment Range</td><td colspan="4">And</td></tr>
<tr><td>Comment Length
17 (0.061)</td><td>Comment Range Length
1 (0)</td><td>Edit Distance
0.03</td><td>Content Keywords
Yes</td><td>Grammatical Keywords
Yes</td></tr>
<tr><td rowspan="3">63</td><td>Comment</td><td colspan="4">It would be easy to read if you describe them as itemization form.</td></tr>
<tr><td>Comment Range</td><td colspan="4">Describe how to make up VLP tools design by Unified Modeling Language (UML) such as class diagrams. Describe the implement the VLP tool design through actual coding of computer language such as CakePHP and SQL. Describe the way of conduct validation and verification testing for each VLP developed tool and also educational content…</td></tr>
<tr><td>Comment Length
13 (0.045)</td><td>Comment Range Length
89 (0.112)</td><td>Edit Distance
0.18</td><td>Content Keywords
Yes</td><td>Grammatical Keywords
Yes</td></tr>
<tr><td rowspan="3">64</td><td>Comment</td><td colspan="4">After definition, you should use “OVLs” in the text. Check all contents.”</td></tr>
<tr><td>Comment Range</td><td colspan="4">online virtual labs</td></tr>
<tr><td>Comment Length
12 (0.041)</td><td>Comment Range Length
3 (0.003)</td><td>Edit Distance
0.26</td><td>Content Keywords
Yes</td><td>Grammatical Keywords
Yes</td></tr>
<tr><td rowspan="3">75</td><td>Comment</td><td colspan="4">Once you define a clipped word, you should use it directly to reduce redundancy. Check all of the text.</td></tr>
<tr><td>Comment Range</td><td colspan="4">virtual learning platforms (VLPs)</td></tr>
<tr><td>Comment Length
19 (0.070)</td><td>Comment Range Length
4 (0.004)</td><td>Edit Distance
0.21</td><td>Content Keywords
Yes</td><td>Grammatical Keywords
Yes</td></tr>
<tr><td rowspan="3">80</td><td>Comment</td><td colspan="4">Delete it and fill your contents.</td></tr>
<tr><td>Comment Range</td><td colspan="4">Describe specifically what you aim to achieve, and to what extent during the period of your doctoral program, and then state the expected outcome and its significance.</td></tr>
<tr><td>Comment Length
6 (0.016)</td><td>Comment Range Length
27 (0.033)</td><td>Edit Distance
0.21</td><td>Content Keywords
Yes</td><td>Grammatical Keywords
Yes</td></tr>
</table>

Another way to reduce the classification errors is by using structural information such as section title or figure captions, for example if a comment is made to the section title, it is likely to be *content-related*. However, some articles may not have such structural information.

Lastly, the method of selecting keywords is important. While the list of keywords was manually compiled, another method for selecting keywords is described that may improve the algorithm performance. As the number of words in all the comments may be as many as tens of thousands, to select which keywords are likely to indicate that a comment is *content-related* or *not*, a frequency count method can be applied. First, the total number of comments is divided into 10 equal blocks. Then one block at a time can be used to pick the keywords which will be tested on the remaining 9 blocks. During each testing process, about 50 words with the highest frequencies are picked as the keywords, and the results from testing the remaining 9 blocks of the comments are recorded. This process is repeated until all the comments have been included in the sample providing the keywords, and also in the testing sample. In this way, the keywords may be optimized for all the comments.

In summary, if a *non content-related* comment is misclassified as being *content-related*, it will be just a minor distraction. On the other hand, if a *content-related* comment is misclassified as being *non content-related*, this could have a more serious impact because an important comment may be ignored until later. It is desirable to minimize the latter errors.

The results obtained mean that using the approach outlined in this chapter, it can be predicted to a reliable extent whether a reviewer comment in an academic article is *content-related* or not. There is a limitation in that these are the comments in one laboratory. Different reviewers have different styles of writing reviewer comments. However, this research believes that the described model can be extended with minimum modifications to reliably classify the comments of other reviewers in other fields.

5.10 Chapter Conclusion

With performance measures of 89% that were achieved for both recall and precision, in this chapter, it was demonstrated that machine learning can be applied to automatically and reliably predict whether a reviewer comment in an academic article is *content-related* or *not*. The classification method was incorporated into TRONA to provide adaptation by estimating the student's revision skill level.

Furthermore, by classifying the comments in the article draft uploaded by the student as content-related or not, the student can start revision with the comments that encourage global revision, with the knowledge that the remaining comments are mostly simple spelling and grammatical errors that can be quickly dealt with. Take for example, a student who has 50 comments in their draft, of which 25 are content-related and 25, not. Once the student uploads their document to TRONA, about 90% of the *non content-related* comments can be filtered out. This gives the student an overview of the kind of global, content-related revision that needs to be done.

In Chapter six, the overall conclusion of this dissertation is discussed, as well as the contributions of this research and opportunities for future work.

6. Conclusion and Future Work

This research investigated the design, development and evaluation of a software tool to support the improvement of revision skill of graduate students. This dissertation began by introducing the background of this research regarding the improvement of writing and revision skill, and a review of tools that support the writing process. The problem statement and research gap were identified. The following research questions were proposed to address the problem statement:

1. How can students improve their revision skill?
2. How can a tool be designed and developed to support the improvement of revision skill?
3. How can the tool be adaptive to the various needs of students depending on their revision skill level?
4. How can machine learning techniques be applied to classify the comments in revision article drafts as either content-related or not?

This chapter begins with a synthesis of how the research findings addressed the research questions. Thereafter, a discussion follows on the implications of this research. Finally, the chapter discusses the limitations of this research and recommendations for future work.

6.1 Summary of Findings

6.1.1 How can students improve their revision skill?

Revision is a complex cognitive skill and due to this, one of the ways to learn it is by practice with a teacher in a traditional classroom setting. However, graduate students often have time limitations because they are focused on carrying out research. The findings of this research show that in the absence of a traditional classroom, research students can improve revision skill, while writing their own academic articles, through a software tool that fulfils two conditions:

- The tool provides a specialized revision corpus which contains not only the published articles in their research domain, but also the initial and subsequent drafts, including any reviewer comments, that led to those final articles.
- The tool is adaptive to the cognitive skill level (in this case the revision skill level) of the students. The tool should enable novices to increase their awareness of, and to focus on global revision, as this may improve their revision skill.

6.1.2 How can a tool be designed and developed to support the improvement of revision skill?

By modeling the revision process, it was possible to determine the type of input, content and output necessary to support the learning of revision skill by students. The type of input was determined to be the student's own draft containing comments, which they can upload to get hints on how to resolve those comments. The content required is the revision corpus containing not only the published articles in their research domain, but also the initial and subsequent drafts, including any reviewer comments, that led to the final articles. The output given includes various kinds of support such as hints to resolve comments, a schedule manager, a search and lookup function for the revision corpus, classification of the comments in their drafts, and other utilities. The tool should be web-based for easy access and navigation. This research developed such a tool, which is named TRONA – **T**opic-**R**elated Revisi**ON** **A**ssistant.

6.1.3 How can the tool be adaptive to the various needs of students depending on their revision skill level?

This research presented a way to apply the teaching methods of modeling, coaching coupled with fading described by the cognitive apprenticeship theory (CAT) to provide adaptation to the skill level of the students. The TRONA interface provides novices with a model of the revision process and a self-check list of the most common comments in the drafts. For intermediate students, it provides hints in form of revision histories of similar comments by previous students in the same laboratory (coaching). For advanced students, it also gives hints but they are less detailed (fading) than those presented to intermediate students. The Item Response Theory (IRT) was proposed as a suitable way to estimate the revision skill level of students. The Pearson's correlation analysis results showed a statistically significant correlation between the student scores by IRT and supervisor estimations.

6.1.4 How can machine learning techniques be applied to classify the comments in revision article drafts as either content-related or not?

At the core of the IRT are the *content-related* comments to be used as item models. The content-related comments are also used to provide hints to students to help them resolve reviewer comments in their own drafts. This research proved that it is possible to use machine learning to successfully classify comments in research article drafts as *content-related* or not by using the following features: comment length, comment range length, similarity between comment text and comment range text, content-related keywords and grammatical keywords. These content-related comments were applied in the estimation of revision skill using the IRT. Classification of comments also achieved a second objective: it enables students to increase their awareness of and focus on global revision. This is because the content-related comments are highlighted i.e. comments that trigger global revision are highlighted.

6.2 Contributions and Implications of the Study

The overall implication is that this research proposed method of preserving laboratory knowledge in the form of a collection of drafts, comments and articles written by students in one laboratory. It is generally difficult for educators to teach students revision as there is no standard revision curriculum. While there are many books on academic writing, it is quite difficult to find a general guide suitable for everyone because in academic writing, different fields have different writing styles, and even each laboratory has its own style. This research proposed how to preserve laboratory knowledge and how to use the archived knowledge for the benefit of new students in a laboratory. This ensures there is knowledge accumulation in a laboratory even as older students graduate and leave.

The contribution of this research is in the following specific areas:

- This research proposed and developed a software tool that enables students to improve their revision skill in the absence of a traditional classroom, which is a useful contribution in the current age of distance or e-learning.
- It proposed a way to build a specialized revision corpus which contains not only the published articles in a particular laboratory, but also the initial and subsequent drafts, including any reviewer comments, that led to those final articles.
- Machine learning classification of comments in research article drafts as content-related or not was achieved for the first time. This contribution is not only in the area of revision tools that use artificial intelligence to support the revision process of graduate students, but also in areas where comments may need to be classified such as in the review of technical documents.
- A method of providing adaptation to the revision skill level of students was proposed, through the teaching methods of the Cognitive Apprenticeship Theory.
- The Item Response Theory (IRT) was proposed as a suitable way to estimate the revision skill level of students. An example of the implementation was given and the evaluation of the IRT application affirmed the suitability of IRT.

6.4 Research Limitations

Revision skill is a cognitive skill, which takes a long time to learn. It is also hard to evaluate and what this research proposed was an estimation method. Time limitation implies that it may not be possible to truly know whether the skill level of the students improved with subsequent drafts. Other ways of estimating the skill level of the student include evaluating the quality of the output, but this may be more complicated because it may be hard to define what constitutes good quality, depending on the type of article or field of study.

This dissertation includes “improvement of revision skill” in the title, but was the improvement of revision skill actually achieved? The dissertation’s main contribution is in creating a basis for a software tool that what would be necessary in order to improve revision skill in academic writing. What was achieved was answering the necessary questions related to the background and definition of revision skill, the design and development of such a tool, and the application of machine learning in improving the efficiency of the tool, how to achieve adaptation in the tool, as well as a discussion on how revision skill could be evaluated. However, there is still need for a long term experimental setup to test and adjust for the actual improvement of revision skill. In summary, the dissertation answered the questions of how a software tool can be applied to improve revision skill for students carrying out research in one laboratory.

The focus of this dissertation was on laboratory education and therefore this study focused on implementing the tool, TRONA, in one laboratory. Different laboratories may have different writing styles, and the reviewers may also have a different way of giving feedback to the students. However, following the methodologies discussed in this dissertation, it may be possible to apply the tool to support students in other laboratories or fields.

6.5 Opportunities for Future Work

The following presents opportunities for future work, based on the research limitations identified in section 6.4:

6.5.1 Evaluating the Long Term Effectiveness of Adaptation in TRONA

To evaluate the long term effectiveness of adaptation in TRONA, subjective feedback can be obtained by asking students whether the hints and feedback they receive are useful or not. In addition, the students can be presented with a series of hints, some adaptive to their skill level while some will be non-adaptive, and then they will be asked which feedback they prefer. They have the option of requesting more or less feedback, and of answering whether the feedback is useful or not. The responses obtained and other interaction data could be analyzed for insights on the effectiveness of the adaptive interface. For further evaluation, each subject in the experiment could be given a number of comments to resolve. Some of the hints will be provided through a random approach and some through the adaptive interface. For example, a student can revise five comments through the random approach, and five comments by the interface. Based on the result of the revision, the quality of the revised text can be given a score. Thereafter, the average score of the random approach can be compared with the adaptive interface approach. In addition, not all the methods of the CAT were implemented. Extending TRONA to accommodate all the CAT methods could improve the effectiveness of adaptation in the tool.

6.5.2 Evaluating the Long Term Effectiveness of TRONA in Improving the Students' Revision Skill

Future work can also focus on evaluating the effectiveness of the approach discussed in this thesis on the outcome of the revision process, such as whether an increase in the awareness of global revision leads to better quality articles. This evaluation can be done by pre-tests and post-tests of the same students over a long period of time, say the duration of a master's or PhD course. It can be confirmed whether their revision skill has improved depending on the quality of their revised drafts before and after using the revision tool after a certain period of time.

References

Aluisio, S. M., & Gantenbein, R. E. (1997). Towards the Application of Systematic Functional Linguistics in Writing Tools. *Computers and their applications: Proceedings of the ISCA 12th international conference* (pp. 181-185). Tempe, Arizona, US: ISCA.

Amiangshu, B., Greiler, M., & Bird, C. (2015). Characteristics of useful code reviews: An empirical study at microsoft. *Mining Software Repositories (MSR), 2015 IEEE/ACM 12th Working Conference on* (pp. 146-156). Florence: IEEE and ACM.

Anzai, H., Kaminaga, H., Morimoto, Y., Yokoyama, S., Miyadera, Y., & Nakamura, S. (2012). A system for management of research resources focusing on grasping the polish processes. *Internet Technology And Secured Transactions, 2012 International Conference* (pp. 816-817). IEEE.

Assessment Systems. (2018, 04). *XCalibre*. From Assess Systems: http://www.assess.com/xcalibre/

Atreya, B., Walters, C., & Shepherd, M. (2003). Support Vector Machines for Text Categorization. *Proceedings of the 36th Hawaii International Conference on System Sciences.* Hawaii.

Baker, F. (2001). *The Basics of Item Response Theory.* Retrieved 1 19, 2018 from http://echo.edres.org:8080/irt/baker/

Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python.* O'Reilly Media Inc. Retrieved 9 11, 2017 from http://www.nltk.org/

Blake, R. (2011). Current trends in online language learning. *Annual Review of Applied Linguistics, 31*, 19-35.

Burgos, D., Tattersall, C., & Koper, R. (2007). How to represent adaptation in e-learning with IMS learning design. *Interactive Learning Environments, 15(2)*, 161-170.

Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes 30.3* , 222-234.

Charles, M. (2014). Getting the corpus habit : EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30-40.

Chee, S. Y. (1995). Cognitive apprenticeship and its application to the teaching of Smalltalk in a multimedia interactive learning environment. *Instructional Science 23*, 133–161.

Cherry, L. (1982). Writing Tools. *IEEE Transactions on Communications, Volume: 30*(Issue: 1), 100-105.

Chih-Chung, C., & Chih-Jen, L. (2011). LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology,* 2:27:1-27:27.

Cho, K., Schunn, C. D., & Kyungbin, K. (2007). Learning writing by reviewing in science. *Proceedings of the 8th iternational conference on Computer supported collaborative learning* (pp. 141-143). International Society of the Learning Sciences.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. *Knowing, learning, and instruction: Essays in honor of Robert Glaser 18* , 32-42.

Crossley, S. A., Kyle, K., & 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, no. 4,* 1227-1237.

De Boeck, P., & Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach.* New York: Springer.

De-Smet, M., Broekkamp, H. S., & Brand-Gruwel, P. K. (2011). Effects of electronic outlining on students' argumentative writing performance. *Journal of computer assisted learning,* 57-74.

Fahnestock, J., & Secor, M. (1988). The stases in scientific and literary argument. *Written communication 5.4* , 427-443.

Fankfort-Nachmias, C., & Nachmias, D. (1992). *Research Methods in the Social Sciences, Fourth Edition.* New York: St. Martin's.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication 32(4),* 365-387.

Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes 32.4* , 208-220.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System 32, no. 3,* 301-319.

Ghadessy, M., Henry, A., & Roseberry, R. L. (2001). *Small corpus studies and ELT: theory and practice. Vol. 5.* John Benjamins Publishing.

Gilmore, A. (2008). Using online corpora to develop students' writing skills. *ELT journal 63.4*, 363-372.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers 36, no. 2*, 193-202.

Grami, G. M., & Alkazemi, B. Y. (2016). Improving ESL writing using an online formulaic sequence word-combination checker. *Journal of Computer Assisted Learning 32.2* , 95-104.

Hartley, J. (2008). *Academic writing and publishing: A practical handbook.* Routledge.

Hasegawa, S., & Yamane, K. (2011). An Article/Presentation Revising Support System for Transferring Laboratory Knowledge. *19th International Conference on Computers in Education* (pp. 247-254). Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education.

Hayes, J. R., Linda, F., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. *Advances in applied psycholinguistics (2)*, 176-240.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing.* Harlow: Person Education.

Iwai, H., Hijikata, Y., Ikeda, K., & Nishida, S. (2014). Sentence-based Plot Classification for Online Review Comments. *IEEE/WIC/ACM International Joint Conference on Web Intelligencce (WI) and Intelligent Agent Technologies (IAT)* (pp. 245-253). Warsaw: IEEE Computer Society Press.

Joty, S., Alberto, B.-C., Giovanni, D. S., Simone, F., Lluís, M., Alessandro, M., & Preslav, N. (2015). Global thread-level inference for comment classification in community question answering. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 573-578). Lisbon: Association of Computational Linguistics .

Kakh, S. Y., & Wan Mansor, W. F. (2014). Task-based writing instruction to enhance graduate students' audience awareness. *Procedia-Social and Behavioral Sciences, 118*, 206-213.

Kashihara, A., Shinya, M., Sawazaki, K., & Taira, K. (2008). Cognitive apprenticeship approach to developing meta-cognitive skill with cognitive tool for Web-based navigational learning. *Seventh IASTED International Conference on Web-based Education.* Innsbruck, Austria.

Kaszuba, T., Albert, H., & Adam, W. (2009). Comment classification for internet auction platforms. *East European Conference on Advances in Databases and Information Systems* (pp. 129-136). Orhid: Springer.

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition 14.2* , 355.

Kobayashi, H., & Rinnert, C. (2001). Factors Relating to EFL Writers'discourse Level Revision Skills. *International Journal of English Studies, 1*(2), 71-101.

Kozma, R. B. (1991). Computer-based writing tools and the cognitive needs of novice writers. *Computers and composition, 8(2),* 31-45.

Lehr, F. (1995). Revision in the Writing Process. *ERIC Digest*.

Linacre, M. (2018, 6). *Ministep*. From Winisteps: http://www.winsteps.com/ministep.htm

Lynch, S. M. (2013). *Using Statistics in Social Research: A Concise Approach.* New York: Springer.

MacArthur, C. A. (2009). Reflections on research on writing and technology for struggling writers. *Learning Disabilities Research & Practice, 24*(2), 93-103.

Maydeu-Olivares, A. (2015). Evaluating fit in IRT Models. In S. P. Reise, & D. A. Revicki, *Handbook of Item Reponse Theory Modeling: Applications to Typical Performance Assessment* (pp. 111-127). New York: Routledge.

Merrouch, F., Hnida, M., Idrissi, M. K., & Bennani, S. (2014). Online placement test based on Item Response Theory and IMS Global standards. *International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2*.

Mosab, F., Abdulla, N., Al-Ayyoub, M., Jararweh, Y., & Quwaider, M. (2014). Cross-lingual short-text document classification for facebook comments. *Future Internet of Things and Cloud (FiCloud)* (pp. 573-578). Barcelona: IEEE.

Mueller, C. M., & Jacobsen, N. D. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL, 28*(1), 3-21.

Mukherjee, A., & Bing, L. (2012). Modeling review comments. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 320-329). Jeju Island: Association for Computational Linguistics.

Narita, M. (2012). Developing a corpus-based online grammar tutorial prototype. *The Language Teacher*, 23-29.

Neuwirth, C. M., & Kaufer, D. S. (1989). The role of external representation in the writing process: implications for the design of hypertext-based writing tools. *Proceedings of the second annual ACM conference on Hypertext. ACM* (pp. 319-341). ACM.

Ocharo, H. N., & Hasegawa, S. (2017). Resource Description Framework (RDF) Models for Representing the Revision Process in Research Support Systems . *International Conference on Computers in Education*, (pp. 252-260). Christchurch.

Ocharo, H. N., Hasegawa, S., & Shirai, K. (2017). Topic-based Revision Tool to Support Academic Writing Skill for Research Students. *Proceedings of The Tenth International Conference* (pp. 102-107). Nice: ThinkMind.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction 1.2*, 117-175.

Pifarré, M., & Fisher, R. (2011). Breaking up the writing process: how wikis can support understanding the composition and revision strategies of young writers. *Language and Education, 25*(5), 451-466.

Refaeilzadeh, P., Lei, T., & Huan, L. (2009). Cross-validation. *Encyclopedia of database systems*, 532-538.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College composition and communication, 31(4)*, 378-388.

Suhaily , H., Rozainun, A. A., & Azmi, A. H. (2015). Postgraduate Tracking System: Student Research Progress Tracking Tool. *International Research in Education, Vol. 3*(No. 1), 47-53.

Sun, Y., Ma, L., & Wang, S. (2015, February). A Comparative Evaluation of String Similarity Metrics. *Journal of Information & Computational Science, 12*(3), 957–964.

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System 29 no. 1*, 91-102.

Tzu-Hsi, Y., Wu, J.-C., Jim, C., Boisson, J., & Jason, C. (2015). Writeahead: Mining grammar patterns in corpora for assisted writing. *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, (pp. 139-144).

Wakabayashi, R. (2013). The Effects of the Peer Feedback Process on Reviewers' Own Writing. *English Language Teaching*, *6*(9), 177-192.

Witte, S. (2013). Preaching What We Practice: A Study of Revision. *Journal of Curriculum and Instruction (JoCI)*, *6*(2), 33-59.

World Economic Forum. (2016). *The Future of Jobs.* World Economic Forum. From http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Wright, C. H. (2010). Technical writing tools for engineers and scientists. *Computing in Science & Engineering 12.5*, 98-103.

Yao, Y. Y. (2003). A framework for Web-based research support systems. *Computer Software and Applications Conference* (pp. 601-606). Dallas, Texas, US: IEEE.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing 13.4*, 257-283.

Yue, L., Zhai, C., & Sundaresan, N. (2009). Rated aspect summarization of short comments. *Proceedings of the 18th international conference on World wide web* (pp. 131-140). Madrid: Association for Computing Machinery (ACM).

Appendices

Appendix 1 – List of grammatical keywords

abbreviation	grammar	page	small
add	guideline	paragraph	space
arrange	indent	passive	spelling
article	itemization	past	structure
Calibri	letter	plural	style
capitalization	limitation	present	subject
caption	line	pronoun	table
check	lowercase	proof	tense
color	merge	reduce	Times
conjunction	misspell	redundant	typo
delete	move	remove	uppercase
error	New	Roman	verb
figure	object	sentence	word
font	order	singular	zoom

Appendix 2 – List of Content keywords

algorithm	difficult	structure
answer	discussion	target
catch	evaluate	true
clearly	example	understand
concept	explain	update
content	explanation	what
context	focus	where
contribution	learner	which
correct	logic	who
define	logical	why
definition	meaning	wondering
describe	originality	
description	question	
design	recommend	
detail	requirement	
development	results	
difference	revise	

Appendix 3 – List of content-related comments for IRT Models

1. What is **originality/original point** of your research?
2. **Define concepts** clearly/What's the **definition** of **concept**
3. Need more **explanation** through **examples/**Add some **examples**
4. What is the **research question**(s)?
5. The **structure** of the chapter/section is quite hard to understand (This paragraph/chapter/paper is **not well organized**)
6. I recommend you should **describe** expected **effectiveness** of the proposed **system**/**methodology**
7. **Logical structure** of the abstract should be considered/**Lack of logic** in paragraph/abstract/section
8. What is your **contribution** by this **article**/**paper**?
9. What do the **results** mean?
10. You should use **same word** for **same meaning**, especially in noun. Please check all of the document
11. It is not so clear about the **purpose/goal of** the **test**
12. **Effect assessment** is more important than **usability test**.
13. Do you have any **references**?/Include **references**
14. You should **describe** the **requirements**
15. **Audience** may find it hard to **understand/catch the meaning**
16. **Describe** not only **design** but also **development** of the system
17. Here you should **answer** the **question** simply. Check all chapters
18. You should **describe** the **motivation** of the research more clearly
19. **Evaluate** by using well thought **criteria** & **methods**
20. **Who** is the **target audience**/**users**?