Web-Based Education in Systems-Based Practice

A Randomized Trial

B. Price Kerfoot, MD, EdM; Paul R. Conlin, MD; Thomas Travison, PhD; Graham T. McMahon, MD, MMSc

Background: All accredited US residency programs are expected to offer curricula and evaluate their residents in 6 general competencies. Medical schools are now adopting similar competency frameworks. We investigated whether a Web-based program could effectively teach and assess elements of systems-based practice.

Methods: We enrolled 276 medical students and 417 residents in the fields of surgery, medicine, obstetrics-gynecology, and emergency medicine in a 9-week randomized, controlled, crossover educational trial. Participants were asked to sequentially complete validated Web-based modules on patient safety and the US health care system. The primary outcome measure was performance on a 26-item validated online test administered before, between, and after the participants completed the modules.

Results: Six hundred forty (92.4%) of the 693 enrollees participated in the study; 512 (80.0%) of the participants completed all 3 tests. Participants’ test scores improved significantly after completion of the first module (P = .001). Overall learning from the 9-week Web-based program, as measured by the increase in scores (posttest scores minus pretest scores), was 16 percentage points (95% confidence interval, 14-17 percentage points; P = .001) in public safety topics and 22 percentage points (95% confidence interval, 20-23 percentage points; P = .001) in US health care system topics.

Conclusion: A Web-based educational program on systems-based practice competencies generated significant and durable learning across a broad range of medical students and residents.

Arch Intern Med. 2007;167:361-366

THE ACCREDITATION COUNCIL FOR GRADUATE MEDICAL EDUCATION (ACGME) now requires that all US residency programs teach and assess their residents in each of 6 general competencies. Many medical schools are now adopting similar competency frameworks. The competency of systems-based practice requires awareness of and responsiveness to the larger context and system of health care and the ability to effectively call on system resources to provide care that is of optimal value. Topics that fall under the domain of systems-based practice include patient safety, health policy, structure of the health care system, health care access, and health care quality. This competency is one of the most challenging to teach and assess because few faculty members have specific expertise in this area, it has traditionally received little attention in medical education, and few valid and reliable assessment tools on this topic have been developed.

Evidence is accumulating that Web-based teaching can be an effective pedagogical tool for delivering and evaluating curricular content across multiple institutions and levels of training. As computers with high-speed Internet access have become ubiquitous in the clinic and at home, dispersed residents and students can easily access Web-based materials regardless of their location and at times that do not conflict with their clinical responsibilities or duty-hour restrictions. In this study, we conducted a multi-institutional randomized, controlled trial to investigate whether a Web-based program could effectively teach and assess elements of systems-based practice in medical students and residents.
col. Residents and students (417 and 276, respectively) from 7 Harvard-affiliated residencies and 2 Harvard Medical School courses were enrolled in the Web-based program between August and October 2005 (Table). All program directors agreed to have their residents and students participate. Participants were free to withhold their data from the research data set.

**DEVELOPMENT OF A VALIDATED CURRICULUM**

Web-based educational materials developed by the Risk Management Foundation (RMF) and the Kaiser Family Foundation (KFF) were selected on the basis of their curricular relevance to the competency of systems-based practice and their perceived educational value for students and residents; each granted permission for the use of their content. Three Web-based educational modules from the RMF were selected to cover topics in patient safety, error prevention, and systems theory. This material is delivered using interactive Web pages that include multiple-choice questions (with answers and explanations), short audio and video clips, and simple animations. Each module takes an average of 23 to 35 minutes to complete. Four Web-based educational modules from the KFF were selected to address topics pertaining to the structure of the US health care system: Medicare, Medicaid, women’s health policy, and the new prescription drug benefit. These modules are delivered online as narrated slide presentations and take 12 to 17 minutes each to complete.15 The content validity of the materials was established by 2 RMF and 4 KFF content experts.

**DEVELOPMENT OF A VALIDATED TEST INSTRUMENT**

Two investigators (B.P.K. and G.T.M.) developed a provisional set of 33 multiple-choice questions (16 on patient safety and 17 on the US health care system) based on the curricular content in the Web-based modules. A panel of content experts at RMF and KFF established content validity of the test items. To determine the psychometric properties of the validated test questions, the 33 items were pilot tested with a group of 18 medical students (years 2-4) and 16 medical residents (postgraduate years 1-3). Point-biserial correlation and Kuder-Richardson 20 calculations were performed for each test item (Integrity Software; Castle Rock Research Corporation, Edmonton, Alberta) and 7 poorly performing items were eliminated to optimize the reliability of the instrument. The resulting validated test contained 26 items: 14 questions on patient safety (PS) and 12 questions on the structure of the US health care system (HS).

**STUDY OBJECTIVES AND DESIGN**

The objectives of this study were to determine whether a Web-based program could effectively teach and assess the competency of systems-based practice, and whether medical students and residents would perceive this online educational program as acceptable and appropriate. The program was constructed with a randomized crossover design to provide an effective control group for measuring initial and overall learning, to estimate the test-retest reliability of the test instrument, and to measure the medium-term retention of the educational content (years 1-3). The curriculum was offered during a 9-week period and comprised 2 sets of topic-specific modules surrounded by a pretest, midtest, and posttest. A pretest covering both topic areas was administered during week 1. During weeks 2 through 4, the first set of Web-based educational materials was distributed (PS materials to group 1 and HS materials to group 2). At week 5, participants completed a midtest covering both topics. During weeks 6 through 8, the remaining set of modules was distributed. At week 9, participants completed the posttest covering both topics. Participants were randomized to the order in which they received the educational modules. Completion of the entire Web-based program of tests and educational modules required 3 to 4 hours.

Hyperlinks to the online tests and educational modules were distributed to the residents and students via weekly e-mail mes-
sages. The test questions were administered online and the test responses were collected online using the SurveyMonkey platform (SurveyMonkey.com, Portland, Ore). The test items and their order on the pretest, midtest, and posttest were identical to increase the reliability of the instrument. Explanations for the test answers were provided after submission of the posttest. Participants self-reported their time spent with the educational materials.

OUTCOME MEASURES

Primary outcome measures were the change in the test scores in each of the 2 topic areas from the pretest to the midtest (initial learning), the stability of the topic-specific test scores between the midtest and the posttest (retention), and the overall educational efficacy as judged by the change in score from the pretest to the posttest (overall learning). The mean differences in scores of those in the control group were subtracted from those in the intervention group to calculate the improvement in scores attributable to the Web-based curriculum on that topic.

Secondary outcome measures included the subjective perceptions of the residents and students of the acceptability of the Web-based educational program and the appropriateness of the educational content for their level of training.

STATISTICAL METHODS

The reliability of the validated test instrument was determined with Cronbach α, a measure of internal consistency. In addition, the averaged 4-week test-retest reliability (indicating stability of the measurement over time) was estimated with a Spearman-Brown adjustment.

Although only 206 subjects would have been required for a 0.9 power to detect a 10% difference in learning, we recruited a larger sample to assess the generalizability of the Web-based program across different institutions, specialties, and levels of training. Enrollees were stratified by program/course and year of training and then block randomized at a single time point by 1 investigator (B.P.K.) between 2 groups. Program directors were blinded to group assignment. Participation in the study was defined as submission of baseline data and completion of the pretest. To allow a conservative intention-to-treat analysis, these baseline data on participants were carried forward, if needed, to impute any missing midtest and/or posttest data; this fixed gains in knowledge at zero for those subjects who did not complete the midtest or posttest. Completion of the Web-based program was defined as submission of all 3 tests.

Overall and topic-specific test scores were normalized to a percentage scale, with a minimum score of 0% and a maximum of 100%. Two-tailed t tests and Wilcoxon rank sum tests were used to determine the statistical significance of changes in learning. Intervention effect sizes for learning were measured by means of Cohen d, which was calculated by dividing mean scores or score increases by pooled standard deviations. Cohen d expresses the difference between the means in terms of standard deviation units, with 0.2 generally considered a small effect, 0.4 considered a moderate effect, and 0.8 considered a large effect. Because the intention-to-treat structure might bias in favor of improved retention by including participants with imputed midtest scores, a secondary analysis of retention was performed in the subset of participants who completed both the midtest and the posttest. Potential associations between treatment effects and subject characteristics were examined by graphical and tabular exploration and formally assessed by multiple linear regression analyses. Statistical calculations were performed with Stata 9.0 (StataCorp, College Station, Tex) and SPSS for Windows 13.0 (SPSS Inc, Chicago, Ill) statistical software.

RESULTS

The 693 enrolled residents and students who were randomized to groups 1 and 2 (344 and 349 enrollees, respectively) were similar with respect to a wide range of characteristics (Table). Six hundred forty (92.4%) of the enrollees participated in the study, and 512 (80.0%) of the participants completed all 3 tests. Of the participants, 128 (20.0%) did not complete the midtest (99 [15.5%]) and/or the posttest (70 [10.9%]). Dropout rates for the different groups were comparable (Figure 2). No enrollees in group 1 and two enrollees in group 2 elected to remove their data from the research data set and were included in those designated as having declined to participate. Under an intention-to-treat analy-
The Cronbach α for the 26-item online test instrument was 0.76 (posttest). Four-week test-retest reliability was 0.63. Of 539 responding participants, 453 (84.0%) completed at least 1 of the PS modules, and 372 (69.0%) completed all 3. Of 553 responding participants, 473 (85.5%) completed at least 1 of the HS modules, and 357 (64.6%) completed all 4. Minimal crossover between groups was reported: 52 (9.6%) of 541 responding participants reported completing 1 or more of the alternate modules before taking the midtest.

INITIAL LEARNING

Mean (SD) pretest scores were 58% (16%) and 47% (14%) in the PS and HS topics, respectively. Pretest scores were comparable between groups (Figure 3). Topic-specific midtest scores were significantly higher for those who received the educational intervention on that topic (P < .001 in each topic, 2-tailed t test). Wilcoxon rank sum tests, which are less susceptible to departures from normality (found to be mild in the data), produced identical results. Increases in initial learning attributable to the Web-based curricula were 14 (95% confidence interval [CI], 11-16) and 20 (95% CI, 17-23) percentage points in the PS and HS topic areas, representing relative increases of 24% and 43% over PS and HS pretest scores, respectively. These changes correspond to Cohen’s d effect sizes of 0.75 (95% CI, 0.59-0.91) and 0.95 (95% CI, 0.79-1.11). A multivariate regression model controlling for the effects of program type, year of training, degree, sex, specialty, and age produced no significant changes in intervention effects.

RETENTION

Group 1 participants displayed strong retention of the PS curriculum, with no significant change in their topic-specific scores from midtest to posttest (mean change, 1% [95% CI, 0%-2%]; P = .10). A small but statistically significant decay was seen in the HS score in group 2 participants from midtest to posttest (mean change, −3% [95% CI, −5% to −2%]; P < .001), which represented a 4% relative decline in HS test scores. A secondary analysis of retention in the subset of participants who completed both the midtest and the posttest demonstrated similar results. Adjusting both models for participant covariates did not alter these retention findings.

OVERALL LEARNING

The overall mean increases in scores (posttest scores minus pretest scores) during the entire 9-week program were 16 (95% CI, 14-17; P < .001) and 22 (95% CI, 20-23; P < .001) percentage points in the PS and HS topic areas, representing relative increases of 28% and 47% over pretest scores, respectively. These changes correspond to effect sizes of 1.00 (95% CI, 0.84-1.16) and 1.22 (95% CI, 1.08-1.32), respectively. These results did not change in models that controlled for participant covariates.

ACCEPTABILITY AND APPROPRIATENESS OF THE WEB-BASED PROGRAM

Participants were asked to rate on a 5-point Likert-type scale the acceptability of the online program as a means of fulfilling their competency education requirements in systems-based practice (in the scale, 1 indicated not at all acceptable; 5, very acceptable). The participants’ median rating was 3. When asked to similarly rate the degree to which the content in the online modules was appropriate to their level of training (on a 5-point Likert-type scale in which 1 indicated not at all appropriate; 5, very appropriate), the participants’ median rating was 4 for both the PS and HS modules.

COMMENT

This multi-institutional randomized, controlled trial demonstrates that a Web-based program generated significant and durable learning in the competency of systems-based practice, one of the more challenging competency areas to teach and assess. Our results confirm that Web-based teaching may be an effective method for delivering and assessing curricular material of this type across a wide range of medical specialties, institutions, and levels of training.
Previous efforts to teach elements of systems-based practice have included in-house workshops, interdisciplin ary learning groups, independent study projects, large-group collaborative projects, Web-broadcast workshops, large-group lectures, community projects, outcomes cards, simulations, and team competitions. The reports cited have been small or specialty-focused, they often fail to validate or assess the reliability of their evaluation instruments, and their use of control groups to assess programmatic efficacy is infrequent. In contrast, we tested the efficacy of a Web-based program across a wide spectrum of trainees, used a carefully designed and tested evaluation instrument, and set up a controlled crossover design to assess the initial learning attributable to each module and the retention of that learning. In this setting we showed significant and durable increases in learning.

No definitive answer is available as to whether a Web-based program to teach and assess systems-based practice is “good enough” to fulfill the ACGME competency requirements because valid standards for competency in this domain have yet to be established. Ideally, these competency standards would be defined and assessed on the basis of actual trainee practice within relevant systems. We have not set a competency standard in our knowledge-based test, nor do we suggest that performance in such a test should be sufficient for competency. Nevertheless, this study demonstrates that knowledge deemed important by content experts in the field can be effectively learned by using Web-based education. This model could be readily implemented across a broad range of programs and institutions.

Participants rated the acceptability of the Web-based program as neutral and rated the content as appropriate. This neutral level of acceptance may reflect a cultural rift between the competency-based educational agenda promoted by the ACGME and the pressing educational needs as perceived by the trainees. Until trainees appreciate the clinical relevance of systems-based practice competencies, educational programs in this domain may be perceived as unwelcome training requirements.

Several factors should be considered in the interpretation of our findings. Learning outcomes assessed by multiple-choice questions cannot supplant practice-based measures of trainee performance. While the test’s reliability precludes making high-stakes decisions based on trainees’ scores, this reliability level is not unexpected given the brevity of the test (26 items) and the disparate nature of the topics covered. Although the participant dropout rate of 20% is well within acceptable standards for educational studies, the possibility of some degree of dropout bias cannot be excluded. Although the identical test was used at 3 time points to maintain assessment reliability, the inclusion of a control group enabled us to confirm that test score increases in the intervention groups were directly due to the educational program, not merely due to priming from prior knowledge of the test questions. Strengths of this study include the randomized, controlled study design; the methodologic rigor of the test development; and the inclusion of trainees from a wide range of specialties, institutions, and levels of training.

In summary, this multi-institutional randomized, controlled trial establishes the principle that Web-based programs can be effectively used across a wide range of medical specialties, institutions, and levels of training to generate substantial learning and retention in the competency of systems-based practice. Further work is needed to establish valid standards for competency in systems-based practice and to promote acceptance of competency education in this domain at the trainee level.

Accepted for Publication: November 15, 2006.

Correspondence: B. Price Kerfoot, MD, EdM, Veterans Affairs Boston Healthcare System, 150 S Huntington Ave, 151DIA, Jamaica Plain, MA 02130 (price.kerfoot@gmail.com).

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. Study concept and design: Kerfoot, Conlin, Travi son, and McMahon. Acquisition of data: Kerfoot and McMahon. Analysis and interpretation of data: Kerfoot and Travi son, and McMahon. Drafting of the manuscript: Kerfoot and Travin son. Critical revision of the manuscript for important intellectual content: Conlin and McMahon. Statistical analysis: Kerfoot and Traversal. Obtained funding: Kerfoot and Conlin. Administrative, technical, and material support: Conlin, Travi son, and McMahon. Study supervision: Conlin and McMahon.

Financial Disclosure: None reported.

Funding/Support: This study was supported by a grant from the RMF. Additional support was obtained from the Research Career Development Award Program and by research grants TEL-02-100 and IIR-04-045 from the Veterans Affairs Health Services Research & Development Service, grants from the American Urological Association Foundation and Astellas Pharma US, Inc, grants K24 DK63214 and R01 HL77234 from the National Institutes of Health, and grants from the Academy at Harvard Medical School.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position and policy of the US government or the Department of Veterans Affairs. No official endorsement should be inferred.

Acknowledgment: We thank the RMF and the KFF for use of their Web-based educational materials; Robert B. Hanscom, JD, and Elizabeth G. Armstrong, PhD, for their support of the program; Lu cean L. Leape, MD, and Saul N. Weingart, MD, for editing and content validation of the PS test items; Alina Salganicoff, PhD, Juliette Cubanski, MPP, MPH, Caya Lewis, MPH, Tricia Neuman, ScD, and Usha Ranji, MS, of the Kaiser Family Foundation for editing and content validation of the HS test items; Ronald A. Arky, MD, Stanley W. Ashley, MD, Christopher C. Baker, MD, Eugene Beresin, MD, Lori R. Berkowitz, MD, Charlie M. Ferguson, MD, Joel T. Katz, MD, Hope A. Ricotti, MD, William Taylor, MD, and Carrie D. Tibbles, MD, for including their programs and courses in the Web-based program; Daniel D. Federman, MD, for support in the conception of the program and assistance in its financial administration; and Susan Herlihy, Jessica E. Hyde, and Colleen E. Graham for administrative support.