CHAPTER 1

INTRODUCTION

This chapter discusses rationale of the research, definition of the problem, objective and hypothesis, scope and delimitation, and the importance of the study

1.1 Rationale

As digital technology advances, the education sector also significantly adapts to technological developments. Evaluation is a crucial stage in the learning process to measure the achievement of learning objectives. Several educational institutions have adopted Computer Assisted Assessment (CAA) [1–3]. CAA refers to the use of computers and technology to facilitate and enhance the assessment and evaluation of students' knowledge, skills, and performance. CAA aims to simplify the assessment process, provide quick feedback, and make evaluation more efficient and effective. Evaluation is an essential stage in the learning evaluation [4, 5]. Evaluation can be done using various methods such as multiple-choice, essay, true-false, matching, etc. Each test has its advantages and disadvantages [6]. E-learning assessments usually use multiple-choice assessments because they are easy to assess. However, multiple-choice tests have the possibility of correct answers by guessing. Therefore, evaluation with multiple-choice tests can only partially measure students' abilities and is easy to cheat [4, 7].

Problems arise when there are parts requiring students to understand where the test must be conducted in essay form. According to the written test guidelines, an essay test is a test where the answer requires the students to recall and organize ideas or things they have learned by presenting or expressing ideas in written description form. The essay test format allows students to express their thoughts so that the answers will show their thinking ability. However, in practice, there are challenges in the essay test assessment process, such as requiring a relatively long time and being more difficult in the correction process, making it challenging to use for large-scale tests. The evaluator's subjective nature can influence essay assessment, so tests graded by different people can yield different results. Essay assessment also needs to improve on low validity and reliability issues, even when graded by the same person [4]. To address these issues, an automatic essay assessment program is needed to solve the essay assessment process, making the correction time faster and more objective. An automatic assessment system is used to assist in implementing tests and assessments, making them practical and efficient. Assessment using this automated system allows students to answer questions through the program and directly obtain grades [8–11].

Automatic essay test is limited for the Indonesian language, making it crucial to design

an Automated Short Answer Scoring (ASAS) system capable of handling essays in this language with high accuracy. Research and development in this field are also still limited, so only a few systems can automatically assess essays in Indonesian. This becomes an obstacle for educational institutions that want to implement this technology in learning and evaluation. Another area for improvement is the limited research on implementing automatic essay assessment in schools. Most research and development of automatic essay assessment systems are conducted at higher education levels or for the English language. Therefore, further research is needed to develop and implement an automatic essay assessment system that can be used in schools, especially in Indonesia, to improve the efficiency and effectiveness of the assessment process.

1.2 Statement of the Problem

A critical aspect of education is the evaluation stage, which measures the achievement of learning objectives. Despite the advantages of multiple-choice assessments in e-learning, they fail to fully capture the depth of students' understanding and critical thinking skills, often leading to issues such as guessing and cheating. On the other hand, essay assessments provide a more accurate measure of a student's ability to remember, organize, and express ideas. However, manual grading of essays is time-consuming, subjective, and challenging to scale, leading to problems with reliability and consistency. Along with the problem, ASAS in Indonesian is still very limited, so there is a need to develop ASAS for the Indonesian language. To address these challenges, this research aims to create an Automated Short Answer Scoring (ASAS) system utilizing sentence embedding techniques to automate the essay evaluation process. The system will compare student scores between ASAS and teachers' manual scores. This research will focus on the essays from basic lessons in the Computer and Network Engineering Program class XI at SMKN 1 Rao Selatan, Pasaman Regency, West Sumatera. The primary goal is to develop an ASAS system that correlates closely with teacher scoring.

The research questions of the research are: "How is the performance of ASAS for the Indonesian language with sentence embedding, in comparison with ASAS for the Indonesian language applying word embedding, and manual scoring by teachers ?"

1.3 Objective and Hypotheses

1.3.1 Objective

This study aims to study and develop automated short answer scoring, for Indonesian language, which applies sentence embedding. The comparison will focus on critical evaluation criteria correlation coefficient and Mean Absolute Error (MAE), about manual scoring by teachers. The goal is to determine how accurately these models can replicate human scoring

in educational settings. Additionally, the research involves designing and implementing an automatic short answer scoring system tailored explicitly for school assessments, thereby enhancing the efficiency of the evaluation process.

1.3.2 Hypotheses

Premis 1: Some previous studies have shown that sentence embedding correlates better with teacher manual assessment than word embedding [12]

Premis 2: The theory states that sentence embedding can produce better correlation coefficients than word embedding because it is able to capture contextual and semantic nuances more effectively [13, 14].

Hypotheses : The sentence embedding model will exhibit a higher correlation with teacher manual assessments than the word embedding model, due to its superior ability to capture contextual and semantic nuances, resulting in a more accurate reflection of human grading patterns.

1.4 Scope and Delimitation

This research focuses on the automated scoring of short answer questions within the context of the "Fundamentals of Computer Network Engineering Program Phase F Vocational School." The subject matter addressed pertains to "wired and wireless networks," and the learning objective is to ensure students understand these types of networks. The study will be constrained to questions, reference answers, and student responses written in Indonesian. Any content involving images, mathematical symbols, or characters outside of the Indonesian language will not be considered. Additionally, responses must be 50-200 words [4].

1.5 Significance of the Study

The contributions of this research are: First, a comprehensive comparative analysis of the effectiveness of sentence embedding models versus word embedding models in automated scoring for Indonesian test. The study provides insights into each approach's relative strengths and weaknesses by evaluating how well these models align with human judgment in grading short answer questions. This analysis will contribute valuable knowledge to the field of automated scoring systems.

Second, the tool ASAS for the Indonesian language based on the findings from the comparative analysis. The system will leverage the model with the highest correlation to manual scoring by teachers, explicitly targeting short answer responses in the "Fundamentals of Computer Network Engineering Program Phase F Vocational School," focusing on wired and wireless networks. The ASAS for the Indonesian language system is designed to enhance

the grading process by streamlining evaluations, reducing teachers' workload, and ensuring consistent and objective assessments of student performance.