

A REVIEW ON THE EFFECTS OF TEST FORMATS ON ACHIEVEMENT AND RETENTION

AISHA ALHARBI

***ABSTRACT_** Primary objective of the study is to analyze the educational learning facilities and how they are providing positive impact to the student learning and development. On the other hand, different psychological test and strength identification test has been performed to understand their capabilities. The mixed research approach has been undertaken in this study which include the collection of literature review related to the topic as well as performing different tests among the students to identify the level of accuracy and capabilities and achieve an optimum outcome of the study. Randomized crossover design has been undertaken for the study. For conducting the study 38 undergraduate students in a physiology class has been undertaken and they were divided into two groups to gather the desired data through collaborative testing process. The results of the stud is quite optimistic which indicates that the strategy and capabilities for the students is directly proportional to their educational development and learning. As per recommendation it has to be understood by the educators which process of learning is effective for a group of students by analysing their psychological behavior and different testing results.*

***KEYWORDS:** Educational learning, Test Formats, retention, collaborative learning.*

A REVIEW ON THE EFFECTS OF TEST FORMATS ON ACHIEVEMENT AND RETENTION

I. INTRODUCTION

There are significant test formats has been discovered by the educators in order to facilitate the educational system which will provide better pathway for the individuals to achieve the educational goals. Moreover, there is a supreme chance of the development of retention of the study materials even more bet through the different testing procedure.

Testing, testing, and more testing is the current trend in education today. There is increased testing at every level from preschool through college. There are pretests, posttests, formative tests, summative tests, standardized tests, achievement tests, end of course tests, and the list goes on and on. All of these tests are designed to measure academic achievement, aptitude for a certain subject, teacher effectiveness, and so on and so on. Increased testing and the use of the scores to judge the effectiveness of teachers (and the future of his or her career) has shifted the focus from real student learning to testing strategies and “teaching to the test” in order to boost scores. The result may be higher scores but not increased learning.

Instead of adding another test to the list, it is time to examine the types of tests given in regards to test format, content, delivery method, frequency or scheduling within the unit of study to determine how or if the test could improve student learning. The independent variables will be the format of the test, its content, and delivery method as well as frequency and timing of administering the tests. The dependent variable to be studied is the performance of the students in regards to higher scores and retention of the tested material. The purpose of this literature review is to examine the literature that relates to testing and student achievement and answer several questions: Do different test formats such as multiple choice test, short answer tests, open-ended essay tests, and performance based tests significantly affect student learning in regards to higher scores and retention of content? Is there a difference in achievement between students using collaborative testing with the various formats mentioned above in regards to higher scores and retention? This issue is important to study because school districts are spending huge amounts of money on testing and teachers are spending huge amounts of time in the classroom “building student abilities for taking standardized tests,” yet whether or not true learning is increasing is unknown [1]. The hypothesis guiding this review is that tests could be designed and implemented in ways to actually increase learning because learning would take place during the test taking resulting in higher scores and increased retention of the tested material.

II. DEFINITION OF TERMS

- Collaborative testing: Collaborative attesting is a particular phenomenon which facilitate the collaborative capabilities within the individuals to cope up within a new environment and work effectively as a team. an assessment in which students are allowed to collaborate or

work together

- Test anxiety: Anxiety test determines the capabilities of the individuals while having a tough situation which results in rushing adrenalin and how they can handle the situation determine the level of accuracy. stress or nervousness experienced while taking a test that affects performance
- Testing effect: testing effect can be described as the potential benefits of retention the learning materials which facilitate the educational learning and test format affects retention of material
- Consolidation: consolidation is the combination of impact that has been developed after learning process and it has strengthening of what was learned because of mental review during a test
- Social loafing: It is a phenomenon which indicates an individual who is much familiar in working alone that that of a team and not preparing on an individual basis and instead relying on the work and study of peers

III. Research objectives

The fundamental objective of the study is to analyze the educational learning facilities and how they are providing positive impact to the student learning and development. On the other hand, different psychological test and strength identification test has been performed to understand their capabilities.

Significance

This study will provide positive significance towards the educators to understand the best possible learning methods and retention process for the students that can facilitate their career goals. On the other hand, the overall societal development could be ached as well as per the capabilities of individual student identification will open better future educational pathways.

RESEARCH QUESTIONS

- 1) What are the process of testing that can be performed for individual students capabilities determination?
- 2) What is the fundamental objectives of learning?
- 3) How the psychological determinants plays a crucial one in learning and retention of knowledge and concepts?

IV. METHODOLOGY

The mixed research method approach has been undertaken in this study which include the collection of literature review related to the topic as well as performing different tests among the students to identify the level of accuracy and capabilities and achieve an optimum outcome of the study. For executing the study the Randomized crossover design has been undertaken and it includes 38 undergraduate students in a physiology class has been undertaken and they were divided into two groups to gather the desired data through collaborative testing process.

PURPOSE OF THE RESEARCH

The purpose of this research is to examine the effects that various test formats have on student achievement as measured by test scores and learning as measured by long-

term retention of material. This literature review seeks to examine previous research that relates to the questions being asked and the thesis proposed. While searching for studies, gaps in the research were found and future research is needed, especially in the area of how the different testing formats affects long term retention. More research should be done in this area so students are getting the most benefit from testing especially since testing takes up so much time in today's classroom.

V. DISCUSSION

The first group of studies reviewed deal with the question of whether or not different test formats such as multiple-choice, short answer, and open-ended essay, etc. affect student learning in regards to higher scores and retention of content. The first study by Burt and Hutchinson [2] investigates the effects of case mixing (mixing lower case and upper case letters in the same word) and size mixing of letters on students' abilities to recognize the correct spelling of regular and irregular words. The purpose is to investigate how the case mixing/size mixing, which is a part of the test format, impacts performance on tests. This resource is very good because it is a primary source that gives a very clear example of how test format affects performance. It is especially good because it builds on a prior experiment by replicating it, refining it, and using item analyses to account for variables. The study used 56 university students that completed spelling test that were formatted with varying cases and sizes of letters and the overall results showed that this did have a negative impact on performance. This is important because teachers and test writers should be aware of how test formats affect student performance because the assessment of learning could be inaccurate as a result of the format.

The second study by Crisp, et al [3] examines how specific features of test formats affect the performance of students with dyslexia. Though this study deals specifically with students who have the learning disability, dyslexia, it is relevant because it considers the learning styles and their relationship to test performance and enhanced learning. Since students with dyslexia have to use special learning techniques, their test performance is easily affected by format, the results could be applicable to other students who do not have dyslexia. The study cites previous research that says that the main issues that affect test performance of students with dyslexia is the way they access and process information during an exam, slower reading and writing speeds, and the test anxiety that results from these difficulties. The study used a group of 54 dyslexic students and a control group of 51 students. The results did not show a significant difference between the two groups or between different question versions. The major finding from the study was that test anxiety can be reduced for students with dyslexia by making changes to the format of a test such as making it easier to read by using larger fonts and adequate spacing reduces test anxiety and can improve test performance. More research could show that changes like these could also help reduce test anxiety for students without a learning disability.

The third study dealing with test format is a study by Duchastel and Nungester [4] that looks specifically at which test formats enhance retention of material; this is referred to in this study and others as the "testing effect." Of particular importance is the examination of

"consolidation" (students having to retrieve lessons from memory and respond in a different format which results in a reinforcement and strengthening of what was learned). Consolidation is one of the very important factors that can make testing a learning experience and not just a tool for assessment. This process of consolidation while completing the test is actually another way to learn the material for long-term retention. The study uses 125 high-school students in the tenth grade who were divided into three groups and were tested using two different formats as well as a "filler task" for the control group. After students read a text, they were given either a short-answer test, a multiple-choice test, or no test. Two weeks later a retention test was given. In the results, both experimental groups did better on the retention test than did the control group, thus confirming the testing effect. The group that took the short-answer test had the best retention of the three groups. This is important because it shows that testing is indeed a learning tool, not just an assessment tool.

The next study by Hohensinn and Kubinger [5] examines the effects of different types of test response formats when administering a Language Awareness Test using Stem-Equivalent Items. The study looks at how the level of difficulty changes with different response formats. It is significant because the study sample is large, consisting of 2,285 students across the nine districts in Austria. Great care was taken to reduce the influence of variables such as item position (the order that an item appears on a test) which could cause a bias. Constructed response formats in which the test taker must formulate his or her answer, essay formats, and multiple choice formats are examined. All of these tests formats require different "cognitive processes." The "latent traits" of the "item stems" (the test question prompt) is examined in several different ways which allows this study to account for this variable and its effect on the examinees' responses. Furthermore, the "item response theory" (IRT) is examined and applied in interpreting the results of the study. The Rasch model is used because it defines the probability of each student's ability to correctly answer the item with regards to its difficulty.

The study found that different cognitive processes are used depending on the response format and therefore the various formats measure different latent traits. Furthermore, it was found that the response formats themselves can contribute to the difficulty of the tested item. Teachers should be aware of this when designing tests.

Leithner [6] asks the question of whether or not students have different testing styles just as they have different learning styles. It is well known from previous research that students do have different learning styles (some learn visually, orally, etc.), so they probably have differences in how they perform on different types of tests in regards to format, method of delivery or response, etc. The study aims at showing that giving all students the same kind of test only shows how well they each can take a test and not necessarily show what they really know. Previous research is cited showing that students score differently on different test styles even when the various formats cover the same material. Student perceptions of test styles and their study habits were also examined. Students least prefer essay exams and perceive that they

do better on multiple-choice exams and short-answer exams than essay exams, but the data does not support this. The students actually did better on the essay exams than on the multiple-choice exams. The reason remains unproven, but researchers theorize that students may prepare less for a multiple-choice exam than for an essay exam because they think the text will be easier. Because of this, I would like to study more and find out if students can learn to prepare in different ways so they feel more confident in taking essay exams. Also, would students actually learn and retain more if they were given essay exams simply because they would study more?

Another study that compares test formats is Mobalegh and Barati's [7] study that directly compares two tests formats, multiple true-false (MTF) and multiple-choice (MC). The main goal of the study was to see which type of test the test-takers preferred. It concluded that the multiple true-false (MTF) test was preferred over the regular multiple-choice type test. The article also addresses the issue of the strength of a multiple choice test versus an essay test. The study points out that multiple choice tests can be structured in such a way that the test taker cannot avoid difficult problems as they can in an essay. Also, the multiple choice tests have better reliability and validity. Because of scoring machines the multiple choice test can also be scored very accurately and objectively. A negative aspect of multiple choice tests is that they cannot assess partial knowledge because the answer is simply right or wrong. To address this issue researchers propose using a new version of multiple choice tests called the "liberal" multiple choice tests. In this type test the test takers can choose more than one answer if they are uncertain and receive partial credit for partial knowledge. The study used a good sample size of 200 students, but all of them were male which is a significant limitation because the effects on females are not known. In the end, if different test types result in basically the same amount of learning, then using a student preferred test type could reduce test anxiety and therefore more accurately reflected content mastery.

The next study that deals with test format effects on learning and how test context affects the learning strategies students choose to use is Samuelstuen et al's [8] study. It examines how the context of testing affects the learning strategies students use to retrieve information. This may then affect learning or retention rates. The sample size for the study is 258 students which makes this study valuable. Some of the context factors include the time of the test administration. For example, how does fatigue affect the way students answer questions at the end of a three hour tests. Factors like these influence which strategies students use, for example, the cognitive and metacognitive strategies of memorization, elaboration, organization, and monitoring. Responding differently because of context is referred to as "response bias." This affects the accuracy of the measure of mastered content. Also, the study says there is a correlation between students' reading literacy and their choice of strategy used during testing. For example, a student may do poorly on a math test because of his poor reading skills and not necessarily his math skills. Again, this reinforces the idea that test format does affect performance.

Ulusoy's [9] study aims to show how six different cloze formats measure reading levels. The reliability levels of each type of test are examined. The student reading

levels that are determined from each format is examined as well as the students' scores on the various closed test formats. Also, the students' opinion about the difficulty of each test type is examined. The sample size was 288 students. The results showed the students scored the highest on the matching cloze (MC) and the lowest on the regular cloze (RC) and dashed cloze (DC). This clearly shows a connection between test formats and scores. The different formats seem to cause students to use different strategies and focus on different cues in the text to determine the correct answers. The results of this study are important because it shows how adjusting the learning environment, including test formats, can be used to meet the needs of individual students.

The final study that deals with test format is a study by van de Watering et al. [10]. This primary study is shows a difference between what students perceive and reality. According to this study and the survey used students said that multiple choice tests are easier than essay tests, but they actually scored lower on the multiple choice tests. The students perceive that written assessments such as multiple-choice are easy to prepare for and reduce stress or test anxiety. The fact that they prefer this type of test but score lower on is attributed to the idea that perhaps the students are more accustomed to this format and feel comfortable with it. Further research could examine if situations like this result in less learning, perhaps by lower scores and lower retention rates of information. The cognitive processes used are examined such as assessing the application of knowledge, problem solving, drawing conclusions, analyzing, and interpreting versus assessing simple reproduction of knowledge. For my research this may be applicable for the "testing effect" examined in other studies.

The next group of studies deals with the test format of collaborative testing. It has been well established that collaborative learning benefits students. Can collaborative testing be beneficial extension of collaborative learning?

The first study to deal with collaborative learning is Cortright et al's [11] study that investigates whether or not collaborative-group testing (testing in which students work in groups to complete the assessment) improves retention. It is important because several studies have shown that collaborative testing improves performance (raises the scores for the test takers), but they do not show the effects on retention of the tested material. The study uses 38 undergraduate students in a physiology class, divided into two groups. There was a randomized crossover design to control for the variables of time and order of exam completions. Using standard individualized testing the students' retention of the material after one month went down significantly from 63.5% accuracy to 46% accuracy. When collaborative testing was used the percentage of correct responses was 81.3% on the first exam and 52.9% on the repeated exam. The study concludes that collaborative testing did improve both test performance and retention. This is important because collaborative activities are being used more in the classroom, so there needs to be evidence that the collaborative activities actually result in facilitating learning. This study provides real evidence on retention and not just performance as most prior studies have done.

Cranney et al. [12] examines how the act of taking a test affects retention of tested material. Specifically how

collaborative testing helps students perform better the first time an assessment is given to the groups as well as when those same students worked individually when the test was repeated. First year university psychology students participated in the study. Students who first took the test in collaborative groups do better the second time as individuals than students who took the first test as individuals and the second test as individuals. It shows that repeated testing increases retention more than repeated or increased studying. The findings are very important to understand because in education today there seems to be increased testing every year with more and more classroom time devoted to it. Also, more collaborative activities are being used in the classroom. Since this is the case, it is good to know that testing itself can enhance learning. More research should be done to find out more about which formats of collaborative testing result in the most student learning and retention of content material.

The next study to deal with collaborative testing is Kapitanoff's [13] study that is very specific about the various aspects of collaborative testing. According to the study, tests can be an active form of learning due to peer mediated learning. Also, collaborative testing results in higher test scores but the research was done in order to determine why. It is proposed that three things cause the higher scores: cognitive processes, interpersonal interactions, and reduced test anxiety. There is a lot of focus on the relation between reduced test anxiety and improved test scores. The possible negative effects such as "social loafing" and grade inflation are examined. Various ways of ensuring individual responsibility and prior test preparation to prevent "free-riders" are addressed. Collaborative groups consisted of only two students and their grades were based on a combination of an individual test and collaborative test to increase individual responsibility. A limitation of the study is that it is a small one with only thirty-three students and four exams comprised of fifty multiple choice and fifty short-answer questions. The study did provide educators with strategies to increase individual responsibility and learning.

The next study by Keselyak et al. [14] examines students' and faculty members' perception of collaborative testing in a dental hygiene course. Individual and group assessment performances were compared using a paired samples t-test. Students viewed the collaborative testing as positive saying that it enhanced their learning, allowed them to defend their answers, and decreased stress. Faculty and students had the same concerns as seen in other studies of individual accountability. More ways to increase individual preparation and accountability are needed. But, most significant in this study is that the collaborative testing is regarded as an actual learning strategy and not just an assessment tool. A goal of the study is to provide instructors with evidence that teaching techniques other than traditional lecturing can be very effective if not more effective in helping students learn. This study talks about tests being "educative assessments" which means that students will actually learn while taking the assessments. With all of the classroom time used for testing today, it is important to make the best use of the testing time by helping students learn and retain more while they are being assessed.

Leight et al.'s [15] study examines the effects of collaborative testing. The study reflects on previous

studies that show that collaborative testing does improve test scores, but says that more research is needed to see whether or not it affects content retention. The research is done at a university in a biology class. At this university it is common for 200 or more students to be enrolled in a class. This condition usually causes the instructor to use multiple choice tests that can be graded quickly by computer. The authors say that this format does not allow the student to get feedback from their exams, so they cannot learn from their mistakes. Testing can be a learning strategy as well as an assessment strategy if students can get feedback. Also, the "testing effect" is examined showing how testing can actually be a study strategy because of the repeated recall efforts. Also, the "negative testing effect," the recalling of incorrect responses is also examined. Furthermore, group testing is examined as a way of collaborative learning because students can analyze and discuss exam questions after they have completed the exam as individuals, thus getting valuable feedback.

The final study in this group is Slusser and Erickson's [16] study that aims to show that group quizzes are not only an assessment tool, but also a learning strategy. Learning can take place during testing, but how much may depend on format and implementation. Group learning has already been proven effective, now group testing is being studied. This study shows that group or collaborative testing improves the behavior and attitude of students towards the course. The study shows that's students learn more from collaborative groups than the traditional lecture method and are trying to see if the same is true for collaborative tests. According to this study there is research that proves that students have improved learning, better retention of the tested material, and reduced test anxiety. Another benefit proposed by the study is that students will come to class well-prepared and will have read the assigned reading when they know there will be collaborative testing.

VI. CONCLUSION

While there is some research about test formats and their effects on achievement and long-term retention of material, there is not nearly enough. With so much time in the classroom being used for testing, it is important that educators understand the effects of test formats and other concepts like test anxiety if they want to get a true measure of content mastery. One test format that especially needs more study is collaborative testing. The research is mixed about the benefits of this type of testing. The debate is whether or not retention is increased along with achievement, or higher test scores. Another question that needs to be examined more extensively is the problem of "social loafing." If a student receives a higher test score only because he or she is able to work in collaboration with other students who studied more, then his true learning may not be increased. More research should be done to help educators improve their testing content and procedures. Furthermore, with all the focus on standardized testing today, test writers need research to help guide them in creating test that most fairly assess student mastery of content.

REFERENCES

- [1] Ediger, M., & Bhaskara Rao, D. (2001). Teaching science successfully (1st ed.). New Delhi: Discovery Pub. House.

- [2] Burt, J. S., & Hutchinson, B. J. (2000). Case-mixing effects on spelling recognition: The importance of test format. *Journal of Psycholinguistic Research*, 29(4), 433-451.
- [3] Crisp, V., Johnson, M., & Novakovic, N. (2012). The effects of features of examination questions on the performance of students with dyslexia. *British Educational Research Journal*, 38(5), 813-839.
- [4] Duchastel, P. C., & Nungester, R. J. (1982). Testing effects measured with alternate test forms. *The Journal of Educational Research*, (5), 309.
- [5] Hohensinn, C., & Kubinger, K. D. (2011). Applying item response theory methods to examine the impact of different response formats. *Educational and Psychological Measurement*, 71(4), 732-746.
- [6] Leithner, A. (2011). Do student learning styles translate to different "testing styles"?. *Journal of Political Science Education*, 7(4), 416-433.
- [7] Mobalegh, A., & Barati, H. (2012). Multiple true-false (MTF) and multiple-choice (MC) test formats: a comparison between two versions of the same test paper of Iranian NUEE. *Journal of Language Teaching And Research*, (5), 1027.
- [8] Samuelstuen, M. S., Bråten, I., & Valås, H. (2007). Context effects in Norwegian 10th-grade students' reports on learning strategies using the cross-curricular competencies instrument. *Scandinavian Journal of Educational Research*, 51(5), 511-529.
- [9] Ulusoy, M. (2008). Evaluating sixth graders' reading levels with different cloze test formats. *Eurasian Journal of Educational Research (EJER)*, (31), 115-133.
- [10] van de Watering, G., Gijbels, D., Dochy, F., & van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education: The International Journal Of Higher Education And Educational Planning*, 56(6), 645-658.
- [11] Cortright, R., Collins, H., Rodenbaugh, D., & DiCarlo, S. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education*, 27(3), 102-108.
- [12] Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, 21(6), 919-940.
- [13] Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education*, 10(1), 56-70.
- [14] Keselyak, N., Saylor, C., Simmer-Beck, M., & Bray, K. (2009). Examining the role of collaborative assessment in a didactic dental hygiene course. *Journal of Dental Education*, 73(8), 980-990.
- [15] Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE - Life Sciences Education*, 11(4), 392-401.
- [16] Slusser, S. R., & Erickson, R. J. (2006). Group quizzes: an extension of the collaborative learning process. *Teaching Sociology*, (3), 249.