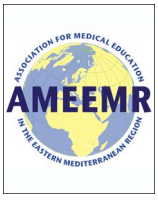

Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education ▮ (▮▮▮▮) ▮▮▮–▮▮▮

www.elsevier.com/locate/hpe

Effect of repeated/spaced formative assessments on medical school final exam performance

Edward K. Chang*, Paul F. Wimmers

Educational Development & Research, Dean's Office, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Received 17 May 2016; accepted 1 August 2016

Abstract

Purpose: Formative assessments, especially if spaced, encourage effective study habits such as retesting. The individual and combined effect of weekly formative assessments and cumulative assessments on final exam performance was studied.

Methods: Students were placed into 6 groups by weekly formative assessment performance and practice exam utilization/performance. Students who scored below the median on weekly assessments comprised groups 1/3/5. Groups 2/4/6 scored above the median. Groups 1/2 did not use the practice exam. Groups 3/4 scored below the median on the practice exam while Groups 5/6 scored above. Multiple comparisons were made using ANCOVA.

Results: Adjusted analysis showed weekly assessment and practice exam performance had a significant relationship with final exam performance ($F[7,145]=18.765$, $p<0.0005$). Groups 2/4/6 performed better on the final exam than groups 1/3/5 respectively (1v2, 80.8% vs 88.0%, $p<0.0005$ ‖ 3v4, 83.4% vs 88.6%, $p<0.0005$ ‖ 5v6, 84.1% vs 90.1%, $p<0.0005$). Group 1 performed worse than group 3 (80.8% vs 83.4%, $p=0.072$) and group 5 (80.8% vs 84.1%, $p=0.047$).

Discussion: Performance on weekly formative assessments was predictive of final exam scores. Struggling medical students will benefit from extra cumulative practice exams while students who are excelling do not need extra practice.

Keywords: Test-enhanced learning; Medical student education; Repeated testing; Spaced learning

1. Introduction

Every student has his or her own unique way of studying. Students vary when deciding what, how long and how to study.[1] Unfortunately many medical students do not necessarily use study habits that promote optimal learning. A survey of 254 medical students in two different medical schools revealed that 62.5% of students crammed for exams, only 17% would prepare a list of probable questions and answers before exams, and only 47.8% of students maintained a daily schedule of study hours.[2]

Certain methods of learning are more effective than others. For example, randomized controlled trials demonstrate that repeated testing promotes better retention of information than repeated studying, a phenomenon known as the testing effect.[3,4] This may be because tests introduce desired difficulties that force

*Correspondence to: Centre for Health Services, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA.

E-mail address: ekchang@mednet.ucla.edu (E.K. Chang).

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region.

http://dx.doi.org/10.1016/j.hpe.2016.08.001

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001

us to invoke our memory and retrieve information. Utilizing our memory creates new memories, perhaps by producing new mediators (clues that help us remember the original information).[5,6]

Tests are typically used as mandatory summative assessments given at the end of an instructional period to evaluate how well a student has learned[7]. Summative assessments are not tools to aid and promote learning but rather measure it. Therefore if educators want to encourage better learning, the role of traditional testing must be expanded. Larsen et al. recommend that "tests would no longer be considered neutral tools of measurement, but rather active instruments to aid in the acquisition and retention of knowledge"[3]. Formative assessments do not contribute to a student's final grade but have distinct purposes specific to the student or educator. They should be given soon after the information is learned [6] and these tests should be equally spaced and repeatedly administered over time.[8]

Students can use formative assessments to guide future learning by reflecting on feedback, identifying strengths and weaknesses, and understanding their teachers' expectations. Educators, on the other hand, can use formative assessments to evaluate and modify their teaching and also to monitor their students' progress[3,7,9]. Whereas summative assessments, when associated with final grades, may encourage competition, anxiety and desire for extrinsic rewards rather than the intrinsic rewards of personal learning, formative assessments can foster a non-judgmental and collaborative learning environment because students can learn without incurring academic penalty.[7]

We believe equally spaced and repeated formative assessments give students the benefit of repeated testing over simply restudying. This study explores the role of weekly formative assessments on final exam performance of first year medical students. We hypothesize that students who consistently perform well on these assessments are better prepared for final exams than those who do not consistently perform well. We also believe that those who utilize an additional cumulative practice test before the final exam as a final formative assessment score higher than students who do not use the tests, regardless of study strategy and weekly assessment performance.

2. Methods

2.1. Medical school curriculum

David Geffen School of Medicine (DGSOM) at UCLA follows the traditional 4-year medical school model, with two initial preclinical years followed by two clinical years. DGSOM students follow a "block" schedule for the first two preclinical years with a total of nine blocks. Student learning includes lecture, laboratory (including anatomy, histology, doctoring, and clinical skills), and Problem-Based Learning (PBL)[10,11]. At the end of each block, students take a cumulative final that covers the material that is taught in lecture, lab, and PBL. No other assignments throughout the block contribute to the final grade.

Although the final grade depends only on the cumulative final, first and second year students are required to do weekly formative assessments at the end of each week. Despite being mandatory, these assessments do not contribute to students' final grades. Their assumed purpose is to allow student and faculty to monitor the student's learning throughout the block. These assessments contain the material covered that particularly week in lecture, lab and PBL. They are done without supervision and at anytime between Friday 5 p.m. and Monday 8 a.m. at the student's discretion. Once the assessment is finished at the end of the week, the questions are closed, preventing the students from retaking them. At the end of the block, the questions from the assessments are compiled and randomized to make a comprehensive practice exam. This comprehensive practice exam was optional and administered in three parts. Medical students' performance on weekly assessments, practice exams, and final exam for only the first block were evaluated. The first block was chosen because at this point, the medical school has not influenced the students' uses and beliefs about studying.

2.2. Participants

All of the students' information was de-identified. This study was also reviewed and exempt by the IRB. Out of 178 medical students, all students took the weekly assessments and final. 97 (54.4%) took all three parts of the practice exam, 25 (14%) took one or two parts, and 56 (31.4%) did not take any parts. Since we were only interested in the full effect of the practice exam, we excluded students started only one or two parts but did not finish them all. Therefore, only students who took all three parts or no parts of the practice exam were included in the study to create a final population of 153 students.

The 153 students were first divided into two groups based on whether the student scored above or below the median on the weekly assessments. These two groups were each subdivided into three more groups – no practice exam, lower, or higher than the median on

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001

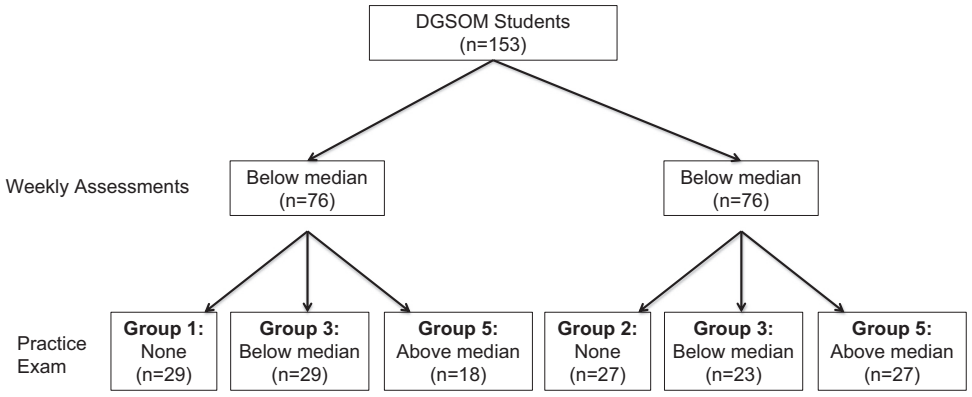

Fig. 1. Study population broken into 6 groups based on weekly assessment and practice exam performance.

the practice exam– to create a total of 6 groups. The odd numbered groups, Groups 1, 3, and 5 scored lower than the median on weekly assessments whereas Groups 2, 4, and 6 (even numbered groups) scored higher. Students who did not use the practice exam comprised Groups 1 and 2. Groups 3 and 4 included students who scored below the median on the practice exam while Groups 5 and 6 scored higher than the median. These groups are shown in Fig. 1.

2.3. Analysis

Multiple comparisons between the groups were made to understand the effect of weekly formative assessments and the practice exam on final exam performance. The odd numbered groups were compared to the even numbered groups to study the effect of the weekly assessments on final exam performance. Then Group 1 was compared to groups 3 and 5 while Group 2 was compared to Groups 4 and 6 to study the effect of the practice exam on final exam performance. We were interested in the individual *and* combined effect of weekly assessment and practice exams on final exam performance.

GPA and MCAT are known to strongly account for the variance in test performance in medical school[12]. They are good predictors for medical school GPA and USMLE scores with the MCAT often being the better predictor[13]. Therefore GPA and MCAT are two factors that were used in this study to control for pre-existing differences between the 6 groups. GPA and MCAT were changed to z-scores.

Statistical analysis was done with SPSS 22.0. Data were analyzed by using repeated measures (ANCOVA), and Fisher's least significant difference (LSD) was used to make post hoc multiple pairwise comparisons between the 6 groups. An alpha level of 0.05 was used for all statistical analyses.

3. Results

The difference in GPA and MCAT between students who scored lower than the median and students who scored higher than the median on the weekly formative assessments was statistically significant ($p < 0.0005$ for both). Pearson correlation for MCAT score or undergraduate GPA to final exam performance was 0.59 ($p < 0.0005$) and 0.50 ($p < 0.0005$).

After accounting for MCAT score and GPA, analysis of the 6 groups showed that the practice exam and weekly formative assessments still had a significant effect on final exam performance (F^{7},145] $= 18.77$, $p < 0.0005$). Groups 2, 4, and 6 performed better, on average, on the final exam than groups 1, 3, and 5 regardless of practice exam status. Employing the LSD post-hoc test, significant differences were found when comparing a group to its counterpart based on weekly assessment (Table 1 and Fig. 2).

Group 1 (19% of the study population), which consisted of students who scored below the median on weekly assessments and did not take the optional practice test had the lowest average score on the final exam. Group 6, which consisted of students who scored highest on the weekly assessments and practice exam, had the highest final exam average out of all other groups. Pearson correlation for average of weekly assessments to final exam performance was 0.58 ($p < 0.0005$).

Out of the groups that scored below the median on the weekly assessments, Group 1 (no practice exam) did not perform as well as group 3 which scored below median on practice exam (80.78% vs 83.44%, $p = 0.072$) and group 5 which scored above median on practice exam (80.78% to 84.15%, $p = 0.047$).

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001

Table 1
Students separated by weekly assessment and practice exam status – mean MCAT, GPA, and final exam scores are provided.

Group	N	Mean MCAT (z score)	Mean GPA (z score)	Weekly Quiz	Practice Exam	Mean Final Exam	95% CI Lower	95% CI Upper
1	29	−0.55	−0.52	Low	None	80.78	77.99	83.58
2	27	0.23	0.14	High	None	88.04	85.94	90.15
3	29	−0.13	−0.11	Low	Low	83.44	81.06	85.83
4	23	0.25	0.14	High	Low	88.63	86.78	90.49
5	18	−0.57	−0.30	Low	High	84.15	81.48	86.81
6	27	0.66	0.61	High	High	90.01	88.56	91.46
Total	153					85.8	84.7	86.8

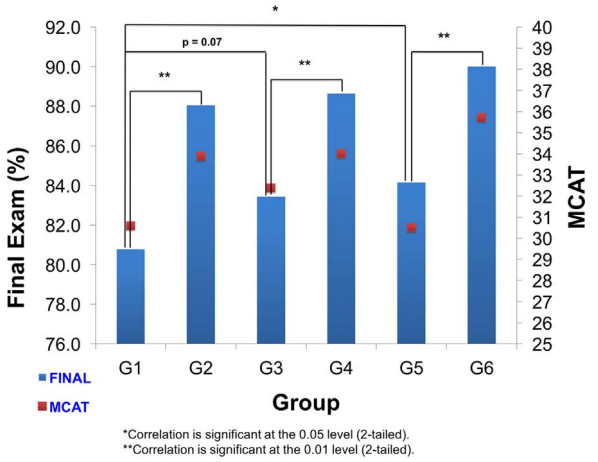

Fig. 2. A comparison of Groups 2, 4, and 6 to their respective counterparts Groups 1, 3, and 5 and a comparison between Group 1 and Groups 3 and 5.

Pearson correlation for practice exam to final exam performance was 0.36 ($p < 0.0005$). Pearson correlations for the relationship between GPA, MCAT, weekly quizzes, practice final and final exam are shown in Table 2. There was no statistical difference between groups 2, 4, and 6 in final exam performance even though group 2 did not take a practice exam.

4. Discussion

In our study, weekly formative assessments were highly predictive of final exam performance. Students who scored above the median on the weekly assessments did not score lower than 88% on the final exam while students who scored below the median did not score higher than 84%. There are a few possibilities why these assessments were so predictive.

First, students who are consistently performing well on the weekly assessments may be spacing their learning better. Medical students do not necessarily have effective study habits as evidenced by Vu and Galofre's study[2]. However, in our study, students were required to take weekly assessments. Larsen et al. argued that students when tested frequently, i.e. weekly, are more likely to keep up with readings and space out their study periods[3]. Research has shown that students who space out their studying rather than cramming have better long-term performance likely due to enhanced retention from multiple exposures[14]. It is possible that the students who are performing better on the weekly assessments are also the ones who are keeping up with the readings and spacing their learning, while the students who perform worse on the weekly assessments are more likely to cram for the final. Cramming works well if the student's goal is simply to do well on one exam. But if medical students want to perform well on an upcoming test *and* want to have good long-term retention for future board exams and clinical practice, they should space their studying[8,15]. All this suggests that ideal studying should incorporate a combination of repeated testing and spaced studying. Unfortunately, it is difficult to know whether or not medical students are truly spacing in an uncontrolled real life medical school setting. Nevertheless, weekly formative assessments may be one of the few effective ways to promote spaced learning.

Second, since students are required to take these weekly formative assessments, they are utilizing repeated

Table 2
Pearson correlations between different measurements of academic performance.

	GPA	MCAT	Weekly Quiz	Practice Final	Final
GPA	1	0.56**	0.44**	0.18	0.50**
MCAT		1	0.46**	0.29**	0.59**
Weekly Quiz			1	0.20*	0.580**
Practice Final				1	0.36**
Final					1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001

testing rather than only repeated studying, which is more beneficial[4,16]. Research suggests that the additional processing of information via retesting after initial learning plays an important role in long-term retention. Practice in retrieving information from memory seems to enhance future recall significantly[17,18]. If students are given more spaced formative assessments, they get the dual benefit of repeatedly testing themselves and having the incentive to space their studying.

Lastly, it is possible that certain students are simply performing better than others because they are smarter. Since those with higher GPAs and MCAT scores generally scored higher on weekly assessments, one could argue that these students may be scoring higher on the weekly assessments because they are simply smarter. There is no clear answer however because GPA and MCAT are flawed ways to measure intelligence. There is correlation between GPA, MCAT, weekly assessments scores, practice exam scores, and final exam scores but the relationship is not completely clear. It is possible that students with higher GPAs and MCAT scores are smarter and/or utilize better study habits such as spaced learning and repeated testing. Any one of those reasons could enable better performance on formative or summative assessments. Regardless of the reason, performance on practice exam and the weekly assessments strongly influenced final exam score even after correcting for GPA and MCAT, suggesting that they may have a mediating effect but do not completely predict exam success. Therefore, a combination of intelligence and study habits most likely explains why students who scored above the median on weekly assessments perform better on the final exam.

Students may score consistently lower on the weekly assessments for a variety of reasons, but the most likely reasons are that the students are either having difficulty with the study material itself, the pace of the learning and/or they are willingly procrastinating. Regardless of the reason for their lower scores, our study proposes that these students are at higher risk for failing final exams. Our results advocate that these students would benefit from an extra cumulative formative assessment such. Among students who scored lower than the median on the weekly assessments, Groups 3 and 5, simply by taking a taking the practice test, scored at least 2.5 percentage points better then the Group 1. Group 5 scored 3.3 percentage points higher ($p = 0.047$). And even though the difference between Group 1 and 3 was not statistically significant at the 0.05 level ($p = 0.072$), the difference is still significant enough to suggest that students who are struggling should at least attempt a practice test given that the difference between failing and passing the final exam could simply be a couple of percentage points. The lack of statistically significant difference between Groups 2, 4 and 6 suggest that the practice exam is not as beneficial to those who are learning well throughout the block because they do not necessarily need the extra practice or feedback.

Our study has a few limitations. First, the study was only done at DGSOM, meaning our conclusions cannot necessarily be extrapolated to all medical schools. Second, we did not have a randomly selected control group that does not do weekly assessments throughout the block. Unfortunately this does not necessarily fall in line with school policy and might be unethical to put one group of students at a disadvantage. Another limitation was that even though students were encouraged to take the weekly assessments closed book by themselves, ultimately students chose to utilize the assessments as they wished. If all students were forced to take the assessments by themselves closed book, our study may have been more powerful. Nevertheless, regardless of how the student took the weekly formative assessments and practice exam, our study still shows their influence on final exam performance. Lastly, our study assumes behavior based on performance but we don't necessarily know if that is the truth for every individual. This does not, however, prevent us from drawing out the cumulative effect of the weekly assessments and the practice exams.

Our study suggests that weekly formative assessments serve as a good way to monitor a student's study progress throughout a block. This is beneficial for both the student and educator. If a student is consistently performing poorly on weekly assessments, an educator can intervene and assist the student, in order to prevent him or her from failing the final exam. The most problematic group is this study is Group 1 (19% of the study population). They had lower scores on the weekly assessments and did not utilize the practice exam even when given the opportunity. Perhaps they were stressed and did not want another test telling them they need to study more. Regardless of their reason for worse performance, they need additional help. For educators, knowing who is performing worse on formative assessments provides them the perfect opportunity to intervene. They can mediate by providing tutoring or, in the context of this study, offering more practice tests opportunities for all students (since repeated testing is beneficial overall) knowing that these practice tests will be especially helpful to struggling students. Educators should also encourage struggling students to use practice tests whenever given the opportunity because most students are not adept at evaluating their own study habits.

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001

There must be a fine balance between self-regulated learning and influence from educators. Educators may administer graded summative assessments frequently because they want their students to space their learning and since repeated testing promotes learning. But this strategy may backfire, according to Neame and Powis, because the structure and frequency of examinations in the a medical school curriculum "may inhibit even the most self-confident and independent student from learning his own way"[19]. If a student becomes extrinsically motivated, simply wanting higher exam scores or avoiding punishment from superiors, rather than intrinsically motivated, engaging in tasks because the individual finds them challenging, interesting and enjoyable, then he or she is more likely to have a superficial approach to studying[20]. The best way to avoid this situation is to provide repeated formative assessments instead of repeated summative assessments because formative assessments allow students to control their own learning without extrinsic motivation and utilize effective study tactics like spaced learning and repeated testing.

Disclosures

Ethical approval: Exempted on 3/7/14 by the Institutional Review Board 14-000354.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Other disclosures

No conflicts of interest.

Acknowledgments

None.

References

1. Kornell N, Bjork RA. The promise and perils of self-regulated study. *Psychonomic Bulletin Review* 2007;14:219–224.
2. Vu NV, Galofre A. How medical students learn. *Academic Medicine* 1983;58:601–610.
3. Larsen DP, Butler AC, Roediger III HL. Test-enhanced learning in medical education. *Medical Education* 2008;42:959–966.
4. Larsen DP, Butler AC, Roediger III HL. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical Education* 2009;43:1174–1181.
5. Pyc MA, Rawson KA. Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2012;38:737.
6. Roediger HL, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science* 2006;1:181–210.
7. Rolfe I, McPHERSON J. Formative assessment: how am I doing? *The Lancet* 1995;345:837–839.
8. Karpicke JD, Roediger III HL. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2007;33:704.
9. Cox M, Irby DM, Epstein RM. Assessment in medical education. *New England Journal of Medicine* 2007;356:387–396.
10. Wood DF. Problem based learning 971-971. *Bmj* 2008;336.
11. Schmidt HG, Rotgans JI, Yew EH. The process of problem-based learning: what works and why. *Medical Education* 2011;45:792–806.
12. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *Bmj* 2002;324:952–957.
13. Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine* 2005;80:910–917.
14. Kornell N. Optimising learning using flashcards: spacing is more effective than cramming. *Applied Cognitive Psychology* 2009;23:1297–1317.
15. Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology* 2013;64:417–444.
16. Pyc MA, Rawson KA. Why testing improves memory: mediator effectiveness hypothesis 335-335. *Science* 2010;330.
17. Karpicke JD, Roediger HL. The critical importance of retrieval for learning. *Science* 2008;319:966–968.
18. Kornell N, Hays MJ, Bjork RA. Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2009;35:989.
19. Neame R, Powis D. Toward independent learning: curricular design for assisting students to learn how to learn. *Academic Medicine* 1981;56:886–893.
20. Moneta GB, Spada MM. Coping as a mediator of the relationships between trait intrinsic and extrinsic motivation and approaches to studying during academic exam preparation. *Personality and Individual Differences* 2009;46:664–669.

Edward K. Chang is a third year medical student at the David Geffen School of Medicine at UCLA.

Dr. Paul F. Wimmers is Director of Evaluations in the Center for Educational Development & Research at David Geffen School of Medicine at UCLA.

Please cite this article as: Chang EK, Wimmers PF. Effect of repeated/spaced formative assessments on medical school final exam performance. *Health Professions Education* (2016), http://dx.doi.org/10.1016/j.hpe.2016.08.001