

TESTING NEW INSTRUCTIONAL METHODS

Interactive Spaced-Education to Teach the Physical Examination: A Randomized Controlled Trial

B. Price Kerfoot, MD EdM^{1,2,3}, Elizabeth G. Armstrong, PhD^{2,3}, and Patricia N. O'Sullivan, MD^{3,4}

¹Veterans Affairs Boston Healthcare System, Boston, MA, USA; ²Harvard Medical International, Boston, MA, USA; ³Harvard Medical School, Boston, MA, USA; ⁴Hebrew Rehabilitation Center, Boston, MA, USA.

BACKGROUND: Several studies have documented that physical examination knowledge and skills are limited among medical trainees.

OBJECTIVES: The objective of the study is to investigate the efficacy and acceptability of a novel online educational methodology termed 'interactive spaced-education' (ISE) as a method to teach the physical examination.

DESIGN: The design of the study is randomized controlled trial.

PARTICIPANTS: All 170 second-year students in the physical examination course at Harvard Medical School were eligible to enroll.

MEASUREMENTS: Spaced-education items (questions and explanations) were developed on core physical examination topics and were content-validated by two experts. Based on pilot-test data, 36 items were selected for inclusion. Students were randomized to start the 18-week program in November 2006 or 12 weeks later. Students were sent 6 spaced-education e-mails each week for 6 weeks (cycle 1) which were then repeated in two subsequent 6-week cycles (cycles 2 and 3). Students submitted answers to the questions online and received immediate feedback. An online end-of-program survey was administered.

RESULTS: One-hundred twenty students enrolled in the trial. Cycles 1, 2, and 3 were completed by 88%, 76%, and 71% of students, respectively. Under an intent-to-treat analysis, cycle 3 scores for cohort A students [mean 74.0 (SD 13.5)] were significantly higher than cycle 1 scores for cohort B students [controls; mean 59.0 (SD 10.5); $P < .001$], corresponding to a Cohen's effect size of 1.43. Eighty-five percent of participants (102 of 120) recommended the ISE program for students the following year.

CONCLUSIONS: ISE can generate significant improvements in knowledge of the physical examination and is very well-accepted by students.

KEY WORDS: educational technology; medical education; medical students; physical examination.

J Gen Intern Med 23(7):973-8

DOI: 10.1007/s11606-008-0533-0

© Society of General Internal Medicine 2008

INTRODUCTION

Knowledge of the physical examination is a critical component of developing effective physical examination skills and of becoming a competent physician. Even so, several studies have documented that physical examination knowledge and skills are quite limited among medical trainees.¹⁻⁵ For example, when over 500 medical students and residents were asked to identify 12 cardiac auscultatory findings, trainees recognized only 20% of the findings, and the number of correct identifications improved little with year of training.² Even though these residents had almost certainly received prior training on how to recognize heart sounds, the trainees did not effectively learn from this prior training and/or did not retain this learning over the course of their training.

The 'spacing effect' has demonstrated promise as a method to improve the acquisition and retention of learning.⁶⁻⁸ The spacing effect is the psychological finding that educational encounters that are spaced and repeated over time (spaced distribution) result in more efficient learning and improved learning retention, compared to massed distribution of the educational encounters (bolus education).⁶⁻⁸ This psychological finding has direct application to medical training. For example, a recent randomized trial compared massed (1 day) versus distributed (weekly) training of surgical residents in microvascular anastomosis skills. Those residents in the weekly training sessions had significantly greater skill retention and were better able to apply these skills on a live, anesthetized rat model.⁹ A distinct neurophysiologic basis for the spacing effect has been identified. A recent study demonstrated that spaced learning by rats improves neuronal longevity in the hippocampus and that the strength of the rats' memories correlates with the number of new cells in this region of their brains.¹⁰

'Spaced education' refers to online educational programs that are structured to take advantage of the pedagogical benefits of the 'spacing effect'.^{6,7} We recently completed a randomized trial of spaced education using daily noninteractive (static) e-mails to 537 urology residents in the United States and Canada.¹¹ In this trial, residents received in-service examination study questions (1) in a bolus format 6 months before their in-service examination or (2) in a spaced-education format consisting of daily e-mails of 1-2 study questions over these 6 months. On a validated test administered at staggered time points to both groups, the spaced-education cohort demonstrated significantly greater knowledge acquisition and retention than those in the bolus cohort.¹¹ Of note, acceptance of the program was excellent, with 95% requesting to participate the following year.

The 'testing effect' also holds promise as a means to bolster retention of learning. Studied since the early 20th century, the

testing effect refers to the psychological finding that initial testing of learned material does not serve merely to evaluate a student's performance, but actually alters the learning process itself to significantly improve retention.^{12,13} The impact of the testing effect was recently demonstrated in a study of 120 college students. Those students who studied a prose passage for 7 minutes and then were immediately tested on the passage over 7 minutes had significantly better retention of the material 1 week later, compared to those students who spent 14 minutes studying the prose in the absence of testing.¹⁴

Interactive spaced-education (ISE) combines the pedagogical merits of both the 'spacing effect' and the 'testing effect.' Instead of delivering static educational material, our new ISE system repeatedly tests and educates students on curricular material over spaced intervals via e-mail. We investigated whether an online ISE program administered concurrently with a year-long physical examination course could improve medical trainees' knowledge of the physical examination.

METHODS

Study Participants

All 170 second-year students enrolled in the Patient-Doctor 2 (PD2; Introduction to Clinical Medicine) course at Harvard Medical School were eligible to enroll in the study. Students were recruited via e-mail announcement in October 2006. Participation was voluntary. There were no exclusion criteria. Faculties at all institutions were blinded to student participation and cohort assignment. Institutional review board approval at Harvard Medical School was obtained.

Development and Validation of the Spaced-Education Materials

Each spaced-education item consists of an evaluative component (a multiple choice question based on a clinical scenario) and an educational component (the answer and explanation). Seventy-seven questions were constructed by PNS and BPK in core physical examination topics areas within the PD2 course. The items were independently content-validated by two PD2 faculty content experts. Seven questions were eliminated due to limited validity, and the remaining 70 questions were divided into two tests which were pilot-tested online by 31 and 32 unique third-year student volunteers. Psychometric analysis of the questions was performed using the Integrity test analysis software (<http://integrity.castlerockresearch.com>; Edmonton, Alberta, Canada). Based on item difficulty, point-biserial correlation, and Kuder-Richardson 20 score, 36 of the questions were selected for inclusion in the ISE program. Six questions were selected from each of 6 core physical examination domains: head/neck, nerves, heart/vessels, thorax/lungs, abdomen/pelvis, and muscles/bones. The educational components of these 36 spaced-education items were then constructed by PNS, and content-validated by BPK and the two PD2 faculty content experts.

Interactive Spaced-Education Online Delivery System

These spaced-education items were delivered to the students via a new interactive spaced-education online delivery system

developed in collaboration with the programmers at the Harvard Medical School Center for Educational Technology. With this new system, students receive spaced-education e-mails at designated time intervals which contain a clinical scenario and question (evaluative component). Upon clicking on a hyperlink in this e-mail, a web-page opens which displays pertinent images and allows the student to submit an answer to the question. Upon downloading this answer to a central server, students are then immediately presented with a web-page displaying the correct answer to the question and an explanation of the curricular learning point (the educational component). By having the provider submit a response before receiving the correct answer and an explanation, this process requires greater interactivity, which educational theory argues may improve learning outcomes. The submitted answers of students were recorded using the MyCourses™ web-based education platform.

Survey Development and Administration

The short answer and Likert-type questions on the end-of-program survey were developed by BPK and content-validated by PNS. Students were asked if they would recommend this spaced-education program to PD2 students next year (yes/no) and whether they would want to participate if similar spaced-education programs on other clinical topics were offered during their ward rotations (yes/no). In addition, students were asked how long each spaced-education e-mail took to complete, what would be the optimal number of e-mails they would want to receive each week, and what would be the optimal number of cycles (repetitions) for the program. On a Likert-type scale, they were also asked to rate the educational effectiveness of the spaced-education program. The survey was constructed and administered online using the SurveyMonkey web-based platform (www.surveymonkey.com; Portland, OR).

Study Design and Organization

This randomized controlled trial was conducted from November 2006 to June 2007. During this time, all enrolled students participated in the year-long Patient-Doctor 2 physical examination course. The students were assigned to 1 of 9 clinical sites where they attended weekly/biweekly 4- to 8-hour sessions. The course was structured to improve students' knowledge of the physical examination through didactic sessions and to allow students to practice and develop their physical examination skills on patients.

Enrolled students were stratified by clinical site and block randomized (block size=4) to one of two starting dates. Students in cohort A started the program on November 6, 2006, while students in cohort B started the program 12 weeks later on January 27, 2007. To take advantage of the educational merits of the spacing effect, the educational material for each cohort was distributed in 3 6-week cycles or repetitions (Fig. 1). During each cycle, students would be sent a daily (Monday-Saturday) interactive spaced-education e-mail, each of which contained a single question. The identical educational material was repeated in each subsequent cycle. For example, the first spaced-education item in cycle 1 was presented in week 1, presented again in week 7 (as a 6-week cycled review), and presented for a final time in week 13 (as a 12-week cycled review). It has been our experience from other trials that the

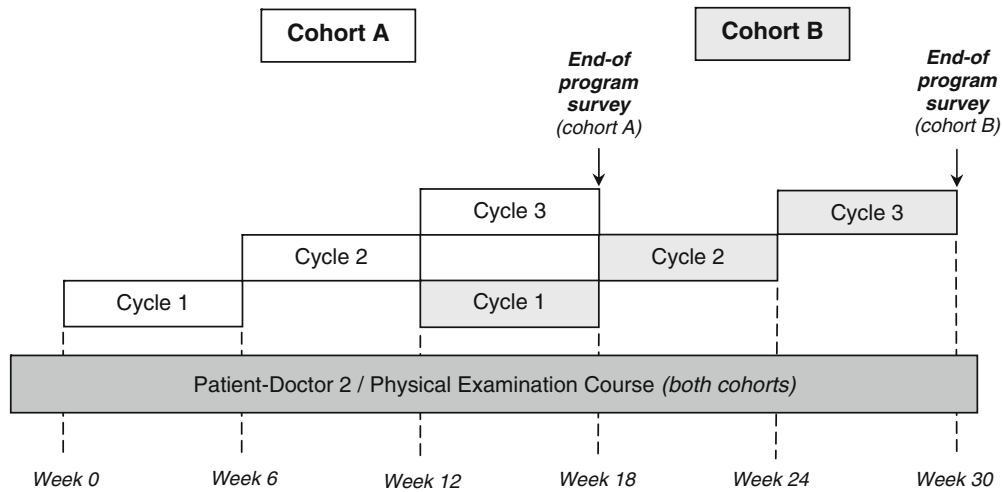


Figure 1. Structure of the randomized controlled trial. To take advantage of the educational merits of the spacing effect, the educational material for each cohort was distributed in three 6-week cycles or repetitions. During each cycle, students were sent a daily (Monday–Saturday) interactive spaced-education e-mail. The 12-week delay in the start of the program for cohort B allowed these students to act as a control group for cohort A: the learning gains of students who had completed 2 cycles of the ISE program in addition to the PD2 course (cohort A) were compared to those students who had only received the PD2 course (cohort B/controls). Thus, this trial structure controlled for students' learning from the PD2 course and allowed the specific learning gains from ISE to be identified.

repetition is not considered burdensome, but rather is seen by the participants as a means to test and reinforce their prior learning^{11,15} Over each 6-week cycle, students would be sent 36 spaced-education e-mails. The entire ISE program ran 18 weeks for students in both cohorts. In an ISE program such as this, evaluation and education are inextricably linked due to the question–answer format of the material. The 12-week delay in the start of the program for cohort B allows these students to act as a control group for cohort A: the learning gains of students who completed 2 cycles of the ISE program in addition to the PD2 course (cohort A) could be compared to those students who only received the PD2 course (cohort B/controls). Thus, this trial structure controls for students' learning from the PD2 course and allows the specific learning gains from ISE to be identified. While this structure enables us to assess the educational efficacy of only two ISE cycles across cohorts (not all three), it allows all of the students in the trial to receive the ISE program.

The end-of-program survey was administered via e-mail to students in each cohort upon completion of cycle 3 (Fig. 1). Upon survey completion and submission of answers to >90% of the spaced-educational e-mails, students received a \$25 gift certificate to an online bookstore.

Outcome Measures

The primary outcome measure was the difference in physical examination knowledge over the same 6-week period (January 27–March 10, 2007) between students who had received 2 cycles of the ISE program (cohort A) and students who had received no prior spaced education (cohort B). The secondary outcome measure was the acceptability of the ISE program to the students. An exploratory analysis was performed to determine the degree to which students who answered a spaced-education item incorrectly in cycle 1 were remediated in cycles 2 and 3 (remedial efficacy).

Statistical Analysis

Students were defined as discontinuing participation if they submitted answers to <80% of spaced-education items. Scores for each cycle were calculated as the number of spaced-education items answered correctly normalized to a percentage scale. An intention-to-treat analysis was performed by including all students for whom baseline data was available (those students who completed the $\geq 80\%$ of items in cycle 1). Percentage scores in cycle 1 were carried forward, if needed, to impute any missing scores from cycles 2 and/or 3. This conservative approach fixed all interval gains in knowledge at zero for those students who did not complete $\geq 80\%$ of questions in cycles 2 and/or 3. A per-protocol analysis was also performed using the data from students who completed $\geq 80\%$ of items in all three cycles. Two-tailed *t* tests were used to test the statistical significant differences in scores between cycles. Intervention effect sizes for learning were measured by means of Cohen's *d*.¹⁶ Cohen's *d* expresses the difference between the means in terms of SD units, with 0.2 generally considered as a small effect, 0.5 as a moderate effect, and 0.8 (and above) as a large effect.¹⁷

Remedial efficacy of the ISE program for both cohorts combined was calculated by first identifying, for each question in cycle 1, those students in the per-protocol dataset who answered that question incorrectly. For these students, their average percentage of correct answers was determined for this same question in cycles 2 and 3. These percentage scores were then averaged for all of the questions in a cycle. Statistical calculations were performed with SPSS for Windows 13.0 (Chicago, IL, USA).

RESULTS

One hundred twenty of the 170 students in the PD2 course enrolled in the trial. The baseline demographic characteristics of the randomized students were similar between cohorts (data

not shown). Cycles 1, 2, and 3 were completed by 88% (106 of 120), 76% (91 of 120), and 71% (85 of 120) students, respectively (Fig. 2). Attrition was similar between cohorts.

Under an intent-to-treat analysis of data from both cohorts combined, the ISE program caused students' baseline scores to increase from a mean 57.9% (SD 10.7) in cycle 1 to 74.4% (SD 13.9) in cycle 3 ($P < .001$), representing a 28.6% relative score increase and a Cohen's effect size of 1.54. Average Cronbach alpha reliability of the 36-item cycles was 0.66 (range 0.47–0.74); with all three cycles combined (108 items), the average reliability rose to 0.86. For our primary outcome measure, the cycle 3 scores for the students in cohort A [mean 74.0 (SD 13.5)] were significantly higher than the cycle 1 scores for the students in cohort B [controls; mean 59.0 (SD 10.5); $P < .001$], corresponding to a Cohen's effect size of 1.43 (Fig. 3). The per-protocol analyses yielded similar results.

Of those aggregated students from both cohorts who answered questions incorrectly in cycle 1, 64.7% (SD 17.8) and 73.4% (SD 14.3) answered the same question correctly in cycles 2 and 3, respectively.

Eighty-six percent of participants (103 of 120) completed the end-of-course survey. Students rated the ISE program to be educationally effective (mean 4.3 [SD 0.7] on a 5-point Likert-type scale, 1=not at all effective, 5=extremely effective). Eighty-five percent (102 of 120) of participants recommended the ISE program to PD2 students the following year, and 83% (100 of 120) wanted to participate themselves in a similar ISE program the following year. Each spaced-education e-mail required a mean 2.7 (SD 1.9) minutes to complete. Students indicated that the optimal number of spaced-education e-mails to receive each week was a mean 4.9 (SD 2.1) and that the optimal number of cycles for the ISE program was a mean 2.7 (SD 0.6).

DISCUSSION

This study demonstrates that an adjuvant ISE program can generate substantial improvements in physical examination knowledge above-and-beyond those generated by a year-long

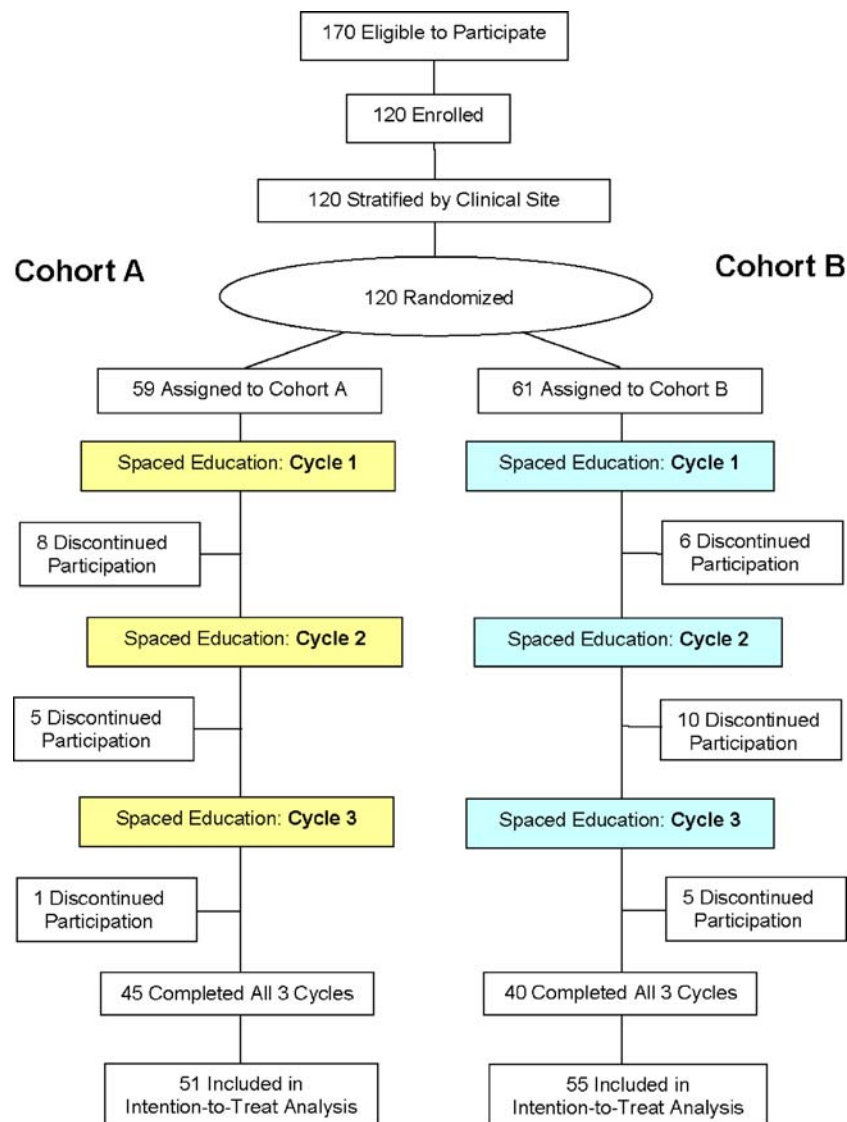


Figure 2. CONSORT flow chart of randomized controlled trial. Students were defined as discontinuing participation if they submitted answers to <80% of spaced-education items in a cycle. Attrition was similar between cohorts.

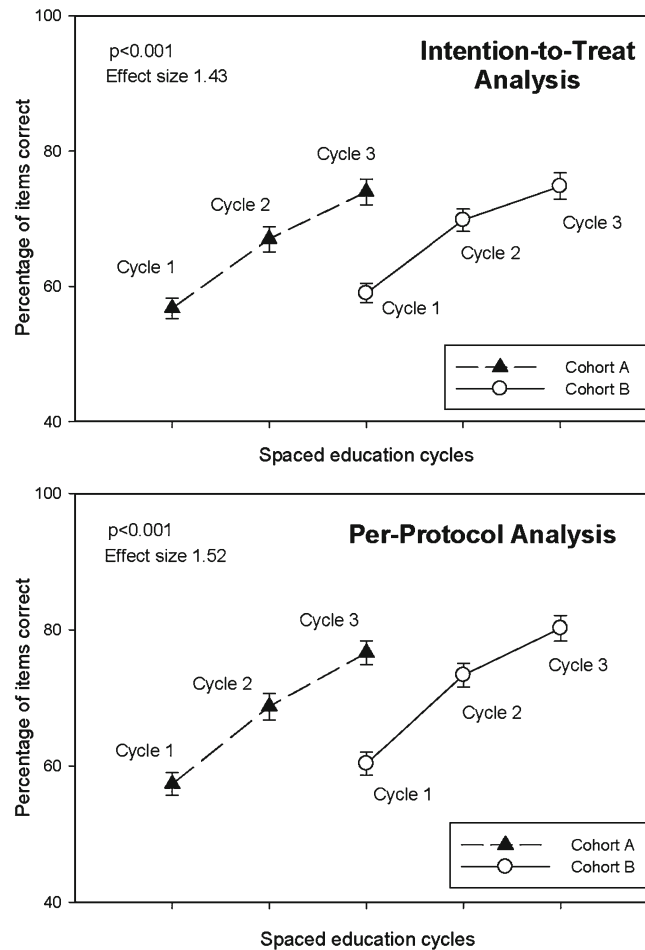


Figure 3. Physical examination knowledge generated by the ISE program. In the intention-to-treat analysis (above), the cycle 3 scores for the students in cohort A (mean 74.0 (SD 13.5)) were significantly higher than the cycle 1 scores for the students in cohort B (controls; mean 59.0 (SD 10.5); $P < .001$), corresponding to a Cohen’s effect size of 1.43. The per-protocol analysis yielded similar results (below). The P values and effect sizes in the diagrams refer to these cross-cohort analyses. Bars represent SE.

physical examination course. In addition, ISE is very well-accepted by students and is an effective form of remediation for those students who were unfamiliar with the material upon initial presentation.

While many studies have documented the dearth of physical examination knowledge and skills among trainees,¹⁻⁵ ISE has the potential to remediate these deficiencies across the spectrum of medical education. Even so, our ISE program is designed as an adjuvant to traditional methods of teaching the physical examination and is not meant to replace the critical interaction between trainees and patients as the trainees develop their physical examination skills. As ISE utilizes traditional web-pages for the submission of answers and for the presentation of learning points, it should be possible to use all of the functionalities of web-pages within the ISE program to meet the training needs of care providers. For example, physician trainees learning how to auscultate the heart can be presented with ISE items which contain an audio recording of an unknown heart sound, and then, trainees can be asked to identify the murmur.

The high acceptance of the ISE program by students is remarkable, especially given the high frequency at which the students received the e-mails. This finding is in stark contrast

to the strong resistance we encountered when conducting a recent trial of web-based teaching modules among 693 medical residents and students.¹⁸ In this trial focusing on systems-based practice competency education, trainees were expected to spend 20 minutes per week over 9 weeks completing web-based teaching modules (interactive web-pages and online narrated slide presentations). Even though trainees reported that the educational content was appropriate, resistance to the program was stiff: one director of a surgical residency program had to threaten his residents with dismissal from the operating room to induce them to complete the modules! In spite of requiring a similar total amount of time by trainees, the ISE program on the physical examination was very well-accepted by students. This high acceptability also likely reflects the ease of use of the spaced-education delivery system, the immediate relevance of the content, and the importance that students attribute to learning the physical examination.

At face value, ISE may appear similar to the ‘programmed instruction’ or ‘programmed learning’ techniques which grew to prominence in the 1960s and 1970s and are grounded in behaviorist psychological theory.¹⁹ While students do work through a structured sequence of question-answer encounters, ISE is not specifically designed to condition the learner

through positive or negative feedback (utilizing the terminology of the behaviorists). However, immediate feedback to the learner must clearly play a role in the learning process. Further research is needed to assess the degree to which behaviorist conditioning may be contributing to ISE's educational efficacy.

There are several limitations to this study, including the fact that it was performed at a single institution and was narrowly focused on a single domain of the medical school curriculum. The limited reliability of each 36-item cycle precludes using the results from a single cycle for high-stakes evaluation of individual students. Even so, the aggregate reliability of the entire program (3 cycles \times 36 items = 108 items) reaches a level ($\alpha=0.86$) which may be appropriate for individual assessment. Cohort cross-over was not specifically assessed but, in our previous studies, has been on the order of 1–5%.^{11,15} The levels of cross-over in the current study are likely even lower, as the ISE delivery program allows for the submission of only one answer per e-mail address. Also, it is not clear from our study which components of the ISE program contributed most significantly to students' learning; future research is needed to isolate the relative impacts of the ISE components (spacing intervals, number of cycles, etc). In addition, our outcome analysis focused on physical examination knowledge, not the higher-order outcome variable of physical examination skills. Future studies are needed to evaluate the impact of ISE on this important downstream outcome measure.

In conclusion, this randomized controlled trial demonstrates that ISE can generate substantial improvements in knowledge, is an effective form of remediation, and is very well-accepted by students. Further work is needed to assess the efficacy of ISE in other content domains and in other trainee populations.

Acknowledgements: We thank Ronald Rouse, Jason Alvarez, and David Bozzi of the Harvard Medical School Center for Educational Technology for the development of the ISE online delivery platform utilized in this trial; Sarah Grudberg and Zaldy Tan for content-validation of the spaced-education items; Arlene Moniz and Karin Vander Schaaf at Harvard Medical International for administrative support; Lippincott Williams & Wilkins for use of images from Bates' Guide to Physical Examination and History Taking (9th Edition) by L.S. Bickley & P.G. Szilagyi; and William Taylor and the Patient-Doctor 2 course faculty for their support of this project.

The views expressed in this article are those of the authors and do not necessarily reflect the position and policy of the United States Federal Government or the Department of Veterans Affairs. No official endorsement should be inferred.

Conflict of Interest: None disclosed.

Funding/Support: This study was supported by Harvard Medical International and the Harvard University Provost's Fund for Innovation in Instructional Technology.

Financial Disclosures: None of the authors have relevant financial interests to disclose.

Author Contributions: Dr Kerfoot had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conception and design: Kerfoot, O'Sullivan

Acquisition of data: Kerfoot

Analysis and interpretation of data: Kerfoot

Drafting of the manuscript: Kerfoot

Critical revision of the manuscript for important intellectual content: Armstrong, O'Sullivan

Statistical analysis: Kerfoot

Obtaining funding: Armstrong

Administrative, technical, or material support: Armstrong

Supervision: Armstrong

Ethical Approval to Perform the Study: The study protocol was approved by the institutional review board at Harvard Medical School.

Corresponding Author: B. Price Kerfoot, MD EdM; Harvard Medical International, One Renaissance Park, 1135 Tremont Street, Suite 900, Boston, MA 02120, USA (e-mail: price.kerfoot@gmail.com).

REFERENCES

1. **Jauhar S.** The demise of the physical exam. *N Engl J Med.* 2006; 354(6):548–51.
2. **Mangione S, Nieman LZ.** Cardiac auscultatory skills of internal medicine and family practice trainees. A comparison of diagnostic proficiency. *Jama.* 1997;278(9):717–22.
3. **St Clair EW, Oddone EZ, Waugh RA, Corey GR, Feussner JR.** Assessing housestaff diagnostic skills using a cardiology patient simulator. *Ann Intern Med.* 1992;117(9):751–6.
4. **Marcus GM, Vessey J, Jordan MV, et al.** Relationship between accurate auscultation of a clinically useful third heart sound and level of experience. *Arch Intern Med.* 2006;166(6):617–22.
5. **Vukanovic-Criley JM, Criley S, Warde CM, et al.** Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study. *Arch Intern Med.* 2006;166(6):610–6.
6. **Glenberg AM, Lehmann TS.** Spacing repetitions over 1 week. *Mem Cogn.* 1980;8(6):528–38.
7. **Toppino TC, Kasserian JE, Mracek WA.** The effect of spacing repetitions on the recognition memory of young children and adults. *J Exp Child Psychol.* 1991;51(1):123–38.
8. **Landauer TK, Bjork RA.** Optimum rehearsal patterns and name learnin. In: Morris PE, Sykes RN, eds. *Practical Aspects of Memory.* New York: Academic Press; 1978:625–32.
9. **Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R.** Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Ann Surg.* 2006;244(3):400–9.
10. **Sisti HM, Glass AL, Shors TJ.** Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learn Mem.* 2007;14(5):368–75.
11. **Kerfoot BP, Baker HE, Koch MO, Connelly D, Joseph DB, Ritchey ML.** Randomized, controlled trial of spaced education to urology residents in the United States and Canada. *J Urol.* 2007;177(4):1481–7.
12. **Spitzer HF.** Studies in retention. *J Exp Psychol.* 1939;30(9):641–56.
13. **Gates AI.** Recitation as a factor in memorizing. *Arch Psychol.* 1917;6(40):1–102.
14. **Roediger HL, Karpicke JD.** Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci.* 2006;17(3):249–55.
15. **Kerfoot BP, DeWolf WC, Masser BA, Church PA, Federman DD.** Spaced education improves the retention of clinical knowledge by medical students: a randomised controlled trial. *Med Educ.* 2007;41(1):23–31.
16. **Cohen J.** *Statistical power analysis for the behavioural sciences* (2nd edition). Hillsdale, NJ: Erlbaum; 1988.
17. **Maxwell SE, Delaney HD.** Designing experiments and analyzing data: A model comparison approach. Belmont, CA: Wadsworth; 1990.
18. **Kerfoot BP, Conlin PR, Trivison T, McMahon GT.** Web-based education in systems-based practice: a randomized trial. *Arch Intern Med.* 2007; 167(4):361–6.
19. **Skinner BF.** The shame of American education. *Am Psychol.* 1984:947–54.