2012

Online tutor for research writing

Nandhini Ramaswamy
Iowa State University

Online tutor for research writing

by

Nandhini Ramaswamy

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Co-majors: Human Computer Interaction; Computer Science

Program of Study Committee:

Stephen B. Gilbert, Co-major Professor

Jin Tian, Co-major Professor

Elena Cotos

Iowa State University

Ames, Iowa

2012

DEDICATED

I would like to dedicate this thesis to my parents, S. K. Ramaswamy and Latha Ayyar, and my sister Anu Radha & her family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

English is the most prominent second language used in educational programs throughout the world. Unfortunately, there is a limitation of time and skill to guide students with learning the language and for evaluating their writings. Automated Writing Evaluation (AWE) tools would help in addressing this gap. AWE is a relatively new field which has been of interest to a lot of researchers. It is believed to have the potential to aid in teaching and learning of writing skills in English. Recent improvements in Artificial Intelligence, Machine learning and Natural Language Processing has also enabled researchers to try new approaches to improve the features and performance of AWE systems. Although AWE could be of great use to both students and teachers, research analyzing the usage of such systems in classroom environments has been insufficient.

In this thesis, I document a contribution to the field of Automated Writing Evaluation in the form of a new AWE tool called the Research Writing Tutor (RWT). The system design, user interface design, and features of this tool are introduced first, and then the findings obtained from an user evaluation study are reported.

RWT is a web-based tool that provides instantaneous feedback on students' research writing using a machine-learning algorithm trained on an annotated collection of articles from 30 different academic fields. The system analyzes individual sections of the research article (Introduction, Methods, Results, and Discussion/Conclusion) and categorizes each sentence as a particular rhetorical shift with a functional meaning. The goal of the RWT, as explained by Cotos et al. (2010) in a proposal for research funding, is to make the students think of their writing as a series of rhetorical strategies to convey a particular functional

meaning. As the research writing norms differ depending on the discipline, the system is designed to be discipline specific. This tool could be of great use to graduate students and undergraduates in writing research reports, articles, and thesis or dissertations.

In this thesis, I develop the user interface, present the design and usability evaluation of a new web-based AWE tool, RWT, which uses the *Intelligent Academic Discourse Evaluator* (IADE) (Cotos, 2010) as its prototype. While *IADE* analyzes only the Introduction section of the research article, RWT is capable of analyzing all the sections of this genre. The website has been designed and developed to be user friendly. Unlike most studies that concentrate on the accuracy of the AWE systems, this study aims at the usability and utility of the RWT.

The study reported here focused on evaluating the two major modules of RWT: the analysis and demonstration module by three groups totaling 39 participants. The three groups of participants were used in order to see the difference in the evaluation of usability and utility of the tool based on their background training and practice with academic writings. As expected, the users who did not have a background or understanding of academic research writing found the tool to be less useful than the users who understood the purpose of the tool. Also the level of trust on automated systems was low among this group when compared to the other groups which attended courses or workshops on academic writing. The knowledge acquired in this study makes a significant contribution to the field of AWE and to the study of usability of such educational tools.

CHAPTER 1 INTRODUCTION

1.1 Overview

The Research Writing Tutor (RWT) is a web-based Automated Writing Evaluator (AWE) that provides formative automated evaluation to guide graduate students and undergraduates in writing research reports, articles, and thesis or dissertations. Once the user chooses the discipline and section of paper that he or she is writing (Introduction, Methods, etc.), and inputs the draft text to the system, RWT automatically categorizes each sentence in the student's writing as a particular rhetorical shift with a specific functional meaning. The writing is analyzed by the system with a machine-learning algorithm. The system is trained on an annotated collection of journal articles chosen from 30 different academic fields.

1.1.1 History of Automated Writing Evaluators

In order to understand RWT and its features and to see the improvements in the field of AWE over the years, it is important to know the history and features of previous AWE tools. Automated evaluation is a relatively recent research field with only 50 years of history. The first generation of automated systems was developed in the mid-1960s by Ellis Page and a few others. It was called the Project Essay Grader (Page, 1966, 1994; Yang, Buckendahl, Juszkiewicz, & Bhola, 2002). Although the pioneering works by Page seemed promising, AWE tools did not gain popularity for the next two decades. It was only in the 1980s that a few tools were developed to provide feedback on student writing. Since then, there have been a limited number of intelligent technologies and approaches developed to evaluate the

writing. There have been a variety of tools designed for various purposes over the years. The *Writer's Workbench* tool (WWB) developed by AT&T provided feedback on "spelling, diction, and readability" (Kukich, 2000, p.23). *Writer's Help* (WH) was a tool designed to give feedback on word frequency, sentence variety, transition word, and paragraph development. The *Intelligent Essay Assessor* (IEA), and *Concept Rater* (C-Rater) evaluate the content of the writing. *Project Essay Grade* (PEG) evaluates the style of the writing. The *Electronic Essay Rater* (E-Rater), *BETSY*, *Intelligent Essay Marketing System*, *SEAR*, *Paperless School free text Marking Engine* (PS-ME), and *Automark* evaluate both the content & style of the writing. The e-Rater, developed by the Educational Testing Services, has been the most popular of all as it provides the most accurate score of all (Burstein, Kukich, Wolff, Chi, & Chodorow, 1998; Burstein, Leacock, & Swartz, 2001). While there have been a lot of systems that concentrates on various aspects of essay writing, there are not many systems to aid in research writing with genre awareness. *IADE* serves as the prototype to RWT to build such a tool to help in academic writing.

Most research on AWE has focused on the accuracy of these systems. Usually this is done by having the system and a human rater score a set of writings and find the correlation between the scores generated. In addition to accuracy, there is also a high research interest to know about how these systems could be used for formative assessment of the writings by students (Charman & Elmes, 1998; Sambell, Sambell, & Sexton, 1999; Shermis & Burnstein, 2003; Myers, 2003). While there have been a lot of research papers on such systems, none of them, to our knowledge, has focused on evaluating usability of their respective tools. This study tries to close this gap in the research of automated systems.

1.1.2 Benefits of Automated Writing Evaluators

Given the advantages of previous systems in improving student writing, we believe RWT will also lend advantages to writers. The following section describes specific benefits that we may expect. AWE has a number of advantages for use in classrooms. The immediate, individual, automated feedback can motivate students towards better writing. Research by Grimes and Warschauer (1996) on the AWE tools *MyAccess!* and *Criterion* indicated that these tools motivated the students to practice writing and provided easier classroom management for teachers. The automated systems are capable of providing feedback on different aspects of writing. *Criterion* can even provide feedback on prompts that were not rated by human scores. It may also reduce the uncertainty of the students about their writing. Teachers' feedback on students writing is extremely valuable, but there are certain limitations to human ability as well. Also, as pointed out by Cotos (2010), there are times when the teachers feedback is vague like "Why?, Relevance?, Explain?" (Ferris, 2003, p.26), which could make the students frustrated (Ferris, 1995, Straub, 1997). AWE systems could be helpful in such cases as they could be designed to provide specific feedback on various aspects of the writing.

Automated feedback is reported to have a greater impact on revision of drafts compared to oral or written feedback (Tuzi, 2004). It can help in understanding the cognitive and social process involved in writing (Kukich, 2000). Also, the automated feedback is accurate, unbiased and provides no pressure to respond quickly. These systems could generate a wide range of feedback starting from individual feedback on grammatical

problems in writing to a holistic feedback on various features of writing like structure, style, content (Brock, 1990, 1993; Hyland & Hyland, 2006).

There has been an increasing demand of large-scale educational testing organizations to evaluate large number of writings in standardized tests conducted nationally and internationally. Human grading of these large scale writings involves training the evaluators for inter reliability training and regulation of grading using reliability checks (Hyland & Hyland, 2006). The cost involved for these makes the AWE an attractive and cheaper alternative to human evaluators. With an increase in class size, the AWE is looked as an economical feasible supplement to instructions given by teachers (Hyland & Hyland, 2006).

As the AWE systems are capable of storing large database of user writings, these could be used by the instructors for in-class discussions or for other purpose of analysis. Reviewing the previous drafts and corresponding errors could help students to monitor and self-correct their own writings (Yuan, 2003).

The RWT has been developed with the motive to help students in a classroom environment providing individualized automated feedback. It is believed to help students as an effective supplement to instructions provided by the instructors and to help in creating genre awareness.

1.1.3 Issues of Automated Writing Evaluators

There are quite a few opponents to AWE as there are proponents. This opposition may also apply to RWT, so it is worth understanding these issues. AWE tools are criticized

for their “over-reliance on surface features of responses, the insensitivity to the content of responses and to creativity, and the vulnerability to new types of cheating and test-taking strategies” (Yang, Buckendahl, and Juszkiewicz, 2002, p. 393). The system might not be able to provide a fair evaluation on writings that are well organized but with poor mechanics (Calfee, 2000, p. 35). Also, students might adjust their writing to the formulaic essay that would obtain the maximum score from the algorithm (Baron, 2005, p. B14).

AWE systems could be misused and thus “reinforce artificial, mechanistic, and formulaic writing disconnected from communication in real-world context” (Grimes and Warschauer, 2006). There is an opinion that the AWE could make students think of their writing as discrete stylistic components that operate independently of communicative contexts (Fitzgerald, 1994, p. 16). Also, the actual impact of the feedback generated by the automated systems is not clearly understood yet (Warschauer & Ware, 2006). Although the advantages of feedback and these automated systems are deemed true, they remain hypothetical at this point. There is a need for further research yielding solid statistical evidence for all anticipated advantages.

Finally, most studies of AWE tools seem to focus on the technical improvements and accuracy of the software, ignoring the learning and teaching process involved. According to critics, this issue makes these research studies methodologically unsound and outcome based, and "it leaves the educational process involved as a black box" (Warschauer and Ware, 2006, p. 14).

1.1.4 Problem Statement

RWT attempts to address part of the difficulties faced by students in academic writing, which are synthesized in Cotos (2010). Some of the most common problems faced in writing are poor structuring, which includes missing linking words or transition sentences, informal language selection, too many quotes, and ineffective paraphrasing. Both native and non-native speakers (NNS) find articulating ideas in the conventions of written English for academic writing as the most frequent challenges (Llosa, Beck, & Zhao, 2011). Specifically, NNS have reported less facility of expression, restriction to simple style, difficulty to provide appropriate amount of force on the claims for their research (Flowerdew, 1999).

Given that these are the general issues faced by students in writing, there are a number of AWE tools to address these surface level problems. But one of the major issues faced by graduate students in academic writing, which is competence in discipline specific writing (Huang, 2010), is an area that requires further attention due to the limitation of tools to aid in discipline specific writing. The main goal of this thesis is to develop the user interface and back end database for such a tool, RWT that attempts to address this gap in AWE tools. The corpus-based approach is considered to be powerful for creating genre awareness, which would aid discipline specific writing (Aston, 2002). Hence, RWT uses the corpus based approach which consists of training the machine-learning system with a wide set of real world discipline specific examples of use of language. RWT could be effective in addressing these difficulties by providing anytime feedback on the rhetorical development in response to individual leaner output.

1.2 Research Questions

A goal of this research work is to help develop a new AWE tool called RWT. This thesis attempts to answer the following questions:

1. Do the users find RWT to be user friendly?
2. Do the users find the various features of RWT to be useful for improving their academic writing?
3. Do the users trust automated systems? What is the level of trust on automated systems among graduate students?

1.3 Significance of the Study

While studies in the field of AWE have been done to investigate the accuracy of the system in evaluating the students writing, this study emphasizes more the usability aspect of one such AWE tool called RWT. This research could serve as an example for building more user friendly AWE software in the future. The importance of trust on automated systems will have a critical response on the way the tool is perceived by the students. The study reports important findings on the level of trust on automated systems and discusses how trust plays an important role in the usage of such tools. Certain features in the software, for example, have been designed to display the results of search in a format users are familiar with from the web. For example, the search feature of the demonstration module has been developed to closely match the Google search page design in order to make it easier for the user to relate the various features of the webpage.

1.4 Thesis Organization

This chapter provides an introduction to the field of automated writing evaluators by providing the history of AWE and discussing both their benefits and issues. It also describes the difficulties faced by students in academic writing, explaining how a part of the difficulties could be solved by RWT and the accompanying research questions. In the rest of the thesis, I document several contributions to the field of automated writing tutors in the form of research studies on how useful (utility), and easy it is to use (usability) the various features of the web-based RWT.

Chapter 2 reviews a list of existing AWE systems. It also reviews previous research work on various automated systems explaining the main concepts that were inspected and those that are yet to be explored. The *Intelligence Academic Discourse Evaluator* (IADE) that serves as the prototype for RWT, its results are also explained. Chapter 3 explains the improvements in RWT over IADE. The system design, database architecture, features of RWT and other technical details behind the functioning of the tool are explained. Chapter 4 explains the research approach, participants, materials, procedure, results and limitations of the research. Chapter 5 summarizes the results of the study and discusses directions for future research.

CHAPTER 2 LITERATURE REVIEW

This chapter is a literature review that elaborates on various existing AWE tools, their features, and their major utility for students. A detailed review of various studies and their results on various AWE systems is presented, as AWE has both supporters and opponents. Finally, as this research focuses on a user evaluation of the RWT interface, an overview of user-centered design approach which is used in the design of RWT is explained. In later chapters, the design decisions made for RWT, based on the best practices outlined here, is explained.

2.1 Existing AWE Tools

A variety of AWE systems have been developed over the years and there are a few systems that have gained commercial success in schools and universities across the world. An overview of a few other AWE systems is provided in this section.

SaK (Wiemer-hastings & Graesser, 1999) is an interesting writing tutoring system that speaks to the writer using avatars during the process of composition. Each avatar provides feedback on different aspects of the composition identifying strengths and weaknesses in the text but without offering corrections. *Summary Street* (Wade-stein & Kintsch, 2004) is another automated system that focuses on drills. It is not based on disciplinary concepts but based on skills like learning to summarize. It incorporates cognitive research on the development of summarization skills. The system generates feedback on the content of the summaries. While using Summary Street in classroom trials, students were found to be more engaged in writing task and showed improvement in writing skills.

Glosser is an automated feedback system that provides contextualized feedback to students about their professional writings. It uses textual data mining and computational linguistics algorithm to quantify features of the text to provide descriptive information about patterns in the text to the users (Calvo, 2010). The feedback from this system aims to provide different perspective on the writing and helps students stay engaged. The version 1.0 of Glosser provides feedback on the structure, coherence, topics and key words of the writing.

Writing-Pal is an intelligent tutoring system that incorporates many AWE elements in order to provide feedback on student essays (McNamara, Crossley, & McCarthy, 2010). It is targeted to provide writing strategy instructions to high school and college students. The *W-Pal* is capable of providing feedback to students' natural language input (Mcnamara et al., 2001). It consists of two principle components: Strategy Training and Essay Training. Strategy Training includes lessons to help in prewriting, drafting and revising an article. In the Essay Training module, students are provided feedback and suggestions to use a particular strategy to improve the essays. Usability evaluation indicates that the feedback from *W-Pal* improved the student essays after revision (Roscoe, Varner, Cai, & Weston, 2011).

The web based AWE tool, *Criterion,* was developed by the Education Testing Service with the capability to score and provide evaluation. It was targeted for use in K-12, colleges, US and foreign universities, national job training programs and military institutions. The software is capable of identifying errors in the syntax, discourse, topical content, lexical complexity, grammar, usage, mechanics and style of writing. A corpus based and statistical approach is used to detect these errors (Attali, 2004). In the corpus-based approach, the

system is trained on a large corpus of text from which the count sequence of adjacent words and part-of-speech pairs called bigrams are extracted. The system searches for bigrams that occur much less often than expected in a student essay based on the corpus frequencies (Chodorow & Leacock, 2000). The system recommends a five-paragraph essay strategy according to which an essay should contain an introductory paragraph, a three-paragraph body and a concluding paragraph. Studies indicate that the reliability between *Criterion* and experts is very high, approximately 0.98 (Shermis, Burstein, & Leacock, 2006.)

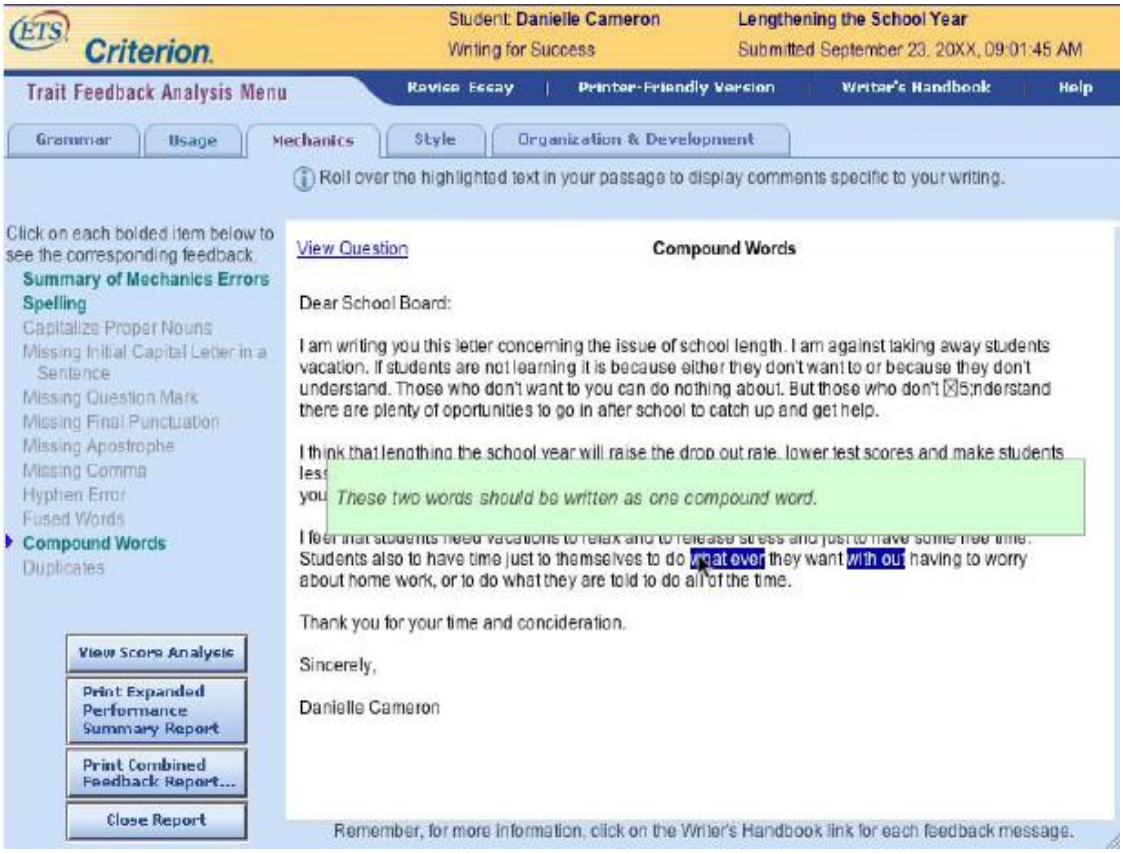


Figure 2.1: Sample Automated Feedback-Criterion

Criterion has two main components: E-Rater, an automated essay scoring system and Critique Writing Analysis Tool. This tool detects errors in grammar, usage, and mechanics. It

also identifies discourse elements and undesirable style in the essay. The various discourse elements in a well-written essay are introduction, thesis statement, main idea, supporting ideas and conclusion. The corpus-based approach was used to train the system to identify such discourse elements and provide feedback on their usage. The other features available in *Criterion* are explained next. The context-sensitive Writer's Handbook provides additional definitions and lessons to the user. This tool highlights the usage of passive sentences and very short or very long sentences and overly repetitious words, so the user could make changes. The system also provides a list of eight templates to aid students in planning their writing. The program has the ability to display both automated and teacher feedback to the users at different stages of their writing.

MyAccess!, developed by Vantage learning is one other popular AWE software that is used for the development of writing skills. The scoring engine is called as Intellimetric, which is capable of analyzing and providing feedback on the organization, content and development, focus and meaning, language use and style, and mechanics and conventions of writing. Immediate score along with diagnostic feedback is provided. It is used for both formative and summative assessment purposes. It is an artificial intelligence scoring engine, which is capable of analyzing some 300 semantic, syntactic and discourse features of an essay, and compares them to features of sample essays scored by humans (Elliot, 2003). It is majorly targeted towards K-12, colleges, the US and foreign universities. A large set of pre-scored essays with scores assigned by human raters is used to calibrate the scoring system.

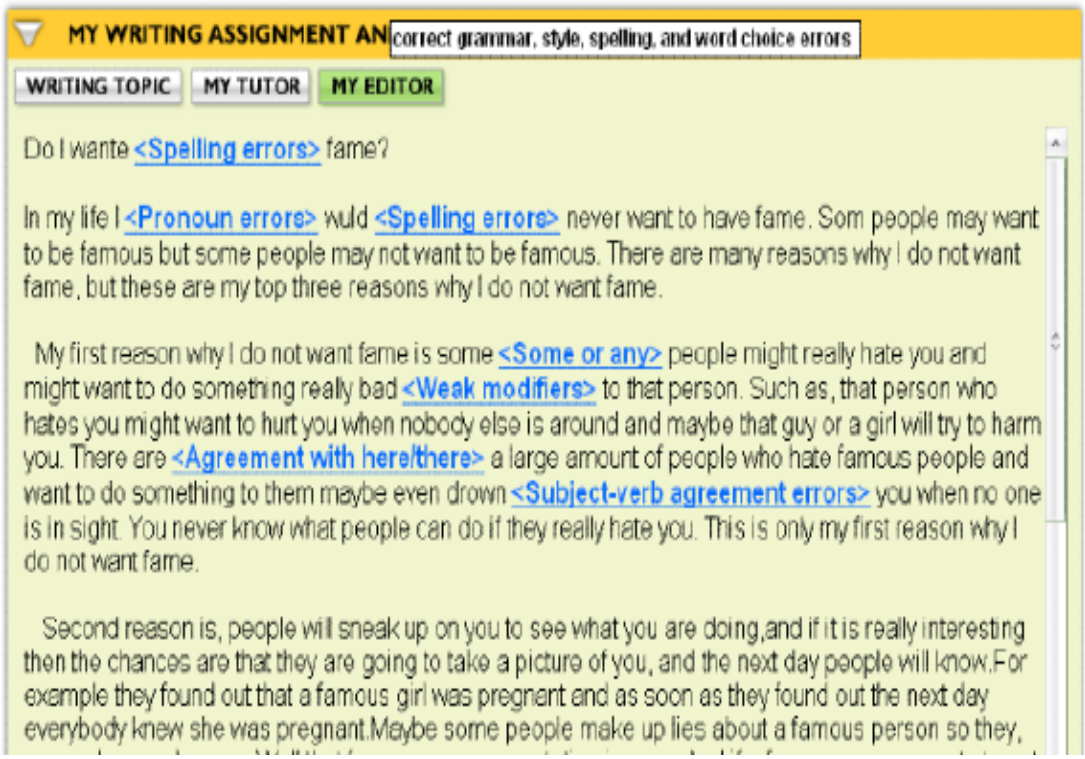


Figure 2.2: Sample Automated Feedback- MyAccess!

The software offers a variety of writing assistance features that includes My Editor, Thesaurus, Word Bank, My Portfolio, Writer's Checklist, Writer's Guide, Graphic Organizers, and Scoring Rubrics. My Editor highlights the error in writing and provides editing suggestions. The scoring rubric is used for self-assessment. Thesaurus is an online dictionary that offers a list of synonyms for the word being consulted. Word Bank offers words and phrases for a number of writing genres. My Portfolio allows students to access their previous works and view their progress. Writer's Guide is an online writing coach to evaluate and provide revision goals on student writings. Also for students who are native Spanish or Chinese speakers, the feedback can be provided in the native language. It also provides the ability for the teachers to create their own writing assignment or to choose a

topic from a list of over 700 prompts that are in line with state standards. These prompts include a variety of genres including narrative, persuasive, informative and literary.

SAGrader is an online learning environment. It uses the principles of artificial intelligence, NLP, computational linguistics and fuzzy logic to form a semantic network that stores the presence or absence of key features. The tool provides a natural language feedback on the content of the writing. It is used to encourage the students to work on multiple drafts by providing direct and specific feedback. The goal of this tool is to help students understand course-specific concepts and to enable effective communication of the content. It also allows the instructors to comment proactively on student writings.

The *Intelligent Academic Discourse Evaluator (IADE*) (Cotos, 2010), the prototype for RWT, is a web based AWE program that was developed as an additional tool for students to practice with and make incremental improvements on their drafts of research article introductions. The goal as explained by Cotos (2010) was to help the learners write academically, abiding by the writing conventions in their field, and to help them achieve a certain rhetorical purpose by learning to express functional meaning. The system terms the communicative strategies as "moves" and the functional meanings as "steps". For example, the moves for the Methods section are (1) contextualizing the study methods, (2) describing the study, and (3) analyzing the data (Cotos et al. 2012)

Figure 2.3: IADE color-coded feedback (Cotos, 2010)

In *IADE*, the students enter their article for analysis in the web browser that is encoded by the Hypertext Preprocessor (PHP) module and sent to the response processor where the drafts are analyzed. The PHP gets the analyzed data and displays it back to the user after saving them to the database. The *IADE* interface with the color-coded feedback is shown in Figure 2.3.

The *IADE* also provides definitions, step statistics, annotated corpus and revision tips on demand if the students need additional guidance. The numerical feedback generated by IADE is shown in Figure 2.4.

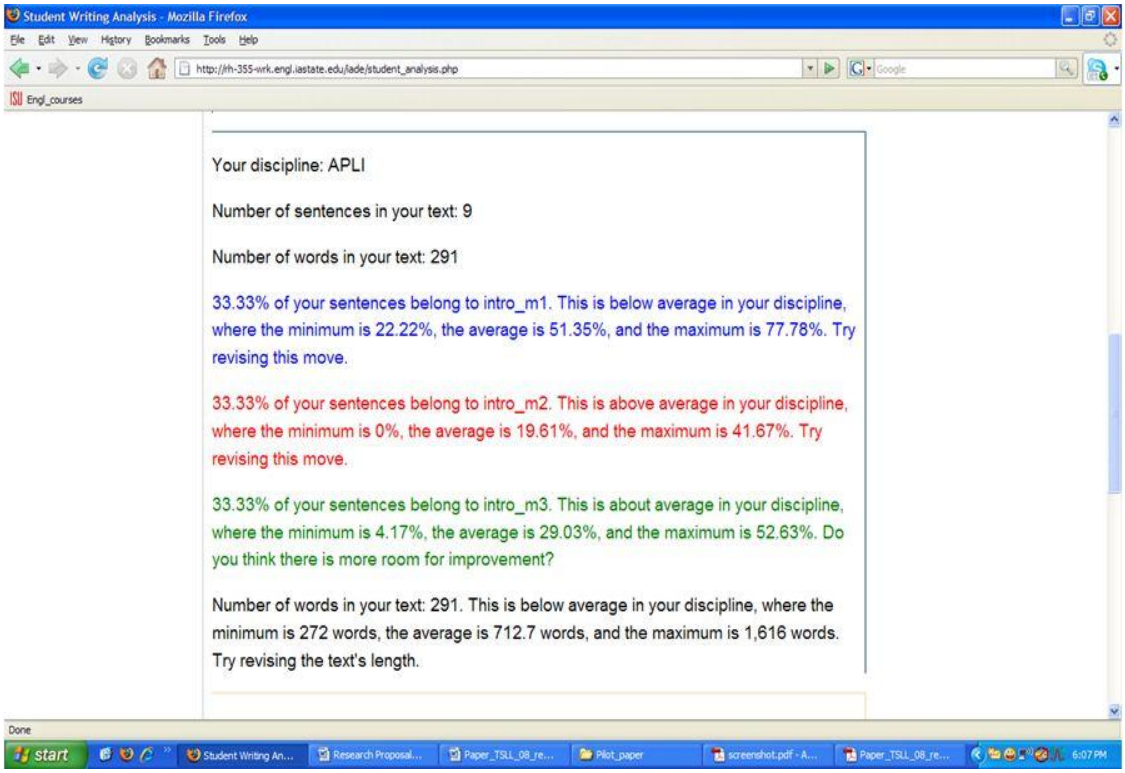


Figure 2.4: IADE Numerical Feedback (Cotos, 2010)

AWE Systems	Assessed Constructs
BETSY (Rudner and Liang, 2002)	Content, grammar, style, mechanics
Automark (Mitchell et al., 2002)	Content, grammar, style, mechanics
Project Essay Grade (Page, 2003)	Grammar, fluency, content, grammar, style, mechanics, plagiarism
E-Rater (Burstein, 2003)	Plagiarism, content, rhetorical structure, syntactic complexity
IntelliMetric (Elliot, 2003)	Focus, content development, elaboration, organization, structure, mechanics
Intelligent Essay Assessor	Organization, ideas, conventions, sentence fluency, word choice, writer's voice, spelling, copying, redundancy, irrelevancy

Table 2.1: AWE Tools

The various AWE tools and its features are summarized in Table 2.1. From the brief overview of AWE tools presented in this section, it could be seen that there are not much systems that aids graduate students with discipline specific writing. Hence, RWT could be really helpful in bridging this gap in AWE tools.

2.2 Related Studies on AWE Tools

A number of studies were conducted to address questions on the accuracy of the scores generated, the efficiency of the feedback, the utility of the tool, and the negative impact of AWE systems on writing. Researchers are also keen on knowing if the AWE systems actually improve students' writing or not, and how they are used (Warschauer & Ware, 2006). There are results that indicate that these systems could help in improving writing skills.

Now, the various studies conducted on the usage of *Criterion* for essay writing is discussed. System generated feedback is one of the important features in automated systems. Studies have been conducted to see if the feedback generated by *Criterion* helps in subsequent revisions of the essay. The study conducted by Attali (2004) evaluated the effectiveness of automated feedback and revision features of *Criterion* by focusing on the improvements in feedback from first to last submission of an essay. The drafts from the large-scale production environment were used for the study. The participants were between six and twelve grade. As the study involved large-scale production data, the user information and details of intermediate drafts were not available. Only 29% of the users used the software for analyzing more than one draft. Hence, only 29% of the users' data was used for analysis.

The results indicate that the students were sensitive to the feedback for most types of errors and were able to correct them in subsequent versions of their essay. In case of discourse elements, students were able to improve all the discourse elements in the essay except for the thesis statement.

Another user evaluation study of *Criterion* was conducted in one of the public schools in Miami (Burstein & Chodorow, 2004). In this study, the teachers who used *Criterion* in class responded to a survey in which they agreed that the immediate scores and feedback were useful. They also found the system easy to learn. Another study conducted by Mark Shermis (Burstein & Chodorow, 2004) reveals that there was no significant difference in the FCAT writing assessment scores between the students who were trained using *Criterion* and those who were not. But the study indicates that the students who used *Criterion* had significant improvements on later prompts. A similar study on *Criterion* to find the difference in writing skills between students in ninth grade across four states who used and did not use the software over a four week period was conducted (Rock, 2007). The results revealed that the usage of *Criterion* for a short time period improved the mechanical aspects of writing and found a small but statistically significant increase in the analytic score of essays.

There is also an argument on the validity of the results of the various researches on automated systems for classroom environments as it is highly different from standardized test environments. While automated evaluators are meant to improve the writing practice among students, results indicate that relatively few revision of drafts were carried out in a classroom

environment which could be because of the little time left for multiple revisions of single essay in classrooms (Attali, 2004).

Next, the various studies on MyAccess! is discussed. In order to understand the effective usage of AWE, MyAccess! was implemented in three writing classes in Taiwan as a part of a study (C. Chen, 2008). The main motive of the implementation was to reduce the instructors' workload. The MY Editor, Thesaurus, Work Bank and My Portfolio were used the most by the users in the study and had a positive reaction from the users. 50% of the users found the feedback generated to be helpful. The users commented on problems like: it favors lengthiness, over emphasizes the use of transition words, ignores coherence and content development, and discourages unconventional ways of essay writing. Although the students did not perceive the AWE very positively, the study shows that the teachers' pedagogical practices with AWE software can affect student perceptions of the effectiveness of AWE. It might not be a good idea to use AWE as a surrogate writing coach without human feedback as it could frustrate students.

The software is reported to have increased the writing ability of students in school districts that implemented MyAccess! (Suite, 2007). There has been significant increase in performance in statewide writing assessments of fifth through eleventh grade students (Elliot & Mikulas, 2004). A study conducted by Vantage Learning also shows that 85% of the users view MyAccess! as an effective tool for preparing for state mandated assessments. The users also reported the software to be user-friendly. The usage of MyAccess! for six weeks was reported to increase the overall score from 2.00 to 2.84 on a four point scale (Elliot & Mikulas, 2004).

Finally, the result of the study on *IADE* is explained. The results of the study conducted with 105 graduate students from Iowa State University (Cotos, 2010) revealed that the *IADE'S* feedback drew the participants' attention to the discourse form of their draft helping them notice the negative evidence in their writing, motivating them to revise it. It was also observed that there was an improvement in the quality of the discourse as the users changed the content, vocabulary, grammar, structure and mechanics of the writing. The users believed that the interaction with *IADE* and its feedback helped them learn about moves.

In the early stages, AWE research was primarily focused on the accuracy of automated scores given to the students (Warschauer & Ware, 2006). Although accuracy of the results produced by the algorithm is important, students can only benefit if provided with a clear, timely and usable feedback (Shute, 2008; Cowie, 1995). This is similar to what Douglas Reeves refers as "The Nintendo Effect" that is experienced by kids: "Kids respond to feedback from electronic games because it is immediate, accurate, and incremental." Similarly, when the students receive instantaneous feedback they get motivated to submit a revised essay as they would be motivated to get to the next level of a videogame (Suite, 2007). The immediate formative feedback is necessary for the students to understand what constitutes quality writing. Additionally, varying combination of feedback style might accommodate students who learn in different ways (Shute, 2008).

The RWT has been designed with care to provide formative feedback to the users in order to improve their writing practices and partially address the problems experienced by the users with existing tools. The evaluation presented in this thesis aims to address not just

the utility of RWT but also the usability of the RWT interface which is least explored by researchers in AWE.

2.2.1 Trust on Automated Systems

Trust is an important factor when it comes to using automation. The trust on automation of systems in various domains like cockpit automation, automated navigation has been studied by researchers. There are also a few studies that explored the trust of users on AWE systems. Studies report that some users prefer understanding how the tool works, its capacity and limitations to help them understand how much they can rely on an automated writing tool (Scharber & Dexter, 2008). A few users who originally had no trust on AWE systems preferred using the tool for a longer time to make a decision about using the tool (Scharber & Dexter, 2008) . With tools like AWE used in a classroom environment, the trust of the teachers on automated tools could also influence the trust of the students on these systems. A study on the use of one such AWE tool showed that the students found the tool to be more useful in cases where the teachers trusted automated systems(C. Chen, 2008). In this study, a student also reported that the reason for his/her distrust in the AWE system could be because of the distrust the teacher had in the system.

Given the importance of trust on automated systems for proper usage, the trust on automated systems among the user groups is also studied. The impact of trust on the perception of the usability and utility of the tool is analyzed.

2.3 Related Usability Research

Usability in human–computer interaction (HCI) is ‘‘the capability to be used by humans quickly, easily and effectively’’ (Shackel, 1991, p. 24). In simple terms it refers to the ease of use of an interface. A website or a software is considered to be usable if (1) it is easy to learn when the user visits it for the first time, (2) enables the users to perform the task efficiently once they learn the features of the website, (3) pleasant to use, easy to remember and use with ease for a returning user, and (4) if the user makes minimal errors using the website/software (Nielsen, 1995). Any software needs to be easy to use (usability) and provide the features the user wants (utility) for it to be useful. Any software that is not usable but provides the utility of the software or one that is user friendly without proper utility cannot be useful to the users (Nielsen, 1995). According to Krug (2000), some key features that could possibly make a software more usable are: identifying the difference between what is clickable and what is not, using buttons instead of hyperlinks, using a proper visual hierarchy for UI components, using minimum number of clicks to reach a destination page, using meaningful names, and using minimum possible scroll.. Drop down boxes are a good choice for providing users with a list of options in a space-constrained environment, but the drop boxes should be used with care as it involves an additional click and less readable compared to other menu options. Also it is better to use web conventions instead of reinventing the wheels. All these best practices in UI design has been taken into account while designing for the user interface of the RWT. The design decisions made are explained in the next chapter.

User interfaces account for 48% of the code (Myers & Rosson, 1992) in a software product developed, but takes only 6% of the development budget (Nielsen, 1993) on usability engineering. Almost one third of the review is about the user interface of the software (Nielsen, 1993). Hence improved usability testing methods could help in greatly improving the quality of products. In usability testing, a group of participants are recruited in order to observe the difficulties faced by the users and see how they perceive the software. In this process, the users are asked to either figure out what a webpage is or to try using the website to do a particular task. It is always a good idea to conduct usability studies with different set of participants, if the audience fall in different groups or require some form of background knowledge about the software or website (Krug, 2000). As suggested by Krug, the user study of the RWT was conducted using three groups of participants, as the target audience of RWT could have varying background knowledge about the tool and its usage. The research about usability of AWE software is not presented here as none of them, to our knowledge, has focused on evaluating usability of their respective tools.

One other topic of interest for this study is the correlation between mouse hovers and eye movement in the webpage. The RWT software tracks the mouse hovers on the various features of the user interface in order to compare its usage rate. A recent study conducted by comScore, Inc and Pretarget (2012) for over nine months on about 263 million impressions indicates that the correlation between hovers and purchase rate in online advertising is 0.49 where 0 indicates no correlation and 1 indicates maximum possible correlation. As the mouse movements are known to have a correlation of about 84 to 88% with eye movements (Chen,

Anderson, & Sohn, 2001) we decided to track the hovers on the user interface, the results of which are provided in Chapter 4.

CHAPTER 3 ONLINE TUTOR FOR RESEARCH WRITING

This chapter is outlines the structure of a research article, explains the previous work done with the *Intelligent Academic Discourse Evaluator (IADE)* (Elena Cotos, 2010) and the improvements over *IADE*. Then the various modules of the RWT, its system and UI design and database architecture is explained in detail.

3.1 IADE to RWT

The RWT is an extension of IADE with the ability to analyze and provide feedback on the Introduction, Methods, Results, and Discussion & Conclusion sections of a research article. Also to make it more interactive the user interface of the *IADE* (see Figure 2.3) has been totally redesigned and developed to include interactive discipline-specific, color-coded & numerical feedback. RWT consists of the analysis and demonstration modules. The analysis module enables the user to browse through the history of drafts and provide different visual and textual feedbacks. The demonstration module serves as an additional help module to aid self-learning. Using this module, discipline specific examples for a particular section, move and a step combination in the corpus could be retrieved. While *IADE* was meant for integrated classroom use, RWT is designed to be supportive of longer-term usage for numerous and diverse learners.

3.2 RWT Implementation

The current implementation of RWT has several improvements over the previous version, *IADE*. The modules that are implemented in RWT along with the system and UI

design and database architecture is explained in this section. The database and UI design for the two modules of RWT are my major contributions in the development process.

3.2.1 System Design

The system design of the RWT consists of programming logic and modules for the web browser and the web server as shown in Figure 3.1. In the figure, blue color represents my contributions to the system.

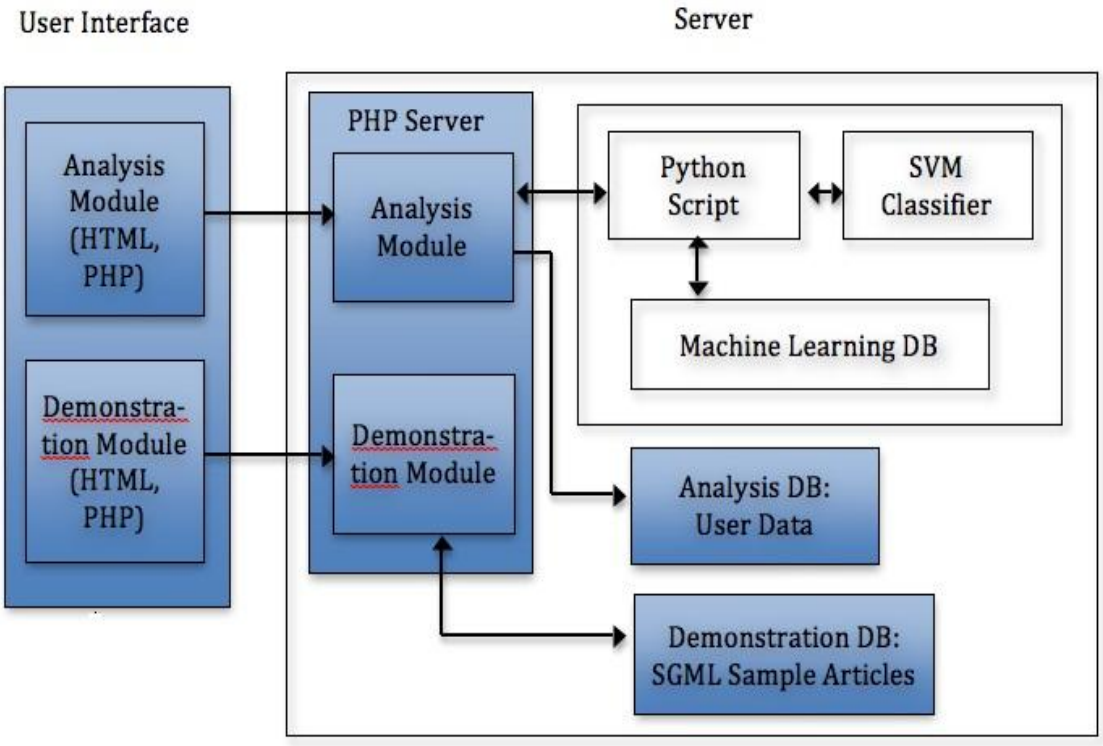


Figure 3.1: RWT System Design

The presentation module of the RWT supported by web browsers consists of the display logic for the analysis and the demonstration module. The user logs in to the system, drafts research articles, receive color coded and numerical feedback on hitting analyze

button, provides comment on system generated feedback and looks for history of articles written in the analysis page of the presentation module. The presentation module also facilitates the interface to search for examples in the demonstration module. The presentation module is developed using HTML, PHP, Java script and AJAX. The demonstration module was developed and added after the program was evaluated with a group of users.

Once the user enters their research article for analysis in the presentation module and hits on the analyze button, the data is sent over to the web server. The web server consists of PHP scripts, UI database for both the analysis and demonstration module, Python scripts, machine learning database, and a Support Vector Machine (SVM) classifier (see Pendar & Cotos, 2008). The PHP script takes care of the encoding and decoding of data, creating sessions, saving results of analysis and user information to the database. It also transforms the text submitted by the users into a format suitable for processing, acquires the results of analysis and converts the text into color coded and for display on the web interface.

The python script receives the data from the PHP script and preprocesses the text like breaking the text into sentences and sends it to the classifier for automated analysis. The SVM classifier analyzes and classifies each sentence as belonging to a particular move and a step. The lexical features indicative of a certain move are identified for this classification. The percentage for move distribution is calculated in the user article to provide feedback on the distribution and percentage of various moves in the writing. The distribution of moves in the annotated corpus of his/her academic field is also calculated. This data is presented as feedback to the user to enable comparison of the distribution of moves in his/her writing and the annotated corpus. This makes the feedback individualized and discipline-specific.

Lastly, the results of the analysis sent from the python script to the PHP script are stored onto the RWT database. After writing the results to the database, the PHP script again reads the data from the database and displays it back in the web interface.

The RWT is designed to provide evaluative comments on how the rhetorical moves in the writer's article approximate the norm in his/her discipline and not provide scores in order to avoid misuse which is one of the most important issues of AWE tools. This design would never encourage students to adjust their writing to the assessment criteria of the tool. Also, the tool helps the student acquire the knowledge of rhetorical functions in writing. The social factor or the influence of the teacher is not avoided as this tool serves as a supplementary tool in revision and not an assessor for academic writing. Studies indicate that students prefer systems that provide feedback on their writing than systems that only score their writing (Riedel, Dexter, & Scharber, 2006).

3.2.2 User Interface Components

The user interface was designed to make it user friendly and intuitive. The RWT has two major modules, namely the analysis module and demonstration module. The analysis module provides the ability to analyze and receive system generated feedback on the various sections of a research article. The user enters the text in the text area of the analysis webpage (shown in Figure 3.1) of the RWT website and hits on the analyze button on the webpage.

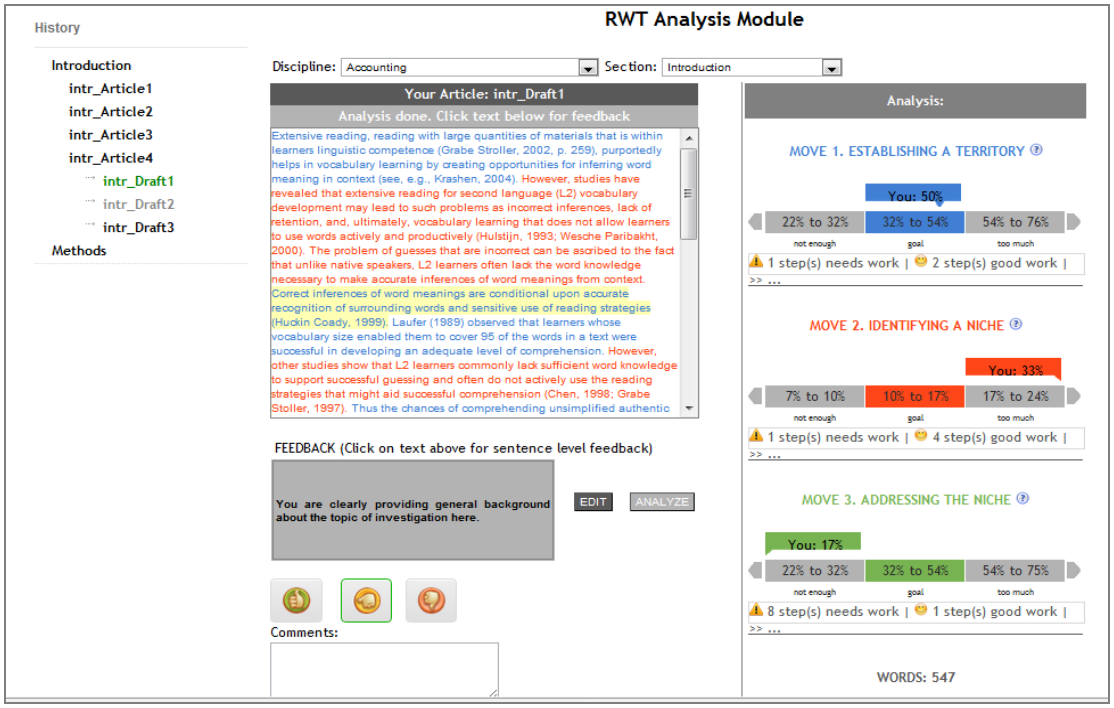


Figure 3.2: Analysis Module of RWT

The analysis webpage sends the user text to the machine-learning algorithm in the back-end, receives the system-generated feedback, and displays the results of the analysis back to the user. The components of the user interface screen are explained next.

The analysis module has various features that were designed mainly to communicate the feedback effectively to the users. A user centered design approach was used throughout the design process. The various features in the interface of the analysis module that a user sees when he/ she uses the software are: discipline and section selection, color coded text, sentence level feedback, thumbs button and user comments, range bars, pie charts and history of drafts.

After logging in, the user can choose the discipline and section for the writing from a drop down list. The various sections of a research article as supported by RWT are

Introduction, Methods, Results and Discussions & Conclusion. The current version of RWT supports up to 30 disciplines listed below.

- Agricultural and Bio-Systems Engineering
- Agronomy
- Animal Science
- Applied Linguistics
- Art and Design
- Bioinformatics and Computational Biology
- Biochemistry and Biophysics
- Biomedical Sciences
- Business
- Curriculum and Instruction
- Economics
- Food Science
- Forestry
- Geological and Atmospheric Sciences
- Horticulture
- Immunobiology
- Molecular Biology
- Mechanical Engineering
- Meteorology
- Microbiology

- Physics and Astronomy
- Plant Physiology
- Psychology
- Sociology
- Special Education
- Synthetic Chemistry
- Community and Regional Planning
- Veterinary Medicine
- Power Systems Economics
- Chemistry

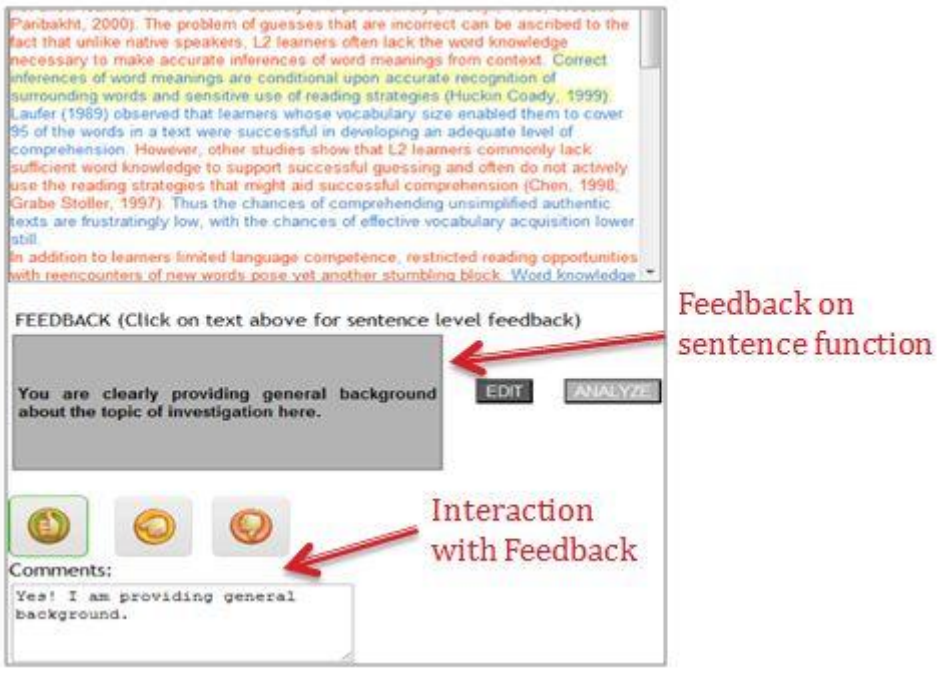


Figure 3.3: System feedback and User comments

After the user chooses the discipline and section of their choice, he/she enters the text to be analyzed into the text editor. When the user clicks on the analyze button, the text becomes color coded to represent the different steps and moves. The user can access the various drafts of the writing for one or more articles from the history of drafts. Clicking on each sentence of the color coded text gives feedback in a text area below. There is also a thumbs up, thumbs down, and thumbs neutral button where the user can provide comments about the system generated feedback (shown in Figure 3.3).

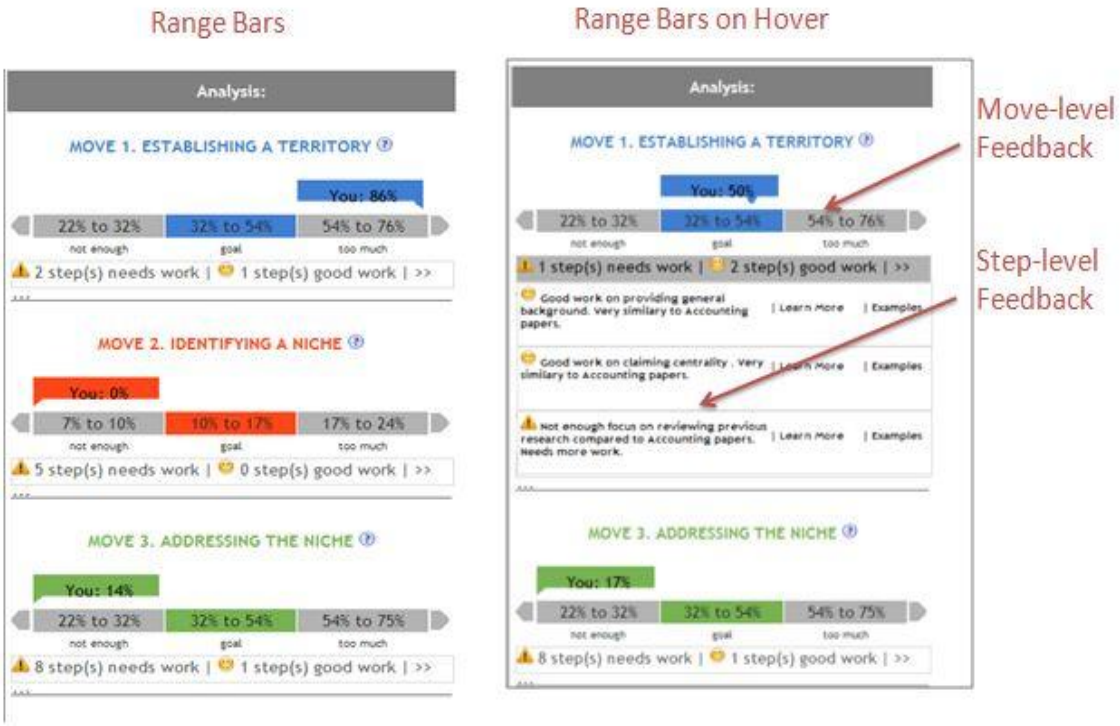


Figure 3.4: Feedback from Range bars on Hover

The users can use the range bars to view a visual analysis of his/her writing in a range of one percent to 100% for different moves. The arrowhead in the top of the range bar indicates the percentage of a move in the users' writing. The user can hover over the range

bar to view the drop down box containing step level feedback. The feedback consists of text explaining if a particular step needs attention or if the user has done a good job with a particular step, accompanied by a caution sign or a smiley face respectively. Research indicates that students favor comments that indicate the positive and negatives in his or her writing (Land & Evans, 1987). Hence comments like good job with a step might encourage the students to work further on their drafts. There are two other options available along with it. The user can click a learn more button to pop up description about the step along with an example. The writer can click the examples button to view the demonstration module loaded with examples that falls under the discipline, section, move and step that the user was working-on in the analysis module. The range bars are shown for all the different moves starting with move 1. Next, the feedback on word count is displayed in a similar fashion. The range bars indicating the different moves and the feedback on steps on mouse hover are shown in Figure 3.4.

The writers can see the percentage of different moves in his or her writing and the average number of moves in most research articles from their discipline using the pie charts presented at the bottom of the page. The above aids them in comparing their writing with established research articles. The pie charts indicating the percentages are shown in Figure 3.5.

The writers can use the demonstration module (shown in Figure 3.6) to look for examples of writings from the corpora. The demonstration module enables the user to search for discipline specific examples from various sections like Introduction, Methods, etc., that

fall under a particular move and step combination. For example, when the user receives a feedback from the analysis module that he/she needs to work on the step "Providing general background" in the introduction section of a research article in Agronomy, the user could search for the same using the demonstration module. This helps users understand how to construct sentences that provides general background for the introduction section of their research paper.

Figure 3.5: Pie Charts

In the demonstration module shown in Figure 3.6, searching for examples displays a list of all sentences that belong to the particular search criteria accompanied by the context of the text. The user can choose from a drop down list of 30 disciplines and four sections similar to the Analysis module. Choosing the section populates a dynamic drop down box for the moves. Similarly choosing the move populates a dynamic drop down for the steps in a particular move. Ten results are displayed per page and a hyperlink to the page numbers is displayed to view further examples. Clicking on an example displays the entire section of the research article as color-coded text on a display box right next to the sentence. Further, the user can click on every sentence in this color-coded article to understand its functionality. The results could span across various pages and the design presents several challenging tradeoffs like "Should the samples be from several articles or one?" and "How much context

should be shown around the sample sentence?" and "How many results should be displayed per page?".

Figure 3.6: Demonstration Module of RWT

3.2.3 UI Design Decisions

The various components of the RWT user interface were designed and implemented with care after previous research on the best practices in web designing explained in the previous chapter. In the analysis module, the drop down boxes was used to list the discipline and section choice. Using drop down is a good design decision as it provides the optimal design for displaying the list of 30 disciplines supported by RWT and is also considered to be the most efficient for alphabetized list (Krug, 2000). The entire width of the screen is divided into three parts with the first half displaying the history of drafts, the second half displaying the text area for the users to enter their article, and the third half displaying the visual feedback to the user. Having such a three-column view instead of a larger text area for the user article is better as it avoids scrolling to look at the feedback for each sentence. Also, as medium sized line length is considered to be better than longer lines for reading, we decide on 60-70 character lengths per line for the text area of the article (Baker, 2005). A simple text area was chosen for article entry instead of rich text or WYSIWYG editors as users could get distracted with formatting instead of concentrating on working on the feedback from the system. On hitting the analyze button, the user writing is converted into color codes as they could enhance the notice and focus on discourse form. The system-generated feedback is displayed below the text area in grey color in order to differentiate user writing from system feedback. The thumbs up, down and neutral button and the comments box are used to crowd source information about the system. Colors and fonts of different weight are used in the feedback column for gaining user attention. Hovering over the feedback provides detailed step level feedback. A drop down feedback on hover is used instead of a pop up feedback on

click in order to minimize the number of user clicks, one for opening the pop up feedback box and the other for closing the pop up box.

In the demonstration module, the drop box serves a similar purpose for selecting the criteria for search from the list. The search page is designed to closely match a known web convention, the Google search interface, as it would be helpful in minimizing the initial learning curve and be intuitive to the users. The entire context of the research article is shown in the second column when the user clicks on an example, instead of displaying a pop up with the entire article or redirecting the user to a new page. This design was done in order to avoid additional clicks in closing the pop up or going back to the search page. The results of the search are also displayed in randomized order in order to minimize the repetition of the same research article next to each other.

3.2.4 Database Architecture

The RWT database forms an important part of the overall system architecture. A part of the data for evaluation of the RWT is extracted from the database. The database stores the user information that includes the first name, last name, login name, email address and discipline, annotated corpora, list of all drafts submitted by a user, results of analysis of each draft, number of hovers on the step and move feedback in the form of range bars and pie charts, number of agreements/disagreements/neutral opinions, and other comments to the sentence level feedback. The various tables in the database are shown in Figure 3.7.

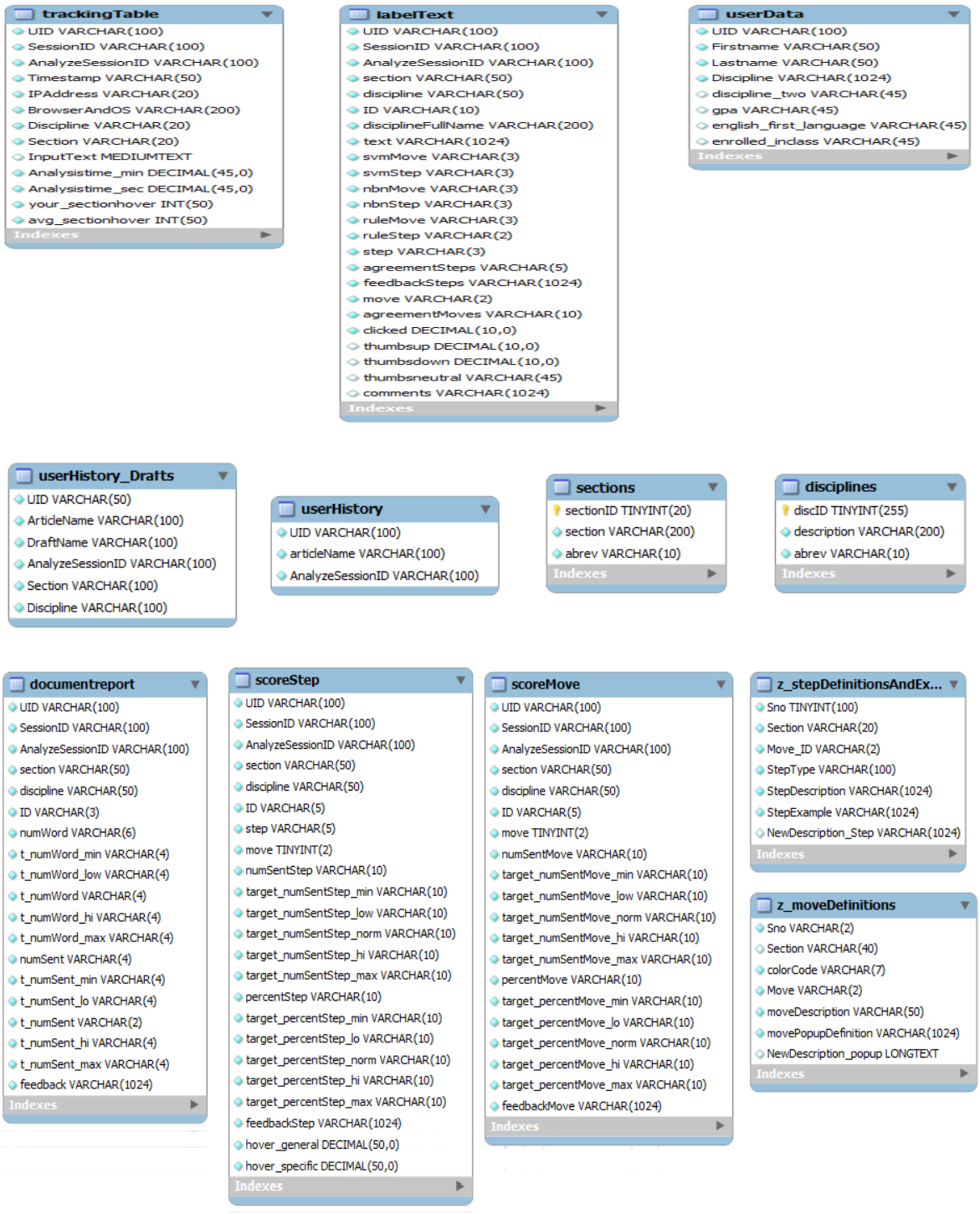


Figure 3.7: Tables in RWT Database

Chapter 4 EVALUATION OF THE TUTOR

This chapter elaborates about the approach undertaken to evaluate the usage of two primary modules of RWT: Analysis Module for the Introduction section, and Demonstration Module as a supplement to instruction. Both quantitative and qualitative data were collected from three different sets of evaluation with a total of 39 participants. The participants in the first evaluation set consisted of graduate students enrolled in English 508 coursework in the Spring of 2012 at Iowa State University*(n=9)*. The participants in second set of evaluation consisted of students who attended the academic writing seminar conducted monthly for graduate students *(n=24)*. The third set of participants consisted of graduate students who have not attended any courses or seminars related to academic writing, having little background understanding about the purpose and the concepts associated with the tool *(n=6)*. Following that, the materials and other data collection instruments is explained. The results that were obtained from multiple data sources and analyses are also explained in this chapter.

4.1 Participants

The participants for the study were split into three different groups. All the participants in the study were graduate students at Iowa State University differing by the amount of training and practice in implementing the move/step schema for academic writing explained in Chapter 3. The main motive of having these three different set of evaluation groups was to identify if there were any difference in the perception of the usefulness of RWT by different group of users who were given different levels of training and practice on academic writing.

4.1.1 User Evaluation – Group I

Users for group I of evaluation consisted of nine graduate students at Iowa State University enrolled in the "Advanced Workshop in Empirical Research Writing" course in the Spring of 2012. There were three male and six female participants, two native English speakers. The students from the ENGL 508 course were from the following disciplines:

1. Civil Construction & Environmental Engineering
2. Plant Pathology And Microbiology
3. Sociology
4. Textiles
5. Animal Science
6. English
7. Apparel Events & Hospitality Management
8. Physics & Astronomy
9. Economics

4.1.2 User Evaluation – Group II

Group II consisted of 24 graduate students at Iowa State University who attended monthly academic writing seminars conducted by the graduate college. The evaluation results of both the Analysis and Demonstration module for this group are presented later in the chapter. These students were also enrolled in the graduate programs at this university in the following disciplines:

1. Animal Science

2. Food Science & Human Nutrition-H Science
3. Chemical & Biological Engineering
4. Human Development & Family Studies
5. Agronomy
6. English
7. Materials Science & Engineering
8. Aerospace Engineering
9. Educational Leadership & Policy Studies
10. Kinesiology
11. Electrical/Electronic Manufacturing - I
12. Plant Pathology And Microbiology
13. Mechanical Engineering

4.1.3 User Evaluation- Group III

Group III consisted of 6 graduate students at Iowa State University who did not attend or take any courses related to academic writing. There were three male and three female participants, four native language speakers. These students were also enrolled in the graduate programs at this university in the following disciplines:

1. English (2)
2. Art & Design (2)
3. Interdisciplinary Graduate Studies (2)

Table 4.1 summarizes the differences between the groups in one glance.

	Significant training on moves/steps	Significant practice on moves/steps	Watch RWT training video
Group I	X	X	
Group II	X		
Group III			X

Table 4.1: Evaluation Groups

4.2 Materials

The RWT software itself is one of the core materials that were used. It was used both as a formative assessment tool for analyzing students writing and as a data collection instrument. The RWT database recorded the usage of on screen RWT features by tracking mouse hovers over various features to track the amount of usage. It also stored the feedback given by users, the different drafts and usage pattern by the users and the time taken for analysis. Data was also collected using Likert-scale surveys, which provided more quantitative analysis of the tool, and open-ended survey responses, which provided qualitative feedback.

4.2.1 Likert-scale & Open Ended Survey Questions

There were two different versions of the survey used. Group I evaluated just the Analysis module unlike the groups two and three. Hence Group I was given version 1 of the survey, which consisted of 14 questions for just the Analysis module (See Appendix A.1). The version 2 of the survey was then created which contains five additional questions for

Analysis module and 18 questions for the demonstration module (See Appendix A.1 & A.2). Both the surveys contained open-ended questions about the Analysis and Demonstration modules. A certain number of survey questions were required to evaluate and analyze the different aspects of the four research questions: In version 1, four questions for the usability of RWT, three questions for trust on automated systems, six questions for the utility of RWT and one question for the working of the RWT website. In version 2, five questions for the usability of RWT, two questions for trust on automated systems, eleven questions for the utility of RWT and one question for the working of the RWT website. The responses in the survey scaled from 1 to 4, 1 being strongly disagree and 4 being strongly agree.

4.2.2 RWT Database

The RWT database was used as an instrument to identify the frequency of usage of various UI features on the Analysis module of RWT. The RWT database contains information about the students name, id, discipline, history of drafts submitted, the results of the analysis for each draft, the hover information over the visual bars representing the feedback on steps and moves, hover information over the pie charts representing the results of analysis, the time taken for analysis, students feedback on the analysis by saving if the user hit a thumbs up or thumbs down or neutral, students written feedback on the analysis. The data from the database were all exported to worksheets, one for each group of users. Specifically the worksheets contained the following data: The time taken for analysis, the number of thumbs up, thumbs down or neutral, the number of hovers on the range bars of Move 1, Move 2 and Move 3, the number of hovers on the pie charts.

4.3 Procedure

As there were three different groups of users evaluated for the RWT, all the three procedures are explained. In group I, RWT was implemented as a part of classroom instruction in the Spring of 2012. They were given extensive training in the move/step schema for each of the sections of the research article. The training for each section lasted about three and a half weeks. During these sessions the students were shown lectures about, had knowledge quizzes on and did interactive exercises to practice applying the move/step schema for the different sections. After introducing the theories and providing enough practice materials, the instructor introduced the RWT tool to the students and explained how it worked with a demonstration. The demonstration consisted of submitting a draft for analyzing, receiving feedback and resubmitting the new draft to the system. The study lasted for about an hour in class when they were asked to use the RWT to analyze their own research article. Only the Analysis Module of RWT was evaluated by this set of participants.

The purpose of the study was explained in the participants' consent form, which was handed out prior to data collection. The ISU Institutional Review Board approved the consent form, data collection process using the survey and RWT database. The data collection was done in class after the students signed the consent form. The students filed an electronic survey out at the end of the session.

In group II, RWT was implemented as a part of the academic writing seminar arranged by the Graduate College. The seminars for each section of research article lasted for about two hours each. So this group of students had lesser practice time and training time compared to the group I students, but had an understanding of the theories behind academic

writing. After the series of seminars about the theories of academic writing, the RWT software was introduced to the participants by the facilitator with a small demonstration. The students used the RWT to analyze their research article after signing the consent form. Both the analysis module and demonstration module were presented to the users of this group. The electronic survey was filled at the end of the seminar.

In group III, RWT was implemented to a group of six participants. The study lasted for about an hour during which they were shown a short video presentation about what RWT is and a hand out, which explained what different steps and moves are. The students were then asked to use RWT to analyze their own research article. This group also evaluated both the Analysis and Demonstration module. The same consent form and electronic survey was used.

4.4 Results

The following section explains the results of the Analysis module and Demonstration module. The findings are reported as they address the research questions about RWT:

1. Do the users find RWT to be user friendly?
2. Do the users find the various features of RWT to be useful for improving their academic writing?
3. Do the users trust automated systems? What is the level of trust on automated systems among graduate students?

First for the Analysis module, results of the data collected from survey and the RWT database are provided for all the three groups of participants. The findings are shown for each

user groups. The results reveal that a majority of users from all the three groups found the tool to be useful but the level of trust on automated systems was very low among the users in group III. Next, results of the data collected from the survey are presented for the demonstration module. Results show that group II of users from the academic seminar found the demonstration module to be really helpful to look for examples.

4.5 Analysis Module

The Analysis module of the RWT has a variety of formats in which the results of the analysis and the feedback given on their writing are presented to the users. It includes the range bars for each move representing the percentage of user's writing that falls under a particular move. In addition to this, it also gives the step level feedback as a drop down when the user hovers over a particular move. There were also two pie charts in the bottom of the page that explains the percentage of different moves in the students writing and the average number of moves in most research articles from their discipline aiding them to compare their writing with established research articles. While the survey questions were meant to get the user's response on the utility of such features, the RWT database was used to track if the user actually used these features. Although the students submitted various drafts of their writing to the system, there was not much difference in the drafts in terms of the content. This could be because of the limited time given for the user evaluation session. Hence, results pertaining to the number of drafts submitted are not presented. All the other results provided in this section is based on the total number of drafts submitted by the users without considering the actual

difference in content between the drafts into account. In the following section, the results of the data analysis of the Analysis module for all the three groups of users are presented.

4.5.1 Group I

User comments on the system-generated feedback and usage data about various visual feedbacks were obtained from the RWT database. The average time taken for analysis of the various drafts (*n*=38) of user's introduction section of the article varied from three seconds to 30 seconds, averaging 8.5 seconds with a standard deviation of 6.6 seconds.

The pie chart showing the distribution of various moves in the user's article and the pie charts showing the distribution of various moves in an average article in the particular discipline were hovered over in 26% of the drafts. The range bars for the different moves, explaining the percentage of a particular move in the student's article along with smiley faces, has the following hovers: Range bar for Move 1 has been hovered over in 69% of the drafts, Move 2 and 3 in 54% of the drafts. Also the percentage of thumbs up, thumbs down, and neutral among those sentences that received user's feedback was 32%, 55% and 13% respectively. For all these values about hovers, the number of hovers on individual drafts was not taken into account as they varied significantly between drafts and users. The reason for such huge differences in the number of hovers needs further analysis.

Next, the data from the user survey is explained in order to answer following research questions.

- Do the users find RWT to be user friendly?

Four questions were used to answer the research question about usability. 11% of the users agreed that the website was difficult to use although none strongly agreed to it. Remaining 89% of the participants disagreed/strongly disagreed for the website being difficult to use. It could be seen that a majority of the users did not find the website as difficult to use.

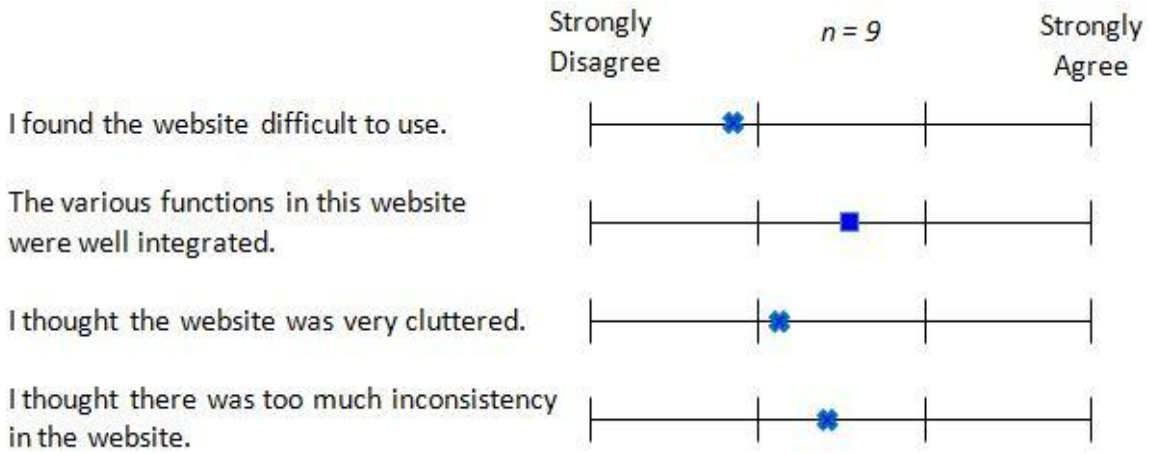


Figure 4.1: RWT usability data for Group 1

There was no strong agreement or disagreement in the response for the question "The various functions in this website were well integrated." But there was a 56% agreement. Also there was none who strongly felt that the website was cluttered. There was a 11% strong disagreement and 67% disagreement. For the question "I thought there was too much inconsistency in the website", there was an equal percentage (11%) of strong agreement and strong disagreement and a 44% disagreement. It is important to note that three out of four of these questions (shown as crosses in Figure 4.1) are negative and hence a lower score is a better. The mean value for all these questions is shown in Figure 4.1.

- Do the users find the various features of RWT to be useful for improving their academic writing?

Six questions were used to answer the research question about the utility of the tool. There was no strong disagreement for any of the questions about the utility of the tool. 89% of the users reported that: they understood the feedback, the feedback made them think twice about their writing, they looked at the feedback for the steps they needed to work on and that they felt this website will be very useful to the students. 100% of the users looked at the system generated feedback for each sentence. The mean values for all these questions are shown in Figure 4.2.

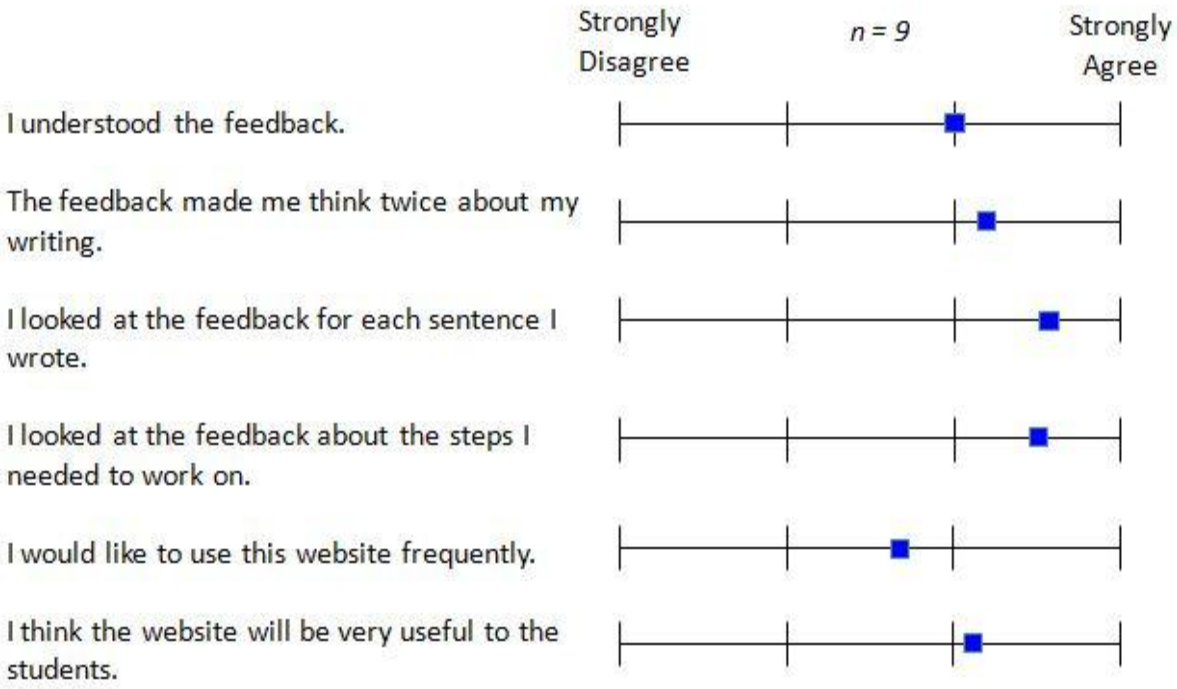


Figure 4.2: RWT utility data for Group I

4.5.2 Group II

For group II, the average time taken for analysis of the drafts (n=102) of the user's introduction section of the article varied from a few microseconds to 23 seconds, averaging 2.7 seconds with a standard deviation of 3 seconds.

The pie chart showing the distribution of various moves in the user's article and the pie charts showing the distribution of various moves in an average article in the particular discipline were hovered over in 41% and 51% of the drafts respectively. The range bars, for the different moves explaining the percentage of a particular move in the student's article along with smiley faces, has the following usage: Range bar for Move 1 has been hovered over in 72% of the drafts, Move 2 in 67% of the drafts and Move 3 in 56% of the drafts. Also the percentage of thumbs up, thumbs down and neutral among those sentences that received user's feedback were 58%, 25%and 16% respectively.

Next, the data from the user survey is explained in order to answer following research questions. Few questions were added to this survey in addition to the survey questions for group I.

- Do the users find RWT to be user friendly?

Five questions were used to answer the research question about usability. None of the users strongly disagreed with the questions about usability. All the users found the analysis module easy to use and thought that the various functions in the website were well integrated. 95% of the users thought that the website was consistent. 82% of the users found the analysis module easy to navigate and the website to be user friendly.

- Do the users find the various features of RWT to be useful for improving their academic writing?

Eleven questions were used to answer the research question about the utility of the tool. Only one of these questions received a strong disagreement. 82% of the users looked at the system-generated feedback for each sentence and 100% of the users reported that they understood the feedback. 90% of the users felt that the feedback made them think twice about their writing. 100% of the users looked at and understood the feedback for the steps they needed to work on. 95% of the users felt this website will be very useful to the students like them. 100% of the users looked at the visual feedback in the form of charts while 95% of the users understood them. 86% of the users would like to use the website frequently. 86% of the users agree that the automated feedback was helpful and 100% of users agree that it would help them improve their writing.

4.5.3 Group III

For group III, the average time taken for analysis of the drafts *(n=9)* of the user's introduction section of the article varied from a few microseconds to twenty three seconds averaging 1.7 seconds with a standard deviation of 1.8 seconds.

The pie chart showing the distribution of various moves in the user's article and the pie charts showing the distribution of various moves in an average article in the particular discipline were hovered in 89% of the drafts. The range bars, for the different moves explaining the percentage of a particular move in the student's article along with smiley faces, has the following usage: Range bar for Move 1 has been hovered over in 89% of the

drafts, Move 2 in 78% of the drafts and Move 3 in 56% of the drafts. Also the percentage of thumbs up, thumbs down and neutral among those sentences that received user's feedback were 69%, 17%and 14% respectively.

Next, the data from the user survey is explained in order to answer following research questions. The survey questions for this group are same as group II.

- Do the users find RWT to be user friendly?

Five questions were used to answer the research question about usability. None of the users strongly disagreed with the questions about usability. 67% of the users found the analysis module easy to use. 34% of the users thought that the various functions in the website were well integrated. 84% of the users thought that the website was consistent.50% of the users found the analysis module easy to navigate and user friendly.

- Do the users find the various features of RWT to be useful for improving their academic writing?

Eleven questions were used to answer the research question about the utility of the tool. Four of these questions received a strong disagreement of 16% each. 60% of the users reported that they understood the feedback. 83% of the users felt that the feedback made them think twice about their writing with a strong agreement of 17%. 50% of the users looked at and understood the feedback for the steps they needed to work on. 50% of the users felt this website will be very useful to the students like them.

57% of the users looked at the system-generated feedback for each sentence and 60% of the users understood the feedbacks. 50% of the users looked at the visual feedback in the

form of charts while 67% of the users understood them. 50% of the users would like to use the website frequently. 84% of the users agree that the automated feedback was helpful and that it would help them improve their writing.

Figure 4.3 compares the responses for the questions on usability of the Analysis module between group II and III. It can be seen that the group II has a better response than group III. Figure 4.4 compares the responses for the questions on utility between group II and III. Group II has a better user response than group III in this category as well. The results for group I is shown separately as the survey questions differed slightly from the other two groups.

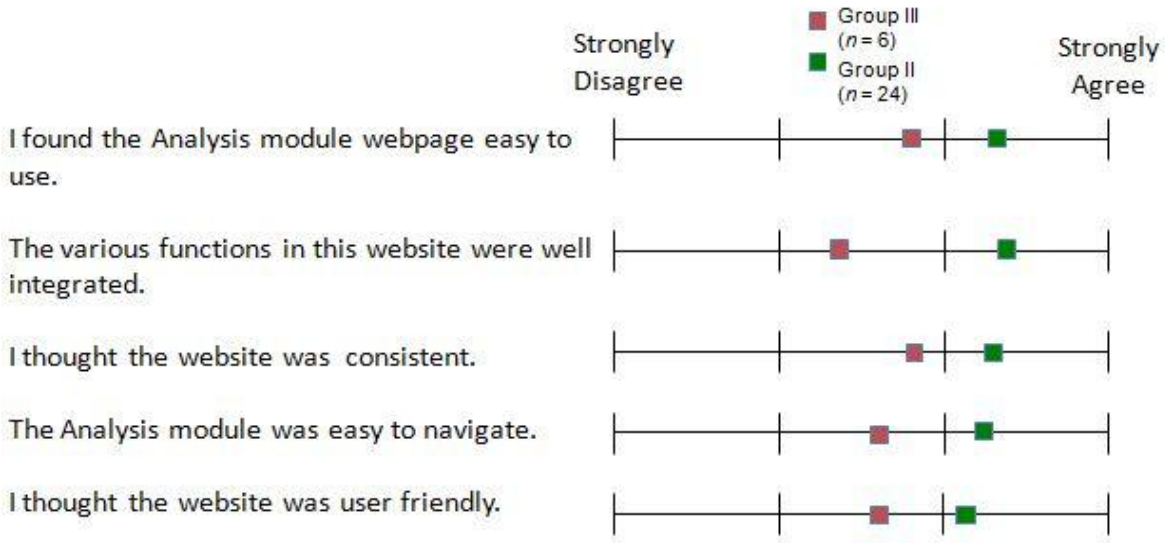


Figure 4.3: Usability data for Group II & III

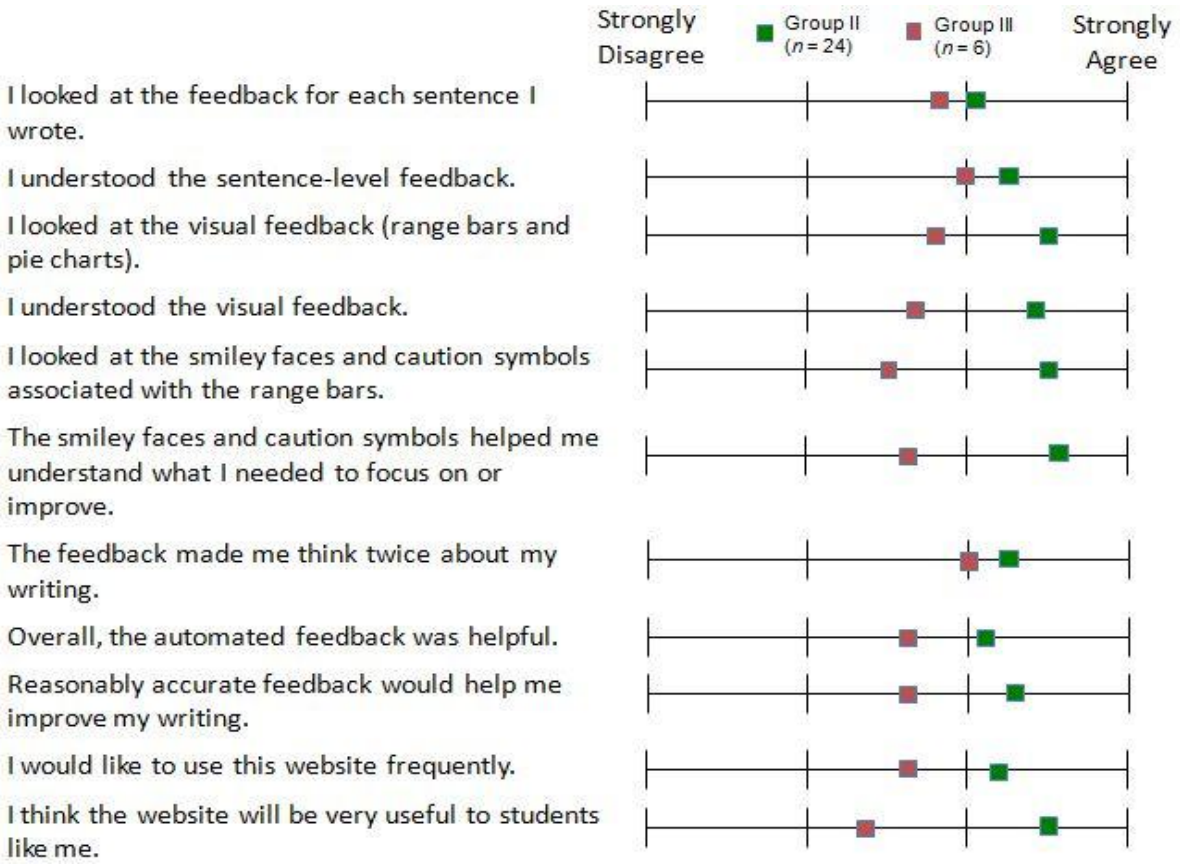


Figure 4.4: Usability data for Group II & III

The comparison between the percentages of drafts in which the various features of the Analysis module were hovered over is shown in Table 4.1. It is clear that the number of hovers is lower for features in the bottom of the screen. But overall, all the three groups seem to have hovered over the various features in the UI screen. It is interesting to note that the number of hovers for most features is high for group III although the users in this group dint find the tool to be as useful as group I and II. This could be because of the lower number of drafts and users in group III or because the users in group III wanted to try out all the features as they were first time visitors.

Feature	Percentage of drafts hovered on		
	Group I (*n*=38)	**Group II (*n*=102)**	**Group III (*n*=9)**
Pie charts – Your Article	26%	41%	89%
Pie charts – Average Article	26%	51%	89%
Range Bars - Move 1	69%	72%	89%
Range Bars- Move 2	54%	67%	78%
Range Bars- Move 3	54%	56%	56%
Thumbs up	32%	58%	69%
Thumbs Down	55%	25%	17%
Thumbs Neutral	13%	16%	14%

Table 4.2: RWT hover statistics for Analysis

The various qualitative feedbacks about the Analysis module of RWT comprised of the following five common themes indicated with examples:

- Utility of the software

"This interface gave me another prospect for looking into my own writing. (User 2) "

"It's very good for helping researcher analyzing their papers so that they have a clear picture of their paper compared to the discipline field. (User 7)"

- Support for more disciplines

"Add more selections for discipline. So students won't have to choose a closest discipline and then find out some paper examples that are actually not relevant to them. (User 4) "

- Accuracy of analysis (Mainly in Group I)

"I think it's an interesting tool & could be useful especially if the accuracy is improved. (User 8) "

- Additional feature request

".. , If this system integrate other system, such as Criterion, it would be great. In this case, we can edit both grammar errors and rhetorical meaning to improve our writing… (User 5)"

- UI Comments

"A better text editor would be helpful. (User 22)"

4.6 Demonstration Module

The demonstration module explained in Chapter 3 was evaluated only with group II and group III of users as the group I evaluation was done before the development of the demonstration module. The demonstration module has data only from the survey responses and is explained in the following section.

4.6.1 Group II

Of the 24 participants, there were only nine responses recorded for the demonstration module in this group. The data from the user survey is explained in order to answer following research questions.

- Do the users find RWT to be user friendly?

Five questions were used to answer the research question about usability of the demonstration module. None of the users strongly disagreed with the questions about usability. 89% of the users found the demonstration module easy to use, consistent and user friendly. 100% of the users thought that the various functions in the website were well integrated and easy to navigate.

- Do the users find the various features of RWT to be useful for improving their academic writing?

Eleven questions were used to answer the research question about the utility of the tool. 78% of the users agreed that they looked at a lot of examples and 100% understood them. 89% of the users used the search feature to look for more examples in different steps and moves. 56% of the users looked at more than one page of search results. 89% of the users felt this website will be very useful to the students like them and 78% of them would like to use the demonstration feature of RWT again.

4.6.2 Group III

The data from the user survey is explained in order to answer following research questions.

- Do the users find RWT to be user friendly?

Five questions were used to answer the research question about usability of the demonstration module. None of the users strongly disagreed with the questions about usability. 80% of the users found the demonstration module easy to use, 60% agreed to be consistent and user-friendly. 80% of the users thought that the various functions in the website were well integrated and easy to navigate.

- Do the users find the various features of RWT to be useful for improving their academic writing?

Eleven questions were used to answer the research question about the utility of the tool. 40% of the users agreed that they looked at a lot of examples and 80% understood them.

60% of the users used the search feature to look for more examples in different steps and moves. 40% of the users looked at more than one page of search results. 80% of the users felt this website will be very useful to the students like them and 80% of them would like to use the demonstration feature of RWT again.

Figure 4.5 compares the mean of the responses for the questions on usability of the Demonstration module between group II and III. It can be seen that the group II has a better response than group III. Figure 4.6 compares the responses for the questions on utility between group II and III. Group II has a better user response than group III in this category as well.

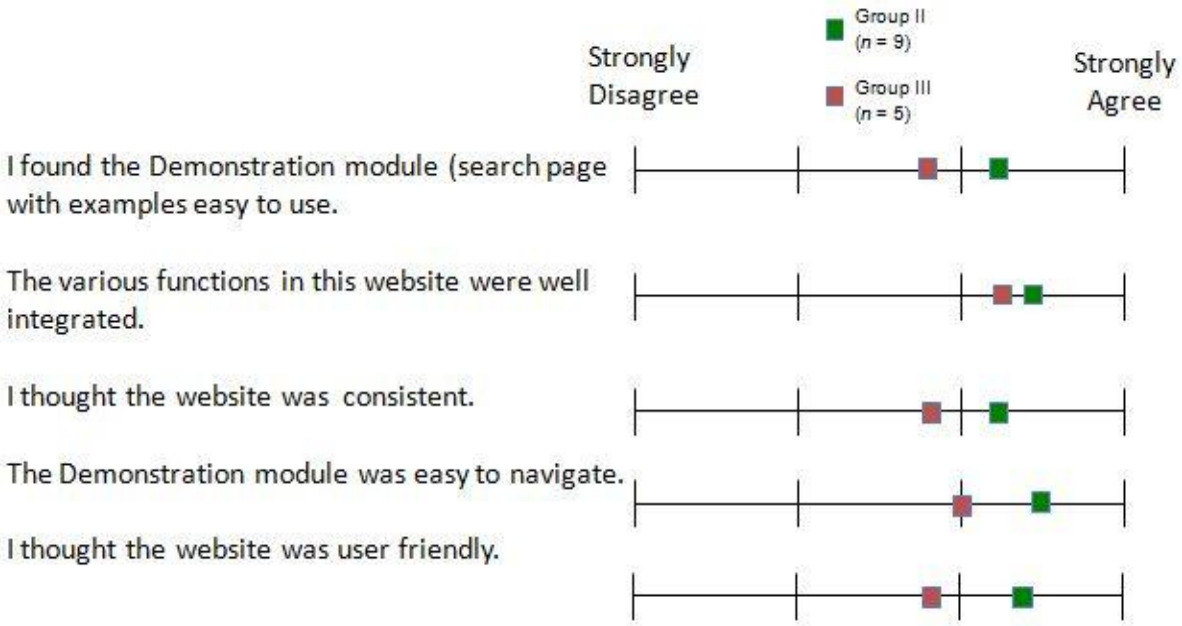


Figure 4.5: RWT usability data for Group II & III (Demonstration Module)

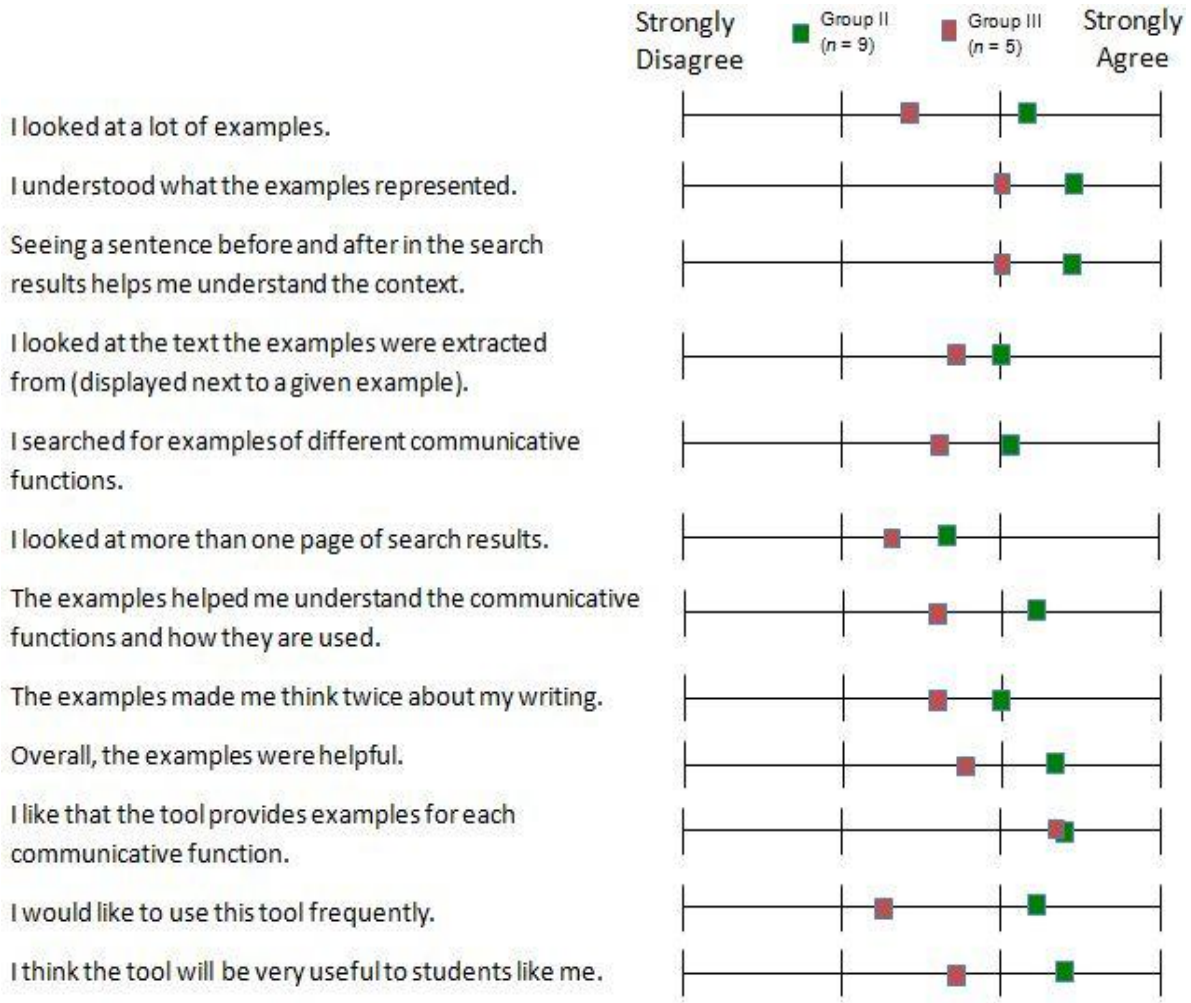


Figure 4.6: RWT utility data for Group II & III (Demonstration Module)

The various qualitative feedbacks about the demonstration module of RWT comprised of the following two common themes indicated with examples:

- Utility of the software

" I absolutely love this part of the program! Seeing how successfully published authors have done what I'm trying to do is enormously helpful.. (User 20) "

- Support for more disciplines

"Add more examples, especially the ones in my specific areas of research.. (User 8) "

4.7 Trust on Automated Systems

Trust is an important factor while presenting automated systems to the users. Hence there were survey questions to measure the level of trust among the participants and to see its impact on the usability and utility review. First the results of the survey for group I are discussed. 22% of the users strongly agreed that they do not like the computer evaluating their writing and there was no strong agreement on the trust on automated systems. 67% of the user disagreed on "I do not like the computer evaluating my writing". 78% of the users felt that they would change their writing if accurate feedback is produced. The mean values for all these questions are shown in Figure 4.7.

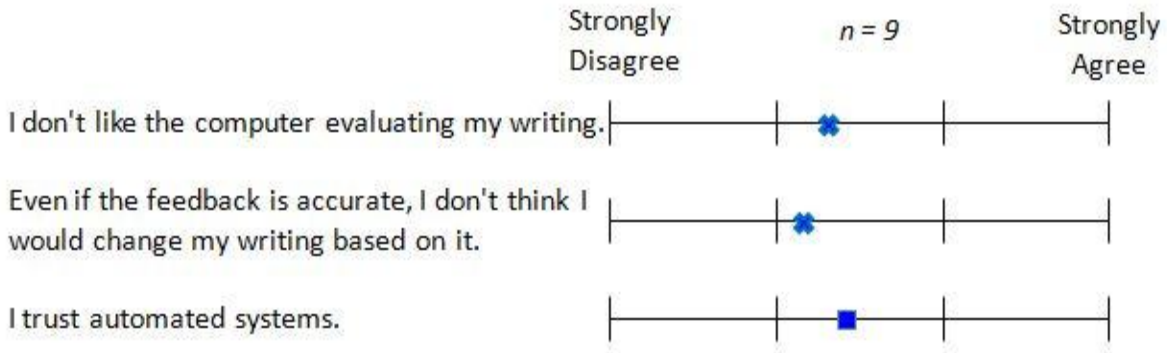


Figure 4.7: Trust on automated systems for Group I

In group II, 86% of the users strongly agreed that they like the computer evaluating their writing and 71% trusted automated systems. There was no strong disagreement for both the question. The mean values comparing group II and III is shown in Figure 4.8.

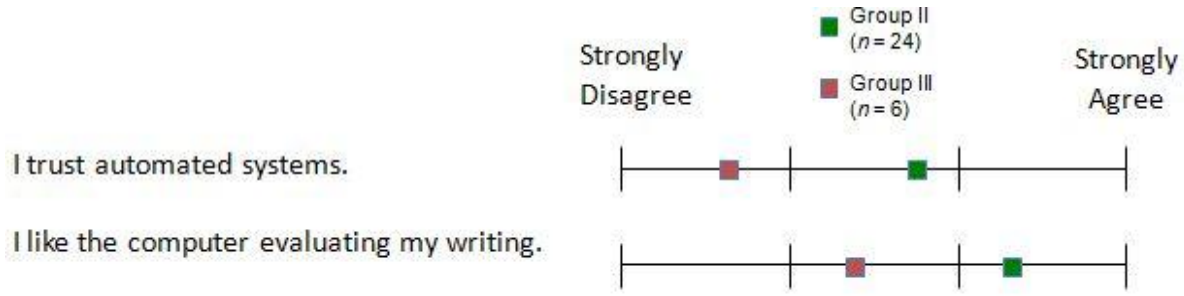


Figure 4.8: Trust on automated systems for Group II & Group III

In group III, 67% of the users strongly agreed that they like the computer evaluating their writing with no strong agreements and none of the users trusted automated systems. Clearly the trust on automated systems is lower among the groups, which did not undergo any courses or workshops related to automated writing systems for research articles. This clearly shows that there are chances of difference in perception about the utility of the tool among these three groups and that the results are subject to the variation in trust on AWE tools.

4.8 Limitations

Although the results from the user surveys are encouraging, there are certain limitations that restrict their generalizability. The limitations are related to the data obtained and the data collection tool used. There were two versions of the survey that was used to collect data and hence it is difficult to generalize the results for all the three groups of participants. The group I users did not evaluate the demonstration module, and the demonstration module survey for group II was not taken by all participants; leading to loss of data. The number of participants was limited and hence it might be necessary to perform a

study that involves a larger group of participants to gain statistical power. There were a few technical issues during the study by the group I participants, which could have some impact on the survey responses regarding the usability of the system. The data about hover on the various features does not necessarily mean that the users used the respective features although previous research indicates a high correlation between hovers and eye movements. Also there was significant difference in the number of hovers between users and between drafts which would require further research as the time of the hovers or the difference in time between the hovers were not tracked.

4.9 Discussion

We presented the evaluation results of the web based AWE tool, RWT based on the data collected from the three groups of participants. First, there is an important observation that needs to be pointed out about the groups. Groups I and II used the tool for improving their writing as a part of a course and a seminar respectively after which they evaluated the tool. With Group III the students were asked to use the tool for the purpose of evaluation. Hence, it is important to notice that the main motive behind the usage of the software varied between the three groups. Also, Group I and II were explained about the goals that could be accomplished using the software by the instructor. It was closely aligned with the goals of their course and the seminar. With Group III the expectations from the software were not set clearly as they were asked to use and evaluate the software after viewing a demo video.

Next, as the mouse hover is known to have a high correlation with the eye movements (Chen, Anderson, & Sohn, 2001), the hover on the various UI elements was tracked and

analyzed. A number of interesting observations were made during the analysis of the data. In a few cases, the number of hovers on various features reduced as writers worked on additional drafts. The number of hovers was lesser for features found on the bottom of the page than in the top of the page as expected which could be due to the additional scrolling involved. But, it is also important to note that the pie charts are passive UI component while the range bars are active components displaying the feedback on hover. Hence the hover over range bars is a must in order to use the feature while it's not the case with the pie charts. Inside the range bars the number of hovers on the step and move level feedback varied significantly. There was no pattern observed in the number of hovers across various drafts. It might be interesting to do further investigation on the hovers and its associated timings to see if any other usage pattern could be obtained.

For the user comments about the performance of the analysis engine, the number of thumbs up increased from group I to III as the accuracy of the engine was improved gradually from the time group I used the software. There is also a possibility of bias against thumbs up, as it indicates no problem in analysis, versus thumbs down, which the users might be interested more in because they are interested in improving or registering complaints about the system.

Overall, the percentage of hovers along with the results from the survey indicates that the users from group I and II found the features to be useful overall. The users from group II reported a better usability experience than the group I. Group III reported the lowest

values for questions on usability. All the three groups gave a positive response about the utility of the tool.

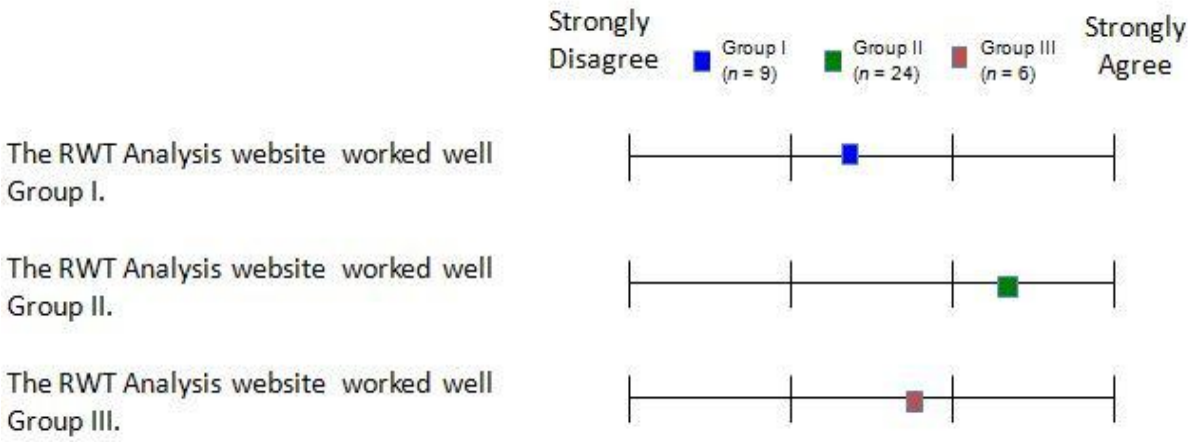


Figure 4.9: RWT's website performance

One other important topic for discussion is the comparison of technical difficulties faced by the three groups to see if that has an impact on the user reviews. From Figure 4.9, it could be seen that the group I had the lowest mean for the question "The RWT analysis module worked well". The results of ANOVA table shows that the responses of the three groups are statistically different with a very low p-value ($p<0.001$). The t-test indicates that group I and II are extremely statistically different (two-tailed p value less than 0.0001) and group II and III are statistically significant (two-tailed p value equals 0.0459) and group I and III are not statistically significant (two-tailed p value equals 0.2137). Group I had a few technical issues during the try out which could have affected the data regarding the usability and utility of the tool. The accuracy of evaluation was also improved from the time of the evaluation by the group I.

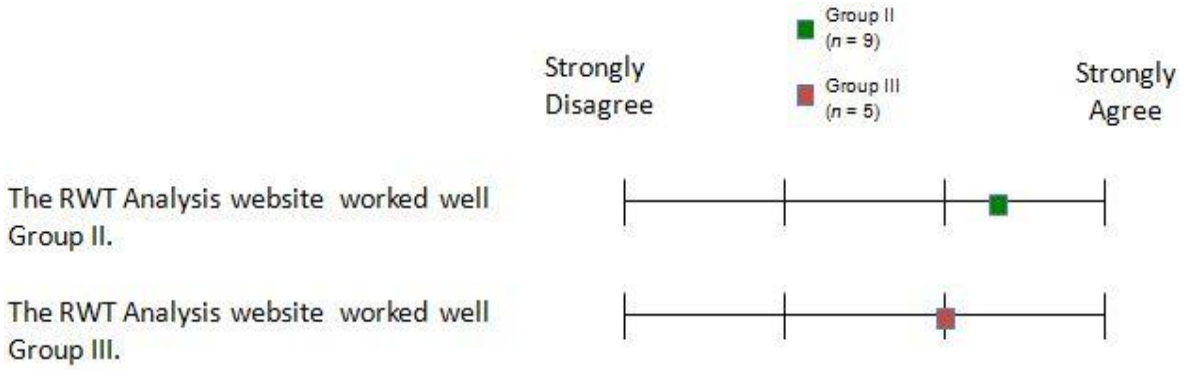


Figure 4.10: RWT's website performance (Demonstration Module)

There was no statistical difference between Group II and Group III (p=0.3266) in the response about the performance of the demonstration module website. Overall the group II gave better user reviews than the other two groups. Also both the groups had a positive response about the usability of the demonstration module than the analysis module. This could be because of the usage of the standard web convention design for displaying the search results unlike the analysis module that requires an initial learning curve.

As expected, the trust on automated systems had a huge impact on the user's review about the utility of RWT. It is clear that the group III users who, were never guided on automated systems, had no or less trust on automated systems. Group I and II had a higher trust on automated systems which could be because of their knowledge about the norms of academic writing and the capabilities of the tool via the demo shown by the instructor during the session. It could be beneficial to educate users on what automated tutors are capable of in order to use it in the best possible way, as trust is an important factor in systems like RWT.

Chapter 5 SUMMARY AND FUTURE WORKS

This chapter summarizes the research work done related to the questions raised in section 1.2, and proposes future work related to RWT and automated tutors.

In order to answer the research questions related to the usability and utility of the RWT software, three sets of user evaluation was performed. There were two primary modules of RWT which was evaluated: the analysis module which enables the users to enter research articles and receive feedback on their writing and the demonstration module which enables the users to search for discipline specific examples of a particular move and a step. To evaluate these two modules, a total of 39 participants were divided into three groups based on the training and practice provided with academic writing skills. Group I consisted of graduate students who were enrolled in one of the academic writing course at Iowa State University and received special training and practice sessions for improving academic writing skills. Group II consisted of graduate students who attended monthly seminars on academic writing, had the background understanding of academic writing, but lesser practice and training than group I. Group III consisted of graduate students who did not attend any course or seminar related to academic writing and used the software after viewing a demonstration video of what the tool is all about. Each participant in the group used the RWT to analyze their research articles. All the three user studies lasted for about an hour each. The results show that the group I and group II found the RWT to be more useful for their improving their academic writing skills than group III as expected. From the mean values obtained, Group II found the tool to be more user-friendly and useful than the other

two groups. Another question of research interest other than usability and utility of the RWT is about the trust on automated systems. The goal was to see how important trust is, in order to find an automated writing evaluation tool like RWT to be useful in guiding academic writings. Also we wanted to see if there is a difference in the level of trust on automated systems among these three groups. As expected, the results from group III show that there is totally no trust on automated systems. This could be because the users from group I and II have a background understanding of what RWT is capable of and how it could help them with academic writing. Users from group I had the maximum trust on automated systems of all the three groups. This results shows that the users would be better benefited by the usage of systems like RWT if they are educated about what these systems are capable of. There was also a question in the survey which was aimed at knowing if there were any technical difficulties faced with the website to see if it has an impact on the response to the questions related to the usability. The results indicate that group I had the lowest mean for the question about the working of the website. Group I gave lower response about the usability although they gave good reviews about the utility of the RWT site. This could be because of the technical problems faced by this group.

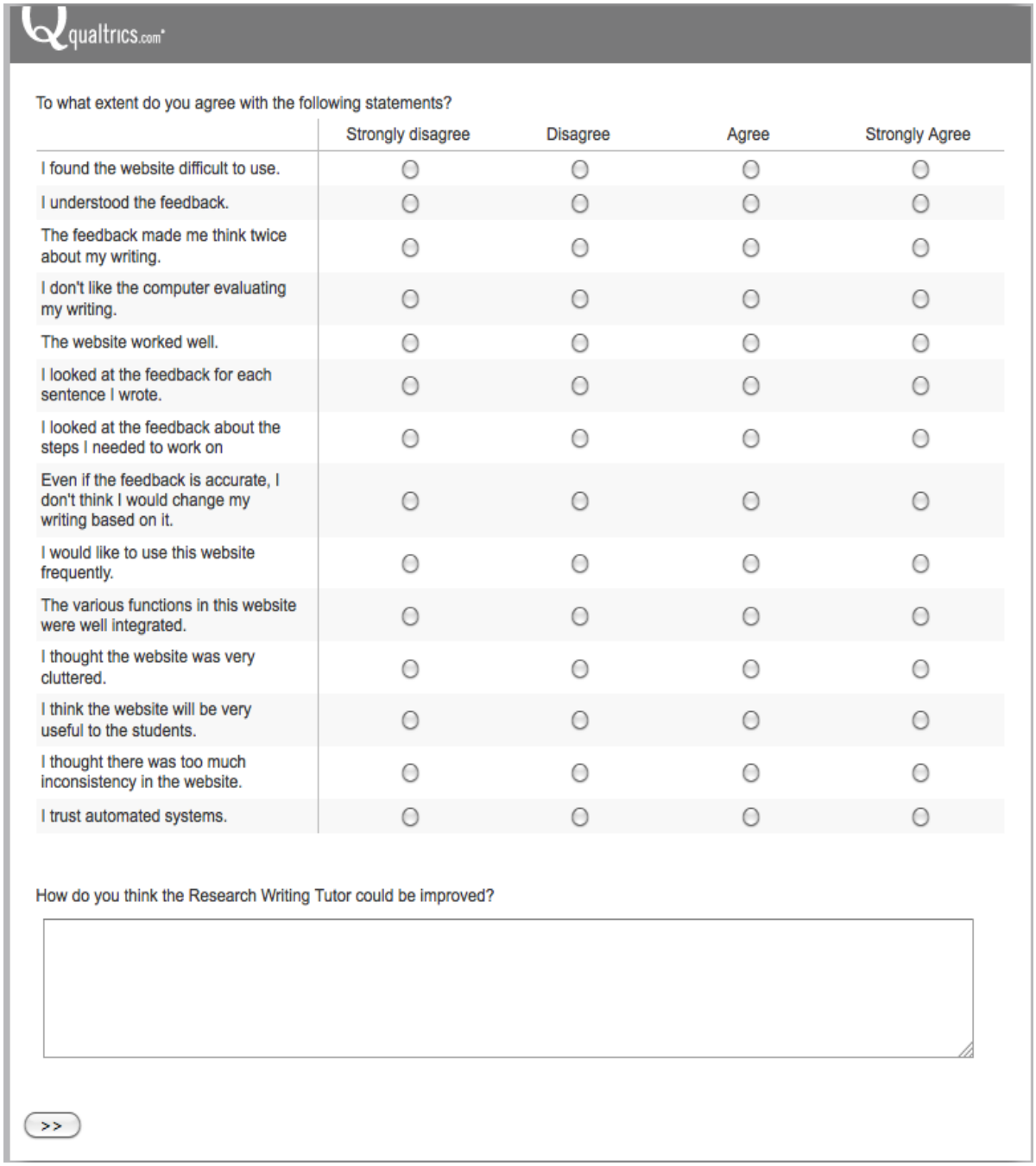
In addition to the survey, there was also data collected from the RWT database to see the usage amount of different features on screen in the analysis module. Overall results from all the three groups indicate that the range bars showing move and step feedback have been viewed by higher number of users than the pie charts in the bottom of the page. Within the range bars feedback for move 1 was hovered over higher number of times than the others

below it. The pie charts which are at the bottom most area of the page has been hovered the least as expected.

There is scope for further improvements in this tool. In terms of the UI, the editor for analyzing the drafts could be replaced with better text editors, additional features like support for uploading documents to the tool for analysis, saving or exporting writings and their feedback from the tool could be added. A few other recommendations that need design decisions for implementing are simplifying the way analyzed drafts are edited. Currently the "Edit" button needs to be clicked on to edit an analyzed draft which could be simplified. The users could also be provided with options to delete, rename or move a draft to another article helping to better organize the writings. Other tools like plagiarism check and grammar corrections could be provided optionally to the users. Also additional features like the one stated above could be given in an advanced tab and not in the home screen in order to maintain the analysis page simple. In terms of the analysis engine, the accuracy could be further improved and it would be interesting to add the potential to analyze the whole research article instead of analyzing it section by section.

APPENDIX A SURVEY QUESTIONS

A.1 Analysis Module



qualtrics.com

To what extent do you agree with the following statements?

	Strongly disagree	Disagree	Agree	Strongly Agree
I found the website difficult to use.	○	○	○	○
I understood the feedback.	○	○	○	○
The feedback made me think twice about my writing.	○	○	○	○
I don't like the computer evaluating my writing.	○	○	○	○
The website worked well.	○	○	○	○
I looked at the feedback for each sentence I wrote.	○	○	○	○
I looked at the feedback about the steps I needed to work on	○	○	○	○
Even if the feedback is accurate, I don't think I would change my writing based on it.	○	○	○	○
I would like to use this website frequently.	○	○	○	○
The various functions in this website were well integrated.	○	○	○	○
I thought the website was very cluttered.	○	○	○	○
I think the website will be very useful to the students.	○	○	○	○
I thought there was too much inconsistency in the website.	○	○	○	○
I trust automated systems.	○	○	○	○

How do you think the Research Writing Tutor could be improved?

>>

qualtrics.com

English

What do you think about the Analysis module of RWT?

	Strongly disagree	Disagree	Agree	Strongly Agree
I found the Analysis module webpage easy to use.	○	○	○	○
The RWT Analysis website worked well.	○	○	○	○
The various functions in this website were well integrated.	○	○	○	○
I thought the website was consistent.	○	○	○	○
The Analysis module was easy to navigate.	○	○	○	○
I thought the website was user friendly.	○	○	○	○
I looked at the feedback for each sentence I wrote.	○	○	○	○
I understood the sentence-level feedback.	○	○	○	○
I looked at the visual feedback (range bars and pie charts).	○	○	○	○
I understood the visual feedback.	○	○	○	○
I looked at the smiley faces and caution symbols associated with the range bars.	○	○	○	○
The smiley faces and caution symbols helped me understand what I needed to focus on or improve.	○	○	○	○
The feedback made me think twice about my writing.	○	○	○	○
Overall, the automated feedback was helpful.	○	○	○	○
I trust automated systems.	○	○	○	○
I like the computer evaluating my writing.	○	○	○	○
Reasonably accurate feedback would help me improve my writing.	○	○	○	○
I would like to use this website frequently.	○	○	○	○
I think the website will be very useful to students like me.	○	○	○	○

How do you think the Analysis module of RWT could be improved?

Did you use the Demonstration module (search page with examples from published articles) of RWT?

○ Yes

○ No

>>

A.2 Demonstration Module

Did you use the Demonstration module (search page with examples from published articles) of RWT?

- (●) Yes
- ( ) No

What do you think about the Demonstration module (search page with examples) of RWT?

	Strongly disagree	Disagree	Agree	Strongly Agree
I found the Demonstration module (search page with examples easy to use.	○	○	○	○
The RWT Demonstration website worked well.	○	○	○	○
The various functions in this website were well integrated.	○	○	○	○
I thought the website was consistent.	○	○	○	○
The Demonstration module was easy to navigate.	○	○	○	○
I thought the website was user friendly.	○	○	○	○
I looked at a lot of examples.	○	○	○	○
I understood what the examples represented.	○	○	○	○
Seeing a sentence before and after in the search results helps me understand the context.	○	○	○	○
I looked at the text the examples were extracted from (displayed next to a given example).	○	○	○	○
I searched for examples of different communicative functions.	○	○	○	○
I looked at more than one page of search results.	○	○	○	○
The examples helped me understand the communicative functions and how they are used.	○	○	○	○
The examples made me think twice about my writing.	○	○	○	○
Overall, the examples were helpful.	○	○	○	○
I like that the tool provides examples for each communicative function.	○	○	○	○
I would like to use this tool frequently.	○	○	○	○
I think the tool will be very useful to students like me.	○	○	○	○

How do you think the Demonstration Module of RWT could be improved?

>>

APPENDIX B USER

HANDOUTS

Reminder: Communicative goals and strategies in research article Introductions

- **Goal 1 (move 1)** — **Establish a knowledge territory**
  Strategies (steps)
  - Claiming centrality
  - Providing general background
  - Reviewing previous research
- **Goal 2 (move 2)** — **Identify a niche**
  Strategies (steps)
  - Indicating a gap
  - Highlighting a problem
  - Raising general questions
  - Proposing general hypotheses
  - Presenting justification
- **Goal 3 (move 3)** — **Address the niche**
  Strategies (steps)
  - Introducing present research descriptively
  - Announcing present research purposefully
  - Presenting research questions
  - Presenting research hypotheses
  - Clarifying definitions
  - Summarizing methods
  - Announcing principal outcomes
  - Stating the value of present research
  - Outlining the structure of the paper

The Research Writing Tutor (RWT) refers to communicative **goals as 'moves'** and to **strategies as 'steps'**. It also uses color codes to visually differentiate the **moves** for you so that you can see how the communicative goals are distributed in the text. Numbers 1, 2, and 3 for **moves** do not represent linear development of the communicative goals.

This practice with RWT is meant to help you think about your writing and analyze what you are doing communicatively in each of your sentences. You may realize that some of your sentences may need to be formulated clearer to make sure that the communicative strategies take effect and impact the reader as intended. If you are not conscious about your communicative goals and use of strategies, the overall argument you are building in your Introduction may not be very effective.

Evaluate your Introduction draft with the Research Writing Tutor (RWT)

To access the RWT website, open Mozilla Firefox and go to https://rwt.vrac.iastate.edu/index.php. Log in with your ISU net ID and password.

1. **Submit your draft for analysis**

Choose your discipline from the drop-down menu. If your discipline is not listed, choose one that is closest to your research area.

Paste your text in the textbox with the heading "Your article / Currently editing text".

Click on **ANALYZE**. Wait till your text is returned color-coded (blue for goal (move) 1, red for goal (move) 2, green for goal (move) 3). Colors and numbers are used for better discrimination of communicative functions.

2. **Use the RWT Analysis module for feedback to self-analyze and evaluate your draft**

When you receive the color-coded feedback, click on a color-coded sentence and check the feedback box below. Do you agree with the feedback prompt?

- If you agree, click on  .
- If you don't agree, click on  and write the communicative strategy (step) you had in mind in the Comments box below.
- If you partially agree (for example, the goal (move) is colored as you intended, but the strategy (step) you are trying to use is different; or, your sentence employs two or more strategies, but the feedback prompt contains only one; etc.) click on  and explain in the Comments box below what you agree with and what the feedback prompt is missing, clarifying what you are trying to communicate.

Also pay attention to the visual feedback (range bars and pie charts). They show a comparison of your draft with the Introductions of research articles published in top journals of the field, showing how close or how far you are from achieving each of the three communicative goals (moves).

If you hover below the range bars, you'll see feedback prompts specifying what you are doing well (with a smiley face ) and what you need to focus on more (with a caution symbol ). You will also see two links: "Learn more" and "Examples." Click on "Learn more" and you will see a brief description of a particular communicative strategy in a pop-up. Click on "Examples", and you'll be taken to the Demonstration module of the RWT to see a selection of all the examples of a particular communicative strategy (step) extracted from published Introduction sections in your field.
As you interact with RWT, you might notice that some parts of your draft may need to be

improved – either at the level of a particular communicative strategy (step) or at a broader level of a certain communicative goal (move). Try to revise in order to make your draft better and more effective. For that, click on **EDIT**, make the changes you want, and then re-submit your modified draft for automated analysis and feedback by clicking on **ANALYZE**.

3. **Use the RWT Demonstration module for examples to guide the revision of your draft**

The Demonstration module of the RWT is a search engine, where you can search for examples of all the communicative goals and strategies (moves and steps) used in research article Introductions published in your field. Search for different steps to see how frequently they are used in your discipline (see the number of pages; e.g., 1 2 3 4 Next ->)

In addition to the list of examples, you can click on the arrows next to each example (>) and you will see the Introduction from which the example was extracted. The text is color-coded just like your Introduction draft in the Analysis module of RWT, and it is also tagged with a move and step, which you can see when you click on a sentence. Look at the color-coded Introductions to see how the communicative goals and strategies (moves and steps) are generally distributed in your field.

The RWT tool is in the earliest stage of its development. Please go to this link http://vrac.us2.qualtrics.com/SE/?SID=SV_bI6x2W38RkRkyNu and answer a few questions. Your input will be extremely useful and informative for the developers.

Thank you!

ISU IRB # 1 11-479
Approved Date: 13 February 2012
Expiration Date: 19 October 2012

CONSENT FORM FOR: RESEARCH WRITING TUTOR (RWT)

This form describes a research project. It has information to help you decide whether or not you wish to participate. Research studies include only people who choose to take part—your participation is completely voluntary. Please discuss any questions you have about the study or about this form with the project staff before deciding to participate.

Who is conducting this study?

This study is being conducted by Stephen B. Gilbert, Elena Cotos.

Why am I invited to participate in this study?

You are being asked to take part in this study because you are enrolled in one of the academic writing courses. You should not participate if you are under the age of 18.

What is the purpose of this study?

The purpose of this study is to discover whether or not automated writing interventions are helpful to students and to learn more about how students perceive the software system.

What will I be asked to do?

If you agree to participate, you will be asked to complete your research paper writing assignments within the RWT web site to get feedback on your writing assignments. The number of clicks, time spent on the site, areas of the web site you visit, and number of adjustments to your writing after tutoring tips will be recorded in a log file. This will take place over the course of a month and following a month of use at the end of which you will be asked to complete a short exit survey asking them to assess the system's ease of use and helpfulness.

Your participation will last for one month.

What are the possible risks and benefits of my participation?

Risks—The possible risks related to your participation in this research are no greater than your daily use of computational technologies.

Benefits—You may not receive any direct benefit from taking part in this study. We hope that this research will benefit society by a better understanding of how software interfaces should be designed as learning aids should benefit society by providing information relevant to the development of new computer interfaces that are easier to use and more effective at helping students learn to write research papers.

How will the information I provide be used?

ISU IRB # 1 11-479
Approved Date: 13 February 2012
Expiration Date: 19 October 2012

The information you provide will be used for the following purposes: the data will be analyzed in order to improve the Research Writing Tutor application and analyzed in aggregate form (nobody will be able to identify you individually) for publication in international conferences and journals.

What measures will be taken to ensure the confidentiality of the data or to protect my privacy?

Records identifying participants will be kept confidential to the extent allowed by applicable laws and regulations. Records will not be made publicly available. However, federal government regulatory agencies, auditing departments of Iowa State University, and the ISU Institutional Review Board (a committee that reviews and approves research studies with human subjects) may inspect and/or copy your records for quality assurance and analysis. These records may contain private information.

To ensure confidentiality to the extent allowed by law, the following measures will be taken: Participant confidentiality will be ensured by not recording names of participants on any data gathering information. All data collected will be stored in a locked lab on a password-protected computer (digital data) or in a locked filing cabinet for non-digital data. Informed consent documents will be kept in a locked filing cabinet in an access-controlled lab. If the results are published, your identity will remain confidential.

Will I incur any costs from participating or will I be compensated?

You will not have any costs from participating in this study. You will not be compensated for participating in this study.

What are my rights as a human research participant?

Participating in this study is completely voluntary. You may choose not to take part in the study or to stop participating at any time, for any reason, without penalty or negative consequences. When completing the survey, you can skip any questions that you do not wish to answer. Your choice of whether to participate or not will have no impact on you as a student or your grade in the course where you have the option of using RWT.

Whom can I call if I have questions or problems?

You are encouraged to ask questions at any time during this study.

- For further information about the study contact Stephen B. Gilbert (gilbert@iastate.edu), 515.294.6782.
- If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, (515) 294-3115, Office for Responsible Research, 1138 Pearson Hall, Iowa State University, Ames, Iowa 50011.

ISU IRB # 1 11-479
Approved Date: 13 February 2012
Expiration Date: 19 October 2012

Consent and Authorization Provisions

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent prior to your participation in the study.

Participant's Name (printed) _____

_____ _____
(Participant's Signature) (Date)

BIBLIOGRAPHY

Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds.), Teaching and learning by doing corpus analysis (pp. 9–25). Amsterdam: Rodopi.

Attali, Y. (2004). Exploring the feedback and revision features of the Criterion service. Paper presented at the National Council on Measurement in Education Annual Meeting, San Diego, CA. Retrieved on November 12, 2007 at http://www.ets.org/Media/Research/pdf/erater_NCME_2004_Attali_B.pdf

Baron, D. (2005, May 6). The College Board's new essay reverses decades of progress toward literacy. The Chronicle of Higher Education, P. B14.

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), Rethinking language pedagogy from a corpus perspective (pp. 207–223). Frankfurt: Peter Lang.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), Teaching and learning by doing corpus analysis (pp. 165–182). Amsterdam: Rodopi.

Brock, M. (1990). Customizing a computerized text analyzer for ESL writers: Cost versus gain. CALICO Journal, 8 (2), 51-60.

Brock, M. (1993). Three disk-based text analyzers and the ESL writer. Journal of Second Language Writing, 2 (1), 19-40.

Brown, S., & Knight, P. (1994). Assessing Learners in Higher Education. Kogan Page, London.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online Writing Evaluation service. AI Magazine, 25(3), 27-36.

Burstein, J., Leacock, C., & Swartz, R. (2001). Automated evaluation of essays and short answers. Proceedings of the 5th International Computer Assisted Assessment Conference (CAA 01), Loughborough University

Burstein, J., Kukich, K., Wolff, S., Lu, C., & Chodorow, M. (1998, April). Computer analysis of essays. Paper presented at the annual meeting of the National Council of Measurement in Education, San Diego, CA. Retrieved September 10, 2005, from http://www.ets.org/research/dload/ncmefinal.

Calfee, R. (2000). To grade or not to grade. IEEE Intelligent Systems 15(5), 35–37.

Calvo, R. (2010). Students' conceptions of tutor and automated feedback in professional writing. *Journal of Engineering Education*.

Charman, D. & Elmes, A. (1998) A computer-based formative assessment strategy for a basic statistics module in Geography, Journal of Geography in Higher Education, 22(3), pp. 381-385.

Chen, C. F, & Cheng, W. Y., (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. Language Learning and Technology, 12(2), 94-112.

Chen, Mon Chu, Anderson, John R., & Sohn, Myeong Ho (2001). What can a mouse cursor tell us more?:correlation of eye/mouse movements on web browsing. *CHI '01*.

Chodorow, M. and Leacock, C. (2000). An unsupervised method for detecting grammatical errors. In Proc. of 1st Meeting of the North America Chapter of ACL, pages 140-147.

comScore, Inc. (2012). For Display Ads, Being Seen Matters More than Being Clicked, Press Release SAN FRANCISCO and RESTON, VA, April 24, 2012.

Cotos, E. (2010). Automated Writing Evaluation for non-native speaker English academic writing : The case of IADE and its formative feedback.

Elliot, S. 2003: IntelliMetric: from here to validity. In Burstein, J.C., editor, Automated essay scoring: a cross-disciplinary perspective. Lawrence Erlbaum, 71–86.

Elliot, S. and Mikulas, C. 2004: The impact of MY Access! use on student writing performance: a technology overview and four studies. Paper presented at the annual meeting of the American Educational Research Association, April 2004, San Diego, CA.

Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. TESOL Quarterly, 29, 33-53.

Ferris, D. (2003). Response to student writing: Implications for second language students. Mahwah, NJ: Lawrence Erlbaum.

Fitzgerald, K.R. (1994). Computerized scoring? A question of theory and practice. Journal of Basic Writing, 13(2), 3–17.

Flowerdew, L. (1998). Concordancing on an expert and learner corpus for ESP. CALL Journal, 8 (3), 3–7.

Flowerdew, J. (1999). Problems in Writing for Scholarly Publication in English : The Case of Hong Kong, *264*(3), 243-264.

Grimes, D. & Mark Warschauer, (2006). Automated Essay Scoring in the Classroom. Paper presented at the AERA, San Francisco.

Huang, L.-S. (2010). Seeing eye to eye? The academic writing needs of graduate and Undergraduate students from students' and instructors' perspectives. *Language Teaching Research*, *14*(4), 517-539.

Hyland, K. (2000). Disciplinary discourses: social interactions in academic writing. Harlow, Essex: Longman.

Hyland, K. (2003). Second language writing. Cambridge: Cambridge University Press

Hyland, K., & Hyland, F. (Eds.). (2006). Feedback in Second Language Writing: Contexts and issues. New York: Cambridge University Press.

Kettemann, B., & G. Marko (Ed). (2002). Teaching and learning by doing corpus analysis. Amsterdam: Rodopi.

Krug, S. *Don't Make Me Thin, A Common Sense Approach to Web Usability.* New Riders, 2005 (2nd edition) 216 pages

Kukich, K. (2000, September/October). Beyond Automated Essay Scoring. In M. A. Hearst (Ed.), The debate on automated essay grading. IEEE Intelligent systems, 27–31. Retrieved November 12, 2004, from http://que.info-science.uiowa.edu/~light/research/mypapers/autoGradingIEEE.pdf

Land, R.E., & Evans, S. (1987). What our students taught us about paper marking. English Journal, 76, 113-

Leah Rock, J. (2007). The impact of short-term use of CriterionSM on writing skills in ninth grade. Research report. Educational Testing Service, Princeton, NJ. Retrieved on December 8, 2008 from http://www.ets.org/Media/Research/pdf/RR-07-07.pdf

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. English for Specific Purposes, 25, 56–75.

Llosa, L., Beck, S. W., & Zhao, C. G. (2011). An investigation of academic writing in secondary schools to inform the development of diagnostic classroom assessments. *Assessing Writing*, *16*(4), 256-273. Elsevier Inc.

McNamara, D., Crossley, S., & McCarthy, P. (2010). The linguistic features of quality writing. Written Communication, 27, 57-86.

Mcnamara, D. S., Raine, R., Roscoe, R., Crossley, S., Jackson, G. T., Dai, J., Cai, Z., et al. (2001). The Writing-Pal : Natural Language Algorithms to Support Intelligent Tutoring on Writing Strategies.

Myers, M. (2003). What can computers and AES contribute to a K–12 writing program? In M. D. Shermis & J. Burstein (Eds.), Automated essay scoring: A cross-disciplinary perspective (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates.

Nielsen, J. *Usability 101: Introduction to Usability - Top Ten Mistakes in WebDesign.* 1995

Page, E.B. (1966). The imminence of grading essays by computers. Phi Delta Kappan, 47, 238–243.

Page, E. B. (1994). Computer Grading of Student Prose, Using Modern Concepts and Software, Journal of Experimental Education, 62, 127–142.

Roscoe, R., Varner, L., Cai, Z., & Weston, J. (2011). Internal Usability Testing of Automated Essay Feedback in an Intelligent Writing Tutor. *Twenty-Fourth*, 543-548.

Sambell, K., Sambell, A., Sexton, G., 1999. 'Student perception of the learning benefits of computer-assisted assessment: A case study in electronic engineering' in Brown, S., Race, P., Bull, J. (Eds.), Computer Assisted Assessment in higher education. Kogan Page, Birmingham, pp. 179-191.

Scharber, C., Dexter, S., 2008. Students' Experiences with an Automated Essay Scorer, The Journal of Technology, Learning and Assessment, 2008, volume 7, issue 1

Shermis, M., & Burstein, J. (2003). Introduction. In M. D. Shermis & J. C. Burstein (Eds.), Automated essay scoring: A cross-disciplinary perspective. (pp. xiii-xvi). Mahwah, NJ: Lawrence Erlbaum.

Shermis, M. D., Burstein, J., & Leacock, C. (2006). Applications of computers in assessment and analysis of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (pp. 403-416). New York: Guilford.

Shute, V. (2008). Focus on formative feedback. Review of Educational Research, 78, 153-189.

Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. Research in the Teaching of English, 31, 91-119.

Suite, T. D. (2007). Vantage Learning, (September).

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, *21*(2), 217-235.

Wade-stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. Language Teaching Research, 10(2), 1–24.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? British Educational Research Journal, 22, 537-548.

Wiemer-hastings, P., Graesser, A. C., & Wiemer-hastings, P. (1999). Select-a-Kibitzer : A computer tool that gives meaningful feedback on student compositions. Interactive Learning Environments 8 (2): 149–69.

Wiemer-hastings, P., Graesser, A. C., & Wiemer-hastings, P. (1999). Select-a-Kibitzer : A

computer tool that gives meaningful feedback on student compositions. Interactive Learning Environments 8 (2): 149–69.

Yang, Y., Buckendahl, C., Juszkiewicz, P., & Bhola, D. (2002). A review of strategies for validating computer automated scoring. *Applied Measurement in Education, 15*(4), 391-412.

Yuan, Yi. (2003). The use of chat rooms in an ESL setting. Computers and Composition, 20.2. 194-206.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those who helped me at various stages of my research work and with the writing of this thesis. Firstly, I would like to thank my major professor, Dr. Stephen Gilbert for his guidance and support throughout this research and for being a wonderful mentor. He helped me learn about Automated Writing Evaluators and Intelligent Tutoring Systems by providing me an opportunity to work in various exciting projects. I owe a lot to Dr. Elena Cotos who provided great insights and feedback throughout the development of the RWT. Her suggestions and feedback were invaluable. I would also like to thank Dr. Jin Tian for his guidance, encouragement, and for being very patient with me. I would like to thank my peers Sarah Huffman, Deepan Babu, Vijay Kalivarapu, Andrew Vernon and Ryan Kirk from the RWT team for their help in various stages of the study. Finally, I would like to thank my friends who provided valuable assistance while completing my research work: Karthik Babu, Rama Srinivasan and Shrenik Devasani.

The research presented in this thesis was supported by the Computer Advisory Committee at Iowa State University.