

An Assessment Tool for Aseptic Technique in Resident Physicians: A Journey Towards Validation in the Real World of Limited Supervision

MONICA L. LYPSON, MD
STANLEY J. HAMSTRA, PhD
PAULA T. ROSS, MA
LARRY D. GRUPPEN, PhD
LISA M. COLLETTI, MD

Abstract

Purpose The purpose of this study was to describe the validation process for assessing an instrument to assess residents' aseptic technique skills.

Methods The validation study entailed comparisons of the performance of aseptic technique procedures between postgraduate year-1 (PGY-1) surgical residents and PGY-2/3 surgical residents. We also compared the performance of PGY-1 surgical residents from 2 different academic years for the same procedures. Finally, we compared the performance of novices (medical students) and experts (operating room nurses) in an effort to determine validity.

Results Our initial analysis found no significant difference between the performance of PGY-1 (mean score, 75.8) and PGY-2/3 (mean score, 75.6) surgical residents for aseptic technique ($t_{(55)} = 0.84$, $P = 0.404$). Further investigation of validity was obtained to determine whether the no difference results reflected a lack of reliability or validity or a true equivalence between the 2 cohorts. The comparison of novices and experts produced the following findings. For reliability, the internal consistency of the checklist for each of the 2

raters was 0.87 and 0.71 (Cronbach α), interrater reliability was 0.74, with $P < 0.001$ (intraclass correlation coefficient) for the global scale. (Internal consistency was done within instrument, ie, between items not between raters.) For validity, operating room nurses outperformed students on the global scale ($t_{14} = 7.47$, $P < 0.0001$ and $t_{(14)} = 10.66$, $P < 0.0001$ for the 2 raters, respectively) and on several checklist items. The effect size values for raters were large (Cohen $d = 3.0$ and 4.4), providing validity evidence for the ability of this assessment to detect difference in performance on this task.

Conclusion The validation study showed that the instrument exhibited reliability and evidence for validity, making it useful for the assessment of aseptic technique skills in different specialties. Programs may want to consider using a validated instrument to check competence given that appropriate use of sterile technique frequently occurs in the context of unsupervised activities. Further work is needed to enhance resident skills in the area of aseptic technique because of limited improvement despite additional clinical experience.

Introduction

Each year new residents begin their clinical education and medical career by performing procedures under limited

supervision and with varied expertise. Unfortunately, many residency programs have traditionally assumed residents arrive with or become proficient in these technical skills without an explicit assessment of specific competencies. Baseline skill assessment thus is critical, as it allows program directors to adequately address resident skill deficits and as it creates appropriate remediation programs for bedside procedures that are performed with minimal or no faculty supervision.

The Accreditation Council for Graduate Medical Education (ACGME) has mandated the assessment of resident competency in a variety of domains, including patient care.¹⁻³ For most residents, this domain includes basic technical skills, such as aseptic technique. Published reports on the assessment of aseptic technique are scarce in surgical and medical literature, although references to proper technique are commonly found in the nursing literature. To date, no large-scale assessment of these skills, in residents in general and surgical residents in particular,

Monica L. Lypson, MD, is Assistant Dean of Graduate Medical Education and Associate Professor of Internal Medicine at University of Michigan Medical School and staff physician at Ann Arbor VA Healthcare System; **Stanley J. Hamstra, PhD**, is Adjunct Associate Professor of Medical Education at University of Michigan Medical School, Research Director at University of Ottawa Skills and Simulation Centre, and Acting Assistant Dean of Academy for Innovation in Medical Education in Faculty of Medicine at University of Ottawa, Ottawa, Ontario, Canada; **Paula T. Ross, MA**, is Clinical Research Coordinator in Pediatrics at University of Michigan Medical School; **Larry D. Gruppen, PhD**, is Josiah Macy Jr Professor of Medical Education and Chair of Medical Education at University of Michigan Medical School; and **Lisa M. Colletti, MD**, is Associate Dean of Graduate Medical Education and C. Gardner Child Professor of Surgery at University of Michigan Medical School.

Corresponding author: Monica L. Lypson, MD, University of Michigan, Office of Graduate Medical Education, 2600 Green Rd, No. 150, Ann Arbor, MI 48105, 734.764.3186, mlypson@umich.edu

Received October 14, 2009; revision received January 5, 2010, accepted January 13, 2010.

DOI: 10.4300/JGME-D-09-00071.1

has been developed.⁴⁻⁹ As aseptic technique is a skill we expect all clinicians to master, its inclusion in training programs and role modeling by attending staff is strongly recommended.¹⁰ Nevertheless, adequate bedside technique continues to be one of the many undetected educational gaps in the transition from undergraduate to graduate medical education.¹¹

At the University of Michigan, we developed the Postgraduate Orientation Assessment (POA) in response to the ACGME mandate, with particular attention given to the development of assessment instruments to evaluate entry-level skills often performed in situations with limited supervision.¹¹ Using the best evidence available in the area of aseptic/bedside procedure technique, we specifically designed 1 of the 9 testing stations in our POA program to assess residents' ability to use aseptic technique and to provide formative feedback. This scenario involves a standardized patient needing an incision and drainage of a large, red, irritated forearm abscess oozing small amounts of purulent fluid. All the materials typically available on an inpatient ward stockroom are accessible in the simulated patient room; in addition, the nurse observer can assist as needed, if directed to do so. The trainees are required to perform all aspects of cleansing and "prepping" without actually performing the procedure (ie, opening the wound) and then to dress the wound as if they had removed the infectious materials. Throughout the procedure, faculty and nursing experts reviewed the realism with which the abscess was portrayed on the standardized patient to ensure it approximated real-world experiences.

Before using the station with trainees, the expert raters spent 2 to 3 hours reviewing the checklist and training protocol and material before their participation. Rater training focused on several key areas: (1) review of the checklist, (2) overview of the expectations of the case given by the station-lead, and (3) if requested, review of prior year's videos of residents' performance. The raters subsequently met several times to discuss discrepancies and ensure they had agreement on the scoring criteria.

Information on baseline ability to perform aseptic technique, as well as formative, immediate feedback was provided verbally to each trainee. Written remediation materials documenting proper technique, as outlined by the checklist, and cooperating material similar to those from the Joint Commission also were provided.

While many traditional methods exist for assessing competency across the range of medical specialties in postgraduate medical education—in addition to many newly developed instruments reported in the recent literature—very few studies make use of standard psychometric practices for evaluating the quality of assessment instruments, including the analysis of reliability and validity evidence, especially in the area of technical skills.^{4-9,12,13} In addition, some program

directors may not be aware of how to apply a modern conceptualization of validity or conduct validity studies. The process of validation should (1) include multiple sources of evidence to support or refute meaningful judgments, (2) be based on a hypothesis for which data is collected to accept or reject it, and (3) only refer to the process of interpretation and not to the assessment.¹⁴ The aims of this study were to establish evidence of validity from the internal structure of the assessment and from predicted patterns of results related to differences in expertise.¹⁴

The 2 studies described in this article entailed the development of a tool to assess residents' aseptic technique skills. To determine the tool's validity, we first compared the performance of residents at various levels. This was followed by further assessment of the tool by comparing the performance of individuals with extreme differences in expertise, namely, medical students and nurses. All studies received exemption status from the University of Michigan Institutional Review Board.

Methods

Data Collection

The study began with a formal assessment of validity evidence for the checklist and global rating scale, with the aim of further developing and refining the instrument. We sought to determine the instrument's ability to detect expected differences in aseptic technique performance on the basis of variation in examinee expertise.

Preliminary Study

In June 2004 and June 2005, we assessed all incoming residents' (postgraduate year-1 [PGY-1]) baseline ability to perform aseptic technique by using a previously developed checklist and global rating scale (TABLE). Checklist items 1 through 4, 12, 14, and 20 were used in the comparisons for both studies; the nontechnical items indicated in italics in the TABLE were omitted from the data analyses. The total possible score was 100, with all items having been given equal weight. We assessed 291 interns from 13 different specialties during the 2004 and 2005 residency orientation sessions. The mean score in 2004 was 71.6 ± 13.5 , with an average global score of 5.1 ± 1.1 ; the mean score in 2005 was 74.4 ± 12.6 , with an average global score of 3.9 ± 1.5 .

The assessment was also administered to 24 second- and third-year surgery residents (mean score, 75.6) in an effort to determine whether more advanced trainees would demonstrate competency for this task. These residents took the POA at the beginning of their internship and then again as either a second- or third-year surgery resident. The same nurse experts were used to assess the residents during the different years the assessment was administered.

TABLE	ASEPTIC TECHNIQUE CHECKLIST AND GLOBAL RATING SCALE ^{a-c}		
	Done	Needs Improvement	Not Done
1. <i>Identify patient</i>			
2. <i>State that informed consent has been obtained</i>			
3. <i>Side/site verification</i>			
4. <i>Patient/family informed as procedure advances</i>			
5. Place pad under patient in appropriate location			
6. Assemble supplies			
7. Put on protective eyewear/mask			
8. Wash hands			
9. Preparation area: Povidone-Iodine Prep Swabsticks			
10. Open supplies			
11. Open and don sterile gloves			
12. <i>Time out</i>			
13. Apply sterile drapes			
14. <i>State that procedure was performed</i>			
15. Apply sterile dressings			
16. Dispose of waste materials			
17. Remove gloves			
18. Wash hands			
19. Secure dressings			
20. <i>Explain outcome of procedure to patient</i>			
21. Your assessment of the resident's overall performance on the above items (1 = novice, 9 = master):			
Circle your score: 1 2 3 4 5 6 7 8 9			

^a Checklist and global rating scale (item 21) used in the preliminary study and the validation study.

^b Items in italics represent nontechnical skills and were not included in the final score for performance.

^c Items 1–4, 12, 14, and 20 were used in the assessment for both the preliminary study and the current study, but these nontechnical items were not included in the data analyses for either study.

Preliminary Results

No significant difference was found between the performance of surgical PGY-1 residents (mean, 75.8 ± 10.6) and surgical PGY-2/3 residents (mean, 75.6 ± 10.9) at the aseptic technique station ($t_{(55)} = 0.84$, $P = 0.404$) during the June 2005 assessment. Although the data strongly suggest no difference between the 2 groups, it is important to note that the assessment instrument had yet to be formally evaluated for the psychometric standard of construct-internal structure validity. Therefore, a plausible interpretation could be that the *internal structure* of the assessment was flawed and the assessment instrument was insensitive to any existing differences in ability.¹⁴ Given the iterative process of item development, we believed that our tool was generalizable; however, additional information

was necessary to determine whether the negative finding between surgical PGY-1 residents and PGY-2/3 residents at the aseptic technique station was due to insufficient training or insensitivity of the instrument.

Second Study

Acknowledging that differences between residents of different training years may be small and difficult to detect, we used the same station to compare the performance of medical students (novices) and nurses (experts), as 2 groups with a greater contrast in expertise and experience. To further refine the instrument according to accepted psychometric standards, we conducted a formal assessment of construct validity of the checklist and global rating scale.

For this second study, e-mails were sent to operating room nursing supervisors (approximately 225) and all first-year

(M1) through third-year (M3) medical students (approximately 480). As only 1 participant could be assessed at a time, participants were recruited until the group comparison data showed a statistically significant difference in group means on performance. The final sample included 10 operating room nurses and 6 medical students (M1–M3) and participants received a \$10 gift card each for their participation.

Data Analysis

Individual performance was assessed by 2 independent raters (M. L. and S. H.), by using the checklist (items 1–20) and global rating scale (item 21) in a real-time setting with a trained standardized patient. Both raters received the aforementioned training. The psychometric analyses included a statistical assessment of reliability (interrater reliability, Cronbach α) and validity evidence (group differences between nurses and M1–M3 students) on the global rating scale, as well as individual items from the checklist. The aims of this study were to establish evidence of validity from the internal structure of the assessment and from predicted patterns of results related to differences in expertise.¹⁴

Results

Reliability

The internal consistency of the checklist was high to moderate for the 2 raters: Cronbach α was 0.87 and 0.71 for raters S. H. and M. L., respectively. Interrater reliability for the checklist items ranged from poor ($\kappa < 0.40$ for items 8, 9, 13, 15, and 17) to good ($\kappa = 0.40$ – 0.75 for items 5–7, 10, 11, 18, and 19) and high ($\kappa = 0.78$ for item 16). For the global rating scale, interrater reliability was good (intraclass correlation coefficient [single measure], 0.742; $P < 0.001$).

Validity Evidence

Both raters judged that the nurses performed much better than the M1 through M3 students on the global rating scale ($t_{(14)} = 7.47$; $P < .0001$ for rater S. H. and $t_{(14)} = 10.66$; $P < .0001$ for rater M. L.), providing the predicted validity evidence that the assessment could distinguish between examinees with large differences in expertise. The effect size values (Cohen d , 19) for both raters were extremely large ($d = 3.0$ for S. H.; $d = 4.4$ for M. L.). Significant group differences (independent group t tests) were found for 8 of the 13 checklist items for rater S. H. (items 5, 6, 9, 10, 11, 16, 17, and 19) and for 4 of the 13 items for rater M. L. (items 6, 10, 16, and 19).

Discussion and Conclusion

Developing meaningful assessment tools that reflect actual clinical practice across specialties is challenging. The need to use aseptic technique is common in today's medical environment; however, no useful or trustworthy assessment tool has been developed that provides valid guidance for teaching and providing useful feedback to learners.

Our results provide evidence for the reliability of this assessment instrument and the validity of the formative feedback based on these scores. We observed predicted differences in aseptic technique performance between experienced nurses and M1 through M3 students who had received little formal training in this skill. Additionally, as described above, this instrument detects differences between novice and expert skill levels. Considered independently, the statistically significant group differences for the global rating scale provide strong validity evidence for the assessment instrument scores. Even with such small groups, both raters detected cues to the level of competence, which reflected the different levels of training. The findings for the checklist are not as clear. While one of the raters (S. H.) found group differences for most of the 13 items, the other rater (M. L.) found group differences for only 4 of the 13 items.

Our finding that group differences were found for global ratings of performance but not for the individual aspects of technical skill that are captured by the checklist perhaps reflects the importance that expert observers place on the overall approach to the procedure, rather than the constituent steps. One interpretation is that novices tend to remember and focus on details, while experts exhibit a more natural flow to their performance, which is consistent with greater experience and practice.¹⁵ These findings are consistent with other studies¹⁶ showing that experts' global ratings of performance exhibit greater construct validity.

Given the validity evidence above, we can now refer to our preliminary study showing no difference in performance between the PGY-1 and PGY-2/3 surgical residents. Cast in this new light, these findings imply that there was no discernable difference in performance between residents at these 2 levels of training. What is perhaps a surprising implication is that the PGY-2/3 residents did not exhibit mastery of this basic skill (as did the operating room nurses), which suggests a need for focused training in this area at the postgraduate level. We often assume that, as residents advance, they become more experienced; however this is an ongoing assumption that has yet to be challenged in a fully competency-based educational paradigm.

Limitations

Although large differences in performance were detected between nurses and medical students, it could be argued that nurses are not the ideal comparison for medical practitioners. Our study used operating room nurses as the "gold standard" for aseptic technique. A potential fallacy of this approach is the assumption that residents or students who do not learn aseptic technique to ensure sterile procedure in the operating room have not received comprehensive training in this skill. Also, our ratings of performance were not blinded. Although we attempted to keep the group identity of the participants from the raters, it was generally apparent from dress, maturity level, and

department which participants were nurses and which were M1 through M3 students.

Educational Significance

Given that incoming residents often perform numerous tasks without supervision, these findings suggest a need for focused training in this area at the postgraduate level. The current study offers a model for developing a validated assessment instrument to assist programs in addressing the mandate of the ACGME Outcome Project. Additionally, this instrument may serve as a valuable tool, in that each day, hospital staff face recurrent issues stemming from nosocomial infections due to improper aseptic technique. In future studies it may be helpful to obtain the perspective of the surgery PGY-2/3 residents from the preliminary study to learn why their performance did not significantly differ from that of the PGY-1 residents. This finding offers further information to improve the content validity of the assessment instrument. Our instrument affords programs an opportunity to assess a skill that often occurs with limited supervision.

References

- 1 Accreditation Council for Graduate Medical Education (ACGME). Outcome project. Available at: http://www.acgme.org/outcome/project/OPintrorev1_7-05.ppt#25. Accessed February 15, 2009.
- 2 Accreditation Council for Graduate Medical Education (ACGME). Outcome project timeline. Available at: http://www.acgme.org/outcome/project/timeline/TIMELINE_index_frame.htm. Accessed February 15, 2009.
- 3 Accreditation Council for Graduate Medical Education (ACGME). Outcome project, competencies. Available at: <http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>. Accessed February 2, 2010.
- 4 Tetzlaff JE. Assessment of competency in anesthesiology. *Anesthesiology*. 2007;106(6):812–825.
- 5 Swick S, Hall S, Beresin E. Assessing the ACGME competencies in psychiatry training programs. *Acad Psychiatry*. 2006;30(4):330–351.
- 6 Emken JL, McDougall EM, Clayman RV. Training and assessment of laparoscopic skills. *JSL*. 2004;8(2):195–199.
- 7 Rumack CM. Diagnostic radiology residency review committee program citations and how to avoid them. *J Am Coll Radiol*. 2005;30(2):617–621.
- 8 Taylor JB, Binenbaum G, Tapino P, Volpe NJ. Microsurgical lab testing is a reliable method for assessing ophthalmology residents' surgical skills. *Br J Ophthalmol*. 2007;91(12):1691–1694.
- 9 Park JP, MacRae H, Musselman LJ, et al. Randomized control trial of virtual reality simulator training: transfer to live patients. *Am J Surg*. 2007;194(2):205–211.
- 10 Beyea SC. Keeping patients safe from infection. *AORN J*. 2003;78(1):133–134, 137.
- 11 Lypson ML, Frohna JG, Gruppen LD, Wooliscroft JO. Assessing residents' competencies at baseline: identifying the gaps. *Acad Med*. 2004;79(10):564–570.
- 12 Savoldelli GS, Naik VN, Joo HS, et al. Evaluation of patient simulator performance as an adjunct to oral examination for senior anesthesia residents. *Anesthesiology*. 2006;104(3):475–481.
- 13 Hamstra SJ. Predicting the technical competence of surgical residents. *Clin Orthop Relat Res*. 2006;449:62–66.
- 14 Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(1):1–8.
- 15 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10):S70–S81.
- 16 Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med*. 1998;73(9):993–997.