Audience Response System: Effect on Learning in Family Medicine Residents

T. Eric Schackow, MD, PhD; Milton Chavez, MD; Lauren Loya, MD; Michael Friedman, MD

Background and Objectives: *The use of an electronic audience response system (ARS) that promotes active participation during lectures has been shown to improve retention rates of factual information in nonmedical settings. This study (1) tested the hypothesis that the use of an ARS during didactic lectures can improve learning outcomes by family medicine residents and (2) identified factors influencing ARS-assisted learning outcomes in family medicine residents.* Methods: *We conducted a prospective controlled crossover study of 24 family medicine residents, comparing quiz scores after didactic lectures delivered either as ordinary didactic lectures that contained no interactive component, lectures with an interactive component (asking questions to participants), or lectures with ARS.* Results: *Post-lecture quiz scores (maximum score 7) were 4.25 ± 0.28 (61% correct) with non-interactive lectures, 6.50 ± 0.13 (n=22, 93% correct) following interactive lectures without ARS, and 6.70 ± 0.13 (n=23, 96% correct) following ARS lectures. The difference in scores following ARS or interactive lectures versus non-interactive lectures was significant (P<.001). Mean quiz scores declined over 1 month in all three of the lecture groups but remained highest in the ARS group. Neither lecture factors (monthly sequence number) nor resident factors (crossover group, postgraduate training year, In-Training Examination score, or post-call status) contributed to these differences, although post-call residents performed worse in all lecture groups.* Conclusions: *Both audience interaction and ARS equipment were associated with improved learning outcomes following lectures to family medicine residents.*

(Fam Med 2004;36(7):496-504.)

Despite impressive advances in medical knowledge over the past century, the process of delivering medical information to postgraduate medical trainees (residents) remains largely unchanged, relying primarily on didactic lectures. However, lectures in which audience members remain passive participants in the learning process yield disappointingly low retention rates of factual information.[1] Medical trainees fail to retain significant percentages of essential teaching points presented during the course of a traditional lecture, and retention worsens further with both increases in "information density" and the passage of time after lecture delivery.[2]

Recently, the development of compact electronic wireless audience response systems (ARS) has allowed for increased audience participation during lectures. An ARS is comprised of a handheld radio-frequency response or "voting" keypad for each audience member, a radio-frequency base station connected to a laptop computer for the lecturer, and software that manages the communication process, vote tallying, and real-time results display. When using ARS during lectures, the lecturer poses multiple-choice questions to the audience members, who answer the questions on their keypads, after which votes are tallied and displayed to the audience.

The use of a modern ARS to promote active participation in the lecture process has been shown to improve retention rates of factual information in general educational (nonmedical) settings.[3] Increasingly, such a tool is being used in medical education settings in an effort to realize similar benefits among medical trainees.[4,5] However, formal evaluations of ARS outcomes in medical education have been few. Published studies have generally been limited to self-reports and observational data demonstrating positive attitudes toward ARS by both audiences and instructors, while more-rigorous assessment of actual learning outcomes has not yet been described.[6-10]

From the Family Practice Residency Program, St Elizabeth Hospital, and the Department of Family Medicine, University of Illinois at Chicago.

The present study (1) prospectively determined whether the use of an ARS during lectures can improve learning outcomes in a particular group of postgraduate medical trainees (family medicine residents) and (2) if improvement occurred, explored factors that might influence or account for this benefit.

Methods

Lectures

We conducted a prospective controlled crossover study between May 2002 and January 2003. The study involved 24 family medicine residents from a community-based, university-affiliated family medicine residency training program in Chicago. Each month, a 1-hour lecture topic from the family medicine residency midday core didactic lecture series was selected for inclusion in the study.

This lecture was then presented twice during the month. The first delivery of the lecture was in a traditional format—referred to as either a "basic" lecture if the audience was not given the opportunity to interact with the speaker and was not presented with multiple-choice questions throughout the lecture (May-August) or as an "interactive" lecture if the audience verbally interacted with the speaker by being presented with multiple-choice questions during the course of the lecture (September–January). The second delivery of the lecture was an ARS lecture, in which the audience physically interacted with the speaker by being presented with multiple-choice questions for which each audience member anonymously "voted" for a correct answer using their ARS keypad.

Each lecture pair was delivered by the same faculty lecturer using presentation software (PowerPoint 2000, Microsoft Corporation, Bellevue, Wash) in an identical small conference room equipped with a laptop computer and a digital projector. A total of four different family medicine faculty members delivered eight monthly paired lectures (16 lectures total) throughout the entire study. Slides were identical in the paired lectures, except that each interactive and ARS lecture included seven additional slides that contained multiple-choice questions interspersed throughout the lecture, with correct answers to each question discussed as part of the presentation. Multiple choice question slides were arranged in pre-test fashion, preceding the lecture slides containing the relevant content.

ARS lectures made use of a basic ARS setup (RSVP hardware with Connect 1.0 software, Meridia Audience Response, Plymouth Meeting, Pa), which included the laptop computer running PowerPoint, a radio-frequency interface box connected to the laptop computer, up to 20 wireless audience response keypads, and computer interface/display software. Audience votes were automatically tallied and displayed in summary histogram format after presentation of each multiple-choice question.

Subject Assignment

Family medicine residents were initially assigned according to first alphabet letter of their last name (A–J versus K–Z) to either the control group (either basic or interactive lecture, depending on the month) or to the experimental (ARS lecture) group. Residents subsequently crossed over between basic/interactive and ARS groups each successive month in an attempt to ensure equal participation by each resident in all lecture methods. This crossover assignment scheme is depicted in detail in Figures 1A and 1B.

Numbers of attendees at each lecture ranged from 8 to 15. The majority of attendees were family medicine residents, while the remainder were medical students and attending faculty physicians. Separately tabulated responses for each of these non-resident groups are not included in this study. Occasionally, an individual resident during a given month had to switch from his/her preassigned group to the opposite group because of a scheduling conflict, leading to small deviations from the expected overall basic:ARS:interactive lecture attendance ratio of 1:2:1. Nonetheless, by the conclusion of the study, residents had achieved actual attendance ratios of 1.33 to 2.08 to 0.92.

Data Collection

At the conclusion of each lecture, a 10-question multiple-choice quiz (five choices per question, single correct answer) was immediately administered to residents participating in the lecture (initial quiz). Seven of the 10 questions were based on content displayed in that lecture's PowerPoint slides (lecture-related questions). These seven questions were essentially identical to the seven questions displayed throughout the interactive and ARS lectures as described above; however, these questions were not displayed during the basic lectures. The remaining three questions were control queries on general medical information unrelated to lecture content (lecture-unrelated questions). A new set of three lecture-unrelated questions was selected every month, adapted from continuing medical education (CME) multiple-choice questions published in past issues of the *American Family Physician*. Residents were also asked to specify whether they were "post-call" at the time of lecture attendance; those who indicated that they were would have been at work continuously for 28–31 hours at the time of lecture participation.

Readministration of the same quiz was performed 1 month later to assess durability of responses (1-month follow-up quiz). There was some attrition in 1-month follow-up quiz numbers because of residents being unavailable at the appropriate time.

Quizzes were scored and individual question responses were entered into a database (Access 2000, Microsoft Corporation, Bellevue, Wash). The total numbers of correct responses in various subgroups of inter-

Table 1

Quiz Scores by Lecture Type

Lecture Type		*Quiz Score (max. 7)*	*n*	P *Value (1-month Differences)*	P *Value (Versus Other Lecture Types)*
Basic	Initial	4.25 ± 0.28	32	} .05	—
	1-month follow-up	3.39 ± 0.33	18		—
ARS*	Initial	6.70 ± 0.13	23	} < .001	< .001 (versus basic initial)
	1-month follow-up	4.67 ± 0.45	12		< .05 (versus basic 1 month)
ARS**	Initial	6.56 ± 0.19	27	} < .001	—
	1-month follow-up	5.07 ± 0.34	14		—
Interactive	Initial	6.50 ± 0.13	22	} < .001	< .001 (versus basic initial) .31 (versus ARS** initial)
	1-month follow-up	4.22 ± 0.37	18		.11 (versus basic 1 month) .11 (versus ARS** 1 month)

ARS—Audience Response System

* ARS lectures from May 2002–August 2002 (Figure 1A).
** ARS lectures from September 2002–January 2003 (Figure 1B).

est were tabulated using custom-designed database queries.

Statistical Methods

Quiz score data described above were imported into statistical software (SPSS 11.5, SPSS Science, Chicago) that was used for data analysis. Data values are presented as means ± standard error of the mean (SEM)s.

Reliability analysis by calculation of Cronbach-alpha coefficients was retrospectively performed on the seven lecture-related questions from each of the eight quizzes. This permitted an assessment of internal reliability of the measuring instrument (ie, the quizzes).

Post-lecture quiz scores from basic/interactive and ARS lecture groups were compared, both initially and 1 month after lecture administration of the quizzes. For the analyses depicted in Figures 2 and 3 and in Table 1, ARS score data were "segregated" into two separate ARS groups corresponding to the first and second halves of the study; each ARS group was compared only against its respective control group (either basic or interactive; see Figures 1A and 1B). However, for the analyses depicted in Figures 4–5 and Tables 2–5, ARS score data from both halves of the study were combined into a single ARS group (designated "pooled" in the text). To assess whether lecture style could have had an unintended effect on quiz performance independently of lecture content, results of the three-question control queries on general medical information (lecture-unrelated questions) were compared in basic, interactive, and ARS lecture groups, both initially and 1 month after lecture administration, using one-way analysis of variance (ANOVA) (Table 2).

Differences in quiz scores between pairs of subgroups were assessed using independent sample *t* tests (Figures 2, 3, and 5), while differences between three or more subgroups were assessed using one-way ANOVA (Table 2). In instances where quiz score data demonstrated a non-normal quiz score distribution, the Mann-Whitney Rank Sum Test or Kruskal-Wallis One-Way ANOVA on Ranks was performed instead. Multiple linear regression analysis was performed to evaluate the interactions of quiz scores with lecture sequence, group assignment, and postgraduate year (PGY) of training (Table 3). Single linear regression analysis was separately performed for In-training Exam (ITE) score correlation (Figure 4, Table 4).

Finally, to evaluate whether post-call status might have influenced residents' performance, *t* tests were used to compare scores of residents who were and were not post call in each lecture group (basic, ARS, or interactive) (Figure 5, Table 5).

Figure 1

Overview of Study Design—Lecture Types, Resident Assignments, and Crossover Protocol are Detailed for May–August 2002 (Basic Versus ARS Lectures) (Figure 1A) and September 2002–January 2003 (Interactive Versus ARS Lectures) (Figure 1B)

Figure 1A

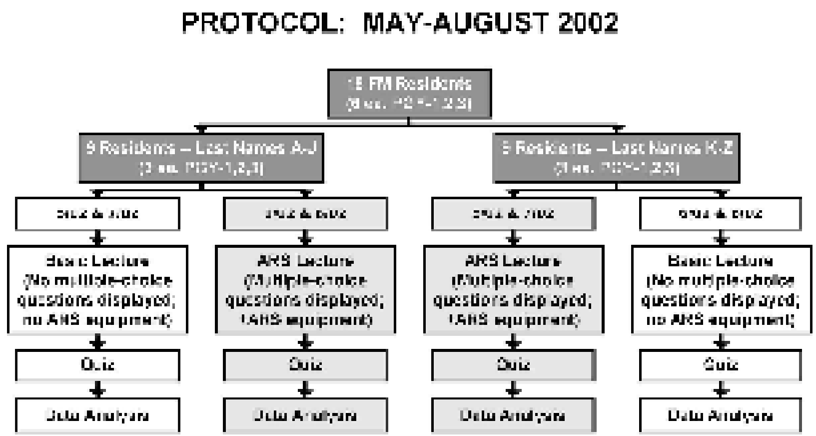
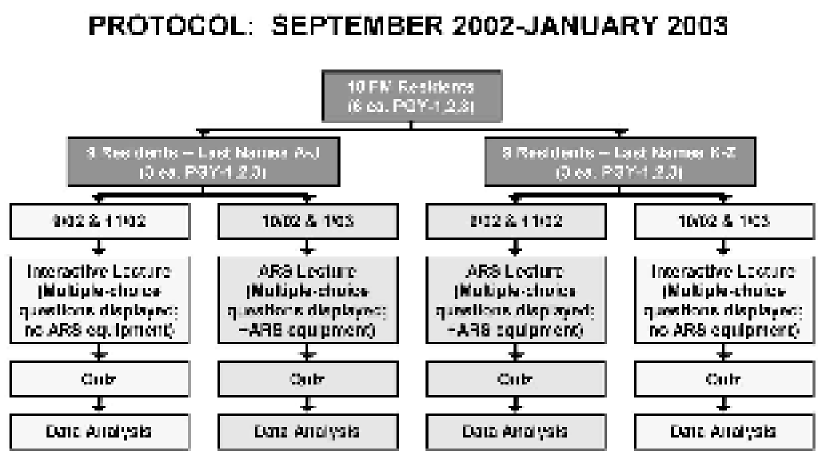


Figure 1B



ARS—Audience Response System

Results

Reliability Analysis

The mean Cronbach-alpha value for the eight monthly quizzes, calculated from a composite of all initial and 1-month-follow-up lecture-related responses each month, was 0.62 ± 0.05 (median=0.62, range=0.42–0.86, n=eight quizzes).

Basic Versus Interactive Versus ARS Lecture Performance

Mean initial interactive quiz scores were significantly higher than initial basic scores (Table 1). At 1-month follow-up, however, the difference between interactive and basic lecture groups did not achieve statistical significance. One-way ANOVA revealed no differences between six groups for the lecture-unrelated questions (Table 2).

Mean quiz scores in the ARS lecture group were significantly better than those in the basic lecture group, both initially and at the 1-month follow-up (Figure 2, Table 1). Quiz scores declined significantly in both groups over 1 month.

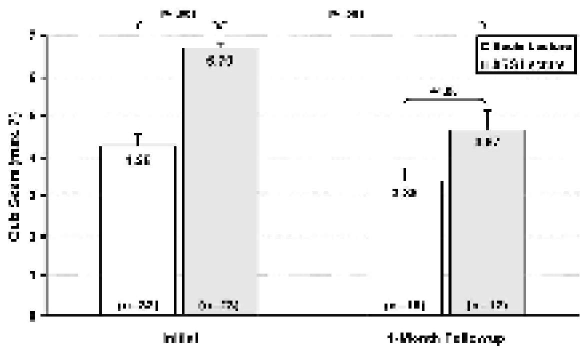
Mean initial quiz scores in interactive and ARS lecture groups did not differ significantly, although at 1-month follow-up there was a nonsignificant trend (P=.11) favoring the ARS group (Figure 3, Table 1). Once again, quiz scores declined significantly in both groups over 1 month.

Effects of Lecture Assignment Scheme (Lecture Sequence Number and Crossover Group) and Training Level (PGY)

A number of other variables that could not easily be controlled for within the context of the study design might have contributed to differences in quiz score performance. Three were of particular interest: (1) lecture sequence number (whether the lecture was given first or second during a given month), which might have contributed to differences in quiz performance as "lecturer experience" improved, (2) subjects' last name (A–J versus K–Z), which determined crossover group assignment, and (3) PGY of training. Results of the multiple regression analysis to adjust for variables are shown in Table 3. The regression models were poor predictors of quiz score, as evidenced by lack of statistical significance for any of the coefficients for the aforementioned three variables, low R^2 values, and lack of overall regression value significance by ANOVA.

Figure 2

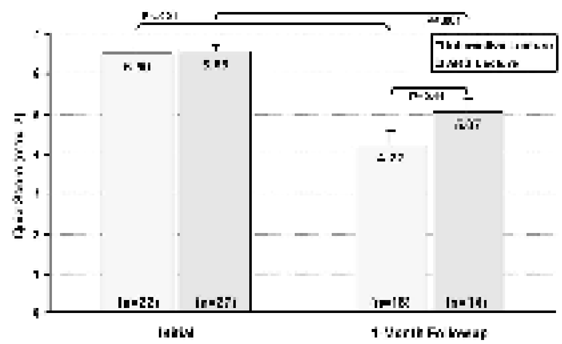
Comparison of Quiz Score Performance
for Basic Versus ARS Lecture Groups



ARS—Audience Response System

Figure 3

Comparison of Quiz Score Performance
for Interactive Versus ARS Lecture Groups



ARS—Audience Response System

Figure 4

Comparison of Individual Resident Quiz Score Performance Versus ABFP ITE Score
for Basic (Figure 4A), ARS (Figure 4B), and Interactive (Figure 4C) Lecture Groups

Figure 4A

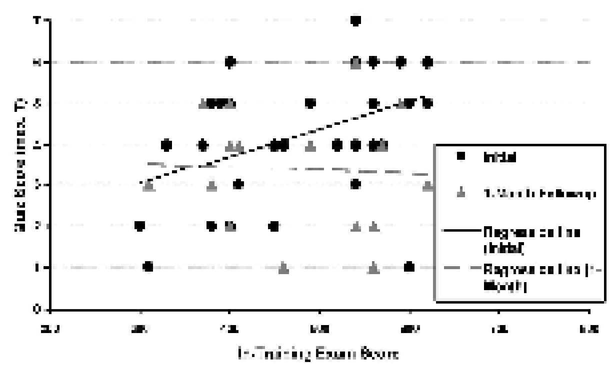


Figure 4B

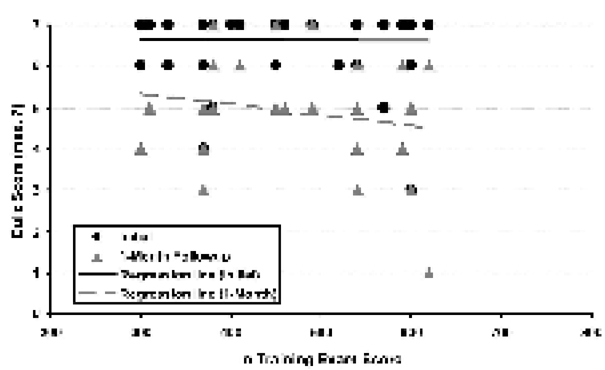
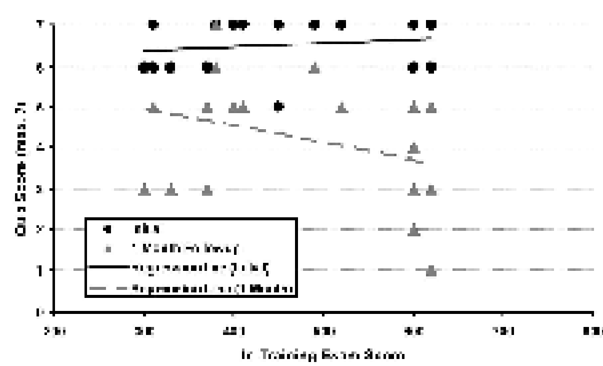


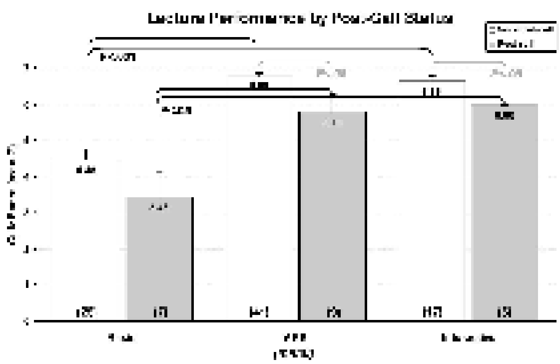
Figure 4C



ABFP ITE—American Board of Family Practice In-training Exam
ARS—Audience Response System

Figure 5

Comparison of Quiz Score Performance for Basic, ARS, and Interactive Lecture Groups According to Resident Post-call Status



ARS—Audience Response System

Table 2

Lecture-unrelated Control Question Scores by Lecture Type

Lecture Type		*Control Questions Quiz Score (Max. 3)*	*n*	P *Value*
Basic	Initial	1.41 ± 0.15	32	
	1-month follow-up	1.44 ± 0.25	18	
ARS*	Initial	1.32 ± 0.12	50	.38
	1-month follow-up	1.50 ± 0.14	26	
Interactive	Initial	1.55 ± 0.23	22	
	1-month follow-up	1.89 ± 0.21	18	

ARS—Audience Response System

* Pooled ARS lectures from May 2002–January 2003 (Figures 1A and 1B).

Table 3

Confounding Factors: Multiple Linear Regression Analysis

Lecture Type		*Confounding Factor*	P *Value (Individual Regression Coefficients)*	R^2 *(Overall Regression Model)*	P *Value (Overall Regression Model)*
Basic	Initial	Sequence #	.12	.15	.20
		Last Name	.11		
		PGY	.07		
	1-month follow-up	Sequence #	.20	.31	.15
		Last Name	.10		
		PGY	.35		
ARS*	Initial	Sequence #	.57	.02	.83
		Last Name	.94		
		PGY	.88		
	1-month follow-up	Sequence #	.67	.04	.82
		Last Name	.68		
		PGY	.40		
Interactive	Initial	Sequence #	.56	.18	.29
		Last Name	.95		
		PGY	.08		
	1-month follow-up	Sequence #	.80	.13	.59
		Last Name	.35		
		PGY	.69		

ARS—Audience Response System

* Pooled ARS lectures from May 2002–January 2003 (Figures 1A and 1B).

Effect of Resident ITE Performance

Another consideration is that residents with different medical "funds of knowledge" might have been expected to perform differently in this study. An individual resident's performance on the American Board of Family Practice ITE is one readily available standardized measure (albeit imprecise) of medical fund of knowledge. To further evaluate the degree to which ITE scores correlated with quiz scores, every resident's quiz score in each lecture group (basic, ARS, and interactive) was graphed against that resident's most recent composite score on the ITE, both for initial and 1-month follow-up quizzes (Figures 4A, 4B, and 4C). Linear regression analysis was performed for each scatter plot, with overall regression model summary results shown in Table 4. Once again, the regression models were poor predictors of quiz score, as evidenced by low R^2 values and general lack of regression value significance by ANOVA. Only initial basic lecture results showed significant (although extremely weak) correlation with ITE score.

Effect of Post-call Status

The effect of post-call status on initial quiz performance is shown in Figure 5 and Table 5. Overall, residents who were post call performed significantly worse than their non-post call counterparts during initial ARS and initial interactive lecture quizzes and had a nonsignificant (*P*=.12) trend toward doing so during initial basic quizzes. Nonetheless, both ARS and interactive lecture participants significantly outperformed basic lecture attendees after initial quiz completion, regardless of whether they were non-post call or post call. Finally, differences in initial quiz performance between ARS and interactive lectures could not be demonstrated regardless of post-call status. There were insufficient numbers of post-call residents completing 1-month follow-up quizzes to permit meaningful analysis of those results.

Table 4

Confounding Factors: ABFP ITE Score Linear Regression Analysis

Lecture Type		R^2	P *Value*
Basic	Initial	.18	.02
	1-month follow-up	.00	.81
ARS*	Initial	.00	.96
	1-month follow-up	.04	.35
Interactive	Initial	.04	.36
	1-month follow-up	.11	.19

ABFP ITE—American Board of Family Practice In-training Exam
ARS—Audience Response System

* Pooled ARS lectures from May 2002–January 2003 (Figures 1A and 1B)

Discussion

The present study adds to a limited literature on ARS in medical education by documenting improvement in a learning outcome (post-lecture quiz score) in postgraduate medical trainees (family medicine residents) using an ARS-enhanced lecture format versus a traditional non-interactive (basic) lecture format. This effect appeared to be durable, with factual retention in ARS lecture attendees continuing to surpass that of

Table 5

Lecture Performance by Post-call Status

Lecture Type	*Post-call Status*	*Quiz Score (Max. 7)*	*n*	P *Value (Non-post Call Versus Post Call)*	*P Value (Versus Corresponding Basic Lecture)*	*P Value (Versus Corresponding ARS* Lecture)*
Basic initial	Non-post call	4.48 ± 0.28	25	} .12	—	—
	Post call	3.43 ± 0.72	7		—	—
ARS* initial	Non-post call	6.80 ± 0.07	41	} < .05	< .001	—
	Post call	5.78 ± 0.49	9		—	—
Interactive initial	Non-post call	6.65 ± 0.15	17	} < .05	< .001	.46
	Post call	6.00 ± 0.00	5		< .05	.79

ARS—Audience Response System

* Pooled ARS lectures from May 2002–January 2003 (Figures 1A and 1B)

basic lecture attendees 1 month after initial lecture administration (Figure 2, Table 1). Notably, quiz scores for basic lecture attendees were poor—both initially and 1 month later (averaging 61% and 48%, respectively)—which speaks to the necessity for improvement on the traditional didactic lecture method. These scores were similar to those documented elsewhere for midday conferences in family medicine residents.[1] ARS lecture participants fared somewhat better, with average quiz scores of 96% and 67%, respectively. Not surprisingly, lecturer interaction with trainees—in the context of the highlighting of essential teaching points via multiple-choice question presentation (which occurred during both ARS and interactive lectures)—also resulted in improved post-lecture quiz scores relative to non-interactive basic lectures (Figures 2 and 3, Table 1).

At least two plausible explanations for these results exist: (1) improved retention occurs with active participation in the lecture process and (2) improved retention occurs when key learning points are highlighted prior to testing. Both of these conditions occurred during ARS and interactive lectures. A separate benefit of the ARS technology itself—beyond what was achieved simply by making the lectures "interactive"—could not be clearly demonstrated, although the data hint at one being present at 1 month of follow-up. Nonetheless, because the interactive lecture method may not be scaleable to larger groups, there remains the possibility that ARS technology would be more effective than interactive lectures in those settings.

Limitations

This study has several limitations. First, the study did not lend itself to blinding of experimental groups but instead relied on a crossover design that was intended to minimize differences in baseline characteristics of subjects. Various regression analyses failed to disclose any significant correlation between lecture assignment scheme, postgraduate training year, or ITE score and quiz performance in any of the six experimental subgroups. However, our study was not sufficiently powered to detect small influences of these factors on our results.

Similarly, the small size of the study and insufficient statistical power may have impaired our ability to resolve small differences in quiz score outcomes between subgroups of particular interest (for example, the difficulty in discerning whether a true difference existed between 1-month follow-up ARS and interactive lecture groups).

Additionally, data were limited to a single family medicine residency training site at a single hospital. Whether these data would translate to (1) other family medicine residency training sites in other locales, (2) resident trainees in disciplines other than family medicine, (3) other medical education settings such as commercial CME events, and (4) subjects other than resident trainees (such as medical students or attending-level physicians) remains unknown but worthy of further investigation.

Our finding that residents who identified themselves as being post call at the time of a lecture performed moderately worse on their post-lecture quizzes (particularly if they participated in an interactive-style lecture format, namely ARS or interactive lecture groups) also has limitations. One difficulty in interpreting these data is that no attempt was made to objectively quantify degrees of fatigue or sleep deprivation among post-call quiz respondents. An additional shortcoming was the insufficient number of respondents to allow for analysis of post-call 1-month follow-up data. Nonetheless, even the findings from initial lectures may have implications for the future conduct of medical education, particularly since the post-call respondents in this study were already operating essentially within the limits of the new ACGME work-hour rules.[11] Importantly, post-call status did not alter this study's fundamental findings of improved post-lecture quiz performance in either ARS or interactive lecture groups relative to the basic lecture group.

An additional limitation is that data presented in this study apply only to a small audience setting (a small conference room with a small (<20) number of lecture attendees). Another important direction for future research will be the assessment of learning performance in larger audiences.

A final cautionary note must be sounded. A paramount objective of medical education is to effect measurable improvement in practice outcomes (either altered physician behavior or altered health care outcomes among patients).[12] While it may be enticing to assume that an active learning tool such as the ARS-enhanced lecture could improve practice outcomes, the present study does not address this fundamental issue; it only studied subjects' ability to correctly answer questions on a quiz. Future research should be directed at the question of whether improved learning during ARS-enhanced lectures can ultimately be linked to improved clinical outcomes.

Conclusions

ARS-enhanced lectures (1) improved post-lecture quiz performance in family medicine residents, both initially and up to 1 month after lecture administration and (2) reliably delivered essential learning points in such a way that good post-lecture factual retention rates were achieved. Both audience-lecturer interaction and ARS equipment usage seemed to contribute to these improved learning outcomes.

Acknowledgments: We gratefully acknowledge the enthusiastic participation of the St Elizabeth Hospital Family Practice residents in this research study.

This project was supported by an equipment grant (audience response system hardware and software) from Meridia Audience Response (Plymouth Meeting, Pa).

Portions of the data in this study were presented at the 2003 American Academy of Family Physicians Scientific Assembly in New Orleans and the 2003 Association of American Medical Colleges RIME Meeting in Washington, DC.

Corresponding Author: Address correspondence to Dr Schackow, Family Practice Residency Program, St. Elizabeth Hospital, 1431 N. Claremont, Chicago, IL 60622. 312-633-5842. Fax: 312-633-5936. schackow@uic.edu.

References

1. Picciano A, Winter R, Ballan D, Birnberg B, Jacks M, Laing E. Resident acquisition of knowledge during a noontime conference series. Fam Med 2003;35(6):418-22.
2. Russell IJ, Hendricson WD, Herbert RJ. Effects of lecture information density on medical student achievement. Acad Med 1984;59:881-9.
3. Draper S, Cargill J, Cutts Q. Electronically enhanced classroom interaction. Paper presented at the Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2001), December 2001.
4. Blandford L, Lockyer J. Audience response systems and touch pad technology: their role in CME. J Contin Educ Health Prof 1995;15:52-7.
5. Robertson LJ. Twelve tips for using a computerized interactive audience response system. Med Teach 2000;22:237-9.
6. Gagnon RJ, Thivierge R. Evaluating touch pad technology. J Contin Educ Health Prof 1997;17:20-6.
7. Kristjanson C, Magwood B, Plourde P. Utilization of keypad technology in medical education teaching. CGEA RIME facilitated poster session #2, Association of American Medical Colleges CGEA RIME Meeting, 2002.
8. Nasmith L, Steinert Y. The evaluation of a workshop to promote interactive learning. Teach Learn Med 2001;13:43-8.
9. Copeland HL, Longworth DL, Hewson MG, Stoller JK. Successful lecturing: a prospective study to validate attributes of the effective medical lecture. J Gen Intern Med 2000;15:366-71.
10. Copeland HL, Stoller JK, Hewson MG, Longworth DL. Making the continuing medical education lecture effective. J Contin Educ Health Prof 1998;18:227-34.
11. Accreditation Council for Graduate Medical Education. Resident duty hours and the working environment. Common program requirements, Section IV D. Chicago: ACGME, July 2003.
12. O'Brien T, Freemantle N, Oxman AD, Wolf F, Davis DA, Herrin J. Continuing education meetings and workshops: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2003;3:CD003030.

Editor's Note: This paper was acknowledged at the 2003 American Academy of Family Physicians Scientific Assembly as an award-winning paper in the category of research by family physicians and fellows primarily in academic medicine.