



The effect of test-enhanced spaced learning on the otolaryngology board and annual examination results: A quasi-experimental study

SASAN DABIRI¹, AEEN MOHAMMADI^{2*}, RITA MOJTAHEDZADEH²

¹Department of Otolaryngology-Head and Neck Surgery, Amir Alam Hospital, Tehran University of Medical Sciences, Tehran, Iran;

²Department of E-Learning in Medical Education, Virtual School, Tehran University of Medical Sciences, Tehran, Iran

Abstract

Introduction: In the test-enhanced spaced learning, educational contents are presented in small packages of well-developed test questions with a defined frequency to the learners. It is not clear that applying this educational style might have a positive impact on the summative assessment. Therefore, in this study we assessed the effect of the test-enhanced spaced learning on the otolaryngology board and annual examinations of residency training.

Methods: In a quasi-experimental study with consecutive sampling, all forty-four residents of otolaryngology in four levels of training in 2016 at Tehran University of Medical Sciences (TUMS) received daily-standardized multiple-choice questions with a twice-repeated frequency of 10 days. Individual feedback according to one's response to each test was provided. The results of national board and annual exam were compared with the same results of all residents of other universities and previous year TUMS' residents for whom spaced learning were not applied and they were considered as the control groups. The board exam had two parts, multiple-choice questions, and computer-based clinical examination. The annual exam format was multiple-choice questions. The total score for each one was 150. Student's t-test and Mann-Whitney U test were used for comparative analyses.

Results: The mean of the board exam results showed statistically significant improvement compared to other medical schools (113.6±10.7 vs. 102.9±13.4 in multiple-choice questions, p=0.048, 118.7±12.5 vs. 54.1±60.0 in the computer-based clinical examination, p<0.001), while similar comparison results in the previous year did not show any significant difference.

Conclusion: Spaced learning with testing effect may be useful in the clinical education setting to improve the learning outcomes.

Keywords: Learning, Cognition, Internship, Residency, Otolaryngology

*Corresponding author:

Aeen Mohammadi,
Department of E-Learning
in Medical Education,
Virtual School, Tehran
University of Medical
Sciences, Dolatshahi, Naderi
St, Keshavarz Blvd, P.O box:
14166-14741, Tehran, Iran
Tel/Fax: +98 21 8896 9594
Email: a_mohammadi@
tums.ac.ir, aeen@yahoo.com

Please cite this paper as:
Dabiri S, Mohammadi A,
Mojtahedzadeh R. The
effect of test-enhanced
spaced learning on the
otolaryngology board and
annual examination results:
A quasi-experimental study.
J Adv Med Educ Prof.
2019;7(3):131-137. DOI:
10.30476/JAMP.2019.74696.

Received: 1 January 2018

Accepted: 4 June 2019

Introduction

The spaced effect that uses repeated exposures to a content or a practice to improve memorization was defined by Ebbinghaus in the late 19th century (1). Recently, biological studies have shown that repetition with the gap between exposures has significant positive

effects on memory (2, 3). Spaced learning, which is also known as distributed practice, is based on this effect and is effective in various types of learning in humans (4-6). Repeating the content presentation to the learner seems to improve their long-term memorization. The term "spaced" is used to distinguish it from the "massed" learning.

In massed learning, all educational contents are presented in one stage, but in spaced learning, contents are fractionated to small parts and delivered repeatedly in diverse time intervals. As effective memorizing of information is very critical for clinical practitioners, some researchers in medical education have focused on the impact of the spaced learning in recent two decades (7-12). Repetitive presentation of small packages instead of massed exposure is the key and critical element in this setting for the improvement of long-term memory.

On the other hand, students could remember most parts of the delivered material when they use tests during educational courses. This phenomenon, which is known as the testing effect increases the percentage of correct retrieval of information (13, 14). The potential of the test, being a small package of information, could be helpful in spaced learning, especially in medical education (15). On the other hand, social networks could be an effective means of transferring small packages of information. Generally, the emergence of e-learning in surgical education, mainly as a blended approach, could enhance the results of learning (16). Social networks, especially in the context of mobile learning, seem to improve learning in higher education as a new e-learning tool (17, 18). Telegram® is one of these social networks that are popular in Iran and has overgrowing but user-friendly capabilities.

One of the acceptable tools for evaluation of the educational programs and any change in them are summative assessments (19). A few studies focused on the summative assessments to evaluate the results of the test-enhanced spaced learning as a new educational structure in medical students (20). In this study, we delivered test questions based on spaced learning strategy via Telegram and assessed its effect on the results of annual and board examinations in otolaryngology residents.

Methods

In this quasi-experimental research with consecutive sampling, all 44 residents of Otolaryngology-Head and Neck Surgery at Tehran University of Medical Sciences (TUMS) participated in the study in 2016. The duration of the residency program in Iran is four years. Residents should participate in the annual exams to upgrade from the previous level of residency (levels 1 to 3) or be nominated for board examination (level 4). These annual exams have two parts: clinical work assessment (total score of 150 with minimum pass level of 105) and written multiple-choice question (MCQ) exam (total score of 150 and the minimum pass level

depends on the level of residency).

Seven last-year residents were the candidates for the national board exam. The board exam of Otolaryngology-Head and Neck Surgery specialty has two parts: the first one is written MCQ exam, and the second step is a computer-based clinical examination (CCE). The MCQs exam contains 150 questions with four choices. The CCE includes various types of questions that assess mainly the first four levels of Bloom's taxonomy in the cognitive domain (remembering, understanding, applying, analyzing). Each part has a score of 150, and anyone who obtains the minimum of 105 in MCQ exam would be eligible to participate in the CCE.

All the mentioned participants were invited to a private electronic group in Telegram®. As mentioned earlier, it is a phone-based media network with an instant message service that was released in 2013. Its real-time performance, the possibility of easy file transfer, simple group creation with multiple administrating and participating options, browsing media types like text, photos, and videos, and compatibility with various platforms renders it a favored social media network in Iran.

This collaborative program was added to the other educational programs of the department. Participation was not mandatory for the residents, and it was assured that participation would not affect their annual clinical work assessment scores. Residents received one similar question daily, responded to it until the end of that day, and received feedback individually according to the response. Also, there was an opportunity for discussion in the mentioned private group. There was not a uniform time interval for the repetition of the questions in spaced learning in the literature. On the other hand, there was a large amount of medical content as a reference study for annual and board examination. As a result, it was decided to perform two repetitions in this study. For each question, this process was repeated on day 11th and day 21st after the first time of exposure to the question (for example, the question of day one was repeated on day 11 and 21; question 2 on day 12 and 22; and so on). One of the authors (SD) gave all feedbacks and managed the discussions in the private group.

For covering the most parts of the reference content of examinations, the total 120 standardized MCQs were designed based on the reference book of the annual and board exams (Cummings Otolaryngology – Head and Neck Surgery, 6th edition, 2015, Elsevier, Saunders Co.) by the expert members of the national otolaryngology specialty board committee and

validated. These questions were delivered to the residents for five months.

Also, an organized answer for each question was developed in the portable document format from the main reference book and other related educational resources. Selection of the control group within the university was not possible because the participants could communicate with each other and separation of the groups was not earned. Thus, all residents of other universities and previous year TUMS' residents, none of whom had received educational contents with spaced structure, were selected as the control groups and the results of the annual and board exams were compared.

Participants' satisfaction was assessed with a researcher made scale which was designed and delivered by Google form builder. Participants responded to the questionnaire anonymously to increase the reliability of the responses. The questionnaire was validated by ten experts in the field of Otolaryngology or medical education. Its Cronbach alpha was 0.94. Other than demographic questions such as the level of residency or total time of participation in the project, the questionnaire included six questions with a five-point Likert scale from strongly agree to strongly disagree. Half of these questions were about the social media used for education and its impact on the learning of the students (satisfaction of social media use in education, usefulness of social media in education, agreement with social media use in education in the future), while other questions were mentioned to the direct point of the spaced learning structure (satisfaction from spaced learning in education, usefulness of spaced learning, agreement with spaced learning in the future).

Statistics

Student's independent t-test was applied for comparisons. For the groups in which the distribution was not normal, calculation of the differences with Mann-Whitney U test was also done. P-value of less than 0.05 was considered a statistically significant difference.

For improving the methodologic design of the study, postgraduate students of the other universities had access to the questions and comprehensive related responses in a same daily release for the intervention group, but without the repeated structure. As there was a limitation for selecting the control group within the intervention university group (due to the risk of the mixture of the participants and the impacts on the outcomes), results of the other universities in that year were considered as the control group.

Comparison of the participants' results with the counterpart results of the otolaryngology residents in other universities in the same year was considered as a horizontal comparison. There were 11 universities other than TUMS that were training otolaryngology residency. The horizontal comparison was made with the mean score of the assessments of all other universities globally. Also, for better controlling of confounding factors and better evaluation of the results, a vertical comparison was performed as well, i.e. comparing the participants' results of otolaryngology residents in the previous year from the same university.

Results

The total number of postgraduate students in the otolaryngology specialty training in the two years of 2015 and 2016 in Iran was 578 in 12 universities that were categorized into two groups: TUMS (40 and 44 students in the two years of 2015 and 2016, respectively), and other medical schools (241 and 253 students, respectively).

All TUMS residents participated in answering the questions according to their daily free time during the five months of the intervention. However, their participation was not a diurnal activity, and the average active days in a spaced learning interactive process were not similar for different levels of residency. The mean percentages of participation in each level for the total 120 question were 0%, 40%, 33% and 22% for level 1 to 4 residents, respectively.

Annual examination

Descriptive results of the analysis of the annual examination are shown in Table 1. Horizontal comparisons of the mean MCQ exam scores did not reach any significant differences in the two years of the study. Also, no significant difference was found in the vertical comparison of the mean MCQ exam scores in the TUMS group. However, this comparison in other universities showed a significant reduction in the MCQ exam score. Comparisons of the annual clinical work assessment scores as a covariant in vertical and horizontal directions did not show statistically significant differences unless in the vertical comparisons of the other universities group.

Board examination

Descriptive results of the analysis of the board examination are shown in Table 2. The analysis showed that there was a statistically significant better result in MCQ exam in the spaced education group in comparison to other universities. There was no significant difference

Table 1: Annual examination results of the otolaryngology specialty in TUMS and other medical schools

Type of exam	Medical School (number of residents)	2015	p-value			2016
			Between schools - 2015	Within a school group	Between schools - 2016	
MCQ score: mean±SD	TUMS (31)	91.5±10.0	0.193	0.990	0.506	91.6±12.2
	Other medical schools (180)	94.9±13.6		0.001*		89.8±14.6
CWA score: mean±SD	TUMS (37)	122.1±23.9	0.461	0.686	0.678	124.0±11.5
	Other medical schools (192)	118.8±23.6		0.040*		123.0±13.3

TUMS: Tehran University of Medical Sciences, MCQ: Multiple Choice Question, CWA: Clinical Work Assessment, SD: Standard Deviation, * statistically significant

Table 2: Board examination results of the otolaryngology specialty in TUMS and other medical schools

Type of exam	Medical School (number of residents)	2015	p-value			2016
			Between schools - 2015	Within school groups	Between schools - 2016	
MCQ score: mean±SD	TUMS (9)	107.3±14.7	0.391	0.362	0.048*	113.6±10.7
	Other medical schools (61)	102.6±15.4		0.905		102.9±13.4
CCE score: mean±SD	TUMS (7)	75.2±58.2	0.460	0.058	<0.001*	118.7±12.5
	Other medical schools (61)	60.7±54.0		0.520		54.1±60.0

TUMS: Tehran University of Medical sciences, MCQ: multiple choice question exam, CCE: computer-based clinical examination, SD: standard deviation, * statistically significant

when the horizontal comparison was done in the previous year. Analysis of the CCE scores in horizontal comparisons was also accompanied by the same results (Figure 1). A significant difference was found with the intervention group but not in the previous year.

In vertical comparison, CCE scores were better in the intervention group, but this difference was not statistically significant (p-value: 0.058). The non-parametric analyses with Man Whitney-U test did not reach a significant difference (p-value: 0.142). The observed power was 0.435. MCQ exam results also did not show a significant result in vertical comparison in the TUMS group. Vertical comparisons in other universities did not show significant differences in CCE and MCQ scores.

These vertical comparisons with Mann-Whitney U test also did not show significant differences.

Satisfaction Survey

From the total of 44 residents in the TUMS, 75% filled the satisfaction survey form. 97% agreed that spaced learning was effective or very effective (two top points from 5-point Likert scale) and others believed that it was neither good nor bad. While 85% completely agreed or agreed with the usefulness of the spaced learning system, 9% did not have an opinion, and 6% disagreed. 49% preferred to continue this educational system without any changes for the future, while 36% agreed with its continuation with little changes. Others did not have any

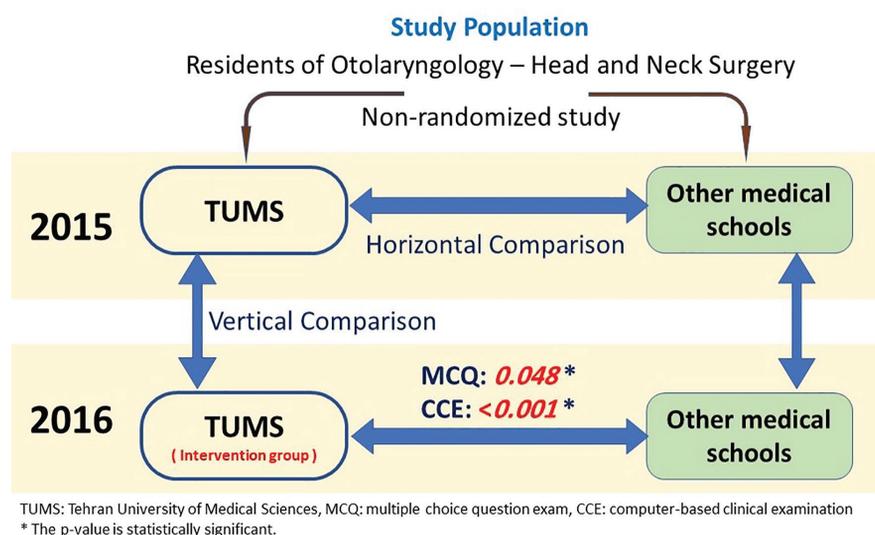


Figure 1: The simplified structure of the horizontal and vertical comparisons of the summative assessment results.

opinion or agreed with the continuation but with significant changes. Overall, there was nobody with severe disappointment on various questions of the survey, noting that even the selection of disagreement choices in the Likert scale was less than 10%, if present in different questions.

Discussion

Despite more than a century after Ebbinghaus presented the spacing effect, the recent electronic world gives a good infrastructure for applying this method. However, few studies have evaluated its impact on clinical education, especially on summative evaluations like the annual examinations. This study tried to assess the effect of the test-enhanced spaced learning on the final summative evaluation in the Otolaryngology residents.

There are not any statistically significant differences between the results of TUMS and other medical schools in annual and board examination results in the year before the intervention. Therefore, it could be assumed that these two groups did not have significant differences in competency and they were probably similar in educational structure and personal potencies for these examinations. Figure 1 simplifies the structure of the comparisons used in this study. As the Figure shows, the analysis was performed in vertical and horizontal directions.

Statistically significant differences were seen in the board examination results between TUMS and other medical schools in 2016, while such a difference was not seen in 2015 (horizontal comparison). As there was not a significant difference between the two groups in one year earlier, this superiority of the results in the target group might be assumed secondary to the intervention. On the other hand, the vertical comparison could not show any significant difference in both groups.

In annual examination results, no significant changes were seen in horizontal comparisons of both. Also, no significant differences were seen in two TUMS groups in vertical comparison, while a significant reduction in MCQ exam score of the other medical schools was found in the second year (2016). This might be secondary to the difference between the competencies of the participants. However, statistical analyses of the ward score did not show significant differences, and it could be assumed that the personal factors do not have a significant impact. Another reason might be the changes in the difficulty index of the examination. Maintenance of the mean score in TUMS after spaced education might be indirectly interpreted as the effectiveness of the

spaced learning. In a cohort study by Smeds and colleagues (19), the results of the final annual examination of the undergraduate medical students in the rotation of surgery were improved significantly in comparison to the previous year. The materials they used for spaced learning were MCQs designed according to educational objectives for the surgery ward rotation and were accessible via smartphone app., email, or online. This difference, i.e. the current study could not show a significant improvement in the results, might be due to the sample size as the observed power analysis of the current study was 44%.

Two important questions in the spaced learning are the time duration among the exposures, and the number of the exposures to reach long-term memorization. Cepeda et al. (3) showed that despite the effectiveness of duration from 1 second to more than a month, a few days of separations between the exposures had better results in comparison to exposures within a day. In the study by Smeds et al. (19), three and seven days for incorrect and correct responses respectively seemed to be an effective spaced time. Larsen (20) emphasized that intervals around a week to a month had very long stability of memorization of the information for months to years. A ten-day fixed interval with three repetitions in the current study showed some beneficial effects on summative assessment.

Participation in this program was not mandatory for residents. Most of the participants were the second and third-year residents that were 40% and 33%, respectively. In a similar research by Blazek et al. (10), email was used for spaced learning in medical students, and the participation rate was 75%. Task and responsibility overload in the first year of residency in a referral non-private teaching hospital may have a significant impact on the residents' participation. On the other hand, postgraduate residents usually are very busy in comparison to undergraduate medical students. As emphasized by Mammarella and the colleagues (12), the effectiveness of the spaced learning could be affected by emotional status. Finally, it is worth mentioning that despite the absence of active participation (responding to the messages), residents could answer the questions for themselves without any visible activity in the social media network.

Residents were very satisfied with this new educational program, and 87.5% declared that it was effective or very effective. About 97% or even more agreed that it was favorable. Half of the participants preferred to continue this program without any change in design. Similar results were achieved by Smeds in a student

survey which showed that helpfulness of the intervention for final course exam, annual exam and preparation for the oral exam were 95%, 90%, and 60%, respectively (20).

Spaced learning could be effective in surgical skills. Cecilio-Fernandes and colleagues (21) have reviewed the literature systematically for the role of spacing on the acquisition of surgical skills. The results showed that spacing of the education of the skills had improved the final score, and this point was helpful even for simple procedures like suturing. Nevertheless, the optimal gap time, which was between 5 minutes and one month and was not clear. Kim et al. (22) tried to reach the optimal time for spacing and retrieval in the real world. They analyzed the data collected using the learning management system (LMS) from 10514 employees who had answered to some work-related questions. The results showed a significant direct relationship between spacing intervals and retention intervals using a developed dataset model.

One of the limitations of this study was the low sample size, especially for participants of the board exam. Lack of a concurrent control group may also be important. Different questions in two years of annual and board examination might have a confounding impact on the results.

On the other hand, randomization was also impossible due to ethical educational issues. Also, the participation of residents was not mandatory. The only factor that motivated them to be active in this program was their incentive along with a load of their obligatory tasks.

The time interval duration and number of repetitions have remained a challenge in the literature (23). Comparison of two groups with different time durations and frequency of repetition in a randomized trial or a multi-center trial could be helpful for more clarification of the effectiveness of spaced education. Spaced learning could also be assessed in the psychomotor abilities that are important in surgical skills, especially for fine surgical skills like otolaryngology surgeries.

Conclusion

According to the results of this research on the postgraduate students of Otolaryngology, the test-enhanced spaced learning may improve learning considering the changes in the summative assessment results. However, further research on this subject is recommended.

Acknowledgment

This research was extracted from the thesis for the Master of Science degree of e-Learning

in Medical Education at Virtual School of Tehran University of Medical Sciences (thesis number: 9214658009). This research was approved by the ethical committee of Tehran University of Medical Sciences (number: 1396-3125). We are honored to send our special thanks to the otolaryngology residents of the TUMS; this study could not be performed without their kind participation. We also appreciate the National Center for Assessment in Medical Sciences Education of the Ministry of Health and Medical Education for providing data of national exams.

Conflict of Interest: There is not any conflict of interest.

References

1. Byrne JH. Learning and Memory. 2nd ed. New York: Macmillan Reference; 2003. 124 p.
2. Fields RD. Making memories stick. *Sci Am.* 2005;292(2):75-81.
3. Smolen P, Zhang Y, Byrne JH. The right time to learn: mechanisms and optimization of spaced learning. *Nat Rev Neurosci.* 2016;17(2):77-88.
4. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol Bull.* 2006;132(3):354-80.
5. Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. Spacing effect in learning: a temporal ridge of optimal retention. *Psychol Sci.* 2008;19(11):1095-102.
6. Lee TD, Genovese ED. Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Res Q Exerc Sport.* 1988;59(4):277-87.
7. Kerfoot BP. Interactive spaced education versus web based modules for teaching urology to medical students: a randomized controlled trial. *J Urol.* 2008;179(6):2351-6.
8. Mathes EF, Frieden IJ, Cho CS, Boscardin CK. Randomized controlled trial of spaced education for pediatric residency education. *J Grad Med Educ.* 2014;6(2):270-4.
9. Gandhi M, Beasley A, Vinas E, Sangi-Haghpeykar H, Ramin SM, Kilpatrick CC. Electronic Learning-Spaced Education to Facilitate Resident Knowledge and Guide Program Didactics. *Obstet Gynecol.* 2016;12(Suppl 1):23s-26s.
10. Blazek MC, Dantz B, Wright MC, Fiedorowicz JG. Spaced learning using emails to integrate psychiatry into general medical curriculum: Keep psychiatry in mind. *Med Teach.* 2016;38(10):1049-55.
11. Andersen SA, Konge L, Caye-Thomasen P, Sorensen MS. Learning Curves of Virtual Mastoidectomy in Distributed and Massed Practice. *JAMA Otolaryngol Head Neck Surg.* 2015;141(10):913-8.
12. Mammarella N, Fairfield B, Di Domenico A. Does emotion modulate the efficacy of spaced learning in recognition memory? *Cogent Psychol.* 2014;1:986922.
13. Glover JA. The testing phenomenon: not gone but nearly forgotten. *J Educ Psychol.* 1989;81(3):392-9.
14. Roediger HL, Karpicke JD. The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci.* 2006;1(3):181-210.
15. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ.* 2008;42(10):959-66.
16. Evgeniou E, Loizou P. The theoretical base of

- e-learning and its role in surgical education. *J Surg Educ.* 2012;69(5):665-9.
17. Maertens H, Madani A, Landry T, Vermassen F, Van Herzeele I, Aggarwal R. Systematic review of e-learning for surgical training. *Br J Surg.* 2016;103(11):1428-37.
 18. Yu AY, Tian SW, Vogel D, Kwok RC. Can learning be virtually boosted? An investigation of online social networking impacts? *Comput Educ.* 2010;55(4):1494-503.
 19. Biggs J. Assessment and classroom learning: a role for summative assessment? *Assessment Educ.* 1998;5(1):103-10.
 20. Smeds MR, Thrush CR, Mizell JS, Berry KS, Bentley FR. Mobile spaced education for surgery rotation improves National Board of Medical Examiners scores. *J Surg Res.* 2016;201(1):99-104.
 21. Cecilio-Fernandes D, Cnossen F, Jaarsma DADC, Tio RA. Avoiding Surgical Skill Decay: A Systematic Review on the Spacing of Training Sessions. *J Surg Educ.* 2018;75(2):471-80.
 22. Kim ASN, Wong-Kee-You AMB, Wiseheart M, Rosenbaum RS. The spacing effect stands up to big data. *Behav Res Methods.* 2019;51: 8.
 23. Larsen DP. Picking the right dose: the challenges of applying spaced testing to education. *J Grad Med Educ.* 2014;6(2):349-50.