HONORS COLLEGE THESIS

A Comparison of the Effectiveness of Different Evaluation Techniques

BY

VICTOR NDIRITU

SENIOR

MAJOR: STATTISTICS

OKLAHOMA STATE UNIVERSITY

05/1/2015

ACKNOWLEDGEMENT

I would like to thank Dr. Tamer Elbayoumi for his support throughout the Thesis preparation and research. I approached him with an idea for research that I had for a long time wanted to investigate. Without his guidance, support, organization, and inspiration the thesis research would not have been nearly as enjoyable as it was.

I would like to thank Dr. Goad for not only being my second reader but also my academic advisor. Through her guidance, I acquired valuable skills that made the analysis of my thesis possible. I look forward to our future interactions as a graduate student.

I would like to thank all of my honor's college advisors, present and past, for their unwavering and unending support throughout my career as a member of the honors college at Oklahoma State University.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

II. CHALLENGES 3

III. EXPERIMENT 5

IV. RESULTS 8

V. DISCUSSION 26

VI. CONCLUSION 30

VII. REFERENCES 31

CHAPTER 1

INTRODUCTION

The United States has for a long time been heralded as an academic powerhouse. Many students all over the world aspire each year to join the ranks of any of its tertiary academic institutions. However, the US has seen a dip in its ranking in recent years. According to a report published by Harvard University's Program on Education Policy and Governance, "Students in Latvia, Chile and Brazil are making gains in Academics three times faster than American students while those in Portugal, Hong Kong, Germany, Poland, Liechtenstein, Slovenia, Colombia and Lithuania are improving at twice the rate." This is but one report from a myriad of reports that suggest the same. One begins to wonder what it is about the American education system that is lacking and needs improvement if America is to hold its own in the world.

According to an online journal called Children of the Code, some key features of a good and effective learning experience include: Teacher quality, class size and better preparation for both students and teachers. While all these are valid points, they do not effectively account for the large gap in performance between children in American schools and their foreign counterparts. I sought out the one factor that might be making all the difference. I was in a unique position to do so because I have grown up in a completely different education system. To do this, I compared my learning experience in the university I attended in Kenya and sought out similar information from fellow international students from a variety of countries. My research yielded a key observation that I believe advantages different education systems over America. I observed that in

most other countries, their methods of academic evaluation and learning varied from those used here in America.

In other countries, in most of their subjects, students are evaluated using a mixture of exam and hands-on projects. When a concept is taught, students are required to be prepared to apply it in case scenarios that might be presented, on a regular basis, by their professors. While this does happen in some subjects in America, it happens on a much lower scale. More often than not, the only means of evaluation used is solely examinations. The lack of a hands-on dimension to these classes may adversely affect students' retention and application capacity.

Armed with this preliminary research, I formulated the hypothesis that students would in general perform better if materials requiring the application of theory were included in the teaching and evaluation procedures of each class.

CHALLENGES

The biggest challenge during my research was the lack of prior sufficient study of materials related to my topic. No one had done research in the past comparing the effectiveness of evaluation methods. I ran across a few papers talking about effective ways to learn, but none that identified examinations as perhaps not the most effective of evaluation tools. This meant that I was charting new waters and had to formulate research from the ground up.

Additionally, another big challenge in conducting this research was trying to formulate the methodology of collecting useful and relevant data. The most straight forward way was to access the students academic records and compare their performance in classes that are heavily exam oriented and those that having a project or independent study component. Unfortunately, doing that would be in violation of several school regulations. This necessitated coming up with a new way to collect data that would be more time consuming but just as effective.

Student participation in my research was another issue that I had to deal with. I was not able to offer any incentives for them to participate in the survey. To deal with this, I reviewed persuasive language used by other researchers in their email communications to get students to participate. I sent out periodic email reminders and eventually got the predetermined sample size amount of responses.

Last but not least, ensuring that my research questions were fair and unbiased was very daunting. I needed to phrase the questions in a professional and concise manner. To ensure fairness, I copied the questioning format from different surveys.

Having addressed these challenges as best as I could, I proceeded to outline the experiment as detailed in the next section.

EXPERIMENT

Study Design:

Recruitment Methods

Potential subjects were selected from different Oklahoma State classes at the beginning of the fall semester. The subjects were chosen at random without regard for the testing format of the classes that they were currently in. Students were reminded to participate in the study via periodical emails. The emails were not descriptive of the research objective so as to avoid any bias. They were merely a kind request to complete a survey for an undergraduate research thesis.

Eligibility was restricted to graduate and undergraduate students at Oklahoma State University. Also, undergraduate students were limited to the sophomore level and above since a lower level student would not have taken enough classes to answer the questions in the survey. As illustrated in figure 2, the distribution of academic classification was 42% senior, 36% junior, 14% sophomore and 8% graduate students. An inclusion and exclusion clause as detailed below was included in the email:

Inclusion and Exclusion Criteria

Adults unable to consent, individuals who are not yet adults and prisoners will be excluded from the research. If you're unsure whether you belong to the aforementioned, please email the researcher or contact the relevant authority. Pregnant women, given they are students will not be excluded from the study.

Sample Size determination

The sample size was determined to be 40 using a power test with power at 0.95. Out of 51 respondents, 11 provided incomplete and hence erroneous data (given the connectedness of the survey questions). As shown in Figure 1, the respondents included 58% females and 42% males.

Data Collection and Survey Design

Data was collected through the use of Survey Monkey, an online survey publisher and data analysis tool. The survey (attached to the end of the thesis) had a total of 10 questions. Out of the ten questions, six were used to classify the students according to various categories such as their gender, academic classification, academic performance and their State residency. The data collected was heavily reliant on student opinion due to the University Policy restricting access to students' academic records. The researcher complied with the Internal Review Board (IRB) requirements of completing a CITI course prior to the experiment.

In order to come up with more conclusive results, the survey divided the hands-on aspect a class could incorporate into three categories: projects, independent study and a mixture of all three formats.

The survey was extended to account for difference in performance given that the examinations given in classes were of different formats. These formats include, multiple-choice, open-ended questions, a mixture of both and other formats. 'Other formats' refers to formats such as gap and matching questions. Perhaps, if exams would prove to be more effective, learning their most effective format would be beneficial.

Data Management

The data was stored without any identifying information given the sensitive nature of the data. The data's integrity was protected according to the terms and conditions of Survey Monkey. A link to these terms and conditions is posted on the reference section.

Data Analysis

Data Analysis was done through the use of excel, SPSS and graphs from Survey Monkey. The results were divided according to the demographic variables that were described in the data collection and management section. The survey distinguished performance and comprehension in the classes. Students were asked about their performance in the different evaluation procedures and then asked about their comprehension of subject matter in the same procedures. The experiment makes the assumption that there aren't any assumptions about the relationship between comprehension of material and performance in the class. This is done because the experiment is heavily reliant upon student opinion and recollection as mentioned in the data collection and survey design stage.

METHODS

Seeing as how the data is non-parametric, the parametric methods of Friedman test and Tukey-HSD test were employed. They are relevant because they analyze data with blocks. An alpha level of 0.05 is employed throughout the research.

RESULTS

Sample Categories

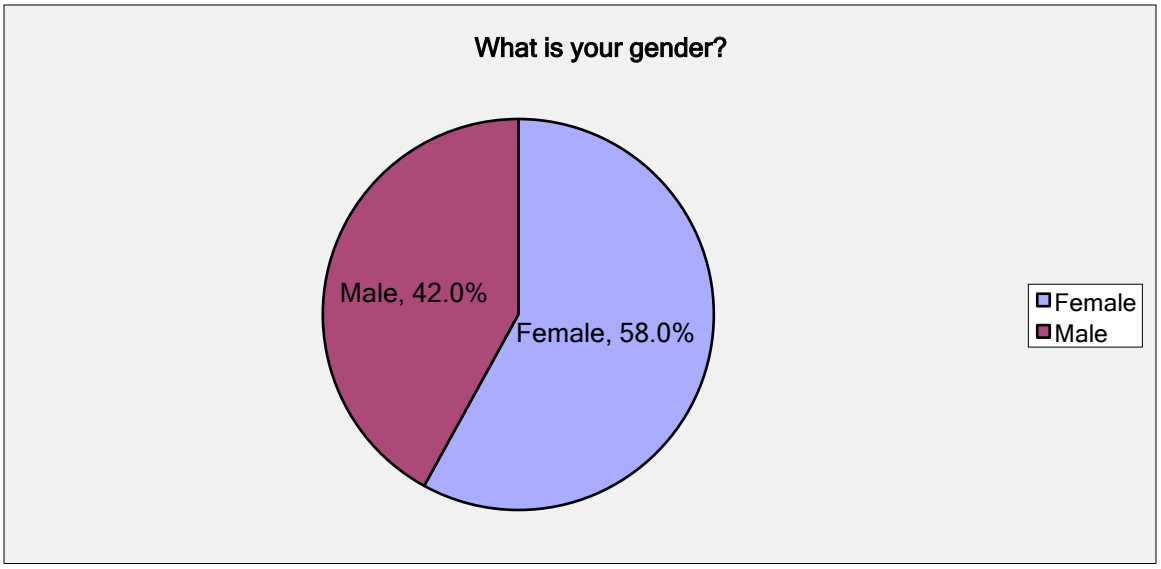

Figure 1 shows the sample distribution of gender in percentage.

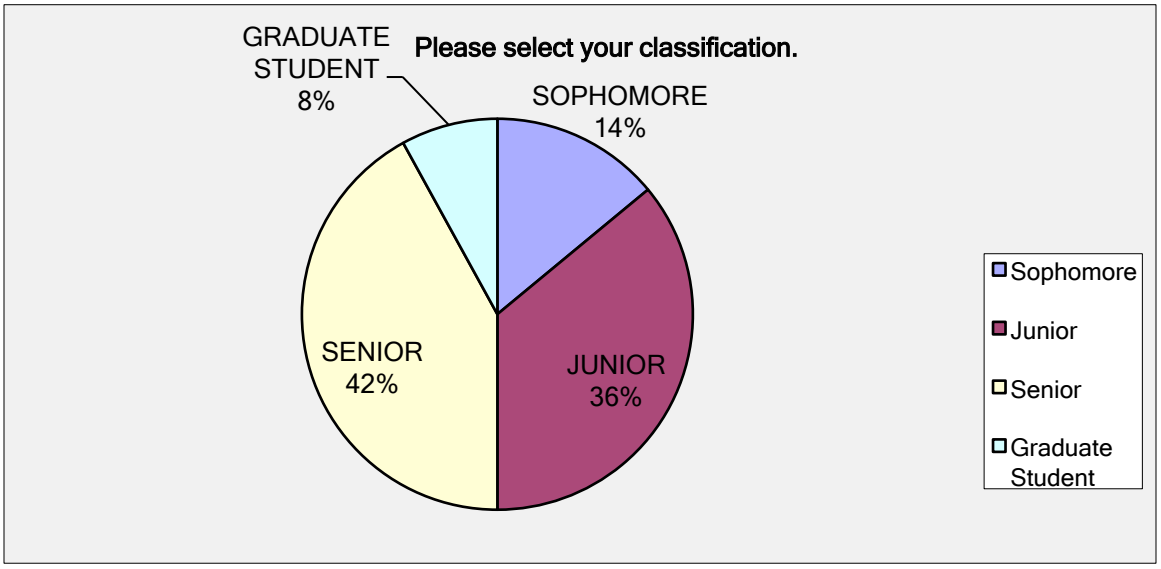

Figure 2 shows the sample distribution of academic classification in percentage.

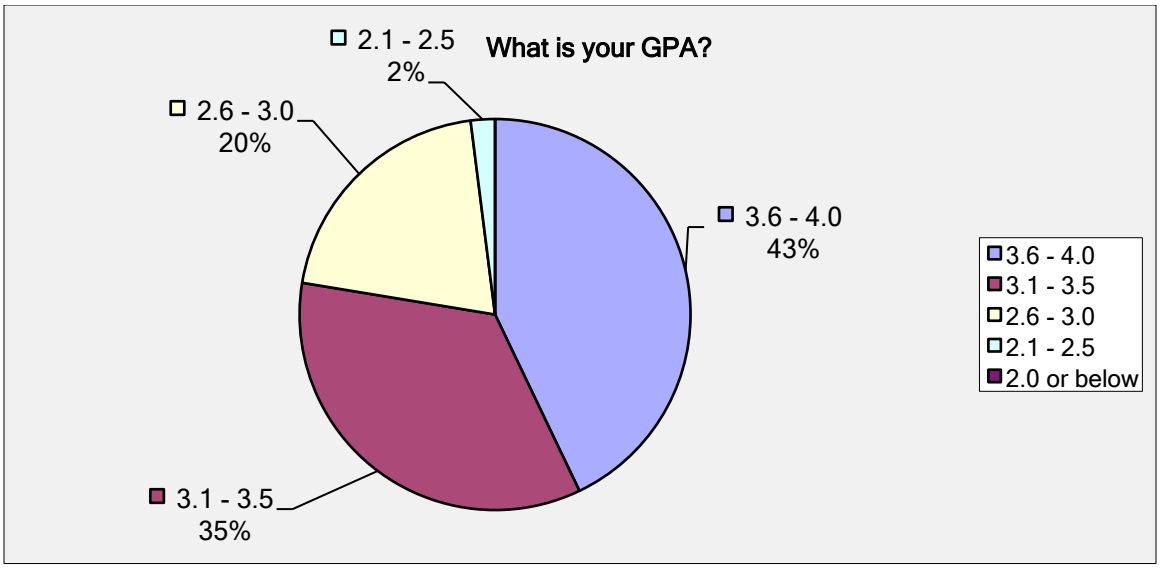

Figure 3 shows the sample distribution of the academic performance by the GPA.

Performance Among the different evaluation Procedures

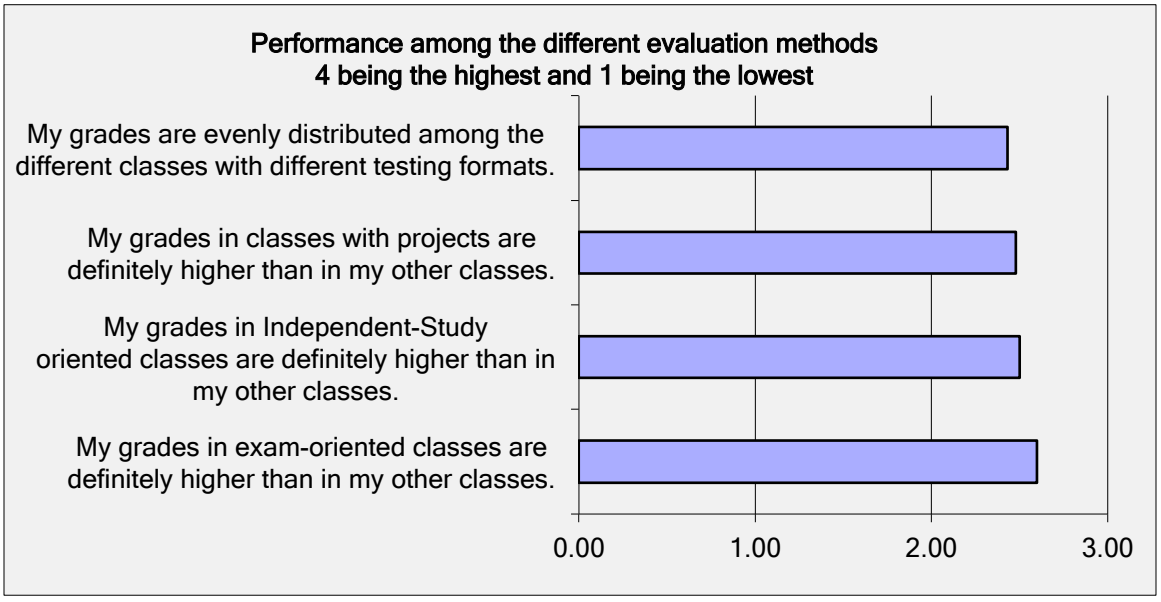

Figure 4 shows a summary of the responses to questions regarding students' performance in classes with different evaluation procedures. The horizontal axis represents the mean rank.

Data of performance of Respondents in classes using the following evaluation formats

Table 1 shows the individual ranks that each of the 40 respondents gave to the different evaluation formats with respect to performance in the class.

Respondents	Exams	Independent Study	Projects	Mixture of Formats
1	1	4	2	3
2	1	2	3	4
3	4	2	1	3
4	3	4	2	1
5	3	2	4	1
6	1	2	4	3
7	4	1	2	3
8	2	1	3	4
9	1	4	3	2
10	2	4	3	1
11	2	3	4	1
12	3	4	1	2
13	1	3	4	2
14	2	4	3	1
15	3	2	4	1
16	3	4	2	1
17	4	1	3	2
18	3	1	2	4
19	4	2	1	3
20	4	3	2	1
21	4	1	2	3
22	4	3	1	2
23	1	2	3	4
24	3	4	1	2
25	3	4	2	1
26	3	4	2	1
27	2	3	1	4
28	1	2	3	4
29	1	2	3	4
30	4	2	1	3
31	4	2	3	1
32	3	4	1	2
33	3	2	4	1
34	1	3	2	4
35	2	1	3	4
36	1	2	3	4
37	4	1	3	2
38	2	3	4	1
39	3	1	2	4
40	2	3	4	1

Friedman Test for the difference in Performance among the different Testing Formats

Table 2 shows the results of the Friedman test that was used to detect a difference in performance among the different evaluation procedures.

Table 2

Friedman's Test	Value
Alpha	0.05
H-stat	0.51
df	3
p-value	0.916689026
sig	no

H_o: There is no significant difference in performance among the different evaluation procedure.

H_a: There is a significant difference in the in the performance among the different evaluation procedures.

Conclusion**:** Fail to reject the null hypothesis at alpha=.05 and conclude that there is insufficient evidence to show that performance differs according to the evaluation formats.

Comprehension of Material Among the Different Evaluation Formats

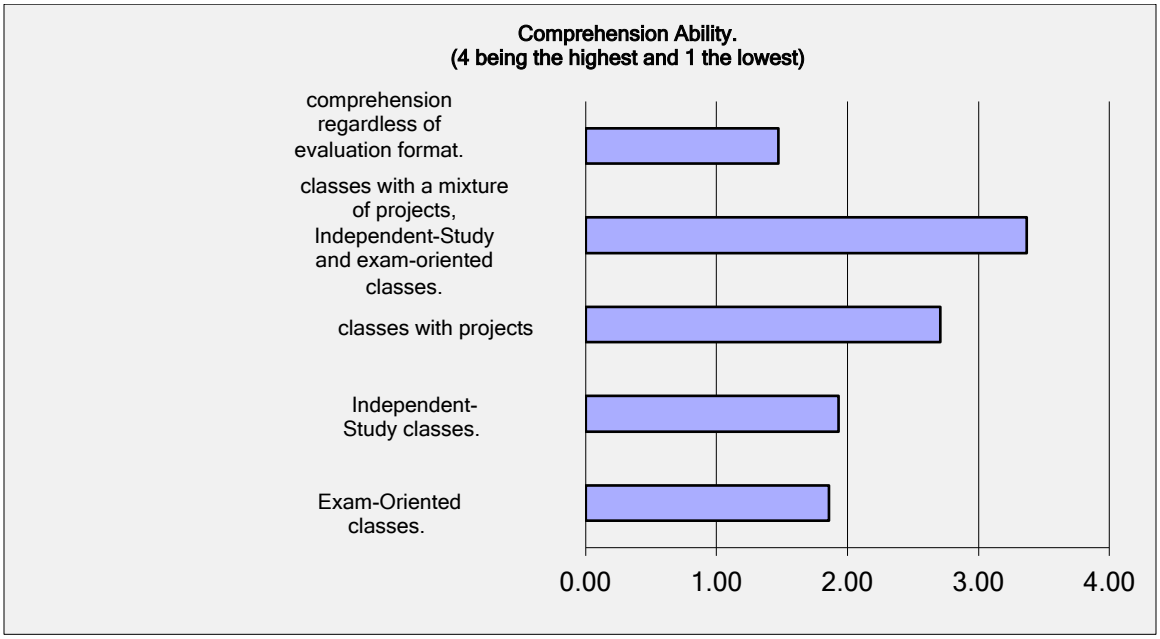

Figure 5 shows a summary of the responses to questions regarding students' comprehension in classes with different evaluation procedures. The horizontal axis represents the mean rank.

Data of performance of Respondents in classes using the following evaluation formats

Table 3 shows the individual ranks that each of the 40 respondents gave to the different evaluation formats with respect to comprehension.

Respondents	Exams	Independent Study	Projects	Mixture of Formats
1	2	2	3	2
2	3	2	2	4
3	3	3	2	4
4	3	3	3	3
5	2	3	4	4
6	3	4	3	4
7	3	3	4	3
8	2	4	3	3
9	3	4	3	3
10	4	3	3	4
11	3	3	3	3
12	2	3	3	4
13	2	3	3	4
14	1	3	2	3
15	2	2	1	2
16	3	3	3	3
17	2	2	3	3
18	3	2	3	3
19	2	2	4	3
20	2	3	3	3
21	2	3	2	4
22	3	3	1	2
23	3	3	3	3
24	2	3	2	4
25	3	2	3	3
26	2	3	4	3
27	1	3	3	3
28	2	3	4	3
29	3	4	4	4
30	2	3	3	3
31	3	2	3	3
32	2	3	3	3
33	3	2	3	3
34	2	3	3	3
35	2	3	4	3
36	2	3	3	3
37	2	3	4	3
38	1	3	3	4
39	3	3	4	3
40	3	3	3	3

Friedman Test to Test the difference in Comprehension Among the Different Evaluation Formats

Table 4 shows the results of the Friedman test that was used to detect a difference in comprehension among the different evaluation procedures.

Table 4

Friedman's Test	Values
Alpha	0.05
H-stat	18.2775
df	3
p-value	0.000386
sig	yes

H_O: There is no significant difference in comprehension among the different evaluation procedures.

H_a: There is a significant difference in comprehension among the different evaluation procedures.

Conclusion**:** Reject the null hypothesis and conclude that there is a significant difference in comprehension ability among the different evaluation formats

Post-hoc Analysis since the Friedman test showed a significant difference among the Evaluation Procedure

1) Exams Vs. Independent Study

Table 5 shows the Tukey-HSD test comparing exams Vs. Independent- study.

Table 5

TUKEY'S HSD / TUKEY-KRAMER					Alpha	0.05
Groups	***c***	***mean***	***n***	***ss***	***c^2/n***	***c*mean***
Exams	1	2.4	40	17.6	0.025	2.4
Independent Study	-1	2.875	40	12.375	0.025	-2.875
Projects		3	40	22	0	0
Mixture of Formats		3.2	40	12.4	0	0
			160	64.375	0.05	-0.475
Q TEST						
std err	***q-stat***	***df***	***q-crit***	***lower***	***upper***	***sig***
0.101570204	-4.67657	156	3.6772	-0.84849	-0.10151	yes

H_O: There is no significant difference between comprehension between Exams and Independent Study.
H_a: There is a significant difference in comprehension between Exams and Independent Study.

Conclusion**:** There is a significant difference in comprehension ability between exams and Independent Study formats at alpha = 0.05. Comprehension in Independent study is higher than comprehension in examinations.

2) Exams Vs. Projects

Table 6 shows the Tukey-HSD test comparing exams Vs. Projects.

Table 6

TUKEY'S HSD / TUKEY-KRAMER					Alpha	0.05
Groups	***c***	***mean***	***n***	***ss***	***c^2/n***	***c*mean***
Exams	1	2.4	40	17.6	0.025	2.4
Independent Study		2.875	40	12.375	0	0
Projects	-1	3	40	22	0.025	-3
Mixture of Formats		3.2	40	12.4	0	0
			160	64.375	0.05	-0.6
Q TEST						
std err	***q-stat***	***df***	***q-crit***	***lower***	***upper***	***sig***
0.101570204	-5.90724	156	3.6772	-0.97349	-0.22651	yes

H_O: There is no significant difference between comprehension between Exams and projects.

H_a: Projects result in higher comprehension rates compared to examinations.
Conclusion**:** There is a significant difference in comprehension ability between exams and project formats at alpha = 0.05. Comprehension in a project-oriented class is higher than comprehension in an exam-oriented class.

3) Exams Vs. Mixture of Formats

Table 7 shows the Tukey-HSD test comparing exams Vs. Mixture of formats.

Table 7

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	c	mean	n	ss	c^2/n	c*mean
Exams	1	2.4	40	17.6	0.025	2.4
Independent Study		2.875	40	12.375	0	0
Projects		3	40	22	0	0
Mixture of Formats	-1	3.2	40	12.4	0.025	-3.2
			160	64.375	0.05	-0.8

Q TEST

std err	q-stat	df	q-crit	lower	upper	sig
0.101570204	-7.87633	156	3.6772	-1.17349	-0.42651	yes

H_O: There is no significant difference between comprehension between Exams and a mixture of the different evaluation procedures.

H_a: A mixture of the different evaluation formats results in higher comprehension rates compared to classes solely using examinations.

Conclusion**:** There is a significant difference in comprehension ability between exams and a mixture of the evaluation formats at alpha = 0.05. Comprehension in a class with mixtures of different evaluation procedures is higher than comprehension in an exam-oriented class.

4) Independent Study Vs. Projects

Table 8 below the Tukey-HSD test comparing Independent study Vs. Projects.

Table 8

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	c	mean	n	ss	c^2/n	c*mean
Exams		2.4	40	17.6	0	0
Independent Study	1	2.875	40	12.375	0.025	2.875
Projects	-1	3	40	22	0.025	-3
Mixture of Formats		3.2	40	12.4	0	0
			160	64.375	0.05	-0.125

Q TEST

std err	q-stat	df	q-crit	lower	upper	sig
0.101570204	-1.23068	156	3.6772	-0.49849	0.248494	no

H_o: There is no significant difference between comprehension between Projects and Independent Study.

H_a: Projects result in higher comprehension than Independent Study.

Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in comprehension ability between projects and Independent Study formats at alpha = 0.05.

5) Independent Study Vs. Mixture of Formats

Table 9 shows the Tukey-HSD test comparing Independent Study Vs. Mixture of formats.

Table 9

TUKEY'S HSD / TUKEY-KRAMER					Alpha	0.05
Groups	*c*	*mean*	*n*	*ss*	*c^2/n*	*c*mean*
Exams		2.4	40	17.6	0	0
Independent Study	1	2.875	40	12.375	0.025	2.875
Projects		3	40	22	0	0
Mixture of Formats	-1	3.2	40	12.4	0.025	-3.2
			160	64.375	0.05	-0.325
Q TEST						
std err	*q-stat*	*df*	*q-crit*	*lower*	*upper*	*sig*
0.101570204	-3.19976	156	3.6772	-0.69849	0.048494	no

H_o: There is no significant difference between comprehension between a mixture of different evaluation formats and Independent Study.
H_a: Mixture of multiple different formats result in higher comprehension than Independent Study.
Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in comprehension ability between a mixture of different evaluation formats and Independent Study formats at alpha = 0.05.

6) Projects Vs. Mixture of Formats

Table 10 shows the Tukey-HSD test comparing Projects Vs. Mixture of formats.

Table 10

TUKEY'S HSD / TUKEY-KRAMER					Alpha	0.05
Groups	*c*	*mean*	*n*	*ss*	*c^2/n*	*c*mean*
Exams		2.4	40	17.6	0	0
Independent Study		2.875	40	12.375	0	0
Projects	1	3	40	22	0.025	3
Mixture of Formats	-1	3.2	40	12.4	0.025	-3.2
			160	64.375	0.05	-0.2
Q TEST						
std err	*q-stat*	*df*	*q-crit*	*lower*	*upper*	*sig*
0.101570204	-1.96908	156	3.6772	-0.57349	0.173494	no

H_O: There is no significant difference between comprehension between a mixture of different evaluation formats and Projects.

H_a: Mixture of multiple different formats results in higher comprehension than Projects.

Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in the comprehension ability between a mixture of different evaluation formats and Projects at alpha = 0.05.

Comparison of Exam Formats

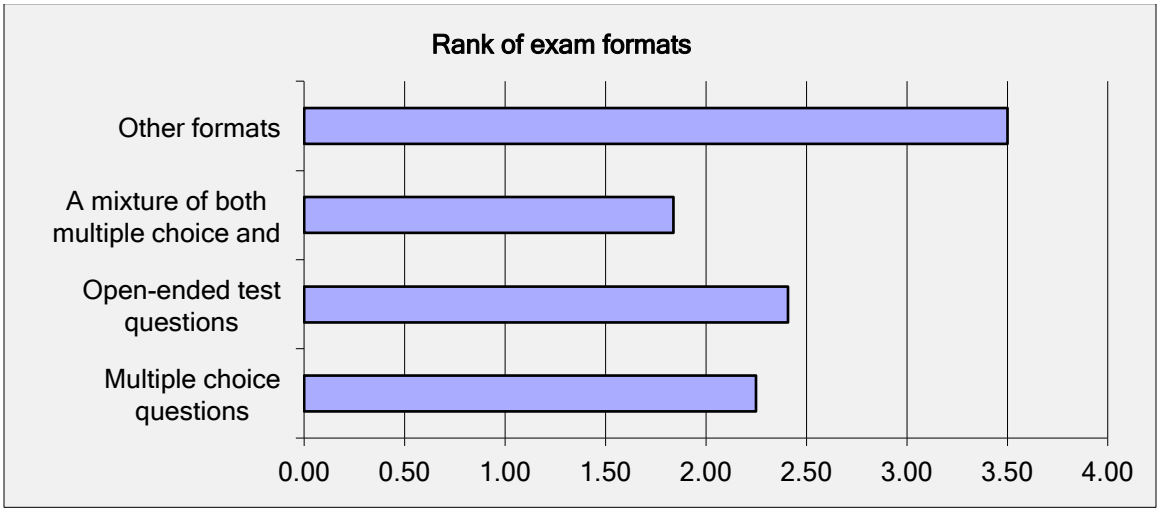

Figure 6 shows a summary of the responses to questions regarding students' preference of exam format. The horizontal axis represents the mean rank.

Data of respondents preference as it relates to exam formats

Table 11 shows the individual ranks that each of the 40 respondents gave to the different exam formats.

Respondents	Multiple Choice	Open-Ended	Mixture-of-Both	Other
1	2	1	3	4
2	1	3	2	4
3	1	2	3	4
4	2	3	1	4
5	1	2	3	4
6	1	3	2	4
7	2	1	4	3
8	1	2	3	4
9	3	1	2	4
10	4	1	3	2
11	1	3	2	4
12	2	3	1	4
13	1	3	2	4
14	2	3	1	4
15	3	4	1	2
16	3	2	1	4
17	3	2	1	4
18	2	3	1	4
19	1	3	4	2
20	2	3	1	4
21	3	2	1	4
22	3	2	1	4
23	1	3	2	4
24	1	4	2	3
25	4	1	3	2
26	4	1	2	3
27	1	4	2	3
28	3	1	2	4
29	2	3	1	4
30	1	3	2	4
31	4	1	3	2
32	1	3	2	4
33	3	2	1	4
34	3	1	2	4
35	2	3	4	1
36	1	3	2	4
37	3	1	2	4
38	2	3	1	4
39	2	3	1	4
40	4	2	1	3

Friedman Test for Difference in performance among different exam formats

Table 12 below shows the results of the Friedman test that was used to detect a difference in preference of exam format.

Table 12

Friedman's Test	Values
Alpha	0.05
H-stat	37.2
df	3
p-value	4.17E-08
sig	yes

H_O: There is no significant difference in preference of exam format.

H_a: There is a significant difference in preference of exam format.

Conclusion**:** Reject the null hypothesis and conclude that there is a significant difference in the preference of exam format at alpha=0.05.

Post-hoc Analysis since the Friedman test showed a significant difference among the Exam Formats

1) Multiple Choice Vs. Open Ended Formats

Table 13 below shows the Tukey-HSD test comparing multiple-choice Vs. Open-Ended Formats.

Table 13

TUKEY'S HSD / TUKEY KRAMER

					Alpha	0.05
Groups	***c***	***mean***	***n***	***ss***	***c^2/n***	***c*mean***
Multiple Choice	1	2.15	40	43.1	0.025	2.15
Open-Ended	-1	2.35	40	35.1	0.025	-2.35
Mixture-of-Both		1.95	40	33.9	0	0
Other		3.55	40	25.9	0	0
			160	138	0.05	-0.2
Q TEST						
std err	***q-stat***	***df***	***q-crit***	***lower***	***upper***	***sig***
0.148712423	-1.34488	156	3.6772	-0.74685	0.346845	no

H_O: There is no significant difference in preference of exam format.

H_a: There is a significant difference in preference of exam format between multiple choice and open-ended format examinations.

Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in the preference of exam format between multiple choice and open-ended exam formats at alpha=0.05.

2) Multiple choice Vs. Mixture of Open-Ended and Multiple Choice

Table 14 below shows the Tukey-HSD test comparing Open-ended format Vs. multiple-choice format.

Table 14

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	c	mean	n	ss	c^2/n	c*mean
Multiple Choice	1	2.15	40	43.1	0.025	2.15
Open-Ended		2.35	40	35.1	0	0
Mixture-of-Both	-1	1.95	40	33.9	0.025	-1.95
Other		3.55	40	25.9	0	0
			160	138	0.05	0.2

Q TEST

std err	q-stat	df	q-crit	lower	upper	sig
0.148712423	1.344878	156	3.6772	-0.34685	0.746845	no

H_o: There is no significant difference in preference between multiple choice and open-ended formats.

H_a: There is a higher preference of open-ended examinations as opposed to the mixture of both multiple choice and open ended.

Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in the preference of open-ended examinations as opposed to the mixture of both multiple choice and open ended examinations at alpha = 0.05.

3) Multiple Choice Vs. Other Formats

Table 15 below shows the Tukey-HSD test comparing multiple-choice vs. other formats.

Table 15

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	c	mean	n	ss	c^2/n	c*mean
Multiple Choice	1	2.15	40	43.1	0.025	2.15
Open-Ended		2.35	40	35.1	0	0
Mixture-of-Both		1.95	40	33.9	0	0
Other	-1	3.55	40	25.9	0.025	-3.55
			160	138	0.05	-1.4

Q TEST

std err	q-stat	df	q-crit	lower	upper	sig
0.148712423	-9.41414	156	3.6772	-1.94685	-0.85315	yes

H_O: There is no significant difference in preference between multiple choice and other formats.

H_a: There is a higher preference for other types of examination formats as opposed to the Multiple-choice format of examinations.

Conclusion**:** Reject the null hypothesis and conclude there is a higher preference for other types of examination formats as opposed to the multiple choice format of examinations at alpha = 0.05.

4) Open-Ended Vs. Mixture of Both Multiple Choice and Open Ended Formats

Table 16 below shows the Tukey-HSD test comparing both multiple-choice and open-ended formats.

Table 16

TUKEY'S HSD / TUKEY-KRAMER					Alpha	0.05
Groups	***c***	***mean***	***n***	***ss***	***c^2/n***	***c*mean***
Multiple Choice		2.15	40	43.1	0	0
Open-Ended	1	2.35	40	35.1	0.025	2.35
Mixture-of-Both	-1	1.95	40	33.9	0.025	-1.95
Other		3.55	40	25.9	0	0
			160	138	0.05	0.4
Q TEST						
std err	*q-stat*	*df*	*q-crit*	*lower*	*upper*	*sig*
0.148712423	2.689755	156	3.6772	-0.14685	0.946845	no

H_O: There is no significant difference in preference between multiple choice and open-ended formats.

H_a: There is a higher preference of open-ended examinations as opposed to the mixture of both multiple choice and open ended.

Conclusion**:** Fail to reject the null hypothesis and conclude that there is no significant difference in the preference of open-ended examinations as opposed to the mixture of both multiple choice and open ended examinations at alpha = 0.05

5) Open-Ended Format Vs. Other Format
Table 17 below shows the Tukey-HSD test comparing open-ended formats vs. other formats.

Table 17

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	*c*	*mean*	*n*	*ss*	*c^2/n*	*c*mean*
Multiple Choice		2.15	40	43.1	0	0
Open-Ended	1	2.35	40	35.1	0.025	2.35
Mixture-of-Both		1.95	40	33.9	0	0
Other	-1	3.55	40	25.9	0.025	-3.55
			160	138	0.05	-1.2
Q TEST						
std err	*q-stat*	*df*	*q-crit*	*lower*	*upper*	*sig*
0.148712423	-8.06927	156	3.6772	-1.74685	-0.65315	yes

H_o: There is no significant difference in preference between open-ended examination formats and other formats.

H_a: There is a higher preference for other types of examination formats as opposed to the Open-ended format of examinations.

Conclusion**:** Reject the null hypothesis and conclude there is a higher preference for other types of examination formats as opposed to the open-ended format of examinations at alpha = 0.05.

6) Mixture of both open ended and multiple-choice Vs. Other

Table 18 below shows the Tukey-HSD test comparing both open-ended and multiple choice formats with other formats.

Table 18

TUKEY'S HSD / TUKEY-KRAMER — Alpha 0.05

Groups	*c*	*mean*	*n*	*ss*	*c^2/n*	*c*mean*
Multiple Choice		2.15	40	43.1	0	0
Open-Ended		2.35	40	35.1	0	0
Mixture-of-Both	1	1.95	40	33.9	0.025	1.95
Other	-1	3.55	40	25.9	0.025	-3.55
			160	138	0.05	-1.6

Q TEST

std err	*q-stat*	*df*	*q-crit*	*lower*	*upper*	*sig*
0.148712423	-10.759	156	3.6772	-2.14685	-1.05315	yes

H_O: There is no significant difference in preference between a mixture of both open-ended and multiple-choice exam formats and other examination formats.

H_a: There is a higher preference for other types of examination formats as opposed to the mixture of both open-ended and multiple-choice formats.

Conclusion**:** Reject the null hypothesis and conclude there is a higher preference for other types of examination formats as opposed to the open-ended format of examinations at alpha = 0.05.

DISCUSSION

The data analysis used was the Friedman test for equality among two or more non-parametric variables. The Tukey-HSD test was used for analysis when the Friedman test indicated a significant difference. The Bonferroni correction was used to account for the multiple comparisons and avoid erroneously inflating the significance level.

The test distinguished performance and comprehension in classes. The two aspects of learning were tested in the different evaluation types. The results show that performance in classes with different evaluation formats is the same. The p-value of the Friedman test was .91. This is an extremely large value that undoubtedly leads to the conclusion of no significant difference at most alpha levels. This would seem to suggest that students will, on the average, perform just as well in a project-oriented class as they would an exam-oriented classroom.

In regard to subject matter comprehension, some evaluation formats are more effective than others. The Friedman test yielded a p-value of 0.000386. This leads to the rejection of the null hypothesis and a conclusion of a difference in the comprehension level among the different testing formats. This conclusion necessitated the performance of post-hoc tests to determine the exact differences between the individual evaluation procedures.

The Tukey HSD test, as regards comprehension, sheds light on the best evaluation procedure(s) to ensure maximum comprehension of subject material.

Independent study proves to be better than examinations in this regard. The test yields a significant value to prove that integrating independent study components in a classroom with examinations as the sole evaluation procedure would result in higher comprehension rates.

Projects yield a higher comprehension of material than examinations do. This is proven by the significant value that results from the Tukey-HSD test. A mixture of the formats(examinations, independent study and Projects) yields higher comprehension than do examinations when provided on their own. This information comes as no surprise seeing as how examinations scored the least when it comes to comprehension compared to all the other formats.

The question as to which of the three formats is best arises. Doing a Tukey HD test on each of the remaining possible pairs does not yield a significant difference. This leads to the conclusion that the multiple various evaluation formats are beneficial when it comes to increasing comprehension rates.

The conclusion drawn from these results is that students will perform the same regardless of evaluation format. However, when it comes to comprehension, which is arguably more important, students comprehend and retain more information from classes with independent study components, project components and a mixture of these formats compared to the sole use of examinations.

With this being said, it is impossible and unwise to completely rule out examinations when it comes to evaluating procedures. They are more time and cost

efficient. In addition, having them in the mixture of formats did prove beneficial. Because of this, it becomes imperative to investigate the type of examination format that students most preferred.

A distinction was made between multiple-choice, open-ended, a mixture of both multiple choice and open-ended and other exam evaluation formats. Other exam evaluation formats refer to selective questioning, essay questions and other formats not described by the study or researcher.

The study showed that there is no significant difference between multiple choice and open-ended questions in terms of student preference. Students felt indifferent to the two types of formats. This comes as a surprise seeing as how vastly different the two types of formats are. There was also no difference between multiple choice and a mixture of both multiple choice and open –ended exam formats.

There is a significant difference between multiple choice and other exam formats(not including open-ended or a mixture of both open-ended and multiple choice). Not surprisingly, there is no difference between open-ended and a mixture of both open-ended and multiple-choice exam formats. There is a significant difference between the open-ended formats and other formats (not including open-ended or a mixture of both open-ended and multiple choice). Finally, there is a significant difference between a mixture of both open-ended and multiple-choice formats and other(not including open-ended or a mixture of both open-ended and multiple choice) formats.

The conclusion from this portion of the study is that students seem to have no preference of the exam format as it relates to multiple choice, open-ended and a mixture of both. This indicates that students would be fine with whichever of the two, or a combination of both.

CONCLUSION

The study is eye opening in terms of what it tells educators. It clarifies issues that may have perhaps been hazy. The question over what evaluation procedure improves comprehension is now clear and educators can now incorporate the different evaluation procedures where relevant. The study leaves room for future research. Perhaps researchers could narrow down on the other formats(not including open-ended or a mixture of both open-ended and multiple choice) that students most prefer. In addition, future researchers might find the optimal mixture of evaluation procedures.

REFERENCES

http://www.huffingtonpost.com/2012/07/23/us-students-still-lag-beh_n_1695516.html

http://www.childrenofthecode.org/interviews/hanushek.htm

http://www.healthknowledge.org.uk/public-health-textbook/research-methods/1b-statistical-methods/parametric-nonparametric-tests

https://www.surveymonkey.com/mp/policy/security/

Evaluation of Test Procedures

* 1. What is your gender?

- Female
- Male

2. What is your ethnicity? (Please select all that apply.)

- American Indian or Alaskan Native
- Asian or Pacific Islander
- Black or African American
- Hispanic or Latino
- White / Caucasian
- Prefer not to answer

Other (please specify)

3. Please select your classification.

- Sophomore
- Junior
- Senior
- Other

4. What is your GPA?

- 3.6 - 4.0
- 3.1 - 3.5
- 2.6 - 3.0
- 2.1 - 2.5
- 2.0 or below

Evaluation of Test Procedures

* 1. What is your gender?

- Female
- Male

2. What is your ethnicity? (Please select all that apply.)

- American Indian or Alaskan Native
- Asian or Pacific Islander
- Black or African American
- Hispanic or Latino
- White / Caucasian
- Prefer not to answer

Other (please specify)

3. Please select your classification.

- Sophomore
- Junior
- Senior
- Other

4. What is your GPA?

- 3.6 - 4.0
- 3.1 - 3.5
- 2.6 - 3.0
- 2.1 - 2.5
- 2.0 or below

5. Please select your classification.

- Domestic Student
- International Student
- Other (please specify)

6. Are you a resident of Oklahoma?

- Yes
- No
- If not, please indicate your state of residency

Evaluation of Test Procedures

* 7. Evaluate the following statements.

	Strongly Disagree	Disagree	Kind of Agree	Agree	Strongly Agree
I understand and retain more knowledge from my Exam-Oriented classes.	○	○	○	○	○
I understand and retain more knowledge from my Independent-Study classes.	○	○	○	○	○
I understand and retain more knowledge from classes with projects	○	○	○	○	○
I understand and retain more knowledge from classes with a mixture of projects, Independent-Study and exam-oriented classes.	○	○	○	○	○
I understand and retain more knowledge from classes regardless of their evaluation format.	○	○	○	○	○

Evaluation of Test Procedures

8. Evaluate the following statements. Please refer to your transcript if necessary(4 being the highest and 1 the lowest)

- My grades in exam-oriented classes are definitely higher than in my other classes.
- My grades in Independent-Study oriented classes are definitely higher than in my other classes.
- My grades in classes with projects are definitely higher than in my other classes.
- My grades are evenly distributed among the different classes with different testing formats.

9. Rank the following exam formats (4 being the highest and 1 the lowest)

- Multiple choice questions
- Open-ended test questions
- A mixture of both multiple choice and open ended questions
- Other formats

10. Finally, Please indicate which evaluation format you would like to see more of in your classes.

- Exams
- Projects
- Independent Study
- Mixture of these formats