DOCUMENT RESUME

ED 382 686 TM 023 227

AUTHOR	Chapman, Dane M.; And Others
TITLE	Critical Emergency Medicine Procedural Skills: A Comparative Study of Methods for Teaching and Assessment.
PUB DATE	Apr 93
NOTE	44p.; Paper presented in part at the Annual Meetings of the Society for Academic Emergency Medicine (San Francisco, April 1993) and the International Conference in Emergency Medicine (London, England, April 1994). This study was supported in part by the Society for Academic Emergency Medicine/Emergency Medicine Foundation Educational Grant Award.
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Animals; Comparative Analysis; Competence; Computer Assisted Testing; *Computer Simulation; Educational Assessment; *Evaluation Methods; Experience; Medical Services; *Medical Students; Performance; *Physicians; *Skills; Teaching Methods; Test Construction
IDENTIFIERS	*Emergency Medical Services

ABSTRACT

Three critical procedural skills in emergency medicine were evaluated using three assessment modalities--written, computer, and animal model. The effects of computer practice and previous procedure experience on skill competence were also examined in an experimental sequential assessment design. Subjects were six medical students, six residents, and six medical faculty members. They completed a survey of their prior experience with thoracotomy, were provided with standardized instruction in the procedures, and were tested with the three modalities. Competence was evaluated in terms of performance time and accuracy. Results indicated that critical emergency medical procedural skills can be evaluated most reliably and validly using an animal model assessment with sufficiently complex anatomy, as represented by the pig used in this study. Results also demonstrated that computer simulation practice using visual imagers and sequential ordering of procedural steps shows promise in teaching and assessment of competence. Previous thoracotomy experience was not an accurate predictor of competence. One figure and four tables present study findings. An appendix presents the procedural steps. (Contains 14 references.) (SLD)

Critical Emergency Medicine Procedural Skills: A Comparative Study of Methods for Teaching and Assessment

Authors:
Dane M. Chapman, M.D., PhD., University of California-Davis, currently at York Hospital, Pennsylvania State University at Hershey
John A. Marx, M.D., Carolinas Medical Center
Ben Honigman, M.D., University of Colorado Health Science Center
Dennis Martinez, M.D., University of California-Davis, currently at UCLA Kern County Medical Center
Kenneth Rhee, M.D., University of California-Davis, currently at University of Nebraska Medical Center
B. Tomas Brofeldt, M.D., University of California-Davis
Edward A. Panacek, M.D., University of California-Davis
Sally H. Cavanaugh, PhD., Emig Research Center of York Hospital

This study was supported in part by the Society for Academic Emergency Medicine/Emergency Medicine Foundation Educational Methodology Grant Award. Manuscript preparation by Sharon Pitzer and Becky Bluett.

Presented in part at the Society for Academic Emergency Medicine Annual Meeting, San Francisco, April 1993 and at the Fifth International Conference in Emergency Medicine, London, April 1994.

Address for reprints:
Dane M. Chapman, M.D., PhD.
Emergency Medicine Residency Program
Department of Emergency Medicine
York Hospital
1001 S. George Street
York, PA 17405
717-851-2450
717-851-3469 (FAX)

Structured Abstract

Critical Emergency Medicine Procedural Skills: A Comparative Study of Methods for Teaching and Assessment

Study Objectives: To objectively evaluate three critical procedural skills using three assessment modalities (written, computer, and animal model), and to determine the effects of computer practice and previous procedure experience upon skill competency.

Design: An experimental sequential assessment design was utilized so examinees could serve as their own controls. Subjects completed a survey of their prior thoracotomy experience, were provided standardized instruction and computer practice, and were subsequently tested using three assessment modalities. The reliability and validity of the three assessment modalities were compared.

Setting: The study was conducted at the animal support facilities of a university medical school with an affiliated emergency medicine residency program.

Type of Participants: Participants were volunteers selected to evenly represent three levels of physician training (medical student, resident and faculty).

Interventions: Level of physician training (medical student, resident, faculty) and type of computer practice (cricothyrotomy, thoracotomy) were independent variables. Procedural competency scores were outcome measures. The defining variables included previous thoracotomy and computer experience.

Measurements: Procedural competency was defined in terms of performance time (animal time scale) and performance accuracy (written accuracy, computer accuracy, and animal accuracy scales) for the three thoracotomy procedures (opening the chest, pericardiotomy, and aortic cross-clamping).

Results: Thoracotomy performance on the animal reliably discriminated among examinees known to differ in level of training. However, computer simulation performance did not significantly differ among levels of training. Computer simulation practice significantly improved later performance on the computer assessment ($p<.05$), but not on the animal assessment. The greatest predictor of procedural performance, in terms of time and accuracy on the animal assessment was the prerequisite ability to sequentially order procedural steps. Previous thoracotomy procedural experience was not a significant predictor of thoracotomy performance time or accuracy.

Conclusions: This study demonstrates that critical emergency medicine procedural skills can be evaluated most reliably and validly using an animal model assessment with

sufficiently complex anatomy (e.g., pig model). It also demonstrates that computer simulation practice using visual images (complex anatomy) and sequential ordering of procedural steps show promise in the teaching and assessment of procedural competency. Previous thoracotomy experience, however, is not an accurate predictor of thoracotomy competency.

Introduction

Little standardization of procedural skill learning exists among medical schools and training programs and often there is no assurance that medical students, residents or faculty physicians can perform procedures well.[1-4] Certain procedures such as cricothyrotomy and thoracotomy must be performed rapidly and expertly to be lifesaving. The bedside is no place to learn or practice such procedures. Since these and other critical procedures are often rarely indicated, medical students and residents have difficulty acquiring procedural competency.[1-3]

In response to this need, various procedural skill teaching workshops have been developed.[1,2,5-7] However, procedural competency assessment remains essentially unexplored, and procedure certification by specialty boards is virtually nonexistent.[1-4,8-10] Typically, the number of procedures performed has been taken as predictive of procedural competency.[2-9] Even faculty ratings of critical procedures performed on live patients are unstandardized, and are problematic due to the relatively low frequency with which procedures are performed before the same faculty observer.

To refine our assessment of critical procedures, we chose thoracotomy as a prototypic critical procedure in a prior study.[2] We then developed objective written, computer simulation and animal assessments of three thoracotomy procedures: opening the chest, pericardiotomy, and aortic cross-clamping.[1,2,11] We next compared the reliability and validity of the three assessment modalities (written, computer, animal) over three levels of physician training (medical students, residents and faculty physicians). The animal (dog) model was used to represent the criterion-standard with the assumption that performance on the animal would be predictive of actual performance on patients in the emergency department.[2] The computer simulation scores correlated significantly with dog thoracotomy performance times. An interesting finding from this same research was that medical students, after memorizing critical procedural steps and watching a 15-minute videotape demonstration of thoracotomy procedures, were able to perform thoracotomy procedures on the animal (dog) assessment with surprising accuracy. It was less clear whether examinees with little thoracotomy expertise such as medical students would do as well on a real patient.

During the prior study, it became clear that the dog model has inherent problems. Most notably, identification of the aorta and other structures critical to successful thoracotomy, pericardiotomy and aortic cross-clamping are readily identifiable on the dog model. In contrast, the pig model more closely resembles the size and anatomy of the human, where structures are more difficult to identify. The pig was not used in the previous study as pigs are considerably more expensive and difficult to work with. Could medical students without previous "hands-on" experience perform complex surgical procedures as accurately as residents and faculty on a model more closely resembling the human?

This paper describes the follow-up study using the pig model that was conducted to: (1) evaluate the reliability and validity of revised written, computer and animal (pig) assessments; (2) determine if computer practice improves procedural skills for performing thoracotomy; and (3) investigate the relationship between the number of thoracotomy procedures performed prior to the study and procedural skill demonstrated on the animal model criterion-standard at the conclusion of the study.

Overview of Procedural Test Parameters

A reliable test must provide reproducible or consistent results. For example, an examinee taking the written test a second time without additional study would be expected to obtain a score similar to the first test score (test-retest reliability). Similarly, an examinee should score consistently on similar procedures within a group of procedures during the same assessment (internal consistency reliability).

A valid test measures what it purports to measure. Typically four types of validity are described:[12]

1. Face validity relates to validity that is based upon face value. For example, animal assessment scores would be expected to have greater face validity since the animal model more closely resembles the human than a written or computer simulation. As such, the animal assessment has been taken as the criterion-standard.
2. Content validity describes validity that is based upon expert opinion. In other words, content experts authorize the validity of the measure. In this study, emergency thoracotomy experts were consulted in the development of all assessments.
3. Construct validity refers to whether a given measure accurately identifies populations that are expected to differ along the assessment. In this study, the construct validity of written, computer and animal thoracotomy assessments would be supported if they can discriminate among levels of physician training (medical student, resident, faculty).
4. Predictive validity/concurrent validity (i.e., criterion-related validity) reflects whether a previously unvalidated measure accurately predicts competency to the same degree as that of an accepted criterion-standard. In this study, written and computer assessments would demonstrate concurrent

validity if assessment scores correlate with the animal criterion-standard scores.

Methods

Study Design

Three thoracotomy assessment modalities (written, computer, animal) were compared over three levels of physician training (student, resident, faculty) using an experimental, sequential assessment design (see Figure 1). Five research questions were addressed by the study design: 1) What are the reliabilities (Cronbach's alpha measure of internal consistency) of the written, computer, and animal model assessments? 2) What are the construct validities of the assessments (i.e., do written, computer, and animal model assessments discriminate among physician training levels as expected)? 3) What are the concurrent validities of the written and computer assessments compared to the animal criterion-standard? 4) What are the effects of computer practice on procedural performance as measured by the three assessment modalities? 5) Is past thoracotomy experience (as reported by subjects) related to procedural competency?

Sample and Setting

Each thoracotomy assessment modality was tested over three levels of training where examinees were expected to differ in procedural competency: six fourth year medical students, six senior emergency medicine residents, and six emergency medicine faculty. The study was conducted at the surgical support facilities of the University of California-Davis School of Medicine. The protocol was approved by the Animal Use Committee.

Thoracotomy Training Laboratory

Prior to procedural testing, all examinees were provided critical content (factual knowledge) instruction during a thirty minute training laboratory. Examinees were instructed to memorize procedural steps, and then viewed a video demonstration of both thoracotomy and cricothyrotomy procedural steps using the computer simulation. This portion of the training was designed to provide all examinees with sufficient content knowledge and visual image (complex anatomy) training to be able to perform the thoracotomy procedures. Participants were next randomized to either thirty minutes of thoracotomy (treatment) or cricothyrotomy (control) computer practice. Examinees also completed a written survey to ascertain their previous thoracotomy and computer experience.

Testing Protocol

All examinees were tested using each of the three thoracotomy assessment modalities (written, computer, animal). Examinees were administered the written assessment immediately following the thoracotomy training laboratory. Upon completion of the written assessment, all examinees were randomized by coin toss to computer and animal assessments using a counterbalanced experimental design to control for sequence effects (see Figure 1). Three examinees from each level of training underwent thoracotomy evaluation using the animal model followed by the computer simulation. The other three examinees from each level of training underwent evaluation using the computer simulation followed by the animal model. Examinees were allowed approximately thirty minutes to complete each evaluation segment.

Thoracotomy Assessment Modalities

The three thoracotomy assessment modalities (written, computer, animal) are described below and in further detail elsewhere.[2] Identification of the thoracotomy procedural critical steps and sequencing information were established by consensus among content experts.

Written Assessment

The written (paper and pencil) simulation required subjects to sequentially order the procedural steps involved in performing a thoracotomy. For example, the thirteen steps for opening-the-chest (See Appendix A) were presented in random order, requiring the subjects to place the numbers 1 through 13 in the appropriate blank preceding each step. This process was repeated for the set of eight steps randomly presented for performing pericardiotomy, and the ten steps for aortic cross-clamping. A total score for proper sequencing of all three thoracotomy procedures was derived as a percentage score, termed written accuracy.

Computer Assessment

An interactive, computer-based multimedia system had been developed previously for teaching and evaluating thoracotomy skills using the IBM InfoWindow system.[1,2,11] This system was updated to replace the touch screen with a mouse function. A physician skilled in using the computer simulation and in thoracotomy assessment stood by to assist examinees with any hardware or software problems. Examinees first entered their name using the keyboard. They then "performed a thoracotomy" using a mouse and menu set to indicate each procedural action step with corresponding instruments and materials needed. They also indicated placement, orientation and extent of each action. Next, a stillframe or motion video displayed the appropriate procedural action according to the previously designated optimal sequence. A computer accuracy score was subsequently derived from the electronically generated data set to describe the percent of correctly completed critical procedural steps.

Animal Model Assessment

Eighteen Hampshire pigs (average weight 39.5 kg.) were premedicated with 0.5 mg./kg. atropine, and 0.25 mg./kg. astepromazine subcutaneously prior to induction with ketamine, 20 mg./kg. intramuscular. They underwent tracheal intubation and were maintained with 1-2 micrograms continuous halothane, using a Droper A-V or Metromotic ventilator. Once anesthetized, pigs were shaven with standard barber's clippers from the left forearm to the left iliac crest. A standard thoracotomy tray was provided with 8 x 11" rib retractors, needle holders, #22, #10, and #11 scapel blades, vascular clamps, small and large Mayo scissors, 8" Mitzenbaum's scissors, alligator tooth forceps, 4 x 4" gauze dressings, Trousseau dilator, tracheal hook, betadine, and suction. A dot was used to mark the pig's mid-thorax to standardize examinees in their initial incision as pigs have multiple nipples. An RCA model 60-60 camcorder was used to videotape the procedures. One of the authors (DM) observed examinees and recorded procedural times for opening the chest, delivering the heart following pericardiotomy, and aortic cross-clamping. Errors of omission of critical steps were noted, as were any errors of commission (e.g. lacerating the myocardium, cross-clamping the pulmonary artery). Examinees were asked to verbalize their actions as they performed the procedures. Once the thoracotomy procedures were completed, animals were euthanized with Sleep-away anesthesia, 390 mg./kg. pentobarbitol.

Measurements

Thoracotomy procedural competency was defined in terms of performance time (minutes) and performance accuracy (percentage of total possible score) for three thoracotomy procedures (opening the chest, pericardiotomy, aortic cross-clamping) as described previously.[2]

Performance times for opening the chest, pericardiotomy, and aortic cross-clamping respectively were defined as the time (in minutes) from skin incision to chest retraction, from chest retraction to delivery of the heart from the pericardium, and from delivery of the heart to aortic cross-clamping. A performance time scale (animal time scale) was obtained by summing the performance times from the three thoracotomy procedures performed on the animal model. Performance time could be measured meaningfully only for the animal model.

Performance accuracy scores reflected the percentages of the total possible points obtained for the three thoracotomy procedures (opening the chest, pericardiotomy, aortic cross-clamping). Three thoracotomy accuracy scales were constructed by summing the accuracy scores for the three thoracotomy procedures on each assessment modality (written accuracy, computer accuracy, and animal accuracy scales).

Reliability and Validity Measures

Reliability estimates for the three performance accuracy scales were obtained using Cronbach's alpha.[12] This coefficient ranges from 0.0 to 1.0, with a value of 0.5 or higher being considered desirable in comparing groups.

Content and face validity of the assessment modalities were described a priori. However, construct and concurrent validities were defined empirically. Construct validity would be supported if a given measure detected expected differences among physicians known to differ in levels of training (student, resident, and faculty). Concurrent validities for the written and computer assessment modalities were calculated by correlating performance scores with performance on the animal model criterion-standard.

Data Collection and Scoring

Data from each testing modality (written, computer, animal) were collected and numerically coded prior to scoring. A standard scoring protocol (See Table 1) was used to score computer and animal data transcripts by the same investigator (DMC) blinded to examinee level of training and identity.

Data Analysis

Four thoracotomy scales (animal time, written accuracy, computer accuracy, and animal accuracy) were constructed as previously described.[2] Descriptive statistics for the four thoracotomy scales were obtained, and are represented as a group mean ± standard deviation. Categorical variables such as previous computer experience were compared using chi-square statistics. Performance scores on the four thoracotomy scales were compared using parametric analysis of variance (ANOVA), with normality and homogeneity of variance assumptions being met. When a significant overall effect was observed, post-hoc pairwise comparisons of means were made by the conservative Sheffé method. Correlations were obtained using the parametric Pearson Product-Moment method. Level of statistical significance was set at $p<.05$. Analyses were conducted using SPSS PC for Windows, version 6.0.[12]

The sample sizes of six students, six residents, and six faculty were chosen based on a previous study using the dog model.[2] A priori power analyses were not conducted as the pig model was expected to yield different results. However, post-hoc power analysis was utilized to estimate sample sizes for non-significant comparisons needed to detect significant differences at a power of 0.80 with an effect size equal to the difference in means between faculty and medical students.

Results

Eighteen examinees (six medical students, six senior emergency medicine residents, and six emergency medicine faculty) completed the protocol. Medical students, residents and faculty did not differ significantly by previous computer experience or by computer ownership. Significantly more medical students had less thoracotomy lab experience ($p<.05$) compared to residents and faculty. There were no statistically significant differences in computer or animal test performance based on the counterbalanced order in which these occurred.

Reliability by Assessment Modalities

The internal consistency reliability (Cronbach's alpha) of the written (alpha=.51) and computer (alpha=.66) accuracy scales were adequate for subsequent analysis, but were not as high as the animal accuracy scale (alpha=.76).

Construct Validity of Assessment Modalities

Animal performance time (animal time) varied significantly as expected across different levels of physician training, thus demonstrating construct validity of the animal (pig) assessment (see Table 2). Average performance times (animal time scale) for medical students were 30-40 percent longer (15.2 ± 1.8 minutes), compared to residents (8.6 ± 2.2 minutes), and faculty (10.6 ± 3.1

minutes) ($p<.001$). While both resident and faculty physicians took significantly less time than medical students to perform the three thoracotomy procedures, no significant performance time differences were found between faculty and residents. The major performance time difference was noted between medical students and residents for opening-the-chest. Two trends were also observed in the data: (1) medical students uniformly took the longest amount of time, and (2) residents uniformly took the least amount of time. The fact that faculty tended to take more time than residents (though not substantially significant) may have been due, at least in part, to faculty having less recent procedural practice compared to residents.

Thoracotomy procedural accuracy on the animal assessment (animal accuracy) was significantly greater for faculty as compared to medical students (mean = 88.8% vs. 70.9%), thus supporting the construct validity of the animal accuracy assessment. While residents scored considerably higher than medical students (mean = 84.2% vs. 70.9%), this difference did not achieve statistical significance (See Table 3).

A statistically significant difference was also found between medical students and faculty on the thoracotomy written accuracy scale suggesting its construct validity. However, considering the very high mean scores achieved by all groups (>95%), the small (3 point) difference between medical students and faculty would seem to have no practical significance (See Table 3).

The computer assessment was considered to have greater face validity than the written assessment as it incorporates the anatomical visual cues of thoracic structures. Mean computer accuracy scores were considerably lower than written accuracy scores for each group, suggesting a more difficult assessment. While not statistically significant, there was a distinct trend in the computer accuracy scores, with faculty performing at a higher level than residents (mean = 70.5 vs. 61.2), and residents higher than students (mean = 61.2 vs 57.2), thus supporting the construct validity of the computer accuracy scale.

Concurrent Validity of Assessment Modalities

As expected, thoracotomy procedural accuracy on written, computer, and animal assessments correlated negatively with thoracotomy performance time (See Table 4). The concurrent validity of the thoracotomy assessment scales is supported by the finding that examinees who scored higher on the accuracy scales also tended to take less time to complete the animal assessment procedures. However, the inverse correlation of animal accuracy and animal time was not as high as previously reported.[2] The lower correlation is likely due to faculty taking more time to perform procedures while still performing with better accuracy compared to residents and students. Faculty performance times are longer probably due to their having less recent procedural practice compared to residents.

Computer Practice

Table 5 presents the four thoracotomy assessment scale results (written accuracy, computer accuracy, animal accuracy, animal time) by type of computer practice (thoracotomy vs. cricothyrotomy). Each group of nine includes 3 medical students, 3 residents, and 3 faculty. Thirty minutes of thoracotomy (treatment) versus cricothyrotomy (control) computer practice had no significant effect on subsequent thoracotomy test performance on the animal model, the criterion-standard. However, those who practiced thoracotomy on the computer simulation performed significantly better on the thoracotomy computer assessment ($p<.05$) compared to those who practiced cricothyrotomy.

During subgroup analyses, medical students, residents and faculty who practiced thoracotomy on the computer all tended to score higher on thoracotomy computer assessments than examinees who practiced cricothyrotomy. However, only the faculty subgroup scored significantly higher (78.9 ± 5.6; thoracotomy practice) versus (62.0 ± 4.6; cricothyrotomy practice) (t-test, $p < .05$).

Previous Experience

Thoracotomy procedural experience was measured in terms of the number of thoracotomy procedures (opening the chest, pericardiotomy, and aortic cross-clamping) previously performed on either animals or patients. All medical students reported no experience in performing thoracotomy procedures, compared to residents reporting an average 2.5 (range 0-5) procedures and faculty an average 21.0 (range 1-48) procedures ($p<.05$).

These self-reported measures of previous thoracotomy experience failed to correlate significantly with any of the performance scores. The correlation between experience and thoracotomy performance time was $r = -0.06$; for animal performance accuracy $r = 0.45$; for written accuracy $r = 0.40$; and for computer accuracy $r = 0.13$. Accordingly, examinees with the greatest thoracotomy experience did not consistently demonstrate the greatest thoracotomy competency.

Discussion

In assessing clinical competence, medical schools, residency programs and specialty boards have traditionally utilized written multiple choice examinations. These tests are optimal for determinations of factual knowledge. There has been increasing emphasis upon measuring clinical reasoning and assessment skills for which multiple choice examinations are no longer useful. Over the past decade reliable and valid measures of clinical reasoning competence have been designed using live simulated patients, oral patient simulations and computer-based patient management simulations. Still, such measures have yet to be used to formally evaluate procedural competency.[1-3] Even though minicourses such as Advanced Trauma Life Support (ATLS), Advanced Cardiac Life Support (ACLS), and Pediatric Advanced Life Support (PALS) teach and assess psychomotor skills to some degree, they do not relieve medical schools, residency programs, and specialty boards of their responsibility to assure a physician's procedural competency.[1-3]

This study confirms the results of a prior study[2] that suggested previous procedure experience a poor predictor of both performance time and performance accuracy. These consistent findings at two academic medical centers do not support our current practice assumption that procedural competence can be predicted by previous procedural experience, (i.e., number of previous procedures performed). Clearly, if procedures have been incorrectly practiced, experience would not be a valid predictor of procedural competency.

This study suggests critical emergency procedures such as thoracotomy can be evaluated most reliably and validly using an animal (pig) assessment than either written or computer assessments. In fact, the pig assessment better differentiated among levels of physician training than the dog assessment used in the prior study.[2] In the prior study, thoracotomy competency was found to be reliably and validly assessed by both computer accuracy and animal (dog) time assessments only.[2] The written procedure assessment of the prior study had poor reliability and validity. The current study demonstrates a revised written assessment that focused on simply the sequencing of critical steps for each procedure, resulting in a more reliable assessment. It appears that this act of sequencing procedural steps is a critical prerequisite to the accurate performance of procedures, a fact supported by moderate positive correlations between written and animal (pig) assessments in the current study. While written procedural tests should not replace hands-on psychomotor assessments (e.g. animal model), these results do suggest that greater emphasis should be placed upon teaching procedural sequences during procedural training. Memorizing the sequential order of performing critical steps is one of eight essential steps of procedural skill learning previously described.[1]

Computer accuracy scores supported the construct validity of the computer assessment, with faculty having a higher mean than residents, and residents a higher mean than students, although not statistically significant. Post-hoc power analysis suggested a sample size of 15 for each level of training to provide sufficient power (0.80) to detect actual differences using the computer assessment. A possible limitation of the computer assessment was that examinees were given visual information and were to indicate their next step based upon the visual cue received. A more sophisticated computer assessment might provide for better discrimination among levels of training.

The computer assessment of the current study was modified slightly from that of the previous study which better discriminated among levels of training.[2] The touch screen used in the prior study was replaced by a mouse function to facilitate using the overlay menu set. Faculty's improved computer assessment performance during the current study compared to the previous study suggests the importance of the mouse capability.

It is unclear why computer accuracy scores of the current study did not discriminate among levels of training as well as in the prior study. Perhaps sampling differences alone would account for the difference. It is also possible that the thirty minutes of computer practice may have masked differences among levels of training by providing increased familiarity with computer sequencing of critical steps and visual cues among all examinees.

This study again demonstrates that critical emergency procedural skills can be taught to novice fourth year medical students through textual information and visual stimuli alone without previous hands-on psychomotor experience. Still, animal (pig) performance times for medical students were 30-40% greater than residents and faculty, and performance accuracy scores varied as expected among levels of physician training.

Practice is needed for student performance times and accuracy to reach resident and faculty levels. This research suggests that thirty minutes of computer practice facilitates thoracotomy procedural learning relating to the memorization of facts and sequencing information, rule using and visual cue recognition.[1] However computer practice does not improve psychomotor coordination which may require hands-on experience. What is learned during computer practice sessions does not appear to be sufficient to improve thoracotomy psychomotor performance on the animal (pig) assessment. It remains to be shown whether additional computer practice beyond the thirty minutes allowed in this study would be sufficient to improve medical student performance on the animal (pig) assessment. The effectiveness of passive viewing of a procedural demonstration versus practicing the procedure on a computer simulation also remains to be determined.

Even though medical students performed the thoracotomy procedures with marginal accuracy on the animal (pig) assessment, it cannot be generalized that they would perform complex surgical procedures on humans. Still, the pig thorax more closely approximates the human thorax compared to the dog, such that thoracotomy performance using the animal (pig) assessment would more likely reflect actual performance on humans.

This study has limitations. First, the study population included volunteers at three levels of training expected to differ in procedural competency (i.e., student, resident, and faculty). Since the examinees were not randomly selected, it cannot necessarily be concluded that volunteers accurately represented the universes of students, residents, and faculty at the University of California, Davis.

The sample size of six students, six residents and six faculty were chosen for logistical reasons. However, several variables demonstrated significance to suggest an adequate sample size. The notable exception was the computer assessment which required an estimated sample size of 15 in each group to achieve a power of 0.80.

Finally, the data regarding previous thoracotomy experience are retrospective survey data and are thus limited by examinee recall of previously performed thoracotomy procedures. Still, thoracotomy procedures are rare and presumably memorable, and more objective data were not available.

These results should be useful for implementing procedural training and competency assessment by medical schools, residency programs and special boards.

Conclusions

Thoracotomy procedural competency can be measured more reliably and validly using the animal (pig) model assessment than either written or computer assessments. Compared to the dog model, the pig model assessment appears to better discriminate among examinees known to differ in training (medical student, resident, faculty), and thus demonstrates greater construct validity. The written and computer assessments in this study demonstrated similar trends, but were not as significant.

Computer simulation practice significantly improves examinee performance on computer simulation assessment of the same skills, but does not improve psychomotor performance of that skill on the animal assessment. Hence, there did not appear to be transfer of skill from the computer to the animal model. Still, computer simulation practice improves performance on the subsequent computer simulation assessment, and probably reflects increased learning of factual information and visual cues, but not increased psychomotor skill.

Clearly, this study has relevance to credentialing issues for all professions in which critical procedural skills are integral to practitioner competence. This study confirmed the prior study, that previous experience (the number of previous thoracotomy procedures performed), is an uncertain predictor of competency as assessed using a reliable and valid animal (pig) model assessment.

Despite the sophistication of interactive computer simulations for teaching or assessing procedural skills, these too may be inadequate. A combination of virtual reality, robotics and computer simulation technologies may provide greater validity where not only factual information, and visual cue discrimination can be learned and assessed, but critical psychomotor skills can be as well. In this regard, computer-based, image-driven robotics have already been introduced to perform precision drilling of a femur,[1,14] to assist in the Automated Endoscopic System for Optimal Positioning (AESOP) during abdominal laporoscopic procedures,[14] and to perform surgery-at-a-distance during battlefield operations where robotics systems are connected via fiber optics to distant surgeons.[14] Such computer-based systems hold promise in facilitating the teaching and assessment of critical procedural competency in the years ahead.

REFERENCES

1. Chapman DM: Use of computer-based technologies in teaching emergency procedural skills. Acad Emer Med 1994;1(4):404-407.

2. Chapman DM, Marx JA, Honigman B, Rosen P, Cavanaugh SH: Emergency Thoracotomy: A comparison of medical student, resident, and faculty performances on written, computer, and animal model assessments. Acad Emer Med 1994;1(4):373-381.

3. Braen GR, Munger BS: Evaluation of procedural skills. Acad Emer Med 1994;1(4):325.

4. Wigton RS: See one, Do one, Teach one. Acad Med 1992;67:743.

5. Powers LR, Draeger SK: Using workshops to teach residents primary care procedures. Acad Med 1992; 67:743-745.

6. Olshakr JS, Brown CK, Authur DC, Tek D: Animal procedure laboratory surveys: Use of the animal laboratory to improve physician confidence and ability. J Emerg Med 1989;7:593-597.

7. Nelson MS: Models for teaching emergency medicine skills. Ann Emerg Med 1990;19:333-5.

8. Bruce NC: Evaluation of procedural skills of internal medicine residents. Acad Med 1989;64:213-216.

9. Gabryel TF, Brierley MA: Credentialing protocols used by internal medicine residency programs in New York State. Acad Med 1990;65:769-71.

10. Kowlowitz U, Curtis P, Sloane PD: The procedural skills of medical students: Expectations and experiences. Acad Med 1990;65(10):656-658.

11. O'Neil PN: Developing video instruction for the health sciences: a consortial approach. Acad Med 1990;65(10):624-27.

12. Norusis MJ: SPSS for Windows Base System User's Guide, Release 6.0. Chicago, IL, SPSS, 1993.

13. Ware JE, Brook RH, Davies AR, Lohr KN: Choosing measures of health status for individuals in general populations Am J Public Health 1981;17:620-5.

14. Edelson, E. Robo Surgeons. Pop Scien 1995;90:62-66.

Figure 1. Study Design

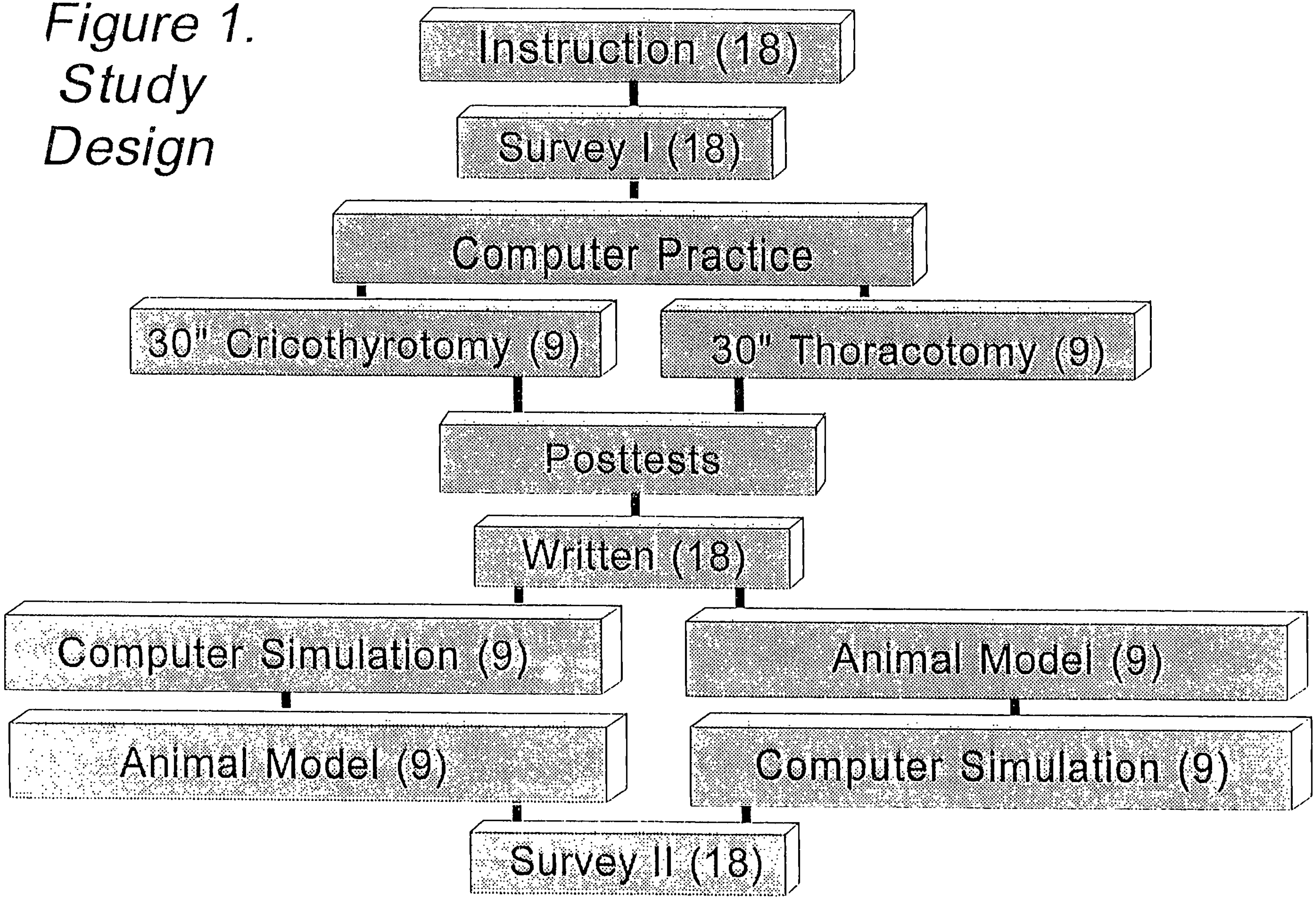

TABLE 1. Procedure Scoring Protocol

Scoring Criteria

Critical Step[1]	Proper Sequence[2]	Performed Correctly[3]	Serious Complication[4]	Points
YES	YES	YES	NO	5
YES	NO	YES	NO	4
YES	YES	NO	NO	3
YES	NO	NO	NO	2
NO	YES	YES	NO	2
NO	NO	YES	NO	1
NO	YES	NO	NO	1
YES	YES	NO	YES	0
NO	YES	NO	YES	0

[1]Critical procedural step such that error of commission or omission would result in unnecessary morbidity or mortality.

[2]Procedural step sequence provided during training session.

[3]Correct instrument, material, orientation and extent for procedural action as taught during training session.

[4]Serious complication resulting in unnecessary morbidity or mortality. (e.g. lacerate myocardium, transect phrenic nerve, etc.)

TABLE 2: Performance Time in Minutes (Mean±SD) for Animal Thoracotomy Procedures by Level of Training

Time Measure	Level of Training			
	Student (n=6)	Resident (n=6)	Faculty (n=6)	ANOVA **
Animal Time Scale*	15.2 ± 1.8^{ab}	8.6 ± 2.2^{a}	10.6 ± 1.8^{b}	$p < .001$
Opening Chest	6.3 ± 2.3^{a}	2.6 ± 0.7^{a}	3.7 ± 2.0	$p < .007$
Pericardiotomy	3.5 ± 1.4	2.3 ± 0.7	2.9 ± 1.3	$p < .234$
Aortic Cross-Clamp	5.4 ± 2.5	3.7 ± 1.5	4.0 ± 0.9	$p < .071$

* Summated scale of the three thoracotomy procedure times.

** ANOVA, F-test, p-value

a,b Denotes significant pairwise comparisons by Scheffe method, $p < .05$.

Table 3. Thoracotomy Procedure Accuracy Scores (mean ± SD) for Written, Computer, and Animal Models by Level of Training

		Level of Training			
Accuracy Measure	Reliability (Cronbach's Alpha)	Student (n=6)	Resident (n=6)	Faculty (n=6)	ANOVA*
Written Accuracy (%)	0.51	95.5 ± 2.6^{a}	97.3 ± 1.7	98.5 ± 0.7^{a}	$p < .037$
Computer Accuracy (%)	0.66	57.2 ± 10.7	61.2 ± 12.2	70.5 ± 10.3	$p < .137$
Animal Accuracy (%)	0.76	70.9 ± 13.6^{a}	84.2 ± 6.4	88.8 ± 6.6^{a}	$p < .014$

* Analysis of Variance, F test, p values
[a,b] Significant pairwise comparisons by Scheffé method, $p < .05$.
+Post-hoc power analysis: n=15 per group to detect expected differences with a power of 0.80.

Table 4. Intercorrelation of Thoracotomy Assessment Scores[+]

	Written Accuracy	Computer Accuracy	Animal Accuracy	Animal Time
Written Accuracy	1.0			
Computer Accuracy	0.46 (.057)*	1.0		
Animal Accuracy	0.61 (.007)	0.30 (.218)	1.0	
Animal Time	-0.57 (.014)	-0.15 (.557)	-0.46 (.055)	1.0

\+ Pearson-Product Moment Correlations
* p-value

Table 5. Effect of Computer Practice on Thoracotomy Test Performance (Mean ± SD)

Thoracotomy Assessment	Type of Computer Practice		
	Thoracotomy (n=9)	Cricothyrotomy (n=9)	t-test, p-value
Written Accuracy (%)	97.9 ± 1.5	96.4 ± 2.5	$p < .141$
Computer Accuracy (%)	69.4 ± 12.1	56.6 ± 7.9	$p < .017$
Animal Accuracy (%)	81.8 ± 11.8	80.8 ± 12.6	$p < .863$
Animal Time (min.)	11.1 ± 3.8	11.8 ± 3.7	$p < .729$

Appendix A
CRITICAL STEPS IN OPENING THE CHEST

Step #	Description of Step	Common Errors of Omission/Commission
1	Optionally, prep thorax with 10% providone-iodine.	-step omitted
2	Optionally, drape field as quickly as possible.	-step omitted
3	Don with sterile surgical gloves, cap and mask.	-step omitted
4	Make left submammary anterior (anterolateral) incision from sternum to mid-axillary line (use #10 blade at 4-5th interspace to incise skin, SQ fat, chest wall musculature sufficient to expose intercostal space).	-incision too small -wrong blade (e.g. #11,#15) - wrong interspace -incision too shallow or too deep, or through female breast
5	Incise small segments of pectoralis major muscle, anterior pectoralis minor and serratus anterior muscle laterally.	
6	Divide intercostal muscles using heavy Mayo scissors or scalpel.	-continue bagging
7	Momentarily stop assisting ventilations while incising pleura	
8	Divide parietal pleura using heavy scissors cutting along superior rib margin to avoid intercostal neurovascular bundle, while ignoring minimal chest wall bleeding.	-lacerate visceral pleura -lacerate lung -cut along inferior rib margin -injure neurovascular bundle -ligate chest wall bleeding with electrocautery
9	Optionally, divide costal cartilage cephalad for additional exposure using heavy scissors.	
10	Insert standard rib retractor with handle directed inferiorly toward axilla.	-insert rib retractor with handle superiorly directed
11	Open chest cavity by twisting rib retractor.	-opening too small
12	Ligate internal mammary artery lacerations once perfusion is established.	-exsanguination when omitted
13	Once stabilized, get patient to OR.	-delay in transport to OR