

An Online Spaced-Education Game to Teach and Assess Medical Students: A Multi-Institutional Prospective Trial

B. Price Kerfoot, MD, EdM, Harley Baker, EdD, Louis Pangaro, MD, Kathryn Agarwal, MD, George Taffet, MD, Alex J. Mechaber, MD, and Elizabeth G. Armstrong, PhD

Abstract

Purpose

To investigate whether a spaced-education (SE) game can be an effective means of teaching core content to medical students and a reliable and valid method of assessing their knowledge.

Method

This nine-month trial (2008–2009) enrolled students from three U.S. medical schools. The SE game consisted of 100 validated multiple-choice questions—explanations in preclinical/clinical domains. Students were e-mailed two questions daily. Adaptive game mechanics re-sent questions in three

or six weeks if answered, respectively, incorrectly or correctly. Questions expired if not answered on time (appointment dynamic). Students retired questions by answering each correctly twice consecutively (progression dynamic). Posting of relative performance fostered competition. Main outcome measures were baseline and completion scores.

Results

Seven-hundred thirty-one students enrolled. Median baseline score was 53% (interquartile range [IQR] 16) and varied significantly by year ($P < .001$, $d_{\max} = 2.08$), school ($P < .001$, $d_{\max} = 0.75$), and gender ($P < .001$, $d = 0.38$). Median

completion score was 93% (IQR 12) and varied significantly by year ($P = .001$, $d_{\max} = 1.12$), school ($P < .001$, $d_{\max} = 0.34$), and age ($P = .019$, $d_{\max} = 0.43$). Scores did not differ significantly between years 3 and 4. Seventy percent of enrollees (513/731) requested to participate in future SE games.

Conclusions

An SE game is an effective and well-accepted means of teaching core content and a reliable and valid method to assess student knowledge. SE games may be valuable tools to identify and remediate students who could benefit from additional educational support.

A game can generally be defined as an outcome-oriented activity that proceeds according to a set of rules and often involves focused decision making.¹ Games can range from the fun pastime of solitaire to serious war games involving thousands of military personnel. The 2011 Horizon Report identifies game-based learning as one of six emerging technologies likely to have a large impact on education over the coming five years.² Multiple educational games for medical students have been developed over the last decade, but the rigorous evaluation of their learning outcomes has been limited.³ In a recent systematic review of the effect of educational games on medical students' learning outcomes, Akl and colleagues³ concluded that the data currently available are insufficient to

confirm or refute the utility of games as an effective teaching strategy for medical students.

We created a novel online educational game by incorporating adaptive game mechanics into an evidence-based form of online education, termed "spaced education" (SE). On the basis of two psychology research findings (the spacing and testing effects), SE has been shown in randomized trials to improve knowledge acquisition, boost learning retention for up to two years, and durably improve clinical behavior.⁴⁻⁷ Further research has shown SE to be a reliable and valid method to assess medical student knowledge and to identify students at risk of performing below standard on their United States Medical Licensing Examination (USMLE) Step 1.⁸ SE is currently delivered using periodic e-mails that contain clinical case scenarios and multiple-choice questions. On submitting an answer, the student is immediately presented with the correct answer and an explanation of the topic. The material is then re-presented in a cycled pattern, ranging from 8 to 42 days, to reinforce the content.

We introduced adaptive game mechanics to SE to individualize the pattern of SE reinforcement and content for each student based on his or her performance on the SE questions. For example, a question is repeated in three weeks if answered incorrectly, repeated in six weeks if answered correctly, and retired (no longer repeated) once answered correctly twice in a row. Additional game mechanics include an appointment dynamic (i.e., questions expire if not answered on time) and a progression dynamic (players work toward a specific goal by retiring questions, and new questions are introduced as older questions are retired).⁹ The SE game also fosters competition between students by displaying how other students have answered each question and how many other students have already retired that question.

We hypothesized that the SE game would be (1) an effective means of teaching core content to medical students and (2) a reliable and valid method of assessing medical student knowledge. To investigate these hypotheses, we conducted a prospective trial of the SE

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Kerfoot, Veterans Affairs Boston Healthcare System, 150 South Huntington Ave., 151DIA, Jamaica Plain, MA 02130; telephone: (774) 286-9230; e-mail: price.kerfoot@gmail.com.

Acad Med. 2012;87:1443–1449.

First published online August 21, 2012
doi: 10.1097/ACM.0b013e318267743a

game across nine months at three U.S. medical schools.

Method

Study participants

Approximately 2,200 medical students from three U.S. medical schools were eligible to participate. The F. Edward Hébert School of Medicine at the Uniformed Services University of the Health Sciences is public; Baylor College of Medicine and the University of Miami Miller School of Medicine are both private. All are four-year MD-granting schools whose curricula are structured to include 1.5 to 2 years of preclinical studies followed by clinical clerkships. We recruited participants via e-mail. We did not exclude any interested participants for any reason. Institutional review board approval was obtained to perform this study. Participation was voluntary. Students' participation and performance had no effect on their grades, standing, or promotion. We replaced students' names with coded identifiers prior to analysis of the data.

Development of content

We selected four topics that covered both preclinical (anatomy and histology) and clinical (cardiology and endocrinology) domains. One physician content expert constructed questions for each topic area targeting core information that every medical student should know upon graduation. Two domain experts/educators independently validated the questions for content.⁸ We restricted the question format to multiple-choice because of the limitations of the SE game delivery system. The questions and explanations used in this trial had been constructed and validated for a 2007–2008 SE trial.⁸ We performed psychometric analysis of the questions using the Integrity test analysis software (Castle Rock Research, Edmonton, Alberta, Canada). We selected 25 questions for each topic area based on question difficulty, point-biserial correlation (assessing how well a question discriminates between students of different ability levels), and Kuder–Richardson 20 score (assessing reliability).

Structure of the game

The game used an automated, interactive e-mail system developed at Harvard

Medical School. On clicking a hyperlink in an e-mail, a Web page opened that allowed enrollees to submit an answer to a multiple-choice question. The system randomized the order of possible answer choices at each presentation. The answer was downloaded to a central server, and students were immediately presented with a Web page displaying the educational components: the correct answer, a summary of the curricular learning points, explanations of why the possible answers were correct/incorrect, and hyperlinks to additional educational material. Because of the question–answer format of the items, evaluation and education were inextricably linked. The adaptive game mechanics would repeat questions in three weeks if answered incorrectly and in six weeks if answered correctly. The spacing intervals between repetitions were established based on psychology research findings to optimize long-term retention of learning.^{10,11} If a question was not answered within three weeks of its arrival, it expired, was marked as answered incorrectly, and was cycled back to the student again (appointment dynamic).⁹ If a question was correctly answered twice consecutively, it was retired and not repeated again (progression dynamic).⁹ The goal of the game was to retire all 100 questions. The length of the adaptive SE course thus varied based on each student's baseline knowledge and his or her ability to learn and retain knowledge from the SE question–answer presentations. To foster a sense of competition and community, students received data showing both how other enrollees answered a given question and how many other students had already retired that question.

Study design

We conducted this multi-institutional trial from October 2008 through June 2009. At enrollment, students self-reported their Medical College Admission Test (MCAT) and USMLE Step 1 and Step 2 Clinical Knowledge (CK) scores. All students received two SE questions each day via e-mail according to the adaptive game mechanics (usual range: one to three questions) outlined above. A server error occurred during six days of the trial that caused duplicate questions to be sent to students who had fallen behind and had allowed some questions to expire. At the end of the program,

students completed a short survey which asked for their most recent USMLE Step 1 and Step 2 CK scores and whether they would like to participate in further SE games (yes/no). On retiring $\geq 80\%$ of the questions and completing an end-of-program survey, students received a \$30 gift certificate to an online bookstore.

Scoring and outcome measures

We measured performance on the SE game using both students' (1) baseline scores and (2) completion scores. Baseline scores measured students' pregame knowledge of content and were calculated as the percentage of questions answered correctly on initial presentation. Completion scores were calculated as the percentage of SE questions retired by students. Completion scores reflected students' ability to master the content by answering the questions correctly twice in a row separated by a six-week interval. The educational effectiveness of the SE game would be indicated by the improvement in the completion scores over baseline scores.

Statistical analysis

We estimated power to be ≥ 0.9 for all planned analyses if 400 students completed the trial, assuming a 0.4 effect size and an alpha of 0.05 (with Bonferroni correction). Reliability (internal consistency) of the 100 SE questions on initial presentation (baseline) was estimated with Cronbach alpha.^{12,13} Cohen *d* provided the intervention effect sizes, with 0.2 generally considered as a small effect, 0.5 as a moderate effect, and 0.8 as a large effect.^{14,15} We included those students who submitted at least one answer to all 100 questions and had not received any duplicate questions in the score analysis. We obtained evidence for construct validity by assessing baseline score performance by year of training.

When a student reported different MCAT or USMLE scores pre- and post-trial, we used the pretrial score reported for analysis. To reduce potential errors, we eliminated the self-reported test-score data from four students with extreme/highly improbable MCAT or USMLE scores. In a prior study, we had demonstrated that self-reported MCAT and USMLE Step 1 scores were quite accurate: 96% of students reported MCAT scores within 2 points of their actual

score, and 99% and 98% of students reported, respectively, their Step 1 and Step 2 CK scores within 10 points of their actual score.⁸ We assessed criterion-based validity of the scores via Spearman correlation and partial correlation controlling for MCAT scores. As a result of the limited number of second- and third-year students who reported their Step 1 and 2 scores (respectively) by the trial's end, we were not able to analyze how their game performance predicted their USMLE scores.

Because of the nonnormal nature of the completion scores, we performed univariate analyses of both baseline and completion scores using the Mann-Whitney *U* and Kruskal-Wallis tests. We have reported these nonnormal data as median and interquartile range (IQR). We used chi-square to assess for cohort-level differences in demographic characteristics (e.g., sex, school, prior degrees).

We submitted both baseline and completion scores to a full effect model analysis of variance (ANOVA) that assesses all the main effects and interactions among the independent variables. None of the interaction effects attained statistical significance at the 0.10 level or even demonstrated a small effect size. In the aggregate, they also failed to add significantly to the model fit beyond the main effects. Hence, we used main effect ANOVA. This model, which is insensitive to violations of normality, adjusts for the simultaneous influence of all of the independent variables, providing a more reliable method of determining the relationship between baseline scores and student characteristics.¹⁵ For our analysis of the completion scores, we applied a Welch-James modification to the ANOVA models to adjust for a moderate violation of homogeneity of variance ($d = 0.32$). In these situations, ANOVA models are still appropriate when used with the Welch-James modification.¹⁶ We did not include degree as a variable in the analysis because of the small number of non-MD students. We performed statistical analyses with SPSS 19.0 (Chicago, Illinois).

Results

Demographics

Of the approximately 2,200 students eligible for participation, 731 students from 3 medical schools (33% of eligible

Table 1

Baseline Demographic Characteristics of Medical Students Participating in a Randomized Control Trial Evaluating an Online Spaced-Education Game, 2008–2009*

| Characteristic | Enrolled, no. (% of 731) | Included in analysis, no. (% of 445) |
|--------------------------------|--------------------------|--------------------------------------|
| Medical school attended | | |
| 1 | 248 (34) | 159 (36) |
| 2 | 244 (33) | 143 (32) |
| 3 | 239 (33) | 143 (32) |
| Year of training | | |
| 1 | 225 (31) | 121 (27) |
| 2 | 233 (32) | 143 (32) |
| 3 | 157 (22) | 107 (24) |
| 4 | 116 (16) | 74 (17) |
| Degree | | |
| MD | 697 (95) | 420 (94) |
| MD-PhD | 22 (3) | 16 (4) |
| MD other | 12 (2) | 9 (2) |
| Gender | | |
| Male | 395 (54) | 232 (52) |
| Female | 336 (46) | 213 (48) |

*Those students who submitted at least 1 answer to all 100 questions and had not received any duplicate questions were included in the score analysis. Percentages may not add to 100% because of rounding.

students) volunteered to participate in the trial (Table 1), of whom 77% (561/731) submitted at least 1 answer to all 100 questions. Among these 561 students, 116 (21%) received duplicate questions because of a server error and were excluded from analysis; thus, we performed score analyses on data for the remaining 445 students (61% of all 731 enrollees). The students excluded from analysis did not vary significantly by school, degree, or gender. They did, however, vary significantly by year of training ($P = .03$; see Table 1). The mean age in years of the 731 students enrolled in the study was 25.1 (standard deviation [SD] 2.5), and the mean age in years of the 445 students included in the analysis was 25.3 (SD 3.5). The mean MCAT score of the 731 students enrolled in the study was 32.1 (SD 3.7), and the mean score of the students included in the analysis was 32.3 (SD 3.8).

Baseline scores

Baseline scores measured students' pregame knowledge with a Cronbach alpha reliability of 0.83. Mean alpha was 0.82 (SD 0.02) across medical schools, 0.75 (SD 0.04) across training years, and 0.68 (SD 0.10) within year within medical school.

Overall, median baseline score was 53% (IQR 16). In univariate analysis, baseline scores varied significantly by year of training, ranging from 43% (IQR 11) for first-year medical students to 61% (IQR 12) for fourth-year medical ($P < .001$, Table 2 and Figure 1). We detected no significant difference in scores between students in years 3 and 4 of medical school [$P = .91$]; see Figure 1). Baseline scores also varied significantly by medical school (ranging from school 2 with 49% [IQR 19] to school 1 with 57% [IQR 15]; $P < .001$) and by gender (male 55% [IQR 17] versus female 50% [IQR 16]; $P < .001$). Baseline scores did not vary significantly by academic degree program.

In multivariate analysis with main effects ANOVA, baseline scores varied significantly by year ($P < .001$, $d_{\max} = 2.08$), medical school ($P < .001$, $d_{\max} = 0.75$), and gender ($P < .001$, $d = 0.38$) but not by age ($P = .14$, $d_{\max} = 0.14$). Similar to baseline scores, the difference in completion scores between students in years 3 and 4 was minimal and nonsignificant.

Baseline scores correlated significantly with MCAT, Step 1, and Step 2 CK scores ($r = 0.25, 0.63$, and 0.55 , respectively;

Table 2

Baseline and Completion Scores of 445 Medical Students Participating in a Randomized Control Trial Evaluating an Online Spaced-Education Game, 2008–2009

| Characteristic | Median baseline scores* (interquartile range [IQR]) | P value [†] | Median completion scores [‡] (IQR) | P value [†] |
|--------------------------------|---|----------------------|---|----------------------|
| Medical school attended | | <.001 [§] | | .001 [§] |
| 1 | 57 (15) | | 95 (12) | |
| 2 | 49 (19) | | 92 (12) | |
| 3 | 51 (15) | | 93 (13) | |
| Year of training | | <.001 [§] | | <.001 [§] |
| 1 | 43 (11) | | 85 (17) | |
| 2 | 50 (13) | | 92 (12) | |
| 3 | 59 (12) | | 96 (6) | |
| 4 | 61 (12) | | 97 (5) | |
| Degree | | .64 | | .03 |
| MD | 53 (16) | | 93 (12) | |
| MD-PhD | 50 (9) | | 86 (16) | |
| MD other | 46 (19) | | 84 (18) | |
| Gender | | <.001 [§] | | .01 |
| Male | 55 (17) | | 95 (12) | |
| Female | 50 (16) | | 92 (13) | |

*Baseline scores measured students' pregame knowledge of content and were calculated as the percentage of questions answered correctly on initial presentation.

[†]The P values listed reflect univariate analyses.

[‡]Completion scores, which were calculated as the percentage of spaced-education questions retired by students, reflect students' ability to master the content by answering the questions correctly twice in a row separated by a six-week interval.

[§]Indicates significant results from multivariate analyses.

$P < .001$). When controlling for MCAT scores, baseline scores continued to correlate significantly with Step 1 scores ($r = 0.56$, $P < .001$) and Step 2 CK scores ($r = 0.43$, $P = .001$).

Completion scores

Median completion score was 93% (IQR 12). In univariate analysis, completion scores varied significantly by year of training, ranging from 85% (IQR 17) among year 1 students to 97% (IQR 5) among year 4 students ($P < .001$; see Table 2 and Figure 2). Completion scores also varied significantly by medical school (ranging from a median score of 92% [IQR 12] for school 2 to a median score of 95% [IQR 12] for school 1; $P = .001$), by gender (a median of 95% [IQR 12] for men and a median of 92% [IQR 13] for women; $P = .01$), and by medical degree (ranging from MD other 84% [IQR 18] to MD 93% [IQR 12]; $P = .026$).

In multivariate analysis with main effects ANOVA, completion scores varied significantly by year ($P = .001$,

$d_{\max} = 1.12$), medical school ($P < .001$, $d_{\max} = 0.34$), and age ($P = .019$, $d_{\max} = 0.43$) but not by gender ($P = .65$, $d = 0.11$). Similar to baseline scores, the difference in completion scores between students in years 3 and 4 was minimal and nonsignificant. Completion scores correlated significantly with MCAT, Step 1, and Step 2 CK scores ($r = 0.18$, 0.50, and 0.45, respectively; $P < .001$). When controlling for MCAT scores, completion scores continued to correlate significantly with Step 1 scores ($r = 0.44$, $P < .001$) and Step 2 CK scores ($r = 0.31$, $P = .008$).

End-of-program survey

Seventy-nine percent of the enrolled students (576/731) completed the end-of-program survey. Eighty-nine percent (513/576) of survey respondents (70% of all enrollees) requested to participate in further SE games.

Discussion and Conclusions

To be useful, an instructional or assessment method should be

effective, reliable, and valid. Our study demonstrates that an SE game is an effective means of teaching core content to medical students and is a reliable and valid method of assessing medical student knowledge. Importantly, we also showed that the SE game is well accepted by medical students, with 70% of all enrollees requesting to participate in future SE games. Ideally, an SE game covering a broad range of content domains would be used as one part of an overall student evaluation program to identify lower-performing students who could benefit from additional educational support. Though performance characteristics of the SE game may change when used as a summative rather than formative evaluation and as a compulsory rather than voluntary program, our results indicate that the 100-question game could be used for moderate-stakes decisions for individual students.

These results are consistent with our prior study of SE progress testing (SEPT), in which students at four medical schools received SE *without* adaptive game mechanics.⁸ The results of that study demonstrated SEPT to be a reliable, valid, and educationally valuable method of longitudinal progress testing for medical students. Because the inclusion criteria for analysis differed substantially between the SEPT and SE game studies, we cannot directly compare the results of these two studies. The one randomized trial conducted to date comparing SE with and without game mechanics showed that the adaptive game mechanics significantly increase learning efficiency among medical students by more than 35%.¹⁷ The findings of multiple SE studies indicate that the cycled reinforcement of question content substantially improves long-term retention of material.^{5,8,18}

In the SEPT study, for example, the cycled reinforcement of content (three presentations of the question content) improved retention five months later by 170%, compared with a single presentation of the question content.⁸ Although we can think of no a priori reason the adaptive game mechanics would limit the retention benefits of SE, further research is needed to assess the impact of these mechanics on longer-term educational and behavioral outcomes.

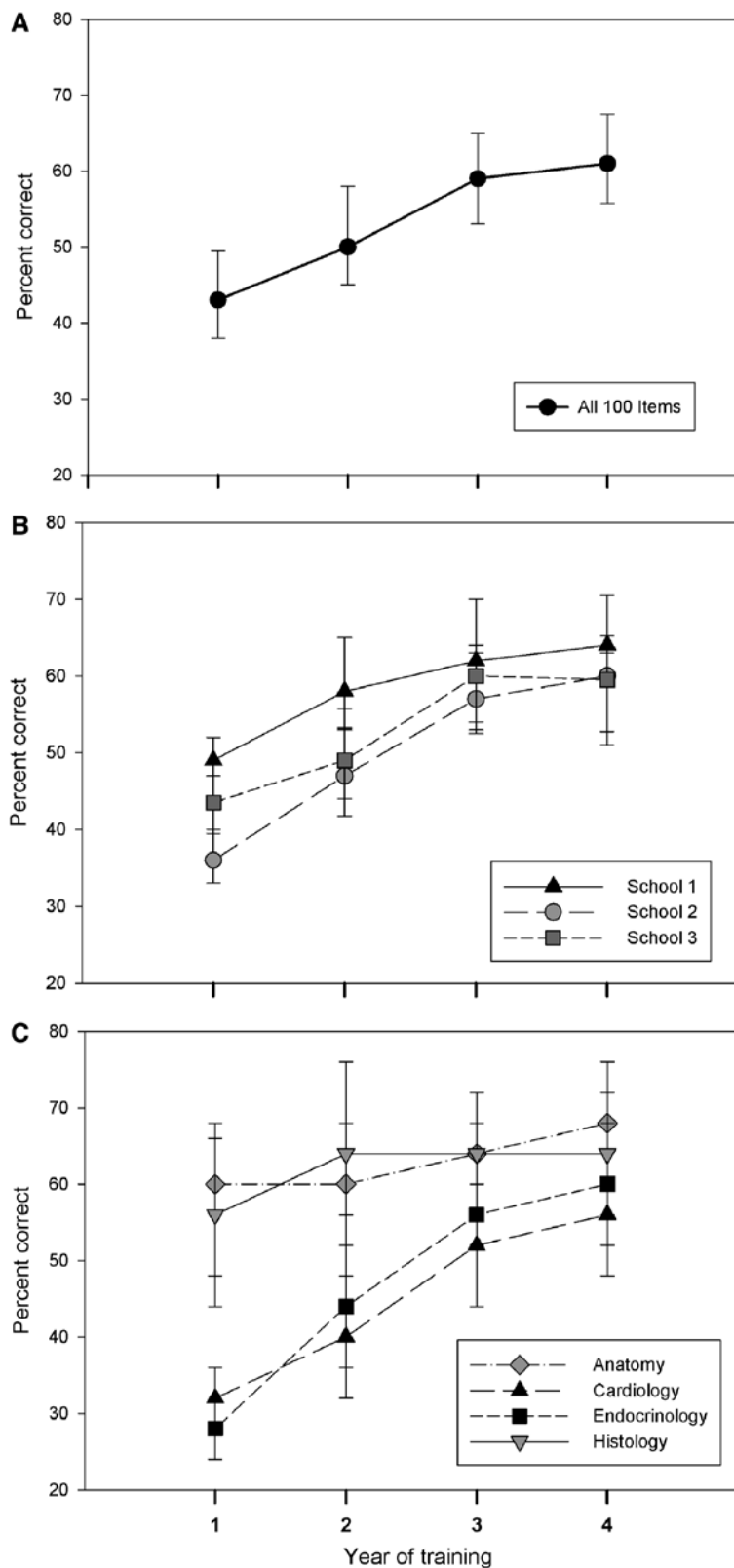


Figure 1 Median baseline scores by year of training. Figure 1A presents the overall scores, Figure 1B presents scores by medical school, and Figure 1C presents scores by topic. Baseline scores measured students' pregame knowledge of content and were calculated as the percentage of questions answered correctly on initial presentation. In 2008–2009, students at three medical schools received 25 validated questions via e-mail in each of the following domains: anatomy, histology, cardiology, and endocrinology (100 questions total). Error bars represent interquartile range.

One benefit of the SE game mechanics is the focus on content mastery (game completion), with students retiring questions by answering them correctly twice in a row separated by a six-week interval. Evidence indicates that SE generates deep learning of the question content, not just memorization of the answers. A randomized trial of 95 primary care clinicians in the northeastern United States demonstrated that SE durably improved their prostate cancer screening by 40% for longer than a year after the SE intervention ended.⁷ Thus, practitioners are not just memorizing answers; rather, they are internalizing the content to optimize their clinical behaviors, consistent with the top of Miller's¹⁹ pyramid.

Evidence for validity of the SE game was obtained both by the strong correlation of students' SE game scores with their USMLE performance and by the increase in scores across years 1 to 3 of medical school. Although our research is limited to only four content domains (cardiology, endocrinology, anatomy, and histology), the fact that SE game scores did not differ significantly between years 3 and 4 adds to the growing evidence that year 4 of medical school may be of limited value in increasing students' fund of knowledge.⁸ The failure of knowledge scores to improve substantially from year 3 to year 4 has now been replicated across seven medical schools.⁸ By suggesting that the elimination of the fourth year of medical school would not significantly reduce the fund of knowledge of graduating students, these data provide empirical evidence to support the recent efforts to introduce flexibility into the duration of medical school.^{20,21}

There are several limitations to this study, including its focus on only four content domains, its use of multiple-choice as the question format, and the server error that required us to discard data from 116 students (16% of 731 enrollees) who received duplicate questions. We have tried to address this last problem by showing that the baseline characteristics of the students in the final sample were not distorted by the loss of the students affected by the server error (Table 1). As a result of the limited number of students in years 2 and 3 who reported their USMLE score by the trial's end, we

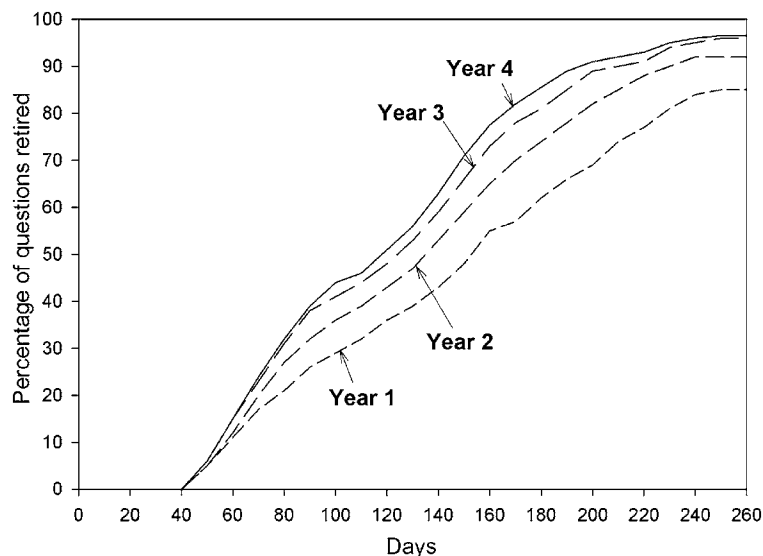


Figure 2 Plot of game progress by year of training. Plots represent the median percentage of questions retired by students in each year of training over the duration of a spaced-education (SE) game (2008–2009). All students received two SE questions each day (usual range: one to three questions) via e-mail according to the adaptive game mechanics. When they clicked on a hyperlink in an e-mail, a Web page opened which allowed enrollees to submit an answer to a multiple-choice question. The answer was downloaded to a central server, and students were immediately presented with a Web page displaying the educational component: the correct answer, a summary of the curricular learning points, explanations of why the possible answers were correct/incorrect, and hyperlinks to additional educational material. The adaptive game mechanics would repeat questions in three weeks if answered incorrectly and in six weeks if answered correctly. If a question was not answered within three weeks of its arrival, it expired, was marked as answered incorrectly, and was cycled back to the student again (appointment dynamic). If a question was correctly answered twice consecutively, it was retired and not repeated again (progression dynamic). The goal was to retire all 100 questions. The length of the adaptive spaced-education course thus varied for each student based on his or her baseline knowledge and ability to learn and retain knowledge from the SE questions.

were not able to analyze how their game performance predicted their USMLE scores. In addition, the participants in this study represent only 33% of the students eligible to enroll, and thus we recommend caution in extrapolating our results to nonparticipating students at these schools. Whereas completion score measures a student's ability to master the game content, this score also reflects several other factors, including the relevance of the content to students' other learning, their baseline knowledge of the content, and the acceptability of the game mechanics and question-answer format. Our study also featured many strengths, including the novelty of the educational intervention, the use of validated content for the game, and the strong generalizability (external validity) of our findings due to the multi-institutional study design.

In summary, our study demonstrates that an SE game is a reliable and valid method to assess student knowledge and is an effective and well-accepted means of teaching core content. As one element

in an overall evaluation program, an SE game may be a valuable tool to identify and remediate students who could benefit from additional educational support.

Acknowledgments: The authors recognize the invaluable work of Ronald Rouse, Jason Alvarez, and David Bozzi of the Harvard Medical School Center for Educational Technology for their development of the spaced-education online delivery platform used in this trial.

Funding/Support: This study was supported by Harvard Medical International, the Harvard University Milton Fund, and the Harvard University Presidential Distance Learning Grant Program.

Other disclosures: Dr. Kerfoot is an equity owner and director of Qstream Inc. None of the other authors have conflicts of interest to disclose.

Ethical approval: The study protocol received institutional review board approval.

Disclaimer: 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, Department of Defense, or Department of Veterans Affairs. No official endorsement should be inferred.

Dr. Kerfoot is associate professor, Surgical Service, Veterans Affairs Boston Healthcare System, Harvard Medical School, Boston, Massachusetts.

Dr. Baker is professor, Department of Psychology, California State University Channel Islands, Camarillo, California.

Dr. Pangaro is professor and chair, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

Dr. Agawal is assistant professor, Department of Medicine, Baylor College of Medicine, Houston, Texas.

Dr. Taffett is Robert J. Luchi, M.D. Chair in Geriatric Medicine and professor, Department of Medicine, Baylor College of Medicine, and chief, Geriatrics Service, Methodist Hospital, Houston, Texas.

Dr. Mechaber is senior associate dean for undergraduate medical education, Leonard M. Miller School of Medicine, University of Miami, Miami, Florida.

Dr. Armstrong is director of educational programs and clinical professor, Harvard Medical School, Boston, Massachusetts.

References

- 1 Salen K, Zimmerman E. *Rules of Play: Game Design Fundamentals*. Cambridge, Mass: MIT Press; 2004.
- 2 Johnson L, Smith R, Willis H, Levine A, Haywood K. The 2011 Horizon Report. Austin, Tex: New Media Consortium; 2011. <http://net.educause.edu/ir/library/pdf/HR2011.pdf>. Accessed May 17, 2012.
- 3 Akl EA, Pretorius RW, Sackett K, et al. The effect of educational games on medical students' learning outcomes: A systematic review: BEME guide no. 14. *Med Teach*. 2010;32:16–27.
- 4 Kerfoot BP. Learning benefits of on-line spaced education persist for 2 years. *J Urol*. 2009;181:2671–2673.
- 5 Kerfoot BP, Armstrong EG, O'Sullivan PN. Interactive spaced-education to teach the physical examination: A randomized controlled trial. *J Gen Intern Med*. 2008;23:973–978.
- 6 Kerfoot BP, Kearney MC, Connelly D, Ritchey ML. Interactive spaced education to assess and improve knowledge of clinical practice guidelines: A randomized controlled trial. *Ann Surg*. 2009;249:744–749.
- 7 Kerfoot BP, Lawler EV, Sokolovskaya G, Gagnon D, Conlin PR. Durable improvements in prostate cancer screening from online spaced education: A randomized controlled trial. *Am J Prev Med*. 2010;39:472–478.
- 8 Kerfoot BP, Shaffer K, McMahon GT, et al. Online "spaced education progress-testing" of students to confront two upcoming challenges to medical schools. *Acad Med*. 2011;86:300–306.
- 9 Schonfeld E. SCVNGR's secret game mechanics playdeck. *TechCrunch*. August 25, 2010. <http://techcrunch.com/2010/08/25/scvng-game-mechanics/>. Accessed May 17, 2012.
- 10 Pashler H, Rohrer D, Cepeda NJ, Carpenter SK. Enhancing learning and retarding forgetting: Choices and consequences. *Psychon Bull Rev*. 2007;14:187–193.
- 11 Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. Spacing effects in learning: A temporal ridge of optimal retention. *Psychol Sci*. 2008;19:1095–1102.

- 12 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
- 13 Pedhazur EJ, Schmelkin LP. *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum; 1991.
- 14 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: L. Erlbaum; 1988.
- 15 Maxwell SE, Delaney HD. *Designing Experiments and Analyzing Data: A Model Comparison Approach*. Belmont, Calif: Wadsworth; 1990.
- 16 Algina J, Olejnik SF. Implementing the Welch–James procedure with factorial designs. *Educ Psychol Meas*. 1984;44: 39–48.
- 17 Kerfoot BP. Adaptive spaced education improves learning efficiency: A randomized controlled trial. *J Urol*. 2010;183:678–681.
- 18 Kerfoot BP, Fu Y, Baker H, Connelly D, Ritchey ML, Genega EM. Online spaced education generates transfer and improves long-term retention of diagnostic skills: A randomized controlled trial. *J Am Coll Surg*. 2010;211:331–337.e1.
- 19 Miller GE. The assessment of clinical skills/ competence/performance. *Acad Med*. 1990;65(9 suppl):S63–S67.
- 20 Cooke M, Irby DM, O’Brien BC. *Educating Physicians: A Call for Reform of Medical School and Residency*. Hoboken, NJ: John Wiley and Sons; 2010.
- 21 Cohen JJ, Hager M, Russell S. Revisiting the Medical School Educational Mission at a Time of Expansion. http://www.josiahmacyfoundation.org/docs/macy_pubs/Macy_MedSchoolMission_proceedings_06-09.pdf. Accessed May 17, 2012.