

A Comparison of the Effectiveness between Multiple-Choice and Essay

Writing Exams in Writing Competency Assessment

การเปรียบเทียบประสิทธิภาพของข้อสอบแบบปรนัยและข้อสอบการเขียนเรียงความ ในการประเมินความสามารถในการเขียน

Naowarat Patipatpakdee
นารัตน์ ปฏิพัทธ์กัตต์

Sawitree Muangyai Reilly Phatcharasorn Noipann* Jantira Koedkham
สาขาวิชานิรัฐ ไรส์ลีย์ พัชราสอร์ โนอี้พันธ์ จันทิรา กีดคำ
Faculty of Arts, Silpakorn University, Thailand

Corresponding author*

e-mail: noipann_p@silpakorn.edu

Received 02-11-2023

Revised 27-11-2023

Accepted 12-12-2023

Abstract

Objectives: This research aims to study the effectiveness of multiple-choice exams when used to assess writing skills compared to essay writing exams.

Methods: The subjects of the study were 79 third-year students who were studying English as their minor subject at the Faculty of Arts, Silpakorn University in 2021 and taking the writing course 412221 Reading and Writing. Two types of exams were used in the study: a multiple-choice exam and an essay writing exam. A cause & effect essay was used in both multiple choice and essay writing exams. The 30-item multiple choice exam covered essay elements and organization, grammar points, and vocabulary. As for the essay writing exam, students were instructed to write one essay on one of the four given topics.

Results: Results showed that the students performed better in the multiple-choice exam, scoring an average of 21.03 points while the average score for the essay writing exam was 19.34 points. Looking at each area of assessment individually, students performed marginally better in the grammar and vocabulary parts of the multiple-choice exam, while they achieved slightly higher scores in the essay elements of the essay writing exam. However, apart from the vocabulary part, differences in the scores were not statistically significant. In addition, the difficulty index mean comparison showed that both types of exams were of comparable difficulty and the discrimination index of the multiple-choice exam and essay writing exam were comparable with no statistical significance—0.27 and 0.28 respectively. One contributing factor to this may be the use of the essay writing exam analytic rubric, as mistakes other than what were included in the analytic rubrics were not recorded. Furthermore, the analytic rubrics had a fixed maximum number of points that could be deducted.

Application of this study: The results suggest that a multiple-choice exam, if constructed well and covering all the topics that are taught in class, can be used as a tool to evaluate writing competency and can have comparable effectiveness to essay writing exams.

Keywords: an essay writing exam, a multiple-choice exam, writing competency

บทคัดย่อ

วัตถุประสงค์: งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของข้อสอบปรนัยเมื่อถูกนำไปใช้ประเมินทักษะด้านการเขียนโดยเปรียบเทียบกับข้อสอบการเขียนเรียงความ

วิธีการศึกษา: กลุ่มตัวอย่างของงานวิจัยคือ นักศึกษาชั้นปีที่ 3 วิชาไทยภาษาอังกฤษ ที่ได้ลงทะเบียนเรียนรายวิชา 412221 การอ่านและการเขียนจำนวน 79 คน ซึ่งทั้งหมดเป็นนักศึกษาของคณะอักษรศาสตร์ มหาวิทยาลัยศิลปากร ประจำปีการศึกษา 2564 งานวิจัยนี้ได้ศึกษาการเขียนเรียงความแบบเหตุและผลโดยใช้ข้อสอบ 2 ประเภท คือข้อสอบปรนัยจำนวน 30 ข้อ มีเนื้อหาครอบคลุมด้านองค์ประกอบและโครงสร้างของเรียงความ ไวยากรณ์และคำศัพท์ และข้อสอบการเขียนเรียงความ โดยนักศึกษาจะได้รับคะแนน 1 เรื่อง จากหัวข้อที่กำหนดให้ 4 หัวข้อ

ผลการศึกษา: ผลจากการวิจัยแสดงให้เห็นว่า นักศึกษาสามารถทำคะแนนข้อสอบปรนัยได้สูงกว่า โดยมีคะแนนเฉลี่ยที่ 21.03 คะแนน ในขณะที่ข้อสอบการเขียนเรียงความ มีคะแนนเฉลี่ยที่ 19.34 คะแนน หากวิเคราะห์ผลการประเมินแยกเป็นแต่ละส่วนจะเห็นได้ว่า นักศึกษาทำคะแนนในส่วนของไวยากรณ์และคำศัพท์ของข้อสอบปรนัยได้ดีกว่าส่วนอื่นเล็กน้อย ในขณะที่ในข้อสอบการเขียนเรียงความ นักศึกษาได้คะแนนค่อนข้างสูงในส่วนขององค์ประกอบของการเขียนเรียงความ อย่างไรก็ตาม จะเห็นได้ว่าจากคะแนนคำศัพท์ ความแตกต่างของคะแนน

ในส่วนอื่นไม่มีนัยสำคัญทางสถิติ นอกจากนี้ จากการเปรียบเทียบดัชนีความยาก ผลการวิจัยแสดงให้เห็นว่าข้อสอบทั้งสองประเภทมีความยากเทียบเคียงกันได้ และจากการเปรียบเทียบดัชนีอ่านจำแนก ผลการวิจัยของข้อสอบปรนัยอยู่ที่ระดับ 0.27 และข้อสอบการเขียนเรียงความอยู่ที่ระดับ 0.28 ซึ่งแสดงให้เห็นว่าข้อสอบทั้งสองประเภทสามารถเทียบเคียงกันได้โดยไม่มีนัยสำคัญทางสถิติ ทั้งนี้ ปัจจัยหนึ่งที่ເອີ້ນດັບผลการวิจัยดังกล่าว อาจมาจากการใช้เกณฑ์การให้คะแนนแบบแยกส่วน ซึ่งข้อผิดพลาดอื่น ๆ ที่ไม่ได้ปรากฏในเกณฑ์จะไม่ได้รับการประเมิน และเกณฑ์การให้คะแนนแบบแยกส่วนจะมีการตั้งคะแนนสูงสุดที่สามารถหักลบได้

การประยุกต์ใช้: ผลจากการวิจัยแสดงให้เห็นว่า การวัดผลโดยใช้ข้อสอบปรนัย หากมีการออกข้อสอบที่ดี และครอบคลุมทั้งหมดที่นักศึกษาได้เรียนในชั้นเรียนจะสามารถใช้เป็นเครื่องมือในการประเมินความสามารถด้านการเขียนได้เป็นอย่างดีรวมถึงมีประสิทธิภาพที่เทียบเคียงกันได้กับข้อสอบการเขียนเรียงความ

คำสำคัญ: ข้อสอบการเขียนเรียงความ ข้อสอบแบบปรนัย ความสามารถในการเขียน

Introduction

English has been taught as a foreign language in Thailand for over a century. Since then, it has played an increasingly important role in the education of Thailand. In Thai universities, certain courses of writing are compulsory for English major and minor students because writing is considered a vital skill that the students are obliged to possess. The reasons why the writing skill is so important to students is to enhance communication skills, heighten levels of creativity and imagination, expand knowledge base, contribute to academic success, and increase their level of confidence (cited in www.easyreadernews.com). Writing skills not only lead to academic success, but are also considered an important tool for exchanging ideas in daily lives (Zeng, 2018). Furthermore, many job opportunities are open to those who excel in English. S. Lee & Schmidgall (2020) further pointed out that writing effectively is a skill needed by any workplace for all professions and business sectors.

Statements and significance of the problems

At the Faculty of Arts, Silpakorn University, in order to meet requirements, all third-year students who minor in English have to take the 412221 Reading and Writing course. Grading written essays, especially when markers have to score a lot of essays, is subject to subjectivity and bias (Schaefer, 2008; Slomp, 2012). Repetitive marking, undoubtedly, undermines the evaluation due to low reliability and validity. To be more specific, marking essays can be problematic for three main reasons. First, it takes a long time to mark each piece of writing, and with the number of students at hand, the accumulated time for marking may be high. This accumulated marking time directly influences another important reason. Erturk et al. (2022) found that boredom seriously affected scores. Their study showed that the higher the boredom was, the lower the scores were. Most importantly, subjectivity might interfere with marking, causing unreliable grading. Generally, in order to achieve reliability in marking, essays need to be marked twice or by at least two markers; however, this is difficult in real examinations (Tisi, 2013). To solve this problem, multiple choice exams could be one of the best alternatives. Multiple choice exams are very different from essay exams in terms of formats and scoring schemes. Walstad & Becker (1994) mentioned that a multiple choice format allows for a wider sampling of the content. It also offers greater efficiency and reliability in scoring than an essay. Whereas in an essay test, students can create their own responses that have the potential to show originality and a greater depth of understanding of the topic. The thought processes of the student can also be assessed through their work. It is unquestionable that the two methods of assessing students' abilities in learning English writing contain both advantages and disadvantages. However, which method is more suitable for assessing students' writing ability is still debatable. Most markers or teachers definitely desire the most effective and reliable method when it comes to designing writing exams.

Objectives of the research

Multiple choice exams have long been used for assessing students' performance. Their effectiveness has been confirmed by a large number of studies. Although essay writing is considered the most effective way to test students on their writing skills, there are times when problems, such as markers' subjectivity and inconsistency, arise. Therefore, this research aimed to investigate whether multiple choice exams are as effective as essay writing exams and can be used as part of the evaluation method.

Literature Review

Of all the language skills, writing is the most difficult challenge for both teachers and students because it is a skill that is based on the cognitive domain. It involves learning, understanding, applying, and synthesizing new knowledge. Writing also encompasses creative inspiration, problem-solving, reflection, and revision that results in a completed manuscript (Sim, 2010).

Graham (2019) stated that writing is a necessary skill for those who desire to be successful in their education, career, and personal lives, and to write effectively, students must learn how to write and practice writing adequately as writing is not a skill that develops naturally. Therefore, teaching writing skills systematically is crucial for students.

Writing assessment and evaluation

Assessment and evaluation are critical processes that examine what students have learnt during the course. Patton (1987) stated that the purpose of evaluation is to improve the effectiveness of the program, or the course. Most importantly, assessment and evaluation can judge students' competency if the methods used are properly designed and chosen. Examinations or tests are one of the most widely used methods. There are different types of exams used as part of an evaluation and assessment scheme. Brown (2005) stated that the main objectives of testing were to enable teachers and administrators to make decisions and plan for curriculum designs. For these purposes, proficiency and placement tests are used. In general, multiple choice exams are used for proficiency and placement tests, where there are a lot of test takers.

In addition to proficiency and placement tests, achievement tests are used to evaluate knowledge retention and to grade students' performance. When English is taught as a foreign or second language, four skills, namely listening, speaking, reading, and writing are emphasized. Instructors of the courses are responsible for constructing achievement tests by designing exams to suit the course objectives.

Essay writing exams

In a writing class, essay writing exams are considered an effective way to make a judgment about students' knowledge and to challenge students to express what they have learnt in their own words (University of North Carolina at Chapel Hill). M Cecil Smith concluded in *The Benefits of Writing*:

Writing is a significant literacy activity in modern life that enables individuals to accomplish a variety of personal, intellectual, occupational, and recreational goals. It has been demonstrated, across a variety of investigations, that writing activities yield a number of intellectual, physiological, and emotional benefits to individuals.

Steele (1997) confirmed that essay writing is advantageous, especially when higher-level cognitive skills are tested. Furthermore, essay writing is critical when the objectives of courses are to encourage students to apply knowledge, analyze, synthesize, and evaluate information or situations. Likewise, essay writing exams enable instructors to assess

students' various skills, particularly reasoning, critical thinking, synthesizing data, and expressing opinions (Bean, 1996 cited in Boyle, 2019). Cameron (2021) mentioned four benefits of essay writing which are improving communication skills, creating and developing the quality of contents, enhancing writing skills, and thinking creatively.

Essay writing will be a more effective evaluation tool if marking is properly done. Mousavi (2011) mentioned that the measurement of writing ability is influenced by four important factors, which are the student, the scoring method, the test administration, and the test itself. While all other three factors are equally significant, the most concerning is the scoring method – the method selected by the rater to pass judgment about the writing ability. The decisions about writing competence that are derived from one scoring method do not always, and do not necessarily, comply with decisions from another scoring method. These scoring systems are very important because they are used to classify test takers (Ghalib & Abdulghani, 2015). There have been many attempts in order to introduce the most effective way of evaluating students' writing ability.

Methods of essay markings

There are two types of essay marking methods, which have been used for years: holistic and analytic scoring methods.

1. Holistic scoring methods

Nordquist (2020) defined holistic grading as "a method of evaluating a composition based on its overall quality." Holistic scoring is very useful to examiners because each writing assignment can be marked quickly (Davies et al, 1999 cited in Nakamura, 2004). Gonzales (2014) also agreed that teachers take less time in marking when using holistic scoring because it is easier to create holistic rubrics. Teachers can mark by looking over a writing assignment and give a score holistically. In other words, holistic scoring considers the entire written response and assigns an overall score to the performance (Hyland, 2002). Holistic scoring rubrics are very practical. They are short, do not include detailed criteria of evaluation, and make the evaluation of an essay by assigning one score to it after only one reading. Therefore, this method is typically used for evaluating written performance in large-scale assessment contexts. Examples of the use of holistic scoring in assessing written performance appear in the computer-based Test of English as a Foreign Language (TOEFL), Graduate Record Examination (GRE), and Graduate Management Admission Test (GMAT) (Weigle, 2002 cited in Ghalib & Abdulghani, 2015).

Gonzalez (2014) claimed that a holistic rubric benefits teachers as it does not take much time to create or to grade each piece of writing. It is most useful when there is a time constraint or a large number of essays to grade, and there is no need to give feedback. Furthermore, a holistic rubric evaluates what students can do, not what they cannot, and it can be used consistently if raters are properly trained, which will increase reliability (DePaul University, 2021-2022). Holistic rubrics are more effective if two or more raters grade each paper; this will increase reliability (Hyland, 2003).

However, holistic rubrics are not an effective method for all writing assignments. Cohen (1994) stated that holistic rubrics provide no diagnostic information, ignore some sub-skills, and confuse writing ability with language proficiency. The reliability of holistic scoring relies heavily on raters; therefore, to achieve this, two or more raters are needed, and these raters must be properly trained (Hyland, 2003). Another weakness of holistic rubrics is that it might be hard for raters to make a decision when a student's writing shows some excellent aspects such as organization or content. At the same time, there might be a lot of mistakes in other parts like grammar or word choice (Nodoushan, 2014).

2. Analytic scoring method

In order to grade more precisely, an analytic scoring method was introduced to mark essays, and it is widely used in academic institutes. This marking method was first used to assess ESL students' writing performance at North American

universities of five elements of writing with a different weight: content (30 points), organization (20 points), vocabulary (20 points), language use (25 points), and mechanics (5 points), and it is mostly used for ESL learners (Jacobs et al, 1981). Frey (2018) explained that analytic scoring is a method of evaluating students' work that requires assigning a separate score for each dimension of a task. Analytic scoring rubrics are often used with performance assessment tasks. It can specify the key dimensions of a task and define student performance relating to a set of criteria across performance levels for each dimension. For example, analytic rubrics used to evaluate students' essay writing often include the following perspectives: development of ideas, organization, language use, vocabulary, grammar, spelling, and mechanics. Therefore, an analytic scoring rubric offers more detail about a test taker's writing performance than the single score of a holistic scoring rubric. Wiseman (2012) claimed that analytic scoring interests language instructors because it involves "the separation of the various features of a composition into components for scoring purposes."

Nodoushan (2007) stated that analytic scoring is a result of the idea that writing competency cannot be assessed holistically, as the skill involves various different features. The analytical scoring method is helpful as a diagnostic and teaching tool, which helps teachers to figure out students' weaknesses in writing, and they can find a more suitable teaching approach to improve students' writing skills. Apart from giving useful feedback, the analytic rubric allows raters to weigh each criterion so that the importance of each dimension can be justifiably reflected (DePaul University, 2021-2022). Nakamura (2004) argued that analytic scoring is more accurate than the holistic method because when marking, teachers are supposed to concentrate on the same part of the writing in order to evaluate the same aspects of that writing piece. Another advantage suggested by Hyland (2003) is that the analytic scoring method can prevent a mix-up of separating components when a piece of writing is being graded.

The analytic scoring method also has some limitations. Hyland (2003) pointed out that this method is time-consuming, the overall performance might be overlooked, and descriptors might be unclear. In addition, if each criterion is not well-defined, each rater may come up with a different score (DePaul University, 2021-2022). Hillocks (1995) pointed out that evaluating a written text considering only the scores of subskills overlooks the importance of the interconnection of written discourse and causes a misunderstanding that each aspect of writing can be fairly assessed separately. Furthermore, it might be difficult even for some experienced raters to numerically score based on particular descriptors (Hamp-Lyons, 1989).

As these two scoring methods have their own strengths and weaknesses, Weigle (2002) recommended using a holistic rubric for large classes with limited time and limited resources. Nakamura's study (2004) suggested having multiple raters and multiple rating items as the most preferred option, but if not possible, having one overall evaluation and multiple raters could be the second best choice. Park (2003) stated that no scoring method can be effectively used to serve all purposes; both holistic and analytic scoring can produce unreliable and invalid results. Thus, instructors or those who are in charge need to make a decision to choose a more appropriate alternative for the specific testing situation.

Multiple choice exams

In contrast to an essay writing exam, which needs a proper scoring method to grade, a multiple choice exam is easy to grade, as it does not rely on the judgment of an examiner. A multiple choice exam is regarded as one of the most effective approaches to evaluating student achievement in tertiary education (Stough, 1993). According to Brame (2013), in a multiple choice item, there are two main parts: a stem (problem) and alternatives (suggested solutions). The alternatives consist of a correct answer and distractors (incorrect answers.) When constructing a multiple choice exam, teachers are advised to use familiar language, avoid trick questions and negative wording, and write questions throughout the semester (University of Waterloo). The stem or problem needs to be clearly written.

The two most important characteristics of a multiple choice exam are its content validity and reliability. Cook & Beckman (2006) defined validity as the range of scores obtained from an assessment tool that truly represents what students have learned. The validity of multiple choice exams depends on what and how a test creator selects to include in the exam. The selection of items concerning content and level of learning must be systematically made in order that the exam will achieve validity (Center for Teaching and Learning, University of North Carolina at Chapel Hill, 1990). For reliability, Brame (2013) defined it as the degree of consistency in measuring a learning outcome. Tuckman (1993) found that different markers are likely to give different grades to the same essay writing response, and the same marker often gives different grades to the same essay writing response on different occasions. Reliability mentioned in the context of multiple choice exams mainly refers to internal consistency, which assesses the correlation between scores students obtain from two parallel sets of an exam (Downing, 2004 cited in Ali, et al, 2016).

Multiple choice exams offer a variety of benefits when they are used as an assessment tool. The first benefit which is widely known is that it is easy to score, as there is only one correct answer for each item. This results in high reliability. Luo and Zhang (2011) claimed that for a large-scale examination, a multiple choice question can help save time, effort, and money; therefore, more questions can be added to enhance reliability. Moreover, its versatility enables the assessment of various levels of learning outcomes, ranging from recalling to analyzing information (Brame, 2013). Weimer (2018) stated that this type of exam includes a wide range of topics in a single exam, allows raters to grade quickly and objectively, and simple statistics can be used to analyze if the items can discriminate between those who really understand what they have learned and those who do not. The findings in Little and Bjork's study (2012) confirmed that multiple choice exams can be used as a tool for learning to reinforce knowledge both short term and long term, and they will be more useful if they are carefully and properly constructed. Tozoglu et al. (2004) pointed out that students preferred multiple choice format to essay writing format and also proposed that when students have a positive tendency toward a multiple choice test format, they are able to collaborate, establish links with teachers, and be motivated when taking tests.

In spite of the fact that multiple choice exams are advantageous in many ways, their downsides are obviously noticed. Critics oppose using this type of exam as it is incapable of assessing higher thinking skills and allows test takers to guess the correct answers. Similarly, Cahill and Leonard (1999) agreed that students can come up with correct answers because of their ability to analyze possible answers, not because of the knowledge that the exam aims to test. Weimer (2018) pointed out that wrong answer alternatives might be sources of misinformation, especially when they are carefully considered and selected as correct answers. Although multiple choice exams are easy to mark and not time-consuming, the procedures of constructing an exam can take a long time and are complex because teachers have to think of what should be tested, the number of items, and how to design plausible distractors (Luo & Zhang, 2011).

All in all, to achieve a goal in teaching writing skills, teachers must be well-trained and well-prepared to choose or adapt a teaching approach that is well-suited to the education context. Not only the approach to teaching but also the assessment and evaluation methods play a vital role in the teaching procedures. Designing and constructing exams with validity and reliability is part of the assessing and evaluating process which will contribute to a precise and accurate evaluation of student performance and writing competency. Furthermore, giving feedback and grading systematically can definitely enhance students' writing skills. Therefore, teachers' roles and responsibilities to ensure success are extensive and not restricted to just teaching.

With writing being a critical skill in both academic and professional fields, teaching and assessing students' writing competency accurately and efficiently are imperative. As a multiple choice exam has several benefits, it could be a practical tool for evaluating students' writing skills.

Research Methodology

Research subjects

The subjects of the study were 79 third-year students who were studying English as their minor subject at the Faculty of Arts, Silpakorn University in 2021.

Ethical considerations

The students were officially informed of the purpose of the study and were asked for their consent. Their participation in the research was voluntary and they could withdraw from the research project at any point for any reason. When conducting the research, the participants would not be physically or psychologically harmed. Moreover, they were aware that their exam results would be kept confidential.

Research instrument design

Two types of exams were used in the study: multiple choice exam and essay writing exam. A cause and effect essay was used in both the multiple choice and essay writing exams.

1. The multiple choice exam was a cloze test comprising 30 items, which were designed to assess the following:

1.1 The understanding of essay elements and organization, which consist of hook, thesis statement, topic sentences, unity and coherence, and conclusion.

1.2 The grammatical points used in the multiple choice exam were chosen from a list of grammatical mistakes that the students had frequently made in their previous essays, which include tenses, active/passive voice, verb forms, parts of speech, relative clauses, pronouns, fragments, and run-on sentences.

1.3 Vocabulary

The exam was reviewed by two native speakers of English. Having been corrected and adjusted, the exam was trialed by ten students to find the ability to discriminate between the good and the less able students and the difficulty of the exam.

2. The essay exam provided four topics from which the students could choose to write.

Data collection and analysis

The participants were given two sets of the exams: a multiple choice exam and an essay writing exam on different days. They had 40 minutes to complete the multiple choice exam, which was on Google Form and 90 minutes for the essay writing exam, whose topics were posted on Google Classroom. The students took both exams online. They were not allowed to use dictionaries or search the Internet for any information. The results of these two exams were part of the course assessment; therefore, the students would do the exams to their full ability.

The essay exams were graded using analytic rubrics, which corresponded with the multiple choice test items. The rubrics were set by the researchers. Mean and Standard Deviation (S.D.) were chosen for the analysis of the collected data. The two-tailed test was used to test for statistical significance of the result.

Results

This research compared the efficacy of multiple choice and essay writing exams in assessing students' competency in writing essays. The results are presented here in three parts as follows:

Population Information

The population of this research are third-year English minor students of the Faculty of Arts, Silpakorn University during the first semester of the 2021 academic year. The data was collected from 79 students.

Table 1 Overall mean comparison

Paired samples statistics				
		Mean	N	Std. Dev.
Pair 1	Multiple Choice Score	21.0253	79	3.53363
	Essay Score	19.3418	79	3.58375

Table 1 shows the students' average scores in the multiple choice exam and essay writing exam. The students performed better in the multiple choice exam with a score of 21.0253 while the same group of students received an average score of 19.3418 in the written exam.

Essay elements results

The following tables are the results of essay elements (see chapter 3) part of the analysis

Table 2 Essay elements mean comparison

Paired samples statistics					
		Mean	N	Std. Dev.	Std. Error Mean
Pair 1	Essay Elements Choice	8.3797	79	1.53007	.17215
	Essay Elements Essay	8.7848	79	1.63825	.18432

As shown in Table 2, students received a score of 8.3797 in the multiple choice exam and a score of 8.7848 in the essay writing exam.

Table 3 Paired sample test

Paired Differences							
95% Confidence							
Interval of the							
Difference							
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper	t	df (2-tailed)
Pair 1	Essay						
	Elements						
	Choice -						
	Essay	-.40506	2.14547	.24138	-.88562	.07550	-1.678 78 .097
	Elements						
	Essay						

As Table 3 illustrates, compared to the multiple choice exam, the average score in the essay writing exam is slightly higher but it is not statistically significant.

Table 4 Grammar mean comparison

		Mean	N	Std. Dev.	Std. Error Mean
Pair 1	Grammar Choice	6.2911	79	1.98812	.22368
	Grammar Essay	5.9494	79	1.95395	.21984

Table 4 shows the average grammar scores in both the multiple choice exam and the essay writing exam. In the multiple choice exam, the average score was 6.2911 with a standard deviation of 1.98812. The essay writing exam grammar average score was slightly lower (5.9494).

Table 5 Paired sample test

Paired Differences								
95% Confidence								
Interval of the								
Difference								
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Grammar							
	Choice– Grammar	.34177	2.55647	.28763	-.23085	.91439	1.188	78
Essay								

According to Table 5, although the average multiple choice grammar score was higher than the essay writing exam, statistically, it is not significant.

Table 6 Vocabulary mean comparison

Paired samples statistics					
		Mean	N	Std. Dev.	Std. Error Mean
Pair 1	Vocab Choice	6.3544	79	1.44149	.16218
	Vocab Essay	4.6076	79	2.33380	.26257

Table 7 Paired samples test

Paired Differences								
95% Confidence								
Interval of the								
Difference								
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Vocab							
	Choice Vocab	1.74684	2.64317	.29738	1.15480	2.33887	5.874	78
Essay								

Table 6 presents the vocabulary scores from both the multiple choice and essay writing exams. Students attained a higher score in the multiple choice exam (6.3544) than in the essay writing exam (4.6076) with standard deviations of 1.44149 and 2.33380 respectively. This difference is statistically significant as seen in Table 7.

Difficulty index mean comparison

Overall DF mean comparison

Table 8 Paired samples test

Paired Differences									
95% Confidence									
Interval of the									
Difference									
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)	
Pair 1	Multiple Choice DF – Essay DF	.05613	.26575	.04852	-.04310	.15537	1.157	29	.257

Table 8 shows the paired difference which was not statistically significant.

Discrimination index mean comparison

Overall DC mean comparison

Table 9 Paired sample statistics

	Mean	N	Std. Dev.	Std. Error Mean
Pair 1	Multiple Choice DC	30	.18281	.03338
	Essay DC	30	.24718	.04513

Table 10 Paired samples test

Paired Differences									
95% Confidence									
Interval of the									
Difference									
	Mean	Std. Dev.	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)	
Pair 1	Multiple Choice DC – Essay DC	-.01050	.30349	.05541	-.12382	.10282	-.189	29	.851

Tables 9 and 10 presents the overall DC (Discrimination Index) of both the multiple choice exam and the essay writing exam. Although the essay writing exam DC average (0.2848) was higher than the multiple choice DC (0.2743), it is not statistically relevant ($p>0.05$). This meant that both exams were discriminating equally.

Conclusion and Discussion

Results showed that the students performed better in the multiple choice exam, scoring an average of 21.03 points while in the essay writing exam, the students received 19.34 points. Isolating each assessment topic, students achieved a marginally better performance in the grammar and vocabulary topics of the multiple choice exam, while they received a slightly higher score in the essay elements of the essay writing exam. However, except for the vocabulary section, the difference in the scores of the remaining parts was not statistically significant. In addition, the difficulty index mean comparison showed that both the types of exam were of comparable difficulty and the discrimination index of the multiple choice exam and essay writing exam were comparable with no statistical significance, 0.2743 and 0.2848 respectively.

As both exams were designed to cover the same topics the students had learned in class, namely essay structure, grammar points, and vocabulary, and have comparable difficulty and discrimination indices, the difference in the scores showed no statistical significance. One contributing factor to this may be the use of the essay writing exam analytic rubric as mistakes other than what were included in the analytic rubrics were not recorded. Furthermore, the analytic rubrics had a fixed maximum number of points that could be deducted. For example, if the grammar section had 10 points, one point would be deducted from the maximum 10 points per mistake. If a student had made more mistakes than the maximum allowed, in this case 10 mistakes, that student would still receive 0 points. The numbers showed that both the designed multiple choice exam and the essay writing exam were comparable in difficulty and both were able to effectively discriminate students.

In conclusion, multiple choice exams, if constructed well and covering all the topics that are taught in class, can be used as a tool in evaluating writing competency and can have comparable effectiveness to essay writing exams. Multiple choice exams can shorten the time spent on marking, improve efficiency, and reduce boredom, which could affect reliability in marking and grading.

Guidelines in constructing and using multiple choice exams for writing classes

Instructors should decide which topics the multiple choice exam would cover, whether it be grammatical points, range of vocabulary, or essay organization.

Limitations and Recommendations for Further Research

Limitations:

- The multiple choice exam did not test students' potential to show creativity, originality, and depth of understanding of the topic.
- Further research is needed to strengthen the reliability and validity of the findings.

Recommendations:

- **Sample:** This research can be conducted on non-English major and minor students at the Faculty of Arts or from other faculties.
- **Exam:** More items covering other topics should be added to the exam. Fill in the blank items could also be added to the multiple choice exam.

Acknowledgement

Our deepest appreciation and gratitude extend to the following people who have contributed to this research. We would like to first and foremost extend our gratitude to the Faculty of Arts, Silpakorn University whose funding in 2021

academic year made this research possible. Secondly, we would like to thank the students who have participated in this research. The data collected from them has been pivotal in our research. Lastly, we would like to thank our family and friends, whose constant and unwavering support motivated us throughout the process of this research.

References

Boyle, A. P. (2019). Writing Better Essay Exams. *IDEA Paper*, 76. Retrieved 15 July 2022, from https://ideacontent.blob.core.windows.net/content/sites/2/2020/01/Writing_Better_Essay_Exams_IDEA_Paper_76.pdf

Brame, C. J. (n.d.). *Writing Good Multiple Choice Test Questions* | Center for Teaching | Vanderbilt University. Vanderbilt Center for Teaching. Retrieved 17 July 2022, from <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. 2nd ed. Addison Wesley Longman.

Cahill, D. R., & Leonard, R. J. (1999). Missteps and masquerade in American medical academy: Clinical anatomists call for action. *Clinical Anatomy*, 12, 220-222.

Cameron, B. (2021). *The Benefits of Essay Writing*. Time Business News. Retrieved 15 July 2022, from <https://timebusinessnews.com/the-benefits-of-essay-writing/>

DePaul University. (n.d.). *Types of Rubrics*. DePaul University Resources. Retrieved 15 July 2022, from <https://resources.depaul.edu/teaching-commons/teaching-guides/feedback-grading/rubrics/Pages/types-of-rubrics.aspx#holistic>

Designing Multiple-Choice Questions | Centre for Teaching Excellence. (n.d.). University of Waterloo. Retrieved 14 July 2022, from <https://uwaterloo.ca/centre-for-teaching-excellence/catalogs/tip-sheets/designing-multiple-choice-questions>

Erturk, S., van Tilburg, W. A. P., & Igou, E. R. (2022). Off the mark: Repetitive marking undermines essay evaluations due to boredom. *Motivation and Emotion*, 46, 264-275.

Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vols. 1-4). SAGE Publications.

Ghalib, K., & Al-Hattami, A. A. (2015). Holistic versus Analytic Evaluation of EFL Writing: A Case Study. *English Language Teaching*, 8(7).

Gonzalez, J. (2014, May 1). *Know Your Terms: Holistic, Analytic, and Single-Point Rubrics*. Cult of Pedagogy. Retrieved 15 July 2022, from <https://www.cultofpedagogy.com/holistic-analytic-single-point-rubrics/>

Graham, S. (2019). Changing How Writing Is Taught. *Review of Research in Education*, 43(1), 277-303.

Hamp-Lyons, L. (1989). Raters respond to rhetoric in writing. In H.W. Dechert & Raupauch (Eds.), *Interlingual processes*, pp. 229-244, Gunter Narr.

Hillocks Jr., G. (1995). *Teaching writing as reflective practice*. Teachers College Press.

How effective is multiple choice... | Durrington Research School. (2019). Research Schools Network. Retrieved 14 July 2022, from <https://researchschool.org.uk/durrington/news/how-effective-is-multiple-choice-quizzing-a-very-effective-b-effective-c-ineffective>

Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse*. pp.115-130, Longman.

Hyland, K. (2003). *Second Language Writing*. Cambridge University Press.

Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.

Lee, S., & Schmidgall, J. (2020). The Importance of English Writing Skills in the International Workplace.

Little, J., & Bjotk, E. (2012). The Persisting Benefits of Using Multiple-Choice Tests as Learning Events. Retrieved 14 July 2022, from https://www.researchgate.net/publication/267028787_The_Persisting_Benefits_of_Using_Multiple-Choice_Tests_as_Learning_Events

Mousavi, H. S. & Kashefian-Naeeini, S. (2011). Academic writing problems of Iranian post-graduate students at national university of Malaysia (UKM). *European Journal of Social Sciences*, 23(4), 593-603.

Luo, S., & Zhang, X. (2011). Multiple-choice Item and Its Backwash Effect on Language Teaching in China. *Theory and Practice in Language Studies*, 1(4), 423-425.

Nakamura, Y. (2004). A comparison of holistic and analytic scoring methods in the assessment of writing. *The Interface Between Interlanguage, Pragmatics and Assessment: Proceedings of the 3rd Annual JALT Pan-SIG Conference*.

Nodoushan, M. A. S. (2007). Thinking on the write path. *Training Journal*, 37-40.

Nodoushan, M. A. S. (2014). Assessing Writing: A Review of the Main Trends. *Studies in English Language and Education*, 1(2), 119-129.

Nordquist, R. (2019). *Holistic Grading (Composition)*. ThoughtCo. Retrieved 15 July 2022, from <https://www.thoughtco.com/holistic-grading-composition-1690838>

Park, T. (2003). Scoring Procedures for Assessing Writing. *Columbia University Libraries*, 3(1), 1-3.

Patton, M. Q. (1987). *Qualitative Evaluation Methods*. SAGE Publications.

Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.

Sim, M. A. (2010). Some thoughts on writing skills. *Annals of Faculty of Economics*, 1(1), 127-133.

Steele, C. W. (1997, Fall). Essays - well worth the effort. *College Teaching*, 45(4).

Stough, L. (1993). Research on Multiple-Choice Questions: Implications for Strategy Instruction. Retrieved 14 July 2022, from https://www.researchgate.net/publication/234623778_Research_on_Multiple-Choice_Questions_Implications_for_Strategy_Instruction

Tisi, J., Whitehouse, G., Maughan, S., & Burdett, N. (2013). A Review of Literature on Marking Reliability Research (Report for Ofqual). Retrieved 15 July 2022, from <https://www.nfer.ac.uk/publications/mark01/mark01.pdf>

Tozoglu, D., Tozoglu, M. D., Gurses, A., & Dogar, C. (2004). The Students' Perceptions: Essay versus Multiple-choice Type Exams. *Journal of Baltic Science Education*, 2, 52-59.

Tuckman, B. W. (1993). "The essay test: a look at the advantages and disadvantages". *Nassp Bulletin*, 77, 55, 20-26.

University of North Carolina at Chapel Hill. (n.d.). *Essay Exams*. UNC Writing Center. Retrieved 15 July 2022, from <https://writingcenter.unc.edu/tips-and-tools/essay-exams/>

Walstad, W., & Becker, W. E. (1994). Achievement Differences on Multiple-Choice and Essay Tests in Economics. *College of Business Faculty Publications University of Nebraska Lincoln*, 84(2), 193-196.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Weimer, M. (2018, February 21). *Multiple-Choice Tests: Revisiting the Pros and Cons*. Faculty Focus. Retrieved 17 July 2022, from <https://www.facultyfocus.com/articles/educational-assessment/multiple-choice-tests-pros-cons/>

Wiseman, S. (2012). A Comparison of the Performance of Analytic vs. Holistic Scoring Rubrics to Assess L2 Writing. *Iranian Journal of Language Testing*, 2(1), 59-92.

Zeng, X. (2018, November). A Study on the Improvement of English Writing Competence for College Students. *Journal of Language Teaching and Research*, 9(6), 1334-1348.