ADDIS ABABA UNIVERSITY

SCHOOL OF GRDUATE STUDIES

AUTOMATIC AMHARIC ESSAY SCORING SYSTEM USING LATENT SEMANTIC ANALYSIS

By: Abel Teklemariam

A THESIS SUBMITTED TO

THE SCHOOL OF GRADUATE STUDIES OF THE ADDIS ABABA UNIVERSITY

IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTERS OF SCIENCE IN COMPUTER SCIENCE

November 2010

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
FACULTY OF COMPUTER AND MATHEMATIAL SCIENCES
DEPARTMENT OF COMPUTER SCIENCE

AUTOMATIC AMHARIC ESSAY SCORING SYSTEM USING LATENT SEMANTIC ANALYSIS

By: Abel Teklemariam

ADVISOR:

Sebsbie Hailemariam (PhD)

APPROVED BY

EXAMINING BOARD:

1. Dr. Sebsbie Hailemariam, (Advisor) _____
2. _____ _____
3. _____ _____

Dedication

Dedicated to my **Mom** & $\boldsymbol{M^2}$!!!

Thanks **God** for giving me the two priceless and precious gifts' that makes my life blissful.

ACKNOWLEDGMENTS

First and foremost I would like to extend my heartfelt thanks to my advisor Dr. Sebsibie Hailemariam for his wonderful and unreserved assistance in every step of this study. I would also like to show my appreciation to Addis Ababa University, Faculty of Informatics, Department of Computer Science post graduate programme for providing the opportunity to conduct this research and acquire a lot of knowledge in conducting the research. It gives me great pleasure to extend many thanks to Wolaita Sodo University for being in no doubt to offer me the grant and finally get the product.

Dr. Tuomo Kakkonen from University of Eastern Finland, without your support I might not have been here to welcome your well done cooperation and massive support in challenges that I had faced. Melese Tamru, my role model, you are very much thanked for reviewing my document and providing useful comments.

For those kindly collaborating staffs of AAU; Ato Derege Haile for preparing the essay questions and W/ro Sinknesh Atale in line with your valuable time, I would like to forward my special thanks. My deep gratitude also goes to the management and staff of Kotebe College and Menelik II secondary and preparatory school for playing a leading and remarkable role in facilitating the exam. All the students who were willing to be part of the study, I really thank you without whom this paper would not have been completed. I would also like to pass my gratefulness to Wondmu Tegengn and Mr Tesema from Akaki campus for your enormous participation to evaluate the student paper round with the whole difficulties.

My kind gratitude is also extended to Computer Science Department secretaries Haymanot and Alemnesh, Kotebe College staff members: Fantanesh, who is head of Linguistic Department, and Ato Yidnekachew from English Department. My friends Teklay, Selama, Moges, Tesfish, Getasew, Mekuanint, Desta, and others whose names I couldn't list down here are also warmly thanked. My acknowledgement is also extended to my brother Abraham who was by my side in many aspects, and my friend Habtamu for his support during the research process.

My friends, what can I say? Thanks a lot; you really deserve to be first-class friends. And my family, and my love, your encouragement during my troubles and continuous support were my engine!!!

Table of Contents

Contents **Page No**

LIST OF TABLES

Tables **Page No**

LIST OF FIGURES

Figures **Page No**

LIST OF ANNEXES

Annexes **Page No**

LIST OF ACRONYMS

AAESS	Automatic Amharic Essay Scoring System
ACES	Automatic Chinese Essay Scoring
AEA	Automatic Essay Assessor
AEE	Automated Essay Evaluator
AES	Automatic Essay Scoring
AESS	Automatic Essay Scoring System
BETSY	Bayesian Essay Text Scoring System
ETS	Education Testing Service
E-rater	Electronic Essay Rater
GMAT	General Management Aptitude Test
GRE	Graduate Record Examination
IntelliMetric	Intelligent Metric System
IEA	Intelligent Essay Assessor
IEMS	Intelligent Essay Marking System
IR	Information Retrieval
JESS	Japan Essay Scoring System
KAT	Knowledge Analysis Technology
LSA	Latent Semantic Analysis
LSI	Latent Semantic Indexing
MBM	Multivariate Bernoulli Model
NLP	Natural Language Processing
PEG	Project Essay Grader
PS-ME	School free Text Marking Engine
SAT	Systematic Aptitude Test
SVD	Singular Value Decomposition
TF–IDF	Term Frequency–Inverse Document Frequency
TOEFL	Test of English as a Foreign Language
VSM	Vector Space Model

ABSTRACT

Essay assessment is considered to play a central role in educational process as essays are the most useful tool to assess learning outcomes. Consequently, essays have been incorporated into many of the standard testing programs like SAT (Systematic Aptitude Test), GRE (Graduate Record Examination), TOFEL (Test of English as a Foreign Language) and GMAT (General Management Aptitude Test).

Though the importance of essay assessment is elevated, the process of assessment using human evaluators is extremely labor intensive and time consuming. Hence, automatic essay scoring systems are developed to overcome time, cost, and generalizability issues in manual essay assessment.

Currently, a number of automatic essay scoring systems using different techniques are available commercially or as a result of research in the field. PEG (Project Essay Grader), E-rater (Electronic Essay rater) and IEA (Intelligent Essay Assessor) are among the most common commercially available automatic essay scoring systems for English language but efforts are also made for other languages as well JESS (Japan Essay Scoring System) and AEA (Automatic Essay Assessor for Finnish) to mention some.

This study is an attempt to develop similar system for Amharic language, the working language of Federal Democratic Republic of Ethiopia, to factual types of essays. The study used Latent Semantic Analysis which is an information retrieval technique to develop the model. LSA is a novel application used to evaluate essay based on the extent to which an essay can be matched against other essays scored by human raters. To achieve this large number of pre-graded essay corpus in three domains are prepared from different educational institutions and used for developing the model and conducting the experiment.

The research conducted a detail set of experiments to measure the performance of the proposed system using the percentage of adjacent agreements between the system score and human score. The result of the experiment varies with the domains involved and found to be 62%, 59% and 52% agreement in three domains which is considered very promising being the first attempt and paves a way for other researchers to participate in automatic essay scoring system.

Keywords: Essay Assessment, Automatic Amharic Essay Scoring, Latent Semantic Analysis

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

The importance of practice in the development of good writing skills has a vital role. Like playing a musical instrument, writing is something that cannot be taught by directions or examples alone. Practicing writing and receiving constructive feedback of students is an integral part of the learning process [1].

In addition, the quality of an essay is regarded as one of the best measures of the student's knowledge of the topic. Writing essays requires more thought than many other forms of testing, such as multiple-choice, matching and short answer. This is because the students must construct their own coherent answers and justifications. Therefore, well-developed and appropriately scored writing assessments can test not only students' skill of language, but also their ability to synthesize and analyze information; to find new connections between ideas and to explain their significance [1].

With regard to this, essay assessment is considered to play a central role in educational process. Essay is considered by many researchers as the most useful tool to assess learning outcomes, implying the ability to recall, organize and integrate ideas, the ability to express oneself in writing and the ability to supply merely than identified interpretations and application of data. It is in the measurement of such outcomes that essay evaluation serves their most useful purpose [2].

Unfortunately, assessing student writing and providing thoughtful feedback is extremely labor-intensive and time-consuming. As a result, instructors are often faced with the difficult decision of assigning fewer writing assignments or marking them less thoroughly. Moreover, essay assessment is a complex and fundamental subjective process. Several findings reported that correlation in grade attribution between two human assessors is around 0.6 to 0.8 [3]. So backing up the human assessment process with automated essay scoring mechanisms is an excellent option to increase both effectiveness and efficiency in the assessment process.

Until recently, little thought was given to the idea of automating the essay evaluation and scoring process especially in Ethiopian national working language, Amharic; as to our knowledge, it has never been tried even if the importance is highly elevated [1].

Computer assisted assessment refers to the use of cognitive power of computers in evaluating students learning outcomes. The need for computer assisted assessment of learning outcome has three main purposes. First, Teachers need to automate the assessment and evaluation process especially in mass courses and having large number of students in each course. Secondly, a student wants to get feedback and assess his or her own learning process before an examination. Thirdly, educational institutions will save a substantial amount of cost incurred by assessment process [4].

Currently essay scoring system is hot area of research and gaining popularity by implementing Natural Language Processing and Machine Learning Techniques. Some of the currently deployed automated essay scoring system including Electronic Essay Rater (E-rater), Intelligent Metric System (IntelliMetric) and the Intelligent Essay Assessor (IEA) have shown to be successful and many standard international exams like General Management Aptitude Test (GMAT), Test of English as a Foreign Language (TOEFL) and others have started integrating them [2]. It has also been developed in other language for example Japan Essay Scoring System (JESS) [5], Automatic Chinese Essay Scoring (ACES) [6] and so on.

It can be understood from the literature that Automatic Essay Scoring System is well-liked and becoming essential tool and it is armed with the natural language processing. However, in the context of Ethiopia, no attempt has been made to develop similar system in any of the local languages including Amharic. Amharic Language is a Semitic language spoken in many parts of Ethiopia and it is the official working language of the Federal Democratic Republic of Ethiopia and thus has large number of speakers nationwide [7]

Current research in Amharic Natural Language Processing (NLP) covers different aspects of the language such as morphological Analysis, syntax and speech recognition and so on. This is very promising, but these researches mainly focused on lower level of NLP. Though morphological analysis is often considered as the first phase of a more complex NLP application, a significant research needs to be done in other areas like **Automatic Essay Scoring Systems** in which syntax, semantic and discourse will be involved in order to make the development of viable NLP applications possible [8].

Moreover, Automatic Essay Scoring System is being extensively researched in English and other languages and has shown good results [2]. But there is no attempt done in Amharic despite the

aforementioned benefits, so it's important to develop Automatic Essay Scoring System for Amharic Language.

Hence, this research is an attempt to study the constructs of Automatic Essay Scoring and to investigate the problem, analyze the cons/pros of different approaches implemented and try to develop the Amharic Essay Scoring System for factual types of essays.

This research focuses on factual essays, which are types of essays based on hard facts not on individual perception or opinions, For example essays on topics "ስለ አደንዛዥ ዕፆች ምንነትና ስርጭት"፣ "የአፄ ቴዎድሮስ የመቅደላ ታሪክ" and other historical and scientific events can be considered as factual essay. Whereas essays on "ለጉብኝት የሚመጡ የውጭ አገር ነዋሪዎች (ቱሪስቶች) ለአንድ አገር ያላቸው ሚና ምን ይመስላል?" ፣ "ስለ አትዮጵያ የትምህርት እድገት ምን አስተያየት አለህ/አለሽ? " is not treated as factual rather it is open for personal opinions and the writer can write anything which he/she feels comfortable with [9]. This study therefore focuses to develop Amharic Essay Evaluation and Scoring system for only factual type of essays.

1.2. Statement of the Problem

Different authors agree that writing skill is developed through constant practice. Thus, revision and feedback are essential aspects of essay writing process. Students need to receive feedback in order to improve their writing ability. However, responding to student's paper can be a burden and also time consuming for a teacher particularly if they have large number of students in a class.

In addition, essay assessment is an inherently subjective process when carried out manually. Evaluators read, analyze and interpret the essays to be scored with different rate of errors and subjective differences. For instance, two teachers may not give the same grade to the same essay and also the same teacher may not give the same grade to the same essay on different occasions. Choosing representative and trained evaluators can circumvent this only to a limited amount. Empirical evidence for this can be found in the usually rather low inter rater correlation of two human assessors which typically floats around 0.6 to 0.8 [3]. Supporting the human assessment process with automated essay scoring mechanisms is an excellent option to increase both effectiveness and efficiency in the assessment process.

In the case of Amharic language, despite the increasing number of electronic and non electronic document, there is no attempt done in developing essay evaluation system so far. In this study, an attempt is made to investigate the problem, analyze the cons/pros of different approaches

implemented and try to develop the Amharic Essay Scoring System for factual types of essays. Factual essay are types of essays which are based on facts not just personal opinion.

1.3. Objectives of the Study

1.3.1. General Objective

The general objective of this research work is to develop Essay Evaluation and scoring model for essays written in Amharic on factual topics.

1.3.2. Specific Objectives

In order to achieve the above general objective, the research will accomplish the following specific objectives;

- Study language specific features of Amharic Essays
- Prepare Amharic essay exams on different domains
- Conduct the Amharic essay examination on large number of students in different educational institutions
- Prepare specific essay scoring criteria and evaluate the Amharic essay by experts for training and testing of the system
- Assess the various approaches for the development of Automatic Essay Scoring System
- Design and model Automatic Amharic Essay Scoring System (AAESS).
- Develop a prototype for AAESS.
- Test the performance of AAESS in different domains
- Conduct analysis of findings.

1.4. Scope and Limitation of the Study

Naturally, Automatic Essay Scoring (AES) is a very complex and rigorous task which needs understanding of natural language. A full-fledged AES system will require a number of Natural language processing tools such as various revisions and editing tools like spell checker, grammar checker, dictionary and thesaurus. Even though some NLP tools have been done by some researchers, they are not publicly available for integrating with our new system. Having these limitations in mind, the scope of the research is as follows:

a. Essays are factual which are based on hard facts
b. Essay Evaluation focus only on the content of the essay rather than grammar, style, and mechanics.

c. AAESS is domain specific
d. Only text without figures, tables and charts is assessed.

1.5. Methodology

In order to achieve the objective of the research, the following methods are employed.

1.5.1. Literature Review

Extensive literature review is performed to get a deeper understanding of automatic essay evaluation and scoring system with due emphasis on the use of Latent Semantic Analysis.

1.5.2. Data Collection

For the successful development of AESS large number data is required for the training and testing of the system. Having this in mind, the research prepared large number of Amharic pre-graded essays in different domains; the following sections discuss the process in detail.

Essay Exam Preparation

Amharic essays are prepared with the cooperation of Amharic department in Addis Ababa University from three different domains that helped us to measure the performance of the system from different domain perspectives. Topics of the essay are selected carefully by giving due concentration on importance, relevance and harmony with the lesson plan of the teaching material of the Amharic department.

Essay Exam Administration

After the exam is prepared, it is administered to more than 268 students in three different institutions: Addis Ababa University, Kotebe College of Education and Dagmawi Mennelik II Secondary School.

Manual Essay Exam Scoring

A specific scoring criterion for each domain is set by experts in Amharic department and based on the criteria the exam is scored by human raters from Addis Ababa University, Akaki campus.

Essay Exam Encoding

After the essay exam is scored, it is encoded using visual Geez Unicode or power geez Unicode and organized so that it can be used for both training and testing purposes.

1.5.3. Development of Tools

In order to successfully achieve the research objectives, a number of tools are used. Java programming language is used for preprocessing the input document. Java is selected for its suitability in developing standalone applications and also it support for Unicode encoding. Besides most of the previous work on preprocessing are available on java, so it is easy to integrate the work of others. The research also used python 3.10 which is an interpreter, object oriented, and high level programming language with dynamic semantics moreover it support for Unicode encoding. Python is used for generating the preprocessed terms by essay matrix. At last MATLAB 7.1 which is powerful matrix processing software is used to decompose the matrix into singular values and semantic representation of the document so that document evaluation and essay grading can be performed.

1.6. Applications of the Result

Amharic Essay Scoring System aims to evaluate the content of Amharic essays. It will save substantial amount of time and energy for instructors; and also saves a lot of cost incurred for evaluation by educational institutions. Overall, students will also get a lot from the system as it allows them to practice writing more essays and develop their writing skills. After the successful completion of the system, it will have the following applications:

- Amharic Essay Scoring System offers students the opportunity to prepare and practice essays, submit the essays to be evaluated, and retrieve an evaluation at the student's convenience. The essays are evaluated in a manner that provides useful instructional feedback to students about their skills so that students can improve their writing skills.
- Amharic Essay Scoring System is intended to be an aid, not a replacement, for classroom instruction. Its purpose is to ease the instructor’s load, thereby enabling the instructor to give students more practice writing essays.
- As plagiarism is becoming a serious problem in education, this study paves a way for developing automatic plagiarism detection in Amharic for student essays.
- It can be used to facilitate admission tests for entrance examination in educational institutions.
- This study can be used as an input in developing a full-fledged Amharic Essay Evaluation system.
- Facilitate standard way of evaluation on essay type topics.

1.7. Organization of the Thesis

Chapter 1 provides general introduction and introduces the main concern of the study. Chapter 2 discusses about educational assessment and different types of assessment then it will direct to automating this task. The chapter explains background information about Automatic Essay Evaluation System, criteria on essay evaluation and various approaches to automatic essay scoring system.

Chapter 3 critically reviews related work on Automatic essay scoring system for different languages such as English, Finnish and Japanese. The review focuses on data set, approach and performance of the system. It also indicates relevant method to that of Amharic language. Chapter 4 presents our proposed approaches to Automatic Essay Scoring System and describes the architecture of the proposed system along with the implementation issues. Chapter 5 presents the empirical results of the proposed system along with their interpretations. Finally, Chapter 6 concludes the thesis with the research findings, conclusions and future works.

CHAPTER TWO

LITERATURE REVIEW

In this chapter, a brief overview of the field of Automatic essay scoring is explained. It begins by introducing the broader topic educational assessment, and among the various types of assessment which are considered to be relevant to the research. Essay, types of essay, essay evaluation criteria and method of essay evaluation are among the topics described in this chapter.

A number of research papers are reviewed to discuss background information related to essay evaluation is addressed in this chapter. Furthermore, the research investigates state of the art techniques used in the area of essay evaluation. Automatic essay scoring is a hot topic of research and hence, there are a lot of works available, but this research only present those works whose contribution made a great progress to the automatic essay scoring field.

2.1. Educational Assessment

Educational Assessment is the broad topic consists of three stake holders' student, teacher and educational institution or organization. It can be defined as the process of documenting usually in measurable terms, knowledge, skills, attitudes and beliefs [10, 11]. Assessment is an important component of educational process that has to be seen in a wider context of educational goals, course design and student motivation. Education in general and specifically higher education should prepare students for life-long and autonomous learning. To achieve this Assessment plays a central role in education and should facilitate this goal. Assessment must be designed in the context of the teaching and learning strategy of a course. There must be a commitment to a pedagogical rationale for adoption of assessment design. In addition, assessment should be student-centered to enable meaningful and relevant learning experiences [12, 13, 14].

Meaningful and constructive assessments need to make students to think critically and should encourage students' interest in learning. As it is widely acknowledged, in the research that assessment drives student learning and directs student effort, assessment design must be planned accordingly and must be an integral part of course design. Furthermore, assessment tasks influence the direction and quality of student learning. Therefore, to move forward students need to be given more responsibility for assessment processes and must be encouraged to participate in this task [15, 16, 17].

Assessment has received attention as many educators, researchers and philosophers have claimed that it is essential to learning. The following selected quotations will support this idea.

"*What has in a question? Everything. It is the way to evoke the stimulating answer or to annihilate the investigation. It is, in essence, the same marrow of education*" [18]

"Assessment is a crucial part of the learning process. It enables students to gauge their progress, tutors to judge the effectiveness of teaching, and can also be used as a teaching tool, to give individuals or groups feedback designed to enable them to improve their performance in the future" [19].

"*It is widely accepted that assessment has a direct impact on students' learning*" [20].

2.2 .Types of Assessments

The term assessment is generally used to refer to all activities teachers use to help students learn and to gauge student progress [21]. Though the notion of assessment is generally more complicated and thus those classifications which are relevant for the research are reviewed.
The first classification is based on the purpose of assessment. Accordingly, there are two types of assessment namely, formative and summative.
Summative Assessment:- is the process of summing up or checking what has been learned at the end of a particular stage of learning, whether it is a module or a course and it is generally carried out at the end of a course or project. In an educational setting, summative assessments are evaluative and typically used to assign students a course grade [22].
Formative Assessment: - is the second type of assessment which is part of teaching and learning and is generally carried out throughout a course or project. It is also referred to as educative assessment which is used to aid learning. In an educational setting, formative assessment might be a teacher or peer or the learner, providing feedback on a student's work, and would not necessarily be used for grading purposes rather it is diagnostic[22].
Educational researcher Robert Stake explains the difference between formative and summative assessment with the following analogy: "When the cook tastes the soup, that is formative. When the guests taste the soup, that is summative." [23].

Summative and formative assessments are often referred to in a learning context as assessment of learning and assessment for learning respectively. Assessment of learning is summative in nature and intended to measure learning outcomes and reports those outcomes to students, parents, and administrators. In addition, Assessment of learning generally occurs at the end of a class, course, semester, or academic year. Assessment for learning is formative in nature and is used by teachers to consider approaches to teaching and next steps for individual learners and the class [22, 24].

As indicated in [22] and others, most of our existing assessment procedures, for example, tests, exams, mark and grades have evolved in relation to the needs of summative assessment. Although formative assessment has always been part of the teaching and learning process, as in the case of teachers comment in the paper, it only very recently that it has become an explicit focus for attention. The educational community is much more confused about what constitutes formative assessment and how it may best conducted than it is in relation to more familiar forms of assessment practice. So the research have noticed this gap and believes that a lot has to be done in supporting formative assessment through various techniques, considering the benefits to the students' improvement as well as to the educational community at large.

The second classification is based on the type of question included in the test. Accordingly, assessment is often categorized as either objective or subjective based on the type of questioning or the response received from the process.

Objective Assessment: - is a type of assessment which requires a form of questioning which has a single correct answer and includes objective types of questions such as true/false answers, multiple choice, multiple-response and matching questions [25]. Objective question can be described as a closed ended question that expects a yes or no, true or false or a choice among several options. For example, the question "አደንዛዥ ዕጾችን ተጠቃሚ ነህ/ ነሽ ወይስ አይደለህም/አይደለሽም?" is objective type of question.

Subjective Assessment: - is a type of assessment which requires a form of questioning which may have more than one correct answer or more than one way of expressing the correct answer. Subjective questions include extended-response questions and essays [25]. Subjective question can be described as open ended question having many right answers. Essays and short answer question are in this category. For example, the question "አደንዛዥ ዕጾችን መጠቀም በምን መልኩ ሊጎዳን ይችላል?" is subjective type of question.

When we compare and contrast the two types of assessment, objective types of assessment are easier to grade but this type question format is widely criticized, because it allows students to blindly guess the correct answer and they may be easy to copy from others. Moreover students may also reduce the writing skills. Subjective types of assessment on the other hand can reveal the depth and breadth of student's knowledge but are much more difficult to grade because of the perceived subjectivity and more effort needed to do the task [21].

Furthermore, objective assessment is well suited to the increasingly popular computerized or online assessment format. Whereas automated technology for analysis and scoring of subjective assessment is still in its infancy. This is largely because the computing power and software sophistication required to do the job has only recently available [26].

A lot of work has to be done in subjective assessment considering its importance and the need for improving the assessment process. Consequently, the next topic is devoted to discuss subjective assessment question particularly essays in detail. Essay is the most commonly used types of subjective question which is used to assess the learning outcomes or knowledge in every domain [27].

2.3 .Basic Notion of Essay Assessment

The word *essay* derives from the French infinitive *essayer*, "to try" or "to attempt". In English *essay* first meant "a trial" or "an attempt". An essay is defined as a well researched and logically structured answer to a particular question, or questions, usually presented as an argument. It is a point of view formulated by critically assessing the information or ideas relevant to the essay topic. Essay is presented in the form of a series of main points which support your direct answer to the question. Each of these points is addressed in a separate paragraph and is supported with evidence, explanation and/or examples. The argument presented in an essay should be supported by referencing in the relevant field. The argument should also form a cohesive whole: this means the paragraphs need to be logically ordered and connections made between the points are presented in those paragraphs. [25]

Essays are used as the most common assessment tool to evaluate your ability to research a topic and construct an argument, as well as your understanding of subject matter. This does not mean that essays are a 'regurgitation' of everything your lecturer has said throughout the course. Essays provide an opportunity to explore in greater depth aspects of the course theories, issues, texts,

etc. and in some cases relate these aspects to a particular context. As a final point Essays create opportunities to articulate ideas, but in a certain way and using formal academic style [25].

2.3.1 Types of Essays

It is difficult to classify essays like the novel; essays are a literary device for saying almost everything about almost anything, usually on a certain topic. But some scholars in the field of writing [25], classified it in to some categories based on their particular purpose and special characteristics. In this classification some essays are written to inform, some to entertain and some to explain and some to persuade. But in general based on the type of question essays basically fall into two broad categories namely, opinion essay and factual essay.

Opinion Essays: - are types of essays which are based on personal opinion or believe and basically written to reflect one's personal judgment on certain issue. They enhance student's writing creativity by allowing them to freely entertain their ideas related to the topic. These types of essay are very subjective as there is no right or wrong answers as long as they are written well and justifiable. However, it is also difficult to evaluate based on a certain frame of reference or guideline. In these types of essay more focus is given on the writing quality rather than the knowledge conveyed [28]. For instance "የ ሰው ልጅ ውበት ወይንም ቁንጅና ምንን ይመስላል?" is an example of such essay topic.

Factual Essays: - is a type of essay which is based on a certain facts or knowledge. Here the knowledge conveyed in the essay is given due attention than that of the writing quality. These types of essays are less subjective and easier to evaluate based on a certain guideline or reference [28]. For instance: "በ ሰውነታችን ውስጥ የደም ዝውውር እንዴት ይከናወናል?" is an example of such essay topic.

The research is directed only to evaluate those essays which are factual types. Prior to performing the task of assessment, it is appropriate to address important issues related with assessment such as: what is to be assessed, the criteria used for assessment, the methods of assessment for effective accomplishment of the task. The next section will discuss in detail about these issues.

2.3.2. Assessing Essay Examination

According to [28, 29], there are basically certain important features that should be considered in evaluating essay. The following are the general features that one should focus when conducting the essay evaluation process.

- **Content:** refers to knowledge of subject and semantic similarity and substantive development of thesis which is relevant to assigned topic. According to [30] content is the most important features and focuses on what is said rather that how it is said.
- **Style:** refers to the way in which sentences or group of sentences put to together. It is very subjective and the focus is on how it is said rather that what is said.
- **Organization:** deals with fluent expression, ideas clearly stated/supported flow of ideas and have logical sequencing of statements.
- **Vocabulary and Language use:** in this case the focus is on knowledge of vocabulary or idiom choice.
- **Grammar usage:** deals with complex sentences, errors of agreement, tense, number, word order/function, articles, pronouns and prepositions.
- **Mechanics:** refers to the correctness of a paper: complete sentences, correct punctuation, accurate word choice
- **Plagiarism** deals with similarity between student essays.

So based on the above features a specific criteria is prepared to perform the evaluation process.

Essay Assessment Criteria

According to [28], the danger of unclear essay grading policies is devastating. It will create a lot of problem in student's performance. To resolve this problem a lot of measures may be taken. Among them, an instruction on how to grade essay papers, or some other clearly defined essay grading criteria is very crucial.

The essay assessment criteria are used as a guideline to evaluate an essay based on the features mentioned above. The grading is based on to what extent a student attempts the specified feature and their corresponding weight or grade is awarded. In relation to this, Amharic department have set their own specific evaluation criteria that can be used to grade an essay. (See the detail information in **Appendix A**).

Other languages also have a specific evaluation criterion which is commonly known as Essay Evaluation Rubric is used to evaluate an essay [31].

Methods of Essay Assessment

Previously Essay evaluation was mostly conducted by human raters, but nowadays the need for automation rises, as there has been a strong interest from the assessment community to introduce increasingly more writing components into standardized tests. Due to this interest, several large scale assessment programs now contain a writing measure. These programs include the Graduate Management Admission Test (GMAT), the Test of English as a Foreign Language (TOEFL), the Graduate Record Examination (GRE), and the College Board's Scholastic Assessment Test II Writing Test and Advanced Placement (AP) exam and so on. Some of these tests have moved to computer based assessment, including GMAT, TOEFL, GRE and Praxis [27]. So the following section will be focused on this new method of assessment which is **Automatic Essay Scoring**.

2.4 Automatic Essay Evaluation[1]/Scoring

Automated Essay Scoring (AES) is defined by [27] as the computer technology that evaluates and scores the written prose. The research on Automated Essay Scoring (AES) has revealed that computers have the capacity to function as a more effective cognitive tool [32].

Most of the research involves the English language, but models are currently being developed to evaluate work in other languages as well. For instance Automatic Essay Scoring System has been done for Japanese [5], Finnish [33], and Chinese [6] languages.

All but the most enthusiastic proponents of AES suspect that there are forms of writing that will always be difficult to evaluate (e.g., poetry). However, for the 90% of writing that takes place in school settings, it should be possible to develop appropriate AES models [27].

AES systems are developed to assist teachers in classroom assessment as well as testing companies in large-scale assessment. They are mainly used to overcome time, cost, reliability, and generalizability issues in writing assessment [34, 35].

A number of studies have been conducted to assess the accuracy and reliability of the AES systems with respect to writing assessment. The results of several AES studies reported high agreement rates between AES systems and human raters [32, 36].

[1] Evaluation and Scoring are the same and used interchangeably throughout the research

AES systems have been also criticized for lacking human interaction, vulnerability to cheating and their need for a large corpus of sample text to train the system [34]. Despite its weaknesses, AES continues attracting the attention of public schools, universities, testing companies, researchers and educators [27].The next point is regarding the type of essays which are suitable for AES systems.

What types of essays are suitable for automatic evaluation?

Among the various types of essay, the most suitable for automatic assessment is that of essays that deals purely on factual knowledge or factual essays. For instance: essays on *scientific facts*, essay on *drug abuse*, essays on *historical events* are types of factual essay which are suitable for automatic evaluation this is due to the ease to assess the essay based on a certain baselines such as reference books and documents [37]. Below an attempt is made to briefly show the historical overview, expected benefits and approaches of AES systems.

2.4.1 Historical Overview Automatic Essay Scoring

Amharic Essay Scoring started in 1966, and has been extremely productive ever since [21]. A limitation of all past work is that the essays or examinations have to be in computer readable format. The pioneer research in the field of automated assessment was by Ellis Page with the Project Essay Grader (PEG) [38]. The aim he pursued was to improve the assessment process.

Advances in other fields such as Natural Language Processing (NLP) and Information Retrieval (IR) gave researchers new techniques to revive the Automatic Essay Scoring. A number of automated essay scoring systems developed and still in development. Currently a “hot topic” in educational measurement is the area of computerized essay evaluation. As essays have been incorporated into many of the standardized testing programs like Systematic Aptitude Test (SAT), GRE, TOFEL and GMAT [39]. Furthermore computerized testing is implemented in more and more instances to encompass a number of areas/discipline for instance computer science, linguistics and cognitive science.

2.4.2 Expected Benefits of Automatic Essay Scoring

A system for automated evaluation would at least be consistent in the way it scores essays, and enormous cost and time savings could be achieved if the system can be shown to grade essays within the range of those awarded by human assessor. Furthermore, according to [27] using computers to increase our understanding of the textual features and cognitive skills involved in

the creation and in the comprehension of written texts, will provide a number of benefits to the educational community. In fact it will help us develop more effective instructional materials for improving reading, writing and other communication abilities. It will also help us develop more effective technologies such as: search engines and question answering systems for providing universal access to electronic information.

2.4.3 Approaches of Automatic Essay Scoring

There is a great diversity of approaches currently used in essay scoring. Besides, there have been a considerable amount of different classifications of techniques to automatically assess essay or free text answers [2, 27]. Table 2.1, summarizes list of most common Automatic essay scoring Systems and their respective approaches.

Table 2.1: Technical approaches in the current existing automated scoring of essays

System	Technique
PEG	Measurement of surface linguistic features
IEA	Latent Semantic Analysis
E-rater	Hybrid approach that combines NLP with SVM
Intellimetric	Artificial Intelligence approach
Larkey's system	Text Categorization Technique
C-rater	Natural Language Processing
AutoMark	Information Extraction
SEAR	Pattern matching based on a schema model
Apex Assessor	Latent Semantic Analysis
Intelligent Essay Marking Systems (IEMS)	Matching with the Indextron clustering algorithm
ATM	Pattern matching based on finding out the concepts and their relationship
Bayesian Essay Text Scoring System (BETSY)	Statistical technique based on classification with the Multivariate Bernoulli Model (MBM) and multinomial event models
School free text Marking Engine (PS-ME)	Natural Language Processing
AutoMarking	Natural Language Processing and pattern matching
CarmelTC	Hybrid approach based on machine learning and a naive Bayesian classification
JESS	Japanese essay scoring system using expert writers
ACES	Chinese AES system using figures of speech

It can be seen from the above table that various approaches are used in the development of automated essay scoring system. When the computer technology advances the approaches used to develop the system also advances, as a result there are now various types of approaches [27]. It is not the aim of the research to review all the approaches currently available rather, we will only give due emphasis to those approaches whose contribution made a great progress to the Automatic essay scoring field. For the simplicity, the research would like to classify the various approaches as general and specific approaches which are relevant to our study.

General Approach of Automated Essay Scoring

According to [27], despite the various numbers of approaches available, the basic procedure is the same. A relatively large set of pre-scored essays responding to one prompt are used to develop or calibrate a scoring model for that prompt. Once calibrated, the model is applied as a scoring tool. Models are then typically validated by applying them to a second, but independent, set of pre-scored items.

The general approach of automated essay evaluation system involves the full Natural Language processing which is the application of computational methods to analyze natural languages. NLP is considered to be one of the most challenging areas of Artificial Intelligence which is defined as the science of making intelligent machines. The research in NLP comprises variety of fields including but not limited to corpus based methods, discourse methods, formal models, machine translation, natural language generation and spoken language understanding [40]. Most AES systems use NLP in one way or another so it is considered as a general approach.

Before we move to specific approaches, automated essays scoring system can be categorized into three types based on their focus on the content, style or both features of the essay. The distinction was first perceived by Ellis Page [38] and later reported by Burstein [41] as the most commonly used features in scoring the models. Each of these is presented below in detail.

Grading essays for style: - the concern here is in the syntax, mechanics and diction of the text. For instance, PEG is a system whose aim is to mark style features such as fluency, grammar etc. by using more indirect measures such as the number of words, the length of the words for example, the use of adjectives. The former are called the intrinsic variables of interest or using the term coined by Ellis Page in 1994, the trins. The problem is that these features are too abstract to be automatically measured by a computer; hence more direct variables are needed as

their representatives. These measures were called proxes. Examples of AES systems that follow this approach are: BETSY, IEMS and SEAR.

Grading essays for content: - in this case the concern is in the meaning of the text. Although previous attempts to develop computational techniques for scoring essays have focused primarily on measures of style [38]: indices of content have remained secondary, indirect and superficial. In contrast to earlier approaches, Latent Semantic Analysis (LSA) methods concentrate on the conceptual content, the knowledge conveyed in an essay, rather than its style, or even its syntax or argument structure. It is also suggested that the most important component in scoring models is content. IEA, Apex Assessor and C-rater have their focus on the content [36].

Grading essay for both content and style:- According to [2], it is necessary to add a new category to this dichotomy, because nowadays most of the researchers are convinced that just evaluating the content or the style is not enough to provide a high quality assessment process. Therefore, the vast majority of the existing systems consider both aspects, such as BETSY, SEAR, IEMS, E-rater and PS-ME.

Specific Approach of Automatic Essay Scoring

There are currently many specific approaches which are used in automatic essay scoring systems as can be shown in table 2.1. These are information retrieval, information extraction technique; pattern matching, clustering and some hybrid approach are also available. But, the research focuses only on that of the Information Retrieval Models which are Vector Space Model (VSM) and Latent Semantic Analysis (LSA). LSA is given due consideration and a detail description of the approach is given as our proposed system is based on it.

Vector Space Model (VSM): it is an old model and originally developed for use in Information Retrieval (IR), which deals with the representation, storage, organization, and access to information items, as defined by [42]; but recently it has been used to measure semantic content in automatic scoring system such as e-rater [29].

The vector vector-space model starts with a co-occurrence matrix where the rows represent terms and the columns represent documents. Terms may be any meaningful unit of information (usually words or short phrases) where as document is any unit of information containing terms, such as sentences, paragraphs, articles, or books. The value in a particular cell may be a simple binary 0 or 1 (indicating the presence or absence of the term in the document) or a natural number indicating the frequency with which the term occurs in the document. Typically, each

cell value is adjusted with an information-theoretic transformation. Such transformations, widely used in IR, weight terms so that they more properly reflect their importance within the document. For example, one popular measure known as Term Frequency–Inverse Document Frequency (TF–IDF) uses the following formula shown in equation 2.1 to compute the degree of relevance of a term with in a document.

$$\text{Wij} = tf_{ij} \log_2(N \div n) \qquad \text{(Eq 2.1)}$$

Here W_{ij} is the weight of term i in document j, tf_{ij} is the frequency of term i in document j, N is the total number of documents, and n is the number of documents in which i occurs.

After the weighting, similarity in the Vector Space Model is determined by using associative coefficients based on the inner product of pairs of documents where word overlap indicates similarity. Several mathematical measures of vector similarity have been proposed in the literature and of these; the most popular measure is the cosine similarity which is based on the principle the wider the angle between the vectors the smaller the cosine value of the angle which indicates the less similar the two vectors. When cosine similarity is applied in essay scoring task, two vectors will be produced one from the bench mark essay and the other from the essay to be scored (new essay). The new essay receives the score of those geometrically closest pre-graded essay benchmarks according to the value obtained using the cosine similarity measure. More specifically, cosine similarity between two documents d1 and d2 is calculated using equation 2.2

$$Sim(d1, d2) = \frac{\text{V (d1)} * \text{V (d2)}}{||V\ (d1)\ ||\ ||V\ (d2)\ ||} \qquad \text{(Eq 2.2)}$$

Where V (d1) and V (d2) are the vector representations of d1 and d2 respectively, the numerator represents the dot product of the vectors while the denominator is the product of their Euclidean lengths.

Latent Semantic Analysis (LSA): the word latent refers to the hidden and the word semantic refers to meaning. Thus it is the meaning of hidden terms and their occurrences. It is also sometimes known as Latent Semantic Indexing (LSI).

LSA is a fully automatic statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. It doesn't use humanly constructed dictionaries, knowledge bases, semantic networks, grammars, syntactic parsers, or morphologies, or the like [1, 43]. LSA is solely based on a collection of documents separated into words or meaningful terms, so it is favorable approach for less resourceful language in terms of language processing tools. Many researches argued that Amharic Language has very few language computational resources. Hence, this research prefers this approach for the language.

LSA is based on the vector-space model discussed previously, but it extends the model further to get more advantage. Specifically, it projects documents into a space with "latent" semantic dimensions [1, 36]. LSA achieves this by performing the following steps:

A. Data Preprocessing
B. Matrix Representation
C. Tuning of the Matrix Weight
D. Singular Value Decompositions
E. Dimensionality Reduction Transformation
F. Reconstruction phase
G. Similarity Scoring phase

A. Data Preprocessing

This is the first step in which input data is preprocessed to produce a set of index terms that represent the input document. In order to achieve this a number of steps are performed for instance stop words, which are words that does not contribute significantly to the topic, are removed and words are stemmed to their original root word, so that there are less words and all of them will carry information and meaning [29, 36].

B. Matrix Representation

In this steps a matrix will be produced from the preprocessed documents where each row in this matrix will be a vector corresponding to a term likewise, column in this matrix will be a vector corresponding to a document, giving its relation to each term. Each cell in the matrix represents the frequency of the term in the document.

For example, Let X be a matrix where element x (i,j) describes the frequency of term i in document j. So this can be represented as a matrix of frequency of terms as shown below in equation 2.3.

$$\begin{array}{ccccc} & & d_j & & \\ & & \downarrow & & \\ & x_{11} & \cdots & x_{1n} & \\ t_i \rightarrow & \cdots & \cdots & \cdots & \\ & x_{m,1} & \cdots & x_{m,n} & \end{array} \qquad \text{(Eq 2.3)}$$

C. The Tuning of the Matrix Weight

In this step the relevance of each word in the passage is measured. This is performed by looking the frequency of the words in all the contexts. The idea is very similar to Term Frequency – Inverse Document Frequency (TF-IDF) which is used in VSM and the same formula can also be applied here.

D. Singular Value Decomposition (SVD)

LSA applies singular value decomposition (SVD) to the matrix. This is a form of factor analysis of a matrix. In SVD, a rectangular matrix is decomposed into the product of three other matrices [43]. For example, the M * N rectangular term sentence co-occurrence matrix (X) is decomposed in to the product of two orthogonal matrices and one diagonal matrix as shown in equation 2.4.

$$\mathbf{X = U\Sigma V^T} \qquad \text{(Eq 2.4)}$$

Where U = $[u_{ij}]$ is an M × M column-orthonormal matrix whose columns are the eigenvectors of the matrix XX^T and they are called left singular vectors. Σ is an M × N diagonal matrix, whose diagonal elements are non-negative singular values sorted in descending order. V = $[v_{ij}]$ is an N×N orthonormal matrix, whose columns are the eigenvectors of the matrix X^TX and they are called right singular vectors [43].

E. Dimensionality reduction transformation

In this step in order to reduce the noise in original matrix X and to capture the hidden semantic structure in the document represented by matrix X dimensionality reduction is performed by reducing the rank of the diagonal matrix. However, there is one big problem in using SVD for dimensionality reduction which is the Selection of the right dimensionality or the value of 'k'. If the value of "k" is too large, it renders the matrix too noisy to be useful whereas if the value of "k" is too small, the reduced matrix will fail to capture the real semantic structure in the documents. Therefore, the optimal dimensionality must be determined empirically [1].

F. Reconstruction phase

In this phase the new diagonal matrix is multiplied by the other two matrices and the outputs of the SVD and the result is the weighted version of the diagonal matrix. It is the LSA representation of the text and in this representation words are located based on their semantic similarity. For example words having similar concepts are located near each other whereas words that are different in meaning are found far apart in the LSA representation. This mapping of words in their concepts is advantage of LSA as it solves the basic problem of information retrieval which is the problem of synonym - the possibility of expressing a given concept in many ways and polysemy - the fact that most words have multiple meaning [43].

G. Similarity Scoring phase

This is the final step in which the similarity score of the document and the query text will be given. The basic idea is once we have the LSA representation of the query text it is compared against the LSA representation of the document text and their similarity is computed. There are different techniques that can be used to compute the similarity of the document and the query vector such as: Dot product, Euclidian distance and Cosine measure [36, 43, 44].

Given the new query vector $Q^T = [w_1, w_2 \ldots\ldots\ldots w_m]$ and one of the document vector from the document matrix is $D^T = [w_{1d}, w_{2d} \ldots w_{md}]$ their similarity measure (Sd) can be calculated as shown in equation 2.5.

$$S(Q^T, D^T) = \frac{\sum_{i=1}^{m} w_i w_{id}}{\sqrt{\sum_{i=1}^{m} w_i^2} \times \sqrt{\sum_{i=1}^{m} w_{id}^2}} \qquad \text{(Eq 2.5)}$$

In conclusion, LSA is also used in various applications like Information Retrieval, Information filtering and currently it is becoming a novel approach in Automatic Essay scoring system [29, 32].

2.5 Performance Evaluation Measurements of AES

The assumption in most of the AES systems is that grades given by human assessors describe the true quality of an essay [33]. Thus the aim of the systems is to simulate the grading process of human raters. Therefore, AES systems is said to perform well if it's able to grade essays as accurately as human raters.

According to [27], there are basically three critical elements of an assessment system: these are validity which deals with worth of measuring whereas the reliability question focused on the acceptable range of score consistency from one rating to another. Finally, the accountability question deals with how testing results are to be reported to the public. These issues should be considered when evaluating the performance of automated systems through various evaluation metrics [29].

Currently, there are a number of evaluation metrics available to measure the performance of AESS. However, common benchmarks and evaluation measures for this application do not currently exist. It is yet impossible to perform a comparative evaluation or progress tracking of this application across similar systems [45].

Moreover, there is no common measure used to make scoring results comparable. Scoring agreement has been reported in terms of exact or adjacent percentages, Pearson or Spearman's correlation, standard deviation, F-score (precision, recall), false positives/negatives, and kappa statistics [45]. The research critically reviewed some of the common evaluation measurements in the following sections.

- **Pearson Correlation or inter-rater reliability:** It measures the standard correlation, that is, how much the teachers' scores (X) are related with the systems' scores (Y) and calculated by applying Equation 2.6.

$$\text{Correlation}(X, Y) = \frac{\text{covariance}(X,Y)}{\text{standardDev}(X) \times \text{standardDev}(Y)} \qquad \text{(Eq 2.6)}$$

- **Spearman rank correlation** is a non parametric test that is used to measure the degree of association between the two ordinal variables reduced to ordinal scale. It uses ranks as opposed to actual values unlike that of Pearson correlation. Equation 2.7 is used to calculate the spearman rank correlation.

$$\rho = 1 - \frac{6 \times \sum d^2}{n \times (n^2 - 1)} \quad \text{(Eq 2.7)}$$

  Where 6 is a constant, n the number of paired ranks and d is the difference between the paired ranks

- **Exact agreement**: - It measures the percentage of times that the system and the human rater have scored just the same value. Human scores are the most reliable source that we have up to date. Hence, counting the number of system scores which have the same value as that of the human score would be the most highly illustrative metric of the system performance [46].
- **Adjacent agreement**: It measures the percentage of times that the system score and the manual score differ with one point [46]. It could be further extended to the equation 2.8 depending on the system performance.

$$Adjacent\ Agreement = \%\ |Manual\_Score - System\_Score| < \partial \quad \text{(Eq 2.8)}$$

Where: ∂ is the accepted threshold between manual score and system score.

Even though correlation is most commonly used measurement in automated essay scoring, the researcher criticizes the use of correlation as a performance measurement. This is because correlation which is a bivariate analysis measures only the strengths of association or linearity relationship between two variables in this case, manual score and system score. Thus the research uses exact or adjacent agreement as a performance measurement which measures the percentage of agreement between system score and manual score which is beyond the linearity relationship.

2.6. Critiques of Automated Essay Scoring

The efficiency of automated essay scoring holds a strong appeal to institutions of higher education that are considering using standardized writing tests graded by AES for placement purposes or exit assessment purposes. However, it is not clear to what extent AES can replace

human raters in judging the quality of essay writing and the validity of AES tools is still a debatable issue. Some researchers criticized AES tools for their over-reliance on surface features of responses, the insensitivity to the content of responses and under value creativity, and the vulnerability to new types of cheating and test-taking strategies [47].

Critics of AES also worried about the consequences of machine grading, which they believed would send the wrong message to students that writing was not important since the audience of student writers was replaced by a machine. In addition, students would be focusing on writing the formulaic essay that matches the computer's highest-score algorithm [48].

At the core of the debate is the issue of whether automatic scoring tools can indeed replace human raters in judging the qualities of writing valued by writing instructors. So far, very few studies have been conducted by independent researchers and users of AES. Thus, it is necessary that more research to be carried out to confirm and shed new light on the existing studies of AES validity and values.

According to [27] automated essay scoring will become more widely accepted when its shifts from that of summative evaluation to a more formative role. It is also proposed by [27] that one mechanism for incorporating AES in electronic portfolios is the possibility of having students "pre submits" their essays before actually turning the work into a human instructor. If this is incorporated as part of a writing class, then more instructors could view AES as a helpful tool, not a competitive one.

2.7 Summary

This chapter reviewed educational assessment and its major classification based on its relevance to our research point of view. The research focused on essays which is subjective type of question and significant to measure the learning outcomes. Two methods of assessment are introduced to evaluate such type of question: manual and automatic grading. But the research is directed to discuss the main ideas related to automatic essay grading system. Further literatures related to historical overview, benefits of automated essay scoring and approaches used to develop the system are reviewed.

Amharic Essay Scoring is a hot area of discussion and there are a number of systems available for commercial systems or as a result of a research in the field. As a result of this, a number of approaches or techniques are available in order to develop the system, but among the most common approaches, the research only discussed VSM and LSA, considering their significance to our proposed method of development. Regarding performance measurement of the system, there are a number of metrics available which are used to measure the performance of the system. The research discusses some of the most common metric systems which are relevant to our work.

Finally the review shows that there are strong critics on automated scoring system. Its validity and reliability is not trusted by some human raters. This is because of wrong perception toward automated essay scoring system. Most human raters perceived automatic essay scoring as a competitive tool and that will replace them from the assessment task. On the contrary, AES is developed to assist and for the effective development of AES the participation of human raters is very vital in both training and testing of the system. Therefore human raters should view AES as the tool that will help them to ease their routine task than a competitive one.

CHAPTER THREE
RELATED WORK

3.1. Introduction

Currently there are a number of Automated Essay Scoring systems available as commercial systems or as a result of research in the field of Educational assessment. Most of the works are limited to English as the language is well advanced in NLP and IR techniques as a result, a number of tools are available to ease the task of automated essay scoring [6]. Nowadays there are also a number of attempts for other language as well. Chinese essays scoring system [6], Japanese essay scoring [5], automated essay scoring for Finnish [33] are some of the attempts made on automated essay scoring system.

The chapter will discuss some of the major essay scoring system in English, Finnish, Chinese and Japanese language focusing on the approach used the data set preparation and the performance of the system. Emphasis will be given to LSA based essay scoring system due to its significance to this research.

3.2. Automatic Essay Scoring Systems for English Language

Most of the work in AES is for English Language such as Project Essay Grade (PEG) [38], Intelligent Essay Assessor (IEA) [49], Electronic Essay Rater (E-rater) [29], Automated Essay Evaluator (AEE) [30], and IntellMetric [50]. Current studies [2, 27, 39] focus on comparison and evaluation of the performance of these systems and the result of the studies is promising in the development of valid and reliable AES system.

The following sections discuss three automatic scoring systems for English language which are relevant to the research we are conducting.

3.2.1. Project Essay Grader (PEG)

PEG is one of the earliest and longest lived implementation of automated essay grading and primarily relies on style analysis of surface linguistics features of a block of text. According to [38], PEG is based on the occurrence of features in computing the score of an essay. PEG, Essays are evaluated based on essential attributes, such as fluency, diction, grammar etc which in general is the writing quality of an essay. As there is no direct measure for these attributes PEG project developed substitutes “proxes” for their measures. For instance if the variable of interest is fluency, PEG measures the number of words.

According to [38] the regression model was developed from a sample of student essays (some with score and some without) and it is separated into two mutually exclusive groups: a design group and a validation group. The design group is used to construct the model and evaluate the parameter. The constructed model is then tested on the validation group. In the regression model, the proxes are the predictor variables and the score is the response variable. By using the proxes a score could be determined and compared to the score from a human judge and the performance can be measured.

According to [28] PEG uses various software products, including a grammar checker, a program to identify words and sentences, software dictionary, a part-of-speech tagger, and a parser were used to gather data about many proxes.

Concerning the performance, PEG's latest experiments achieved results reaching multiple regression correlation as high as 0.87 with human graders [2].

3.2.2. Electronic Essay Rater (E-Rater)

E-Rater is most commonly used essay scoring system in Education Testing Service (ETS) and has been used to score as a second reader for the essay submitted for the Graduate Management Aptitude Test (GMAT) Exam. It is also integrated in different Automated scoring systems as a back end for instance like Intelligent Essay Assessor (IEA) [49], Automated Essay Evaluator (AEE) [30] and for criterion online essay evaluation service [29], which is a web based services developed by ETS to evaluate a student writing skill and provide instantaneous score and diagnostic feedback.

According to [51] E-Rater uses a combination of statistical and NLP techniques to extract linguistic features from the essays to be graded. Essays are evaluated against a benchmark set of human graded essays. E-Rater adopts a corpus-based approach to build a model using actual essay data to analyze the features of a sample of essay responses. The application is designed to identify features in the text that reflect writing qualities specified in human reader scoring criteria. According to [2], E- Rater is currently composed of five main independent modules. Three of the modules identify features that may be used as scoring guide criteria for the syntactic variety, the organization of ideas and the vocabulary usage of an essay. A fourth independent module is used to select and weight predictive features for essay scoring. Finally, the last module is used to compute the final score.

E-rater follows the scoring guide used by expert human evaluators for scoring. This scoring guide has a six-point scoring scale. It checks the essay for the argument structure, syntactic structure and vocabulary structure. The software is based on three general classes of features: Syntactic, rhetorical and topical content features. The features are extracted from the essay texts and quantified using computational linguistics techniques. As pointed out by [51], the features are described as follows:

Syntactic features: - the analysis of syntactic features of an essay can include the quantity and ratio of simple, compound and complex sentence, types of independent clauses, use of auxiliary verbs and other features.

Rhetorical features: - it measures the argument structure of the essay. E- Rater identifies and quantifies the essay use of cue words and other rhetorical features. For example words like "possibly" and "perhaps" are often used while developing an argument. Words like "this" and "these" signify the writer has not changed topics and phrase. Words like "in summary" or "in conclusion" signify the coming of a summary.

Topical content features: - E-rater evaluates the topical content of an essay by comparing the pattern of words it contains to those patterns found in manually graded essays. According to [51], two different measures which are Essay content which is based on the vocabulary usage in the essay as a whole and Argument content which is based on the vocabulary usage for specific arguments found in the essay are point out.

Concerning the score prediction, E-rater is trained on a sample of 270 essays that have been scored by human readers. It measures more than 50 features and then computes a stepwise linear regression to select those features which make a significant contribution to the prediction of essay score. The value is rounded to the nearest whole number to yield the score.

In regard to [2], the performance of e-rater was measured and over 750000 GMAT essays have been scored, with agreement rates between human expert and system consistently above 97%. By comparing human and E-Rater grades across 15 test questions, the empirical results range from 87% to 94%.

3.2.3. LSA based Automatic Essay Scoring

Previous attempts to develop computational techniques for scoring essays have focused primarily on measures of style [38]; indices of content have remained secondary. In contrast to earlier approaches, LSA based methods concentrate on the conceptual content, the knowledge conveyed

in an essay, rather than its style, or even its syntax or argument structure. The following sections reviews two of the most common LSA based essay scoring system Intelligent Essay Assessor (IEA) [49], Automated Essay Evaluator (AEE) [30].

3.2.3.1. Intelligent Essay Assessor (IEA)

The IEA is a computerized essay evaluation program that automatically assesses and critiques electronically submitted essays, providing assessment and instructional feedback useful in almost every subject area. IEA is a back end service using Knowledge Analysis Technology (KAT) and a customer's web interface to evaluate essays as reliably as skilled human readers [52]. It is also defined by [39], as a web based service which is based on LSA that provides an evaluation and advice on the conceptual concept of written work.

The IEA measures abstract factual knowledge based on extensive background readings, texts, sample essay and news sources, not just superficial factors such as word counts, word length, keywords, or punctuation. The IEA's assessment focuses on the understanding of the subject matter that goes into the creation of an essay and the semantic space that is created from all the material in the creation of an essay allows IEA to read any essay in the same domain and understand the many synonyms and alternate ways of stating the same important ideas.

IEA's content measures are based on Latent Semantic Analysis (LSA). IEA computes the overall content similarity in LSA space between a new essay and essays on the same topic that have been graded by humans, and next determines the proximity of the new essays to human graded essays. IEA predicts score based on this proximity in semantic space to the human graded essays and the new essay will be given the score of the closest pre-graded essay using holistic score.

According to [28], the holistic method has been tested on a large number of essays over a diverse set of topics. The essays have ranged in grade level, including middle school, high school, college and college graduate level essays. The topics have included essays from classes in introductory psychology, biology, history, as well as essays from standardized tests, such as analyses of arguments, and analyses of issues from the Educational Testing Service (ETS) Graduate Management Achievement Test (GMAT).

Concerning the performance measurement, according [51], the inter-rater reliability level is the standard criterion used for the Intelligent Essay Assessor. Other systems often use percentage of agreement on the score points, which is a somewhat less sensitive measure. The scores given by the IEA are compared to the scores given by a human reader for a particular set of essays. If the

reliability level between the IEA and a human reader is comparable to the reliability level between two human readers, the model is considered a success. The reliability is evaluated through a correlation coefficient based on the score pairs for the set. A reliability of 1.0 indicates absolutely perfect agreement between readers.

Concerning the performance, a test is conducted on the GMAT essays using the IEA system which resulted in percentage for adjacent agreement with human graders between 85% to 91% [2].

IEA can be applied in various fields as in education as a consistency checker, in which the teacher grades the essay and then the IEA re-grades the essay and indicates discrepancies between the two grades. IEA also used in US army for training a student [51], plagiarism detection and distance education [49], and so on.

3.2.3.2. Automated Essay Evaluator (AEE)

AEE is a system developed to automate the evaluation of large collection of essay type documents using NLP and LSA techniques. Rule based NLP is used for grammar checking while LSA is used to evaluate the content.

To assess quality of the content of a document or essay, LSA techniques is used to model human knowledge and extract words and passage from a text. The AEE was first trained on domain representative text then on human graded essay that would be compared to the new essays to be graded. This comparison will generate the appropriate score for the essay through a matrix manipulation [30].

AEE was trained on a corpus containing pre- graded essays gathered from high school department and the essay were written by third year students to address the topic "*characteristics of Jose Rizal that made him worthy to be called our National hero*" and were graded by at least two human teachers based on its mechanics, organization and content of the essay.

Concerning the grammar checking, E-rater is used which make use of a combination of statistical and NLP to determine the essays feature for evaluation in AEE. Based on the process of the AEE system, AEE system is divided into three major modules [30]:

Training Modules: - This module is comprised of methods for converting essays into matrices and storing them in a corpus which will be used later on the content evaluation modules of the system. Before it employs the process of LSA for the actual training the pre graded essay

undergoes preprocessing to remove stop words and generate keywords that will represent the rows of the matrices. After that actual training will be performed.

Grammar Evaluation: - the grammar evaluation of the essay was implemented by checking its grammatical correctness. Each sentence in the essay is analyzed through three processes. Where lexical analysis, parsing of sentence and semantic analysis stage was performed

Content Evaluation: - in this case new essays are evaluated based on its consistency to the central idea and focusing on the given topic. The content evaluation is done by transforming the essay into a vector and then comparing it to the matrices of the pre graded essay stored in the corpus those essays on which the system was previously trained. The average grade of the pre graded essay most similar to the input essay is computed and is assigned as its content score. Finally after the grades for grammar and content criteria have been retrieved, the overall grade of the essay is computed.

Concerning the performance, a number of experiments are conducted to assess the performance of AEE; from the results performed the accuracy of the system is very promising and doesn't necessarily depend on the number of pre- graded essays. For example the corpus containing 25 pre-graded essays had an average percentage of error 2.48% which is exactly the same as the corpus containing 125 pre-graded essays.

3.3 Automatic Essay Assessor for Finnish Language

Automatic Essay Assessor (AEA) is a java based essay grading system for Finnish Language although the system is designed to work for other languages as well. AEA is based on LSA which provides a means of comparing the semantic similarity between a source and target text. LSA can also be considered as a computational model of human knowledge representation and research has showed that the method is able to simulate leaning [33, 53].

The system preprocesses the essays by morphological parsers of Finnish and also allows the user to adjust a set of parameters, for fine-tuning the accuracy. The grade is computed by using both essays, previously graded by humans, and an assignment-representative text from a text book and all the information is saved into the MySQL database [33, 53].

Concerning the experiment, 143 essays from an undergraduate course in education were collected and the essays were graded by a professor on a scale from zero to six. The grading model was constructed by using a relevant chapter from the course textbook and part of the

essays. The result shows that, the spearman rank correlation between the scores given by human grade and the system ranges between 0.78 and 0.82 [33].

3.4 Automated Japanese Essay Scoring System (JESS)

According to [5], JESS is the first automated Japanese essay scorer and has become most famous since it was introduced in February 2005.

JESS has been created for scoring essays in college-entrance exams and it was based on articles written by experts to develop the model, unlike all of the previous AES system such as PEG[38], E-rater[51], IEA[49], ACES[6] and AEA[33] which they make use of expert raters or human graders to develop their model for evaluating essays.

As indicated by [5], all automated system previously developed are based on the assumption that the true quality of essays must be defined by human judges. But many researchers including [54], criticized the over reliance on human ratings as the sole criterion for evaluating computer performance. The researchers argued that human ratings are typically based as a constructed rubric that may ultimately achieve acceptable reliability at the cost of validity. Thus it was insisted by researchers to use another approach to perform better than ordinary human raters would be to use expert writers than expert raters.

Reputable professional writers produce sophisticated and easy to read essays and by detecting a statistical outlier to predetermined essay features compared with many professional writings for each prompt, the system can evaluate essays based on three features: Rhetoric, organization and content.

Among the scoring criteria of JESS which are rhetoric, organization and content. A technique called latent semantic indexing can be used to check whether the content of a written essay responds appropriately to the essay prompt.

Concerning data set preparation, JESS used some form of electronic media for use as essay models so complete article from the Mainichi Daily Newspaper up to 2005 and the Nihon Keizai newspaper up to 2004 is used for the purposes of linguistic study. Furthermore with regard to morphological analysis, syntactic analysis and parsers are available (some of which are free) and with such resources tools were prepared for computer processing of the articles and columns that was collected as essay models.

Concerning the performance, many experiments are conducted and showing promising results for instance experiment using 143 university students' essays collected at the National Institute

for Japanese Language on the topic “smoking” shows a correlation between Jess and the expert raters was 0.83, which is higher than the average correlation of expert raters (0.70).

3.5 Summary

In this chapter, research work on automated essay scoring developed for three different languages such as: English, Finnish and Japanese is presented. But emphasis was given to that of English as there are a number of AES systems available. The chapter discusses on the basic aspects of automated essay scoring systems focusing on the data set preparation, the techniques used and the experimental result.

To sum up the Automatic Essay Scoring is being implemented in various languages using various techniques to develop the model and to predict essay scores. The performance of the AES systems is also very promising and as a result many assessment companies for instance Education Testing Service (ETS), Knowledge Analysis Technology (KAT) have started to use AES system to grade an essay in large scale in English language. Hence many researches have to done in other language as well and enjoy the benefit.

CHAPTER FOUR

DESIGN AND IMPLEMENTATION OF AUTOMATIC AMHARIC ESSAY SCORING SYSTEM USING LATENT SEMANTIC ANALYSIS

4.1. Introduction

This chapter describes the design and implementation of Automatic Amharic Essay Scoring System (AAESS) using Latent Semantic Analysis. The design and implementation of AAESS requires a large number of corpus or pre-graded essays for developing the model and performing the experiment. Hence in this research, we prepared a large corpus of pre-graded essays and the next chapter discusses on the process of the corpus preparation in detail.

The design and implementation process of the proposed system involves the development of two general architectures which are the LSA Model Architecture and the Amharic Essay scoring system Architecture.

In the first Architecture the LSA Model is generated from the manual scored essays which involves preprocessing, matrix construction, Singular Value Decompositions, Dimension Reduction and Matrix Reconstruction. The second architecture makes use of the LSA model to predict a score to the query essay.

4.2. System Architectures

As shown in Figure 4.1, the general architecture of the LSA model has two modules which are preprocessing module and semantic space module to generate the LSA model that is used later in scoring the query essay.

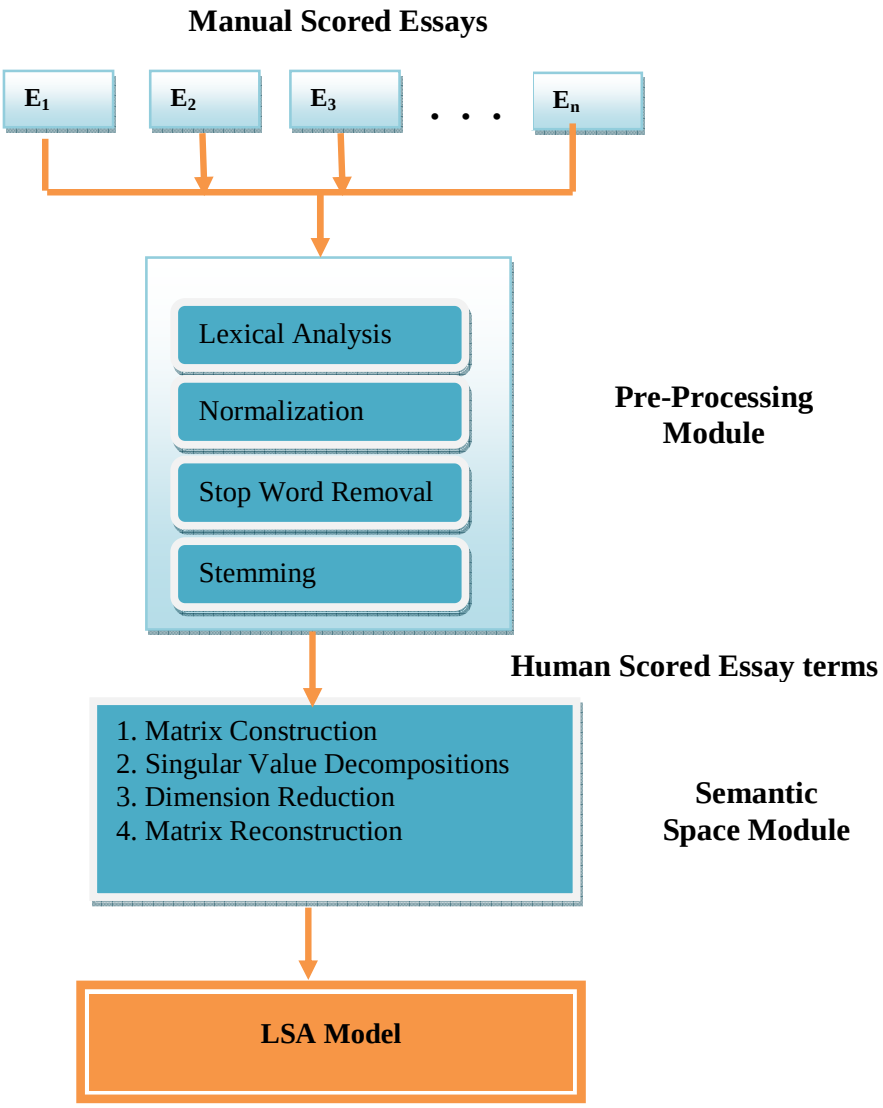

Figure 4.1: the general Architecture for LSA Model

The second architecture has three modules which are preprocessing module, essay vector construction module and essay scorer module that uses the LSA Model to predict a score for the query essay as shown in Figure 4.2.

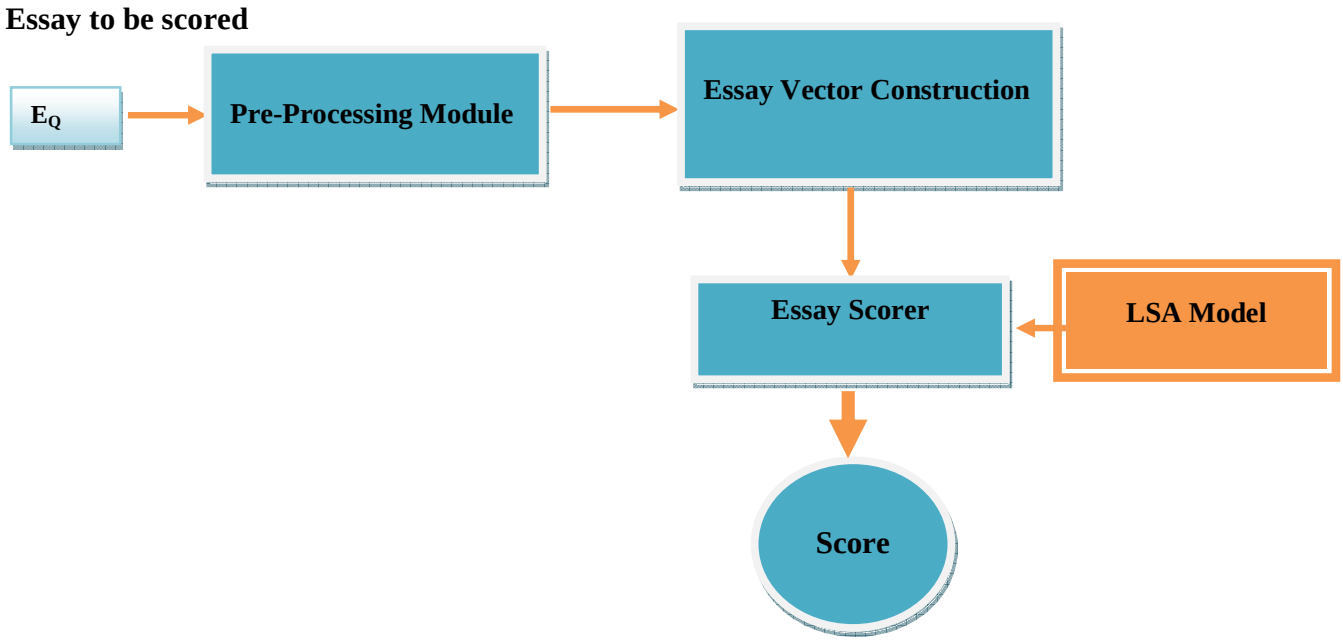

Figure 4.2: the general Architecture of Automatic Amharic Essay Scoring System Using LSA Model

The modules of both architectures are discussed in detail in the following subsections.

4.2.1. Preprocessing Module

It is a module which is found in both architectures that takes an essay document and produces terms which are rich in semantic content. In order to achieve this objective, the module carries out lexical analysis, normalization, stop-word removal and stemming. In this thesis, we have

adopted the preprocessing module from the work of Tesema [55]. In the following sub-sections, detailed explanation of the adopted components is presented.

Lexical Analysis

The first step in the preprocessing of the input document is lexical analysis which is also known as tokenization. The essay evaluation involves producing a matrix of word and essay co occurrence, hence lexical analysis in essay evaluation involves splitting essays into words. In this thesis individual words are extracted from the essay by scanning each sentence for the predefined word delimiters such as new line, space, Amharic punctuations, etc.

Normalization

There are several Amharic characters that have the same pronunciation and use but different symbols. Such characters are automatically replaced by a common character in this component. For example, the different forms of the Amharic word 'Hailu', which are ሀይሉ, ሃይሉ, ሐይሉ, and ኀይሉ are all converted to the common form ሀይሉ by changing the first character of the three words. Table 4.1 shows sample character replacements used by Tessema [55].

Table 4.1: Sample of Normalized Characters [55].

Characters to be replaced	Replacing character
ሐ፣ኅ፣ኀ፣ሓ	ሀ
ዐ፣ዓ	አ
ሠ ሡ . . . ሯ	ሰ ሱ . . . ሰ
ኸ	ከ
ጎ	ገ
ዉ	ው

Furthermore, the normalization component also includes the expansion of words that are written in a short form using "/" or "." into one word or two words. For example, ት/ቤት is expanded to ትምህርት ቤት. To achieve this, each word obtained after tokenization is checked for its presence in

a list of common short words and if a word is found to be in the list, it is expanded into its corresponding form [43].

Stop Word Removal

Some words, referred to as “stop-words” are words that do not contribute significantly to the overall topic of the essay such as conjunctions, articles, pronouns. Such words can be identified and removed by using a predefined list of stop-words [43].

In addition to the common stop words such as ነው, ናቸው, ነበረ, ይገልጻል etc, the research also identified other stop words besides adopting previous work**.** This is performed by reading all the essays and point out some words which we believe that it doesn’t contribute any meaning to the essay topics.

Stemming

The next step in the preprocessing of the input essay after removing of stop-words is stemming. This work adopted the stemming algorithm developed by Tessema. The stemming algorithm removes those affixes that are usually used for changing tense, number, gender and case of a word. Furthermore, in the case of removing suffixes with vowels, the last character of the word after the removal of suffix is changed to sades (the sixth order character in Amharic alphabet). The output of this phase produces terms that represents the essay under consideration. These terms can further be refined to retain only those terms that occur above a certain threshold in the essay. However, since Latent Semantic Analysis works on the idea of word co-occurrences or on the co-relation of words in the document, it is necessary to retain as many of the words in the document as possible [43]. Hence, in this thesis, we have decided to use all the terms in the document following the removal of stop words.

4.2.2. Semantic Space Module

This is the core module which builds the LSA model of the pre graded essays in specific domain and used to compare with the new essay. In this model some latent semantic structure is assumed in the data where LSA attempts to capture in the matrix representation [1].

It involves matrix constructing, Singular Value Decomposition, dimensionality reduction and matrix reconstruction steps. Each of the steps is discussed below.

Matrix construction and term weighting representation

In this step the list of essay text are converted in to a matrix where each row stands for a unique essay term and each column stands for an essay. The cells of the matrix contain the frequency in which essay terms appears in the essay which defines the term weight in that essay. Term weighting refers to the measure of the importance of a term in particular document or essay. There are a number of terms weighting approaches available [43], but the research uses term frequency or local weighing scheme which is the importance of a term with in a document and it is most appropriate for the research we are conducting. This is because, in our research essay to be graded is compared to the human-graded essays and the grade is based on the most similar essay's grade; but in the case of essays which is to be graded by comparing to the essay topic related materials such as textbook the grade is given based on the similarity of all these material so in this case it is appropriate to use the global weighting schemes that demands the importance of a term within the whole document [33].

Singular Value Decompositions

SVD is a form of factor analysis which is used to condense large matrix $\mathbf{A}$ into products of three matrices for analysis. A weighted Matrix $\mathbf{A}$ is decomposed in to the product of two orthonormal matrices ($\mathbf{U}$ and $\mathbf{V}$) and one diagonal matrix ($\mathbf{\Sigma}$) as shown in the equation 4.1.

.

$$\mathbf{A = U\Sigma V^T} \qquad \text{(Eq 4.1)}$$

Where U is an M × M column-orhonormal matrix whose columns are called left singular vectors, ∑ is an M × N diagonal matrix, whose diagonal elements are non negative singular values in decreasing order, ($\sigma 1, \sigma 2, \ldots, \sigma n$) represented as ∑ diag ($\sigma 1, \sigma 2 \ldots \sigma n$), V is N × N orthonormal matrix whose columns are called right singular vectors. The research used matlab libraries for such singular value decomposition.

Dimension Reduction

In order to find the relations between terms and their context it is necessary to reduce the rank of diagonal matrix by removing the list of insignificant eigenvalues from $\mathbf{\Sigma}$ and reduced to $\mathbf{\Sigma_r}$. The corresponding Eigenvectors from $\mathbf{U}$ and $\mathbf{V^T}$ is also reduced to $\mathbf{U_r}$ and $\mathbf{v_r}$ respectively**.** This is

because in order to perform matrix reconstruction in the next step, the Eigenvectors should have the same dimensions as that of the diagonal matrix.

In this transformation of dimension reduction, the hidden relationship of terms is detected. It was also suggested that truncating the singular value decomposition matrix reduces the amount of noise in the data while retaining only the most important structure [43, 53]. Hence, it is the most important step in LSA to acquire the semantic representation of the pre-graded essays after the Singular Value Decomposition of matrix is performed.

In this step, noisy relationships are suppressed and important relationships become very clearly visible. Dimensionality reduction do not have straightforward rule that can be used to select the optimal value of reduced dimension say r. Hence, in this research a simple algorithm is used to selects the best r so that n-r eigenvalues is removed where n is the total number of eigenvalues.

The algorithm first computes the difference in values of the eigenvalues and identifies the eigenvalues that show the maximum difference or gap and the value which shows the maximum gap is considered as the cut of point or r values. The basic idea behind the algorithm is that the absolute difference of each consecutive pairs of singular values are small for terms having similar meaning and in the same way for terms that doesn't have similar meaning the absolute difference of consecutive pairs of singular values are larger than the previous difference of singular value pairs.

Matrix Reconstruction

In this step the new reduced singular matrix $\mathbf{\Sigma_r}$ is multiplied with the other two reduced matrices $\mathbf{U_r}$ and $\mathbf{V_r}$ as shown in equation 4.2 to generate LSA model of the pre-graded essays.

$$\mathbf{A' = U_r \Sigma_r {v_r}^T} \qquad \text{(Eq 4.2)}$$

This matrix ($\mathbf{A'}$) is the Latent Semantic representation or the LSA model of the pre-graded essays and used as inputs for the essay scorer module to score any new essay from the same domain.

4.2.3. Essay Vector Construction Module

This module is involved in preparing a query vector from new essay to be graded. The query vector consists of the same number of terms as that of the LSA representation of the pre-graded essays. This is achieved through mapping the LSA representation of the pre-graded essay terms to that of the new query essay to be graded. The terms that are found in the LSA representation but do not occur in the new query essay have 0 frequencies, where as the terms that appear in the query essay but not in the LSA representation are discarded. In addition, terms that appear in both the query and the LSA representation have their term frequency in the essay vector.

As a result a query vector is created having the same dimensions as that of the LSA representation of the pre-graded essay which is suitable to perform similarity comparison in essay scorer module. The next topic discusses this issue in detail.

4.2.4. Essay Scorer Module

In this module, we performed similarity comparison between the LSA representation and the query vector which is the essay to be-graded in the same semantic space. The score of the new essay is predicted based on the computed similarity measure and various similarity measures exist in the literature [56]. In this research the cosine similarity is chosen because it is the most common similarity measurement used in researches involving LSA and it is easy to implement and shows successful results [36].

The cosine similarity between the query vector **Q** and one of the pre-graded essay document vector $\mathbf{E_i}$ in the LSA matrix $\boldsymbol{A'}$ can be defined as the dot product of the two vectors divided by their magnitude and it is shown in equation 4.3.

$$\cos\boldsymbol{\theta_i} = \frac{\boldsymbol{Q}.\boldsymbol{E_i}}{||\boldsymbol{Q}||\,||\boldsymbol{E_i}||} \qquad \text{(Eq 4.3)}$$

Where Q is the query vector or the essay to be graded, E_i is the i^{th} essay from the pre-graded essays matrix, θ_i is the angle between Q and E_i. Q**.** E_i is the dot product of the two vectors and ||Q|| , || E_i || are the magnitude or the length of the vectors Q and E_i respectively and it can be calculated as $||Q||=\sqrt{Q \times Q}.$, and $||E_i||=\sqrt{E_i \times E_i}$

The similarity between the two vectors Q and E_i is directly related to the cosine of the angle between those vectors which ranges from 0 to 1. The smaller the angle the more similar are the vectors. The cosine similarity is implemented in our research using Matlab 7.1. The module generates all the possible θ values corresponding to the total number of essay in the LSA representation in their decreasing cosine values in other words in their increasing cosine angles.

Once θ values are obtained, the new essay is graded using the θ values and the actual grade assigned for the essays which is used in the LSA modeling. The cosine of the angle θ between the vectors shows the similarity between the query vector and each of the pre-graded essays. The higher the similarity or the cosine of the angle between the query vector and one of the vectors in the pre-graded essay, the closer the match between the two vectors as a result similar score is given to the query essay.

The scoring of new essay is achieved through the following steps; each of the pre-graded essays E_i is represented in a semantic space with their corresponding grades. The query essay Q, which is to be graded is also represented in the same semantic space as shown in the Figure 4.3. The essay query is being compared to each of the pre-graded essays and **M** weighted neighbors of essays with the highest similarity value in other words with the smallest cosine angle to the query essay is selected. Then the score for the query is calculated as the average score of the pre-graded essays weighted by their cosine values and can be calculated using equation 4.4.

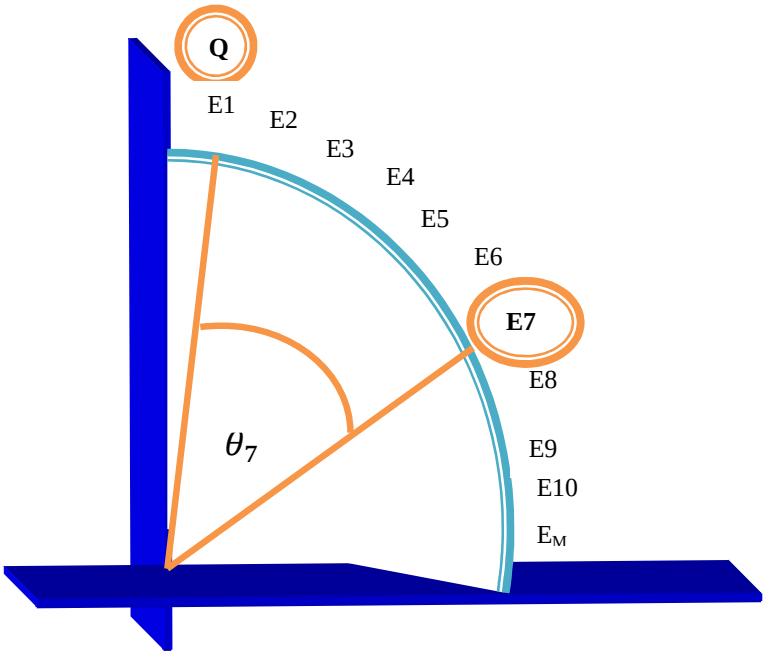

Figure 4.3: Essay scoring represented by N dimensional spaces.

Where E1, E2, E3…E_M are pre-graded essay vector with manually assigned score values S1, S2, S3…S_M respectively.

$$Score(Q) = \sum_{i=1}^{M} w_i S_i \qquad \text{(Eq 4.4)}$$

Where w_i is the weight assigned to the pre-graded essay E_i whose grade is Si defined as:

$w_i = \frac{\cos\theta_i}{\sum_{j=1}^{M} cos\theta_j}$, θ_i is the angle between the query vector Q and the i^{th} essay vector E_i and S_i is the score of the pre- graded essay i, in which i ranges from 1 to M, and M is the number of pre-graded essay that gives the best prediction of the grade.

The number of M- weighted neighbors or M number of pre-graded essays that gives highest cosine similarity with the query is selected empirically through conducting experiment which involves comparing the percentage of adjacent agreement between the predicted score and their corresponding manual score. The number of M-weighted neighbors having highest percentage of agreement is chosen from the total number of pre-graded essay pairs used for essay score prediction.

In this research, 10 pre-graded essays with highest cosine similarity were found sufficient and were used as total number of essays used for score prediction but using large number of pre-graded essay pairs is advantageous.

4.3. Summary

In the design and development of automatic Amharic essay scoring system, the research used LSA or information retrieval technique which is a novel application to score essay based on the extent to which an essay can be matched against other essays scored by human raters. The design involves the development of two general architectures which are the LSA model architecture and the Amharic Essay scoring system architecture.

The first Architecture generates the LSA model from the manual scored essays which is used as an input to the second architecture. It involves preprocessing of the essays, matrix construction, singular value decompositions, dimension reduction and matrix reconstruction.

The second architecture involves query essay preprocessing which includes lexical analysis, normalization, stop word removal and stemming to generate query essay terms which is mapped to the LSA representation to have the same dimension in the essay vector construction module to perform similarity comparison. Finally essay scorer module predicts the score of the query essay by calculating the average score of those pre-graded essays with the highest cosine similarity from the query essay.

CHAPTER FIVE

EXPERIMENTAL RESULTS

5.1. Introduction

In this chapter, we present Amharic essay corpus preparation, the data set, the evaluation metrics, the experimental results and finally discussion of the experimental result of the Amharic essay scoring system proposed in chapter four. Despite the fact that evaluating the performance of the essay scoring system is an important part of the study, it was challenging task as there is no standard method or well defined criteria for essay scoring. This chapter will describe the set of procedures that is used to conduct the experiment.

5.2. Amharic Essay Corpus Preparation

Design and development of automatic essay scoring requires a large number of pre-graded essays both for developing the model and testing the system, so the research gives due emphasis considering its importance to the development of the system. Different institutions, various linguistics professionals and a large number of students in various academic levels are involved in this process. Factual essays which is focused on the content of the knowledge of the topic was prepared by linguistic professionals in three different domains and students are allowed to take the essay exam and then the essay exam was corrected by two Amharic PhD students who developed specific criteria to minimize the subjectivity. Finally the marked essay is encoded and used for developing the model. The following subsections discuss the processes in detail.

Essay Exam Preparation

This is the first task which requires preparing an essay exam to be tested by large number of students in various institutions. In preparing the exam two linguistic advisors[2] from Addis Ababa University have participated and three different specific domains or topics are selected which are: *essay on drug usage*, *essay on trade* and *essay on transportation*. The topics or the domain of the essays are carefully selected by considering its importance or relevance both to the research as well as to the Amharic writing syllabus or curriculum.

From the research point of view the questions are prepared to address only on the content of the topic or the knowledge conveyed in the topic rather than the quality or the style of the language.

[2] Among the linguistic advisors Ato Derege Gebre is well experienced linguistic teachers moreover he is working as vice president for Ethiopian writers association.

On the other hand the topics are frequently appearing in the curriculum of writing lesson in the Amharic linguistic study. Three specific topics are selected in order to enable us to evaluate the performance of the model in three various domains. The following are the three topic of essay exam question prepared for the research:

Topic 1: ስለ አደንዛዥ ዕጾች ምንነትና ስርጭት እንዲሁም ስለሚያደርሱት ማሕበራዊና ኢኮኖሚያዊ ቀዉስ በአጭሩ ግለፁ?

Topic 2: የንግድ ግንኙነት ማን ማንን ያካትታል? ግንኙነታቸውስ እንዴት ይገለፃል?

Topic 3: የአለም የመጓጓጉ ዘዴ ምንን ያካትታል? ስለደረሰበት የእድገት ደረጃም በአጭሩ ግለፁ?

After the topics are selected the essay exam was prepared. Detail instructions are also included in the essay exam such as students are requested to write an essay with total words ranging from 100 to 150 and other relevant information is required to make the exam a real world exam. From past experience[3] students will be reluctant to perform well when they are informed that it is an input to the research. Therefore, the researchers thoroughly discussed the issue with the teachers and agreed that students should be informed as a quiz or test rather than the input to the research. From the research ethics point of view it is not necessarily mean unethical to do so as far as it creates no problems on the students especially on privacy issues. The essay exam is found in **Appendix B**.

Essay Examination Testing

After the essay exam is prepared it was reviewed by other Amharic Linguistic department staffs in order to make sure that the topics of the questions are relevant and can be used to evaluate the content or the knowledge conveyed. Following the revision of the essay exam, it was tested in three academic institutions and large numbers of student in various levels are participated. As a result various ranges of result appeared in the specific domain and used to develop the model that can be used to automatically evaluate new essay, this is the basic idea behind various range of results in the essay corpus preparation.

Three governmental academic institutions: Addis Ababa University, Kotebe College of Education and Dagmawi Mennelik II Secondary School are selected for the purpose of essay

[3] We have tried to test students by clearly informing to the students about the research and the result was not good enough

examinations. In regarding to the various levels of students taking the exam there is also an issue that is well addressed in this research. Hence, from Addis Ababa University degree program students in Amharic Linguistic and folklore department, from Kotebe College of Education in diploma program in Amharic and English department and finally from Dagmawi Mennelik II secondary school, grade 10+1 and 10+2 students have participated in the exam. It was discussed previously that the objective of varying the level of students as well as institutions is in order to obtain good distribution of results. This helps us to develop or design a machine learning model that can able to grade any type of new essay in the specific topic in which it is trained.

Following the various students or target groups (about 300 in number) are selected. They are informed by their teacher prior to the essay exam and students in the same institutions took the exam at the same time. Sample essay exam responses in each domain are found in **Appendix C**.

The share of the exam in each institution is shown below in a table 5.1. It can be seen from the table that Addis Ababa University has the largest share and Dagmawi Menelik being the least share this is due to Addis Ababa University has large number of students available in different faculty where as in the case of Dagmawi Menelik there is no specialization or department and it is added to include variety for the research to have different range of results.

Table 5.1: Number of Essay examination share among three institutions

Institutions	**No of essay exams taken**
Addis Ababa University	150
Kotebe College of Education	100
Dagmawi Mennelik II Sec School	50

Even if most of the students are from the linguistic department, the research can involve other students as well. This is due to the focus of the study is to evaluate the content of the knowledge rather than the style or the language used. Therefore the research is not delimited to linguistic departments other natural and social science student can also be used as a target group. Linguistic department are chosen as they are easily available through the researcher linguistic professional supporters. Moreover it is unusual to find essay written on science topics in Amharic, this is due to the fact that students are learning such fields in English media as a result their essay exam will only be in English. The researcher believes that supporting science fields in

Amharic is also important at least in lower level institutions as it enables them to understand the knowledge in their contextual environment. Essays on science and social studies are based on factual knowledge and it is suitable for the research we are conducting.

Among the 300 students that are planned to take the exam, 268 essay responses are collected. This is due the following reasons: very few students were not available for the test, some exams were totally blank or not written and some exams were lost. Therefore the research considered 268 exam responses and prepared for human evaluation. The next topic will discuss issues related with essay exam grading by human raters in detail.

Essay Exam Manual Marking/Grading

In this step the essay exams are collected for manual evaluations. The research used two PhD students[4] from Addis Ababa University, Akaki campus for the grading process. As we have seen in the literature essay grading process is inherently subjective process [28]. Accordingly to reduce the subjectivity a detail specific criterion on how to evaluate the essay is vital. In this research a detail specific criteria was designed by linguistic advisors to minimize the subjectivity of the process.

The process of conducting manual grading involves classifying the essay responses in to four categories based on its relevance to the research. The categories are shown below in table 5.2.

[4] Those PHD students have done their master degree on TEAM (Teaching of Amharic language) and have vast experience in teaching different Amharic courses

Table 5.2: Essay category after the evaluation process

Categories	Description	Status
Completed Essays	All the three question in the essay are answered	All essays are marked
In completed Essays	Some of the questions are left blank so the students did not attempt all the three questions in the essay.	Essays are marked and combined with other incomplete essays to form completed essay
Limited Essays	The given answer are so limited in reference to the asked question and criteria stated in the instructions (essays below 100 words)	Essays which are too limited are discarded, but some limited essays are marked in order to have large data
Unrelated Essays	Examinations under this category are with unrelated answers.	Some unrelated essay are marked to 0, but some which are very irrelevant are discarded

Based on the above information some papers were believed to be unnecessary and excluded from the research, therefore the research does not include the entire 268 essay exam papers rather around 207 essays are used for human grading. Following the selection of essays which are to be graded, specific criteria is set for each topic and the grading is performed based on the criteria which are shown in table 5.3, 5.4 and 5.5 for topic1, topic 2 and topic 3 respectively.

Table 5.3: Manual Essay Scoring Criteria for topic 1: Essay on drugs

	Criteria (10 points)							
	Unity (2 points)			**Coherence (2 points)**			**Completeness(6 points)**	
	Well unified (2 pt)	Unified (1pt)	Not unified (0 pt)	Well coherent (2 pts)	Coherent (1 pt)	Not coherent (0 pt)		
Question 1:							Definitions (2pts)	
							Drug Transmissions (2 pts)	
							Effects of drug usage (2pts)	Social effect
								Economical effect

Table 5.4: Manual Essay Scoring Criteria for topic 2: Essay on trade

	Criteria (10 points)							
	Unity (2 points)			**Coherence (2 points)**			**Completeness(6 points)**	
	Well unified (2 pt)	Unified (1pt)	Not unified (0 pt)	Well coherent (2 pts)	Coherent (1 pt)	Not coherent (0 pt)		
Question 2:							Trade participants	Merchants
								Customers
								Government
							Trade relationship (3pts)	

Table 5.5: Manual Essay Scoring Criteria for topic 3: Essay on transportation

	Criteria (10 points)							
	Unity (2 points)			**Coherence (2 points)**			**Completeness(6 points)**	
	Well unified (2 pt)	Unified (1pt)	Not unified (0 pt)	Well coherent (2 pts)	Coherent (1 pt)	Not coherent (0 pt)		
Question 3:							Types of transportation (4 pts)	Animal
								Land
								Sea
								Air
							Level of development (2pts)	

The above essay grading criteria is used to evaluate an essay ranging from 0 to 10 points for each topic. It measures unity, coherence and completeness of the essay. Unity refers to discussion of ideas on the same topic in which all points must talk about one single idea, whereas coherence measures the smoothness of flow ideas in the topic. Finally, completeness deals with the factual information conveyed in the topic. The detail criterion of essay grading is found in **Appendix D**.

Essay Exam Encoding

This is the final step of Amharic essay corpus preparation that encodes the marked essays into computer readable format using visual geez Unicode. It is time consuming and tedious task considering the poor and unreadable hand writing of most students. It was performed by two typists and later reviewed by the researcher. We are able to encode around 207 completed and marked essays in three domains each. Some modification is done by the researcher to construct completed essay from those essays which are previously in completed. In addition 7 essays are removed and 200 essays are used as essay corpus this is due to for the suitability in the data set preparation.

Following the essay corpus preparation, 90% of the data is used for developing the model and the rest 10% is used for evaluating the performance of the model using cross validation.

Cross validation is a technique used for developing the model which involves partitioning of a sample of data into complementary subsets, performing the analysis in one subset called training set and validating the analysis on the other subsets called validation set or testing set. Among the cross validation technique the research used 10-fold cross validation technique which is most commonly used in NLP. The next topic deals with experimental procedures in order to evaluate the performance of our system.

5.3. Experimental setup/procedures

In order to evaluate the Amharic essay scoring system, the research carried out the following procedures which are described in the following subsections.

5.3.1. Data set preparation

As we have discussed in the corpus preparation section, the research prepared 200 essays in each of the three topics which is used to measure the performance of the system in three different domains. The data set is prepared using 10-fold cross validation techniques.

The 10-fold cross validation divides the data set into 10 subsets, and the experiment is repeated 10 times in each topic. Each time one of the 10 subsets is used as the test set and the other 9 subsets are put together to form a training set. Figure 5.1 shows the 10-fold cross validation for a topic.

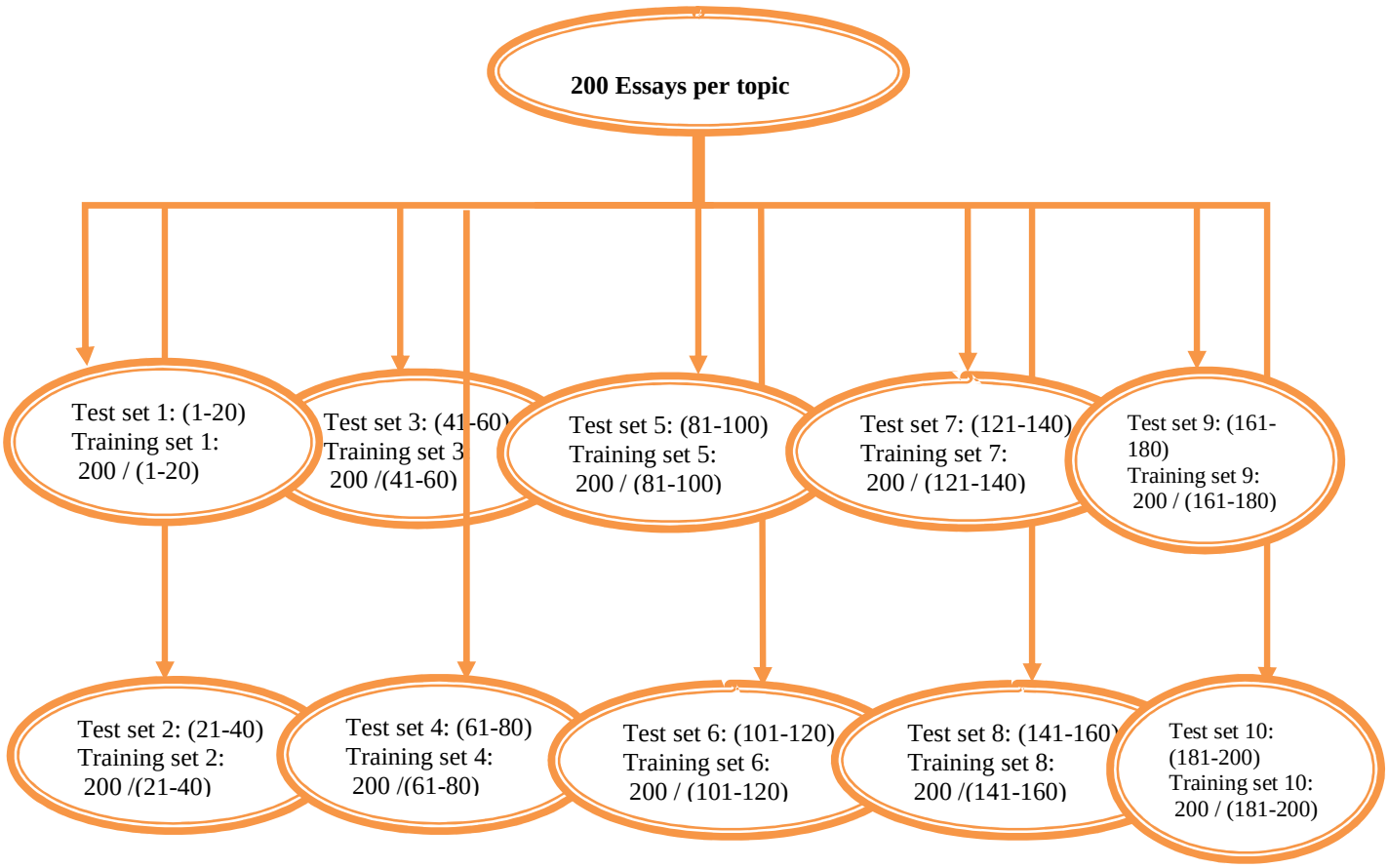

Figure 5.1: 10- fold cross validation for a topic

The same 10-fold cross validation technique is applied for all the topics used in the research. The advantage of this method is that every data point gets to be in a test set exactly once, and gets to be in the training set 9 times. The disadvantage of this method is that the training algorithm has to be rerun from scratch 10 times which means it takes 10 times as much computation to make an evaluation in each topic in the experiment.

5.4. Evaluation Metric

Performance Evaluation of the Amharic essay scoring system is done using the percentage of adjacent agreement between the system score and the human score and it can be defined as follows:

- **Adjacent agreement**: measures the percentage of times that the system score and the manual score differ with some significance value ∂ and it is given in the equation 5.1 in the next page [46].

$$Adjacent\ Agreement = \%\ |Manual\_Score - System\_Score| < \partial \qquad \text{(Eq 5.1)}$$

Where ∂ is the accepted threshold between manual score and system score.

It can be inferred from the equation 5.1 that the smaller the value of ∂; the best performance is to be measured. The research used percentage of adjacent agreement in order to select the best prediction of scores from the number of pre-graded essays pairs or M-weighted neighbors of pre-graded essays which gives highest cosine similarity to the query essay.

Following the selection of the best prediction of system scores, the research used confusion matrix which is commonly used visualization tool in the field of artificial intelligence. In our research confusion matrix is used to map and match the human score and the best predicted system score as shown in the Table 5.6.

Table 5.6: Confusion matrix between manual score and system score

		Manual Score										
System Score		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
	1											
	2											
	3											
	4											
	5											
	6											
	7											
	8											
	9											
	10											
												Total

As shown from the table, the column indicate the ranges of human score where as the rows indicate the ranges of scores predicted by the system and the cell indicates the number of essays or counts having the human score as indicated by the column and also system score as indicated by row. It can be inferred from the table that as we move to the diagonal, the number of essay predicted by the system is the same as that of manual score. On the contrary, as we move away from the diagonal the system predicted essays varies from that of human score.

5.5. Result of the Experiment

The research conducted 10 experiments in each topic by combining one test set and one training set. Table 5.7 shows the number of experiments conducted in the research

Table 5.7: The number of experiments conducted in each topic

No	Topics	Number of experiments conducted
Topic 1	Essay written on Drugs	10
Topic 2	Essays written on Trade	10
Topic 3	Essay written on Transportation	10

In order to simplify, the research combined the experiments based on their topics to have one large scale experiment in each topic. The experiment which is conducted in each topic has two objectives. The first objective is to select the best essay score and the other is to visualize the performance of the selected best essay score with that of manual score.

In the first case, the experiment is used to select the best essay score which is done by taking the average score of the pre-graded essays which have highest cosine similarity to the query. The M^5-weighted pre-graded essay pairs is used to predict best score is determined empirically by measuring the percentage of agreement between their predicted score and the corresponding manual score of the query essay.

In the second case, confusion matrix is produced to visualize the performance of the best predicted score to their corresponding manual score.

5.5.1. Experimental result on topic 1: *Essays written on Drugs*

In this experiment the score of the essays on topic 1, which are 200 essays written on drugs are predicted. Essay score prediction is achieved by taking the average score of the 10 pre-graded essays which have highest cosine similarity with the query. The score prediction starts from the first pre-graded essay (M=1) until the average of the last pre-graded essay pair (M=10). Table 5.8 shows the sample result of the experiment one but, the full experiment is found in **Appendix E.**

[5] M is the number of pre-graded essay pairs which gives highest cosine similarity to the query essay

Table 5.8: Sample essay[6] scoring result on topic 1: Essays written on drugs

No	*Sys.[7] Score M=1*	*Sys. Score M=2*	*Sys. Score M=3*	*Sys. Score M=4*	*Sys. Score M=5*	*Sys. Score M=6*	*Sys. Score M=7*	*Sys. Score M=8*	*Sys. Score M=9*	*Sys. Score M=10*	*Manual Score*
1	5.500	5.254	5.488	5.608	5.683	6.052	5.978	6.276	6.299	6.318	6.000
2	3.000	4.235	4.484	4.368	4.585	4.893	5.046	5.041	4.930	5.319	4.500
3	6.000	6.000	6.800	6.487	6.489	6.098	5.678	5.598	5.534	5.865	6.000
4	4.000	3.505	4.327	4.739	5.371	5.473	5.338	5.297	5.527	5.807	5.500
5	6.500	5.276	5.513	5.632	5.981	5.831	6.112	5.870	5.883	6.940	7.000
6	5.000	5.472	5.801	5.379	5.591	5.338	5.361	5.558	5.605	5.547	7.000
7	5.500	4.762	4.999	4.756	4.803	4.916	4.790	5.054	4.837	4.757	5.000
8	5.000	6.219	6.632	6.485	6.026	6.326	6.219	5.975	5.928	5.935	5.000
9	4.000	5.469	5.954	5.728	5.496	5.259	5.496	5.615	5.655	5.879	5.000
10	6.000	5.510	5.830	6.113	5.716	5.838	5.591	5.639	5.573	5.472	7.000
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
199	5.000	4.513	4.023	4.979	4.790	4.986	4.850	5.169	5.258	5.233	5.500
200	4.000	5.248	4.834	4.875	4.899	5.160	5.347	5.548	5.597	5.683	5.000

From this sample of essay scoring result, it is not possible to decide the best score among the 10 predicted scores by simply observing the table. Hence percentage of adjacent agreement between manual score and each of the predicted essay score is performed to select the best essay prediction.

The research conducted percentage of agreement by varying the threshold ∂ less than or equal to 0.25, 0.5 and 1.0 as shown in Figure 5.2. The threshold $\partial <= 1.0$ is used to determine the best score prediction in the research. This is because in this threshold better performance of adjacent agreement is found without over generalizing the result. In the case of using threshold $\partial >= 1.0$ even if high percentage of adjacent agreement can be found, it will over generalize the result, hence not used in the research.

[6] In this case sample score is selected using the first 10 essays and the last 2 essays from the total of 200 essays in topic 1

[7] Sys.Score is to mean System Score

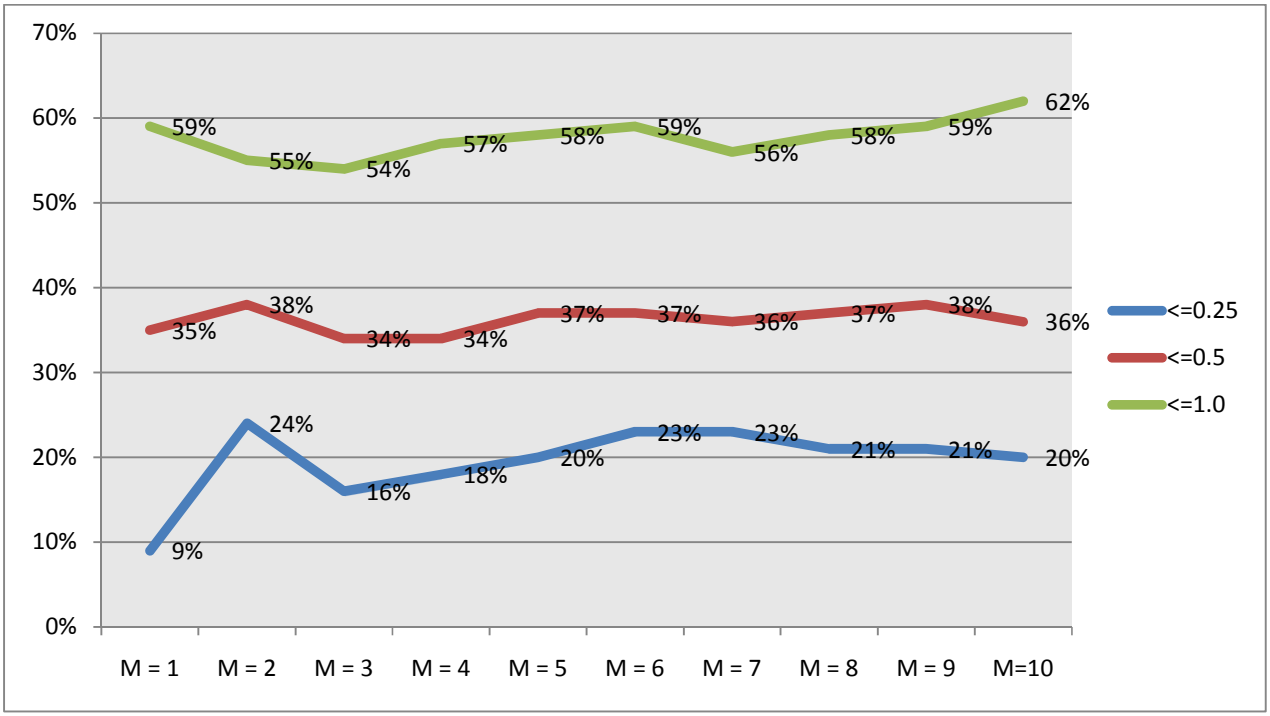

Figure 5.2: Best score prediction using percentage of adjacent agreement with $\partial <= 1.0$ for topic 1

Using the empirical result from Figure 5.2, the predicted score of using the average of 10 pre-graded essay pairs (M=10) is selected as the best prediction of score with maximum percentage of agreement of 62% with the corresponding human score as compared with the others. The line curve further shows that as we increase the numbers of pre-graded essay pairs to compute the average score the performance also increases.

Following the selection of the best prediction of score we can further visualize the performance of the best prediction of the system score with manual score using confusion matrix as shown below in Table 5.9.

Table 5.9: Confusion matrix between Manual score and System score for topic 1

		Manual score												
		3.00	**4.00**	**4.50**	**5.00**	**5.50**	**6.00**	**6.50**	**7.00**	**7.50**	**8.00**	**8.50**	**9.00**	**9.50**	**Total**
System score	**3.00**	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	4.00	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	4.50	1	2	2	0	0	1	0	0	0	0	0	0	0	**6**
	5.00	0	7	2	9	3	6	1	1	2	0	0	0	0	**31**
	5.50	1	7	0	12	12	13	7	6	1	0	0	0	0	**59**
	6.00	0	4	2	9	4	15	8	11	3	7	3	1	1	**68**
	6.50	0	1	0	3	2	2	7	5	0	1	0	1	0	**22**
	7.00	0	0	0	0	1	1	2	2	0	2	0	2	0	**10**
	7.50	0	0	0	0	0	1	0	0	1	0	0	1	0	**3**
	8.00	0	0	0	0	0	0	0	0	0	1	0	0	0	**1**
	8.50	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	9.00	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	9.50	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	Total	**2**	**21**	**6**	**33**	**22**	**39**	**25**	**25**	**7**	**11**	**3**	**5**	**1**	**200**

It can be seen from the confusion matrix most of the score is concentrated around the diagonal escalating the performance of the system on the contrary, there are also some scores which deviate from the diagonal these account for the degrading the performance.

5.5.2. Experimental result on topic 2: *Essay written on Trade*

In this experiment the score of the essay on topic 2, which are 200 essays written on trade are predicted and their performance is measured using percentage of adjacent agreement and visualized in confusion matrix.

Essay score is predicted using the average of 10 pre-graded essay scores which have highest cosine similarity to the query essay. Table 5.10 shows the sample result of the experiment 2 but, the full experiment is found in **Appendix F**:

Table 5.10: Sample of essay scoring result on topic 2: Essay on trade

No	*Sys. Score (M=1)*	*Sys. Score (M=2)*	*Sys. Score (M=3)*	*Sys. Score (M=4)*	*Sys. Score (M=5)*	*Sys. Score (M=6)*	*Sys. Score (M=7)*	*Sys. Score (M=8)*	*Sys. Score (M=9)*	*Sys. Score (M=10)*	*Manual Score*
1	6.500	5.340	4.600	5.629	5.699	5.431	5.696	5.619	5.267	5.420	4.00
2	5.000	4.501	4.830	5.113	5.374	5.393	5.078	5.242	5.218	5.379	4.00
3	5.000	5.500	5.002	4.754	4.991	5.389	5.208	5.470	5.323	5.475	3.00
4	5.000	5.500	5.002	4.754	4.991	5.389	5.208	6.470	5.323	5.475	9.00
5	4.000	4.932	4.654	4.949	4.958	4.964	5.291	5.485	5.486	5.441	6.00
6	7.500	6.516	6.347	6.630	6.896	6.426	6.714	6.628	6.127	6.212	6.50
7	7.000	6.530	6.675	6.401	6.231	6.350	6.370	6.327	6.294	6.267	8.00
8	7.500	6.501	6.666	6.271	5.853	5.255	5.354	5.429	5.385	5.442	5.00
9	4.000	5.209	6.097	6.430	6.253	6.135	5.982	5.865	5.669	5.842	5.00
10	4.000	5.209	6.097	6.430	6.253	6.135	5.982	5.865	5.669	5.842	4.50
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
199	7.500	5.370	4.934	4.725	4.948	4.808	5.327	5.183	5.261	5.282	3.00
200	6.500	6.744	5.536	5.769	6.006	6.161	6.273	5.348	6.498	6.459	4.00

From this sample of essay scoring result, it is not possible to decide the best score among the 10 predicted scores by simply observing the table. Hence percentage of adjacent agreement between human score and each of the predicted essay score is performed to select the best essay score

prediction. Figure 5.3 shows the percentage of agreement by varying the threshold ∂ less than to 0.25, 0.5 and 1.0. The threshold $\partial <= 1.0$ is used to determine the best score prediction.

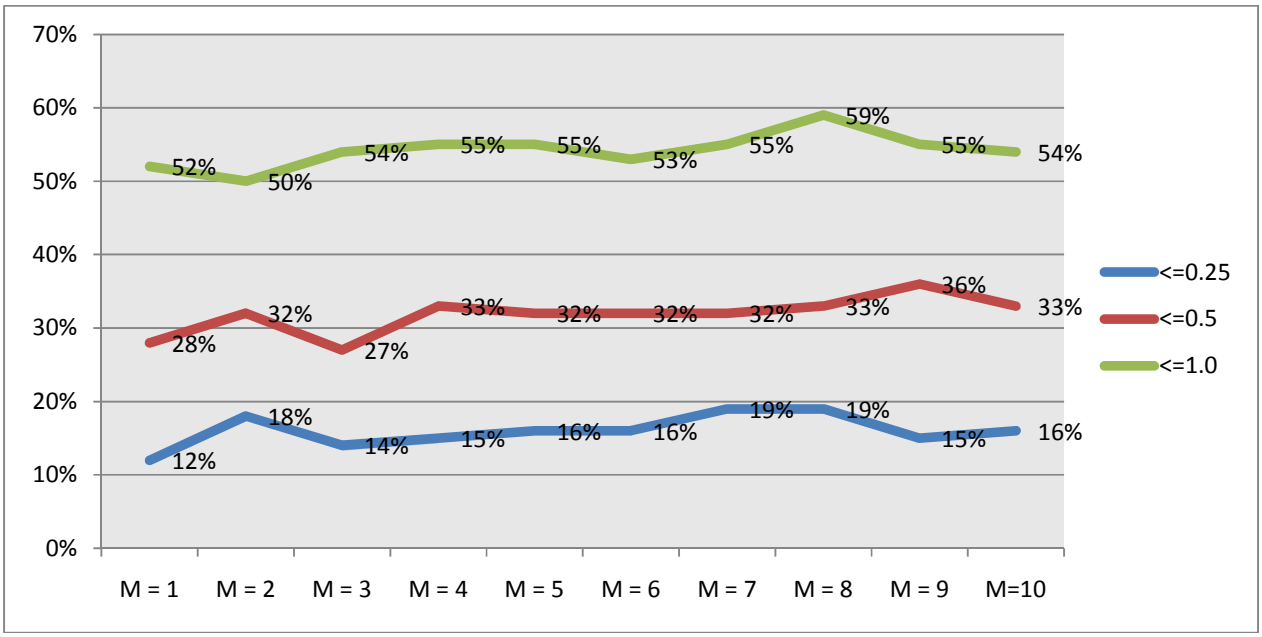

Figure 5.3: Best score prediction using percentage of adjacent agreement with $\partial <= 1.0$ for topic 2

As shown the line curve in Figure 5.3, the predicted score of using the average of 8 pre-graded essay pairs (M=8) have a maximum percentage of agreement of 59% therefore, it is selected as the best score prediction of the query essay. The line curve further shows that as we increase the number of pre graded essay pairs (M) the percentage of adjacent agreement increase until M reaches 8, it then started to decrease. Following the selection of the best prediction of score we can further visualize the performance of the best prediction of system score with manual score using confusion matrix table as shown below in Table 5.11.

Table 5.11: Confusion matrix between Manual score and System Score for topic 2

		Manual score																
		2.00	**3.00**	**3.50**	**4.00**	**4.50**	**5.00**	**5.50**	**6.00**	**6.50**	**7.00**	**7.50**	**8.00**	**8.50**	**9.00**	**9.50**	**10.00**	Total
System score	**2.00**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	3.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	3.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	4.00	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**2**
	4.50	2	1	0	4	0	1	1	0	0	0	0	0	0	0	0	0	**9**
	5.00	2	3	1	6	0	8	2	3	2	3	0	0	0	0	0	0	**30**
	5.50	1	2	1	8	3	8	4	5	1	2	1	0	0	1	0	0	**37**
	6.00	0	1	0	5	1	9	5	12	4	10	4	0	0	2	0	0	**53**
	6.50	1	0	0	2	0	2	3	5	8	9	2	1	1	0	0	1	**35**
	7.00	0	0	0	1	0	1	0	0	0	13	2	6	0	2	0	1	**25**
	7.50	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	**4**
	8.00	0	0	0	0	0	0	0	0	0	0	0	2	1	1	0	0	**4**
	8.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	9.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	9.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	10.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Total		**7**	**7**	**2**	**27**	**4**	**29**	**15**	**25**	**16**	**38**	**9**	**9**	**2**	**7**	**1**	**2**	**200**

It can be seen from the confusion matrix most of the score is concentrated on the diagonal escalating the performance of the system and some even lie on the diagonal itself showing the exact matching between the system and human score. On the contrary, there are also some scores which deviate from the diagonal which account for the degrading the performance.

5.1.1. Experimental result on topic 3: *Essay written on Transportations*

In this experiment the score of the essay on topic 3, which are 200 essays written on transportation are predicted and their performance is measured using the percentage of adjust agreement.

Essay score is prediction is achieved using the average score of pre-grade essay which have highest cosine similarity ranging from 1st pre-graded essay (M =1) to 10th pre-graded essay (M=10). Table 5.12 shows the sample result of the experiment 3, but the full experiment is found in **Appendix G**:

Table 5.12: Sample essay scoring result on topic 3: Essay on transportation

No	*Sys. Score M=1*	*Sys. Score M=2*	*Sys. Score M=3*	*Sys. Score M=4*	*Sys. Score M=5*	*Sys. Score M=6*	*Sys. Score M=7*	*Sys. Score M=8*	*Sys. Score M=9*	*Sys. Score M=10*	*Manual Score*
1	7.000	7.000	6.676	6.284	6.323	6.652	6.697	6.505	6.454	6.459	6.50
2	4.000	4.487	4.649	4.497	4.869	5.199	5.173	5.327	5.139	5.400	4.00
3	8.000	7.510	7.182	6.419	6.530	6.686	6.867	6.883	6.842	6.953	6.00
4	5.000	5.705	6.111	5.618	5.878	5.209	6.586	6.282	6.252	6.134	4.00
5	7.000	6.763	7.154	6.882	6.904	6.074	6.670	6.821	6.890	6.633	4.00
6	6.000	6.499	6.995	7.244	7.392	7.410	7.492	7.433	7.493	7.397	5.00
7	4.000	4.000	4.312	5.637	6.075	6.372	6.520	6.349	6.264	6.375	4.50
8	8.000	7.536	7.996	7.997	7.998	7.763	7.393	7.346	7.362	7.234	6.00
9	6.500	6.747	6.181	5.649	6.293	6.326	6.145	6.247	6.221	6.106	4.00
10	7.500	7.743	8.151	7.154	7.125	7.105	6.678	6.716	6.693	6.862	8.00
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
199	6.500	6.988	6.831	6.754	6.615	6.208	6.045	6.215	6.193	6.221	6.5000
200	4.000	6.346	6.712	6.066	6.243	6.046	5.777	5.745	5.568	5.653	6.0000

From this sample of essay scoring result, it is not possible to decide the best score among the 10 predicted scores by simply observing the table. Hence percentage of adjacent agreement between manual score and each of the predicted essay score is performed to choose the best essay score prediction. Figure 5.4 shows the percentage of agreement by varying the threshold ∂ less than to

0.25, 0.5 and 1.0. The threshold is $\partial <= 1.0$ is used to determine the best score prediction in the research.

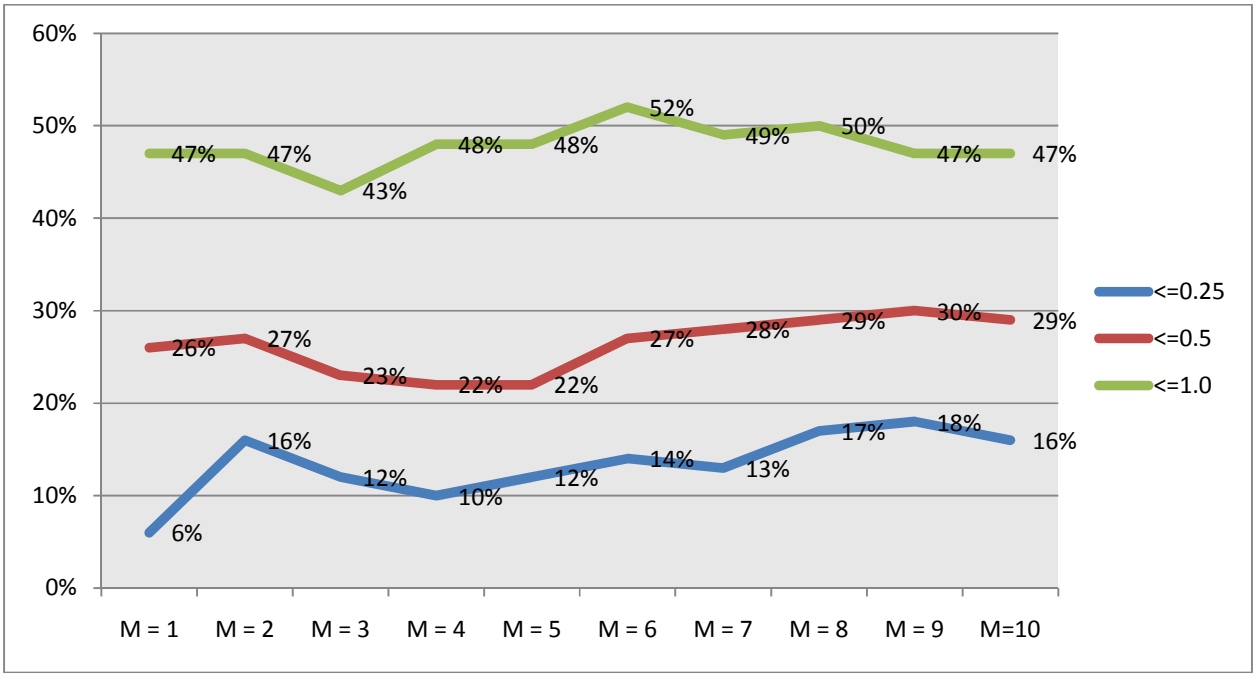

Figure 5.4: Best score prediction using percentage of adjacent agreement with $\partial <= 1.0$ for topic3

Using the line curve in Figure 5.4, the predicted score of using the average of 6 pre-graded essay pairs (M=6) is selected as the best prediction of score with maximum percentage of agreement 52% with threshold value $\partial <= 1.0$. It can also be interpreted from the table that as the number of pairs increase the percentage started to decrease steadily. Following the selection of the best score prediction we can further visualize the performance of the best prediction of system score with manual score using confusion matrix as shown below in Table 5.13.

Table 5.13: Confusion matrix between Manual score and System Score for topic 3

		Manual score																
		2.00	**3.00**	**3.50**	**4.00**	**4.50**	**5.00**	**5.50**	**6.00**	**6.50**	**7.00**	**7.50**	**8.00**	**8.50**	**9.00**	**9.50**	**10.00**	**Total**
System score	**2.00**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	3.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	3.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	4.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	4.50	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	**2**
	5.00	0	1	0	3	0	4	0	0	2	2	1	0	0	0	0	0	**13**
	5.50	0	0	0	3	0	1	5	3	2	4	1	1	0	1	1	0	**22**
	6.00	1	3	0	7	1	2	3	11	5	7	5	3	0	0	0	0	**48**
	6.50	0	1	0	1	0	4	1	10	9	9	3	6	1	3	1	0	**49**
	7.00	0	0	0	0	0	3	1	6	3	4	6	7	3	4	0	0	**37**
	7.50	0	0	0	0	0	1	0	2	0	2	2	5	1	0	0	0	**13**
	8.00	0	0	0	0	0	0	0	2	0	2	0	1	4	1	1	2	**13**
	8.50	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	**2**
	9.00	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**1**
	9.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	10.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
	Total	**1**	**5**	**1**	**14**	**1**	**16**	**10**	**34**	**21**	**31**	**18**	**23**	**10**	**9**	**3**	**3**	**200**

It can be seen from the confusion matrix that most of the score is concentrated around the diagonal escalating the performance of the system. On the contrary, there are also some scores which deviate from the diagonal which account for the degrading the performance.

5.2. Discussion of the Results

The research conducted intensive experiments in three topics and shown promising results. Three topics are selected to measure the performance of the system in three different domains. The variation of results in each domain and the overall results are discussed in this section.

Variations of the performance of the Amharic essay scoring system in each topic

In the first experiment, *essays written on drugs*, the system performs relatively well having percentage of adjacent agreement 62% which is because of the issues raised concerning drugs is purely factual and there is less biasness when grading the essays among the human raters as compared with the other essay topics as a result, the system performs relatively well.

On the contrary, the third experiment, *essay written on transportation*, the system performs relatively low compared with other experiments having the percentage of adjacent agreement 52%. The questions specifically topic 3 which were designed by Amharic department is not purely factual (የአለም የመጓጓዣ ዘዴ ምንን ያካትታል? ስለደረሰበት የእድገት ደረጃም በአጭሩ ግለፁ?). This is because the first part of the question requires listing which is factual however; the second part requires personal justification. Thus the questions is not purely factual which lead to biasness to some extent accordingly, the human raters is also biased when grading the essay hence, the performance is lowest than the others.

Factors that degrade the performance of AAESS

Other issues which have degrading effect on the performance of automated essay scoring system are discussed below.

- There are many situations which accounts for the decrease in the performance of automatic essay scoring systems. Among them, inter-rater agreement accounts most. This is because the assumption in most of the AES systems is that grades given by human assessor describe the true quality of an essay [33] thus essay graded by human is used in prediction of the system score. Consequently, if there is some biasness in manually grading score between the human raters then this accounts for the reduction of the performance of the system score. In order to have better performance of essay grading system, there should be common consensus between human raters in manually grading essay. In the case of the Amharic essay scoring system, the human raters mostly scored essay independently using the specific essay grading criteria. This results in variation of individual subjectivity between the human raters which adds some biasness to the system performance.

- The other factors that accounts for degrading the performance of the system is regarding to the specific scoring criteria which is used to grade essays manually by human raters. The scoring criterion which is found in the Appendix C is mostly based on the content of the topic but there are also some issues included from the basic elements of essay which are unity and coherence. These criteria especially that of coherence or flow of ideas in the topics is hardly evaluated by the system but it is graded and given a score in the case of manual score. This accounts for the variation of the grade predicted by the system and by human raters.
- **Preprocessing Component:-** this component has direct effect on the performance of the system for instance the effect of stemmer component has aggressive effect in which two words which are different in meaning may be stemmed to the same value for instance: ፈተናዎች (exams) and ፈቶች (divorcees) are two different terms but they will be stemmed to the same word ፈት [7].
- **Dimensionality Reduction: -** is one of the important steps in generating the LSA model. There is a big problem related with this issue which is the selection of the right dimensionality or the value of r, if the value of "r" is too large it renders the matrix too noisy to be useful whereas if the value "r" is too small, the reduced matrix will fail to capture the real semantic structure in the document. The performance of the system is affected on the finding of the optimal dimensionality.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

"***Quality Education***" has been enormously thought in minds of all academicians, institutions, governmental bodies, and others. Yes it is true that quality should be the essence for education; nevertheless, the increase intake capacity of educational institutions, shortage of skilled and experienced academic staffs, shortage of academic resources, and many more have degraded the quality of education. Therefore, to bring quality for education it requires a strong commitment and hard work from educational community, the government and the nation at large.

From the educational community perspective, educational assessment plays a major role in bringing quality for education. This is because assessment involves all activities teachers use to help students learn and to gauge their progress [47]. As a result, improving the educational assessment process has direct role in improving the quality of education and this research focused on improving the assessment process through automation techniques which involve using computer technology to perform student assessment. Essays are one of the most accepted forms of student assessment at all levels of education and have been incorporated in many of the standardized testing programs such as SAT, GRE and TOFEL [39].

Currently a "hot topic" in educational measurement is the area of automatic essay evaluation and many researchers have developed automatic essay evaluation systems for different languages using different methods. The results of these research works have shown that essay evaluation can be automated and good results can be obtained.

In this thesis we have proposed a method to develop Amharic essay evaluation for factual topics based on Latent Semantic Analysis. The design and implementation of the proposed system involves the development of two general architectures which are LSA model architecture and the Amharic essay scoring system architecture. The first architecture generates the LSA model form pre graded essays where as the second one uses the LSA model as input to predict the score of the query essay.

.

The research performed intensive experiments in three topics to measure the performance of the proposed system in different domains. The performance of the proposed system is measured using the percentage of adjacent agreement between the system score and manual score. The result is found to be 62%, 59% and 52% for topic 1, topic 2 and topic 3 respectively.

The result shows that our proposed model is promising considering the first attempt to develop AES systems for Amharic language however; further research is also required to increase the performance of our proposed model.

At last, AES is a developing technology and many AES systems are used to overcome time, cost, and generalizability issues in writing assessment. The search for excellence in machine scoring of essays is continuing and numerous studies are being conducted to increase the accuracy and effectiveness of AES systems.

6.2. Contributions of the study

The main contributions of this study are summarized as follows:

- To develop AES system large number of corpus is vital. Accordingly, the study prepared large number of Amharic pre-graded essay corpus.
- The general architecture of LSA-based Automatic Amharic essay scoring is proposed
- The study identifies basic challenges in developing Amharic essay scoring systems and the possible strategies to solve those challenges
- As plagiarism is becoming a serious problem in education, this study paves a way for developing automatic plagiarism detection in Amharic for student essays.
- The research critically reviewed the performance evaluation measurement for AES systems and identified that there is no standard benchmark and common measure used to evaluate AES system. Most of the previous system uses correlation analysis as evaluation metric. The research criticized the use of correlation as evaluation metric in AES systems this is because of correlation doesn't measure the true relationship between the system and the human grade rather it only measure the linearity relationship between two variables in this case the human score and the system score.

6.3. Recommendations and Future Work

The research on Amharic essay scoring being the first work did not extensively exploited all the various and ultimate options that exists in order to develop effective automatic essay scoring systems which can achieve the highest performance. Though, the research has investigated in detail these issues and recommends for further work should be conducted to improve the performance of the system through exploiting the following various alternatives.

- The researcher recommends that automatic essay scoring systems should be used in formative assessment role to improve the performance of students' achievements. To achieve this, automatic essay scoring systems need to be equipped with instructional feedback component which enables students to improve their writing skills.
- Though Amharic essay evaluation involves both content and grammar of the essay, the research focused only on the content evaluation. Therefore further research demands the development of the Amharic grammar evaluation as well in order to have viable automatic essay scoring system.
- In our research we have trained the model using pre-graded essays only but other research should also incorporate domain representative texts such as textbook, articles or chapters of a text during learning besides pre-graded essay which helps to acquire high dimensional semantic representation.
- In our research, to predict a score similarity measurement is performed using the standard cosine similarity measure between the query essay and the LSA representation but the researcher is also aware of other similarity measures such as Euclidean measure and dot product. So other research should see the prospect of involving such similarity measurements, and the effect of using different similarity measurements should be experimented to select the best similarity measurement empirically.
- The researcher adopted preprocessing component from previous work and experienced some degradation of the performance particularly on that of stemmer component. So, other researchers should give serious consideration on this issue and work on improving the performance of the stemmer algorithm.
- From several weighting schemes, the research only used local weighing to determine the importance of a term within the document. This is due to the fact that the research used only pre-graded essays for training, and as a result no global weighting is required. But

other researchers should increase the training data to encompass domain representative texts in addition to pre-graded essays and should use different weighting schemes to determine the importance of a term within the whole documents. Moreover, experiments should be conducted to determine the effect of using different weighting scheme on the performance of the automatic essay grading to select the best weighting scheme.

- One of the contributions of this research is to prepare large amount of Amharic essay corpus. In preparing the corpus the research is aware of the subjectivity and biasness nature (Inter-rater agreement) of the manual essay grading process and to counter this problem the research further prepared specific and detail criteria to grade an essay but this by itself does not able to remove the biasness between the two human raters that the research used. So further research should be conducted to solve such problem on preparing essay corpus. The research identified that using more than two human raters and taking essays score in which all human raters agree up on can circumvent biasness but this demands further research.
- The study used the average of 10 weighted neighbors in order to predict the score of a query essay. It further believes that using the average of more than 10 weighted neighbors with the highest cosine similarity to the query should be assessed empirically in order to have best score prediction.
- The research investigated that there is no standard evaluation metric used to evaluate the performance of AES systems [45].Hence, further research is needed to develop standard and reliable measurement to measure the performance of AES.

References:

[1] Miller T., *"Essay Assessment with Latent Semantic Analysis"*, Toronto University, Canada, 2003.

[2] Valenti S. and etal *"An overview of current research on automated essay grading"*, Journal of Information Technology Education, (2), pp. 319–330, 2003.

[3] Landauer T. and Psotka J., *"Simulating Text Understanding for Educational Applications with Latent Semantic Analysis: Introduction to LSA"*, In: Interactive Learning Environments, vol. 8, no. 2, pp. 73-86(14), 2000.

[4] http://cs.joensuu.fi/edtech/mw/saea.php *"Tools for improving work in learning organization"*, last accessed on: August 17, 2009.

[5] www.aclweb.org/anthology/P/P06/P06-1030.pdf. "Automated Japanese Essay Scoring System: Jess", last accessed on: February 27, 2010.

[6] www.aclweb.org/anthology/Y/Y06/Y06-1004.pdf.,*"Enhancing Automatic Chinese Essay Scoring System from Figures-of-Speech"*, last accessed on: January 17, 2010.

[7] Seid Muhie, "Teteyeq (ተጠየቅ): *Amharic Question Answering for Factoid Questions"*, Master's Thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2009.

[8] http://nlp.amharic.org/wiki, *"Amharic Lingua Wiki"*, Last accessed on: July 10, 2009.

[9] http://www.masterpapers.com/blog/essay-writing-tips/essay-on/23 *"Essay on Factual Topics"*, last accessed on: July 14, 2009.

[10] www.rapidintellect.com, *"Educational Assessment"*, Academic Exchange Quarterly, Last accessed on: February 25, 2009

[11] Kendle A. and Northcote M., *"The Struggle for Balance in the Use of Quantitative and Qualitative Online Assessment Tasks"*. Paper presented at the 17th conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE). Coffs Harbour, NSW, 2000.

[12] Macdonald J., *"Assessing online collaborative learning: process and product Computers & Education"*. *40*(4), 377-391, 2003.

[13] Muirhead B., *"Effective Online Assessment Strategies for Today's Colleges and Universities"*, Artist, 2002.

[14] O'Reilly M., *"Hallmarks of Excellence in Online Assessment"*, Paper presented at the First International Conference on Enhancing Teaching and Learning through Assessment, Hong Kong Polytechnic University, 2005.

[15] Leathwood C., *"Assessment policy and practice in higher education: purpose, standards and equity"*, Assessment & Evaluation in Higher Education, *30*(3), 307-324, 2005.

[16] Maclellan E., *"How convincing is alternative assessment for use in higher education?"* Assessment and Evaluation in Higher Education, *29*(3),311-321, 2004.

[17] Taras M., "*Using Assessment for Learning and Learning from Assessment*", Assessment & Evaluation in Higher Education, *27*(6), 501-510, 2002.

[18] Dewey J., *"How we think, Henry Regnery"*. Reprinted in LW.8:105354, 1933.

[19] Bradney A. and Cownie F., *"teaching legal system"*, Warwick: National Centre for Legal Education, 1999.

[20] Berry R., *"Alternative assessment and assessment for learning"*, in Proceedings of the 29th IAEA Conference, theme: Societies Goals and Assessment, 2003.

[21] www.pdkintl.org, *"Inside the Black Box: Raising Standards through Classroom Assessment"*, Last accessed on: February 28, 2009.

[22] Weeden P., Winter J., etal. *"Assessment what is not for School"*, Routledge Falmer 270 Madison Ave, New York, NY 10016: USA and Canada, 2002.

[23] Scriven M., *"Evaluation thesaurus":* 4th ed., Newbury Park, Canada. Sage Publications. ISBN 0-8039-4364-4, 1991.

[24] Earl, Lorna, *"Assessment as Learning: Using Classroom Assessment to Maximize Student Learning"*. Thousand Oaks, CA, Corwin Press. ISBN 0-7619-4626-8, 2003.

[25] Abate A., *"A practical Writing Approach"*, Far-East-Trading P.L.C (P.P): Ethiopia, 2008.

[26] Steele W., *"Essays - well worth the effort"* College Teaching, Vol. 45, No. 4, 1997.

[27] Shermis M. D. & Burstein J., *"Automated Essay Scoring: A cross disciplinary perspective"*. Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[28] Palmer J., Williams R. & Dreher H. *"Automated Essay Grading System Applied to a First Year University Subject – How Can we do it better?"* Curtin University of Technology, Perth, WA. Australia, 2002.

[29] http://www.ets.org/Media/Research/pdf/erater_iaai03_burstein.pdf, *"Criterion SM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays",l*ast accessed on: January 12, 2010.

[30] http://mathsci.math.admu.edu.ph/~raf/pcsc04/papers/e04.pdf, *"Automated Essay Evaluator"*, last accessed on: February 15, 2010.

[31] *www.edpsycinteractive.org/brilstar/units/.../giggey_homework%20rubric.pdf, "Essay Scoring Rubric"*, last accessed on: May 12, 2010.

[32] Attali Y., *"Exploring the feedback and revision features of Criterion"*, Paper presented at the National Council on Measurement in Education (NCME), San Diego, Canada, April 2004.

[33] http://cs.joensuu.fi/pages/edtech, "Automatic Essay Assessor for Finnish Language", last accessed on: March 12, 2010.

[34] Hamp-Lyons, L. "*Fourth generation writing assessment*", In T. Silva and P. K. Matsuda (Eds.), *On Second Language Writing*, (pp. 117–125). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2001.

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

[36] Landauer T. And etal. *"How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans"*, Proceedings of the 19th Annual Conference of the Cognitive Science Society, (pp. 412–417). Mahwah, NJ: Erlbaum, 1997.

[37] Wang J. & Brown M. S., *"Automated essay scoring versus human scoring: A correlation study"*, Contemporary Issues in Technology and Teacher Education, *8*(4), 310-325, 2008.

[38] Page E.B., *"Computer Grading of Student Prose Using Modern Concepts and Software"*, Journal of Experimental Education. (67), 1994.

[39] http://andromeda.rutgers.edu/~gshafer/little.pdf, *"Computerized Evaluation of Essays"*, last accessed on: January 12, 2010.

[40] Salem A. B. M., *"The potential role of artificial intelligence technology in Education"* (ERIC document reproduction service no ED 477 318), 2000.

[41] Burstein J., Kukich K., Wolff S., Lu C. & Chodorow M., *"Computer analysis of essays"*. Proceedings of the NCME Symposium on Automated Scoring. Montreal, Canada, 1998.

[42] Salton G. and Micheal J. McGill. *"Introduction to Modern Information Retrieval"*, New York: McGraw-Hill Book Company, 1983.

[43] Melese Tamiru, *"Automatic Amharic Text Summarization Using Latent Semantic Analysis"*, Master's Thesis, Addis Ababa University, 2009.

[44] Lemaire b., Dessus p**.** *"A System to assess the semantic content of student essays"*. J Educational computing research. Vol. 24(3) 305-320. *University of Grenoble-II*. 2001.

[45] Jana Z. Sukkarieh, Bolge E., *"Building a Textual Entailment Suite for the Evaluation of Automatic Content Scoring Technologies"*, Educational Testing Service Rosedale Road, Princeton, NJ 08541, 2009.

[46] P´erez D., *"Automatic evaluation of users short essays by using statistical and shallow natural language processing"*, Advanced Studies Diploma Work. 2004.

[47] Yang Y., Buckendahl C.W., Juszkiewicz P.J., & Bhola D.S, *"Applied Measurement in Education"* A review of strategies for validating computer automated scoring, *15*(4), 391–412, 2002.

[48] Baron, D. *"The Chronicle of Higher Education"*, The College Board's new essay reverses decades of progress toward literacy, P. B14, 2005.

[49] http://www.psych.nmsu.edu/~pfoltz/reprints/Edmedia99.html, *"Intelligent Essay Assessor*", Last accessed on: February 23, 2010.

[50] Rudner L. M., Garcia V., Welch C., *"An Evaluation of the IntelliMetricSM Essay Scoring System"*. The Journal of Technology, Learning and Assessment, 4(4), 2006.

[51] http://www.pearsonkt.com/papers/ab-CredGrading2002.shtml, *"The Credible Grading Machine: Automated Essay Scoring in the Dod"*, Last accessed on: July 13, 2010.

[52] http://www.knowledge-technologies.com "*Pearson*", Last accessed on March 21, 2010.

[53] Kakkonen T. & etal, *"Comparison of Dimension Reduction Methods for Automated Essay Grading"*, Educational Technology & Society, 11(3), pp 275–288, 2008.

[54] *www.ets.org/Media/Research/pdf/RM-04-01.pdf. "Moving the Field Forward: Some Thoughts on Validity and Automated Scoring"*, last accessed on: March 17, 2010

[55] Tesema Mindaye,"*Design and Implementation of Amharic Search Engine"*, Master's Thesis Addis Ababa University: Addis Ababa, Ethiopia, 2007.

[56] *Brandon P., "Comparison of Human and Latent Semantic Analysis (LSA) Judgments of Pair wise Document Similarities for a News Corpus"*, Intelligence**,** Surveillance and Reconnaissance Division Information Sciences Laboratory, *DSTO Information Sciences Laboratory, Edinburgh South Australia, 2004.*

ANNEXES

Annex A: Manual Essay Scoring Criteria for Amharic Literature Department, Addis Ababa University.

የጽሀፈት ችሎታ መመዘኛ መስፈርቶች

ነጥብ	አደረጃጀት	ማስፋፊያ	የአ. ነ አወቃቀር	የቃላት ምርጫ	ስርአተ ነጥብ
6	ውጤታማ አደረጃጀት . ውህደት (ወጥነት ግጥምጥምነት) .ጠንካራ አጀማመር፣ አካሄድና አጨራረስ መኖር . የመሸጋገርያ ሀረጋት መኖር .የአንቀጽ አከፋፈል ተገቢነት	. ውጤታማ ማስፋፊያ .የማስረጃዎች የምሳሌዎች ጥንካሬ፣ የዝርዝርና ደጋፊ ማስረጃዎች ብቃት .የትኩረት ግልጽነትን መጠበቅ	. ውጤታማ የ አ. ነ ቅንብር . አረፍተ ነገሮች የራሳቸው ብቃትና ስብጥር መኖር	. ውጤታማ የቃላት አመራረጥ ብቃትና አጠቃቀም . ግልጽነት . ቁጥብ . ቀጭ	. ውጤታማ የ ስ ነ አጠቃቀም . በስርአተ ነጥብ አጠቃቀም ረገድ የሚታዩ ጉድለቶች ከቁጥር የማይገቡበት
5	ጥሩ አደረጃጀት . የሀሳብ ተጠየቃዊ እድገት . ጥሩ አጀማመር፣ አካሄድና አጨራረስ . አጥጋቢ የመሸጋገርያ ሀረጎች መኖር .ጥሩ የአንቀጽ አከፋፈል	. ጥሩ ማስፋፊያ . የማስረጃዎች አጠቃቀም፣ ግልጽነትና የሌሎች ማብራሪያዎች መኖር . ግልጽ ትኩረት	. ጥሩ የአ.ነ ቅንብርና ቁጥጥር ትክክለኛና የተሟላ አ. ነ መሟላት . የ አ. ነ አይነትና ስብጥር	. ጥሩ የቃላት ምርጫ . ቁጥብና ግልጽ	. ጥሩ የስ.ነ አጠቃቀም .በ ስ.ነ አጠቃቀም ረገድ መለስተኛ ጉድለቶች
4	. ልከኛ አደረጃጀት .የሀሳብ ተጠየቃዊነትና ተከታታይነት ልከኛ ስለመሆኑ አንዳንድ መረጃዎች መኖር . መሸጋገሪያዎቹ መኖር . አንቀጽ መኖር	. ልከኛ ማስፋፊያ . ልከኛ ምሳሌዎች፣ ማስረጃዎችና ዝርዝሮች መኖር . ዘላቂ ትኩረት ላይኖር ይችላል	. ልከኛ የ አ.ነ ቅንብርና ቁጥጥር . የ አ.ነ አወቃቀር መለስተኛ መሆን .የ አ.ነ ስብጥር ጉድለት መኖር	. ልከኛ የቃላት ምርጫና አጠቃቀም . ተገቢነት ማነስ .ቀላልነትና ልልነት	. ልከኛ የስ.ነ አጠቃቀም .በስርአተ ነጥብ አጠቃቀም ረገድ ጥቂት ጉድለት መኖር
3	. ውስን አደረጃጀት .ያየአደረጃጀት ግልጽነት አለመኖር .አጀማመር አካሄድና አጨራረስ ላይ ጉድለት መኖር .የመሸጋገሪያዎች አጠቃቀም ውስንነት .አነስተኛ የአንቀጽ አከፋፈል መኖር	. ውስን ማስፋፊያ .የምሳሌዎች የማስረጃዎችና የአንዳንድ ዝርዝሮች አጠቃቀም ውስን መሆን .የትኩረት ውስን መሆን	. ውስን የ አ.ነ ቅንብርና ቁጥጥር . በ አ.ነ አወቃቀር አንዳንድ ስህተት መኖር	. ውስን የቃላት አጠቃቀም .መዋዘቅ . መደጋገም	. ውስን የስ.ነ አጠቃቀም .ተደጋጋሚ ስህተት መኖር
2	. አነስተኛ አደረጃጀት .ያደረጃጀት ጉድለት .ግልጽ የሽግግር ጉድለት	. አነስተኛ ማስፋፊያ .የአጥጋቢ ምሳሌዎች ማስረጃዎችና ዝርዝሮች ጉድለት . የትኩረት ደብዛዛነት	. አነስተኛ የ አ.ነ ቅንብርና ቁጥጥር .ያላለቁ አ.ነ መኖር .ገስጋሽ ወይም የማያቋርጡ አ.ነ መኖር	. አነስተኛ የቃላት አመራረጥና አጠቃቀም .የመምረጥ ድክመት .ደጋጋሚነት	. አነስተኛ የስ.ነ አጠቃቀም .ተደጋጋሚ ያጠቃቀም ጉድለት
1	. ዝብርቅርቅ ደረጃጀት .ለከትትል አስቸጋሪ የሆነ .ምንም መሸጋገሪያ የሌለው	. ከበቂ በታች ማስፋፊያ .ምሳሌዎች ማስረጃዎችና ዝርዝሮች የሌሉት .ትኩረት አልባ	. የ አ.ነ ትክክለኛነት ጉድለት .ያልተቋጨና ተከታታይ አ.ነ መብዛት .ገስጋሽ አ.ነ መብዛት	. ከተገቢ በታች የቃላት አመራረጥና አጠቃቀም መኖር .አ. ተገቢነት .ስህተት	. ከበቂ በታች የስ.ነ አጠቃቀም .በስርአተ ነጥብ አጠቃቀም ረገድ የምርጫ ችግር ያለበት
0	. የማይታረም	. ሀሳቡ የማይገኝ	. ድርሰቱ በአጠቃላይ ስለምን እንደሆነ የማይነገር		

Annex B: Sample Essay Exam

<u>የአማርኛ ፅሁፍ ፈተና:</u>

መመሪያ:- በሚከተሉት ርዕሰ ጉዳዮች ላይ ከ100-150 ቃላት ማብራርያ ፃፍ/ፃፊ።

ሙሉ ስም:-_____ዲፓርትመንት:-_____

ጥያቄ 1- ሰለ አደንዛገር ዕጾች ምንነትና ስርጭት አንዲሁም ስለሚያደርሱት ማሕበራዊና ኢኮኖሚያዊ ቀዉስ በአጭሩ ግለፁ?

ጥያቄ 2፡- የንግድ ግንኙነት ማን ማንን ያካትታል? ግንኙነታቸውስ እንዴት ይገለፃል?

ጥያቄ 3፡ የአለም የመጓጓጉ ዘዴ ምንን ያካትታል? ሰለደረሰበት የእድገት ደረጃም በአጭሩ ግለፁ?

Annex C: Sample Essay Response Exams

Domain 1: *Essay on Drugs*

1. አደንዛዥ ዕጽ ማለት አንድ ሰው ተፈጥሮ በሰጠችው የማሰብና የማሰላሰል ችሎታው እንዳይጠቀም መሰናክል የሚሆን የመድኃኒት አይነት ሲሆን ስርጭቱ በህገ ወጥ መንገድ ከተለያዩ ሀገሮች ወደ ሀገር ውስጥ የሚገባና እንዲሁም በሀገር ውስጥ በተለያዩ አካባቢዎች እጾቹ ተተክለው ለተለያዩ አካባቢዎች በህገወጥ ነጋዴዎች አማካኝነት ይሰፋፋሉ፡፡ እነዚህ እጾች የሚያደርሱትን ማህበራዊ ጉዳት ስንመለከት ደግሞ የአደንዛዥ ዕጽ ተጠቃሚ ሰዎች ብዙ ጊዜ በትክክለኛው አእምሮአቸው ማሰብ ስለማይችሉ ብቸኝነትን ይመርጣሉ መጠጊያቸውን መጠጥ ቤት ያደርጋሉ በዚህም ተገፋፍተው ልቅ የሆነ የወሲብ ግንኙነት ለመፈጸም ለተለያዩ የአባላዘር በሽታዎች ይጋለጣሉ፡፡ በዚህ ምክንያት ቤትና ልጆች ይበተናሉ፡፡ በአጠቃላይ ለተለያዩ ማህበራዊ ቀውስ ይዳረጋሉ፡፡ ኢኮኖምያዊ ሁኔታውን ስንመለከት ደግሞ በአብዛኛው አደንዛዥ ዕጾች በህገ ወጥ ነጋዴ ስለሚመጡ ዋጋቸው ትንሽ ወደድ ይላል በዚህም አንድ ሰው ወርሃዊ በጀቱን ሳይጠቀም ለዕፅ መግዣነት በማዋል የኢኮኖሚ ሁኔታ ይዛባበታል፡፡

2. አደንዛዥ ዕጾች አዕምሮንና አካልን ከመደበኛው በላይ የሚያነቃቁ ነገሮች ናቸው አደንዛዥ ዕጽ የሚለው ስያሜም የተሰጣቸው አንድ ሰው በተደጋጋሚ ከተጠቀመባቸው በኋላ ሱስ ስለሚሆኑና ያለ ዕፁ አነቃቂነት መስራት ስለሚያቅም ወይም ስለሚደነዝዝ ነው ከስማቸው እንደምንረዳው አደንዛዥ ዕጾችየዕፅዋት ውጤቶች ናቸው ለምሳሌ ሄሮይን ማሪዋናና የመሳሰሉትን መጥቀስ ይቻላል እነዚህ ዕፆች በይፋ ጥቅም ላይ ስለማይውሉ በህግ የተከለከለና ሲያዘዋውር የተገኘም ህጋዊ እርምጃ ስለሚወሰድበት ሁልግዜም ወደተጠቃሚዎቹ የሚደርሱት በጣም ስውር በሆነ የስርጭት ዘዴ ነው የሚበቅሉትና የሚዘጋጁት ገጠርማ በሆኑ አካባቢዎች ሲሆን ጥቅም ላይ የሚውሉት ግን በከተማ ነዋሪ ወጣቶች ነው እነዚህ አደንዛዥ ዕጾች በሰዎች ላይ አካላዊ ስነልቦናዊና ኢኮኖሚያዊ ጉዳቶችን ያደርሳሉ ለምሳሌ አንድ የሱስ ሰለባ የሆነ ወጣት ዕፁን የሚገዛበት ገንዘብ ሲያጣ ሊሰርቅ ወይም ሊቀማ አልያም ደግሞ ሊለምን ይችላል ይህ ደግሞ ለእስራት ለስነልቦና ችግር ወይም በማህበረሰቡ መገለልን ሊያደርስበት ይችላል፡፡

3. አደንዛዥ እጽ ማለት አዕምሮን አደንዝዞ የማይገባ ስራን እንድንሰራ የሚያደርግና ለከፍተኛ የጤና መታወክ የሚያጋልጥ ዕጽ ነው፡፡ በአሁኑ ጊዜ በመላ ሀገሪቷ በድብቅ ህገ ወጥ በሆነ መንገድ በከፍተኛ ሁኔታ በመሰራጨት ላይ ይገኛል፡፡ በማህበራዊ ህይወት ላይ ስንመለከት ደግሞ አደንዛዥ ዕጽ የሚጠቀም ሰው ከሰው ጋር ተስማምቶ መኖር አይችልም ምክንያቱም እጹን ሲጠቀም ደፋር፣ ተደባዳቢና የማይገባ ስራን ስለሚያሰራ ከህብርተሰቡ ጋር ለመስማማት ይቸገራሉ በተጨማሪም ያንን የተለመደውን ካጡ ተነጫናጭ እና ድብርታም ስለሚያደርጋቸው ለብቸኝነት ያጋልጣቸዋል፡፡ ህብረተሰቡም በነዚህ ሁለት ምክንያቶች የእጹን ተጠቃሚዎች ያገላቸዋል፡፡ በኢኮኖሚ ረገድ ስንመለከት እነዚህ ዕጽ ተጠቃሚ ሰዎች ትምህርታቸውን ለማቋረጥ ይገደዳሉ እንደምንም ትምህርታቸውን ጨርሰው ስራ ቢይዙ እንኳ የሚያገኙትን ገንዘብ ሱሳቸውን ለማሟላት ሲሉ ስለሚያጠፉ ከቤተሰብ ትከሻ ላይ ሳይወርዱ ቤተሰብ በመጦር ፈንታ የተገላቢጦሽ ቤተሰብ እነሱን ስለሚጦር ኢኮኖሚያዊ ቀውስን ያመጣሉ፡፡ በቤተሰባቸው ላይ ብቻ ሳይሆን በሀገሪቱ ኢኮኖሚ ላይም ችግር ያመጣል፡፡ ምክንያቱም የዚህ ሱስ ሰለባ የሚሆነው በአብዛኛው ወጣቱ ህብረተሰብ ነው ወጣቱ ህብረሰተሰብ ደግሞ አምራች ዜጋ የምንለው ስለሆነ ወጣቱ በዚህ አይነቱ ሱስ ከተያዘ አምራች ዜጋ ስለሚጠፋ ሀገሪቱ በኢኮኖሚ ረገድ ተጎጂ ትሆናለች፡፡

Domain 2: *Essay on Trade*

1. የንግድ ግንኙነት ነጋዴ፣ ሸማች (ገዢ) እና መንግስትን ያካትታል ግንኙነታቸውም የሚገለጸው አንድ ነጋዴ የሚሸጠው እቃ ለማትረፍ ነው፡፡ ይህ ደግሞ ሸማች (ገዢ) ለመግዛትና አንዳንድ ነገሮችን በቅናሽ እንዲገዛ ያደርገዋል፡፡ መንግስት ደግሞ ነጋዴውንና ገዢ በሚሸጡበትና በሚገዙበት ሰአት መንግስት ቀረጥ (ገቢ) ከነጋዴውም ከገዢውም ያገኛል፡፡ በተጨማሪም መንግስት በነጋዴዎች ላይ ጥብቅ ቁጥጥር የሚያደርገው ከአቅም በላይ የሚሸጡትን ነገሮች ተከታትሎ በተመጣጣኝ ዋጋ እንዲሸጡና ገዢውም እንዲጠቀም ያደርጋል በአጠቃላይ ነጋዴ፣ ገዢና መንግስት በነዚህ አይነት መንገዶች ስራቸውን ሊያከናውኑ ይችላሉ፡፡

2. የንግድ ግንኙነት ገዢና ሻጭን ያካተተ ሲሆን ሁለቱንም ባማከለ ደረጃ የሚቆጣጠር ደግሞ መንግስት አለ፡፡ በዚህ ሁኔታ ሻጭ በሚሰራው ስራ አስፈላጊውን ገቢ እንዲያገኝና ገዢም ተመጣጣኝ በሆነ ሸያጭ ግዢውን ያደርጋል፡፡ በነዚህ ሁኔታ መንግስት በተለያዩ ሀገራዊ ስራዎች ከነጋዴዎች የስራ ውጤት ተገቢውን ቀረጥ ያገኛል፡፡ በአንዳንድ ሁኔታዎች ግን ሻጭና ገዢ እንዲሁም መንግስት የሚጎዳዱበት መንገድ አለ፡፡ ማለትም ሻጮች ያለአግባብ ዋጋ በመጨመርና ጥራት የሌላቸውን የምርት ውጤቶች ለገዢው በማቅረብ የማጭበርበር ስራን ሲሰሩ ይታያል፡፡ መንግስትም ይህንን ስራ በአግባብ ያለመቆጣጠርና ከመጠን ያለፈ የግብር ክፍያን ሲጭን ይታያል፡፡

3. ሸማች ሻጭና አምራችን በዋናነት የሚይዝ ሲሆን መንግስት ደግሞ በሶስቱ መካከል ሆኖ የሚሰራው ተግባር ይኖረዋል አምራች የሻጭና የሸማቹን ፍላጎት እንዲሁም የሀገሪቱን አመታዊ የግለሰብ ገቢ በማስላት ከማህበረሰቡ የኪስ ገቢ ጋር ተመጣጣኝ የሆኑ ምርቶችን በማምረት ለሻጮች ማስረከብ ወይም በቀጥተኛ ግንኙነት ለሸማቹ ህብረተሰብ እንዲደርስ ማድረግ፣ ሻጭ ደግሞ የሸማቹን ፍላጎት ከእለት ገቢው ጋር በማጣጣም ሸማቹ በቅርብ ሊያገኝበት በሚችለው ሁኔታ ተፈላጊውን ምርት ማቅረብ ሲሆን፣ ሸማች ደግሞ ፍላጎቱ ከእለት ተእለት ገቢው ጋር በማጣጣም የራሱንም ሆነ የሀገሪቱን ኢኮኖሚ በሚያናጋ መልኩ ምርቶችን ተገቢ ፍቃድ ካላቸው ሻጮች/ነጋዴዎች በመግዛት ነው፡፡ ነገር ግን መንግስት እነዚህን አካላት በጣልቃ ገብነት መቆጣጠር ይችላል፡፡ ይህም ነጋዴዎች ከአምራቾች ያገኙትን ምርት ከተገቢ በላይ ወይም በመንግስት ፍቃድ ከተተመነው ዋጋ ጨምረው በመሸጥ የኑሮ ውድነት እንዲኖር ሊያደርጉ ስለሚችሉ በተቻለ መጠን ቢቆጣጠሩ ካልሆነ ደግሞ የራሱ የሆነ ሌሎች ችግሩን እንደ ችግር ሊያቆሙ የሚችሉ መፍትሔዎችን ማለትም በቀላል ዋጋ ሊገኙ የሚችልባቸውን ነገሮች መንግስታዊ ሱቆች በመክፈት ለህብረተሰቡ ፍላጎቱንና ገቢውን ያገናዘበ አቅራቦት ማድረግ ነው፡፡

Domain 3: *Essay on Transportations*

1. የአለም የመጓጓዣ ዘዴ ብዙ ነገርን ያካትታል እነዚህንም ዘዴዎች ባህላዊና ዘመናዊ ብለን ልንከፍላቸው እንችላለን ከባህላዊ ዘዴዎች መካከል እንስሳት ለምሳሌ ፈረስ በቅሎ አህያ ግመል ወዘተ ዕቃን ወይም የሰውን ልጅ ከአንድ ቦታ ወደ ሌላ ቦታ ለማጓጓዝ አይነተኛ ዘዴዎች እንደሆኑ ይጠቀሳሉ፡፡ ዘመናዊ የመጓጓዣ ዘዴዎች ብለን የምንጠራቸው ደግሞ መኪና መርከብ አውሮፕላን ባቡር ሞተር ሳይክል ወዘተ ናቸው እነዚህ ከባህላዊ የመጓጓዣ ዘዴ የሚለያቸው በጣም ፈጣን መሆናቸው ነው ሌለው ባህላዊ የመጓጓዣ ዘዴዎች በየብስና በውሀ አካላት ላይ ብቻ ሲሆን የሚጠቅሙት ዘመናዊዎቹ ከዚህ በተጨማሪ በአየር ላይ ጉዞ በማድረግ አይነተኛ ሚና ይጫወታሉ እናም የአለም የመጓጓዣ ዘዴ ዛሬ የደረሰበት ደረጃ በጣም የሚያስደንቅ ነው፡፡ በተለይም ዘመናዊ አውሮፕላኖች በሰአታት ውስጥ ከአንድ ሀገር ወደ ሌላ ሀገር እንዲሁም ከአንድ አህጉር ወደ ሌላ አህጉር በመብረር የሚያደርጉት ፍጥነት በጣም አስገራሚ ነው ቀን በቀን የሚፈበረኩ የየተለያዩ የመኪና ሞዴሎች በምቾታቸውም ሆነ በፍጥነታቸው በጣም እያደጉ ነው በአጠቃላይ አለማችን የደረሰችበት የመጓጓዣ ዘዴ አጠቃቀም በሰው ልጅ የኑሮ ሁኔታ ላይ ይህ ነው የማይባል ለውጥ እያሳየ ነው፡፡

2. የመጓጓዣ ዘዴ ሰው ከቦታ ቦታ መዘዋወር ከጀመረ እስከ አሁን ድረስ የቀጠለ ዘዴ ነው፡፡ ስልጣኔ እንደ አሁኑ ባልተስፋፋበት ዘመን የሰው ልጅ ከቦታ ቦታ የሚዘዋወረው በእግሩና በጋማ ከብቶች በመጓጓዝ ነበረ ከጋማ ከብቶች በተለይ በአህያ በፈረስና በበቅሎ እንዲሁም በግመልና ዝሆን ለረጅም ዘመናት ሲገለገልባቸው ቆይቷል አሁን ስልጣኔ በተስፋፋበት ዘመን ደግሞ ሰው በተሽከርካሪ በመርከብና በአውሮፕላን ከቦታ ቦታ ይዘዋወራል በየብስ ከቀላል የቤት መኪና እስከ ከባድ የጭነት ተሽከርካሪዎችን ስንጠቀም በባህር ደግሞ ከትንንሽ ጀልባዎች እስከ ግዙፍ መርከቦች ድረስ ይጠቀማል በአሁኑ ሰአት የመጓጓዣ ዘዴ ከእለት ወደ እለት የዘመነ በመምጣት ላይ ይገኛል በጣም ፈጣን ባቡሮችና አውሮፕላኖችም ለመጓጓዣነት ይውላሉ በአጠቃላይ የመጓጓዣ አገልግሎቱ እንስሳትን ከመጠቀም ተነስቶ አሁን አውሮፕላን ላይ ተደርሷል ይደረሳል ተብሎ የማይታሰበው ቦታ በተወሰነ ሰአትም መድረስ ተችሏል፡፡

3. በአለም ውስጥ ብዙ አይነት የመጓጓዣ ዘዴዎች ይገኛሉ፡፡ በድሮ ግዜ የነበረና እስከ አሁን ያለው ባህላዊ መጓጓዣ ነው ይኸውም የአህያ የበቅሎ የፈረስ እና የመሳሰሉት ባህላዊ መጓጓዣዎችን ያካትታል፡፡ ሌላው ደግሞ ዘመናዊ የመጓጓዣ አይነት ሲሆን በዚህ የሚካተቱትም መኪና፣ አውሮፕላን መርከብ፣ ጀልባ፣ መንኩራኩር እና የመሳሰሉት ናቸው፡፡ በአለም ላይ የመጓጓዣ ዘዴ ከጊዜ ወደ ጊዜ ፈጣን እድገት እያሳየ እና በጣም እየተራቀቀ መጥቷል፡፡ ይኸውም ከባህላዊ መጓጓዣ ወደ ዘመናዊ መጓጓዣ እየተሻሻለ የመጣ ነው፡፡ ከእንስሳት መጓጓዣ ወደ ሞተር መለወጡ ነው፡፡ በሞተር መጓጓዣዎች ላይም የተለያየ ለውጥ እየመጣ መጥቷል፡፡ ለምሳሌ ስናይ ከአውሮፕላን የበለጠ መጓጓዣ ማለትም ፕላኔትን ለማየት የሚያስችል መንኩራኩር ተሰርቷል፡፡ ይህ የሚያሳየው መጓጓዣ ፈጣን እድገት ላይ መድረሱን ነው፡፡

Annex D: Manual Essay Scoring Criteria

<table>
<tr><td>Question Number</td><td colspan="9">Criteria (10 points)</td></tr>
<tr><td rowspan="2"></td><td colspan="3">Unity (2 points)</td><td colspan="3">Coherence (2 points)</td><td colspan="2">Completeness(6 points)</td></tr>
<tr><td>Well unified (2 pt)</td><td>Unified (1pt)</td><td>Not unified (0 pt)</td><td>Well coherent (2 pt)</td><td>Coherent (1 pt)</td><td>Not coherent (0 pt)</td><td colspan="2"></td></tr>
<tr><td rowspan="6">1</td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td colspan="2">Definition (2pt)</td></tr>
<tr><td colspan="2">Transmission (2 pt)</td></tr>
<tr><td rowspan="2">Effects (2pt)</td><td>Social</td></tr>
<tr><td>economical</td></tr>
<tr><td rowspan="3">What and who (3pt)</td><td>Merchants</td></tr>
<tr><td>Customers</td></tr>
<tr><td>Government</td></tr>
<tr><td rowspan="6">2</td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td rowspan="6"></td><td colspan="2">Trade Relationships (3pt)</td></tr>
<tr><td rowspan="5">Types (4 pt)</td><td>Animal</td></tr>
<tr><td>Land</td></tr>
<tr><td>Sea</td></tr>
<tr><td>Air</td></tr>
<tr><td colspan="2">Level of development (2pt)</td></tr>
</table>

Annex E: Full Experiment of Domain 1

Where: θ_i the cosine similarity angle of i where i ranges from 1 to 10.
SG_i is the calculated score of i nearest neighbor pre-graded essays and MG is the manual grade.

	θ_1	SG1	θ_2	SG2	θ_3	SG3	θ_4	SG4	θ_5	SG5	θ_6	SG6	θ_7	SG7	θ_8	SG8	θ_9	SG9	θ_{10}	SG10	MG
t1_1	66.17	5.50	66.87	5.25	68.63	5.49	69.02	5.61	69.17	5.68	69.17	6.05	69.69	5.98	69.81	6.28	69.82	6.30	69.90	6.318	6.000
t1_2	67.81	3.00	68.34	4.24	68.96	4.48	69.56	4.37	69.75	4.59	69.82	4.89	69.83	5.05	69.87	5.04	70.29	4.93	70.44	5.319	4.500
t1_3	67.12	6.00	68.38	6.00	69.14	6.80	69.34	6.49	69.51	6.49	70.20	6.10	70.23	5.68	70.47	5.60	70.49	5.53	70.55	5.865	6.000
t1_4	69.81	4.00	70.21	3.50	70.38	4.33	70.54	4.74	71.01	5.37	71.09	5.47	71.17	5.34	71.43	5.30	72.07	5.53	72.12	5.807	5.500
t1_5	69.84	6.50	70.71	5.28	70.90	5.51	70.93	5.63	72.26	5.98	72.91	5.83	73.36	6.11	73.38	5.87	73.46	5.88	73.77	6.940	7.000
t1_6	65.56	5.00	68.28	5.47	68.40	5.80	69.36	5.38	69.39	5.59	69.46	5.34	69.48	5.36	69.55	5.56	69.65	5.60	69.92	5.547	7.000
t1_7	64.74	5.50	65.56	4.76	66.54	5.00	66.58	4.76	66.98	4.80	67.00	4.92	67.30	4.79	67.58	5.05	67.85	4.84	68.00	4.757	5.000
t1_8	73.18	5.00	74.01	6.22	74.42	6.63	75.40	6.49	75.76	6.03	76.18	6.33	76.38	6.22	77.15	5.98	77.15	5.93	77.24	5.935	5.000
t1_9	66.39	4.00	67.39	5.47	68.70	5.95	69.22	5.73	69.46	5.50	69.53	5.26	69.69	5.50	69.93	5.61	69.96	5.66	70.02	5.879	5.000
t1_10	67.24	6.00	68.17	5.51	68.69	5.83	69.02	6.11	69.96	5.72	70.33	5.84	70.42	5.59	70.79	5.64	70.87	5.57	70.94	5.472	7.000
t1_11	66.79	5.00	67.69	4.51	67.84	4.34	67.88	4.26	67.89	4.31	67.97	4.26	67.98	4.43	68.66	4.68	68.86	4.82	68.88	4.79	5.50
t1_12	65.91	6.50	68.01	6.50	68.02	5.69	68.75	5.41	68.80	5.52	69.19	5.68	69.32	5.58	69.68	5.75	69.69	5.94	69.82	5.90	6.50
t1_13	64.74	5.50	65.42	4.76	65.95	5.00	66.11	5.61	66.27	5.30	66.83	4.93	67.23	4.80	67.28	5.01	67.57	5.11	67.69	5.00	5.00
t1_14	69.23	6.00	69.81	5.01	69.88	5.34	70.30	5.13	70.35	4.91	70.68	5.49	70.86	5.29	70.89	5.37	70.91	5.39	71.04	5.35	5.50
t1_15	67.12	6.00	67.65	6.00	67.67	6.83	67.88	6.50	68.41	6.50	68.46	6.58	68.59	6.22	68.95	6.26	69.12	6.01	69.28	5.82	6.50
t1_16	64.64	4.00	65.43	4.00	66.24	4.16	66.40	4.85	66.51	4.98	66.67	5.31	67.15	5.13	67.50	5.11	67.50	5.26	67.93	5.23	6.00
t1_17	66.62	6.00	67.53	6.25	68.20	5.52	68.44	5.64	68.87	5.32	69.17	5.35	69.25	5.37	69.45	5.21	69.56	5.24	69.67	5.45	5.00
t1_18	66.61	5.50	67.01	5.00	67.15	5.50	67.45	5.62	67.57	6.09	68.41	5.76	68.63	5.52	68.82	5.51	68.92	5.57	69.04	5.42	4.00
t1_19	72.28	6.00	72.50	6.00	73.28	5.84	73.38	6.36	74.35	6.48	74.42	6.48	75.06	6.49	75.24	6.26	75.32	6.23	75.32	6.40	6.00
t1_20	71.13	8.00	72.56	7.04	72.89	6.87	73.11	6.78	73.14	6.24	73.16	6.53	73.26	6.38	73.55	6.10	73.67	6.09	73.74	6.17	6.50
t2_1	67.51	5.00	69.39	5.24	69.76	4.84	70.30	5.24	70.49	5.48	70.54	5.88	70.79	5.83	71.03	5.79	71.04	5.87	71.20	5.78	5.00
t2_2	66.11	6.00	66.11	5.00	67.53	5.48	68.13	5.49	68.27	5.30	68.58	5.33	68.80	5.49	68.87	5.43	69.04	5.49	69.14	5.44	6.00
t2_3	68.22	6.00	68.39	7.25	68.49	6.83	68.77	6.38	69.88	6.02	70.05	6.02	70.21	6.29	70.34	6.14	70.70	6.07	70.83	6.11	7.00
t2_4	68.10	6.50	68.22	6.75	68.23	6.67	68.99	6.63	69.18	6.41	69.21	6.42	69.23	6.22	69.33	6.20	69.38	6.07	69.49	6.11	6.50
t2_5	66.02	4.00	66.16	4.25	66.93	4.49	68.64	4.15	68.75	4.40	69.78	4.57	69.98	4.82	70.29	5.13	70.45	5.22	70.50	5.29	5.00
t2_6	65.53	6.50	67.88	4.83	67.99	5.85	68.15	6.01	68.23	6.10	68.46	6.17	68.47	6.28	68.54	6.31	68.93	6.28	69.51	6.20	5.00
t2_7	66.50	5.50	68.04	5.74	68.05	4.85	68.37	4.89	68.78	4.52	68.95	4.44	69.16	4.38	69.49	4.75	69.69	4.72	69.70	4.56	6.00
t2_8	64.66	5.50	66.17	5.74	66.99	5.67	67.20	5.63	67.54	5.51	67.57	5.67	67.66	5.50	67.76	5.20	67.94	5.07	68.20	5.11	6.00

t2_9	68.70	5.00	69.93	6.70	69.99	6.63	70.05	6.60	70.27	6.09	70.28	6.49	70.66	6.35	70.75	6.13	70.98	6.11	71.04	5.91	5.50
t2_10	66.38	5.50	67.15	6.24	67.22	5.83	67.53	5.99	67.67	6.29	67.84	6.24	67.95	6.35	68.09	6.31	68.51	6.33	68.61	6.10	6.00
t2_11	62.70	5.00	62.75	5.25	63.63	5.17	63.99	5.37	64.29	5.30	64.80	5.33	65.21	5.15	65.28	5.37	65.30	5.28	65.52	5.06	4.00
t2_12	63.44	5.50	63.85	5.75	65.66	5.19	65.69	5.51	66.24	5.41	66.38	5.50	66.40	5.64	66.43	5.62	66.45	5.56	66.48	5.41	5.50
t2_13	64.87	8.00	65.47	8.25	65.59	7.67	66.17	7.02	66.22	6.82	66.23	6.36	66.61	6.24	66.84	6.21	67.24	5.87	67.25	5.98	6.00
t2_14	68.40	4.50	69.17	4.75	69.39	4.17	69.50	4.62	69.64	4.99	70.02	4.83	70.46	4.99	70.51	5.35	70.56	5.42	70.69	5.28	5.00
t2_15	60.79	3.00	61.95	3.74	63.34	4.94	63.53	4.71	64.41	4.58	64.66	4.88	65.32	4.96	65.46	5.08	65.88	5.08	65.93	5.07	5.50
t2_16	66.90	4.00	66.96	5.00	68.21	5.00	68.24	5.00	68.60	5.10	69.21	5.24	69.48	4.94	69.95	5.00	70.11	4.90	70.19	4.95	5.00
t2_17	62.75	4.50	64.27	4.74	64.83	5.15	65.23	4.87	65.25	5.09	65.30	5.32	66.06	5.34	66.53	5.36	66.63	5.32	66.91	5.39	4.00
t2_18	70.44	4.00	70.95	4.74	72.17	6.09	72.75	6.07	72.85	5.96	73.31	6.04	73.69	6.10	73.77	5.86	73.93	5.77	74.21	5.98	5.00
t2_19	71.31	8.00	72.15	8.49	72.36	7.35	72.38	7.14	72.39	6.52	72.83	6.52	72.83	6.38	73.08	6.33	73.24	6.40	73.27	6.41	8.00
t2_20	67.40	6.50	67.68	7.25	69.05	7.49	70.13	7.26	70.53	7.12	70.58	7.33	70.87	6.89	71.53	6.84	71.54	6.76	71.66	6.78	7.00
t3_1	66.52	4.00	68.04	5.45	68.04	5.30	69.03	5.23	69.21	5.28	69.26	5.24	69.26	5.48	69.27	5.60	69.46	5.65	69.50	5.49	6.50
t3_2	71.90	6.50	72.14	6.00	72.83	5.84	73.08	5.88	73.18	6.00	73.35	5.68	73.41	5.86	73.46	6.12	74.02	6.11	74.02	6.00	9.00
t3_3	65.80	5.50	66.07	5.75	66.66	6.16	67.44	6.24	67.58	6.29	67.87	5.76	67.87	5.73	67.98	5.52	68.28	5.57	68.29	5.66	8.00
t3_4	73.18	5.00	74.48	4.04	74.61	5.32	75.07	5.24	75.57	5.29	75.66	5.48	75.87	5.42	76.30	5.48	76.88	5.44	76.89	5.58	5.00
t3_5	67.52	6.50	67.87	5.76	68.53	5.18	69.08	5.38	69.15	5.40	69.19	5.42	69.68	5.36	69.71	5.50	69.86	5.45	69.91	5.40	5.50
t3_6	70.03	4.00	70.85	4.49	70.89	4.82	71.87	5.34	72.07	5.85	72.10	5.63	72.36	5.41	72.39	5.48	72.44	5.70	72.59	5.96	7.00
t3_7	68.76	5.50	69.75	6.48	69.81	6.32	70.06	6.00	70.56	5.61	70.66	5.51	70.70	5.58	70.75	5.39	70.98	5.24	71.23	5.12	7.00
t3_8	69.82	8.50	69.99	7.25	71.00	7.01	71.41	6.77	71.54	6.81	71.93	7.00	71.94	7.28	72.19	7.36	72.26	7.27	72.42	7.10	7.50
t3_9	68.03	6.00	68.06	6.50	68.41	5.84	68.42	5.38	68.61	5.11	68.71	5.34	68.77	5.29	68.83	5.32	68.83	5.23	69.15	5.21	5.00
t3_10	68.92	6.00	69.69	4.53	69.75	4.68	70.65	4.76	70.86	5.19	70.92	5.00	71.44	5.40	71.46	5.53	71.52	5.79	71.54	5.72	7.50
t3_11	67.40	5.00	68.00	5.74	68.53	6.15	68.79	5.63	68.92	5.70	69.11	5.67	69.32	5.58	69.55	5.57	69.63	5.51	69.89	5.55	4.00
t3_12	67.59	5.00	67.72	6.25	68.07	6.33	68.38	5.51	68.87	5.70	69.38	5.43	69.72	5.85	69.77	5.51	69.84	5.56	69.90	5.51	6.50
t3_13	67.86	5.50	68.08	4.26	68.24	4.34	68.73	4.25	69.29	4.59	69.58	4.97	70.03	5.32	70.10	5.40	70.25	5.25	70.45	5.32	6.00
t3_14	69.08	4.00	69.69	4.49	69.77	4.99	69.95	4.75	70.23	5.09	70.26	5.07	70.33	4.78	70.54	4.87	70.57	4.88	70.70	4.90	5.50
t3_15	69.67	7.50	72.49	7.04	72.57	6.08	72.58	6.18	72.72	5.76	72.73	5.48	72.74	5.62	72.80	5.66	72.83	5.59	72.85	5.58	7.00
t3_16	69.68	5.00	71.43	5.24	71.55	6.13	71.61	6.10	71.95	6.17	71.96	6.07	71.97	6.13	72.03	5.87	72.07	5.94	72.24	6.04	8.00
t3_17	72.21	4.00	72.77	4.49	73.03	5.80	73.11	5.73	73.34	6.46	73.40	6.31	73.59	6.19	73.64	6.29	73.67	6.37	73.94	6.14	6.00
t3_18	70.46	7.00	70.56	8.25	70.72	7.34	70.73	6.51	71.28	6.51	71.30	6.10	71.60	6.08	71.72	6.13	72.07	6.12	72.25	6.16	9.00
t3_19	69.32	5.00	69.65	5.74	70.30	5.83	70.52	5.87	70.94	5.51	71.41	5.27	71.47	5.44	71.49	5.57	71.57	5.51	71.65	5.27	6.00
t3_20	70.16	4.00	70.59	6.23	70.80	5.82	70.82	5.62	71.19	5.30	71.27	5.50	71.30	5.50	71.54	5.56	71.59	5.71	71.65	5.74	8.00

t4_1	69.36	5.00	69.88	6.23	70.16	6.16	70.72	6.36	71.04	5.91	71.05	6.24	71.13	6.28	71.18	6.25	71.29	6.11	71.45	5.91	5.00
t4_2	69.16	4.50	70.22	5.48	71.17	5.64	71.25	5.85	71.37	5.88	71.62	5.90	72.03	5.91	72.24	5.92	72.36	5.72	72.49	5.79	6.00
t4_3	68.20	9.50	69.94	6.86	70.11	6.74	70.14	6.81	70.21	6.55	70.38	6.46	70.52	6.40	70.71	6.41	70.82	6.37	71.05	6.33	8.00
t4_4	64.10	5.50	64.43	5.00	64.89	5.17	64.89	4.88	65.97	4.71	66.49	4.76	66.62	5.13	66.72	5.23	67.08	5.10	67.13	5.28	6.50
t4_5	72.76	8.00	72.87	6.26	73.04	6.83	73.73	6.51	73.76	6.41	73.85	6.02	74.40	6.09	74.50	6.14	74.53	6.33	74.54	6.25	5.00
t4_6	72.50	4.00	72.81	4.99	73.15	5.32	73.37	4.75	74.21	4.90	74.23	4.83	74.47	4.92	74.57	4.81	74.67	4.94	74.80	4.94	7.00
t4_7	67.80	4.00	69.04	4.97	69.08	4.65	69.66	4.86	69.90	5.08	70.08	5.23	70.47	5.40	70.52	5.11	70.55	5.31	70.61	5.33	5.50
t4_8	65.08	6.50	65.61	5.26	66.32	5.34	67.05	5.98	67.65	5.98	67.85	5.91	67.90	5.65	67.91	5.63	68.08	5.93	68.19	5.75	5.00
t4_9	66.33	5.50	67.01	6.24	67.61	6.49	67.94	5.88	68.14	5.52	68.35	5.43	68.41	5.72	68.50	5.82	68.68	5.62	68.70	5.66	7.50
t4_10	72.29	6.00	73.55	6.00	73.79	6.32	73.85	5.76	74.15	5.80	74.47	5.99	74.60	5.93	74.83	5.99	74.83	5.78	74.92	5.80	6.00
t4_11	60.14	4.50	61.77	4.74	62.08	4.50	62.97	4.74	64.03	4.88	64.52	5.14	64.54	5.18	64.69	5.04	64.87	5.14	65.70	5.18	4.00
t4_12	67.27	7.00	69.28	6.76	69.47	6.52	69.92	5.91	70.14	5.83	70.37	5.54	71.09	5.54	71.21	5.59	71.31	5.48	71.34	5.57	5.00
t4_13	69.03	6.50	69.81	6.25	70.06	6.33	70.45	6.01	70.64	5.91	70.73	5.93	70.98	5.80	71.16	5.88	71.19	5.90	71.41	5.95	8.00
t4_14	71.17	5.50	71.29	5.50	72.14	6.15	72.60	6.23	72.72	6.57	72.75	6.56	72.83	6.48	73.00	6.36	73.03	6.38	73.05	6.34	5.00
t4_15	66.01	6.00	66.74	5.01	68.56	5.32	68.69	5.37	69.00	5.39	69.41	5.64	69.53	5.69	70.04	5.73	70.29	5.60	70.35	5.45	7.00
t4_16	66.41	7.50	68.28	6.78	68.72	5.89	68.89	5.92	69.10	6.03	69.12	5.54	69.74	5.33	70.56	5.52	70.66	5.62	70.99	5.75	6.50
t4_17	59.22	5.50	61.82	4.78	62.10	4.53	62.70	4.76	62.71	4.81	63.04	4.76	63.05	5.14	63.55	5.00	64.70	5.10	64.79	5.14	3.00
t4_18	67.05	4.00	69.18	4.95	69.31	5.13	69.51	4.98	69.71	4.98	70.19	5.14	70.30	5.19	70.36	5.05	70.62	5.10	70.68	5.18	6.50
t4_19	66.91	6.00	67.31	6.25	67.59	6.00	68.15	5.88	68.19	6.39	68.89	6.49	68.90	6.49	69.04	6.49	69.30	6.44	69.31	6.44	4.00
t4_20	69.84	6.00	70.28	6.00	70.74	6.16	71.24	6.24	71.85	6.39	71.86	5.85	71.93	5.87	72.22	5.77	72.24	5.84	72.40	5.67	7.00
t5_1	70.91	7.50	71.92	6.53	72.21	7.00	72.28	7.00	72.44	7.29	72.63	7.09	72.79	7.21	72.87	7.07	72.94	6.74	73.14	6.57	5.50
t5_2	68.23	6.00	68.89	5.01	69.62	5.49	70.61	5.14	70.69	5.69	70.81	5.74	70.93	5.36	71.20	5.73	71.51	6.08	71.65	6.16	5.50
t5_3	67.28	5.50	67.40	4.75	68.04	4.18	68.99	4.14	69.02	4.50	69.20	4.98	70.19	4.85	70.33	5.16	70.62	5.14	70.78	5.17	4.00
t5_4	65.65	4.00	65.97	4.75	66.07	5.16	66.88	5.12	67.12	4.90	67.98	4.76	68.36	4.99	68.53	5.00	68.62	4.89	68.66	5.00	5.00
t5_5	67.76	5.00	69.38	5.72	70.19	5.97	70.31	5.86	70.52	5.98	71.01	6.06	71.25	6.05	71.32	6.05	71.35	6.04	71.41	5.99	7.00
t5_6	67.72	6.00	67.89	5.00	68.14	5.66	69.06	5.26	69.09	5.70	69.23	5.58	70.14	5.37	70.18	5.39	70.35	5.40	70.65	5.45	6.00
t5_7	66.35	8.00	68.39	7.04	68.55	7.03	68.72	6.66	68.78	6.53	68.91	6.20	69.07	6.17	69.35	5.91	69.69	5.82	70.09	5.79	4.50
t5_8	68.59	5.50	68.62	5.75	68.99	5.17	69.21	5.13	69.25	4.91	69.27	4.59	69.30	4.51	69.48	4.63	69.62	4.56	69.73	4.85	5.00
t5_9	69.03	6.00	69.92	6.24	70.16	6.49	70.63	6.25	70.64	6.69	70.75	6.58	70.81	6.22	70.81	6.07	70.83	5.95	70.96	6.20	7.00
t5_10	67.63	7.00	67.91	7.25	68.02	6.84	68.38	6.38	68.41	6.41	68.74	6.02	69.19	5.74	69.45	5.71	69.65	5.64	69.91	5.67	4.00
t5_11	64.18	4.00	65.02	5.23	65.35	5.16	66.09	4.88	66.25	5.48	67.37	5.56	67.37	5.55	67.54	5.61	67.82	5.59	68.05	5.68	6.50
t5_12	65.84	4.00	68.56	4.94	69.34	4.34	69.39	4.49	69.57	4.78	69.69	4.66	69.70	4.57	69.90	4.80	70.01	4.71	70.06	5.08	4.00

t5_13	66.17	5.00	66.72	4.51	68.10	4.82	68.28	4.38	68.48	4.79	68.81	4.90	68.89	5.12	69.04	5.23	69.06	5.10	69.18	5.18	7.00
t5_14	67.41	6.00	67.44	5.00	67.92	5.00	68.07	4.75	68.64	5.00	68.67	5.00	68.80	5.20	68.86	4.94	69.22	5.05	69.59	5.28	6.50
t5_15	73.13	3.00	75.02	5.12	75.03	5.24	75.28	5.54	75.34	6.02	75.56	5.69	75.61	6.08	75.71	6.43	75.76	6.17	76.00	6.11	6.00
t5_16	68.11	6.50	69.30	5.53	69.70	5.68	70.01	6.00	70.52	6.09	70.80	6.00	70.92	6.00	71.28	5.94	71.33	5.74	71.35	6.00	6.00
t5_17	69.30	6.00	71.25	4.57	71.62	5.03	71.76	5.26	72.24	5.21	72.40	5.02	72.41	5.22	72.51	5.25	72.59	5.44	72.75	5.45	6.00
t5_18	65.43	5.50	65.70	4.75	66.05	4.51	66.45	4.63	66.87	4.51	66.95	4.67	67.62	4.58	67.78	4.39	67.80	4.61	67.85	4.60	6.00
t5_19	65.92	6.50	66.34	5.26	67.96	5.97	68.35	5.62	68.49	5.79	68.63	5.66	68.85	5.71	69.00	5.39	69.60	5.61	69.71	5.83	5.50
t5_20	68.35	6.00	69.43	5.02	69.77	5.34	70.31	5.26	70.53	5.40	70.56	5.34	71.14	5.16	71.71	5.25	72.08	5.48	72.11	5.39	4.00
t6_1	66.58	6.00	68.05	6.00	68.35	6.00	68.46	6.37	69.18	6.58	69.73	6.26	70.16	5.95	70.36	5.61	70.45	5.70	70.62	5.54	6.00
t6_2	68.52	4.50	69.02	4.99	69.87	5.16	70.02	5.24	70.26	4.81	70.47	4.68	70.57	4.79	70.62	5.00	70.73	5.05	70.75	5.19	5.00
t6_3	64.60	5.50	66.25	6.95	66.35	5.99	66.67	5.87	66.93	6.28	67.39	6.56	67.64	6.21	67.76	5.83	67.85	6.11	68.26	5.91	6.00
t6_4	65.60	8.00	67.27	5.58	67.48	6.53	67.82	6.16	68.44	6.22	68.79	6.11	68.81	6.16	68.83	5.90	68.83	6.07	69.14	5.88	6.00
t6_5	66.18	5.00	67.78	5.48	68.31	5.33	68.78	5.01	69.55	5.20	69.59	5.32	69.93	5.41	70.21	5.25	70.23	5.17	70.48	5.29	4.00
t6_6	67.49	5.00	67.95	4.50	68.15	5.16	70.23	5.58	70.32	5.66	70.63	6.10	70.65	5.82	70.75	6.08	70.81	6.12	70.82	5.92	6.00
t6_7	67.48	4.00	68.33	3.51	68.71	4.16	69.11	4.60	69.80	4.87	70.03	4.81	70.04	4.84	70.25	4.92	70.85	5.03	70.91	5.17	6.00
t6_8	66.53	8.00	66.64	7.00	66.88	7.50	66.96	6.63	67.63	6.51	67.64	6.43	67.76	6.37	67.90	6.32	67.91	6.34	68.18	6.31	5.00
t6_9	68.49	6.00	69.39	5.76	69.82	5.83	69.92	5.14	70.15	5.41	70.20	5.50	70.42	5.64	70.87	5.80	71.27	5.82	71.28	5.84	7.00
t6_10	67.23	6.00	67.47	5.75	67.79	5.67	68.90	5.39	69.38	5.32	69.50	5.66	69.88	5.44	69.94	5.15	70.04	5.03	70.18	4.93	7.50
t6_11	69.89	6.50	70.74	5.28	70.86	4.86	70.88	4.89	70.91	4.82	70.92	5.01	70.98	5.01	71.00	4.88	71.15	4.90	71.29	5.01	6.00
t6_12	66.47	5.00	66.97	5.00	67.62	4.67	68.26	4.87	68.40	4.70	68.63	4.99	68.76	5.13	68.77	4.99	68.83	5.10	68.94	5.24	5.50
t6_13	68.50	5.00	68.98	6.48	69.18	6.32	69.51	5.76	69.92	5.90	70.37	6.00	70.43	5.93	70.48	6.06	70.62	6.32	70.82	6.19	5.50
t6_14	66.17	6.00	66.52	5.50	66.79	5.50	67.79	6.46	67.84	6.37	68.96	6.00	69.58	5.61	69.90	5.31	70.03	5.38	70.04	5.25	7.00
t6_15	67.97	5.50	68.24	5.25	69.98	5.64	70.30	5.73	71.06	5.41	71.42	5.42	71.56	5.63	71.61	5.62	71.76	5.45	71.85	5.45	7.00
t6_16	65.63	4.50	68.20	4.26	68.81	5.60	69.57	5.69	69.65	5.85	69.87	6.19	70.02	6.16	70.38	5.79	70.62	5.81	71.01	5.88	8.00
t6_17	59.90	5.50	62.76	4.31	63.27	4.53	64.24	4.64	64.24	4.99	64.54	5.23	64.88	5.34	65.27	5.18	65.39	5.32	65.42	5.19	4.00
t6_18	57.20	5.50	58.13	4.76	58.85	4.84	59.13	4.63	59.47	5.00	60.66	4.84	60.68	5.34	61.09	5.06	61.43	5.00	61.55	4.81	5.00
t6_19	64.48	5.50	64.57	6.00	65.70	6.00	66.00	6.00	66.02	5.61	66.23	5.67	66.35	5.65	66.36	5.33	66.39	5.24	66.52	5.26	5.00
t6_20	64.81	4.00	66.16	4.49	66.62	5.62	66.66	6.33	66.97	6.17	67.17	6.30	67.19	6.61	67.79	6.47	68.02	6.42	68.04	6.19	4.50
t7_1	64.81	5.50	69.96	6.17	70.62	6.27	70.64	5.97	71.06	5.98	71.08	5.90	71.15	5.85	71.28	5.75	71.33	5.83	71.64	5.85	4.00
t7_2	66.65	7.50	67.04	6.76	67.84	6.84	68.22	6.75	68.30	6.41	68.50	6.75	68.61	6.65	68.63	6.81	69.06	6.78	69.15	6.80	6.00
t7_3	66.42	6.00	66.44	5.50	66.83	5.67	67.84	5.63	67.91	5.31	68.05	5.58	68.17	5.50	68.54	5.56	68.60	5.88	68.67	5.70	6.00
t7_4	70.09	8.50	70.21	6.75	70.91	7.16	71.26	6.88	71.52	6.81	71.78	6.36	71.83	5.90	71.84	5.79	71.87	6.13	71.88	6.12	6.50

t7_5	66.82	5.00	67.59	5.49	67.63	5.66	68.34	5.50	69.57	5.69	70.84	6.03	71.11	6.09	71.15	6.02	71.48	5.97	71.49	6.02	4.00
t7_6	66.57	6.00	66.90	6.99	68.35	6.04	68.38	5.79	68.45	5.64	68.63	5.45	68.74	5.46	68.84	5.28	69.39	5.41	69.51	5.51	6.50
t7_7	59.99	6.50	60.04	5.75	62.06	5.19	62.24	5.27	63.35	5.03	63.46	5.10	64.20	4.96	64.52	5.37	65.02	5.33	65.04	5.39	3.00
t7_8	62.74	5.00	63.19	6.24	65.22	6.32	65.69	6.25	66.37	5.83	66.42	5.77	66.61	5.40	66.62	5.59	67.15	5.53	67.17	5.53	4.00
t7_9	64.89	6.00	70.29	5.56	70.30	5.08	70.42	5.29	70.58	5.14	70.92	5.20	70.99	5.38	71.04	5.39	71.46	5.35	71.50	5.32	5.50
t7_10	60.63	5.00	65.81	4.09	65.91	4.06	66.31	4.87	66.41	4.89	67.00	4.99	67.24	5.19	67.36	5.29	67.73	5.36	67.87	5.51	4.00
t7_11	63.89	3.00	64.04	4.25	65.25	4.33	65.76	4.73	66.10	4.78	66.93	4.97	66.97	5.31	67.08	5.28	67.30	5.14	67.40	5.03	4.00
t7_12	66.24	4.50	67.97	4.74	68.45	4.98	69.32	5.11	69.39	5.28	69.71	5.71	70.51	5.81	70.51	6.18	70.75	6.37	71.48	6.34	7.00
t7_13	66.86	6.00	69.07	5.05	69.90	4.72	69.95	4.67	69.95	4.73	70.60	4.77	70.64	4.94	70.65	5.07	70.87	5.38	71.07	5.48	5.00
t7_14	67.33	8.50	67.52	8.25	69.22	7.54	69.25	7.17	69.51	6.37	69.52	6.47	69.67	6.89	69.84	6.78	69.88	6.81	70.14	6.82	6.50
t7_15	69.28	4.00	70.05	6.21	70.39	6.14	70.52	6.35	70.76	6.28	70.88	6.24	71.24	6.14	71.57	6.00	71.58	6.05	71.64	6.10	8.50
t7_16	70.97	5.50	71.17	5.50	71.48	5.01	71.64	5.37	72.29	5.49	72.43	5.26	72.60	5.56	72.73	5.44	73.27	5.55	73.35	5.31	5.00
t7_17	57.23	5.00	63.64	5.00	63.67	5.78	65.11	5.83	65.15	5.30	66.46	5.63	67.18	5.74	67.42	5.55	67.75	5.54	67.78	5.49	5.50
t7_18	56.02	5.00	67.89	5.60	68.37	6.14	68.52	5.89	68.71	6.36	69.02	6.15	69.13	6.13	69.33	5.78	69.46	5.80	69.69	5.63	7.50
t7_19	65.85	5.00	68.15	6.19	68.55	5.97	68.58	5.98	68.71	5.98	68.73	5.90	68.93	5.64	69.22	5.80	69.51	5.77	69.52	5.60	7.00
t7_20	61.32	4.50	62.67	4.26	66.23	4.48	67.96	4.37	67.98	4.13	68.31	4.48	68.96	4.61	69.44	4.71	69.59	4.64	69.80	4.72	5.50
t8_1	53.63	7.50	56.67	6.54	60.80	5.78	66.17	5.43	66.84	5.68	66.96	5.93	67.93	5.82	67.94	5.89	68.06	5.62	68.25	5.65	5.00
t8_2	74.21	8.00	74.60	6.77	74.95	6.35	74.96	6.51	75.19	6.61	75.22	6.75	75.35	6.37	75.73	6.62	75.79	6.45	75.94	6.50	5.00
t8_3	69.88	8.00	70.17	7.01	70.20	7.00	70.31	7.38	70.74	6.52	70.85	6.44	70.88	6.65	70.94	6.51	71.02	6.40	71.09	6.32	9.00
t8_4	71.52	5.50	72.95	5.50	73.29	5.34	73.48	5.26	73.82	5.30	74.12	4.94	74.13	5.02	74.34	4.90	74.45	5.12	74.46	5.01	5.00
t8_5	69.48	7.50	69.92	6.51	69.93	6.18	69.95	6.13	70.19	6.11	70.49	6.17	70.53	5.94	70.78	5.88	70.94	5.68	71.10	5.52	6.50
t8_6	64.20	5.50	65.90	4.29	66.06	4.52	67.28	4.63	67.37	4.80	67.87	4.99	68.00	4.85	68.10	5.29	68.27	5.31	68.30	5.23	6.00
t8_7	68.57	8.00	68.74	6.75	69.77	6.19	70.02	6.26	70.31	5.83	70.72	5.78	70.83	6.15	70.88	6.25	71.34	6.22	71.40	6.16	9.00
t8_8	69.92	7.50	69.96	6.00	69.98	6.17	69.99	6.00	70.10	6.10	70.39	5.92	70.46	6.14	70.63	6.12	70.69	5.89	70.73	5.90	8.00
t8_9	68.82	4.00	71.38	4.00	71.91	4.47	71.98	4.83	72.12	4.87	72.14	4.73	72.47	4.90	72.62	4.91	72.77	5.03	72.84	5.36	7.00
t8_10	66.19	5.50	67.74	5.74	68.29	5.82	68.62	5.39	68.71	5.31	69.01	5.10	69.47	5.09	69.74	5.20	69.86	5.49	69.91	5.54	5.00
t8_11	69.96	7.00	72.52	7.70	72.79	7.16	73.09	7.01	73.26	6.72	73.61	6.77	73.64	6.80	73.76	6.94	73.97	6.79	74.22	6.62	5.50
t8_12	68.46	5.00	70.63	5.47	71.23	5.80	71.42	5.73	71.81	5.87	71.93	6.13	72.05	6.11	72.07	6.04	72.11	6.04	72.25	5.84	9.00
t8_13	65.51	5.00	65.75	5.00	65.81	4.67	65.82	4.50	66.12	4.50	66.50	4.66	66.84	4.78	66.98	4.81	67.37	4.93	67.65	5.03	6.00
t8_14	66.66	5.50	67.23	4.26	69.59	4.18	69.82	4.26	70.08	4.21	70.17	4.18	70.39	4.35	70.60	4.67	71.04	4.70	71.26	4.87	7.50
t8_15	72.72	4.50	72.91	6.49	73.53	5.68	73.97	5.76	73.98	5.81	74.11	5.84	74.25	5.93	74.29	5.94	74.37	5.73	74.78	5.75	8.50
t8_16	69.53	7.00	69.81	6.25	70.08	5.51	70.99	5.51	71.40	5.60	71.46	5.67	71.56	5.85	71.97	6.16	72.04	6.25	72.13	6.27	8.00

t8_17	65.71	6.00	67.52	6.48	67.53	6.00	67.68	6.61	67.80	6.39	67.92	6.25	67.95	6.63	68.38	6.44	68.41	6.07	68.68	5.87	9.50
t8_18	68.11	7.00	70.97	6.07	71.13	5.89	71.44	6.39	71.52	6.51	72.47	6.43	72.49	6.37	73.34	6.55	73.60	6.50	73.61	6.18	6.50
t8_19	68.77	6.00	69.09	6.74	69.11	7.16	69.44	7.00	70.41	6.52	70.56	6.52	70.75	6.31	71.07	6.39	71.43	6.46	71.50	6.37	6.50
t8_20	65.53	7.50	68.31	7.03	68.45	6.06	69.48	5.93	69.69	5.76	69.78	5.95	69.81	5.89	69.86	6.20	69.92	6.18	70.04	6.16	7.00
t9_1	69.34	6.00	69.53	5.75	69.75	5.83	69.94	6.00	70.02	6.20	70.10	6.16	70.18	6.21	70.28	6.49	70.38	6.55	70.40	6.69	7.00
t9_2	66.47	4.00	67.83	5.70	68.14	6.45	68.28	6.09	68.51	6.37	69.03	6.31	69.10	6.20	69.27	5.93	69.47	5.73	69.59	5.80	8.00
t9_3	67.27	6.00	67.73	5.75	68.66	5.51	69.39	5.15	69.73	5.69	70.07	5.66	70.14	5.57	70.17	5.38	70.22	5.34	71.22	5.45	5.00
t9_4	66.16	6.00	66.92	5.02	67.57	5.17	68.30	5.37	68.95	5.87	69.05	5.57	69.05	6.04	69.11	5.97	69.13	5.87	69.14	5.88	4.00
t9_5	67.14	6.00	67.58	6.50	68.17	6.01	68.63	5.52	69.22	5.61	69.54	5.60	69.74	5.38	69.83	5.33	70.15	5.35	70.35	5.41	6.00
t9_6	67.96	8.50	68.42	7.51	68.76	7.02	69.38	7.13	69.62	7.30	69.71	7.09	69.78	7.08	69.87	6.95	69.88	7.06	70.25	7.15	6.00
t9_7	69.47	4.00	70.60	4.00	70.93	4.16	71.95	4.36	72.38	4.57	72.56	4.64	72.59	4.69	72.60	4.61	72.79	4.75	72.85	4.87	4.50
t9_8	67.29	5.50	69.31	4.78	69.64	5.17	69.77	5.37	69.98	5.40	70.17	5.17	70.20	5.43	70.31	5.80	70.33	5.82	70.64	5.79	7.00
t9_9	68.09	5.00	69.40	4.51	69.52	5.32	69.54	5.37	69.58	5.10	69.62	5.08	69.63	5.35	69.64	5.62	69.68	5.33	69.73	5.20	4.00
t9_10	57.27	5.00	57.90	4.75	58.29	4.18	60.56	4.72	60.57	4.59	61.13	4.81	61.71	4.70	61.81	4.50	62.18	4.76	62.66	4.83	4.00
t9_11	66.85	6.00	67.12	6.75	67.35	6.50	67.73	6.13	67.84	6.11	67.86	5.76	68.15	5.86	68.64	6.18	68.65	5.95	68.89	5.86	6.50
t9_12	67.78	6.50	68.21	6.00	69.45	5.37	69.49	5.04	69.92	5.03	70.13	5.26	70.20	5.09	70.33	4.96	70.44	5.34	70.51	5.21	6.00
t9_13	64.24	6.00	65.24	7.23	67.37	6.69	67.72	6.29	67.77	6.33	67.78	6.60	68.60	6.79	69.34	6.75	69.64	6.68	69.75	6.52	6.50
t9_14	61.00	4.00	61.50	4.00	61.84	4.00	62.55	4.36	64.14	4.48	64.60	4.80	65.74	4.82	65.98	4.90	66.20	5.01	66.27	5.01	4.50
t9_15	67.63	7.00	68.90	7.73	68.95	7.00	69.02	6.88	69.07	6.31	69.08	6.10	69.37	5.80	69.45	5.83	69.51	5.90	69.53	6.11	6.00
t9_16	66.32	5.00	67.01	5.49	67.28	5.66	67.49	5.37	67.50	5.50	68.03	5.90	68.51	5.92	68.82	5.98	69.23	6.19	69.33	6.27	4.00
t9_17	64.71	4.00	65.56	4.00	65.81	4.66	65.96	4.99	66.09	4.99	66.75	5.07	66.78	5.27	66.97	5.30	67.43	5.16	67.65	5.14	5.50
t9_18	64.97	4.00	65.21	4.00	65.33	4.00	65.64	4.62	66.01	4.79	66.05	4.91	66.19	4.92	66.29	4.93	67.19	5.04	67.34	5.04	6.00
t9_19	63.92	4.00	64.89	4.49	65.03	4.33	65.27	4.62	65.51	4.79	66.09	5.06	66.21	5.06	66.74	4.93	67.21	4.94	67.43	5.04	6.50
t9_20	65.26	5.00	67.77	4.53	67.91	4.84	68.07	5.00	68.88	5.09	69.36	5.31	69.62	5.47	69.78	5.53	69.85	5.58	69.87	5.57	6.50
t10_1	65.93	5.00	67.77	5.24	68.10	4.84	69.86	5.11	69.87	5.09	69.90	5.39	70.18	5.47	70.39	5.36	70.46	5.11	70.48	5.24	5.50
t10_2	59.70	4.00	60.98	4.49	61.57	5.63	64.00	5.60	64.88	5.40	65.35	5.19	65.47	5.23	65.74	4.97	66.13	4.97	66.55	4.89	4.50
t10_3	71.80	7.50	72.03	7.75	72.65	7.67	72.68	6.77	72.79	6.23	72.88	6.27	72.90	6.09	73.14	6.14	73.15	6.24	73.43	6.26	6.50
t10_4	65.23	4.00	66.93	4.48	67.69	4.33	67.76	4.85	67.99	4.88	68.08	5.14	68.35	5.19	68.57	5.35	68.62	5.36	68.77	5.47	6.00
t10_5	66.98	6.00	67.78	6.25	68.19	6.33	68.87	6.13	69.35	5.72	69.36	5.45	69.38	5.59	69.44	5.28	69.56	5.41	69.84	5.56	7.00
t10_6	68.51	7.00	69.01	5.52	69.43	5.67	69.48	5.88	69.66	6.00	69.74	5.92	69.81	5.79	70.13	5.81	70.15	5.89	70.18	5.85	8.00
t10_7	67.82	4.00	68.85	4.98	69.29	5.31	69.48	5.23	69.83	5.00	69.90	5.00	69.94	4.86	70.06	5.24	70.23	5.21	70.35	5.34	6.00
t10_8	67.20	6.00	67.77	6.99	68.21	6.18	68.91	5.41	69.06	5.53	69.57	5.44	69.68	5.25	69.75	5.10	69.92	5.19	69.95	5.22	7.50

t10_9	67.51	5.50	67.70	6.00	67.76	5.67	68.47	5.50	69.21	5.60	69.38	5.34	69.43	5.16	69.63	5.08	69.69	5.39	69.87	5.50	6.50
t10_10	69.84	7.00	70.37	7.00	70.46	6.67	71.09	6.39	71.16	6.31	71.16	6.34	71.18	6.57	71.22	6.50	71.26	6.34	71.36	6.36	6.50
t10_11	62.43	4.00	62.67	4.75	63.20	4.50	64.46	4.62	65.39	4.69	65.85	4.74	66.24	4.77	66.25	5.04	66.39	4.93	66.49	4.98	6.00
t10_12	71.17	5.50	71.35	4.26	71.59	5.32	71.60	5.24	71.94	5.49	71.96	5.49	72.00	5.84	72.09	5.62	72.17	5.82	72.17	5.84	6.50
t10_13	66.89	5.00	67.13	5.75	67.32	6.00	68.73	6.00	68.84	6.00	68.98	6.08	69.21	5.66	69.32	5.70	69.52	5.89	69.53	5.71	8.50
t10_14	66.35	8.00	69.41	7.07	70.55	6.89	70.68	6.57	70.70	6.74	70.70	6.63	70.77	6.54	70.90	6.53	71.34	6.43	71.37	6.20	7.00
t10_15	69.15	3.00	69.52	3.50	69.57	4.82	69.99	5.23	70.07	5.38	70.15	5.48	70.44	5.28	70.51	5.12	70.67	5.00	70.87	4.95	6.00
t10_16	69.33	4.00	69.49	5.25	70.23	5.17	70.56	5.49	70.69	5.59	70.73	5.65	70.92	5.77	71.16	5.80	71.22	5.66	71.40	5.64	5.50
t10_17	70.39	5.00	70.95	4.51	71.06	5.33	71.54	5.61	71.67	5.69	71.76	5.90	71.98	5.91	72.01	5.92	72.11	5.72	72.15	5.55	6.00
t10_18	66.31	3.00	67.42	4.47	68.29	4.96	68.53	4.73	68.61	4.68	68.71	5.22	68.82	5.26	69.25	5.59	69.91	5.58	70.14	5.80	6.00
t10_19	67.68	5.00	68.86	4.51	69.26	4.02	69.73	4.98	69.77	4.79	69.79	4.99	69.99	4.85	70.22	5.17	70.23	5.26	70.45	5.23	5.50
t10_20	66.56	4.00	66.63	5.25	66.73	4.83	66.76	4.88	67.43	4.90	67.53	5.16	67.61	5.35	67.73	5.55	68.05	5.60	68.29	5.68	5.00

Annex F: Full Experiment of Domain 2

Where: θ_i the cosine similarity angle of i where i ranges from 1 to 10.
SG_i is the calculated score of i nearest neighbor pre- graded essays and MG is the manual grade.

	θ_1	SG1	θ_2	SG2	θ_3	SG3	θ_4	SG4	θ_5	SG5	θ_6	SG6	θ_7	SG7	θ_8	SG8	θ_9	SG9	θ_{10}	SG10	MG
t1_1	65.01	6.50	68.56	5.34	68.62	4.60	69.40	5.63	69.51	5.70	69.66	5.43	71.14	5.70	71.84	5.619	72.17	5.27	72.23	5.42	4.00
t1_2	62.20	5.00	62.36	4.50	62.81	4.83	63.70	5.11	64.98	5.37	65.77	5.39	66.17	5.08	66.43	5.242	67.17	5.22	67.47	5.38	4.00
t1_3	64.67	5.00	64.70	5.50	64.87	5.00	65.05	4.75	66.46	4.99	66.66	5.39	68.02	5.21	68.22	5.470	68.89	5.32	68.89	5.48	3.00
t1_4	64.67	5.00	64.70	5.50	64.87	5.00	65.05	4.75	66.46	4.99	66.66	5.39	68.02	5.21	68.22	6.470	68.89	5.32	68.89	5.48	9.00
t1_5	63.72	4.00	67.27	4.93	69.38	4.65	70.61	4.95	71.23	4.96	71.39	4.96	71.44	5.29	71.58	5.485	71.64	5.49	71.69	5.44	6.00
t1_6	70.55	7.50	71.20	6.52	71.28	6.35	71.47	6.63	71.83	6.90	71.87	6.43	71.95	6.71	72.07	6.628	72.11	6.13	72.22	6.21	6.50
t1_7	64.00	7.00	67.13	6.53	68.42	6.67	68.69	6.40	68.78	6.23	69.42	6.35	69.80	6.37	69.92	6.327	70.42	6.29	70.58	6.27	8.00
t1_8	69.42	7.50	69.45	6.50	69.66	6.67	70.92	6.27	71.94	5.85	71.95	5.26	72.21	5.35	72.56	5.429	72.85	5.38	72.86	5.44	5.00
t1_9	62.77	4.00	64.61	5.21	65.58	6.10	66.17	6.43	66.50	6.25	66.94	6.14	67.04	5.98	67.14	5.865	67.40	5.67	67.57	5.84	5.00
t1_10	62.77	4.00	64.61	5.21	65.58	6.10	66.17	6.43	66.50	6.25	66.94	6.14	67.04	5.98	67.14	5.865	67.40	5.67	67.57	5.84	4.50
t1_11	60.01	6.00	60.57	6.00	61.78	5.35	63.63	5.04	64.24	5.03	64.93	5.18	65.07	5.42	65.65	5.37	65.72	5.33	65.73	5.54	4.00
t1_12	64.95	9.00	66.94	8.04	67.36	7.87	67.60	8.38	67.84	8.11	68.09	8.26	68.16	8.01	68.28	8.07	68.78	8.01	68.79	7.92	8.00
t1_13	58.75	7.00	59.29	7.00	59.75	6.51	60.59	6.63	61.25	6.51	61.44	6.59	61.88	6.64	61.91	6.80	62.38	6.72	62.44	6.65	4.00
t1_14	60.46	6.00	60.77	5.50	61.34	5.34	62.28	5.01	62.69	5.49	62.89	5.25	63.78	5.22	64.02	4.96	64.19	5.12	64.82	5.29	3.00
t1_15	66.74	5.00	67.01	5.99	67.48	5.34	68.01	5.38	68.02	5.11	68.45	5.41	68.45	5.50	68.93	5.73	69.42	5.71	69.45	5.69	7.00
t1_16	65.19	7.00	68.73	6.77	68.95	6.21	69.91	6.39	69.94	6.60	70.11	6.51	70.19	6.03	70.23	5.91	70.28	5.28	70.32	5.45	4.00
t1_17	64.62	6.00	70.44	5.12	70.99	4.94	71.62	5.40	71.88	5.42	71.89	5.51	71.98	5.70	72.77	5.40	73.00	5.36	73.25	5.15	5.00
t1_18	71.57	3.00	72.12	4.48	72.55	4.32	72.69	4.98	72.91	5.37	73.29	5.15	73.52	5.13	74.00	5.11	74.09	5.36	74.19	5.61	5.00
t1_19	67.96	4.00	69.74	5.44	72.49	6.05	73.81	5.93	73.89	6.12	75.64	6.24	76.63	6.38	76.76	6.04	76.79	6.04	76.99	5.95	7.00
t1_20	69.60	6.50	72.05	7.67	74.19	8.06	74.85	7.17	75.20	7.58	75.50	7.42	75.60	6.98	75.87	6.82	76.14	6.34	76.33	6.45	7.00
t2_1	69.37	9.50	71.52	9.26	71.71	8.86	72.29	8.30	72.48	7.77	73.01	7.73	73.28	7.57	73.48	7.27	73.99	7.44	74.00	7.45	9.00
t2_2	70.28	4.00	70.91	5.48	71.67	6.13	72.52	6.10	72.93	6.27	73.19	5.91	73.34	5.79	73.50	5.70	73.54	5.52	73.70	5.47	7.00
t2_3	69.88	6.50	70.58	7.97	71.07	7.98	71.56	7.87	71.94	7.51	72.06	7.43	72.15	7.30	72.22	7.39	72.27	7.18	72.44	7.17	8.00
t2_4	65.21	7.00	66.98	4.59	67.80	5.67	69.08	5.98	69.11	6.17	69.14	6.06	69.64	6.26	69.78	6.11	69.99	5.99	70.36	5.99	7.00
t2_5	65.74	4.00	66.51	5.23	68.17	5.79	69.67	5.72	70.17	5.95	70.38	6.19	71.70	5.79	71.72	5.82	71.86	5.24	72.79	5.13	5.00
t2_6	64.49	9.00	64.56	8.25	65.45	7.84	66.31	7.40	67.87	7.05	68.16	6.82	69.04	6.32	69.08	6.46	69.76	6.46	70.53	6.33	5.50
t2_7	67.08	6.00	67.41	6.50	67.84	5.68	69.15	5.75	69.38	5.80	69.75	5.83	70.09	5.59	71.16	5.52	71.17	4.96	71.24	5.01	5.50
t2_8	64.61	6.50	67.43	6.97	67.46	7.30	68.44	7.23	68.65	6.81	68.76	6.37	69.21	6.18	69.33	6.34	69.55	5.99	69.99	5.62	5.00
t2_9	61.82	7.00	64.26	5.56	66.60	5.54	67.30	5.76	68.92	6.15	69.04	6.21	69.51	6.05	69.64	6.05	69.69	6.04	69.89	6.13	6.00
t2_10	59.69	7.00	61.10	6.51	61.34	6.83	61.74	6.87	62.30	7.00	62.55	6.51	64.12	6.71	64.24	6.57	64.27	6.77	64.58	6.79	7.00
t2_11	65.66	6.00	65.67	6.50	67.53	5.71	68.49	5.08	68.56	5.35	69.14	5.21	69.40	5.19	69.48	5.63	69.61	6.09	70.04	5.89	4.00

t2_12	65.67	6.00	66.44	5.51	66.93	5.99	67.64	6.12	68.56	5.72	69.47	5.84	69.83	5.08	70.03	5.18	70.22	5.27	71.00	5.42	3.00
t2_13	60.99	6.00	65.20	5.07	66.07	5.05	66.87	5.38	67.07	5.31	68.34	5.12	68.56	4.84	68.74	4.63	69.02	4.97	69.08	4.88	4.00
t2_14	61.64	5.00	63.16	5.49	64.43	5.97	65.09	5.50	65.73	5.41	68.70	5.49	68.88	5.62	68.90	5.11	68.91	5.30	69.00	5.40	4.00
t2_15	58.10	7.00	58.99	5.52	65.23	5.08	65.66	5.62	65.86	5.15	66.45	4.98	66.78	4.59	66.81	4.64	67.16	4.73	67.72	4.84	4.00
t2_16	66.55	6.00	66.57	6.50	66.67	6.33	67.23	6.25	67.60	6.40	67.75	6.41	67.77	5.80	67.92	5.95	69.75	5.90	69.76	5.96	6.00
t2_17	62.87	6.50	63.48	5.76	64.07	6.49	64.88	6.73	65.22	6.59	65.53	6.34	66.95	6.62	67.27	6.55	67.39	6.39	67.43	6.40	5.00
t2_18	70.84	6.50	71.25	6.75	71.25	5.84	71.56	5.39	71.79	5.70	72.09	5.83	72.45	5.31	72.65	5.33	72.72	5.08	73.07	5.31	7.00
t2_19	69.48	7.00	69.69	6.00	69.83	6.17	70.08	5.88	70.10	6.20	70.49	6.09	70.81	5.94	71.18	5.71	71.21	5.84	71.21	5.81	5.50
t2_20	68.56	6.00	68.95	6.25	69.05	6.33	70.24	6.37	70.37	5.92	70.75	6.09	70.85	6.01	71.18	6.18	71.28	6.06	71.31	6.20	4.50
t3_1	68.03	5.50	68.23	5.75	69.89	5.51	70.16	5.15	70.44	4.93	70.90	4.79	71.35	4.95	71.41	5.01	71.44	5.22	71.52	5.11	6.50
t3_2	66.52	4.00	68.69	5.43	68.85	4.97	69.49	5.45	70.12	4.81	70.13	4.68	70.68	4.59	71.04	4.87	71.10	4.78	71.15	5.03	7.00
t3_3	65.35	6.50	66.49	5.28	66.56	5.84	67.34	5.76	68.21	5.90	68.51	5.29	69.04	5.58	69.18	5.81	69.29	5.83	70.01	5.93	7.00
t3_4	70.30	5.50	70.66	6.00	71.47	6.32	71.52	6.48	72.09	6.58	72.65	6.65	72.65	6.56	72.80	6.79	72.84	6.81	72.84	6.64	6.50
t3_5	61.45	7.00	63.62	7.00	65.18	6.69	65.34	6.76	65.63	6.71	65.96	6.76	65.98	6.93	66.17	7.17	66.20	6.84	67.07	6.39	7.00
t3_6	62.17	7.00	62.91	6.51	63.36	5.69	64.25	6.00	64.41	5.91	65.48	5.92	66.51	5.93	66.57	6.28	66.91	5.95	67.26	5.91	5.50
t3_7	71.54	7.00	72.21	4.54	72.52	5.18	72.59	5.99	72.92	6.19	73.12	6.16	73.27	6.69	73.31	6.84	73.58	6.86	73.62	6.78	7.00
t3_8	61.74	4.00	63.72	3.03	64.41	3.98	65.73	4.45	65.83	4.37	66.78	4.77	67.17	4.93	67.33	4.94	67.66	5.15	68.05	5.23	6.00
t3_9	63.83	5.50	67.30	5.97	68.12	5.67	68.36	5.74	68.63	5.51	68.70	4.95	68.88	5.10	69.11	5.09	69.75	5.18	69.98	5.26	5.50
t3_10	69.29	4.00	70.56	4.73	70.92	4.98	71.62	4.98	72.48	4.80	72.54	4.52	72.59	4.45	72.61	4.40	72.64	4.57	72.87	4.42	5.50
t3_11	65.95	6.50	66.51	6.99	66.68	7.33	67.92	7.37	68.10	7.68	68.29	6.77	68.55	6.67	68.63	6.71	68.74	6.74	69.83	6.67	7.00
t3_12	60.81	4.00	63.61	4.00	63.96	4.48	64.80	4.72	66.37	4.59	67.37	4.65	67.76	4.89	67.82	4.33	68.76	4.45	69.52	4.67	2.00
t3_13	65.15	7.00	65.54	7.00	66.58	6.68	67.35	6.52	67.90	6.42	68.00	6.43	68.20	6.10	68.52	6.03	68.60	6.24	68.80	6.22	5.00
t3_14	61.61	8.00	63.05	7.02	63.25	6.69	64.36	6.76	64.83	6.62	64.89	6.20	66.02	6.24	67.13	6.33	67.19	6.55	67.32	6.54	7.50
t3_15	63.50	4.00	64.91	4.49	65.13	4.33	66.73	4.72	67.02	4.58	67.27	4.49	67.78	4.29	68.06	4.61	68.38	4.85	68.52	5.01	5.00
t3_16	64.21	2.00	65.45	3.47	65.96	4.77	66.49	5.43	67.31	4.78	67.43	4.82	67.63	4.71	67.71	4.98	67.99	5.03	68.12	4.94	4.00
t3_17	66.36	5.50	67.88	4.77	69.82	5.30	69.84	5.47	70.27	5.85	70.46	5.72	70.78	5.69	71.27	5.03	71.44	4.82	71.55	4.66	7.00
t3_18	71.74	7.00	71.87	4.51	71.99	4.34	72.81	4.74	72.86	4.98	73.29	5.07	73.36	5.06	73.49	5.05	73.62	5.36	74.22	5.14	5.00
t3_19	68.55	5.00	68.93	5.50	75.47	5.75	75.94	5.40	76.72	5.42	76.96	5.36	77.42	5.49	77.88	5.59	78.98	5.63	79.18	5.69	6.00
t3_20	64.15	7.50	64.81	7.50	65.19	7.17	66.12	7.13	68.44	6.21	68.53	6.18	68.86	6.29	68.90	6.43	69.00	6.38	69.25	6.35	8.00
t4_1	68.55	5.50	68.81	3.76	69.25	4.82	69.52	5.84	70.17	6.06	70.61	5.90	70.98	5.91	71.00	6.04	71.72	6.14	72.07	6.31	7.00
t4_2	57.72	8.00	61.23	7.29	61.41	7.20	62.36	7.15	63.07	6.75	63.22	6.79	64.87	7.01	65.17	6.89	65.35	6.40	65.49	6.50	7.50
t4_3	66.27	7.00	67.10	6.51	67.83	5.38	67.85	5.53	69.40	5.62	69.85	5.67	70.79	5.52	70.92	5.63	70.93	5.47	71.32	5.52	4.00

t4_4	61.11	7.00	62.88	7.00	63.47	7.00	65.72	6.77	65.90	6.91	65.98	7.08	66.75	6.54	67.04	6.48	67.54	6.53	67.90	6.44	10.00
t4_5	69.63	6.00	71.06	6.00	71.57	6.32	71.91	6.48	71.93	6.39	72.80	6.41	72.88	6.22	73.08	6.31	73.11	6.38	73.90	6.17	5.50
t4_6	67.05	5.50	67.94	6.24	70.43	5.71	70.50	5.89	71.21	6.00	71.22	5.93	71.76	5.67	71.88	5.71	71.97	5.84	72.33	5.67	7.00
t4_7	66.40	7.50	67.44	5.79	68.89	5.85	68.98	6.13	69.61	5.73	70.20	5.61	70.86	5.80	71.27	5.48	71.50	5.58	71.57	5.76	4.00
t4_8	61.51	7.00	62.58	5.53	63.54	4.39	67.70	5.06	67.91	5.49	68.44	5.49	68.73	5.30	68.85	5.32	69.11	5.29	69.38	5.36	5.50
t4_9	64.15	5.00	64.46	5.25	64.67	5.17	64.78	4.88	65.45	3.93	65.68	3.62	65.89	3.67	66.14	3.83	66.36	4.06	66.55	4.20	4.00
t4_10	72.05	5.00	72.43	5.25	73.74	5.64	73.81	5.73	74.34	5.22	74.51	5.10	74.81	4.96	74.96	5.08	75.00	4.97	75.61	4.70	3.50
t4_11	63.09	8.00	65.55	6.57	66.80	6.70	67.04	6.77	67.11	6.72	67.69	6.29	68.17	6.39	68.30	5.87	68.43	5.94	68.76	5.95	7.50
t4_12	65.89	3.00	68.12	3.95	68.14	3.97	69.27	3.98	69.49	4.36	70.97	4.60	71.42	4.65	71.44	5.03	71.59	4.93	71.92	5.07	4.00
t4_13	69.69	4.00	70.86	3.51	72.30	4.13	72.94	4.22	73.07	4.64	73.18	4.85	73.21	5.08	73.37	5.24	73.97	5.12	74.02	5.20	7.50
t4_14	73.45	6.50	74.94	6.26	75.76	5.87	76.34	6.13	76.38	5.92	76.76	5.62	76.81	5.74	77.36	5.43	77.38	5.39	77.90	5.75	6.00
t4_15	64.41	7.00	66.19	6.28	66.36	6.19	67.53	5.67	67.60	5.73	67.81	5.93	68.28	5.94	68.37	5.71	68.46	5.53	68.56	5.53	5.00
t4_16	65.23	3.00	65.30	4.00	65.60	4.00	66.69	4.12	67.47	4.10	67.88	4.24	68.90	4.40	69.06	4.76	69.53	4.63	70.01	4.57	2.00
t4_17	68.43	6.00	69.08	5.75	69.64	6.00	71.48	5.54	71.64	5.81	72.10	5.53	72.36	5.59	72.41	5.58	72.47	5.63	72.62	5.75	7.00
t4_18	67.73	3.00	67.86	4.25	68.16	4.83	68.74	5.11	69.83	5.09	70.23	5.54	70.76	5.21	71.35	5.30	71.59	5.27	71.61	5.29	4.00
t4_19	61.65	7.00	63.91	7.00	66.35	6.85	67.04	6.88	67.88	6.72	67.92	6.77	68.09	6.93	68.28	6.82	68.44	6.84	68.75	6.85	7.00
t4_20	65.53	7.00	66.10	6.51	66.13	6.67	66.39	6.75	67.06	5.83	67.40	5.86	67.97	6.01	68.26	6.01	69.45	6.11	69.59	6.10	6.00
t5_1	62.78	7.00	64.51	6.52	64.71	7.65	65.24	7.49	65.58	7.20	66.02	7.17	66.30	7.15	67.03	6.96	67.45	6.65	67.47	6.50	7.00
t5_2	69.05	5.00	69.15	5.25	69.24	5.83	69.39	5.75	69.51	5.30	69.80	5.42	70.04	5.64	70.42	5.80	70.97	5.40	71.37	5.41	5.00
t5_3	64.16	5.50	65.01	6.24	65.60	4.86	66.75	4.66	67.76	4.91	68.11	5.38	68.58	5.60	68.89	5.93	69.96	6.13	70.19	6.34	5.00
t5_4	57.36	2.00	58.95	3.96	61.34	4.59	61.61	5.16	61.84	4.94	62.25	5.35	62.53	5.71	62.81	5.86	62.99	5.87	63.86	6.07	6.00
t5_5	64.62	4.00	69.10	4.00	69.41	5.08	69.45	5.53	69.58	5.43	69.80	5.21	69.89	5.25	69.98	5.81	70.15	5.67	70.31	5.42	4.00
t5_6	68.70	2.00	69.39	4.95	69.58	5.46	70.46	5.35	70.68	6.05	71.03	5.80	71.22	5.97	71.43	5.91	71.53	5.92	71.84	5.79	6.50
t5_7	63.64	7.00	63.92	7.00	63.99	7.00	66.71	6.88	67.24	6.91	67.36	6.77	67.44	6.60	67.68	6.47	67.93	6.47	67.96	6.52	5.50
t5_8	59.59	7.50	62.75	7.26	63.24	7.82	63.63	7.14	64.13	7.02	65.50	7.02	65.73	6.88	65.74	6.43	66.32	6.34	66.36	6.40	6.50
t5_9	67.52	6.50	67.65	5.50	69.94	6.28	71.47	6.44	71.68	6.27	72.86	6.09	73.01	6.20	73.27	6.18	73.63	6.26	74.05	6.19	7.00
t5_10	68.70	8.00	68.97	7.75	69.92	5.90	70.87	5.22	70.98	5.55	71.20	5.86	71.95	5.88	72.02	6.01	72.10	6.11	72.66	5.92	6.50
t5_11	66.68	7.00	71.90	7.00	72.23	6.70	73.67	6.11	73.87	5.21	74.06	5.03	74.13	5.02	74.48	4.91	74.58	4.92	75.72	4.93	4.00
t5_12	67.17	7.00	67.48	7.00	68.75	6.36	69.52	6.16	69.82	5.75	70.20	5.63	70.30	5.41	70.72	5.37	70.72	5.59	71.48	5.63	5.00
t5_13	66.97	7.00	68.74	5.56	70.06	6.94	70.68	6.95	71.19	5.86	71.25	5.88	71.30	5.63	71.35	5.43	71.38	5.44	71.51	5.12	7.50
t5_14	69.69	8.00	73.84	7.55	73.88	7.38	74.20	6.83	74.82	7.04	74.84	6.96	75.09	6.97	75.10	6.85	75.46	6.87	75.52	6.42	7.00
t5_15	64.71	5.00	66.39	6.69	66.80	6.79	66.86	6.23	67.16	5.80	67.33	5.91	67.73	5.86	67.89	5.81	68.03	5.73	68.16	5.75	4.00

t5_16	68.67	4.00	68.83	4.50	69.15	4.99	69.54	4.99	70.15	5.00	70.22	5.32	70.37	5.41	70.51	5.24	70.52	5.11	70.65	5.34	7.00
t5_17	68.08	10.0	72.92	8.46	73.97	8.03	74.17	7.69	74.38	7.47	74.93	6.66	75.32	6.32	75.44	6.56	75.75	6.41	76.15	6.64	8.00
t5_18	69.24	5.50	70.45	7.69	70.98	6.50	71.33	6.50	71.75	6.60	71.75	6.50	71.80	6.57	71.85	6.26	72.06	6.34	72.12	6.31	5.00
t5_19	63.55	4.50	68.68	5.40	69.13	5.74	70.93	5.68	72.15	5.06	72.37	5.05	72.57	5.10	73.88	5.25	74.13	5.41	74.42	5.50	6.00
t5_20	67.12	5.50	67.24	6.50	68.72	6.66	69.16	6.74	69.32	6.41	69.50	5.70	69.87	5.95	69.87	5.95	69.88	6.07	70.14	6.20	5.00
t6_1	65.97	7.00	69.08	4.66	70.50	5.53	70.67	5.87	70.92	5.71	70.92	5.52	71.63	5.51	72.03	5.57	72.14	5.41	73.15	5.55	4.00
t6_2	66.70	5.00	67.63	5.49	68.28	5.65	68.57	5.01	68.76	4.82	69.60	5.47	69.75	5.68	69.75	5.60	69.82	5.32	70.11	5.43	5.00
t6_3	65.22	6.00	65.27	6.50	66.00	6.66	66.17	6.99	66.37	7.09	66.71	6.99	66.92	7.06	67.09	6.94	67.66	7.00	68.11	7.00	5.00
t6_4	65.34	5.00	66.31	4.51	67.29	5.31	67.97	5.95	68.83	5.58	69.83	5.65	69.87	5.95	70.35	6.41	70.37	6.27	70.75	6.15	6.00
t6_5	70.82	6.00	72.21	6.48	72.40	6.49	72.79	5.89	73.14	5.91	73.33	5.69	73.81	6.00	74.32	6.00	74.47	5.69	74.74	5.54	6.00
t6_6	56.90	9.00	62.51	8.08	64.24	7.31	64.45	7.35	64.72	6.91	64.73	6.46	64.91	5.99	64.96	6.05	65.23	5.84	65.35	5.95	7.50
t6_7	67.45	9.00	69.42	5.65	71.14	5.76	71.44	6.05	72.46	5.32	72.60	5.20	72.72	5.69	72.72	5.79	73.03	5.60	73.08	5.69	7.50
t6_8	70.53	7.00	71.19	7.25	71.34	6.51	72.09	6.75	72.53	6.61	73.13	6.36	73.24	6.31	73.26	6.63	73.34	6.72	73.37	6.65	6.00
t6_9	63.24	7.00	65.70	5.57	66.65	6.18	67.30	6.49	67.60	6.40	67.67	6.49	68.63	5.90	69.75	6.19	69.78	6.27	69.88	6.34	6.50
t6_10	70.43	6.00	70.50	6.00	70.88	4.68	71.49	4.88	71.82	5.29	72.48	5.56	73.06	5.48	73.15	5.54	73.48	5.69	73.62	5.81	3.50
t6_11	58.67	7.00	61.62	7.72	64.35	7.35	64.52	7.27	66.31	7.22	66.69	6.44	67.33	6.51	67.74	6.57	67.80	6.61	68.57	6.38	7.00
t6_12	66.18	5.50	66.39	6.00	68.90	6.31	69.64	6.70	70.06	6.84	70.35	6.64	70.75	6.82	71.17	6.84	71.64	6.85	71.67	6.87	8.00
t6_13	60.61	6.50	61.48	6.25	62.21	4.87	62.38	5.39	62.47	5.71	62.64	5.75	63.66	5.92	64.13	6.11	64.72	6.20	64.81	6.32	5.00
t6_14	65.98	8.00	66.44	7.50	66.48	7.34	67.20	7.50	67.31	7.40	67.38	7.50	67.94	7.43	68.41	7.44	68.74	7.29	68.89	7.26	9.00
t6_15	60.66	6.00	67.74	6.65	67.91	6.76	67.98	6.12	68.14	6.19	69.35	6.31	69.60	6.01	70.36	6.12	70.41	6.21	71.11	6.37	7.00
t6_16	53.49	4.00	54.99	4.00	56.38	4.64	57.14	4.97	58.80	5.07	58.84	5.21	59.98	5.51	61.43	5.68	62.52	5.81	62.66	5.47	6.00
t6_17	60.79	5.50	61.19	6.25	62.32	6.49	64.42	6.38	64.87	6.31	65.21	6.41	65.22	6.36	65.79	6.55	65.87	6.60	66.39	6.68	5.50
t6_18	60.78	3.00	62.41	3.49	63.21	4.29	63.43	4.22	63.89	4.57	63.93	4.64	64.00	4.96	64.17	5.09	64.28	5.13	64.53	5.26	2.00
t6_19	63.53	7.00	65.23	7.00	66.51	6.37	66.70	5.80	67.25	4.90	67.28	5.07	67.56	4.79	67.95	4.87	68.00	5.04	68.22	5.28	4.00
t6_20	59.61	8.00	63.87	5.21	64.87	5.92	64.98	6.06	65.90	6.04	66.39	6.27	66.79	6.23	66.80	6.32	67.04	6.08	67.14	6.21	7.00
t7_1	49.41	6.00	55.15	6.47	55.53	6.95	56.08	6.84	57.73	6.69	57.84	6.81	58.37	6.71	58.64	6.74	59.45	6.72	59.65	6.56	2.00
t7_2	71.27	5.00	71.33	5.00	71.44	4.34	72.25	4.98	73.50	4.98	73.83	5.14	74.02	5.19	74.06	5.28	74.16	4.74	74.46	4.67	5.00
t7_3	57.73	7.00	59.91	7.00	59.93	6.67	59.99	6.75	60.31	6.90	60.94	6.92	61.64	6.93	61.76	6.82	61.80	6.52	62.22	6.42	6.00
t7_4	60.59	7.00	62.34	7.24	62.90	7.49	63.29	7.37	64.71	7.11	65.90	7.17	66.20	7.15	67.41	7.13	67.57	7.07	67.87	7.02	4.00
t7_5	66.26	4.00	67.65	4.49	68.95	4.33	69.78	4.72	69.85	5.15	70.36	5.28	70.72	5.38	70.82	4.98	71.03	4.88	71.86	5.25	2.00
t7_6	66.12	4.00	66.43	5.74	67.09	5.17	67.70	5.37	67.70	5.11	68.10	5.17	68.27	5.56	68.81	5.26	69.20	5.38	69.25	5.82	3.00
t7_7	55.68	6.00	57.78	6.49	58.13	5.68	59.27	6.11	59.41	6.28	59.74	6.24	60.47	6.21	60.78	6.12	62.16	6.31	62.47	6.01	4.00

t7_8	65.63	5.00	68.73	5.00	69.40	5.62	69.49	5.36	70.15	4.73	70.52	5.39	70.67	5.41	70.72	5.42	70.91	5.58	71.37	5.39	2.00
t7_9	69.53	5.50	70.34	5.25	71.33	4.86	71.54	4.41	71.98	5.00	72.32	5.15	72.57	4.93	72.71	4.82	72.73	4.79	73.01	5.00	2.00
t7_10	64.69	4.00	65.67	4.98	66.08	5.48	67.52	5.13	67.75	5.20	67.95	5.48	68.97	5.49	69.21	5.20	69.32	5.23	69.67	4.75	6.00
t7_11	66.93	7.00	68.43	7.00	70.72	7.00	71.58	6.33	71.77	6.27	71.88	6.38	71.96	6.33	72.00	6.35	72.10	6.21	72.91	6.28	6.00
t7_12	64.71	5.00	65.00	6.49	65.29	6.66	66.61	6.26	68.13	6.40	69.71	6.63	69.72	6.55	69.74	6.94	69.82	6.95	70.29	6.77	7.00
t7_13	61.69	9.00	67.62	7.89	69.30	7.04	70.81	6.93	70.84	6.68	70.86	6.80	71.12	5.92	71.59	6.04	71.64	5.99	72.01	6.08	4.00
t7_14	69.94	5.00	70.19	5.00	70.67	5.65	71.02	5.62	71.40	5.31	71.57	5.10	72.72	4.95	73.08	4.73	73.26	4.96	73.50	5.15	4.00
t7_15	64.22	4.00	64.77	4.99	67.27	5.61	67.31	6.06	67.38	6.05	67.49	6.20	67.55	6.31	67.89	6.45	67.94	6.51	68.19	6.60	5.00
t7_16	62.79	6.00	64.70	6.00	66.64	6.46	66.86	6.82	66.92	6.86	67.17	6.88	67.31	6.69	67.44	6.73	67.64	6.55	67.96	6.45	6.00
t7_17	62.58	6.50	65.18	6.26	66.14	6.34	67.95	5.58	68.17	5.84	69.78	4.99	69.85	5.25	70.05	5.11	70.09	5.05	71.19	5.39	6.00
t7_18	64.16	6.50	67.12	5.79	67.60	5.86	67.91	5.18	68.37	4.96	69.19	5.19	70.03	5.17	70.37	5.43	71.98	5.58	72.01	5.44	5.00
t7_19	68.22	7.00	69.04	6.26	69.49	6.50	70.08	6.74	71.44	6.97	71.67	7.05	71.88	7.05	71.98	7.16	72.06	7.09	72.10	6.80	6.50
t7_20	73.25	7.00	73.28	6.00	73.37	5.83	73.55	6.12	73.62	5.90	73.69	5.75	74.22	5.92	74.24	5.87	74.67	5.99	74.86	5.71	3.00
t8_1	61.04	6.00	66.14	6.00	66.37	5.69	68.89	5.76	69.80	5.28	71.22	5.10	71.90	5.33	72.11	5.24	72.48	5.26	72.99	5.37	6.50
t8_2	68.75	7.00	68.99	7.50	69.75	6.69	71.11	6.53	71.75	5.70	73.01	5.81	73.23	5.84	73.42	5.97	73.49	6.07	73.53	6.15	5.50
t8_3	66.33	5.50	70.51	6.18	71.01	6.43	71.35	6.56	71.60	6.37	71.85	6.62	71.94	6.80	71.98	6.71	72.12	6.74	72.47	6.81	7.00
t8_4	73.24	7.50	73.84	5.78	73.95	6.18	74.02	6.38	74.16	6.21	75.73	6.03	75.77	6.15	75.78	6.25	76.07	6.33	76.09	6.21	4.00
t8_5	72.64	7.00	72.92	7.00	73.11	5.35	73.26	5.51	73.60	5.61	74.23	5.83	74.59	5.78	74.63	5.99	74.76	6.09	74.83	6.08	7.00
t8_6	55.23	2.00	55.65	3.99	57.70	4.63	58.93	5.19	59.44	5.53	59.88	5.76	60.56	6.06	60.83	6.11	60.90	6.26	61.60	6.37	7.00
t8_7	65.37	7.00	66.98	6.76	67.76	6.52	68.25	6.39	68.43	6.51	68.44	6.59	68.86	6.51	69.48	6.57	69.50	6.66	69.53	6.69	8.00
t8_8	70.43	7.00	71.05	7.98	71.29	7.17	71.77	6.65	71.95	6.42	72.00	6.28	72.18	6.44	72.21	6.39	72.23	6.35	72.33	6.41	7.00
t8_9	65.99	4.00	66.21	5.24	66.98	5.98	67.18	5.25	67.75	5.01	67.76	4.85	68.16	4.59	68.60	4.64	68.64	4.89	68.77	4.81	6.50
t8_10	63.82	4.00	64.40	4.99	66.72	4.68	67.39	4.52	68.43	4.43	68.50	4.82	68.64	4.71	68.92	4.75	70.27	4.67	70.35	4.53	5.00
t8_11	65.52	6.50	70.09	6.27	70.64	6.19	71.04	5.92	71.34	5.84	71.46	6.10	71.91	5.82	71.94	5.78	72.10	5.91	72.41	5.92	5.50
t8_12	66.53	7.00	70.15	7.00	70.80	7.15	70.83	7.35	70.87	7.29	71.00	7.08	71.26	6.94	71.91	7.29	71.93	7.31	72.00	7.28	7.00
t8_13	62.20	5.00	62.84	5.99	63.59	5.67	63.97	6.11	64.12	6.29	64.57	6.24	64.64	6.48	64.70	6.55	64.92	6.06	65.80	6.05	6.50
t8_14	69.99	6.00	70.19	5.50	70.23	6.00	70.57	6.25	71.42	6.29	71.46	6.17	71.46	6.28	71.63	6.49	72.28	6.54	72.54	6.58	7.00
t8_15	68.41	7.00	69.30	7.00	69.53	7.16	70.19	7.12	71.06	7.10	71.33	7.09	71.34	7.07	71.46	7.12	71.48	6.95	71.49	7.24	8.50
t8_16	67.37	4.50	67.85	4.75	67.98	5.49	68.10	5.49	68.21	5.59	68.66	5.50	68.78	5.71	69.13	5.68	69.51	5.40	69.66	5.55	6.00
t8_17	68.06	6.00	69.48	5.03	70.24	5.34	70.41	5.26	70.83	4.83	70.99	5.09	71.19	5.22	71.31	4.95	71.33	4.85	72.19	4.96	7.00
t8_18	64.28	7.00	66.58	7.00	66.63	7.00	67.91	6.53	69.76	6.79	70.22	6.67	70.23	6.72	71.10	6.75	71.33	6.72	71.42	6.75	8.00
t8_19	70.27	4.00	70.88	5.23	71.07	4.83	71.78	4.99	71.90	5.09	72.03	5.39	72.09	5.34	72.38	5.54	72.57	5.69	72.93	5.82	7.00

t8_20	68.59	7.00	71.68	5.84	73.18	6.19	73.87	6.26	74.36	6.39	74.59	6.18	74.62	6.29	75.02	6.03	75.09	5.98	75.35	5.62	8.00
t9_1	66.85	6.50	69.33	5.79	69.98	5.54	70.78	5.88	71.76	5.72	71.84	5.61	71.97	5.60	72.34	5.76	72.66	5.78	72.82	5.80	7.00
t9_2	67.83	7.50	69.24	8.71	70.45	8.02	70.54	8.25	71.13	7.83	71.64	6.93	71.79	7.07	72.04	7.00	72.39	6.95	72.41	6.59	9.00
t9_3	67.05	7.50	68.41	7.74	69.03	7.18	69.14	6.41	69.49	5.85	69.63	5.72	69.73	5.89	70.21	6.08	70.26	6.02	70.42	6.11	7.00
t9_4	59.34	4.00	59.44	5.50	59.75	5.00	59.79	5.25	60.73	5.30	61.59	4.77	61.85	5.01	61.94	5.31	61.97	5.33	62.83	5.30	4.00
t9_5	71.47	5.00	72.64	5.97	73.23	5.34	73.37	5.02	73.83	5.30	74.01	4.78	74.03	4.94	74.43	5.01	74.43	4.69	74.47	4.82	6.00
t9_6	67.51	6.50	69.53	6.02	69.79	6.50	70.37	5.90	70.78	5.83	70.79	5.62	70.99	5.74	71.00	5.71	71.19	5.42	71.28	5.38	7.50
t9_7	67.70	7.00	69.30	6.76	69.38	5.86	69.71	5.17	70.26	5.33	70.47	5.51	70.80	4.77	70.90	4.74	70.98	4.55	71.42	4.83	5.50
t9_8	63.30	4.00	64.68	5.22	68.54	5.30	68.93	5.01	69.22	5.10	69.73	5.23	70.46	5.27	70.50	5.29	71.15	5.36	71.22	5.51	5.50
t9_9	69.57	7.00	70.67	6.76	70.82	7.16	73.06	7.13	73.50	7.10	73.81	6.34	73.89	6.55	74.40	6.60	74.48	6.44	74.68	6.58	6.50
t9_10	63.05	5.00	64.63	5.00	65.55	5.48	65.86	5.84	66.44	5.87	66.70	5.89	68.36	5.97	69.31	6.19	69.41	5.98	70.34	5.98	5.00
t9_11	58.69	7.50	68.13	6.67	68.89	5.61	70.43	5.80	71.15	5.92	71.18	5.50	71.41	5.43	71.53	5.61	71.77	5.45	71.77	5.32	9.00
t9_12	72.72	5.00	73.61	5.24	73.70	4.51	73.79	5.12	73.98	5.10	74.05	5.08	74.16	5.00	74.19	5.12	74.81	5.16	74.99	5.24	5.00
t9_13	65.16	7.00	65.34	6.25	67.35	6.17	67.78	6.01	67.90	6.49	68.04	6.09	68.08	5.80	68.28	5.59	68.28	5.74	68.50	5.96	7.00
t9_14	63.25	4.00	64.74	5.70	65.69	5.16	65.90	4.88	67.02	5.28	67.17	4.92	67.55	4.66	67.68	4.88	67.81	4.79	67.85	4.71	6.00
t9_15	69.73	4.00	70.94	4.97	71.03	5.31	71.26	5.35	71.52	5.87	71.67	5.89	71.84	5.63	71.94	5.79	72.22	5.92	72.58	6.02	9.00
t9_16	62.46	6.00	67.67	7.35	68.71	7.25	69.05	7.08	69.89	6.16	70.41	6.21	70.48	6.64	70.99	6.12	71.03	6.11	71.42	6.10	6.50
t9_17	73.01	3.00	73.40	4.98	73.86	4.98	73.98	4.74	75.27	4.79	76.11	4.90	76.51	4.91	77.51	5.13	77.77	5.30	77.94	5.44	5.00
t9_18	65.74	2.00	67.61	4.16	68.67	5.06	69.40	5.16	70.89	5.49	71.01	5.64	71.09	5.82	71.31	6.07	72.84	6.16	75.76	6.22	3.00
t9_19	60.61	6.50	65.07	6.73	66.26	6.51	66.30	5.45	66.66	5.65	67.16	5.55	68.11	5.54	68.57	5.48	69.50	5.58	70.28	5.79	4.50
t9_20	59.86	6.50	64.48	6.73	64.49	6.66	65.48	6.04	65.87	5.84	66.00	5.79	67.59	5.95	67.98	5.84	68.01	5.96	68.02	6.14	5.50
t10_1	71.40	8.00	71.57	7.25	71.79	7.17	72.06	7.25	72.29	6.91	72.48	6.44	72.73	6.79	72.86	7.18	72.96	7.37	72.96	7.34	9.50
t10_2	66.30	6.50	67.31	6.26	67.57	6.50	67.82	6.13	67.89	6.30	68.66	6.25	68.78	6.36	69.02	6.55	69.03	6.49	69.93	6.54	7.00
t10_3	67.85	9.00	69.80	7.57	70.21	7.07	70.75	7.05	70.95	6.85	71.52	5.94	71.79	6.15	72.02	6.25	72.03	6.12	72.03	6.39	8.50
t10_4	63.55	7.00	64.12	5.76	67.59	5.68	67.62	5.87	68.25	5.80	68.98	5.23	70.35	5.39	70.76	5.56	71.03	5.65	71.64	5.68	6.50
t10_5	66.10	5.00	67.44	5.24	67.99	4.84	68.57	5.36	68.92	5.10	69.44	5.01	69.81	5.27	69.96	5.42	69.97	5.58	70.05	5.24	6.00
t10_6	68.91	4.00	68.98	5.00	70.19	4.68	70.25	5.97	70.40	5.39	70.56	5.65	70.68	5.70	70.95	5.62	71.05	5.44	71.27	5.40	7.00
t10_7	62.35	7.00	64.70	7.00	65.87	7.00	66.64	6.65	66.71	6.43	66.77	6.05	67.12	6.04	67.26	6.04	67.70	5.82	68.82	5.84	6.00
t10_8	69.02	5.50	69.21	6.25	71.30	6.17	71.35	5.42	71.58	5.44	71.88	5.21	72.17	5.18	72.36	5.22	72.45	5.09	72.80	4.99	5.00
t10_9	65.00	2.00	66.33	4.44	67.06	5.26	67.90	5.43	68.52	5.54	68.57	5.53	68.88	5.60	69.01	5.53	69.22	5.52	69.27	5.62	6.00
t10_10	64.67	7.00	66.32	7.00	66.76	7.32	66.77	7.24	67.22	6.91	67.39	6.60	68.23	6.65	68.41	6.69	68.42	6.73	68.68	7.04	6.50
t10_11	63.34	4.00	63.34	5.00	63.75	5.33	64.87	5.01	65.39	4.82	65.92	4.85	66.47	4.73	66.75	4.65	66.96	4.68	67.16	4.71	5.00

t10_12	67.69	4.00	67.89	3.00	68.65	3.33	69.24	4.21	70.70	4.72	70.70	5.00	72.07	5.00	72.20	5.00	72.58	5.39	72.62	5.53	5.00
t10_13	74.21	4.00	75.17	5.70	76.11	5.79	76.18	6.55	76.50	6.35	76.66	5.67	76.76	5.45	77.24	5.45	77.81	5.56	77.89	5.33	6.00
t10_14	72.70	8.00	72.71	7.75	73.26	7.51	73.33	6.65	73.54	6.13	73.64	5.95	74.01	5.27	74.24	5.71	74.42	6.06	74.47	6.15	6.50
t10_15	67.02	8.00	69.51	6.11	70.17	7.02	70.93	6.55	71.14	6.63	71.51	6.46	72.03	6.40	72.23	6.41	72.41	5.96	72.77	6.14	10.00
t10_16	68.11	6.50	70.82	7.20	71.70	7.45	72.13	7.00	72.44	6.90	72.95	6.61	73.25	6.93	73.41	6.59	74.03	6.73	74.24	6.75	9.00
t10_17	66.00	7.00	66.12	6.50	66.13	7.00	66.53	6.63	66.59	6.70	66.70	6.83	68.58	6.46	68.75	6.52	69.12	6.62	69.45	6.70	7.50
t10_18	71.42	6.00	71.96	7.48	72.16	7.98	72.99	7.03	73.46	6.46	74.07	5.77	74.70	5.61	74.78	5.71	74.78	5.84	74.83	5.90	4.50
t10_19	56.84	7.50	60.54	5.37	61.04	4.93	63.93	4.72	65.45	4.95	65.66	4.81	66.75	5.33	67.07	5.18	67.68	5.26	67.68	5.28	3.00
t10_20	59.46	6.50	61.12	6.74	61.82	5.54	62.24	5.77	62.60	6.01	63.74	6.16	64.16	6.27	67.82	6.35	68.65	6.50	69.72	6.46	4.00

Annex G: Full Experiment of Domain 3

Where: θ_i the cosine similarity angle of i where i ranges from 1 to 10.
SG_i is the calculated score of i nearest neighbor pre- graded essays and MG is the manual grade.

	θ_1	SG1	θ_2	SG2	θ_3	SG3	θ_4	SG4	θ_5	SG5	θ_6	SG6	θ_7	SG7	θ_8	SG8	θ_9	SG9	θ_{10}	SG10	MG
t1_1	57.28	7.00	59.96	7.00	60.13	6.68	61.91	6.28	63.24	6.32	64.06	6.65	64.27	6.70	64.89	6.51	65.08	6.45	65.27	6.46	6.50
t1_2	47.00	4.00	49.70	4.49	52.00	4.65	53.41	4.50	54.63	4.87	55.04	5.19	55.97	5.17	56.22	5.33	56.64	5.14	57.45	5.40	4.00
t1_3	63.17	8.00	64.35	7.51	64.79	7.18	65.61	6.42	65.92	6.53	65.92	6.68	66.17	6.87	66.37	6.88	66.40	6.84	66.52	6.95	6.00
t1_4	60.92	5.00	64.45	5.71	65.25	6.11	65.96	5.62	66.18	5.88	66.60	5.20	66.61	6.59	66.93	6.28	66.99	6.25	67.26	6.13	4.00
t1_5	63.88	7.00	66.54	6.76	67.19	7.15	67.77	6.88	68.05	6.90	69.86	6.07	69.96	6.67	70.01	6.82	70.06	6.89	70.88	6.63	4.00
t1_6	68.67	6.00	68.73	6.50	68.97	7.00	69.07	7.24	69.44	7.39	69.63	7.41	69.65	7.49	69.93	7.43	70.20	7.49	70.25	7.40	5.00
t1_7	61.20	4.00	64.64	4.00	65.67	4.31	66.31	5.64	66.89	6.08	67.33	6.37	67.88	6.52	68.67	6.35	68.91	6.26	69.20	6.38	4.50
t1_8	56.51	8.00	61.46	7.54	61.82	8.00	62.92	8.00	62.98	8.00	63.35	7.76	63.73	7.39	63.77	7.35	64.01	7.36	64.31	7.23	6.00
t1_9	60.08	6.50	60.83	6.75	61.76	6.18	61.94	5.65	62.69	6.29	63.25	6.32	63.33	6.14	63.35	6.25	63.85	6.22	64.29	6.11	4.00
t1_10	63.98	7.50	65.55	7.74	65.79	8.15	66.49	7.15	66.54	7.12	66.83	7.10	66.98	6.68	67.57	6.72	67.77	6.69	68.16	6.86	8.00
t1_11	37.43	7.00	42.78	6.04	43.49	5.70	44.65	5.30	45.79	5.62	46.07	5.92	48.53	5.67	50.17	5.65	50.93	5.48	51.26	5.80	4.00
t1_12	56.94	4.00	57.19	5.49	57.83	5.99	57.99	5.50	60.46	5.69	60.51	5.97	62.58	5.97	62.66	6.09	63.15	6.13	63.46	6.03	5.00
t1_13	59.58	6.50	61.66	5.29	62.56	6.48	63.09	6.60	63.21	6.68	64.09	6.26	64.61	6.36	64.81	6.55	64.82	6.65	65.59	6.87	6.00
t1_14	64.28	4.00	65.34	4.49	65.40	5.15	65.41	5.24	65.45	5.49	66.09	5.73	67.07	6.17	67.36	6.33	67.68	6.40	67.77	6.45	5.50
t1_15	65.18	9.00	66.25	8.27	67.52	7.86	67.61	7.78	68.45	7.35	69.40	7.60	69.59	7.52	69.79	7.12	70.05	7.16	70.36	6.96	7.00
t1_16	60.93	7.00	61.00	7.00	62.28	7.16	64.01	7.24	64.30	6.63	64.92	6.84	65.99	6.80	66.02	6.77	66.13	6.74	66.18	6.95	9.00
t1_17	66.85	6.50	67.54	6.01	69.45	6.16	69.54	6.83	70.09	6.96	70.20	7.12	70.40	7.24	70.69	7.33	70.73	7.45	70.91	7.50	6.00
t1_18	62.81	6.50	67.30	6.73	67.62	6.50	67.65	6.50	67.82	6.79	68.12	6.98	68.72	7.18	69.90	7.05	70.10	7.09	70.50	6.95	5.00
t1_19	67.02	6.50	67.38	6.00	68.03	6.33	68.71	6.37	69.06	6.49	69.12	6.65	69.28	6.63	69.69	6.56	70.09	6.55	70.18	6.40	7.50
t1_20	63.78	9.00	66.05	8.76	66.60	8.84	67.71	8.52	67.73	8.24	67.74	8.36	67.83	8.17	68.28	8.21	68.52	8.40	68.75	8.17	8.00
t2_1	70.70	7.50	70.96	7.00	71.45	7.00	71.72	7.00	71.81	6.42	71.93	6.67	72.31	6.79	72.43	6.45	72.58	6.40	72.71	6.41	6.00
t2_2	56.57	4.00	60.08	4.48	63.89	4.33	64.52	4.93	64.54	5.40	64.76	5.95	65.00	6.03	65.46	5.68	65.50	5.92	65.72	6.11	6.00
t2_3	66.19	4.00	67.57	4.00	67.68	4.33	68.31	5.45	68.81	5.84	68.96	6.19	69.00	6.43	69.31	6.14	69.43	6.29	69.74	6.36	7.00
t2_4	63.56	9.00	65.57	8.28	65.93	7.38	66.18	7.29	66.51	7.33	67.36	7.51	67.91	7.05	67.99	6.69	68.16	6.72	68.18	6.80	8.50
t2_5	50.58	5.00	50.90	4.50	57.38	4.35	58.27	5.06	59.17	4.87	59.49	4.74	60.38	5.29	60.59	5.03	61.23	5.32	61.50	5.43	6.50
t2_6	65.01	4.00	67.64	4.00	67.80	5.28	68.20	5.33	68.43	5.08	68.61	5.38	68.88	5.94	69.00	6.07	69.84	6.37	69.84	6.52	5.00
t2_7	55.64	4.00	58.15	4.00	59.23	4.80	59.91	4.61	60.75	4.68	61.12	4.57	61.22	4.50	62.12	4.90	62.55	5.12	62.66	5.34	5.00
t2_8	63.53	5.00	65.46	6.21	65.83	6.14	65.94	6.11	66.11	5.70	66.59	5.42	67.51	5.57	67.94	5.39	68.42	5.60	68.42	5.69	7.50
t2_9	60.89	4.00	65.50	6.30	66.92	7.12	67.04	6.40	68.06	5.96	68.09	6.20	68.43	6.37	68.65	6.67	68.68	6.55	68.73	6.59	7.00
t2_10	63.22	7.50	63.30	7.75	64.49	6.54	65.51	5.93	65.93	6.70	66.67	7.06	66.84	7.32	67.16	7.57	67.17	7.57	67.52	7.23	8.00
t2_11	63.02	4.00	64.41	5.22	65.24	4.83	66.03	5.34	66.20	5.28	66.22	5.63	66.30	6.03	66.72	5.67	67.32	5.49	67.34	5.35	6.00

t2_12	55.04	7.00	56.66	7.00	58.31	6.04	58.72	6.27	59.92	5.85	60.59	5.40	60.63	5.35	60.92	5.19	61.01	5.43	61.83	5.30	6.50
t2_13	62.26	6.00	64.48	7.44	65.51	8.09	65.67	7.11	65.82	6.90	66.30	6.92	66.36	7.00	66.44	7.12	66.93	6.79	67.66	6.67	6.50
t2_14	57.41	5.00	57.90	6.49	57.92	5.66	58.82	6.48	59.23	6.58	59.41	6.57	61.50	6.56	61.57	6.78	62.23	6.76	62.36	6.50	6.00
t2_15	51.12	4.00	56.42	4.00	58.64	4.31	58.96	4.23	59.36	4.19	60.66	4.16	61.83	4.53	62.28	4.80	63.05	4.62	63.16	4.93	5.00
t2_16	68.50	9.00	69.90	8.27	70.10	7.70	70.26	6.56	70.51	6.25	70.63	6.21	70.88	6.32	71.17	6.52	71.24	6.68	71.39	6.76	7.50
t2_17	48.87	4.00	58.29	5.33	58.76	4.93	58.89	4.94	60.44	5.32	60.47	5.11	60.72	5.56	61.01	5.38	61.08	5.60	61.37	5.69	7.00
t2_18	53.31	5.50	58.83	6.20	59.22	5.51	61.20	5.16	62.46	5.49	63.05	5.42	63.59	5.62	63.62	5.89	63.68	5.70	63.80	5.82	6.50
t2_19	65.58	9.00	68.26	8.76	68.29	8.04	68.36	7.79	68.55	8.02	69.04	7.38	69.41	7.40	69.54	7.47	69.79	7.52	69.81	7.76	9.00
t2_20	67.16	10.0	68.08	8.53	68.61	7.71	69.24	7.90	69.49	7.54	69.55	7.77	69.57	7.73	69.61	7.77	69.94	7.84	69.97	7.71	10.00
t3_1	68.19	4.00	68.67	5.24	69.58	5.80	69.65	5.37	69.96	5.68	71.06	5.96	71.59	6.16	71.64	6.20	72.05	6.23	72.40	6.12	5.00
t3_2	65.86	5.50	66.20	6.00	67.02	5.67	67.50	5.63	69.10	5.79	69.12	5.51	69.97	5.32	70.12	5.68	70.37	5.71	70.49	6.01	5.50
t3_3	63.01	7.50	63.06	5.75	63.17	6.33	63.23	6.75	64.28	6.70	65.34	7.06	66.52	7.44	66.56	7.05	66.80	7.14	67.11	7.04	9.50
t3_4	68.79	5.50	69.98	7.20	70.07	7.30	70.34	7.23	70.65	7.09	70.98	6.75	71.04	6.65	71.55	6.34	71.71	6.51	71.97	6.51	6.00
t3_5	57.42	7.00	62.50	5.62	62.91	5.58	64.12	6.25	64.16	6.49	64.24	6.80	64.74	6.76	64.86	6.44	64.99	6.44	65.59	6.22	7.50
t3_6	60.71	7.00	63.00	7.00	63.06	7.00	63.48	7.12	63.49	7.00	64.16	7.00	64.28	6.59	64.98	6.28	65.49	6.30	65.55	6.18	8.00
t3_7	68.01	7.00	69.43	7.48	69.87	7.65	70.38	7.49	70.70	7.97	70.86	7.74	71.33	7.50	71.38	7.62	71.80	7.61	71.82	7.60	8.50
t3_8	68.69	5.00	68.89	4.50	68.91	4.34	69.60	4.25	69.80	4.69	70.06	5.22	70.12	5.19	70.28	5.77	70.57	5.90	70.73	5.72	7.50
t3_9	61.43	3.00	62.46	5.46	63.99	5.31	64.79	5.24	64.79	5.67	64.84	5.40	65.26	5.21	66.18	5.42	66.22	5.27	66.54	5.39	3.00
t3_10	63.77	6.00	64.48	6.00	64.51	6.33	65.55	6.97	66.00	6.40	66.89	6.19	67.56	6.42	67.75	6.14	68.21	5.93	68.91	5.93	8.50
t3_11	69.30	8.00	69.72	7.75	70.19	7.34	70.73	6.30	70.87	5.85	70.93	6.04	71.86	6.30	72.01	6.27	72.13	6.45	72.27	6.54	6.50
t3_12	62.40	5.00	62.52	5.75	63.97	5.83	64.45	6.47	64.92	6.57	65.13	6.56	65.79	6.62	66.20	6.31	66.21	6.33	66.28	6.44	6.50
t3_13	63.92	4.00	66.04	4.00	66.50	4.80	66.70	4.97	66.71	5.36	66.86	5.14	67.13	5.40	67.16	5.47	68.12	5.58	68.13	5.67	7.00
t3_14	64.57	8.50	66.86	8.02	67.32	7.86	67.32	6.92	68.49	7.12	68.63	6.95	68.96	6.89	68.98	7.08	69.03	7.28	69.13	7.11	5.50
t3_15	67.63	5.00	69.27	5.96	70.23	6.29	70.47	6.22	70.61	5.80	71.39	5.91	71.93	5.66	72.11	5.47	72.15	5.32	72.42	5.47	8.00
t3_16	58.04	4.00	60.75	4.00	62.38	4.00	63.99	4.57	64.62	5.01	65.12	5.39	65.30	5.67	65.35	5.94	65.82	5.74	65.96	5.86	7.50
t3_17	67.92	5.00	67.93	5.25	68.43	6.65	68.59	6.61	69.09	7.07	69.10	6.90	69.11	6.35	69.75	6.43	69.83	6.39	70.31	6.16	7.00
t3_18	64.75	7.00	65.22	7.25	66.01	7.17	66.40	6.88	67.23	6.91	67.56	6.45	69.15	6.52	69.27	6.57	69.27	6.52	69.62	6.51	6.50
t3_19	66.95	4.00	67.13	5.74	67.33	6.32	67.41	6.99	67.63	6.79	67.70	6.91	68.75	6.38	68.81	6.33	68.98	6.51	69.03	6.75	6.00
t3_20	58.70	7.00	60.08	5.53	60.74	6.01	60.77	6.25	60.90	6.40	62.40	6.02	62.46	6.29	62.84	6.31	63.00	6.44	63.64	6.35	9.00
t4_1	67.70	3.00	69.26	4.45	70.55	4.93	70.85	5.65	71.41	5.90	71.91	6.15	72.20	6.26	72.29	6.23	72.66	6.41	73.00	6.60	2.00
t4_2	68.77	8.00	69.27	6.02	69.64	6.18	69.82	5.65	69.99	5.81	70.15	5.84	70.22	5.59	70.23	5.39	70.32	5.62	70.42	6.05	6.00
t4_3	62.04	6.00	64.63	6.72	65.96	7.12	66.59	6.97	67.24	7.26	67.76	6.99	68.18	6.60	68.30	6.93	68.44	7.14	69.11	7.13	9.00

t4_4	52.53	4.00	56.28	5.43	61.44	5.31	61.58	5.35	61.99	5.10	62.15	4.94	62.22	4.68	62.73	4.60	63.67	4.84	63.82	5.08	7.00
t4_5	68.87	8.50	69.66	9.24	69.69	9.49	71.00	9.49	71.66	8.84	71.95	8.56	72.19	8.75	72.20	8.54	72.27	8.38	72.35	8.30	7.00
t4_6	70.36	10.0	71.38	9.27	72.45	8.87	72.49	8.42	72.77	8.25	74.06	7.91	74.14	7.79	74.18	7.99	74.24	8.04	74.54	7.99	6.00
t4_7	69.55	5.00	70.56	5.49	70.93	5.01	71.01	5.74	71.12	5.98	71.14	5.82	71.54	6.12	71.88	6.17	72.26	5.94	72.28	6.04	4.00
t4_8	51.49	4.00	53.41	4.00	53.88	4.82	56.20	4.62	56.46	5.07	57.15	4.91	57.26	4.65	59.91	4.80	60.79	4.96	60.92	5.23	5.00
t4_9	57.07	4.00	57.80	5.73	58.44	6.31	59.41	5.99	60.04	6.19	60.16	6.08	62.69	6.13	62.71	6.34	62.94	6.51	63.25	6.74	7.00
t4_10	59.27	4.00	61.71	4.00	62.13	5.13	62.43	5.70	63.36	5.57	63.37	5.96	63.45	5.69	63.49	5.97	64.81	6.28	64.86	6.07	7.00
t4_11	62.63	7.50	63.64	5.78	64.27	5.20	64.75	4.91	65.63	5.78	66.58	6.05	66.59	6.17	66.60	6.15	66.89	6.14	67.03	5.94	9.00
t4_12	65.52	6.00	67.09	6.97	67.21	6.98	68.06	6.98	68.81	7.17	69.15	7.15	69.24	7.53	69.52	7.59	69.96	7.83	70.24	7.80	7.50
t4_13	64.37	6.00	64.99	6.00	66.72	6.32	67.35	6.48	67.48	6.67	67.52	6.25	67.60	6.35	68.34	6.60	68.49	6.49	68.80	6.25	6.50
t4_14	70.75	3.50	71.80	4.96	71.87	4.65	72.66	5.09	73.18	5.45	73.23	5.69	73.40	5.87	73.86	5.94	74.11	6.15	74.64	6.05	6.00
t4_15	60.37	4.00	60.49	6.50	61.87	6.50	62.44	6.50	62.67	6.59	63.05	6.58	63.44	6.64	63.63	6.32	63.68	6.08	64.07	6.16	8.00
t4_16	67.17	9.00	67.55	9.00	67.87	8.34	68.18	7.28	69.32	7.32	69.55	7.19	69.68	7.10	69.90	7.26	69.90	7.13	69.96	6.83	6.00
t4_17	62.70	3.00	63.20	5.48	64.68	5.33	64.98	4.77	65.68	5.10	65.73	5.32	65.87	5.14	66.56	5.18	66.64	5.42	66.79	5.48	9.00
t4_18	54.36	7.50	55.46	5.77	57.94	5.21	60.41	5.62	60.51	5.78	62.33	5.96	62.41	5.71	63.37	5.80	63.48	5.62	63.68	5.70	7.00
t4_19	64.99	3.00	67.00	4.92	67.23	5.91	67.44	6.30	67.54	6.05	67.54	6.36	67.62	6.52	67.69	6.21	67.88	6.30	67.96	6.22	7.00
t4_20	64.67	6.00	65.48	6.49	66.87	6.65	67.47	6.38	68.36	6.87	68.45	6.11	69.08	6.29	69.32	6.26	69.51	6.23	69.61	6.30	6.50
t5_1	66.22	7.00	66.87	6.75	69.76	6.98	69.77	6.87	69.83	6.52	70.86	6.51	71.26	6.45	71.55	6.45	71.60	6.46	71.64	6.23	5.00
t5_2	65.04	7.50	65.63	8.24	66.15	7.84	66.54	7.15	66.69	7.12	67.22	7.26	67.62	6.68	68.29	6.66	68.51	6.85	68.60	6.68	8.00
t5_3	63.70	7.00	64.08	7.00	64.61	7.00	65.00	6.63	65.19	6.80	65.87	7.31	66.24	7.27	66.76	7.36	67.21	7.37	67.46	7.34	6.00
t5_4	61.30	4.00	63.07	5.46	63.81	4.99	64.62	5.47	64.70	5.57	64.86	5.88	65.24	6.03	65.64	5.79	66.05	5.71	66.10	5.92	5.50
t5_5	48.94	4.00	53.28	4.00	56.24	4.92	57.79	4.94	58.17	5.32	59.09	5.27	59.27	4.97	61.02	4.86	61.21	4.98	61.32	5.20	5.50
t5_6	61.16	6.50	61.81	6.50	63.04	5.05	63.28	5.89	63.59	5.71	64.66	5.92	66.14	6.12	66.20	6.16	66.58	6.35	66.76	6.32	8.00
t5_7	61.86	6.50	63.55	6.99	66.00	7.30	66.07	7.23	66.94	6.63	67.32	6.23	67.42	6.46	67.80	6.18	67.86	6.21	68.00	6.42	8.00
t5_8	59.53	5.50	61.04	5.99	61.34	5.67	63.29	6.33	64.81	6.36	64.90	6.31	64.92	6.40	65.87	6.47	65.99	6.52	66.44	6.74	6.00
t5_9	54.01	4.00	62.17	5.33	63.58	5.23	63.78	5.41	64.50	5.33	64.67	5.13	64.71	4.98	65.10	4.87	65.13	5.04	65.23	4.94	5.00
t5_10	60.25	6.00	61.74	5.51	64.65	6.27	64.91	6.56	65.04	6.64	65.07	6.07	65.18	6.19	65.35	5.93	65.53	5.73	66.01	5.80	7.00
t5_11	61.12	7.00	65.41	7.00	65.79	6.84	67.21	6.54	67.38	6.90	67.83	6.91	67.85	6.99	67.99	6.64	68.37	6.78	68.68	6.53	8.00
t5_12	70.31	7.00	71.27	7.24	71.96	7.01	72.64	7.12	73.19	6.82	73.68	6.39	73.81	6.40	74.68	6.36	74.91	6.42	74.97	6.47	6.00
t5_13	60.09	7.00	61.75	6.27	62.91	5.86	63.81	6.01	64.10	6.11	65.59	5.79	66.06	5.55	66.25	5.95	66.57	5.75	67.93	5.86	5.50
t5_14	51.99	5.00	54.33	4.51	54.89	5.32	56.59	5.01	56.73	4.81	58.64	5.07	58.72	4.80	59.30	5.05	60.61	5.35	60.86	5.32	4.00
t5_15	65.44	6.50	67.32	6.74	67.96	5.87	68.26	6.02	68.29	6.30	68.73	6.41	68.76	6.56	69.20	6.85	69.58	7.08	69.80	6.93	8.00

t5_16	55.41	7.00	56.70	7.00	58.47	7.16	59.26	6.65	59.71	6.90	59.87	7.24	61.02	7.14	61.15	7.12	61.20	7.11	61.79	7.01	7.00
t5_17	67.12	4.00	67.38	4.00	68.03	4.00	68.42	4.00	68.51	4.29	68.73	4.73	68.74	5.25	69.54	5.58	70.13	5.78	70.51	5.71	7.00
t5_18	62.70	4.00	64.85	5.68	66.52	5.16	67.13	5.70	67.47	5.29	68.20	5.71	68.67	5.88	68.82	5.89	69.44	5.80	69.50	5.73	8.00
t5_19	63.69	8.50	65.21	7.28	66.02	7.19	66.28	7.39	67.07	7.31	67.41	7.19	68.22	7.23	68.36	7.38	68.74	7.54	68.97	7.54	8.00
t5_20	69.95	7.00	70.43	5.27	71.11	5.18	71.56	5.50	71.67	5.69	71.97	5.90	72.75	6.24	73.55	5.99	73.71	6.09	73.86	6.13	4.00
t6_1	65.72	5.00	66.12	5.99	66.69	5.34	67.52	5.86	68.57	5.51	68.77	5.90	68.99	6.45	69.13	5.93	69.16	5.88	69.16	6.13	4.00
t6_2	60.65	5.00	63.04	5.96	65.34	6.89	66.91	7.03	67.07	7.02	67.64	7.17	67.69	6.63	67.87	6.55	68.29	6.29	68.59	6.31	6.00
t6_3	62.46	4.00	64.69	6.40	65.04	6.59	65.62	6.81	66.06	6.28	66.20	6.55	66.27	6.68	66.74	6.60	66.75	6.59	66.98	6.77	9.00
t6_4	59.56	5.00	61.33	5.97	61.41	6.47	61.71	6.60	62.60	6.10	63.10	6.24	64.48	6.34	64.78	6.07	65.49	6.27	66.75	6.29	6.00
t6_5	66.81	10.0	67.56	9.02	67.79	8.68	69.81	8.18	70.95	7.79	71.00	7.52	71.11	7.32	71.53	7.40	71.71	7.10	71.75	7.19	6.00
t6_6	60.80	7.50	66.02	7.27	67.98	6.60	68.17	6.47	68.47	6.29	68.89	6.63	69.06	6.67	69.25	6.83	69.38	6.84	69.41	6.86	7.00
t6_7	69.69	9.50	72.04	9.26	72.21	9.50	72.34	8.90	72.89	7.97	73.05	7.82	73.24	7.64	73.81	7.63	74.42	7.46	74.45	7.38	7.00
t6_8	62.78	7.00	62.79	6.00	63.23	5.34	64.34	5.86	65.31	5.89	65.38	5.59	65.87	5.71	66.85	5.51	67.65	5.76	67.74	5.92	7.50
t6_9	70.41	4.50	70.58	5.74	71.00	5.99	71.38	6.48	71.86	6.01	71.92	6.00	72.05	6.42	72.21	6.55	72.22	6.70	72.29	6.83	3.00
t6_10	68.79	7.50	70.30	5.33	70.38	6.20	70.82	6.63	70.92	6.51	71.04	6.83	71.48	6.79	71.52	6.57	71.60	6.30	71.84	6.55	8.50
t6_11	69.56	6.50	70.73	7.71	71.17	7.81	71.27	7.37	71.71	6.72	71.74	6.69	71.99	6.93	72.29	7.06	72.83	7.16	72.85	7.05	6.00
t6_12	67.57	6.50	67.89	7.24	68.66	6.52	69.84	6.75	70.30	6.05	70.90	6.19	71.04	6.10	71.28	6.32	71.68	6.44	71.69	6.45	3.00
t6_13	58.65	4.00	60.98	6.41	61.16	5.63	62.95	5.48	62.98	5.96	63.73	6.20	64.03	6.04	64.10	5.68	64.86	5.77	65.07	5.88	8.00
t6_14	69.11	7.50	70.14	5.30	70.29	6.51	70.72	6.63	70.94	6.70	71.11	6.91	71.36	6.99	71.41	6.63	71.46	6.83	71.62	6.94	8.00
t6_15	67.82	4.00	67.84	5.50	68.89	6.15	68.95	6.11	70.31	6.00	71.06	6.07	71.44	6.19	71.75	6.17	72.70	6.11	72.91	6.05	6.00
t6_16	71.14	4.00	71.26	6.24	71.97	6.33	72.27	6.73	72.29	6.88	72.84	7.06	72.90	7.05	72.99	6.98	73.21	6.67	73.43	6.75	8.00
t6_17	57.76	7.00	58.62	5.52	59.50	5.02	60.29	5.26	60.47	5.60	61.07	5.82	61.46	6.25	61.95	6.10	62.96	6.20	63.13	6.36	6.00
t6_18	67.15	7.00	67.37	6.75	67.63	7.16	69.54	6.89	69.74	6.82	69.79	6.61	69.80	6.26	70.32	5.99	70.58	5.89	70.66	6.27	3.00
t6_19	65.22	4.00	65.87	4.00	66.27	4.98	67.56	4.75	68.95	5.07	69.09	5.44	69.51	5.38	69.59	5.51	70.14	5.56	70.33	5.56	4.00
t6_20	64.67	8.00	65.01	8.50	65.08	8.33	66.22	7.77	66.82	7.53	67.78	7.14	68.28	6.73	68.98	6.76	69.03	6.78	69.50	6.53	7.00
t7_1	57.80	4.00	62.20	4.00	64.70	5.05	66.35	5.04	67.10	5.21	67.68	5.47	67.73	5.60	68.13	5.76	68.32	5.73	68.68	5.84	5.00
t7_2	66.71	7.00	66.85	7.00	66.92	5.67	67.43	5.75	67.50	6.29	67.51	6.41	67.63	6.77	67.99	6.92	68.05	6.61	68.07	6.60	8.00
t7_3	58.21	7.00	58.64	5.51	59.48	5.67	61.39	5.51	61.69	5.89	62.57	5.59	62.64	5.92	63.80	6.15	63.82	5.93	63.97	5.99	7.50
t7_4	61.75	7.00	64.40	6.76	64.57	6.84	65.46	5.93	66.03	6.22	66.12	6.35	66.22	6.57	66.77	6.62	66.86	6.61	67.48	6.60	6.00
t7_5	58.84	6.50	60.24	6.74	62.46	6.82	62.57	5.92	63.27	6.31	63.96	6.65	64.08	6.76	64.24	6.56	64.38	6.71	64.38	6.64	5.00
t7_6	45.26	4.00	46.93	4.00	47.91	4.00	52.50	4.23	53.23	4.83	54.08	5.16	54.16	5.20	54.32	5.06	55.25	4.96	55.27	5.24	4.00
t7_7	59.16	5.00	61.10	5.97	61.63	5.82	63.25	6.10	63.33	5.51	63.57	5.59	65.01	5.77	65.06	5.86	65.24	5.92	65.78	6.20	4.00

t7_8	54.73	7.00	54.82	7.25	54.99	7.17	55.23	6.63	55.44	6.11	55.58	6.26	56.06	6.36	56.73	6.68	57.56	6.66	57.65	6.55	4.00
t7_9	59.19	4.00	62.99	5.64	65.79	5.15	66.43	5.12	66.49	4.91	67.30	4.78	68.05	5.00	68.63	5.00	69.24	5.05	69.60	5.22	3.50
t7_10	60.81	7.00	62.15	6.27	62.32	5.52	62.75	5.88	63.04	6.20	63.68	6.33	63.84	6.62	64.68	6.43	64.68	6.70	64.95	6.73	7.50
t7_11	65.05	8.50	66.87	8.26	67.38	7.86	68.12	8.01	68.39	7.82	68.41	7.77	68.48	7.59	68.64	7.52	69.02	7.47	69.31	7.14	6.00
t7_12	58.91	4.00	62.55	4.00	62.79	5.12	63.04	5.81	65.28	6.21	65.39	6.56	65.70	6.61	66.24	6.43	66.87	6.44	67.03	6.22	6.50
t7_13	60.89	7.50	61.42	7.00	63.82	7.00	64.61	6.89	65.02	6.35	65.43	6.60	65.72	6.26	65.80	6.70	66.17	6.99	66.44	7.08	6.00
t7_14	59.79	4.00	61.42	5.46	62.11	5.96	62.55	6.09	63.34	5.69	63.45	5.90	63.78	6.05	64.38	6.04	64.41	5.99	64.46	6.23	7.50
t7_15	51.06	6.50	52.05	6.75	58.34	5.93	59.10	6.17	61.03	6.23	61.42	6.34	61.68	6.43	62.02	6.60	62.10	6.69	62.28	6.72	5.50
t7_16	62.38	7.00	63.64	7.00	63.67	6.02	63.96	6.62	64.44	6.79	64.60	6.67	64.98	6.85	65.08	7.23	65.54	7.42	65.66	7.00	7.50
t7_17	56.63	4.00	57.91	6.21	58.48	6.47	58.52	5.86	60.73	6.07	61.82	6.22	62.06	6.39	62.93	6.40	63.30	6.26	63.79	6.19	6.50
t7_18	57.07	8.50	61.52	6.40	63.47	6.58	64.03	6.91	64.54	6.56	64.65	6.31	64.86	6.01	65.01	6.24	65.20	6.42	65.37	6.62	9.00
t7_19	64.99	7.50	66.40	7.99	66.94	7.51	66.95	7.63	67.29	7.51	67.65	7.75	67.66	7.64	67.88	7.50	67.98	7.40	68.18	7.36	8.50
t7_20	63.70	4.00	64.53	5.48	66.88	5.33	67.53	5.72	68.15	5.77	68.27	5.96	68.79	5.70	68.91	5.85	69.54	5.67	69.55	5.83	6.50
t8_1	59.65	9.00	59.93	7.76	63.36	7.52	64.16	7.40	65.56	7.51	65.87	7.51	66.18	7.38	66.19	6.99	67.54	6.70	68.36	6.46	8.50
t8_2	64.89	7.00	65.44	7.25	66.39	7.17	66.72	7.37	67.10	7.49	67.67	6.78	67.76	6.81	67.99	7.07	68.00	7.06	68.14	6.77	8.50
t8_3	62.99	9.00	63.85	7.52	64.65	7.84	64.93	8.12	65.76	7.91	66.47	7.77	66.94	7.66	66.94	7.53	67.55	7.17	67.71	7.06	7.00
t8_4	67.72	7.00	69.30	7.48	69.68	7.97	71.40	6.84	71.48	6.77	71.83	6.73	72.29	6.38	72.54	6.73	72.84	6.71	72.88	6.83	7.50
t8_5	67.13	4.00	68.40	5.95	69.02	6.28	69.07	6.46	71.55	6.64	71.72	6.70	71.75	6.74	72.18	6.43	72.35	6.28	72.77	6.35	7.50
t8_6	65.18	5.50	66.95	4.78	67.67	6.12	68.36	6.21	68.95	6.08	69.36	6.53	69.71	6.79	70.17	6.76	70.34	6.73	70.40	6.76	5.00
t8_7	57.40	9.00	58.53	6.54	58.84	5.71	59.23	6.15	59.43	5.92	59.65	5.61	60.12	6.01	60.17	5.76	60.78	5.68	61.39	5.90	6.50
t8_8	67.60	4.50	69.88	6.16	70.01	6.75	70.70	7.52	70.88	7.61	71.30	6.89	72.15	6.84	72.63	7.08	72.66	6.97	72.74	7.07	5.00
t8_9	55.29	4.00	59.81	4.00	59.95	4.00	61.84	4.58	62.16	4.84	62.41	4.94	62.59	5.22	63.08	5.66	63.27	5.80	63.41	6.00	7.50
t8_10	57.98	5.00	58.72	4.51	59.45	4.34	62.29	4.95	62.60	5.24	64.00	5.05	64.40	5.05	64.97	4.93	65.60	5.03	65.81	5.03	6.50
t8_11	50.97	4.00	60.89	4.00	61.16	4.30	61.96	4.91	63.04	5.56	63.11	5.93	63.72	5.68	63.83	5.66	64.54	5.79	64.58	5.54	4.00
t8_12	65.80	7.00	66.22	7.99	66.26	7.99	66.56	7.75	67.38	7.70	67.96	7.75	68.35	7.85	68.49	7.99	68.51	8.09	68.71	8.13	10.00
t8_13	61.65	6.50	63.15	6.50	63.23	6.66	63.80	6.75	64.29	6.79	64.31	6.67	64.32	6.92	64.69	6.93	64.81	6.62	65.62	6.65	7.50
t8_14	67.27	8.50	67.99	8.01	72.90	8.14	73.16	8.11	73.29	7.21	73.30	7.03	73.79	7.28	73.90	7.08	73.93	7.07	74.10	7.02	6.50
t8_15	67.83	8.50	68.24	7.26	68.40	7.67	69.29	7.99	69.64	7.99	69.95	8.31	70.72	8.20	71.05	8.12	71.27	7.90	71.63	7.82	10.00
t8_16	58.10	6.50	58.24	5.25	58.76	4.51	59.19	5.37	59.39	5.59	59.75	5.49	60.00	5.15	60.25	5.37	60.28	5.39	60.35	5.35	6.00
t8_17	66.84	7.00	68.47	7.00	68.53	6.03	69.34	5.54	69.64	5.25	69.67	5.05	69.68	5.59	69.76	6.00	70.32	6.16	70.58	6.38	9.50
t8_18	59.38	5.50	59.77	7.24	61.49	5.88	63.01	6.14	63.27	6.49	63.77	6.26	63.77	6.29	64.58	5.91	64.60	6.08	65.22	6.07	7.50
t8_19	52.69	5.50	53.50	6.24	54.53	5.51	55.34	6.36	56.18	5.91	56.45	6.00	56.85	6.20	58.37	6.30	59.01	6.32	59.15	6.38	7.00

t8_20	68.07	6.50	68.77	6.75	70.05	7.14	70.48	7.11	70.78	6.33	71.37	6.28	71.76	6.44	71.93	6.51	72.08	6.35	72.09	6.41	6.00
t9_1	58.97	6.50	61.34	7.70	62.75	7.17	63.64	7.48	64.42	7.21	64.68	7.17	65.08	6.89	65.32	7.19	65.87	7.48	66.09	7.52	7.50
t9_2	72.88	5.50	73.11	6.49	73.47	6.99	74.11	6.27	74.35	6.22	74.54	6.11	74.56	6.36	74.70	6.44	75.25	6.50	75.51	6.59	8.00
t9_3	62.22	4.00	65.13	4.47	65.25	5.28	65.46	5.82	65.68	5.95	65.76	5.64	66.38	6.16	66.54	6.26	66.59	6.23	66.69	6.02	6.00
t9_4	66.46	7.50	66.96	7.50	68.97	7.50	69.73	7.73	70.13	7.78	70.26	7.43	70.84	7.63	70.85	7.21	71.40	7.19	71.80	7.08	5.00
t9_5	68.13	5.50	68.43	6.74	69.13	7.48	69.52	8.09	69.59	7.29	70.04	7.25	70.59	7.08	70.62	6.71	70.70	6.69	70.73	6.67	5.00
t9_6	54.96	7.00	57.79	6.52	60.05	5.74	60.64	5.56	60.86	5.27	61.59	5.69	63.33	6.12	63.45	5.99	63.55	6.04	63.60	6.27	7.00
t9_7	68.76	7.50	69.15	6.76	69.49	6.67	71.01	6.75	71.14	6.70	71.49	6.90	71.94	6.98	72.54	6.93	72.73	6.94	73.51	6.77	6.50
t9_8	66.00	4.00	66.24	5.49	66.32	5.99	68.46	6.46	68.82	6.47	69.45	6.32	69.57	6.48	69.62	6.30	70.05	6.07	70.80	5.97	9.00
t9_9	58.59	4.00	60.40	5.22	61.07	5.79	61.85	5.60	61.91	5.68	62.12	5.73	62.43	5.49	62.45	5.86	63.27	5.82	63.52	5.65	7.00
t9_10	61.44	6.00	62.86	5.02	63.93	5.02	63.97	5.50	64.61	6.17	64.91	5.82	65.92	5.58	66.31	5.69	66.74	5.77	66.79	6.07	5.50
t9_11	55.90	6.50	56.61	5.26	56.72	5.84	58.03	6.12	59.17	6.29	59.62	6.08	60.01	6.21	61.06	6.24	61.10	6.16	61.80	5.87	7.00
t9_12	69.86	7.00	71.48	7.00	72.21	7.79	72.49	7.72	73.90	7.23	74.11	7.05	74.26	6.65	74.45	6.92	75.01	6.63	75.37	6.84	8.50
t9_13	63.95	4.00	66.27	4.00	66.70	4.00	67.78	4.23	68.08	5.13	68.19	5.42	68.80	5.63	68.94	6.03	68.97	6.08	69.00	6.02	6.00
t9_14	58.00	4.00	58.44	5.49	59.71	5.98	59.95	5.50	60.03	5.21	60.42	5.01	60.73	4.87	61.20	4.95	61.22	5.38	61.68	5.49	8.00
t9_15	72.03	7.00	72.60	6.02	72.80	6.67	73.07	6.62	73.32	6.70	74.40	5.98	74.41	6.18	74.71	6.04	74.74	6.04	74.82	6.17	6.00
t9_16	58.95	4.00	59.06	5.50	59.95	5.01	59.97	4.76	60.15	4.61	60.53	4.92	61.10	5.20	61.22	5.18	61.46	5.48	61.60	5.68	6.50
t9_17	66.32	4.00	68.30	4.00	68.54	5.61	68.83	5.82	69.28	6.24	69.59	6.28	69.84	6.51	70.00	6.81	70.24	7.09	70.31	6.80	7.00
t9_18	55.48	4.00	56.47	5.48	57.37	5.97	60.70	5.87	60.99	6.08	61.59	5.91	62.54	6.12	62.58	6.22	63.17	5.99	63.77	5.90	6.50
t9_19	62.03	4.00	63.80	4.48	64.62	5.45	65.43	5.58	65.46	5.95	65.70	6.35	66.18	6.44	66.21	6.03	66.57	6.13	66.67	6.03	7.00
t9_20	59.73	7.00	60.69	5.52	61.24	5.03	64.28	5.02	66.22	5.02	67.54	5.01	67.70	4.89	68.13	5.01	68.55	5.15	68.62	5.05	6.50
t10_1	56.81	4.00	61.20	5.40	62.74	4.97	63.30	4.98	64.12	4.98	64.94	5.59	66.13	5.90	66.46	5.96	66.65	5.77	67.09	5.83	5.50
t10_2	60.81	4.00	61.12	4.00	64.04	4.00	65.35	4.80	65.77	5.57	66.15	6.10	66.26	5.95	66.82	5.73	66.95	5.86	67.13	6.06	4.00
t10_3	64.17	7.50	66.01	5.33	66.29	6.19	66.82	6.51	66.90	6.79	67.05	6.83	67.85	7.26	67.91	7.46	68.29	7.36	68.42	7.33	8.00
t10_4	66.10	4.00	66.55	5.98	67.05	6.15	67.70	5.63	68.16	6.18	69.23	6.08	69.81	6.07	70.05	6.06	70.07	6.26	70.57	6.51	6.00
t10_5	70.36	6.50	70.42	7.75	70.78	8.16	70.86	7.87	71.83	7.14	71.84	6.80	72.16	6.42	72.25	6.13	72.26	6.17	72.80	5.92	7.00
t10_6	58.05	6.50	61.02	6.74	62.66	5.88	63.91	6.37	64.49	6.58	64.96	6.03	65.16	5.76	65.42	5.91	65.46	5.81	65.80	5.64	7.00
t10_7	50.86	4.00	52.60	5.96	54.45	6.29	54.59	6.46	55.99	6.00	57.52	5.84	57.99	5.93	58.19	6.00	58.22	6.05	58.37	6.33	7.00
t10_8	61.24	4.00	64.27	4.95	65.88	5.43	66.52	5.33	66.60	5.36	66.72	5.15	67.17	5.33	67.50	5.35	67.87	5.21	68.00	5.28	7.00
t10_9	62.84	6.00	62.98	7.50	63.00	7.16	63.41	7.12	63.67	7.20	63.68	6.67	64.57	6.92	64.90	6.87	66.16	6.73	66.39	6.80	8.50
t10_10	55.66	5.00	59.00	4.52	60.15	5.46	60.25	5.11	61.05	4.71	62.37	5.06	63.12	5.19	64.01	5.33	64.50	5.10	64.89	5.23	7.00
t10_11	60.52	7.00	61.74	7.25	61.98	7.33	62.08	7.12	62.69	7.00	63.82	6.92	64.02	6.53	64.08	6.58	64.21	6.79	64.86	6.81	8.00

t10_12	68.94	7.00	69.25	5.01	69.90	5.66	70.29	6.23	70.47	5.99	71.19	6.62	71.26	6.74	71.32	6.89	71.45	7.06	71.68	7.24	8.00
t10_13	58.09	7.00	58.21	7.25	58.43	6.84	59.18	7.12	59.96	6.71	60.05	6.27	62.11	6.63	63.09	6.90	63.35	6.91	63.76	6.65	8.00
t10_14	64.32	7.50	65.98	7.98	67.81	8.30	68.05	8.46	68.22	8.28	68.53	8.08	68.64	7.87	68.96	8.12	69.67	7.95	69.73	7.77	8.50
t10_15	57.76	4.00	61.55	5.42	64.74	6.03	64.90	6.37	65.04	5.75	65.30	5.86	65.34	5.75	65.51	5.66	65.71	5.75	65.76	5.87	7.00
t10_16	66.53	9.00	69.60	7.37	69.97	8.04	70.39	7.09	70.52	6.98	70.89	6.51	71.03	6.44	71.06	6.74	71.20	6.98	71.23	6.84	5.50
t10_17	65.26	4.00	65.31	4.00	66.02	4.00	66.66	4.85	67.22	4.88	67.41	5.37	68.07	5.19	68.27	5.64	68.27	5.62	68.41	5.80	6.00
t10_18	62.30	8.00	62.99	6.02	64.08	6.02	64.65	6.13	64.84	5.72	65.26	6.00	65.57	6.00	65.98	6.18	66.23	6.27	66.57	6.52	9.50
t10_19	65.74	6.50	66.92	6.99	67.57	6.83	69.10	6.75	69.56	6.61	69.59	6.21	69.66	6.05	70.10	6.21	70.74	6.19	71.00	6.22	6.50
t10_20	54.13	4.00	58.82	6.35	59.08	6.71	59.59	6.07	60.27	6.24	60.53	6.04	61.86	5.78	62.60	5.75	62.84	5.57	63.39	5.65	6.00

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all sources of materials for the thesis have been duly acknowledged.

_____

ABEL TEKLEMARIAM

This thesis has been submitted for examination with my approval as an advisor.

_____

SEBSIBE HAILEMARIAM (PhD)

Addis Ababa, Ethiopia

November, 2010