Effect of Two Retrieval Practices on Anatomy Scores among MBBS Students

Amber Ilyas[1], Syed Moin Ali[2], Urooj Fatima[1], Fouzia Kirmani[3] and Sadia Iqbal[1]

[1]*Department of Anatomy, Jinnah Sindh Medical University, Karachi, Pakistan*
[2]*Department of Educational Development, Aga Khan University Hospital, Karachi, Pakistan*
[3]*Directorate of Medical Education, College of Physicians and Surgeons Pakistan, Karachi, Pakistan*

ABSTRACT

Objective: To compare anatomy scores following retrieval practice (RP) using multiple-choice questions (MCQs) and short-answer questions (SAQs) in the head and neck and endocrinology modules.

Study Design: A quasi-experimental study.

Place and Duration of the Study: Department of Anatomy, Jinnah Sindh Medical University, Karachi, Pakistan, from April 2023 to February 2024.

Methodology: A total of 350 second-year MBBS students were equally divided into two groups: Group A (all the odd roll numbers) and Group B (all the even roll numbers). Group A practiced retrieval using SAQs and Group B using MCQs in the first module and vice versa in the second module. RP exercises were administered by the principal investigator at the end of lectures, followed by immediate feedback. At the end of each module, a formative assessment of 25 MCQs was administered to assess student performance. Data were entered and analysed using SPSS version 20. An independent samples t-test was used to compare scores between the groups within each module, whereas a paired samples t-test was employed to compare MCQ- and SAQ-based performance across the two modules.

Results: Students engaged in RP using SAQs achieved significantly higher scores (mean = 81.734 ± 17.444) than those who practiced with MCQs (mean = 70.978 ± 17.027, p = 0.024). The student deemed the RP exercises to be beneficial.

Conclusion: RP is beneficial for improving academic performance. SAQs were more effective than MCQs in promoting retention.

Key Words: *Retrieval practice, Multiple-choice questions, Short-answer questions, Learning, Memory.*

How to cite this article: Ilyas A, Ali SM, Fatima U, Kirmani F, Iqbal S. Effect of Two Retrieval Practices on Anatomy Scores among MBBS Students. *J Coll Physicians Surg Pak* 2026; **36(07)**:908-911.

INTRODUCTION

Retrieval is the reconstructive process of accessing information from long-term memory.[1] Long-term retention is essential for building knowledge and effective problem-solving,[2] yet instructors often struggle to help students retain large volumes of content.[3] Several learning strategies support retention.[4] These include retrieval practice (RP) and distributed practice.[5] RP involves recalling facts or concepts to strengthen learning.[1] Students can be tested through various methods.[6] RP improves understanding and memory by promoting active recall and application of knowledge.

Various RP tools are described in the literature, including short-answer questions (SAQs) and multiple-choice questions (MCQs),[6-8] often paired with feedback.[9] Using multiple formats is beneficial, as it allows students to access long-term memory through different retrieval pathways.[10]

Correspondence to: Dr. Amber Ilyas, Department of Anatomy, Jinnah Sindh Medical University, Karachi, Pakistan
E-mail: dr.amberilyas@gmail.com
Received: December 01, 2025; Revised: April 06, 2026; Accepted: April 20, 2026
DOI: https://doi.org/10.29271/jcpsp.2026.07.908

This research is based on Atkinson-Shiffrin's information processing model,[11] which proposes that memory comprises sensory, working, and long-term stores.[12]

Information moves through these stages, and retrieval helps bring information back into working memory, which in turn supports long-term retention. RP is considered more effective than re-reading.[13,14] However, evidence is mixed on whether SAQs or MCQs better enhance retrieval. This study, therefore, examines the impact of using SAQs versus MCQs during RP on student performance.

METHODOLOGY

This quasi-experimental study was conducted at the Department of Anatomy, Jinnah Sindh Medical University, Karachi, Pakistan, from April 2023 to February 2024, using a universal sampling technique. A total of 350 second-year MBBS students who fulfilled the inclusion criteria were enrolled in the study. The students who attended all sessions in which RP exercises were conducted and submitted their response sheets were considered eligible for the study. Students who failed to complete all RP exercises or did not provide voluntary consent were excluded. Ethical approval for the study was obtained from the Institutional Review Board of the hospital. The principal investigator (PI) explained the study rationale and procedures to the students before obtaining written informed consent.

Students were encouraged to participate in RP exercises for their academic benefit and were assured of confidentiality, voluntary participation, and the right to withdraw at any time without consequences. Codes were assigned to the roll numbers to maintain anonymity, and all codes were securely maintained by the PI.

The intervention was conducted during anatomy lectures in two second-year MBBS modules: head and neck (H&N) and endocrinology. The H&N module was delivered over 7 weeks, whereas the Endocrinology module spanned 5 weeks. Written permission to conduct the RP intervention was obtained from the Head of the Department of Anatomy before the start of the study.

Two groups of students were formed based on their roll numbers. Group A comprised odd roll numbers, and Group B comprised even roll numbers. In the first module, RP exercises were given as SAQs to Group A and MCQs to Group B, with the formats reversed in the second module. These RP exercises were only implemented in lectures delivered by the PI during the last 15 minutes of class.

In the H&N module, the PI distributed RP questions to students and gave them 5 minutes to respond. Group A attempted 3 SAQs, and Group B attempted 5 MCQs, as SAQs required more time. Students entered their roll numbers on response sheets for data entry, to which codes were assigned and stored in a password-protected file accessible only to the PI. After collecting responses, the PI provided the corrective feedback and clarified the concepts. The PI also developed a 25-item MCQ-based final RP exercise, which was reviewed by a senior professor to establish content validity. This was administered in class at the end of the module, with immediate formative feedback provided to reinforce learning.

In the endocrinology module, the RP formats were switched: Group A attempted 5 MCQs and Group B 3 SAQs. Written consent was obtained again, and the same procedure used in the H&N module was followed. A 25-MCQ final RP exercise was administered at the end of the module using the same protocol. Only 57 students in H&N, 123 in endocrinology, and 36 students common to both modules completed all RP exercises and were included in the analysis. Data were entered in SPSS version 20. Descriptive statistics were calculated, and the normality of the data was assessed using the Shapiro-Wilk test. Hypotheses were tested using independent t-tests, while paired t-tests were applied to compare mean percentage scores of MCQs and SAQs in both modules. A p-value <0.05 was considered significant.

RESULTS

Since 57 students in the H&N module had completed the RP exercise and 123 students in endocrinology, the resultant sample size was 180 (51.4%) of 350 students.

Table I: Descriptive statistics of H&N and endocrine modules.

Descriptive Statistics	% SAQ scores		% MCQ scores	
Modules and student numbers (n)	**H&N (n = 23)**	**Endocrinology (n = 73)**	**H&N (n = 34)**	**Endocrinology (n = 50)**
Mean	60.16	88.23	56.77	86.66
Mode	65.11	91.66	58.13	91.66
SD	15.58	13.73	16.29	12.93
Minimum	30.23	25	25.58	58.33
Maximum	90.69	100	86.04	100

SD: Standard deviation.

Table II: Hypothesis testing of H&N and endocrine module (independent samples test).

Measures	Assumptions	Levene's test for equality of variances		t-test for equality of means		
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**
% scores in the H&N test	Equal variances assumed	0.07	0.781	-0.784	55	0.436
% scores in the endocrine test	Equal variances assumed	0.044	0.834	-0.640	121	0.523

A p-value was calculated using the independent sample t-test.

Table III: Comparison of combined mean scores for SAQs and MCQs.

		Paired differences					t	df	Sig (2-tailed)
		Mean	**SD**	**Std. error mean**	**95% confidence interval of the difference**				
					Lower	**Upper**			
Pair 1	MCQ SAQ	-10.755	27.365	4.560	-20.015	-1.496	-2.358	35	0.024

A p-value was calculated using a paired samples t-test. Note: p <0.05.

Table I compares the descriptive statistics obtained in the two modules and in the two RP exercise methods. The difference in scores between the 2 modules is evident. It is interesting to note that the mean scores for SAQs (60.16 ± 15.58 in the H&N module and 88.23 ± 13.73 in the Endocrine module) are higher than the mean scores for MCQs (56.77 ± 16.29 in the H&N and 86.66 ± 12.93 in the Endocrine) in both modules.

When the results of SAQs and MCQs were analysed for significant differences by independent sample t-tests, no significant result was obtained as the p-value was >0.05 for each module (Table II). On this basis, the null hypothesis was rejected, and it was concluded that the observed differences in scores were by chance.

A paired sample t-test was applied for comparison between the combined mean percentage scores of MCQS and SAQS for both modules (Table III).

Table III reveals a significant difference in the performance of students who answered MCQs and those who answered SAQs in both modules. Students performed better on SAQs compared to MCQs. The consistency of higher scores in SAQs is supported by the confidence interval, which shows that the true difference in mean scores favours SAQs. Additionally, the statistical significance (0.024) of the observed difference, as indicated by the significance level (Sig. 2-tailed) is 0.024, confirms that this difference is not due to random variation but reflects a genuine difference in student performance between the two types of questions.

DISCUSSION

The study showed no significant difference between SAQs and MCQs within each module; however, a significant difference emerged when scores from both modules were combined.

Sundararajan *et al*. reported that all forms of RP, regardless of test type, improved performance,[6] a finding also supported by Chan *et al*., who noted that test format does not significantly alter the testing effect.[15] The slight advantage of SAQs in this study may relate to the active recall and detailed responses they require, which promote deeper learning, as noted by Greving.[16] Overall, the effectiveness of both SAQs and MCQs aligns with prior evidence that RP enhances learning and retention more than re-reading.[11]

Stenlund *et al*. found that both SAQs and MCQs tests improve retention more than re-reading, with SAQs showing higher initial scores and greater long-term retention, especially on delayed assessments.[17] McDaniel *et al*. similarly reported that SAQs offer greater benefits than MCQs.[18]

This study has several limitations. It was conducted at a single public institution; its findings have limited generalisability. The quasi-experimental design without randomisation may introduce bias and limit causal inference. It primarily assesses anatomical knowledge and short-term outcomes, which may not reflect broader learning or long-term retention. Factors such as teaching methods, student engagement, and study habits were not analysed. Limited seating and spacing during assessments may have allowed academic misconduct, potentially affecting validity.

CONCLUSION

This study highlights the impact of two RP exercises and the importance of evaluating RP methods in medical education, emphasising the benefits of diverse learning strategies in anatomy. These findings are consistent with the retrieval effort and transfer-appropriate processing hypotheses. SAQs demonstrated superior retention, likely owing to effortful retrieval, while the observed cross-format transfer suggests that the RP format can influence performance on different final assessments.

ETHICAL APPROVAL:
Ethical approval was obtained from the Institutional Review Board of Jinnah Sindh Medical University, Karachi, Pakistan (Ref. No. JSMU/IRB/2023/700; dated: 18 March 2023).

PATIENTS' CONSENT:
Written informed consent was obtained from all the study participants.

COMPETING INTEREST:
The authors declared no conflict of interest.

AUTHORS' CONTRIBUTIONS:
AI: Conception, design, acquisition, data analysis, and writing of manuscript.
SMA: Mentorship, study design, manuscript review, and intellectual input.
UF: Content review, data interpretation, and manuscript editing.
FK: Medical education input, framework refinement, and manuscript review.
SI: Anatomical expertise, study design, and manuscript review.
All authors approved the final version of the manuscript to be published.

REFERENCES

1. Bruning RH, Schraw GJ, Ronning RR. Cognitive Psychology and Instruction. Prentice-Hall, Inc., One Lake Street, Upper Saddle River, New Jersey; 1999. pp. 4-10. Available from: https://www.scirp.org/reference/referencespapers?referenceid=971482.
2. Dobson JL, Perez J, Linderholm T. Distributed retrieval practice promotes superior recall of anatomy information. *Anat Sci Educ* 2017; **10(4)**:339-47. doi: 10.1002/ase.1668.
3. Shahbaz A, Inam F, Wasay A, Mohsin S, Siddiqi Z, Qureshi F. Study through memes: Using memes as a teaching tool in anatomy. *Proc Sheikh Zayed Med Comp Lahore* 2024; **38(2)**:95-100. doi: 10.47489/szmc.v38i2.483.
4. Weinstein CE, Underwood VL. Learning strategies: The how of learning. In Thinking and Learning Skills 1st ed. Routledge; 2014 Jun 3. p. 241-58. Available from: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315060149-10/learning-strategies-learning-claire-weinstein-vicki-underwood.

5. Durrani SF, Yousuf N, Ali R, Musharraf FF, Hameed A, Raza HA. Effectiveness of spaced repetition for clinical problem solving amongst undergraduate medical students studying paediatrics in Pakistan. *BMC Med Educ* 2024; **24(1)**:676. doi: 10.1186/s12909-024-05479-y.
6. Sundararajan N, Adesope OO, Trevisan D. Rethinking the use of tests: A meta-analysis of practice testing. *Rev Educ Res* 2017; **87(3)**:659-701. doi: 10.3102/0034654316689306.
7. Rowland CA. The effect of testing *versus* restudy on retention: A meta-analytic review of the testing effect. *Psychol Bull* 2014; **140(6)**:1432-63. doi: 10.1037/a0037559.
8. Karpicke JD. Retrieval-based learning: A decade of progress. In: Byrne JH, Ed. Cognitive Psychology of Memory: Learning and Memory: A Comprehensive Reference (2nd ed), Academic Press, 2017, pp. 487-514. doi: 10.1016/B978-0-12-809324-5.21055-9.
9. Butler AC, Karpicke JD, Roediger 3rd HL. The effect of type and timing of feedback on learning from multiple-choice tests. *J Exp Psychol Appl* 2007; **13(4)**:273-81. doi: 10.1037/1076-898X.13.4.273.
10. Kang SH, McDermott KB, Roediger HL. Test format and corrective feedback modify the effect of testing on long-term retention. *Eur J Cogn Psychol* 2007; **19(4-5)**:528-58. doi: 10.1080/09541440601056620.
11. Atkinson RC, Shiffrin RM. Human memory: A proposed system and its control processes. In Psychology of learning and motivation. Vol. 2. New York: Academic Press; 1968. pp. 89-195. doi: 10.1016/S0079-7421(08)60422-3.
12. Shiffrin RM. Human Memory: A Proposed System and Its Control Processes! 30 Years of Memory. Inon Human Memory. Psychology Press; 1999. pp. 31-48. Available from: https://app.nova.edu/toolbox/instructionalproducts/edd8124/articles/1968Atkinson_and_Shiffrin.pdf.
13. Wixted JT. Atkinson and Shiffrin's (1968) influential model overshadowed their contemporary theory of human memory. *J Mem Lang* 2024; **136(5044)**:104471. doi: 10.1016/j.jml.2023.104471.
14. Karpicke JD, Blunt JR. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 2011; **331(6018)**:772-5. doi: 10.1126/science.1199327.
15. Chan JCK, Meissner CA, Davis SD. Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychol Bull* 2018; **144(11)**:1111-46. doi: 10.1037/bul0000166.
16. Greving S, Richter T. Practicing retrieval in university teaching: Short-answer questions are beneficial, whereas multiple-choice questions are not. *J Cognit Psychol* 2022; **34(5)**:657-74. doi: 10.1080/20445911.2022.2085281.
17. Stenlund T, Sundstrom A, Jonsson B. Effects of repeated testing on short- and long-term memory performance across different test formats. *Educ Psychol* 2014; **36(10)**:1-8. doi: 10.1080/01443410.2014.953037.
18. McDaniel MA, Roediger 3rd HL, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 2007; **14(2)**:200-6. doi: 10.3758/BF03194052.