

Knowledge Retention after an Online Tutorial: A Randomized Educational Experiment among Resident Physicians

Douglas S. Bell, MD, PhD¹, Charles E. Harless, MS¹, Jerilyn K. Higa, MS¹, Elizabeth L. Bjork, PhD², Robert A. Bjork, PhD², Mohsen Bazargan, PhD³, and Carol M. Mangione, MD, MSPH^{1,4}

¹Division of General Internal Medicine and Health Services Research, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; ²Department of Psychology, University of California, Los Angeles, CA, USA; ³Department of Family Medicine, Charles R. Drew University, Los Angeles, CA, USA; ⁴Department of Health Services, UCLA School of Public Health, Los Angeles, CA, USA.

BACKGROUND: The time course of physicians' knowledge retention after learning activities has not been well characterized. Understanding the time course of retention is critical to optimizing the reinforcement of knowledge.

DESIGN: Educational follow-up experiment with knowledge retention measured at 1 of 6 randomly assigned time intervals (0–55 days) after an online tutorial covering 2 American Diabetes Association guidelines.

PARTICIPANTS: Internal and family medicine residents.

MEASUREMENTS: Multiple-choice knowledge tests, subject characteristics including critical appraisal skills, and learner satisfaction.

RESULTS: Of 197 residents invited, 91 (46%) completed the tutorial and were randomized; of these, 87 (96%) provided complete follow-up data. Ninety-two percent of the subjects rated the tutorial as "very good" or "excellent." Mean knowledge scores increased from 50% before the tutorial to 76% among those tested immediately afterward. Score gains were only half as great at 3–8 days and no significant retention was measurable at 55 days. The shape of the retention curve corresponded with a 1/4-power transformation of the delay interval. In multivariate analyses, critical appraisal skills and participant age were associated with greater initial learning, but no participant characteristic significantly modified the rate of decline in retention.

CONCLUSIONS: Education that appears successful from immediate posttests and learner evaluations can result in knowledge that is mostly lost to recall over the ensuing days and weeks. To achieve longer-term retention, physicians should review or otherwise reinforce new learning after as little as 1 week.

Electronic supplementary material The online version of this article (doi:10.1007/s11606-008-0604-2) contains supplementary material, which is available to authorized users.

Trial registration: NCT00470860, *ClinicalTrials.gov*, U.S. National Library of Medicine.

Received June 18, 2007

Revised January 31, 2008

Accepted March 11, 2008

Published online April 30, 2008

KEY WORDS: knowledge retention; online tutorial; randomized educational experiment; resident physicians; educational technology; learning theory.

J Gen Intern Med 23(8):1164–71

DOI: 10.1007/s11606-008-0604-2

© Society of General Internal Medicine 2008

INTRODUCTION

Nearly all physicians invest time in educational activities aimed at maintaining their knowledge, but their efforts are often unsuccessful.^{1,2} Many educational activities engender immediate knowledge or performance gains, but early studies that measured subsequent recall among physicians found it to be poor after 6–18 months.^{3,4} A recent review identified 15 studies of physician education in which at least some knowledge gains persisted over 3 months or longer, but the report concludes that the quality of evidence for factors that promote retention is weak.⁵

Forgetting previously learned information is a universal human experience. People tend to view instances of forgetting as personal failings, but forgetting over time is actually an essential mental function, enabling us to access more current information in preference to older, typically less-relevant information.⁶ Memories formed during educational activities are also susceptible to forgetting, despite one's intention to remember them.

Reinforcement, for example, through restudy of the original material or through practice in recalling it, is an important method for improving one's retention of learning.⁷ Laboratory experiments show that reinforcement, sometimes referred to in the literature as *relearning*, is more effective at engendering long-term storage if it takes place after enough competing experiences have occurred to diminish the original memory's retrieval strength.⁸ Online educational programs offer the potential to deliver reinforcement at prespecified time intervals, but the amount of time that should elapse to optimize reinforcement for physicians is not known. Thus, a more detailed understanding of the retention curve after physicians' initial learning experiences could inform the design of online and other educational programs, but prior studies have measured knowledge retention among physicians after only a single time delay.⁵

Other factors that can modulate forgetting include the extent to which the learner forms associations with other knowledge,⁹ which may, in turn, be influenced by one's self-efficacy.¹⁰ Interactive online education that is personalized to

one's current beliefs and level of self-efficacy may have considerable potential to improve the depth of physicians' learning.^{11,12} Graduate and continuing medical education programs are beginning to make use of online educational methods, but more evidence is needed to guide the design of these experiences.¹³⁻¹⁷

In the current study, our goal was to chart the time course of retention after initial learning from an online tutorial covering principles of diabetes care and to examine the role of learner characteristics such as critical appraisal skills and self-efficacy that might modify physicians' learning and retention.

METHODS

Study Population and Recruitment

During the period December 2004 through May 2005, we recruited residents from the family medicine and Internal Medicine residency programs at 2 academic medical centers. We did so by sending a lead letter, announcing the study at teaching conferences, and then sending each resident an e-mail invitation containing a personalized web link that led to an online consent form. Residents who did not respond were sent up to 5 follow-up invitations. We paid \$50 for completing the study. The institutional review boards for each participating university approved the study.

Enrollment Survey and Intervention Scheduling

Consenting subjects took an online enrollment survey (Online Supplement) assessing their demographic characteristics, prior rotations in endocrinology and cardiology, computer attitudes (using 8 items from a published scale),¹⁸ self-efficacy toward diabetes care (using 9 items adapted from 2 other physician surveys),^{19,20} and critical appraisal skills (using 3 items we authored). We had pilot-tested these items among the coinvestigators. After the enrollment survey, subjects could begin the tutorial immediately or schedule themselves for a future available date in the ensuing 2 months. We required all subjects to complete the tutorial by June 2005.

Learning Objectives and Knowledge Questions

With the permission of the American Diabetes Association (ADA), we created a tutorial based on their guidelines for managing hypertension and lipids in patients with diabetes.^{21,22} We wrote 20 learning objectives,²³ each based on a specific guideline passage, representing knowledge or cognitive skills that would enable better diabetes care (Table 1). We then drafted and refined 2 multiple-choice questions to measure the achievement of each learning objective, many of which involved identifying the best management of a patient case. The questions and objectives were revised based on feedback from an expert diabetologist and on cognitive interviews with 12 randomly selected Los Angeles-area Internists and family physicians. Of the 40 final questions, 32 presented exclusive choices and had a single correct answer, whereas 8 allowed the selection of multiple answers with each answer being correct or incorrect. Overall, the items had from 3 to 9 response options.

Web-based Tutorial System

We constructed an online tutorial system called the longitudinal online focused tutorial system (LOFTS). LOFTS began with a pretest consisting of 1 randomly selected question per learning objective, administered without feedback; it then displayed an overview of the learning objectives and primary guideline documents; and, finally, it reviewed each learning objective individually in an interactive fashion (Appendix). LOFTS's upper pane showed feedback on the user's pretest response for the learning objective, and its lower pane showed the related guideline passage as highlighted segments within the complete guideline document.

For incorrect responses, LOFTS told users to review the guideline passages and then try another answer. Users had to select the correct answer(s) before moving on to review the next learning objective. After they selected the correct answer, the system gave additional explanation for the question. Below each explanation, it asked users to rate, on a 7-point scale, the learning objective's importance to providing excellent overall diabetes care. The tutorial system ended by assessing learners' satisfaction using a scale we adapted from the American College of Physicians.²⁴

The LOFTS system warned users after 5 minutes of inactivity and logged them off after another 30 seconds. It sent users who interrupted the tutorial an e-mail asking them to return as soon as possible. Upon returning, users resumed their tutorial session from the point where they had left off. We programmed LOFTS mostly in Java with additional functions performed by JavaScript in the user's browser; MySQL was its database.

Random Assignment, Posttest

LOFTS randomly assigned participants completing the interactive tutorial to 1 of 6 follow-up intervals for taking the posttest: 0, 1, 3, 8, 21, or 55 days. It used a randomized permuted block algorithm to balance the number of subjects in each group,²⁵ and it concealed the randomization sequence from personnel involved in recruiting. Subjects assigned to the "0 days" group went immediately into the posttest; the other subjects were informed of their assignment and then sent, on the appropriate day, an e-mail invitation with a web link to the posttest. Subjects had 24 hours after receiving the posttest e-mail invitation to complete the test; those who did not respond within 21 hours were paged with an additional reminder. For each subject, the posttest consisted of the 20 questions that they had not seen on the pretest. Thus, the random assignment of questions made the posttest different from the pretest for each subject, but averaged across subjects it made the pretest and posttest equivalent.

Analysis

We constructed scale scores for computer attitudes, diabetes care self-efficacy, and learner satisfaction by summing the ratings given for each item. Critical appraisal skills were initially scored from 0 to 3 by giving 1 point for each correct answer. Internal consistency reliabilities for scales were estimated using Cronbach's coefficient alpha.²⁶

We scored knowledge items as 0 for incorrect or 1 for correct with partial credit for multiple-answer items based

on the proportion of all response options that were correctly chosen or not chosen. We classified items as malfunctioning (in the context of the tutorial) if scores among the participants answering them at 0 or 1 days after the tutorial were lower than scores among those answering them on the pretest, and the learning objectives with malfunctioning items were excluded in defining the final knowledge scale. We used means to represent the central tendency of knowledge scale scores because Shapiro–Wilk tests indicated that their distribution was consistent with the normal distribution.

We used chi-square tests to assess associations among categorical variables and *t* tests or ANOVA tests to compare the values of continuous variables among groups. To assess the bivariate relationship between the delay time and posttest scores, we used linear regression with a 1/4-power transfor-

mation of time to account for the expected curvilinear relationship.^{9,27} To explore the association of learner characteristics with the amount learned and rate of retention loss, we used multivariate linear regression modeling of posttest scores as a function of (delay time)^{1/4}, characteristics, and (delay time)^{1/4}–characteristic interactions. In these analyses, we dichotomized the critical appraisal skills as 0 vs ≥ 1 because average performance was very similar across participants in the latter category.

All reported *P* values are two-tailed. Power calculations showed that a sample size of 78 participants (13 in each group) would provide a power of 81% in a linear regression analysis to detect a loss in retention of 1 standard deviation after 55 days with an alpha significance level of 0.05. Statistical calculations were carried out in SAS version 8 or in JMP version 5.1 (both SAS Institute, Cary, NC, USA).

Table 1. Success of the Test Items and the Tutorial for Each Learning Objective

Guideline Topic Learning objective	Proportion correct on pretest*	Proportion correct on posttest at 0, 1-day†	Importance rating‡
Hypertension management in adults with diabetes			
•Blood pressure targets			
Identify how blood pressure goals differ for patients with DM	0.81	1.0	5.9
Identify specific blood pressure goals recommended by the ADA	0.48	0.79	6.0
Identify the evidence for achieving blood pressure goals§	0.55	0.24	5.8
•Nondrug management of blood pressure			
Identify the role of sodium restriction	0.66	0.69	5.2
Identify how should exercise be prescribed	0.47	0.52	5.8
Identify the evidence for the benefits of weight loss	0.36	0.76	5.6
•Drug management of blood pressure			
Identify when drug management of hypertension should be initiated	0.50	0.79	6.0
Identify the strategy that the ADA recommends for initial drug therapy of high blood pressure	0.42	0.63	5.6
Identify the antihypertensive medications that have been proven to improve outcomes for diabetes patients	0.56	0.59	5.9
Identify which antihypertensive agents have proven superior to others in direct comparisons	0.32	0.46	5.8
Dyslipidemia management in adults with diabetes			
•Lipid testing			
Identify dyslipidemia patterns that are associated with type 2 DM and their significance	0.31	0.82	5.6
Identify the intervals at which patients with DM should have lipid testing	0.24	0.38	5.9
•Lipid management			
Identify the lipid goals recommended by the ADA	0.74	0.93	5.8
Identify the role of lifestyle modification in treating dyslipidemias among patients with diabetes	0.53	0.69	5.7
Identify what should trigger initiation of lipid-lowering medication§	0.72	0.62	5.8
Identify the ADA recommendation for treating LDL elevations	0.94	0.93	6.0
Identify the ADA recommendation for treating low HDL levels	0.52	0.71	5.6
Identify the ADA recommendation for treating elevated triglycerides	0.38	0.57	5.7
Identify the ADA recommendation for treating combined dyslipidemias§	0.55	0.52	5.8
Identify the special considerations necessary in using niacin therapy for dyslipidemias	0.23	0.83	5.8

Each learning objective was assessed by 2 test items. For each subject, one of these items was randomly selected for the pretest and then the other item was used on the posttest. For the 8 knowledge items in which multiple responses could be selected, partial credit was given for each correct response to a maximum score of 1 point if all responses were correct for the item.

*Mean proportion correct for the learning objective when assessed on the pretest. (About half of subjects were assessed with 1 item and about half with the other item for the learning objective).

†Mean proportion correct for the learning objective when assessed on the posttest among the subgroup of 29 subjects who were tested immediately or 1 day after the tutorial.

‡Mean rating of the learning objective's importance for "providing excellent overall diabetes care" on a scale from 1 ("extremely unimportant") to 7 ("extremely important") given by subjects after viewing the guideline passage, pretest answer, and explanation.

§Learning objectives for which the items and tutorial, in combination, appeared to malfunction with performance significantly worse after the tutorial (within 1 day) than it was on the pretest. In most cases, this effect appeared to be because of potentially confusing questions. These learning objectives were excluded from the final knowledge scale.

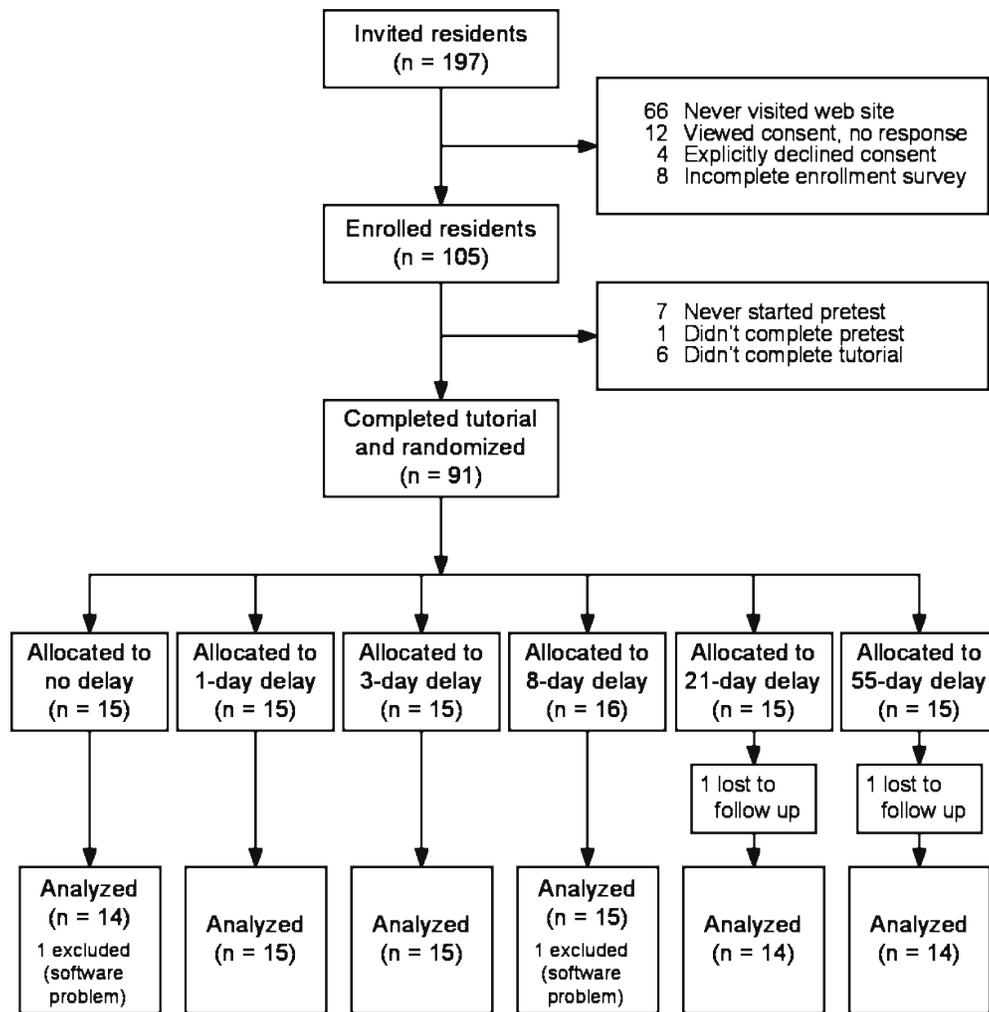


Figure 1. Disposition of all individuals eligible for study participation through recruitment, tutorial participation, randomization, posttest, and analysis.

RESULTS

Participant Characteristics

Of 197 residents invited, 131 visited the study website; 113 consented and 4 explicitly declined participation; 105 completed the enrollment survey, and of these, 91 completed the initial tutorial and were randomized (Fig. 1). Two subjects failed to complete the posttest and 2 other subjects were excluded in the analysis because a technical problem allowed them to skip part of the tutorial, leaving 87 who fully completed the study (44% of those invited, 96% of those randomized). Participants who completed the study were similar to the remainder of those recruited in terms of specialty (Internal Medicine versus family medicine) and residency year, but participation rates differed between universities (Tables 2 and 3). No subject characteristic was significantly associated with the posttest time delay.

Our scales for critical appraisal skills, computer attitudes, and self-efficacy toward diabetes care each demonstrated good to excellent internal consistency reliability (Cronbach's alphas of 0.75, 0.86, and 0.89, respectively). The computer attitude score correlated with a single-item computer sophistication question ($r=0.49$, $P<.0001$). On average, males had higher

scores than females for critical appraisal skills (1.4 vs 0.9 out of 3; $P=.04$) and for computer attitudes (23 vs 21 out of 28, $P=0.002$), but not for diabetes self-efficacy (21 vs 20 out of 36, $P=.33$). Residency year was correlated with diabetes self-efficacy (+1.5 points/year, $P=0.02$), but not critical appraisal skills or computer attitudes.

Table 2. Characteristics of Study Participants and Non-participants

Characteristics known for all recruited	Completed study(n=87)	Never enrolled or failed to complete(n= 110)	P value
Residency			.78*
training year			
First year (R1)	34	38	
Second year (R2)	31	32	
Third year (R3)	35	30	
Internal Medicine (versus family medicine)	71	71	.95*
Female	41	54	.09*
University A (versus B)	75	59	.02*

Data are within-column percentages except where noted otherwise. *Chi-square tests; test of significance

Table 3. Additional Characteristics of Study Participants

Characteristics known for enrollees	Completed study (n=87)	Failed to complete (n=18)	P value
Age, yrs, mean±SD	31±5	31±5	.65*
Completed any rotation in			
Endocrinology	41	44	.77†
Cardiology	74	44	.02†
Critical appraisal skill score, mean±SD	1.20±1.2	0.94±1.3	.43‡
Computer attitude score, mean±SD	22±4	22±4	.77‡
DM self-efficacy score, mean±SD	21±5	19±6	.18‡

Data are within-column percentages except where noted otherwise.

*t tests; test of significance

†Chi-square tests; test of significance

‡Wilcoxon rank sum tests; test of significance

Learners' Use of the Tutorial, Question Ratings, and Satisfaction

Subjects completing the study spent an average of 18 minutes browsing the 2 guideline documents (range=0.4 to 55 minutes, median=13) and an average of 16 minutes going through the interactive tutorial for each pretest question (range=4.3 to 51 minutes, median=14). Subjects rated all learning objectives as important for providing excellent diabetes care—mean ratings were above 5.5 out of 7 for 19 of the 20 learning objectives (Table 1). The mean learner satisfaction scale score was 17 on a 0–20 scale (Cronbach's α =0.92, SD=2.9). The tutorial's overall quality was rated as "very good" or "excellent" by 80 of the 87 final subjects (92%).

Knowledge Results

Including all 20 of the learning objectives, the overall mean knowledge score increased from before to after the tutorial (Table 4). Among the groups assigned to different delay times, scores did not differ significantly before the tutorial (the pretest), but were much higher after the tutorial for those tested immediately than for those tested at longer time delays. For 3 learning objectives (identified in Table 1), we classified

the test items as malfunctioning in combination with the tutorial. For these, average performance was poorest immediately after the tutorial and improved with increasing delay time toward the pretest mean of 1.8 for these items. The remaining 17 learning objectives formed our final knowledge scale. Overall mean scores on this scale increased from 8.5 (50%; SD=1.8 points) before the tutorial to 10.8 (63%; SD=2.5) afterward.

Before the tutorial, the final knowledge scores correlated with residency year (+0.7 points/year, $P=.005$) and with completion of a prior cardiology rotation (+0.9 points, $P=.05$), but these pretest scores were not associated with any other subject characteristic we assessed (age, gender, specialty, prior endocrine rotation, critical appraisal skill, diabetes self-efficacy, and computer attitudes). After the tutorial, increasing time delays were associated with lower scores; at 3 and 8 days, performance reflected about half of the gains seen immediately; and at 55 days, performance was equivalent to the pretest mean. Posttest scores were negatively correlated with a 1/4-power transformation of the time delay after the tutorial (Fig. 2). Performance on the malfunctioning items followed the opposite pattern.

In multivariate regression modeling, in addition to the power-transformed time delay, final posttest scores were significantly associated with critical appraisal skill, having completed a cardiology rotation, and subject age, and there was a trend toward better performance for Internist versus family medicine trainees (Table 5). Adjusting for these effects modestly blunted the effect of delay time on performance. The remaining characteristics (listed above) were not significantly associated with posttest knowledge in this model. Furthermore, there were no significant interaction effects between any characteristic and the time delay, indicating that these characteristics did not appear to modify the rate of retention decline. These effects differed only slightly when all 20 learning objectives were included.

DISCUSSION

After a one-time online tutorial, residents' ability to recall what they had learned diminished relatively quickly, despite large initial knowledge gains and positive learner evaluations. No

Table 4. Mean Knowledge Scores for each Time Delay after the Tutorial

	n	Score including all 20 items on each test		Score for the 3 malfunctioning items*		Final 17-item knowledge score	
		Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Overall mean	87	10.3 (0.2)	12.4 (0.3)	1.8 (0.1)	1.6 (0.1)	8.5 (0.2)	10.8 (0.3)
Assigned delay							
None	14	9.6 (0.4)	14.2 (0.6)	1.9 (0.2)	1.2 (0.2)	7.8 (0.4)	13.0 (0.5)
1 day	15	10.1 (0.6)	12.7 (0.7)	1.5 (0.1)	1.5 (0.2)	8.5 (0.6)	11.2 (0.6)
3 days	15	9.6 (0.5)	12.4 (0.6)	1.7 (0.3)	1.5 (0.2)	7.8 (0.4)	10.9 (0.5)
8 days	15	10.8 (0.5)	12.8 (0.6)	1.8 (0.1)	1.8 (0.2)	9.0 (0.4)	11.0 (0.6)
21 days	14	10.4 (0.5)	12.0 (0.9)	1.6 (0.3)	1.9 (0.3)	8.8 (0.4)	10.1 (0.7)
55 days	14	11.1 (0.6)	10.2 (0.7)	2.4 (0.2)	1.7 (0.2)	8.8 (0.6)	8.5 (0.6)
P value†		.20	.006	.10	.27	.31	<.0001

Values are the mean (standard error) number of questions answered correctly, for each category of knowledge items.

*Items for the learning objective were classified as malfunctioning if subjects were less likely to answer correctly after they completed the tutorial (at 0 or 1 day) than they were beforehand (on the pretest).

†Tests of significance are one-way ANOVA tests for differences between assignment groups on the pretest and the posttest.

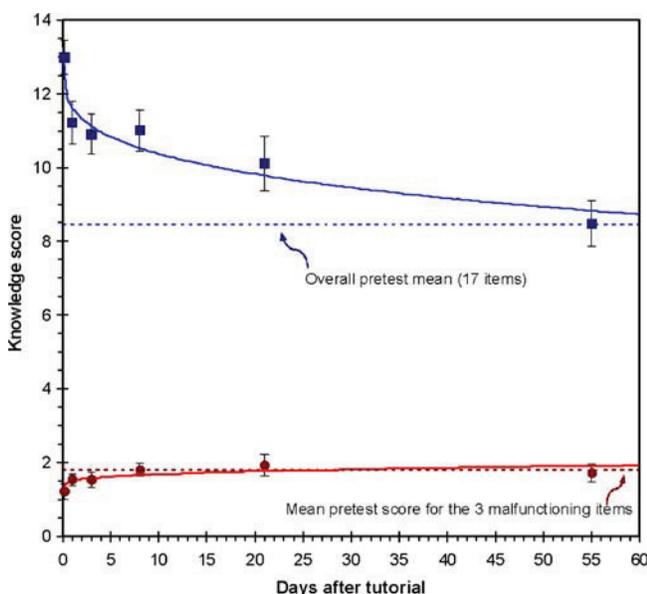


Figure 2. Chart of knowledge test results for each time interval after the tutorial with squares representing the mean scores on the final 17-item posttest and circles representing the sum of scores on the posttest for the 3 malfunctioning items. Horizontal dotted lines represent subjects' mean knowledge scores before the tutorial for each of the 2 item groups. Vertical bars show the standard error of the mean. Solid lines represent the best linear fit for each outcome versus the time delay, after fourth-root transformation of the time interval. (17-item posttest score = $13.3 - 1.63t^{1/4}$, $R^2 = 0.27$, $P < .0001$; posttest sum of 3 malfunctioning items = $1.26 + 0.23t^{1/4}$, $R^2 = 0.05$, $P = .04$.)

subject characteristics were associated with slower retention loss, although residents with critical appraisal skills and prior cardiology rotations achieved greater initial learning, and having thus started from a higher base, they had greater retention overall. Subjects with more negative attitudes toward computers and those with less self-efficacy toward diabetes care gained and retained as much from the tutorial as subjects with higher levels of these characteristics. This study's strengths include (a) its use of a single time interval for measuring each subject's retention without the reinforcement that would arise from repeated testing and (b) its use of a common item pool for each test, thus balancing the difficulty of pretest and posttest.

The retention loss we observed is greater than in some prior studies of online education, including our own prior study in which 58 residents retained 40% of their initial knowledge gains 4–6 months after a tutorial on postmyocardial infarction care (on which they spent an average of 28 minutes).²⁴ For the current study, we streamlined aspects of the LOFTS tutorial process based on user feedback. Perhaps as a result, subjects in the current study spent 40% less time per learning objective than did subjects in the online arm of our previous study. In retrospect, we may have removed some “desirable difficulties”—aspects that make initial knowledge acquisition more difficult and less appealing, but enhance long-term retention.²⁸

Other online learning studies have differed more significantly in their design. One partially comparable study used an online tutorial to reinforce learning 1 month after a lecture on bioterrorism.²⁹ It found no measurable retention of learning from this educational exercise at either 1 or 6 months. Another study involved an internet tutorial consisting of 40 brief

modules on herbs and dietary supplements, each designed to take about 5 minutes. Among 252 practitioners (including physicians and other health professionals) who completed a follow-up test after 6–10 months, 41% of initial knowledge gains were retained.³⁰ In a final study, 44 primary care physicians spent an average of 2.2 hours on a lipid management tutorial that included online lectures and interactive patient management cases. The tutorial was also reinforced after 1 month by a live, instructor-led web conference. At about 12 weeks after the tutorial, subjects showed a small but statistically significant knowledge gain compared with an immediate posttest.³¹ It is important to note that both of these studies used the same knowledge questions for pre, post, and follow-up tests, making it possible that some of the effects seen were because of the questions increasing in familiarity over repeated administrations. Our study avoided this possibility, whereas keeping the average difficulty of the pretest and posttest equivalent, by randomly splitting the questions for each participant.

One limitation of the current study is that we did not ascertain the service residents were on when they participated. Those assigned to longer delay times could have changed services in the interim, and if their new service were more stressful, their posttest performance could have been differentially affected. All residents, however, who were randomized to longer delay times no doubt had learning experiences in the interim that might have interfered with their recall, regardless of whether they had changed rotations. Another possible limitation is that retention functions following more traditional educational activities, such as lectures and self-directed reading, might differ from those we observed, but most controlled trials have not found online learning to differ from content equivalent activities using other media.^{24,31} Finally, residents also may differ from other physicians in their sleep patterns, and sleep is important for long-term memory consolidation.³²

Table 5. Multivariate Correlates of Posttest Knowledge Scores

Participant characteristic	17-point final knowledge score		20-point all-item score	
	Parameter estimate (SE)	P value	Parameter estimate (SE)	P value
(Days delay after tutorial) ^{1/4}	-1.3 (0.3)	<.0001	-1.1 (0.3)	<.0001
Critical appraisal skill >0	2.0 (0.4)	<.0001	2.2 (0.5)	.0001
Completed any rotation in cardiology	1.0 (0.5)	.04	1.1 (0.6)	.06
Age (yrs)	0.1 (0.05)	.02	0.1 (0.05)	.02
Internal Medicine (versus family medicine)	0.8 (0.5)	.10	0.9 (0.6)	.11

Multivariate linear regression model for final (17-item) posttest knowledge scores; $R^2 = 0.47$. The following participant characteristics were excluded from the model because of a lack of association ($P > 0.20$): postgraduate year, gender, prior endocrinology rotation, computer attitude score, diabetes care self-efficacy score, and pretest score. The same model was arrived at using either a backward or forward stepwise approach.

Such limitations notwithstanding, our findings have several important implications. First, physicians need to understand that one-time educational activities can provide a misleading sense of permanent learning even when, as in the present study, the evaluation of the activity is very positive and the initial gains in learning very substantial. Educators, too, need to be aware that activities producing rapid learning and high learner satisfaction may nonetheless result in poor retention. It is important to introduce “desirable difficulties,” such as making learners reorganize the information to apply it, varying the medical contexts in which key principles and techniques are presented, and spacing, rather than massing, follow-up reinforcement. Our results demonstrate the feasibility of using online methods to deliver exercises that could automate the introduction of such desirable difficulties.

Second, for newly learned knowledge to remain accessible, it needs either to be accessed regularly in one’s medical practice or refreshed as part of an institutional or personal continuing-education program. It is a fundamental fact of human memory that existing knowledge, without continuing access and use, becomes inaccessible in memory, but can be relearned at an accelerated rate. Furthermore, learning and accessing new knowledge accelerates the loss of access to old knowledge, even when that old knowledge remains relevant and useful. Long-term learning is a cumulative process through which new knowledge needs to replace or become integrated with prior knowledge. Given these properties of human memory, the cognitive demands that accompany a physician’s everyday life and a medical world that is rapidly changing, it becomes crucial to optimize both the efficiency of new learning and the reinforcement of valuable old learning.

Finally, policy-makers should make greater investments in research that strengthens the role of education in practice improvement.³³ Improving physicians’ understanding of recommended practices could encourage their adoption, but optimizing instructional practices in continuing education has received little emphasis among quality improvement options. Reliable and research-supported methods for reinforcement might enable education to play a greater role in quality improvement and, given the time that physicians spend in educational activities and the “quality chasm” that persists in health care delivery,³⁴ the need is urgent.

Acknowledgements: *This study was funded by the Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program (San Antonio, TX, USA) and the Diabetes Action Research and Education Foundation (Washington DC, USA). Dr. Mangione received additional support from the National Institute on Aging through the UCLA Resource Center for Minority Aging Research (NIA AG-02-004). Dr. Bazargan received additional support from the National Center for Research Resources (G12 RR 03026-16). We are grateful to Drs. Diana Echeverry, Lisa Skinner, and Mayer Davidson for the assistance with developing educational content and to Drs. Jodi Friedman, Michelle Bholat, Mohammed Farooq, and Nancy Hanna for facilitating our access to residents in their training programs.*

Conflicts of interest: *Dr. Bell has performed consulting for Google. This activity is related to physician education. Mr. Harless is currently employed as a software engineer by the Walt Disney Corporation; his contributions to this manuscript are solely his own and are unrelated to his employment.*

Corresponding Author: Douglas S. Bell, MD, PhD; Division of General Internal Medicine and Health Services Research, Department of Medicine, David Geffen School of Medicine at UCLA, 911 Broxton Avenue, Room 227, Los Angeles, CA 90024, USA (e-mail: dbell@mednet.ucla.edu).

REFERENCES

- Davis D, O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *JAMA*. 1999;282(9):867-74.
- Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005;142(4):260-73.
- Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ*. 1977;52(7):578-81.
- Evans CE, Haynes RB, Birkett NJ, Gilbert JR, Taylor DW, Sackett DL, et al. Does a mailed continuing education program improve physician performance? Results of a randomized trial in antihypertensive care. *JAMA*. 1986;255(4):501-4.
- Marinopoulos SS, Dorman T, Ratanawongsa N, Wilson LM, Ashar BH, Magaziner JL, et al. Effectiveness of Continuing Medical Education. Rockville, MD: Agency for Healthcare Research and Quality; 2007. Evidence Report AHRQ Publication No. 07-E006, January.
- Bjork EL, Bjork RA. On the adaptive aspects of retrieval failure in autobiographical memory. In: Grueneberg MM, Morris PE, Sykes RN, eds. *Practical Aspects of Memory: Current Research and Issues: Vol 1 Memory in Everyday Life*. New York: Wiley; 1988.
- Bjork RA, Bjork EL. A new theory of disuse and an old theory of stimulus fluctuation. In: Healy A, Kosslyn S, Shiffrin R, eds. *From Learning Processes to Cognitive Processes: Essays in honor of William K Estes*. Hillsdale, NJ: Lawrence; 1992.
- Schmidt RA, Bjork RA. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychol Sci*. 1992;3(4):207-17.
- Anderson JR, editor. *Cognitive psychology and its implications*. New York: Freeman; 1995.
- Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*. 1977;84(2):191-215.
- Chumley-Jones HS, Dobbie A, Alford CL. Web-based learning: sound educational method or hype? A review of the evaluation literature. *Acad Med*. 2002;77(10 Suppl):S86-93.
- Cook DA, Dupras DM. A practical guide to developing effective web-based learning. *J Gen Intern Med*. 2004;19(6):698-707.
- Embi PJ, Bowen JL, Singer E. A Web-based curriculum to improve residents' education in outpatient medicine. *Acad Med*. 2001;76(5):545.
- Sisson SD, Hughes MT, Levine D, Brancati FL. Effect of an internet-based curriculum on postgraduate education. A multicenter intervention. *J Gen Intern Med*. 2004;19(5 Pt 2):505-9.
- Cook DA, Dupras DM, Thompson WG, Pankratz VS. Web-based learning in residents' continuity clinics: a randomized, controlled trial. *Acad Med*. 2005;80(1):90-7.
- Bennett NL, Casebeer LL, Kristofco RE, Strasser SM. Physicians' internet information-seeking behaviors. *J Contin Educ Health Prof*. 2004;24(1):31-8.
- Accreditation Council for Continuing Medical Education. ACCME Annual Report Data 2004; 2005.
- van Braak JP, Goeman K. Differences between general computer attitudes and perceived computer attributes: development and validation of a scale. *Psychol Rep*. 2003;92(2):655-60.
- Chin MH, Cook S, Jin L, Drum ML, Harrison JF, Koppert J, et al. Barriers to providing diabetes care in community health centers. *Diabetes Care*. 2001;24(2):268-74.
- Cabana MD, Rand C, Slish K, Nan B, Davis MM, Clark N. Pediatrician self-efficacy for counseling parents of asthmatic children to quit smoking. *Pediatrics*. 2004;113(1 Pt 1):78-81.
- Arauz-Pacheco C, Parrott MA, Raskin P. Hypertension management in adults with diabetes. *Diabetes Care*. 2004;27(Suppl 1):S65-7.
- Haffner SM. Dyslipidemia management in adults with diabetes. *Diabetes Care*. 2004;27(Suppl 1):S68-71.
- Kern DE, Thomas PA, Howard DM, Bass EB. *Goals and Objectives. Curriculum Development for Medical Education: A Six Step Approach*. Baltimore: Johns Hopkins University Press; 1998.

24. Bell DS, Fonarow GC, Hays RD, Mangione CM. Self-study from web-based and printed guideline materials. A randomized, controlled trial among resident physicians. *Ann Intern Med.* 2000;132(12):938–46.
25. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials.* 1988;9(4):327–44.
26. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297.
27. Rubin DC, Wenzel AE. One hundred years of forgetting: a quantitative description of retention. *Psychol Rev.* 1996;103(4):734–60.
28. Bjork RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe J, Shimamura A, eds. *Metacognition: Knowing about Knowing.* Cambridge, MA: MIT; 1994:185–205.
29. Chung S, Mandl KD, Shannon M, Fleisher GR. Efficacy of an educational Web site for educating physicians about bioterrorism. *Acad Emerg Med.* 2004;11(2):143–8.
30. Beal T, Kemper KJ, Gardiner P, Woods C. Long-term impact of four different strategies for delivering an on-line curriculum about herbs and other dietary supplements. *BMC Med Educ.* 2006;6:39.
31. Fordis M, King JE, Ballantyne CM, Jones PH, Schneider KH, Spann SJ, et al. Comparison of the instructional efficacy of internet-based CME with live interactive CME workshops: a randomized controlled trial. *JAMA.* 2005;294(9):1043–51.
32. Gais S, Born J. Declarative memory consolidation: mechanisms acting during human sleep. *Learn Mem.* 2004;11(6):679–85.
33. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med.* 2004;79(10):955–60.
34. Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academy Press; 2001.

APPENDIX

The LOFTS interactive tutorial

This screen-capture image illustrates the operation of the main interactive tutorial portion of the longitudinal online focused tutorial system (LOFTS). The upper frame of the tutorial window shows 1 pretest question at a time with feedback on the user's pretest response shown to the right. Above the question, the relevant learning objective is stated in italics. The

lower frame of the tutorial window contains the guideline document in its entirety with the relevant passages highlighted. In this example, the user had answered the question incorrectly, resulting in feedback that encouraged a careful review of the guideline passages and instructions to then try another answer.

