

Development of a Test to Evaluate Residents' Knowledge of Medical Procedures

Shilpa Grover, MD^{1,2}
Paul F. Currier, MD^{2,3}
Jason M. Elinoff, MD^{2,3}
Kevin J. Mouchantaf, MD⁴
Joel T. Katz, MD^{1,2}
Graham T. McMahon, MD, MMSc^{1,2}

¹ Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts.

² Harvard Medical School, Boston, Massachusetts.

³ Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts.

⁴ Department of Medicine, North Shore Medical Center, Salem, Massachusetts.

Sponsored in part by an Innovation Grant in Instructional Computing from the Harvard University Provost.

The authors have no conflicts of interest to report.

BACKGROUND AND AIM: Knowledge of core medical procedures is required by the American Board of Internal Medicine (ABIM) for certification. Efforts to improve the training of residents in these procedures have been limited by the absence of a validated tool for the assessment of knowledge. In this study we aimed to develop a standardized test of procedural knowledge in 3 medical procedures associated with potentially serious complications.

METHODS: Placement of an arterial line, central venous catheter, and thoracentesis were selected for test development. Learning objectives and multiple-choice questions were constructed for each topic. Content evidence was evaluated by critical care subspecialists. Item test characteristics were evaluated by administering the test to students, residents and specialty clinicians. Reliability of the 32-item instrument was established through its administration to 192 medical residents in 4 hospitals.

RESULTS: Reliability of the instrument as measured by Cronbach's α was 0.79 and its test-retest reliability was 0.82. Median score was 53% on a test comprising elements deemed important by critical care subspecialists. Increasing number of procedures attempted, higher self-reported confidence, and increasing seniority were predictors of overall test scores. Procedural confidence correlated significantly with increasing seniority and experience. Residents performed few procedures.

CONCLUSIONS: We have successfully developed a standardized instrument to assess residents' cognitive competency for 3 common procedures. Residents' overall knowledge about procedures is poor. Experiential learning is the dominant source for knowledge improvement, but these experiences are increasingly rare. *Journal of Hospital Medicine* 2009;4:430–432. © 2009 Society of Hospital Medicine.

KEYWORDS: medical procedures, residency training, test development.

Medical procedures, an essential and highly valued part of medical education, are often undertaught and inconsistently evaluated. Hospitalists play an increasingly important role in developing the skills of resident-learners. Alumni rate procedure skills as some of the most important skills learned during residency training,^{1,2} but frequently identify training in procedural skills as having been insufficient.^{3,4} For certification in internal medicine, the American Board of Internal Medicine (ABIM) has identified a limited set of procedures in which it expects all candidates to be cognitively competent with regard to their knowledge of these procedures. Although active participation in procedures is recommended for certification in internal medicine, the demonstration of procedural proficiency is not required.⁵

Resident competence in performing procedures remains highly variable and procedural complications can be a source of morbidity and mortality.^{2,6,7} A validated tool for the assessment of procedure related knowledge is currently lacking. In existing standardized tests, including the in-training examination (ITE) and ABIM certification examination, only a fraction

of questions pertain to medical procedures. The necessity for a specifically designed, standardized instrument that can objectively measure procedure related knowledge has been highlighted by studies that have demonstrated that there is little correlation between the rate of procedure-related complications and ABIM/ITE scores.⁸ A validated tool to assess the knowledge of residents in selected medical procedures could serve to assess the readiness of residents to begin supervised practice and form part of a proficiency assessment.

In this study we aimed to develop a valid and reliable test of procedural knowledge in 3 procedures associated with potentially serious complications.

Methods

Placement of an arterial line, central venous catheter and thoracentesis were selected as the focus for test development. Using the National Board of Medical Examiners question development guidelines, multiple-choice questions were developed to test residents on specific points of a prepared curriculum. Questions were designed to test the

TABLE 1. Subject Characteristics

	Number (%)
Total residents	192
Males	113 (59)
Year of residency training	
First	101 (52)
Second	64 (33)
Third/fourth	27 (14)
Anticipated career path	
General medicine/primary care	26 (14)
Critical care	47 (24)
Medical subspecialties	54 (28)
Undecided/other	65 (34)

essential cognitive aspects of medical procedures, including indications, contraindications, and the management of complications, with an emphasis on the elements that were considered by a panel of experts to be frequently misunderstood. Questions were written by faculty trained in question writing (G.M.) and assessed for clarity by other members of faculty. Content evidence of the 36-item examination (12 questions per procedure) was established by a panel of 4 critical care specialists with expertise in medical education. The study was approved by the Institutional Review Board at all sites.

Item performance characteristics were evaluated by administering the test online to a series of 30 trainees and specialty clinicians. Postadministration interviews with the critical care experts were performed to determine whether test questions were clear and appropriate for residents. Following initial testing, 4 test items with the lowest discrimination according to a point-biserial correlation (Integrity; Castle Rock Research, Canada) were deleted from the test. The resulting 32-item test contained items of varying difficulty to allow for effective discrimination between examinees (Appendix 1).

The test was then administered to residents beginning rotations in either the medical intensive care unit or in the coronary care unit at 4 medical centers in Massachusetts (Brigham and Women's Hospital; Massachusetts General Hospital; Faulkner Hospital; and North Shore Medical Center). In addition to completing the on-line, self-administered examination, participants provided baseline data including year of residency training, anticipated career path, and the number of prior procedures performed. On a 5-point Likert scale participants estimated their self-perceived confidence at performing the procedure (with and without supervision) and supervising each of the procedures. Residents were invited to complete a second test before the end of their rotation (2-4 weeks after the initial test) in order to assess test-retest reliability. Answers were made available only after the conclusion of the study.

Reliability of the 32-item instrument was measured by Cronbach's α analysis; a value of 0.6 is considered adequate

TABLE 2. Number of Procedures Performed by Year of Internal Medicine Residency Training

Year of Residency Training	Median Number of Procedures (Interquartile Range)		
	Arterial Line Insertion	Central Venous Line Insertion	Thoracentesis
First	1 (0-3)	1 (0-4)	0 (0-1)
Second	8.5 (6-18)	10 (5-18)	2 (0-4)
Third/fourth	13 (8-20)	14 (10-27)	3 (2-6)

and values of 0.7 or higher indicate good reliability. Pearson's correlation (Pearson's r) was used to compute test-retest reliability. Univariate analyses were used to assess the association of the demographic variables with the test scores. Comparison of test scores between groups was made using a t test/Wilcoxon rank sum (2 groups) and analysis of variance (ANOVA)/Kruskal-Wallis (3 or more groups). The associations of number of prior procedures attempted and self-reported confidence with test scores was explored using Spearman's correlation. Inferences were made at the 0.05 level of significance, using 2-tailed tests. Statistical analyses were performed using SPSS 15.0 (SPSS, Inc., Chicago, IL).

Results

Of the 192 internal medicine residents who consented to participate in the study between February and June 2006, 188 completed the initial and repeat test. Subject characteristics are detailed in Table 1.

Reliability of the 32-item instrument measured by Cronbach's α was 0.79 and its test-retest reliability was 0.82. The items difficulty mean was 0.52 with a corrected point biserial correlation mean of 0.26. The test was of high difficulty, with a mean overall score of 50% (median 53%, interquartile range 44-59%). Baseline scores differed significantly by residency program ($P = 0.03$). Residents with anticipated careers in critical care had significantly higher scores than those with anticipated careers in primary care (median scores critical care 56%, primary care and other nonprocedural medical subspecialties 50%, $P = 0.01$).

Residents in their final year reported performing a median of 13 arterial lines, 14 central venous lines, and 3 thoracenteses over the course of their residency training (Table 2). Increase in the number of performed procedures (central lines, arterial lines, and thoracenteses) was associated with an increase in test score (Spearman's correlation coefficient 0.35, $P < 0.001$). Residents in the highest and lowest decile of procedures performed had median scores of 56% and 43%, respectively ($P < 0.001$). Increasing seniority in residency was associated with an increase in overall test scores (median score by program year 49%, 54%, 50%, and 64%, $P = 0.02$).

Increase in self-reported confidence was significantly associated with an increase in the number of performed procedures (Spearman's correlation coefficients for central

line 0.83, arterial lines 0.76, and thoracentesis 0.78, all $P < 0.001$) and increasing seniority (0.66, 0.59, and 0.52, respectively, all $P < 0.001$).

Discussion

The determination of procedural competence has long been a challenge for trainers and internal medicine programs; methods for measuring procedural skills have not been rigorously studied. Procedural competence requires a combination of theoretical knowledge and practical skill. However, given the declining number of procedures performed by internists,⁴ the new ABIM guidelines mandate cognitive competence in contrast to the demonstration of hands-on procedural proficiency.

We therefore sought to develop and validate the results of an examination of the theoretical knowledge necessary to perform 3 procedures associated with potentially serious complications. Following establishment of content evidence, item performance characteristics and postadministration interviews were used to develop a 32-item test. We confirmed the test's internal structure by assessment of reliability and assessed the association of test scores with other variables for which correlation would be expected.

We found that residents performed poorly on test content considered to be important by procedure specialists. These findings highlight the limitations in current procedure training that is frequently sporadic and often variable. The numbers of procedures reported over the duration of residency by residents at these centers were low. It is unclear if the low number of procedures performed was due to limitations in resident content knowledge or if it reflects the increasing use of interventional services with fewer opportunities for experiential learning. Nevertheless, an increasing number of prior procedures was associated with higher self-reported confidence for all procedures and translated to higher test scores.

This study was limited to 4 teaching hospitals and further studies may be needed to investigate the wider generalizability of the study instrument. However, participants were from 3 distinct internal medicine residency programs that included both community and university hospitals. We relied on resident self-reports and did not independently verify the number of prior procedures performed. However, similar assumptions have been made in prior studies that physicians who rarely perform procedures are able to provide reasonable estimates of the total number performed.³

The reliability of the 32-item test (Cronbach's $\alpha = 0.79$) is in the expected range for this length of test and indicates good reliability.^{9,10} Given the potential complications associated with advanced medical procedures, there is increasing

need to establish criteria for competence. Although we have not established a score threshold, the development of this validated tool to assess procedural knowledge is an important step toward establishing such a goal.

This test may facilitate efforts by hospitalists and others to evaluate the efficacy and refine existing methods of procedure training. Feedback to educators using this assessment tool may assist in the improvement of teaching strategies. In addition, the assessment of cognitive competence in procedure-related knowledge using a rigorous and reliable means of assessment such as outlined in this study may help identify residents who need further training. Recognition for the necessity for additional training and oversight are likely to be especially important if residents are expected to perform procedures safely yet have fewer opportunities for practice.

Acknowledgements

The authors thank Dr. Stephen Wright, Haley Hamlin, and Matt Johnston for their contributions to the data collection and analysis.

Address for correspondence and reprint requests:

Graham T. McMahon, MD, MMSc, Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, MA 02114; Telephone: 617-732-5666; Fax: 617-732-5764; E-mail: gcmcmahon@partners.org Received 14 July 2008; revision received 13 November 2008; accepted 21 January 2009.

References

1. Nelson RL, McCaffrey LA, Nobrega FT, et al. Altering residency curriculum in response to a changing practice environment: use of the Mayo internal medicine residency alumni survey. *Mayo Clin Proc.* 1990;65(6): 809–817.
2. Mandel JH, Rich EC, Luxenberg MG, Spilane MT, Kern DC, Parrino TA. Preparation for practice in internal medicine. A study of ten years of residency graduates. *Arch Intern Med.* 1988;148(4):853–856.
3. Hicks CM, Gonzalez R, Morton MT, Gibbon RV, Wigton RS, Anderson RJ. Procedural experience and comfort level in internal medicine trainees. *J Gen Intern Med.* 2000;15(10):716–722.
4. Wigton RS. Training internists in procedural skills. *Ann Intern Med.* 1992; 116(12 Pt 2):1091–1093.
5. ABIM. Policies and Procedures for Certification in Internal Medicine 2008. Available at: <http://www.abim.org/certification/policies/imss/im.aspx>. Accessed August 2009.
6. Wickstrom GC, Kolar MM, Keyserling TC, et al. Confidence of graduating internal medicine residents to perform ambulatory procedures. *J Gen Intern Med.* 2000;15(6):361–365.
7. Kern DC, Parrino TA, Korst DR. The lasting value of clinical skills. *JAMA.* 1985;254(1):70–76.
8. Durning SJ, Cation LJ, Jackson JL. Are commonly used resident measurements associated with procedural skills in internal medicine residency training? *J Gen Intern Med.* 2007;22(3):357–361.
9. Nunnally JC. *Psychometric Theory*. New York: McGraw Hill; 1978.
10. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.