

II. LITERATURE REVIEW

This chapter presents literature review related to the research problem. Therefore, a number of relevant topics are reviewed here, they are; (1) the previous research, (2) the concept of writing, (3) the concept of argumentative essay writing, (4) the concept of writing assessment, (5) writing and technology involvement, (6) theory on raters, (7) theoretical assumption, and (8) hypothesis.

2.1 Previous Research

Research comparing handwriting with word processing for the presentation of assessment answers can broadly be identified as having two main phases. In the initial phase of the early 1990s, a small number of groups investigated the effects of two modes of answering an examination questions – handwritten or word processed – upon students' scores. One notable work in this phase was by Powers, Fowles, Farnum and Ramsey (1994). Powers, et al converted a sample of original handwritten essay answers into word processed versions as well as transcribed a sample of original word processed essay answers into handwritten versions. Powers, et al found the surprising finding that: Raters awarded higher scores to responses presented in handwritten form as compared to the exact same responses presented as computer-printed text, although he anticipated that responses presented in hand-written form would receive lower scores.

To explain this seemingly contradictory finding, Powers, et al (1994) offered several hypotheses, some of which drew upon their work as well as the work of Arnold, Legas, Obler, Zpacheco, Russell, and Umbdenstock (as summarized by Powers et al., 1994). These hypotheses included:

- A. Readers may have expected fully edited and polished final products when presented as computer-printed text and thus had higher expectations for these essays;
- B. Handwritten text caused the reader to feel closer to the writer which “allowed for a closer identification of the writer’s individual voice as a strong and important aspect of the essay” (as quoted in Powers et al.);
- C. Readers may have given handwritten responses the benefit of the doubt when they encountered sloppy or hard-to-read text;
- D. Hand-written responses appeared longer and thus appeared to have been the result of greater effort.

To examine the final hypothesis, Powers et al. (1994) conducted a small follow-up study during which computer-printed responses were double-spaced to make them appear longer. During this follow-up study, training procedures were also modified such that readers were informed of the presentation effect and were instructed to apply the same criteria to handwritten and computer-printed responses. The combination of supplemental training and double-spacing of computer-printed responses reduced the presentation effect, but did not eliminate it.

This replicated findings of Arnold, Legas, Obler, Pacheco, Russell and Umbdenstock (1990) that student papers converted to word-processed versions received lower scores than did the original handwritten versions. Arnold et al suggested that the reviewers may have had higher expectations of the word processed work, less empathy with the authors of word processed work or may have been less likely to give the benefit of any doubt. Powers et al suggested other possible factors including lack of evidence in the word processed versions of evidence to revise work, greater visibility of typographical errors, or apparently shorter answers.

A second phase of research in this area resulted in a series of publications from 2004 onwards. Principal authors have included Russell and Tao (2004a, 2004b) who found some evidence in support of the Powers et al work of 1994, suggesting that the lower scores for computer printed work may have resulted from factors identified by increased visibility of any errors, higher expectations of readers and less empathy with students (“felt a stronger connection to the writer because of the handwriting”). Russell and Tao (2004) explored several types of errors in students writing when essays are presented as computer text. According to them the different formats of presentation would result in different scores. The essay presented in original handwritten received higher score than those presented as computer text. As Powers identified that the difference in length between computer-printed and handwritten essays also has an influence on the assessment of essays, Russell and Tao replicated it by manipulating the length of the essay by double spacing the computer-printed essay so it better matched the amount of

space taken up by the handwritten essay. The computer-printed scores still received lower scores; however, the effect was reduced when the computer essay was double spaced rather than single spaced. Therefore, this study has also added evidence that the difference in scores is due to the visibility of errors. Through interview with raters, this study also suggests that the higher standard and expectations raters have for text presented as computer print and the ability of some raters to identify with students and see their effort when they handwrite essays may impact the scores they award.

More recent work in the United Kingdom by researchers at the University of Edinburgh (Mogey, Paterson, Burk and Purcell, 2010) compared transcribed scripts of first year students in a mock examination: handwritten scripts were transcribed into typed format and typed scripts were transcribed into handwritten format. Mogey et al found “weak evidence” that handwritten scripts generally scored slightly higher than typed scripts.

Since this study is partly replicated the previous research done by Russell and Tao (2004), therefore, the methodology also follow what have been done by them. Unlike Powers et al (1994) who gathered the sample essays from students of Business and Law, this study focused on sample essays from students of Language and Arts Department.

2.2 The Concept of Writing

According to Walters (1999:90), writing is a complex process since it is made of a large number of skills which involved various language elements such as:

punctuation, spelling, grammar, diction, etc. In other words, a writer should achieve the ability of crafting a good piece of writing by combining those elements in free-mistake as well as possible.

A simple evidence how knowing proper punctuation matters in composing a writing product can be shown in this example.

Let's eat Grandma!

Such simple sentence may lead to a big disaster semantically when the writer unaware of the proper punctuation needed to make the sentence meaning correct; a comma is definitely needed in that sentence; thus:

Let's eat, Grandma!

When we compare with the first sentence, the meaning is totally changed. In the first sentence, the meaning is to eat grandma; while in the second one, the meaning is to invite grandma to eat. If it is not merely sort of joke, punctuation indeed can save people live!

When writers should deal with spelling matters, they are basically dealing with the prescribed meaning of the sentence. For student writers or beginners, relying solely on *spell checker* is not a wise way since spell checker will permit any other related words even though the words appear on screen are not the intended words; e.g. *She is a diligence student*. In that case computer will just allow 'diligence' in that sentence without the red mark. However, the writer intended to write 'diligent student' since it is 'adjective' which explains 'student', not 'diligence student' which

will make 'double noun'. If it is the case, then the sentence is also grammatically incorrect; therefore, a thorough understanding of correct spelling is important to create a correct sentence.

Diction also plays a major role when composing a good piece of writing. A good writer know what proper diction to use in different context and setting. This ability can be achieved by reading a lot of writing products and understand the different usage from each diction. The improper use of diction may lead the writer to be underestimated by his readers simply because he writes 'thou' instead of 'you' in modern setting of writing. Readers can even judge the writer as less capable in carrying out the 'proper diction' in the right setting; therefore writers should be careful when choosing the correct dictions in their writing.

Similar with the concept proposed by Walters (1999:90), Tarigan (1981:1) put the skill of writing as the last skill learners should master, following the other three skills in language composition; listening, speaking, and reading. It suggests that writing is the most difficult skill to master a language since it has three prerequisites of skills mentioned above. It is not surprising that students often feel reluctant when they are assigned a writing assignment.

Raimes (1983:93) states that writing also involves a systematic way of thinking. It means that it takes much effort to gather information and write them down in a systematic way to get a good piece of writing. Since it involves a high capacity of thinking, the students are expected to train this skill periodically to maintain their ability in high-stake thinking.

There are two general purposes of writing according to Ellis (1990:93); to spread the message to others and to keep it for personal use. As in academic setting, writing, particularly writing an essay, is aimed to communicate the writers' thought to the readers. Therefore students are expected to be able to produce a good piece of writing and deliver the message effectively.

2.3 The Concept of Argumentative Essay Writing

Essay as a form of written product is very familiar in academic setting as an attempt to high critical thinking of a composition. In college level, particularly in English major, composition is part of curricula that has met the standardized material and assessment. Students are expected to comprehend the skill in composition from basic writing level to advanced level by the end of their college years. Even before students enter the college or university, they are familiarized by various writing forms and texts when they are in middle and high school. Therefore when they enter the college or university, they are fully ready to digest every small aspect in writing.

As this research employed essay to gather the data, a thorough explanation is needed to complete the understanding about essay, particularly argumentative essay.

According to Merriam Webster online dictionary, argumentative essay is a genre of writing that requires the student to investigate a topic; collect, generate, and evaluate evidence; and establish a position on the topic in a concise manner. Argumentative essay assignments generally call for extensive research of

literature or previously published material. Argumentative assignments may also require empirical research where the student collects data through interviews, surveys, observations, or experiments. Detailed research allows the student to learn about the topic and to understand different points of view regarding the topic so that she/he may choose a position and support it with the evidence collected during research. Regardless of the amount or type of research involved, argumentative essays must establish a clear thesis and follow sound reasoning.

The structure of the argumentative essay is held together by the following.

- **A clear, concise, and defined thesis statement that occurs in the first paragraph of the essay.**

In the first paragraph of an argument essay, students should set the context by reviewing the topic in a general way. Next the author should explain why the topic is important (**exigence**) or why readers should care about the issue. Lastly, students should present the thesis statement. It is essential that this thesis statement be appropriately narrowed to follow the guidelines set forth in the assignment. If the student does not master this portion of the essay, it will be quite difficult to compose an effective or persuasive essay.

- **Clear and logical transitions between the introduction, body, and conclusion.**

Transitions are the mortar that holds the foundation of the essay together. Without logical progression of thought, the reader is unable to follow the essay's argument, and the structure will collapse. Transitions should wrap up the idea

from the previous section and introduce the idea that is to follow in the next section.

- **Body paragraphs that include evidential support.**

Each paragraph should be limited to the discussion of one general idea. This will allow for clarity and direction throughout the essay. In addition, such conciseness creates an ease of readability for one's audience. It is important to note that each paragraph in the body of the essay must have some logical connection to the thesis statement in the opening paragraph. Some paragraphs will directly support the thesis statement with evidence collected during research. It is also important to explain how and why the evidence supports the thesis (**warrant**).

However, argumentative essays should also consider and explain differing points of view regarding the topic. Depending on the length of the assignment, students should dedicate one or two paragraphs of an argumentative essay to discussing conflicting opinions on the topic. Rather than explaining how these differing opinions are wrong outright, students should note how opinions that do not align with their thesis might not be well informed or how they might be out of date.

- **Evidential support (whether factual, logical, statistical, or anecdotal).**

The argumentative essay requires well-researched, accurate, detailed, and current information to support the thesis statement and consider other points of view. Some factual, logical, statistical, or anecdotal evidence should support the thesis. However, students must consider multiple points of view when collecting evidence. As noted in the paragraph above, a successful and well-rounded

argumentative essay will also discuss opinions not aligning with the thesis. It is unethical to exclude evidence that may not support the thesis. It is not the student's job to point out how other positions are wrong outright, but rather to explain how other positions may not be well informed or up to date on the topic.

- **A conclusion that does not simply restate the thesis, but readdresses it in light of the evidence provided.**

It is at this point of the essay that students may begin to struggle. This is the portion of the essay that will leave the most immediate impression on the mind of the reader. Therefore, it must be effective and logical. Do not introduce any new information into the conclusion; rather, synthesize the information presented in the body of the essay. Restate why the topic is important, review the main points, and review your thesis. You may also want to include a short discussion of more research that should be completed in light of your work.

A common method for writing an argumentative essay is the five-paragraph approach. This is, however, by no means the only formula for writing such essays. If it sounds straightforward, that is because it is; in fact, the method consists of (a) an introductory paragraph (b) three evidentiary body paragraphs that may include discussion of opposing views and (c) a conclusion.

2.4 The Concept of Writing Assessment

According to the Merriam-Webster online dictionary the word *assessment* comes from the root word *assess* which is defined as:

1. to determine the rate or amount of (as a tax)
2. to impose (as a tax) according to an established rate b: to subject to a tax, charge, or levy
3. to make an official valuation of (property) for the purposes of taxation
4. to determine the importance, size, or value of (assess a problem)
5. to charge (a player or team) with a foul or penalty

The term *assessment* is generally used to refer to all activities teachers use to help students learn and to gauge student progress. Though the notion of assessment is generally more complicated than the following categories suggest, assessment is often divided for the sake of convenience using the following distinctions:

1. initial, formative, and summative
2. objective and subjective
3. referencing
4. informal and formal.

As in writing, the assessment is categorized as subjective assessment. It is a form of questioning which may have more than one correct answer (or more than one way of expressing the correct answer). There are various types of subjective questions, include extended-response questions and essays.

Writing assessment refers to an area of study that contains theories and practices that guide the evaluation of a writer's performance or potential through a writing task. Writing assessment can be considered a combination of scholarship from Writing Theory and Measurement Theory within educational assessment. Writing

assessment can also refer to the technologies and practices used to evaluate student writing and learning.

In “Looking Back as We Look Forward: Historicizing Writing Assessment as a Rhetorical Act,” Kathleen Blake Yancey offers a history of writing assessment by tracing three major shifts in methods used in assessing writing. She describes the three major shifts through the metaphor of overlapping waves: “with one wave feeding into another but without completely displacing waves that came before”. In other words, the theories and practices from each wave are still present in some current contexts, but each wave marks the prominent theories and practices of the time.

The first wave of writing assessment (1950-1970) sought objective tests with indirect measures of assessment. The second wave (1970-1986) focused on holistically scored tests where the students’ actual writing began to be assessed. And the third wave (since 1986) shifted toward assessing a collection of student work (i.e. portfolio assessment) and programmatic assessment.

Bob Broad in *What We Really Value* points to the publication of *Factors in Judgments of Writing Ability* in 1961 by Diederich, French, and Carlton as the birth of modern writing assessment. Diederich, French, and Carlton based much of their book on research conducted through the Educational Testing Service (ETS) for the previous decade. This book is an attempt to standardize the assessment of writing and, according to Broad, created a base of research in writing assessment.

In the first wave of writing assessment, the emphasis is on reliability: reliability confronts questions over the consistency of a test. In this wave, the central concern was to assess writing with the best predictability with the least amount of cost and work.

The shift toward the second wave marked a move toward considering principles of validity. Validity confronts questions over a test's appropriateness and effectiveness for the given purpose. Methods in this wave were more concerned with a test's construct validity: whether the material prompted from a test is an appropriate measure of what the test purports to measure. Teachers began to see incongruence between the material being prompted to measure writing and the material teachers were asking students to write. Holistic scoring, championed by Edward M. White, emerged in this wave. It is one method of assessment where students' writing is prompted to measure their writing ability.

The third wave of writing assessment emerges with continued interest in the validity of assessment methods. This wave began to consider an expanded definition of validity that includes how portfolio assessment contributes to learning and teaching. In this wave, portfolio assessment emerges to emphasize theories and practices in Composition and Writing Studies such as revision, drafting, and process.

Methods of writing assessment vary depending on the context and type of assessment. The following is an incomplete list of writing assessments frequently administered:

1. Portfolio

Portfolio assessment is typically used to assess what students have learned at the end of a course or over a period of several years. Course portfolios consist of multiple samples of student writing and a reflective letter or essay in which students describe their writing and work for the course. “Showcase portfolios” contain final drafts of student writing, and “process portfolios” contain multiple drafts of each piece of writing. Both print and electronic portfolios can be either showcase or process portfolios, though electronic portfolios typically contain hyperlinks from the reflective essay or letter to samples of student work and, sometimes, outside sources.

2. Timed-Essay

Timed essay tests were developed as an alternative to multiple choice, indirect writing assessments. Timed essay tests are often used to place students into writing courses appropriate for their skill level. These tests are usually proctored, meaning that testing takes place in a specific location in which students are given a prompt to write in response to within a set time limit. The SAT and GRE both contain timed essay portions.

3. Rubric

A rubric is a tool used in writing assessment that can be used in several writing contexts. A rubric consists of a set of criteria or descriptions that guides a rater to score or grade a writer. The origins of rubrics can be traced to early attempts in education to standardize and scale writing in the early 20th century. Ernest C Noyes argues in November 1912 for a shift toward assessment practices that were

more science-based. One of the original scales used in education was developed by Milo B. Hillegas in *A Scale for the Measurement of Quality in English Composition by Young People*. This scale is commonly referred to as the Hillegas Scale. The Hillegas Scale and other scales used in education were used by administrators to compare the progress of schools.

In 1961, Diederich, French, and Carlton from the Educational Testing Service (ETS) publish *Factors in Judgments for Writing Ability* a rubric compiled from a series of raters whose comments were categorized and condensed into a five-factor rubric:

Ideas: relevance, clarity, quantity, development, persuasiveness

Form: Organization and analysis

Flavor: style, interest, sincerity

Mechanics: specific errors in punctuation, grammar, etc.

Wording: choice and arrangement of words

As rubrics began to be used in the classroom, teachers began to advocate for criteria to be negotiated with students to have students stake a claim in the how they would be assessed. Scholars such as Chris Gallagher and Eric Turley, Bob Broad, and Asao Inoue (among many) have advocated that effective use of rubrics comes from local, contextual, and negotiated criteria.

A scoring rubric is an attempt to communicate expectations of quality around a task. In many cases, scoring rubrics are used to delineate consistent criteria for grading. Because the criteria are public, a scoring rubric allows teachers and

students alike to evaluate criteria, which can be complex and subjective. A scoring rubric can also provide a basis for self-evaluation, reflection, and peer review. It is aimed at accurate and fair assessment, fostering understanding, and indicating a way to proceed with subsequent learning/teaching. Another advantage of a scoring rubric is that it clearly shows what criteria must be met for a student to demonstrate quality on a product, process, or performance task.

Douglas H. Brown (2004:335) states that in teaching writing, the compositions is supposed to:

- a) meet certain standards of prescribed English rhetorical style;
- b) reflect accurate grammar;
- c) be organized in conformity with what the audience would consider to be conventional.

It means that a good deal of attention was placed on “model” composition that the students would emulate and how well a student’s final product measured up against a list of criteria including content, organization, vocabulary use, grammatical use and mechanical consideration such as spelling and pronunciation.

According to Maley (1998) as quoted in Rudy (2011:15), a good written composition also should consist of several degrees in order to get a good quality of writing; a high degree of organization in the development of ideas and information, a high degree of accuracy so that there is no ambiguity of meaning, the use of grammatical features, and a careful choice of diction, grammatical pattern, and sentence structure.

In this study, the researcher employed a content based rubric which was taken from the documentary of English Language Center (ELI) of Missouri State University. The rubric itself was aimed for grading the English Foreign Learners (EFL) in the university and had been applicable to the wider scope of EFL assessment in all around the states. Thus, the researcher was adopting the rubric to grade the essay performances in this research (see **Appendix 4**).

2.5 Writing and Technology Involvement

In this age of technology, the involvement of computer has been widely used in helping students to write down their ideas. Instructors and teachers have introduced computer as an assistive tool to learn writing. The students at junior level have to master at least the basic skill of typing and editing in Microsoft to ease them in completing the assignments.

In the level of college or university, the students have been very familiar with the instruction through computer assignment. Essay assignment or writing response should be submitted as a printed piece rather than as a handwritten piece. Unfortunately, although composing essays on computers is becoming more common, studying its effect on writing assessment has received little attention as reported by Chase (1979).

It is therefore, this study focused on comparing the two different formats of writing products, they are essays presented as handwritten format and essays presented as computer text format. The researcher employed three pairs of raters

in order to assess the students' essays to find out whether there is difference on raters' scores for different format of essay.

2.6 Theory on Raters

Introduced by Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963), Generalizability theory which is also known as G Theory, is a statistical framework for conceptualizing, investigating, and designing reliable observations. It is used to determine the reliability (i.e., reproducibility) of measurements under specific conditions. It is particularly useful for assessing the reliability of performance assessments. In G theory, sources of variation are referred to as *facets*. Facets are similar to the "factors" used in analysis of variance, and may include persons, raters, items/forms, time, and settings among other possibilities. These facets are potential sources of error and the purpose of generalizability theory is to quantify the amount of error caused by each facet and interaction of facets. The usefulness of data gained from a G study is crucially dependent on the design of the study. Therefore, the researcher must carefully consider the ways in which he/she hopes to generalize any specific results. Is it important to generalize from one setting to a larger number of settings? From one rater to a larger number of raters? From one set of items to a larger set of items? The answers to these questions will vary from one researcher to the next, and will drive the design of a G study in different ways.

In addition to deciding which facets the researcher generally wishes to examine, it is necessary to determine which facet will serve as the object of measurement (e.g. the systematic source of variance) for the purpose of analysis. The remaining

facets of interest are then considered to be sources of measurement error. In most cases, the object of measurement will be the person to whom a number/score is assigned. In other cases it may be a group or performers such as a team or classroom. Ideally, nearly all of the measured variance will be attributed to the object of measurement (e.g. individual differences), with only a negligible amount of variance attributed to the remaining facets (e.g., rater, time, setting).

The results from a G study can also be used to inform a decision, or D, study. In a D study, we can ask the hypothetical question of “what would happen if different aspects of this study were altered?” For example, a soft drink company might be interested in assessing the quality of a new product through use of a consumer rating scale. By employing a D study, it would be possible to estimate how the consistency of quality ratings would change if consumers were asked 10 questions instead of 2, or if 1,000 consumers rated the soft drink instead of 100. By employing simulated D studies, it is therefore possible to examine how the generalizability coefficients would change under different circumstances, and consequently determine the ideal conditions under which our measurements would be the most reliable.

2.7 Theoretical Assumption

Since the researcher set up two alternatives of hypotheses in this study, there should be two strong assumptions in conducting the study:

The first assumption is to prove the first hypothesis that raters gave higher score on essays presented as handwritten text as to essay presented as computer-text. There are several possible causes upon why raters set the different rating scores on

the three modes of presentations. 1) Mechanical errors, 2) Higher expectation on essay as computer printed text, and 3) Stronger connection when reading essay as handwritten text rather than as computer printed text.

The second assumption is to prove the second hypothesis that the length of essay can eliminate the presentation effect. In another word, the double space essay would likely to receive higher score than the single space essay. The assumption comes up from the idea that handwritten essays tend to appear longer than computer text essays.

2.8 Hypothesis

H_0 : There is no difference on raters' scores between essay presented as handwritten form and essay presented as computer printed text

H_1 : There is difference on raters' scores between essay presented as handwritten form and essay presented as computer printed text

H_0 : The length of essay does not eliminate the presentation effect;

H_1 : The length of essay eliminates the presentation effect.