Comparison of the effect of post-instruction multiple-choice and short-answer tests on delayed retention learning

Ramraje SN[1], Sable PL[2]

1. Department of Pathology, Grant Medical College, Mumbai, India

2. Department of Preventive and Social Medicine, G.S. Medical College, Mumbai, India

RESEARCH

Corresponding Author:
Dr. Sushma N. Ramraje,
2/15, Dhanvantari, Doctors Quarters,
J.J. Hospital Campus, Byculla,
Mumbai-400008
Email: sushmaramraje@yahoo.com

Abstract

Background

People forget much of what they learn, therefore students could benefit from learning strategies that yield long-lasting knowledge. Yet surprisingly, little is known about how long-term retention is most efficiently mastered. We studied the value of teacher made in class tests as learning aids and compared two types of teacher-made tests (multiple choice and short-answer tests) with a no test (control) to determine their value as aids to retention learning.

Method

The study was conducted on two separate batches of medical undergraduate students. This study compared two types of tests [multiple choice questions (MCQs) and short answer questions (SAQs)] with a no test (control) group. The investigation involved initial testing at the end of the lecture (post instruction), followed by an unannounced delayed retention test on the same material three weeks later. The unannounced delayed test comprising of MCQs and SAQs on the same material was given three weeks later to all the three groups.

Results

In batch I, the MCQ group had a higher mean delayed retention score of 10.97, followed by the SAQ group (8.42) and the control group (6.71). Analysis of variance (ANOVA) test and least significance difference (LSD) post hoc test revealed statistically significant difference between the means of the three groups. Similar results were obtained for batch II

Conclusion

Classroom testing has a positive effect on retention learning; both short-answer and multiple-choice tests being more effective than no test in promoting delayed retention learning, however, multiple-choice tests are better.

Key Words

Initial testing, delayed retention tests, retention learning

What this study adds:

1. Much discussion is already ongoing in this field. There are plenty of reliable references. Although studies have been published analysing knowledge retention in other fields of study, a literature search failed to yield any published research on this topic in the field of medical education.

2. Medical school curriculum is vast and information taught is difficult to retain in future years or in clinical practice. Classroom testing will definitely help in retention of information. Our data will add to the knowledge pool.

Background

Cognitive learning is best assessed with traditional classroom tests and it is important to maximise the learning value of the time spent in testing regardless of the type of tests used. Testing assesses what students have learned and also improves long-term memory. The act of taking a test helps move information from the short-term memory to a deeper level. Whether this effect is caused by the mere fact that taking a test provides one additional opportunity for rehearsal or there is some unknown factor (such as the kinesthetic act of writing the answers) at work has not been determined yet. Students remember information which has been tested more, than non-tested information, even without additional study of the to-be-tested material hence the need for undertaking such a study.

Method

Aims and objectives

1. To discover if initial testing post instruction helps retention learning.

2. To improve delayed retention learning by introducing post instruction multiple assessment systems like MCQs and SAQs.

3. To compare the effect of post instruction MCQ and SAQ tests on delayed retention learning.

Ethics committee approval was taken.

Definition of Terms:

 "Initial testing" means testing which occurs at the time of instruction or immediately thereafter.

"Delayed retention tests" are research instruments which are administered two or more weeks after instruction and initial testing to measure retained knowledge.[1–3]

"Retention learning" means learning which lasts beyond the initial testing and it is assessed with tests administered two or more weeks after the information has been taught and tested.[4]

A delay period of three weeks was used in this study.

Design

1. Comparative study.
2. Target group: Two batches of students of Second MBBS (Pathology) class
3. Sample size: Batch I (n1 =84), divided into three groups:

   a)   multiple-choice test
        (Group A,   n=35);
   b)   short-answer test (Group B, n=35);
         and
   c)   no test (Control, Group C, n=14).

Batch II (n_2 =72), divided into three groups:

   a)   multiple-choice test (Group A, n=30);

   b)   short-answer test (Group B, n=30); and

   c)   no test (Control, Group C, n=12).

Total sample size= $n_{1 +} n_2$ =156. The study was conducted on two separate batches of medical undergraduate students.

This study compared two types of tests (multiple choice and short-answer) with a no test (control) group.

At the beginning of the lecture it was announced that an experimental study was being conducted. Participation was voluntary and participants should give their best.

All instructional and testing procedures were done in the same room, by the same teacher thus helping to control environment variables.

Process

The investigation involved initial testing at end of lecture (post instruction), followed by an unannounced delayed retention test on the same material three weeks later. The multiple-choice tests were a 10 item test and had four response alternatives. The short answer versions were

identical to the multiple-choice tests, however there were no alternatives and brief answers were required. The same information was reflected in both tests.

An initial test of MCQs and SAQs was given at the end of lecture (post instruction) to the MCQ and SAQ group respectively. The control group was not given any test initially.

An unannounced delayed test comprising of MCQs and SAQs on the same material was given three weeks later to all the three groups.

Results

Only the delayed test scores were statistically analysed. Statistical analysis was done using SSPS (Version 17). Mean test scores for all the groups were calculated. Statistical difference between different groups was calculated using ANOVA and post hoc comparisons.

The total students (n=84) participating in the project were divided into three groups: MCQ, SAQ and control group. The MCQ group (n=35) had a higher mean delayed retention score of 10.97, followed by the SAQ group (8.42) and the control group (6.71). The total mean retention test score was 9.20 (n=84) (see Table 1).

The GLM procedure of ANOVA was used to discover the difference between the three groups. It showed a statistically significant difference between the means of delayed retention test scores of the MCQ, SAQ and control group (p < 0.05) (see Table 2).

The LSD post hoc test (post comparison method) was used to calculate the difference between the means of the three groups. Statistical analysis revealed significant difference between the groups (p value < 0.05). Hence, the mean score of the MCQ group is higher than the SAQ group as well as the control group (see Table 3).

The total students (n=72) participating in the project were divided into three groups: MCQ, SAQ and control group. The MCQ group (n=30) had a higher mean delayed retention score of 10.20, followed by the SAQ group (8.16) and the control group (5.66). The total mean retention test score was 8.59 (n=72) (see Table 4).

The GLM procedure of ANOVA was used to discover the difference between the three groups. It showed a statistically significant difference between the means of delayed retention test scores of the MCQ, SAQ and control group (p < 0.05) (see Table 5).

LSD post hoc test (post comparison method) was used to calculate the difference between the means of the three groups. Statistical analysis revealed significant difference between the groups (p value < 0.05). Hence, the mean score of the MCQ group is higher than the SAQ group as well as the control group (see Table 6).

In batch I, the MCQ group had a higher mean delayed retention score of 10.97, followed by the SAQ group (8.42) and the control group (6.71). ANOVA test and LSD post hoc test revealed a statistically significant difference between the means of the three groups. Similar results were obtained for batch II (see Figure 1).
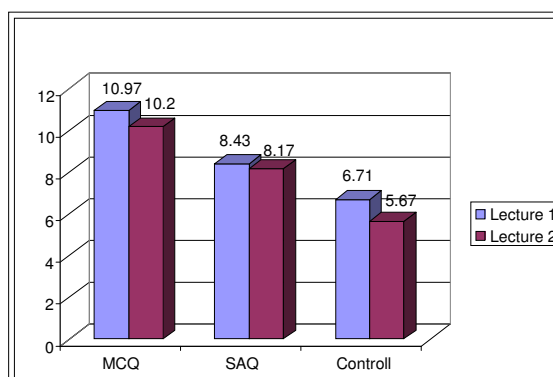

Figure 1: Mean delayed retention test scores

Discussion

Testing in education i.e. time devoted to testing and the effects of testing are becoming very important. Testing assesses students' knowledge; also, it improves long-term memory. This retention of course-related information resulting from test-taking is called the 'testing effect'.[5] The testing effect has been studied in laboratory and classroom settings, dating back to the early 20th century[6] and continuing to the present day.[7,8]

Much research on testing has been done in recent years and has been focused on standardised tests.[9] In schools, teacher-made tests are used for evaluation. [3,10,11,12] Data is available on the quality of these tests creating doubt as to whether teachers are able to perform evaluation effectively or not. [11,13,14,15] But teacher-made tests are important in the classroom for evaluation of taught material.[12] Teacher-made tests are an important part of the educational system and should be researched. [3, 11,12]

Evaluation by teacher-made tests in schools is an important and needed part of the educational system and a crucial area for research. [11,12,16,17]

The use of short-answer tests is popular. Short-answer tests are used more often and are more effective with lower level types of learning.[10] They have certain qualities, such as they are relatively easy to prepare[10] and scored more quickly. However, they are not objective like multiple-choice tests and cannot adequately test students who have read the subject well. Despite this inadequacy, they are useful because many teachers cannot make good multiple-choice items.[10,11]

Multiple-choice tests promote retention.[2,3] Literature has shown that only test-taking enhances retention learning. The announcement of an upcoming test did not have a positive effect on retention learning. Hence increased studying for a slated test did not result in better retention, only the act of taking the test did.[18]

The first of the three objectives of this study was: to discover if initial testing post instruction helps retention learning. The testing of instructional material did help in retention learning. This finding has been very consistently observed among several studies.[2,3,4,18] Two types of tests were shown to be effective in supporting retention learning. Another objective of this study was: does introducing post instruction multiple assessment systems like MCQs and SAQs aid retention learning? It was noted that testing of instructional material did promote retention learning. Whether the act of test writing helped in retention learning or the knowledge of an upcoming test motivated students to study better is still a matter of controversy. Haynie put forth his findings in which announcements of the intention to test did not promote retention learning unless actual tests or reviews were taken.[18]

Another objective was to compare the effect of post instruction MCQ and SAQ tests on delayed retention learning. It was found that short-answer and multiple-choice tests were both effective in the promotion of retention. Multiple-choice tests promote retention learning better. Both Group A and B scored significantly higher than the control (no test) group. However, multiple-choice tests appeared to be more effective in promoting retention learning than short-answer tests as shown by the finding of significantly higher scores for Group A, the reason being the presence of the correct answer along with the distractors in the multiple-choice items, which was not the case with short-answer test items. Similar findings were noted by Haynie.[4]

Conclusion

Examinations and assignments are the two most commonly used approaches to assessment in education.

Since testing consumes a large amount of students' and teachers' time, the evaluation process should be made more efficient.

Classroom testing has a positive effect on retention learning; both short-answer and multiple-choice tests being more effective than no test in promoting delayed retention learning, however, multiple-choice tests are better.

Multiple-choice tests are good in promoting retention learning of the information actually contained in the immediate post-tests as compared to the short-answer tests because the correct answer to each item is provided along with the distracters in the multiple choice items. The advantage of short-answer tests is that students do not have to choose from a set of responses. Teachers who find constructing multiple-choice items difficult could make do with short-answer items which are relatively easy to prepare. Their usefulness in the promotion of retention learning should be researched. It is recommended to choose the type of test which meets the educational objectives best. Continued research must be conducted and the best ways to test sought so as to maximise retention of important learning in all disciplines.

References

1. Duchastel P. Retention of prose following testing with different types of test. Contemporary Educational Psychology 1981; 6: 217-226.

2. Nungester R.J, Duchastel PC. Testing versus review: Effects on retention. Journal of Educational Psychology 1982; 74: 18-22.

3. Haynie WJ. Effects of take-home and in-class tests on delayed retention learning acquired via individualized, self-paced instructional texts. Journal of Industrial Teacher Education 1991; 28:52-63.

4. Haynie WJ. Effects of multiple-choice and short-answer tests on delayed retention learning. Journal of Technology Education 1994; 6: 32-44.

5. Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science 2006; 17:249 – 255.

6. Gates A. Recitation as a factor in memorizing. In: Woodworth RS, editor. Archives of Psychology. No.40. New York: The Science Press; 1917: 1-104.

7. Carpenter S, Pashler H, Vul E. What types of learning are enhanced by a cued recall test? Psychonomic Bulletin & Review 2006; 13:826-830.

8. Karpicke J, Roediger H. Repeated retrieval during learning is the key to long-term retention. Journal of Memory and Language 2007; 57:151-162.

9. Stiggins R.J,Conklin NF, Bridgeford NJ. Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice 1986; 5:5-17.

10. Haynie WJ. Student evaluation: The teacher's most difficult job. Monograph Series of the Virginia Industrial Arts Teacher Education Council, Monograph no. 11. 1983.

11. Haynie WJ.Post hoc analysis of test items written by technology education teachers. Journal of Technology Education 1992; 4:26-38.

12. Mehrens WA, Lehmann IJ. Using teacher-made measurement devices. NASSP Bulletin 1987; 71:36-44.

13. Carter K. Do teachers understand the principles for writing tests? Journal of Teacher Education 1984; 35:57-60.

14. Gullickson AR, Ellwein MC. Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice 1985; 4:15-18.

15. Stiggins RJ, Bridgeford NJ.The ecology of classroom assessment. Journal of Educational Measurement 1985;22: 271-286.

16. Ellsworth RA, Dunnell P, Duell OK. Multiple choice test items: What are textbook authors telling teachers? Journal of Educational Research 1990; 83:289-293.

17. Nitko AJ. Designing tests that are integrated with instruction. In: Linn RL, ed. Educational measurement.3rd ed .New York:Macmillan 1989:447-474.

18. Haynie WJ. Effects of tests and anticipation of tests on learning via videotaped materials. Journal of Industrial Teacher Education 1990; 27: 18- 30.

PEER REVIEW
Not commissioned. Externally peer reviewed

CONFLICTS OF INTEREST
The authors declare that they have no competing interests

ETHICS COMMITTEE APPROVAL
Institutional Ethics Committee,
Dept of Pharmacology,
Grant Medical College, Mumbai, Maharashtra, India.

Tables

Table 1: Means, Standard deviations and sample sizes.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	Upper Bound	Minimum	Maximum
MCQ	35	10.9714	1.12422	.19003	10.5852	11.3576	9.00	13.00
SAQ	35	8.4286	1.06511	.18004	8.0627	8.7944	6.00	12.00
Control	14	6.7143	1.13873	.30434	6.0568	7.3718	5.00	8.00
Total	84	9.2024	1.94985	.21275	8.7792	9.6255	5.00	13.00

Table 2: GLM procedure of Analysis of Variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig. (p value)
Between Groups	217.160	2	108.580	89.380	.000
Within Groups	98.400	81	1.215		
Total	315.560	83			

Table 3: LSD multiple comparisons

(I) Test category	(J) Test category	Mean Difference (I-J)	Std. Error	Sig. (p value)	95% Confidence Interval Lower Bound	Upper Bound
MCQ	SAQ	2.54286*	.26347	.000	2.0186	3.0671
	Control	4.25714*	.34854	.000	3.5637	4.9506
SAQ	MCQ	-2.54286*	.26347	.000	-3.0671	-2.0186
	Control	1.71429*	.34854	.000	1.0208	2.4078
Control	MCQ	-4.25714*	.34854	.000	-4.9506	-3.5637
	SAQ	-1.71429*	.34854	.000	-2.4078	-1.0208

*. The mean difference is significant at the 0.05 level.

Table 4: Means, Standard deviations and sample sizes.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
MCQ	30	10.2000	.99655	.18194	9.8279	10.5721	9.00	12.00
SAQ	30	8.1667	.91287	.16667	7.8258	8.5075	7.00	10.00
Control	12	5.6667	.88763	.25624	5.1027	6.2306	4.00	7.00
Total	72	8.5972	1.86638	.21995	8.1586	9.0358	4.00	12.00

Table 5: GLM procedure of Analysis of Variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig. (p value)
Between Groups	185.686	2	92.843	103.940	.000
Within Groups	61.633	69	.893		
Total	247.319	71			

Table 6: LSD multiple comparisons

(I) Test category	(J) Test category	Mean Difference (I-J)	Std. Error	Sig. (p value)	95% Confidence Interval	
					Lower Bound	Upper Bound
MCQ	SAQ	2.03333*	.24403	.000	1.5465	2.5202
	Control	4.53333*	.32282	.000	3.8893	5.1773
SAQ	MCQ	-2.03333*	.24403	.000	-2.5202	-1.5465
	Control	2.50000*	.32282	.000	1.8560	3.1440
Control	MCQ	-4.53333*	.32282	.000	-5.1773	-3.8893
	SAQ	-2.50000*	.32282	.000	-3.1440	-1.8560

*. The mean difference is significant at the 0.05 level.