Critical analysis of a computer-assisted tutorial on ECG interpretation and its ability to determine competency

J. F. Burke, E. Gnall, Z. Umrudden, M. Kyaw & P. K. Schick MD, FACP

Critical analysis of a computer-assisted tutorial on ECG interpretation and its ability to determine competency

J. F. BURKE[1], E. GNALL[1], Z. UMRUDDEN[1], M. KYAW[1] & P. K. SCHICK[2]

[1]Lankenau Hospital, Wynnewood, [2]Lankenau Hospital, Drexel University and Jefferson Medical College, USA

Abstract

Background: We developed a computer-based tutorial and a posttest on ECG interpretation for training residents and determining competency.

Methods: Forty residents, 6 cardiology fellows, and 4 experienced physicians participated. The tutorial emphasized recognition and understanding of abnormal ECG features. Active learning was promoted by asking questions prior to the discussion of ECGs. Interactivity was facilitated by providing rapid and in-depth rationale for correct answers. Responses to questions were recorded and extensively analyzed to determine the quality of questions, baseline knowledge at different levels of training and improvement of grades in posttest. Posttest grades were used to assess improvement and to determine competency.

Results: The questions were found to be challenging, fair, appropriate and discriminative. This was important since the quality of Socratic questions is critical for the success of interactive programs. The information on strengths and weakness in baseline knowledge at different levels of training were used to adapt our training program to the needs of residents. The posttest revealed that the tutorial contributed to marked improvement in feature recognition. Competency testing distinguished between residents with outstanding grades and those who needed remediation.

Conclusions: The strategy for critical evaluation of our computer program could be applied to any computer-based educational program, regardless of topic.

Background

A number of computer-based programs on electrocardiographic interpretation are available that vary in design, interactivity, speed of response time and the target audience (Balzan 2007; Crimando 2007; Medi-Smart 2007). These programs have been evaluated primarily from feedback and some by improvement in grades. There is large number of computer-based programs on a wide range of medical topics. Some have been assessed by feedback and improvement in grades (Lewis 1999; Al-Rawi et al. 2007; George et al. 2007). However, most educators with experience with computer based education agree that there still is a need for in-depth evaluation of the value of these programs (Lewis 1999; Al-Rawi 2007).

Several programs were designed to test for competency. For example, the American College of Cardiology (ACC) electrocardiogram (ECG) proficiency test (Kadish et al. 2001). ACC also developed ECGSAP I, II, and III to provide physicians with a means to compare their proficiency in ECG interpretation to their peers and to improve their own proficiency (Mason et al. 2001). The Institute for Clinical Evaluation, a foundation of the American Board of Internal Medicine, has developed a certification examination in ECG interpretation that has also been offered to fellows training in general cardiovascular disease. However, there is currently a

Practice points

- A computer-based tutorial can improve skills for the identification of abnormal features in ECGs.
- The tutorial and posttest generated information about strengths and weaknesses in baseline knowledge in residents at various stages of training.
- The strategy for critical evaluation of the ECG program could be applied to any computer-based program, regardless of the topic.

need for new approaches to teach and test for competency for ECG interpretation (Kadish et al. 2001, Salerno et al. 2003a, b) have suggested that previous attempts to develop competency testing were influenced by inclusion of too many details of clinical history and did not emphasize interpreting abnormal details in ECGs. The Accreditation Council for Graduate Medical Education has mandated that training programs establish measurable thresholds for competency in ECG interpretations.

In response, we developed a computer-based tutorial that emphasized interpreting details and abnormal features in ECG's. A posttest that was designed to determine competency that addressed the concerns of Salerno et al. (2003a, b).

Correspondence: P. K. Schick, MD, FACP, Lankenau Hospital, Wynnewood, Pennsylvania, USA. Tel: 610-645-3305; email: pksone@comcast.net

1. The tutorial and the posttest were more extensively evaluated. In addition to evaluating improved grades and feedback, our study scrutinized the quality of questions since they are critical for the success of interactive computer-based programs. Also, responses to questions in the tutorial were graded and revealed strengths and weakness in baseline knowledge in residents at different stages of training. This information was used to adapt our training program to resident needs. The educational value of the tutorial and the validity of using the programs for determining competency were also analysed.

2. The tutorial and posttest focused on the recognition of abnormal features in ECGs since this emphasis was important to develop the ability to interpret complex and ambiguous ECGs.

3. The program was highly interactive.

Testing for competency distinguished between residents who achieved outstanding grades in ECG interpretation and those who needed additional training. Detailed grades in competency were shown to participants in a confidential manner to stimulate self-improvement.

The study demonstrates the importance of using computer-based programs to train residents to identify abnormal clinical findings and to understand pathophysiology. The study also demonstrates a strategy for in-depth evaluation of computer-based educational programs that can serve as a model for the evaluation of all medical educational computer programs, regardless of topic.

Software

Microsoft Visual Basic was ideal software for preparation of the program since it supported rapid processing of images and immediate response to answers. The program recorded time spent and graded answers to questions. It was available on the hospital Intranet.

Methods

Tutorial. 53 ECGs of disorders commonly encountered by primary physicians were presented in the tutorial. They were grouped in modules containing 3 to 5 ECGs. The modules cover the following topics:

1. ST-T abnormalities;
2. MI locations;
3. Myocardial injury, old & recent MI;
4. Axis deviation;
5. AV conduction disturbances;
6. Wide QRS complexes;
7. Chamber enlargement (Atrial);
8. Chamber enlargement (Ventricular);
9. Irregular rhythms;
10. Arrhythmias (narrow QRS);
11. Arrhythmias (wide QRS);
12. Metabolic disorders.

Fifty participants used the tutorial: 40 residents, 6 cardiology fellows, 3 cardiologists and 1 hospital-based internist.

Participants were asked to identify all abnormal features in the ECGs. Minimal clinical details were provided to avoid the influence of clinical history when interpreting abnormal features and details. The correct and incorrect choices are instantly displayed prior to showing a list of possible underlying disorders. Asking questions about abnormal features prior to showing a list of possible diagnoses helped participants focus on recognizing abnormal features. Next, participants were asked to diagnose the underlying disorder from a list of possible diagnoses. For educational purposes, both the ECGs and a list of the relevant abnormalities were available to help participants identify all abnormalities when choosing the diagnosis. After participants chose a diagnosis, the program immediately displayed the rationale for the correct diagnosis

The program graded and analysed participants' responses to questions and recorded the time spent using the tutorial and posttest. Since the questions were answered prior to providing for the correct answers, the program established baseline knowledge. Only initial responses were recorded by the program. The data were stored in a database.

Feedback. The program gave participants the opportunity to evaluate the program and their comments were recorded in a database. Feedback was also obtained by exit interviews.

Posttest. A computer-based posttest was given within 3 weeks after a participant finished the tutorial. Twenty of the ECGs from the tutorial were chosen for the posttest because they represented core knowledge for primary care physicians. Forty-six participants took the posttest, 40 residents and 6 cardiology fellows.

ECGs used in the Posttest:

1. Acute injury pattern;
2. Pericarditis;
3. Acute anteroseptal MI;
4. Acute inferior MI;
5. Acute myocardial Injury;
6. Left axis with fascicular block;
7. 2nd degree AV block Mobitz I;
8. 1st degree HB;
9. WPW;
10. RBBB;
11. Right atrial abnormality;
12. Right atrial hypertrophy;
13. Left ventricular hypertrophy;
14. Atrial fibrillation;
15. Atrial flutter;
16. Sinus arrhythmia;
17. Atrial tachycardia with aberrant conduction RBBB;
18. Ventricular tachycardia with RBBB;
19. Hyperkalemia;
20. Digitalis effect.

Participants were asked to identify all abnormal features in the ECGs in the posttest as in the tutorial, but, unlike the tutorial, answers were not provided. Next, they were asked to

diagnose the underlying disorder. When making the diagnoses, participants were not provided the list of abnormal features as was displayed in the tutorial. Therefore, the challenge was to only rely on one's ability to identify abnormal features in the ECGs to diagnose the underlying disorder without the benefit of viewing a list of abnormal features. As with the tutorial, the answers to questions were recorded and graded. Time spent was also recorded.

Analysis of data: Individual questions were evaluated for difficulty and discriminative value. We considered a question to be moderately difficult and ideal if between 50 to 85% of participants answered it correctly. Questions were thought to be difficult or easy if less than 50% or more than 85%, respectively, answered them correctly. Discriminative value was determined by a ratio of the number of residents in the upper quintile vs the number in the lower quintile who answered a question correctly. The upper and lower quintiles were defined by grades in the tutorial. A question with a ratio greater than 1.25 was considered discriminative, a question with a ratio greater than 1.5 was considered very discriminative and questions with ratios of 1 or less were not discriminative.

In the tutorial, grades for the ability to detect abnormal features and diagnostic skills were analysed to establish baseline knowledge at different levels of training, 1st (PGY1), 2nd (PGY2) and (PGY3) 3rd year medical residents and experienced physicians. Experienced physicians included cardiology fellows, and cardiologists and a hospital-based internist. Grades in feature identification and diagnostic skills in the posttest to were used to determine whether the tutorial had improved knowledge and had educational value.

Statistical analysis. T-test; One Way ANOVA and Scheffe post-hoc analysis.

Results

The average time spent on the tutorial by residents was 1 hr and 51 min, by cardiology fellows was 1 hr and 48 min, and by cardiologists and hospital-based internists was 1 hr and 25 min. Most participants completed the tutorial in 3 separate sessions. The average time spent on the posttest by residents was 34 minutes and by cardiology fellows was 25 minutes.

Figure 1 demonstrates the difficulty and discriminative value of questions on identifying abnormal features and on diagnosing the underlying disorders in the tutorial and in the posttest.

Figure 1a demonstrates that most questions on feature identification in the tutorial were moderately difficult. The average difficulty was 56%. The average discriminative value was high with an average value of 1.6.

Questions in each module were evaluated. For example, questions on feature identification in the Tutorial in the ECGs in Module 11 (wide QRS arrhythmias) were difficult since only 37% of the participants chose correct answers. However, the choice of features in this module was highly discriminative with a value of 3.1, and therefore they were fair questions.

Figure 1b demonstrates that most questions on diagnosis in the tutorial were moderately difficult, and the average difficulty

was 79%. The average discriminative value was extremely high with an average value of 2.1.

Figure 1c demonstrates that most questions on feature identification in the posttest were moderately difficult. The average difficulty was 56%. The average discriminative value was 1.63.

Figure 1d demonstrates that most questions on diagnosis in the posttest were moderately difficult, and the average difficulty was 77.4%. Also, the questions were discriminative and that the average value was 1.91.

Questions about each ECG were evaluated. For example, the diagnosis of 2 ECGs in the posttest was difficult, and both demonstrated myocardial injury. None of the participants made the diagnosis for the ECGs that is circled in Figure 1d. The ability to diagnose this ECG was not discriminating with a value of 1.0. Therefore, this question was considered to be unfair and was not graded. The other question designated by an arrow in Figure 1d was also very difficult since only 22% of the residents made the diagnosis. However, the ability to diagnose this EKG was very discriminating with a value of 4.0. Therefore, while difficult, it was a valid question.

Baseline grades established in the tutorial in feature identification and diagnosis in residents at different years of training and a more experienced group (cardiology fellows, cardiologists and hospital-based internists) are shown in Table 1.

Table 1a shows that in the tutorial baseline grades in feature identification progressively increased from 52.6% for PGY1 to 68.8% for the experienced group. Cardiology Fellows' grades for feature identification were 70%. Differences in baseline grades in feature identification between all residents and the experienced group were significant.

Table 1b also shows that in the tutorial baseline grades in diagnostic skills progressively increased and ranged from 70.3% for PGY1 to 96.2% for the experienced group. Cardiology Fellows' grades for diagnosis were 95.3%. The difference in grades in diagnosis for all residents and cardiology fellows was significant. Differences between PGY1 or PGY2 and experience physicians and differences between PGY1 and PGY3 were significant.

Table 1b demonstrates that in the posttest grades in features identification progressively increased and ranged from 64.1% for PGY1 to 79.0% for cardiology fellows. Table 1b also shows that grades in diagnostic skills progressively increased and ranged from 69.9% for PGY1 to 90.5% for cardiology fellows. The difference in grades in diagnosis between all residents and cardiology fellows and between PGY1 and experience physicians were significant. Only residents and cardiology fellows but not cardiologists took the posttest.

Grades for the identification of features for all participants in the posttest were markedly higher than in the tutorial. Baseline knowledge in feature identification in the tutorial was $56.4 \pm 19.7\%$ and in the posttest were $68.6 \pm 13.1\%$. The difference was significant.

The improvement of grades in feature identification was further assessed in 3 groups participants based on their grades in the tutorial (lower, middle and upper third). Figure 2 demonstrates that improvement in grades in feature identification was most striking in residents whose grades in
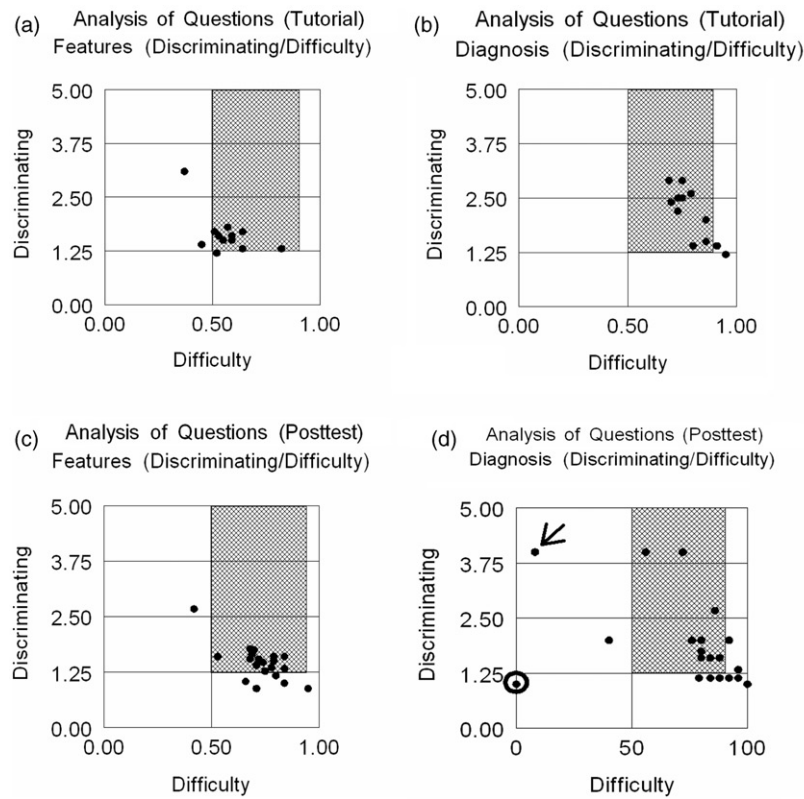
Figures 1a–d. Evaluation of questions in the tutorial and posttest:
● Each data point represents the data on the difficulty and discriminative value of questions. Difficulty was judged by how many participants answered questions in each module correctly. From 50 to 85% was considered an ideal range and indicated that the questions were moderately difficult. Discriminative value was determined by the ratio of the number of participants who ranked in the upper quintile in all the questions in the tutorial who answered the questions correctly vs. the number in the lower quintile. The upper and lower quantiles were defined by grades in the tutorial. A question was considered discriminative if the ratio was greater than 1.25 and very discriminative if over 1.5. The ideal range of difficulty of discriminative questions is shown in the shaded areas.
● Figure 1a: Data from the evaluation of questions of feature identification in the 12 Modules in the Tutorial. Each Module contains 3 to 5 ECGs relevant to the topic of the Module. Each data point represents the mean of difficulty and discriminative values for all the ECGs in a Module.
● Figure 1b: Data on the evaluation of questions on diagnosis in the 12 Modules in the Tutorial.
● Figure 1c: Data points represent data on the evaluation of the difficulty and discriminative value of questions on feature identification each of the 20 ECGs in the Posttest.
● Figure 1d: Data points represent data on the evaluation of the difficulty and discriminative value of questions on the identification of abnormal features in each of the 20 ECGs in the posttest. The ECG that is circled was difficult and non-discriminative, and therefore was not graded. Feature identification in the EKG designated by the arrow was difficult but very discriminative and was graded.

	PGY1 (n = 18)	PGY2 (n = 10)	PGY3 (n = 12)	*Experienced physicians (n = 10)	ANOVA (CF = Critical F)
Table 1a. Baseline grades in tutorial (n = 50).					
Features	52.6 ± 22.7	56.8% ± 12.8	55.0% ± 21.9	68.8% ± 15.8	F = 1.584032
					CF.$_{05}$ = 2.8.06843
Diagnosis	70.3 ± 12.3	78.9% ± 11.5	85.2% ± 15.1	96.2% ± 13.1	F = 9.174631
					CF.$_{05}$ = 2.806843

*Experienced physicians: Cardiology fellows, cardiologists and hospital-based internists.

the tutorial were in the lower third. Baseline grades and posttest grades for the lower 1/3 groups were $32.0 \pm 19.7\%$ and $59.9 \pm 12.0\%$, respectively ($p < 0.001$ per T-test). The improvement in the middle third, 60.3 ± 5.9 and $68.1 \pm 10.7\%$, respectively, was significant ($p = 0.02$). The improvement in the upper third was not significant.

The correlation of grades in diagnosis vs. feature identification for all participants in the posttest is shown in Figure 3a.

	PGY1 ($n = 18$)	PGY2 ($n = 10$)	PGY3 ($n = 12$)	Cardiology Fellows ($n = 6$)	ANOVA (CF = Critical F)
Table 1b. Mean ± SD for grades in the posttest ($n = 46$).					
Features	64.1 ± 10.7	69.8% ± 11.5	69.2% ± 15.8	79.0% ± 13.0	F = 2.125157
					CF.₀₅ = 2.827051
Diagnosis	69.9 ± 10.3	77.3% ± 11.8b	79.3% ± 12.5	90.5% ± 7.15	F = 5.747487
					CF.₀₅ = 2.827051

Table 1a: Grades in diagnosis in the tutorial
- Difference in grades in diagnosis between for all residents and experienced physicians was significant, 75.3 ± 15.5% and 96.2 ± 13.1%, respectively ($p < 0.001$ per t-test). Differences between PGY1 or PGY2 and experienced physicians were significant.
- Differences between PGY1 and PGY3 were significant. (ANOVA and Scheffe post-hoc analysis)

Table 1b: Grades in features in the tutorial
- The difference in grades in feature identification between all residents and cardiology fellows was significant, 54.4 ± 19.7% and 68 ± 15.8%, respectively ($p < 0.03$ per t-test).

Table 1c: Grades in diagnosis in the posttest
- Differences between PGY1 and Experienced Physicians and between other groups were not significant. (ANOVA)
- Diagnosis: Difference in grades in diagnosis between for all Residents and Experienced Physicians was significant, 74.6 ± 11.7% and 90.5 ± 7.2%, respectively ($p < 0.005$ per t-test).
- Differences between PGY1 and Experience Physicians were significant (ANOVA and Scheffe post-hoc analysis).

Table 1d: Grades in features in the posttest
- Difference in grades in feature identification between all residents and cardiology fellows was significant, 67.1 ± 12.4% and 79.0 ± 13.0%, respectively ($p < 0.05$ per t-test).
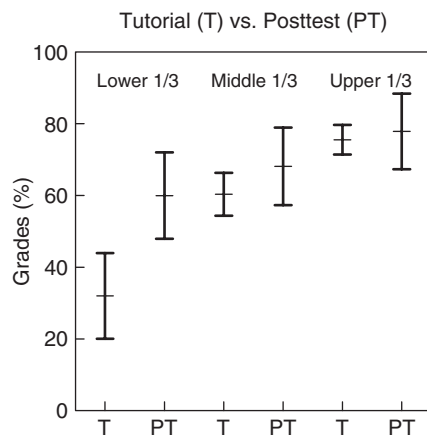- There was no significant difference between the groups (ANOVA).

Figure 2. *Improvement of Grades for Feature Identification for participants at different stages of training:* As described in the results there was a significant improvement grades in the Posttest from that in the Tutorial for all residents. This figure shows the improvement in participants at different stages of training. Only data from cardiology fellows and residents are shown since cardiologists and hospital based physicians did not take the posttest.

The same information for Cardiology Fellows and PGY1, PGY2 and PGY3 residents are shown in Figures 3b–e). It is clear that diagnostic skills correlated with feature identification in the posttest. Also, differences in performance between residents at different levels of training and cardiology fellows were apparent.

Table 2a summarizes data on grades in diagnosis in the posttest. The mean grade for all residents and cardiology fellows was 76.5%. Grades above 89.3% (mean + one SD) were considered to be "Outstanding". Eleven Participants ranked in the outstanding group: one PGY1, two PGY2, four PGY3 and four Cardiology Fellows. Grades below 63.6% (Mean – one SD) were considered "Poor Performance". Twelve participants were in this group and thought to need remediation. Seven PGY1, three PGY2, two PGY3 and no Cardiology Fellows were in this category.

Table 2b summarizes data on grades in abnormal feature identification in the posttest. The mean grade for all residents and cardiology fellows was 69.9%. Grades above 83.1% (mean + one SD) were considered to be "Outstanding". Seven participants ranked in the outstanding group: No PGY1, one PGY2, four PGY3 and two Cardiology Fellows. Grades below 56.8% (mean- one SD) were considered "Poor Performance". Eight participants were in this group and thought to need remediation. Three PGY1, two PGY2, two PGY3 and one Cardiology Fellow were in this category.

The use of some of the same ECGs in the posttest that had been used in the tutorial did not appear to have biased the grades. We tested for conceptual understanding by asking participants to select all abnormal features from extensive lists and we avoided multiple choice questions. Eighty percent of residents who had highest grades in the posttest also had the highest grades in the tutorial indicating that residents with "photographic memories" did not have a selective advantage. However, we plan to begin to use different ECGs in the posttest.

Competency was based on grades in diagnosis as well as in features. In the posttest, six out of 11 participants that had outstanding grades in diagnosis also had outstanding grades in feature identification. Seven out of 12 participants who had poor grades in diagnosis also had poor grades in feature identification. Therefore, residents who had outstanding or poor grades in feature identification were also likely to have had outstanding or poor grades in diagnostic skills.

Feedback was favorable. Residents felt that the programs were aimed at senior residents and fellows, but they generally
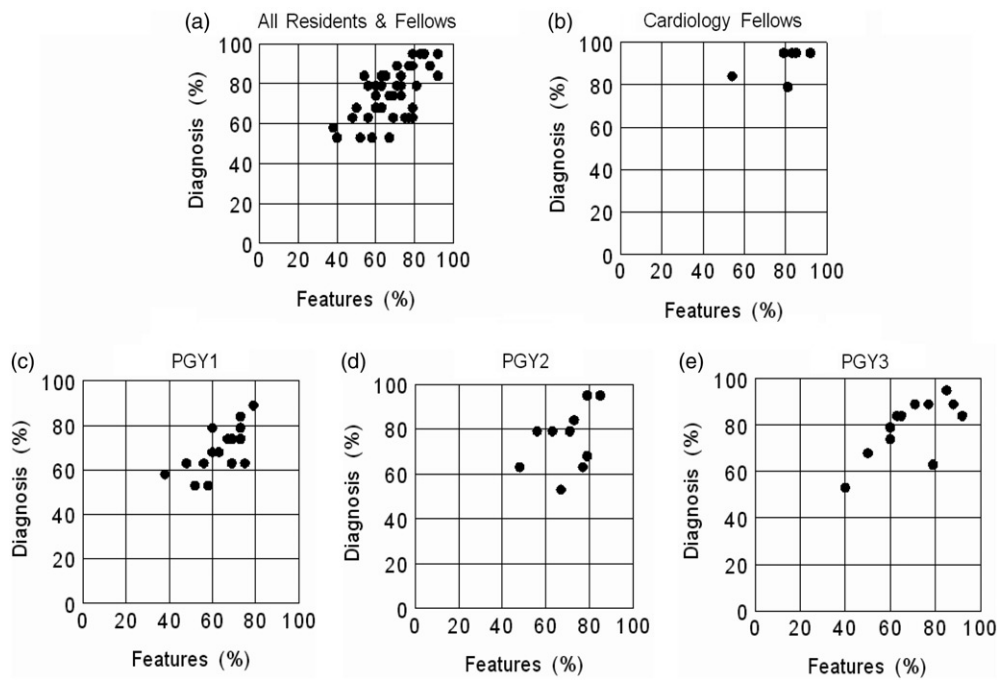
Figures 3a–e. *The correlation of grades in feature identification and diagnosis*: (a) For all participants; (b) For PGY1; (c) For PGY2; (d) For PGY3; (e) For cardiology fellows (cardiologists and hospital-based physicians did not take the posttest).

Table 2a. Participants with ''Outstanding'' & ''Poor'' grades for diagnosis in the posttest.					
Mean grade = 76.5%	PGY 1	PGY 2	PGY 3	Cardiology fellows	ALL
''Outstanding Performance'' >89.3 (Mean + one SD)	1/18	2/10	4/13	4/6	11
''Poor Performance'' <63.6% (Mean − one SD)	7/18	3/10	2/13	0	12

Table 2b. Participants with ''Outstanding'' & ''Poor'' grades for feature identification in the posttest.					
Mean grade = 69.9%	PGY 1	PGY 2	PGY 3	Cardiology Fellows	ALL
Outstanding Performance'' >83.1 (Mean + one SD)	0/18	1/10	4/13	2/6	7
''Poor Performance'' < 56.8 (Mean - one SD)	3/18	2/10	2/13	1/6	8

Tables 2a & 2b. The mean grade for all participants for diagnosis was 76.5% and for feature identification was 69.9%. Mean grades ± One SD were used a criteria for outstanding and poor grades. The table shows the number of participants that were in the ''Outstanding'' and ''Poor'' groups per the total number of PGY1, PGY2, PGY3 and cardiology fellows who took the posttest.
Six out of 11 in the outstanding category for diagnosis also were in the outstanding in feature identification. This is an impressive correlation since only residents were outstanding in feature identification
Seven out of 12 in the poor performance category for diagnosis were also in the poor category in feature identification.

felt the tutorial was interactive and instructive program. Several suggestions will be implemented such as the addition of calipers and using yellow indicators of green for color-blind participants. We plan to use calipers.

However, the lack of calipers did not invalidate the posttest grades for interpretation of the ECGs on arrhythmias since the average grade for diagnosis (74%) and feature identification (78%) was not significantly different from grades for all ECGs (72%) in the posttest

Discussion

The program emphasized the recognition of all abnormal details and features of ECGs. The rationale was that it is necessary to understand all abnormalities to diagnose complex and ambiguous ECGs. This was supported by evidence that the ability to recognize abnormal features correlated with diagnostic skill as shown in Figure 3. Emphasis was achieved by providing minimal clinical details and asking participants to

identify abnormal features prior to asking them to diagnose the underlying disorder. This encouraged the careful analysis of ECG patterns.

The study revealed a significant increase in the grades in the recognition of features in the posttest from that in the tutorial. Most striking was the marked improvement in the posttest for residents with grades in the lower 1/3 percentile for feature identification in the tutorial. This information clearly showed that the tutorial was effective in increasing the ability to recognize and understand abnormal features. It was especially helpful for residents who had low scores in baseline knowledge in feature identification. Thus, one of our major goals had been achieved.

Since residents will be asked to study the tutorial and take the posttest each year of their 3-year residency, we should be able to evaluate the long-term retention of the ability to recognize abnormal features. Current baseline grades for residents at different stage of training will serve as controls since this is the first year that the tutorial was available. Residents transferring into our training program from other hospitals will also serve as controls.

The tutorial was highly interactive. It provided rapid and detailed rationale for correct answers to questions asked in the program. Rapid response time was in part due to use of Visual Basic as the programming language.

Participants were asked to evaluate the program. The feedback was favorable and indicated that the program was useful for learning how to interpret ECGs. We plan to improve the program by responding to constructive comments by providing tutorial, a sample practice program to reinforce the instructions, add calipers and choose colors to accommodate color-blind participants.

The study was unique since all responses and time spent were recorded, and this information was used to carry out an extensive post-hoc evaluation of the tutorial and posttest.

The quality of questions was evaluated. Overall, objective evaluation revealed that almost all the questions in the tutorial and the posttest were ideal in respect to difficulty and discriminatory value, and they therefore were appropriate and fair. This was confirmed by participants' comments that they thought the questions were appropriate and Socratic. This was gratifying since the success of interactive computer-based programs is dependent on the quality of the questions.

The evaluation of individual questions revealed useful information. In the tutorial, questions about ECGs of arrhythmias associated with wide QRS complexes were shown to be difficult, but highly discriminative. Therefore, we plan to emphasis the interpretation of these arrhythmias in our training program.

In the posttest, two ECGs were difficult to diagnose, and they were both from patients with myocardial injury. One question had very little discriminative value and was therefore considered to be an unfair questions and was not graded. Even though participants had difficulty making the diagnosis of myocardial injury, they were able to identify most of the abnormal features in these ECGs. Apparently, they did not appreciate the significance of the features. It is clear that we will have to spend more time reviewing how to diagnose acute myocardial injury in our training program. These are examples of how the program can identify which questions were constructive and which were inappropriate.

Post hoc evaluation of data defined baseline knowledge that was useful for our training program. There was a wide distribution in grades. Thirteen out of 40 residents and all experienced physicians scored grades in diagnosis over 90%. Therefore, knowledgeable participants could obtain high grades. The fact that experienced physicians and knowledgeable residents could score high grades indicated that questions were fair and the ECGs were appropriate.

Strengths and weakness in baseline knowledge and expertise in interpreting ECGs were identified in residents at different levels of training. Generally, the ability to interpret ECGs improved during their training period. PGY3 developed considerable expertise in evaluating ECGs during their 3-year training period. The information about strengths and weaknesses will be used to adapt our training program for residents at different levels of training.

The posttest grades were used for determining competency in compliance with the mandate of the Accreditation Council for Graduate Medical Education. The grades in diagnosis as well as feature identification in the posttest were used to determine competency. There was close correlation between grades in feature identification and diagnosis. In the posttest, mean grades $\pm$ one SD in feature identification and diagnostic skills were used to establish criteria for outstanding and "poor performance.

A total of 11 participants had outstanding grades in diagnosis. The percent of participants in the "outstanding category" increased with the length of training.

Twelve participants were in the "poor" category for diagnostic skills and needed additional help interpreting ECGs and remediation. The percent of participants in the "poor performance category" decreased with the length of training.

We applied the same approach (mean $\pm$ one SD) to establish an "outstanding" and "poor" category in feature identification at different levels of training in the posttest. We found the pattern off results was similar to that in diagnostic skills. As noted above, grades in features correlated with grades in diagnosis.

We used the data on competency to help residents improve the skills in the interpretation of ECGs. We sent a copy of Figures 3a–c that showed grades in diagnosis and feature identification to each resident with his/her grades circled in red. The approach provided his/her grades in a confidential manner and allowed him/her to see their grades relative to other residents. We hoped that would encourage self-improvement. For residents in the "poor" category, we recommended remediation.

In summary, the tutorial improved skills for the identification of abnormal features in ECGs. Questions asked in the tutorial and posttest was found to be discriminative and Socratic, and therefore appropriate for the success of an

interactive computer program. The tutorial and posttest generated information about strengths and weaknesses in baseline knowledge in residents at different stages of training that facilitated introspective assessment in order to improve our training program. A strategy for determining competency in ECG interpretation and identifying residents who need additional help was devised.

Most computer assisted instruction programs have not been adequately evaluated. Our study demonstrated the value of analysing the quality of questions used in the program and the demographic analysis of baseline knowledge in residents at different levels of training. The study describes a strategy for critical evaluation that could be applied to any computer-based program, regardless of the topic.

Notes on contributors

JAMES BURKE MD, Cardiologist. Major intellectual input.

ERIC GNALL, DO. Cardiology Fellow. He provided intellectual input, assisted assembling the program.

ZIA UMRUDDEN MD and MOE KYAW, MD. Chief Residents. Implemented the study, obtained feedback.

PAUL SCHICK, MD. Professor of Medicine. Organized intellectual content, developed the program and analysed data.

Acknowledgements

References

Al-Rawi WT, Jacobs R, Hassan BA, Sanderink G, Scarfe WC. 2007. Evaluation of web-based instruction for anatomical interpretation in maxillofacial cone beam computed tomography. Dentomaxillofac Radiol 36:459–464.

Balzan MV. 2007. Astonix Project - Volume 1 - ECG interpretation. Available online at: http://www.sharewarejunkies.com/8ef3/ecgint.htm (Accessed 15 December 2007).

Crimando J. 2007. EKG arrhythmia review. Available online at: http://www.gwc.maricopa.edu/class/bio202/cyberheart/ekgqzr0.htm (Accessed 15 December 2007).

George S, Pachev GS, MacDonald WA. 2007. Embedding Medical Student Computer Tutorials into a Busy Emergency Department. Acad Emerg Med 1:138–148.

Kadish AH, Buxton AE, Kennedy HL, Knight BP, Mason JW, Schuger CD, Tracy CM, Winters WL Jr, Boone AW, Elnicki M, Hirshfeld JW Jr, Lorell BH, Rodgers GP, Tracy CM, Weitz HH. 2001. ACC/AHA clinical competence statement on electrocardiography and ambulatory electrocardiography: a report of the ACC/AHA/ACP-ASIM task force on clinical competence (ACC/AHA Committee to develop a clinical competence statement on electrocardiography and ambulatory electrocardiography endorsed by the International Society for Holter and noninvasive electrocardiology. Circulation 104:3169–78.

Lewis D. 1999. Computer-based Approaches to Patient Education: a review of the literature. Am Med Inform Ass 6:272–282.

Mason JW, Froehlicher V, Gettes LS. 2001. *Electrocardiography Self-Assessment Programs. In Print and CD-ROM [computer program]. ECGSAP III* (Bethesda, Maryland, American College of Cardiology).

Medi-Smart (2007) ACLS and EKG Guides. University of Cincinnati. Available online at: http://medi-smart.com/software.htm (Accessed 15 December 2007).

Salerno SM, Alguire PC, Waxman HS. 2003a. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. Ann Intern Med 138:751–760.

Salerno SM, Alguire PC, Waxman HS. 2003b. Training and competency evaluation for interpretation of 12-lead electrocardiograms: recommendations from the American College of Physicians. Ann Intern Med 138:747–750.