Background: Cardiac examination is an essential aspect of the physical examination. Previous studies have shown poor diagnostic accuracy, but most used audio recordings, precluding correlation with visible observations. The training spectrum from medical students (MSs) to faculty has not been tested, to our knowledge.

Methods: A validated 50-question, computer-based test was used to assess 4 aspects of cardiac examination competency: (1) cardiac physiology knowledge, (2) auditory skills, (3) visual skills, and (4) integration of auditory and visual skills using computer graphic animations and virtual patient examinations (actual patients filmed at the bedside). We tested 860 participants: 318 MSs, 289 residents (225 internal medicine and 64 family medicine), 85 cardiology fellows, 131 physicians (50 full-time faculty, 12 volunteer clinical faculty, and 69 private practitioners), and 37 others.

Results: Mean scores improved from MS1-2 to MS3-4 ($P=.003$) but did not improve or differ significantly among MS3, MS4, internal medicine residents, family medicine residents, full-time faculty, volunteer clinical faculty, and private practitioners. Only cardiology fellows tested significantly better ($P<.001$), and they were the best in all 4 subcategories of competency, whereas MS1-2 were the worst in the auditory and visual subcategories. Participants demonstrated low specificity for systolic murmurs (0.35) and low sensitivity for diastolic murmurs (0.49).

Conclusions: Cardiac examination skills do not improve after MS3 and may decline after years in practice, which has important implications for medical decision making, patient safety, cost-effective care, and continuing medical education. Improvement in cardiac examination competency will require training in simultaneous audio and visual examination in faculty and trainees.

Arch Intern Med. 2006;166:610-616

CARDIAC EXAMINATION (CE) is a multisensory experience that requires integration of inspection, palpation, and auscultation in the context of initial symptoms and patient history. When CE is performed correctly with attention to all of these modalities, most structural cardiac abnormalities can be accurately detected or considered in a differential diagnosis. This practice enables more appropriate and expedient diagnostic and therapeutic management decisions. However, CE skills are seemingly in decline,1-9 and trainees often perform physical examinations inaccurately.1 One study2 reported serious errors in two thirds of the patients examined.

Despite widespread recognition of this problem,3,6,7 efforts to improve CE skills are hampered by many obstacles: the scarcity of “good teaching patients”; the lack of teaching time at the bedside; the promotion of newer, more expensive diagnostic modalities; and the shortage of clinically oriented instructors competent in CE. Several decades ago, patients’ hospital stays were long, providing trainees and their instructors frequent opportunities for bedside teaching rounds. Today, hospital admissions are short and intensely focused, with fewer opportunities for trainees to learn and practice bedside examination skills. Attending physicians, having been trained in this environment, further amplify the problem if their own CE skills are not well developed.

Teaching strategies designed to mitigate these problems include audio recordings, multimedia CD-ROMs, electronic heart sound simulators, and mannequins, in order of increasing cost from less than $50 to more than $75 000. Each of these modalities can be used for training and testing of CE proficiency. However, mannequins, no matter how sophisticated, cannot replace contact with patients.

See also pages 603 and 617

Author Affiliations are listed at the end of this article.
When audio recordings and electronic simulators are used as surrogates for patients, the assumption is promulgated that cardiac auscultation with eyes closed is sufficient for diagnostic purposes. In contrast, the expert clinician relies on ancillary visible and palpable clues while listening to establish the timing of audible events in the cardiac cycle and to glean additional diagnostic information from the contours of the arterial, venous, and precordial pulsations. The skills required to process multiple senses simultaneously cannot be effectively taught with textbooks and are best acquired by exposure, practice, and testing for competence.

Until now, no convenient, reliable, and objective method of measuring CE skills has been available. Previous studies of CE skills\(^5,7,8,10\) have evaluated only auscultation, making the results difficult to extrapolate to actual patient encounters, where a palpable or visible pulse can aid in timing systolic and diastolic events heard through a stethoscope. The studies commonly focused on 1 or 2 training levels, \(^3,5,7-12\) and none studied the entire spectrum of physicians from students to faculty. Studies of practicing physicians\(^3,14\) are few, and no studies have evaluated the CE skills of internal medicine faculty, who largely teach this skill. Finally, it is difficult to compare results among different studies owing to the variety of methods used.

To address these needs we developed and validated\(^15\) a test\(^16\) of competency in CE skills that uses audiovisual recordings of actual patients with normal and abnormal findings and characteristic precordial and vascular pulsations. Questions tested (1) knowledge of cardiac physiology, (2) auditory skills, (3) visual skills, and (4) integration of auditory and visual skills using recordings of actual patients. To allow meaningful comparisons of competency at all training levels, we tested medical students (MSs), internal medicine (IM) and family medicine residents, cardiology fellows (CFs), full-time faculty (FAC), volunteer clinical faculty, and private practice physicians.

We hypothesized that by using a more realistic test, trainees, faculty, and practicing physicians would score higher than students, as suggested in a preliminary study.\(^10\) Students and trainees commonly ignore the precordial, carotid, and jugular venous pulsations, and they also tend to identify every murmur and extra sound as systolic. Therefore, we also determined sensitivity and specificity for detecting diastolic and systolic events.

### METHODS

#### CE TEST

The CE Test (Blaufuss Medical Multimedia) is a 50-question, interactive, multimedia, computer-based test. It combines computer graphics animations and virtual patient examinations (VPEs) (actual patients filmed at the bedside). Previous research established the reliability and validity of this measure of cardiac auscultation competency.\(^15\) The first questions were computer graphics animation based and were intended as an introductory “warm-up” that required combining observations with auscultation. These questions were an “open book” review of normal pressure-sound correlations and the expected auscultatory findings (graphically and audibly displayed) and their causation in mitral and aortic stenosis. The remaining questions consisted of audiovisual recordings of patients (VPEs).\(^16-18\) Only scenes with clearly visible arterial pulsations and discernible heart sounds and murmurs were selected. These seamlessly looped scenes were filmed from the examiner’s perspective, with the heart sounds recorded from the stethoscope. The VPEs require recognition of pathologic alterations in sounds and murmurs, establishing their timing by correlation with visible pulsations, and differentiating carotid from jugular venous pulsations. Synchronous electrocardiograms and pulse sweeps were not available for VPEs because they are not available at the bedside.

Test content was determined using a 1993 published survey of IM program directors that identified important cardiac findings\(^5\) and Accreditation Council for Graduate Medical Education training requirements for IM residents\(^9\) and CFs.\(^20\) We tested for recognition of (1) sounds (ejection sound, absent apical first sound, opening snap, and split sounds) and (2) murmurs (systolic [holosystolic, middle, and late], diastolic [early, middle, and late], and continuous murmur). Examinees were not asked for a diagnosis but rather for bedside findings that provided pertinent diagnostic information. Six academic cardiologists reviewed the test, and minor content revisions were made accordingly.

To test for knowledge of cardiac physiology, participants were required to interpret animations of functional anatomy with graphical pressures and phonocardiograms synchronized with heart sounds at the apex and base. To test auditory skills, participants were required to identify the presence and timing of extra sounds (eg, near first or second sound) and murmurs (as systolic, diastolic, both, or continuous). More than 1 listening location was provided when appropriate. To test visual skills, participants were required to differentiate carotid and jugular pulsations in audiovisual recordings. For the integration of auditory and visual skills, participants were required to place the sounds and murmurs properly within the cardiac cycle or, conversely, to use the sounds to time visible pulsations.

### SAMPLE AND STUDY SITES

Between July 10, 2000, and January 5, 2004, 860 volunteers at 16 different sites (15 in the United States and 1 in Venezuela) were tested. The sites included 8 medical schools, 7 teaching hospitals, and 1 professional society continuing medical education meeting. Table 1 summarizes the participants and study sites: 318 MSs, 225 IM residents, 64 family medicine residents, 85 CFs, 131 physicians (50 FAC, 12 volunteer clinical faculty, and 69 private practice physicians), and 37 other health professionals. Most of the practicing physicians were internists (Table 2). The “other” group included 10 nurses, 15 other physicians (1 in research, 2 in administration, 1 geriatrics fellow, 3 of unknown specialty, 5 internists with an unreported training level, and 3 part-time/retired), and 12 participants who did not identify their level of training. Six incomplete tests were omitted from the analysis.

### TESTING

The examination consists of 34 true-false and 16 four-part multiple-choice questions and requires approximately 25 minutes to complete. In Venezuela, the test questions were translated into Spanish by a coauthor (L.G.-M.). Participants listened through stethophones or their own stethoscopes applied to individual speaker pads while simultaneously observing a computer monitor or digitally projected image.

Institutional review boards determined that the study was exempt research under clauses 1, 2, and 4 of the Code of Federal Regulations\(^14\) and did not require obtaining written consent. Most testing was performed in conference rooms. Indi-
individuals could take the examination by entering answers on paper or directly into the computer. Continuing medical education meeting attendees were tested at the beginning of a cardiac auscultation workshop. Answers from both testing modalities were collected and maintained in a secure file to prevent access to the answers. All the tests were scored by 2 independent graders and were confirmed by automated scoring on a computer. Two points were awarded for each correct answer, 1 point was subtracted for each incorrect answer, and blank answers were counted as 0 points, for a possible total of 100 points.

STATISTICAL ANALYSIS

To test for differences in CE competency, we compared the mean test scores of the different groups using 1-way analysis of variance (F test). The Levene statistic was computed to test for homogeneity of group variances. After a significant F score, post hoc pairwise mean comparisons were made using the Newman-Keuls test (for homogeneous group variances) or the Games-Howell test (for heterogeneous group variances). Statistical significance was set at $P < .05$. Analyses were performed using statistical software (SPSS version 13.0; SPSS Inc, Chicago, Ill). To test for sensitivity and specificity for detecting systolic and diastolic events, questions that directly asked for the timing of a heart sound or murmur were analyzed separately.

RESULTS

CE COMPETENCY SCORES

Table 3 lists descriptive statistics for each subgroup of students, residents, fellows, and physicians by year of training or academic affiliation. The mean ± 95% confidence interval for mean scores were as follows: MS1-2, 52.4 ± 2.6 (n = 95); MS3, 58.5 ± 2.46 (n = 157); MS4, 59.1 ± 3.8 (n = 66); IM residents, 61.5 ± 1.9 (n = 225); family medicine residents, 56.6 ± 3.4 (n = 64); CFs, 71.8 ± 3.5 (n = 85); FAC, 60.2 ± 4.1 (n = 50); volunteer clinical faculty, 56.1 ± 9.8 (n = 12); and private practice physicians, 56.4 ± 4.1 (n = 69). The mean ± 95% confidence interval for mean score for “other” was 47.51 ± 6.51 (n = 37).

Mean CE competency scores by training level are plotted in Figure 1. Mean scores improved from MS1-2 to MS3-4 ($P = .003$). However, no improvement was observed thereafter: mean scores for practicing physicians, including faculty, were no better than those for MSs and residents. Only CFs tested significantly better than all other groups ($P < .001$).

SUBCATEGORIES OF COMPETENCY

Figure 2 plots comparisons of mean scores for each level of training. Training levels that fall into a distinctly similar grouping after statistical comparisons of mean scores are plotted in the same horizontal stratum, with the best-performing group at the top and lower-performing groups at lower strata. Overall competency scores are plotted at the top of the figure, followed by subcategories that measure competence in 4 aspects of physical examination: basic cardiac physiology knowledge (interpretation of pressures, sounds,
and flow related to cardiac contraction and relaxation), auditory skills, visual skills, and integration of auditory and visual skills. For overall competency, and for each subcategory, CFs scored the best of all the groups tested and were in the top test stratum, that is, the best-performing group. They especially excelled in visual skills compared with other groups. At the other end of performance, MS1-2 was consistently found in the lowest test stratum, along with the “other” group. Mean scores for full-time faculty were not significantly better than those for students, residents, or other practicing physicians in any subcategory.

SENSITIVITY AND SPECIFICITY FOR DETECTING SYSTOLIC AND DIASTOLIC EVENTS

To test the accuracy of the study participants’ ability to identify a heart sound as systolic or diastolic, we separately analyzed questions that required differentiating these events from the professional society continuing medical education meeting (n=192). Twenty-six questions asked participants to place a sound or murmur in the correct phase in the cardiac cycle: only 66% of these questions were answered correctly (Table 4).

We observed a large difference in their ability to recognize an isolated systolic (84%) vs a diastolic (49%) murmur (P<.001). The sensitivity for systolic murmurs was relatively high (0.84), but with a very low specificity (0.35). When tested with diastolic murmurs, the sensitivity was no better than chance (0.49); specificity was higher (0.67).

In this evaluation of CE competency across the medical training spectrum, we found that CE skills reached a plateau in MS3 and did not improve thereafter. An important exception to this trend was CFs, the highest-performing group overall and in the top test stratum for each of the 4 subcategories of competence tested. Administering the same test to the teachers was revealing: faculty performance overall was not statistically different from that of students, residents, or other practicing physicians. High sensitivity and low specificity for detecting systolic murmurs revealed a “systolic bias” among practicing physicians, including most faculty tested, and suggest that participants were not using the carotid pulse to establish the timing of murmurs in the cardiac cycle. These results identify areas for improvement in undergraduate, graduate, and continuing medical education. However, skills at any training level are not likely to improve without first addressing the shortage of instructors who are competent in CE.

Previous studies showed, that one hears is influenced by what one sees, and vice versa. For cardiac auscultation, it is easier to hear a murmur in systole when inspecting an arterial pulse or precordial lift, just as it is easier to see these pulsations while listening.

By implementing a more realistic test, we expected that study participants would recognize diastolic and systolic events more easily and that faculty would be better at integrating sight and sound than students. However, we observed that most participants listened with their eyes closed or averted, actively tuning out the visual timing reference that would help them answer the question. (Relying on the electrocardiogram from a monitored bed during auscultation is actually misleading because the displayed electrocardiogram for a patient is variably delayed by at least half a cardiac cycle.) Listening without the benefit of a visual timing reference may be the source of a common misconception that diastolic murmurs are rare and difficult to hear (obvious murmurs must, therefore, be systolic) and that loud murmurs are most likely holosystolic. Analysis of test questions revealed that par-
Participants exhibited this systolic bias when listening. The clinical consequence is that diastolic murmurs, which are always pathologic, may be underdetected, as they were in this test.

Markedly lower scores for the “other” group may reflect the diversity of participants, which included nurses, researchers, and administrators, and the relatively large number of physicians (11 of 12 overall) who did not identify their training level at the IM society meeting: these scores were among the lowest tested.

Schools and training programs, as well as the licensing and accreditation agencies, are motivated to improve CE skills teaching and testing. The number of training programs reporting auscultation teaching has doubled in the past 10 years, the United States Medical Licensure Examination has added a clinical skills component to the Step 2 Examination, and the American Board of Internal Medicine has added a physical diagnosis component to its recertification program. The United States Medical Licensure Examination’s Step 2 Clinical Skills Examination relies on standardized patients: healthy individuals (actors) who are trained to present a consistent history and symptoms but who cannot present appropriate cardiac findings. Currently, this assessment of CE skills is limited to documenting the process of what was done, not any physical findings. To recreate cardiac findings, the American Board of Internal Medicine’s clinical skills module uses digital video of a mannequin with simulated heart sounds. In these simulations, the pulse is visualized with moving cotton swabs placed obliquely on the mannequin because the mechanical pulse is not visible in the compressed video.

The superior scores from CFs could be explained by considering them a special population of trainees with a greater interest and aptitude in CE. A more likely explanation is that CFs are using more information from the patient. As Figure 2 shows, CFs excelled in all test subcategories, especially visual skills. One should not expect that these skills are unique to CFs: MS3s have improved to the level of first-year CFs after minimal training in VPE.

To improve CE skills, we recommend that MS1-2 be introduced to normal and abnormal findings that include visual, auditory, and palpable examples from actual patients. Students should be able to distinguish venous from arterial pulsations and should become accustomed to looking while listening. This multimedia training should be reinforced in MS3 and MS4 by using visual and palpable information to distinguish systole from diastole using multimedia programs and

---

**Figure 1.** Mean test scores for cardiac examination competency by training level. The dotted horizontal line indicates the mean score for all participants (59.24). The mean score for full-time faculty (FAC) was not significantly different from that of medical students, internal medicine (IM) residents, family medicine (FM) residents, or other practicing physicians (volunteer clinical faculty [VCF] and private practice [PP]). Mean scores were improved in third- and fourth-year students compared with first- and second-year students ($P=.003$), but they did not improve thereafter. Asterisk indicates $P=.045$. Error bars represent 95% confidence intervals.
during patient encounters. In residency training, findings from cardiac laboratory studies should be compared and correlated with bedside findings. Finally, faculty and other practicing physicians must be included in multimedia training throughout their career to ensure improvement in patient safety and cost-effective patient care and teaching.

The relatively low scores for faculty may be explained by the prevalence of internists over cardiologists in the study sample. By design, this study focused on faculty who now largely teach physical examination and diagnosis because cardiologists have become less involved in teaching MSs and residents.20 Although the test presented physiologic and pathologic bedside findings using audiovisual recordings of actual patients, further studies may confirm whether test scores can be directly equated with the ability to make an appropriate observation or diagnosis in an actual patient.

With the previously mentioned limitations noted, this study contributes to the ongoing efforts in improving CE skills teaching and testing. These findings provide detailed assessment of CE skills in a broad sample across the medical training spectrum, showing that CE skills did not improve after MS3, with a particular weakness in identifying diastolic events. Failure to use both visual and auditory information from patient examinations is one explanation for this poor performance, and it may be a consequence of how most have been trained with audio-only recordings. Finally, faculty performance indicates that they must be included in any training efforts.
Participants had high sensitivity for detecting systolic murmurs (84%) but by the sum of positive, negative, and blank answers for false. The same for true. Systolic specificity is the measure of false-negative answers divided asked for correct timing. Systolic sensitivity is the measure of true-positive answers.

before we can expect better CE skills in physicians as a whole.

Accepted for Publication: October 9, 2005.

Author Affiliations: Stanford University School of Medicine, Stanford, Calif (Dr Vukanovic-Criley); Blaufuss Medical Multimedia, Palo Alto, Calif (Mr Criley); Internal Medicine Residency Program (Dr Warde) and Department of Family Medicine and Medical Education Research (Dr Boker), University of California, Irvine College of Medicine; Cardiology Fellowship Program, Jose Maria Vargas School of Medicine, Central University of Venezuela, Caracas (Dr Guevara-Matheus); Department of Hematology, Harvard Medical School, Boston, Mass (Dr Churchill); Department of Cardiology, University of Colorado, Denver (Dr Nelson); and Department of Medicine and Radiological Sciences, Harbor-UCLA Medical Center and UCLA School of Medicine, Torrance and Los Angeles, Calif (Dr Criley).

Correspondence: Jasminsina M. Vukanovic-Criley, MD, 270 Alameda de las Pulgas, Redwood City, CA 94062 (jmcriley@gmail.com).

Financial Disclosure: None.

Funding/Support: This study was supported by grants 1R43HL062841-01A1, 2R44HL062841-02, and 2R44HL062841-03 from the National Heart, Lung, and Blood Institute, Bethesda, Md.

Role of the Sponsor: The funding source was not involved in the design, conduct, or reporting of the study or decision to submit this manuscript for publication.

Acknowledgment: We thank our patients for their willingness to have their sounds and images used in a teaching program for the benefit of many patients to come; the students, residents, fellows, and practicing physicians who participated in the studies; David Criley, BA, who developed the test used in the study; Jonathan Abrams, MD, Rex Chiu, MD, MPH, Gregg Fonorow, MD, Victor F. Froelicher, MD, Andrea Hastillo, MD, Steve Lee, MD, Joseph P. Murgo, MD, Ronald Oudiz, MD, Shobita Rajagopalan, MD, George Vetrovec, MD, and Jan H. Tillisch, MD, for collaborating in the multilevel proficiency study; the following individuals at UCLA: LuAnn Wilkerson, EdD, Patricia Anaya, Kimberly A. Crooks, PhD, Sylvia A. Merino, MBA, MPH, and Anita Skaden; and Patrick Alguire, MD, and Edward B. Warren, BA, of the American College of Physicians for allowing us to participate in the clinical skills workshops at the 2001 annual meeting.

### Table 4. Sensitivity and Specificity for Detecting Systolic and Diastolic Events

<table>
<thead>
<tr>
<th>Test Answer</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systolic murmurs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>677</td>
<td>235</td>
</tr>
<tr>
<td>Negative</td>
<td>84</td>
<td>143</td>
</tr>
<tr>
<td>Left blank</td>
<td>47</td>
<td>26</td>
</tr>
<tr>
<td>Subtotal</td>
<td>808</td>
<td>404</td>
</tr>
<tr>
<td>Diastolic murmurs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>298</td>
<td>145</td>
</tr>
<tr>
<td>Negative</td>
<td>271</td>
<td>403</td>
</tr>
<tr>
<td>Left blank</td>
<td>37</td>
<td>58</td>
</tr>
<tr>
<td>Subtotal</td>
<td>666</td>
<td>606</td>
</tr>
</tbody>
</table>

*Systolic or diastolic murmurs were used to construct questions that asked for correct timing. Systolic sensitivity is the measure of true-positive systolic answers divided by the sum of positive, negative, and blank answers for false. The same method is used to calculate specificity and sensitivity for diastolic murmurs. Participants had high sensitivity for detecting systolic murmurs (84%) but with low specificity (35%). For diastolic murmurs, the reverse was true: a low sensitivity (48%) was combined with a higher specificity (67%).

### REFERENCES

21. 45 CFR § 1, subpart 101 (revised October 1, 2004).


