# Simulation-Based Medical Education: Development of an Assessment Tool for Novice Use

Sasha M. Ulrich, MD; Joseph C. L'Huillier, MD; Sarah A. Jung, PhD; Laura K. Krecko, MD; Alexandra A. Rosser, BS; Amy K. Schulze, MD; Amy E. Liepert, MD; Ann P. O'Rourke, MD, MPH

## ABSTRACT

**Background:** Simulation-based medical education, an educational model in which students engage in simulated patient scenarios, improves performance. However, assessment tools including the Oxford Non-Technical Skills (NOTECHS) scale require expert assessors. We modified this tool for novice use.

**Methods:** Medical students participated in 5 nontechnical simulations. The NOTECHS scale was modified to allow for novice evaluation. Three novices and 2 experts assessed performance, with intraclass correlation used to assess validity.

**Results:** Twenty-two learners participated in the simulations. Novice reviewers had moderate to excellent correlation among evaluations (0.66 < intraclass correlation coefficients [ICC] < 0.95). Novice and expert reviewers had moderate to good correlation among evaluations (0.51 < ICC < 0.88).

**Discussion:** The modified NOTECHS scales can be utilized by novices to evaluate simulation performance. Novice assessment correlates with expert review. These tools may encourage the use of simulation-based medical education.

## BACKGROUND

The use of simulation in medical education has increased significantly over the past decade.[1] Simulation-based medical education (SBME) improves students' clinical skills and patient outcomes compared to traditional education methods.[2-5] Student skill improvement is attributable, in part, to performance assessment and feedback from low-stakes, high-fidelity simulations. One commonly used validated assessment tool of SBME is the Oxford Non-Technical Skills (NOTECHS) scale, which was designed to assess nontechnical skills, such as cognitive reasoning and communication, of an individual within a group setting (eg, assessment of a surgeon in the operating room).[6]

Despite its utility in assessing individual nontechnical skills, the NOTECHS scale is limited by the need for assessors with prior experience in the specific nontechnical skill areas being tested.[6] Thus, NOTECHS scale assessments typically require faculty participation. This presents challenges for implementation, as it requires coordination with busy academic and clinical schedules and, thus, potentially limits use of SBME. For this reason, using nonclinical personnel as assessors (ie, crowdsourcing clinical novices) could improve the efficiency and frequency of SBME. Crowdsourcing novices for assessment of student technical performance in simulation can be time-efficient and achieve high interrater reliability between novices and clinicians.[7-10]

In order to address this limitation of existing simulation assessment tools, our team of surgeons and clinical novices modified the NOTECHS scale as part of an exploratory study to assess the ability to eliminate the need for experienced assessors and allow for successful performance evaluation by novices.

## METHODS

This project was determined to be exempt from review by the institutional review board. Fourth-year medical students at the

**Author Affiliations:** Mayo Clinic Health System, Department of Family Medicine, Eau Claire, Wisconsin (Ulrich); University of Wisconsin School of Medicine and Public Health, Department of Surgery, Madison, Wis (Ulrich, L'Huillier, Jung, Krecko, Rosser, O'Rourke); University at Buffalo, Jacobs School of Medicine and Biomedical Sciences, Department of Surgery, Buffalo, New York (L'Huillier, Schulze); University of California San Diego Health, Department of Surgery, San Diego, California (Liepert).

**Corresponding Author:** Ann P. O'Rourke, MD, MPH, University of Wisconsin School of Medicine and Public Health, Department of Surgery, Clinical Science Center, 600 Highland Ave, Madison, WI 53792; phone 608.262.6246; email orourke@surgery.wisc.edu.

**Figure 1.** a-NOTECHS Assessment Tool

Faculty Name: _____
Student Code: _____

**UW Madison Surgery IPC Intern Prep Course SIMCenter Evaluation Tool – 2017**

Please circle the case being performed: Oliguria A Fib Chest Pain Change in Mental Status

| Topic | 5 = High Performing | 4 = Above Average | 3 = Satisfactory | 2 = Below Average | 1 = Poor Performing |
|---|---|---|---|---|---|
| **Overall Approach** Interaction with patient and nursing staff to successfully evaluate the patient and obtain pertinent clinical information in a timely and organized fashion | Organized approach results in efficient obtainment of clinical data, medical decision making, clinical treatment and notification of senior resident | Generally organized, however misses an important aspect of care | Becomes sidetracked in evaluation of the patient, resulting in loss of care efficiency. | Provides appropriate medical care but disorganized approach hinders the timing or delivery of that care | Scattered or disorganized approach interferes with timely care and management of the patient |
| | **5 = High Performing** | **4 = Above Average** | **3 = Satisfactory** | **2 = Below Average** | **1 = Poor Performing** |
| **Data Collection** Orders diagnostic tests appropriate to and focused on the clinical setting | Obtained all critical data points | Obtained most of the critical data | Obtained some critical data | Required prompting to obtain critical data | Failure to obtain critical data |
| | **5 = High Performing** | **4 = Above Average** | **3 = Satisfactory** | **2 = Below Average** | **1 = Poor Performing** |
| **Differential Diagnosis** Creates a differential diagnosis appropriate to the clinical setting | Comprehensive Included Critical Diagnosis | Appropriately Broad | 3-4 Diagnoses | One or Two Diagnoses | No Differential Diagnosis |
| | **5 = High Performing** | **4 = Above Average** | **3 = Satisfactory** | **2 = Below Average** | **1 = Poor Performing** |
| **Medical Decision Making** Appropriate to the clinical setting | Initiated all appropriate therapies Appropriate disposition of patient | Initiated some appropriate therapies | Delayed decision making resulted in delay of care | Some ideas but no clear decisions made | Delay in treatment of urgent condition Stumped |
| | **5 = High Performing** | **4 = Above Average** | **3 = Satisfactory** | **2 = Below Average** | **1 = Poor Performing** |
| **Communication and Interaction w/ RN** | Requested appropriate work up at initial call Professional behavior. Good Communication. | Provided some Communication | Incomplete communication Didn't request additional testing/info | Delayed or repetitive communication | Excessively delayed communication Required prompting Disregarded RN or Rude |
| | **5 = High Performing** | **4 = Above Average** | **3 = Satisfactory** | **2 = Below Average** | **1 = Poor Performing** |
| **Communication with Senior Resident** | Called Senior in timely, appropriate fashion Concise report, with all needed information | Called within appropriate time period. Missing minimal information | Call delayed or missing some info | Required prompting to call senior | Excessively delayed Communication Required excessive prompting |

**Comments:-**

_____
_____
_____

University of Wisconsin (UW) who matched into surgical residency programs completed an Internship Preparation Course (IPC). Additional components of the course are reported elsewhere.[11,12] The curriculum incorporated SBME by conducting SimMan 3G simulations at the UW Health Clinical Simulation Program. Simulation scenario scripts were written by course directors about 5 common on-call clinical presentations: chest pain, atrial fibrillation, oliguria, altered mental status, and trauma. Each scenario was designed to test clinical decision-making rather than technical skills. To simulate the clinical environment, registered nurses and senior resident confederates were simulated by faculty and resident surgeons.[13] The scenarios were designed to last approximately 7 to 8 minutes. The simulations were audio/video recorded for review. The Oxford NOTECHS scale was adapted to allow for assessment of a single student's performance (a-NOTECHS; Figure 1). This initial adaptation was separate from modifying the tool for novice use. Faculty and resident surgeons assessed learner performance using the a-NOTECHS form to provide feedback to students. These assessments were not included in our analysis.

We modified the a-NOTECHS to allow for assessments to be completed without the need for prior training or a clinical background with the input of a first-year medical student and 2 faculty surgeons. The overarching a-NOTECHS domains remained unchanged and included overall approach, data collection, differential diagnoses, medical decision-making, communication and interaction with the nurse, and communication with the senior resident. All simulation assessment tools and scenario scripts are included as appendices. Our modified-NOTECHS (m-NOTECHS) for the chest pain scenario is included as an example (Figure 2). The process of creating scenario-specific m-NOTECHS scales is described by domain.

### Data Collection, Differential Diagnoses, Medical Decision-Making

In the original a-NOTECHS, the 3 domains of data collection, differential diagnoses, and medical decision-making require evalu-

**Figure 2.** m-NOTECHS Assessment Tool, Chest Pain

| **Data Collection**<br>*Orders diagnostic tests appropriate to and focused on the clinical setting* | **Differential Diagnosis**<br>*Creates a differential diagnosis appropriate to the clinical setting* | **Medical Decision-Making**<br>*Appropriate to the clinical setting* | **Communication and Interaction With Registered Nurse (RN)** | **Communications With Senior Resident (SR)** | **Overall Approach**<br>*Interaction with pt and nursing staff to successfully evaluate the pt and obtain pertinent clinical info in a timely and organized fashion* |
|---|---|---|---|---|---|
| Critical Data Points:<br>☐ Vitals (≥ 3)<br>___Heart rate<br>___Blood pressure<br>___Respiratory rate<br>___Temperature<br>___O2 saturation<br>☐ Current medications<br>☐ Confirm pt is on telemetry<br>☐ Confirm pt has an IV<br><br>Labs:<br>☐ Troponin*<br>☐ BMP/electrolytes (≥ 2)<br>___Potassium, K<br>___Magnesium, Mg<br>___Phosphorus, P<br>☐ CBC<br>☐ ABG<br><br>Imaging:<br>☐ ECG*<br>☐ Chest x-ray | Critical Diagnoses:<br>☐ Angina<br>☐ Demand ischemia<br>☐ NSTEMI<br>☐ STEMI<br>☐ Anxiety<br>☐ PE<br>☐ Pleuritic pain/PNA | ☐ Apply Oxygen<br><br>Administer medication:<br>☐ Nitroglycerin<br>☐ Aspirin<br>☐ Reglan<br>☐ 2nd nitroglycerin<br>☐ Beta-blocker<br>☐ Morphine<br><br>☐ Administer IV fluids<br>☐ Transfer to ICU or call code | ☐ Contact occurs w/in first 2 mins and 30 sec of sim*<br><br>☐ Request nurse's sign out/notes on pt<br>☐ Request pt PMH<br>☐ Obtain current/ home medications<br>☐ Order/discuss workup of pt**<br>☐ Communicates concerns/differential diagnosis<br>☐ Communicates medication order(s)<br><br>☐ Professional communication (Please and thank you) | ☐ Contact occurs w/ each evaluation interaction*<br>☐ Pt name<br>☐ Age<br>☐ Gender<br>☐ Hospital/post-op day<br>☐ Reason in the hospital or operation<br>☐ Reason for call<br>☐ Vitals (≥ 2)<br>___Heart rate<br>___Blood pressure<br>___Respiratory rate<br>___Temperature<br>___O2 saturation<br>☐ Lab and/or imaging results (≥2)<br>___Electrolytes<br>___Troponin<br>___ECG<br>___Chest x-ray<br>☐ Treatment thus far + any results (or thoughts/questions) | Scattered or disorganized approach that interferes w/ timely care and management of pt = 1<br><br>Provides appropriate medical care but disorganized approach hinders timing/ delivery of that care; required SR guidance in data and treatment = 2<br><br>Becomes sidetracked in evaluation of pt, loss of efficiency; required SR guidance in data or treatment = 3<br><br>Organized but missed important aspect of care; 1 box; required minimal guidance = 4<br><br>Efficiently completed scenario; all boxes = 5 |
| No tests ordered = 1<br>1-3 checkbox; missed * = 2<br>4-5 checkboxes, incl. * = 3<br>6-7 checkboxes = 4<br>8-10 checkboxes = 5 | No diagnoses stated = 1<br>1-2 stated dx = 2<br>3-4 stated dx = 3<br>5+ stated dx, stumbled, scattered = 4<br>5+ stated dx, concise = 5 | Stumped = 1<br>Verbalized idea(s) but nothing done = 2<br>1-3 checkbox = 3<br>4-6 checkboxes = 4<br>7-9 checkboxes = 5 | Did not contact or had to be told by facilitator to contact RN; was rude, ignored or disregarded RN = 1<br>Delayed, missed *; repetitive = 2<br>3-4 checkboxes, incl. * but missed ** = 3<br>5-6 checkboxes, incl. * & ** = 4<br>7-8 checkboxes = 5 | Did not contact; was rude or disregarded SR = 1<br>Delayed, missed *; had to be told by facilitator to contact SR = 2<br>1-4 checkboxes, incl. * =3<br>5-7 checkboxes, incl. * = 4<br>8-10 checkboxes, incl. * = 5 | |

Abbreviations: pt, patient; $O_2$, oxygen; BMP, basic metabolic panel; CBC, complete blood cell count; ABG, arterial blood gas; ECG, electrocardiogram; NSTEMI, non–ST-segment elevation myocardial infarction; STEMI, ST-segment elevation myocardial infarction; PE, pulmonary embolism; PNA, pneumonia; dx, diagnoses; IV, intravenous; ICU, intensive care unit; PMH, past medical history; post-op, postoperative.

ators to globally judge whether a criterion is met without providing specific benchmarks (eg. "obtained all critical data points"). This language is not suitable for novice reviewers as they do not have the clinical experience to make these expert-level judgements (Figure 1). The process of quantifying these 3 domains was similar. Scenario-specific tests, diagnoses, and treatments were interpreted from the learning objectives of scenario scripts and made into checkboxes. The course directors added other appropriate possibilities and labeled specific data and treatments that must be completed in order to achieve a "satisfactory" m-NOTECHS score (Figure 2). In order to quantitate the "some" and "most" terms used in the a-NOTECHS scoring throughout these 3 domains, each score was quantified by requiring a set number of checkboxes. Within medical decision-making, a delay in treatment was defined as the elapse of 75% of the allotted simulation time before a treatment was initiated.

### Communication and Interaction With the Registered Nurse

"Timely" (Figure 1) was defined as communication occurring prior to elapse of 25% of the allotted simulation time; communication occurring after that was considered delayed. Obtaining other relevant patient information from the nurse, such as past medical history, the nurse's notes, and current medications, were added as checkboxes. Checkboxes also were added for professionalism, communicating medications or other orders, and discussing concerns with the nurse to further quantify the communication (Figure 2).

### Communication With the Senior Resident

In each scenario, a specific action or change in patient status was marked as an essential time for communication with the senior resident (Figure 1). Thus, contact at or before each point defined timely communication with the resident, and communication after that point was considered delayed. Since the scenarios were urgent situations, communication occurring after 75% of the allotted simulation time was defined as excessively delayed, which allowed for consistency in scoring across simulations. The information that needed to be communicated included orientation of the senior resident to the patient, reason for hospitalization, why the student called, any tests or treatments that had been done, and the results. Checkboxes were formed to address each of those points and quantitate communication for scoring (Figure 2).

**Table 1.** Descriptive Statistics for Scores Provided by Novice and Expert Reviewers for Each Simulation by Domain

| | Novice (n=3) | | | Expert (n=2) | | |
|---|---|---|---|---|---|---|
| | Median | Lower Quartile | Upper Quartile | Median | Lower Quartile | Upper Quartile |
| **Domain – Data Collection** | | | | | | |
| Simulation 1: Atrial Fibrillation | 4 | 3 | 4 | 4 | 4 | 4 |
| Simulation 1: Altered Mental Status | 4 | 3 | 4 | 4 | 4 | 5 |
| Simulation 2: Chest Pain | 3.5 | 2 | 4 | 3 | 3 | 4 |
| Simulation 2: Oliguria | 2 | 2 | 2 | 3 | 2 | 3 |
| Simulation 3: Trauma | 2 | 2 | 3 | 3 | 2 | 3 |
| **Domain – Differential Diagnoses** | | | | | | |
| Simulation 1: Atrial Fibrillation | 2 | 2 | 3 | 2 | 2 | 3 |
| Simulation 1: Altered Mental Status | 2 | 1 | 2 | 1.5 | 1 | 2 |
| Simulation 2: Chest Pain | 2 | 2 | 2 | 2 | 2 | 2 |
| Simulation 2: Oliguria | 2 | 2 | 2 | 2 | 2 | 2.5 |
| Simulation 3; Trauma | 2 | 2 | 3 | 2 | 2 | 2 |
| **Domain – Medical Decision-Making** | | | | | | |
| Simulation 1: Atrial Fibrillation | 3 | 3 | 4 | 3 | 3 | 4 |
| Simulation 1: Altered Mental Status | 2 | 1 | 3 | 2 | 2 | 3 |
| Simulation 2: Chest Pain | 4 | 4 | 4 | 4 | 4 | 4 |
| Simulation 2: Oliguria | 3 | 3 | 4 | 3 | 2 | 4 |
| Simulation 3: Trauma | 3 | 2 | 4 | 3 | 2 | 4 |
| **Domain – Communication with the Registered Nurse** | | | | | | |
| Simulation 1: Atrial Fibrillation | 3 | 2 | 4 | 3.5 | 1 | 5 |
| Simulation 1: Altered Mental Status | 3 | 2 | 5 | 3.5 | 2 | 4 |
| Simulation 2: Chest Pain | 3 | 3 | 4 | 4 | 3 | 4 |
| Simulation 2: Oliguria | 3 | 2 | 3 | 4 | 1 | 4 |
| Simulation 3: Trauma | 1.5 | 1 | 3 | 2 | 1 | 4 |
| **Domain – Communication With the Senior Resident** | | | | | | |
| Simulation 1: Atrial Fibrillation | 3 | 1 | 4 | 3 | 1 | 4 |
| Simulation 1: Altered Mental Status | 1 | 1 | 1 | 1 | 1 | 2 |
| Simulation 2: Chest Pain | 4 | 2.5 | 4 | 4 | 4 | 5 |
| Simulation 2: Oliguria | 3 | 2 | 4 | 4 | 1.5 | 4 |
| Simulation 3: Trauma | 4 | 4 | 5 | 4 | 4 | 5 |
| **Domain – Overall Approach** | | | | | | |
| Simulation 1: Atrial Fibrillation | 3 | 3 | 4 | 2 | 2 | 3 |
| Simulation 1: Altered Mental Status | 3 | 2 | 3 | 2 | 1 | 2 |
| Simulation 2: Chest Pain | 3 | 3 | 4 | 3 | 3 | 4 |
| Simulation 2: Oliguria | 3 | 3 | 3 | 3 | 3 | 4 |
| Simulation 3: Trauma | 3 | 3 | 3 | 3 | 3 | 4 |

### Overall Approach

After reviewing several of the recorded simulations, it was observed that facilitators occasionally guided or helped students with data collection and treatments. To account for this variability in the scoring, the amount of help provided was added to the scoring descriptions in overall approach. Help provided in both data and treatment was added to the score of 2, help in either domain was added to the score of 3, and minimal help was added to the score of 4 (Figure 2).

### Analysis

Audio-visual simulation recordings were evaluated by novice and expert reviewers. The novice reviewers were 2 medical student authors who had just completed their first year of medical school and had limited clinical exposure (SU, JL) and a surgical education researcher (AR). While 1 student (SU) helped to create the modified checklists, all 3 were content novices. One surgery resident who completed 2 clinical years of training (LK) and 1 fifth-year surgery resident (AS) served as expert reviewers in the context of these common on-call scenarios. Neither were involved in checklist creation. The novices and experts used the scenario-specific assessments (m-NOTECHS) for each simulation. Descriptive statistics were calculated. One novice and 1 expert were randomly selected to assess inter-item reliability to assess internal consistency of scores. Intraclass correlations were conducted to evaluate reliability among novices and experts. The novice and expert reviewers' scores were used for research purposes only and not provided to students.

**Table 2.** Mean Inter-Item Correlation for 1 Randomly Selected Novice (Top Number) and 1 Randomly Selected Expert (Bottom Number)

| | | Data Collection | Differential Diagnosis | Medical Decision-Making | Communication with RN | Communication wtih SR | Overall Approach |
|---|---|---|---|---|---|---|---|
| **Similation 1: Atrial Fibrillation** | Data collection | 1.0 | 0.37 | -0.14 | -0.13 | -0.38 | -0.01 |
| | | 1.0 | 0.57 | -0.15 | -0.40 | -0.24 | 0.42 |
| | Differential diagnosis | 0.37 | 1.0 | -0.19 | -0.46 | -0.19 | 0.10 |
| | | 0.57 | 1.0 | 0.05 | 0.05 | -0.63 | 0.47 |
| | Medical decision-making | -0.14 | -0.19 | 1.0 | 0.59 | 0.30 | 0.41 |
| | | -0.15 | 0.05 | 1.0 | 0.74 | -0.13 | 0.06 |
| | Comm with RN | -0.13 | -0.46 | 0.59 | 1.0 | 0.06 | 0.62 |
| | | -0.40 | 0.05 | 0.74 | 1.0 | -0.05 | 0.27 |
| | Comm with senior resident | -0.38 | -0.19 | 0.30 | 0.06 | 1.0 | -0.27 |
| | | -0.24 | -0.63 | -0.13 | -0.05 | 1.0 | -0.26 |
| | Overall approach | -0.01 | 0.10 | 0.41 | 0.62 | -0.27 | 1.0 |
| | | 0.42 | 0.47 | 0.06 | 0.27 | -0.26 | 1.0 |
| **Simulation 1: Altered Mental Status** | Data collection | 1.0 | -0.13 | 0.28 | 0.35 | -0.29 | -0.10 |
| | | 1.0 | -0.13 | 0.59 | 0.38 | 0.18 | 0.41 |
| | Differential diagnosis | -0.13 | 1.0 | 0.40 | -0.55 | -0.04 | 0.03 |
| | | -0.13 | 1.0 | -0.19 | -0.18 | 0.66 | 0.13 |
| | Medical decision-making | 0.28 | 0.40 | 1.0 | 0.05 | 0.32 | 0.15 |
| | | 0.59 | -0.19 | 1.0 | 0.70 | 0.18 | 0.69 |
| | Comm with RN | 0.35 | -0.55 | 0.05 | 1.0 | -0.04 | -0.21 |
| | | 0.38 | -0.18 | 0.70 | 1.0 | -0.12 | 0.77 |
| | Comm with senior resident | -0.29 | -0.04 | 0.32 | -0.04 | 1.0 | -0.20 |
| | | 0.18 | 0.66 | 0.18 | -0.12 | 1.0 | 0.07 |
| | Overall approach | -0.10 | 0.03 | 0.15 | -0.21 | -0.20 | 1.0 |
| | | 0.41 | 0.13 | 0.69 | 0.77 | 0.07 | 1.0 |
| **Simulation 2: Chest Pain** | Data collection | 1.0 | 0.24 | -0.35 | -0.20 | 0.13 | 0.13 |
| | | 1.0 | -0.27 | | -0.53 | -0.14 | 0.01 |
| | Differential diagnosis | 0.24 | 1.0 | -0.20 | 0.20 | 0.23 | 0.60 |
| | | -0.27 | 1.0 | | 0.29 | 0.06 | 0.22 |
| | Medical decision-making | -0.35 | -0.20 | 1.0 | 0.00 | -0.15 | 0.42 |
| | | | 1.0 | | | | |
| | Comm with RN | -0.20 | 0.20 | 0.00 | 1.0 | 0.34 | 0.24 |
| | | -0.53 | 0.29 | | 1.0 | -0.15 | 0.13 |
| | Comm with senior resident | 0.13 | 0.23 | -0.15 | 0.34 | 1.0 | 0.56 |
| | | -0.14 | 0.06 | | -0.15 | 1.0 | 0.67 |
| | Overall approach | 0.13 | 0.60 | 0.42 | 0.24 | 0.56 | 1.0 |
| | | 0.01 | 0.22 | | 0.13 | 0.67 | 1.0 |
| **Simulation 2: Oliguria** | Data collection | 1.0 | 0.53 | -0.27 | -0.04 | 0.08 | 0.10 |
| | | 1.0 | 0.54 | 0.18 | 0.22 | 0.23 | 0.00 |
| | Differential diagnosis | 0.53 | 1.0 | -0.28 | 0.22 | 0.00 | 0.21 |
| | | 0.54 | 1.0 | 0.10 | 0.41 | -0.18 | 0.00 |
| | Medical decision-making | -0.27 | -0.28 | 1.0 | 0.46 | -0.27 | -0.24 |
| | | 0.18 | 0.10 | 1.0 | 0.25 | 0.44 | 0.53 |
| | Comm with RN | -0.04 | 0.22 | 0.46 | 1.0 | 0.23 | 0.17 |
| | | 0.22 | 0.41 | 0.25 | 1.0 | 0.29 | 0.53 |
| | Com with Senior Resident | 0.08 | 0.00 | -0.27 | 0.23 | 1.0 | 0.72 |
| | | 0.23 | -0.18 | 0.44 | 0.29 | 1.0 | 0.70 |
| | Overall Approach | 0.10 | 0.21 | -0.24 | 0.17 | 0.72 | 1.0 |
| | | 0.00 | 0.00 | 0.53 | 0.53 | 0.70 | 1.0 |
| **Simulation 3: Trauma** | Data collection | 1.0 | 0.19 | 0.29 | 0.12 | 0.49 | 0.74 |
| | | 1.0 | 0.08 | -0.24 | 0.28 | 0.38 | 0.50 |
| | Differential diagnosis | 0.19 | 1.0 | 0.47 | -0.01 | 0.43 | 0.28 |
| | | 0.08 | 1.0 | 0.45 | -0.27 | -0.04 | -0.07 |
| | Medical decision-making | 0.29 | 0.47 | 1.0 | 0.14 | 0.39 | 0.51 |
| | | -0.24 | 0.45 | 1.0 | -0.15 | 0.17 | -0.23 |
| | Comm with RN | 0.12 | -0.01 | 0.14 | 1.0 | -0.12 | 0.15 |
| | | 0.28 | -0.27 | -0.15 | 1.0 | 0.19 | 0.03 |
| | Comm with senior resident | 0.49 | 0.43 | 0.39 | -0.12 | 1.0 | 0.52 |
| | | 0.38 | -0.04 | 0.17 | 0.19 | 1.0 | 0.27 |
| | Overall approach | 0.74 | 0.28 | 0.51 | 0.15 | 0.52 | 1.0 |
| | | 0.50 | -0.07 | -0.23 | 0.03 | 0.27 | 1.0 |

There was no variability in medical decision-making scores by the expert for the chest pain scenario.

## RESULTS

In total, 22 learners participated in the IPC simulations. Participants were split into 2 groups. For simulation 1, one group participated in the atrial fibrillation scenario (n=11) and the other participated in the altered mental status scenario (n=11). For simulation 2, one group participated in the oliguria scenario (n=10) and the other participated in the chest pain scenario (n=12). For simulation 3, all learners participated in the trauma scenario (n=17). Five learners in the IPC did not participate in the simulations due to scheduling conflicts or were unable to be scored due to facilitator deviation from the script.

Descriptive statistics are shown in Table 1. Learners scored the lowest in the differential diagnosis domain and highest in the data collection domain. Learners scored the lowest in the altered mental status scenario and the highest in the chest pain scenario. Inter-item correlation for Novice #3 and Expert #2 are shown in Table 2.[14]

Intraclass correlation coefficients (ICC) between the 3 novices, 2 experts, and 3 novices and 2 experts were calculated for each scenario and each rubric domain and are shown in Table 3. The novice reviewers had good to excellent correlation in their summation of all domains of 4 simulations and moderate to good correlation in their summation of all domains of 1 simulation (chest pain). The novice and expert reviewers had moderate to good correlation in their summation of all domains of 4 simulations and good correlation in their summation of all domains of 1 simulation (altered mental status).[15] Domain-specific correlations vary. The lowest correlations were seen in the overall approach domain.

**Table 3.** Intraclass Correlation Coefficient (ICC) Calculations with 95% CI Between the Three Novices, Two Experts, and Among the Three Novices and Two Expert Clinicians for Each Simulation and Each Rubric Section

| | 3 Novices | 2 Experts | 3 Novices + 2 Experts |
|---|---|---|---|
| **Summation of all Domains** | | | |
| Simulation 1: Atrial Fibrillation | 0.89 < ICC < 0.95 | 0.56 < ICC < 0.83 | 0.67 < ICC < 0.83 |
| Simulation 1: Altered Mental Status | 0.85 < ICC < 0.93 | 0.85 < ICC < 0.94 | 0.77 < ICC < 0.88 |
| Simulation 2: Chest Pain | 0.66 < ICC < 0.83 | 0.71 < ICC < 0.89 | 0.51 < ICC < 0.72 |
| Simulation 2: Oliguria | 0.78 < ICC < 0.90 | 0.68 < ICC < 0.87 | 0.63 < ICC < 0.81 |
| Simulation 3: Trauma | 0.89 < ICC < 0.95 | 0.56 < ICC < 0.78 | 0.55 < ICC < 0.72 |
| **Domain – Data Collection** | | | |
| Simulation 1: Atrial Fibrillation | 0.02 < ICC < 0.81 | -0.36 < ICC < 0.80 | 0.23 < ICC < 0.83 |
| Simulation 1: Altered Mental Status | 0.51 < ICC < 0.93 | 0.49 < ICC < 0.95 | 0.62 < ICC < 0.93 |
| Simulation 2: Chest Pain | 0.53 < ICC < 0.92 | 0.50 < ICC < 0.96 | 0.46 < ICC < 0.90 |
| Simulation 2: Oliguria | 0.81 < ICC < 0.98 | -0.07 < ICC < 0.87 | 0.51 < ICC < 0.92 |
| Simulation 3: Trauma | 0.68 < ICC < 0.94 | -0.57 < ICC < 0.38 | 0.36 < ICC < 0.80 |
| **Domain – Differential Diagnoses** | | | |
| Simulation 1: Atrial Fibrillation | 0.71 < ICC < 0.97 | 0.00 < ICC < 0.903 | 0.66 < ICC < 0.96 |
| Simulation 1: Altered Mental Status | 0.10 < ICC < 0.79 | -0.15 < ICC < 0.82 | 0.14 < ICC < 0.73 |
| Simulation 2: Chest Pain | 0.25 < ICC < 0.84 | -0.60 < ICC < 0.60 | 0.33 < ICC < 0.86 |
| Simulation 2: Oliguria | 0.68 < ICC < 0.96 | 0.75 < ICC < 0.98 | 0.74 < ICC < 0.96 |
| Simulation 3: Trauma | 0.57 < ICC < 0.91 | 0.10 < ICC < 0.82 | 0.33 < ICC < 0.78 |
| **Domain – Medical Decision-Making** | | | |
| Simulation 1: Atrial Fibrillation | 0.90 < ICC < 0.99 | 0.224 < ICC < 0.94 | 0.75 < ICC < 0.97 |
| Simulation 1: Altered Mental Status | 0.81 < ICC < 0.98 | 0.39 < ICC < 0.94 | 0.56 < ICC < 0.92 |
| Simulation 2: Chest Pain | 0.31 < ICC < 0.86 | - | 0.09 < ICC < 0.70 |
| Simulation 2: Oliguria | 0.80 < ICC < 0.98 | -0.11 < ICC < 0.86 | 0.38 < ICC < 0.87 |
| Simulation 3: Trauma | 0.85 < ICC < 0.98 | 0.13 < ICC < 0.83 | 0.39 < ICC < 0.81 |
| **Domain – Communication with the Registered Nurse** | | | |
| Simulation 1: Atrial Fibrillation | 0.79 < ICC < 0.98 | 0.00 < ICC < 0.90 | 0.48 < ICC < 0.92 |
| Simulation 1: Altered Mental Status | 0.54 < ICC < 0.93 | 0.40 < ICC < 0.94 | 0.53 < ICC < 0.91 |
| Simulation 2: Chest Pain | 0.20 < ICC < 0.82 | -0.46 < ICC < 0.72 | 0.13 < ICC < 0.74 |
| Simulation 2: Oliguria | 0.22 < ICC < 0.86 | 0.42 < ICC < 0.95 | 0.30 < ICC < 0.84 |
| Simulation 3: Trauma | 0.92 < ICC < 0.99 | 0.29 < ICC < 0.87 | 0.19 < ICC < 0.68 |
| **Domain – Communication with the Senior Resident** | | | |
| Simulation 1: Atrial Fibrillation | 0.77 < ICC < 0.98 | 0.71 < ICC < 0.98 | 0.67 < ICC < 0.96 |
| Simulation 1: Altered Mental Status | 0.84 < ICC < 0.98 | 0.37 < ICC < 0.94 | 0.57 < ICC < 0.92 |
| Simulation 2: Chest Pain | 0.25 < ICC < 0.84 | 0.46 < ICC < 0.96 | -0.04 < ICC < 0.55 |
| Simulation 2: Oliguria | 0.92 < ICC < 0.99 | 0.82 < ICC < 0.99 | 0.85 < ICC < 0.98 |
| Simulation 3: Trauma | 0.43 < ICC < 0.87 | 0.35 < ICC < 0.89 | 0.17 < ICC < 0.67 |
| **Domain – Overall Approach** | | | |
| Simulation 1: Atrial Fibrillation | 0.17 < ICC < 0.87 | -0.28 < ICC < 0.83 | 0.27 < ICC < 0.85 |
| Simulation 1: Altered Mental Status | -0.15 < ICC < 0.61 | 0.44 < ICC < 0.95 | 0.12 < ICC < 0.71 |
| Simulation 2: Chest Pain | -0.23 < ICC < 0.48 | 0.17 < ICC < 0.92 | 0.02 < ICC < 0.64 |
| Simulation 2: Oliguria | 0.04 < ICC < 0.79 | 0.04 < ICC < 0.89 | 0.21 < ICC < 0.79 |
| Simulation 3: Trauma | 0.52 < ICC < 0.90 | 0.03 < ICC < 0.79 | 0.13 < ICC < 0.64 |

Of note, the 2 experts gave all learners the same score for Simulation 2: Chest Pain; thus, no ICC could be reported.

## DISCUSSION

This exploratory study shows that our m-NOTECHS can be used by clinical novices to evaluate student performance with little variation in scores between expert and novice raters. Thus, the m-NOTECHS could provide consistent novice scoring of simulated performance, with novice scores correlating with those of an expert reviewer. Although outside of the scope of this study, clinical novices likely gain knowledge from evaluating simulation performance. Simulation participants and novice reviewers alike may benefit from this method of assessment—an area ripe for future study.[16]

In addition to a lack of available expert reviewers, other barriers to SBME implementation exist, including the need for confederate actors, simulation center staff, and reliable equipment. Future work should further reduce the number of other resources required to run educational simulations. Furthermore, our described meth-

odology can be used to adapt and improve novice-friendly scoring systems for a wider variety of nontechnical simulations used across multiple medical specialties.

This study is limited by the small number of students participating in simulations at a single institution and a limited number of reviewers for evaluation of the m-NOTECHS. Despite providing simulation scripts, utilizing multiple facilitators led to subtle variation in the simulation. Additionally, evaluators were not blinded to the names of the medical students they were evaluating, which may introduce bias in scoring. The inter-item correlations varied widely and suggest that further refinement of the scoring rubrics may improve internal consistency.

Regardless, these tools and simulations can be used and modified for SBME and serve as resources for medical educators. Using a modified scoring rubric with explicit, jargon-free criteria can allow for timely and accurate review of complex medical decision-making by novices. This may eliminate 1 barrier to implementation and encourage continued use of simulation in medical education, a resource-intense learning tool.

**Funding/Support:** This work was supported by the National Cancer Institute of the National Institutes of Health (NIH), grant T32CA090217. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors would like to thank the University of Wisconsin School of Medicine and Public Health Shapiro Summer Research Fellowship Program and the Department of Surgery for their financial support of this project.

**Financial Disclosures:** None declared.

**Appendices:** Available at wmjonline.org.

## REFERENCES

**1.** American Board of Surgery; American College of Surgeons; Association of Program Directors in Surgery; Association for Surgical Education. Statement on surgical preresidency preparatory courses. *Am Surg*. 2014;80(11):1085-1086. doi:10.1007/s00268-014-2749-y

**2.** Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. doi:10.1080/01421590500046924

**3.** McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med*. 2011;86(6):706-711. doi:10.1097/ACM.0b013e318217e119

**4.** Cook DA. How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Med Educ*. 2014;48(8):750-760. doi:10.1111/medu.12473

**5.** Antonoff MB, Swanson JA, Green CA, Mann BD, Maddaus MA, D'Cunha J. The significant impact of a competency-based preparatory course for senior medical students entering surgical residency. *Acad Med*. 2012;87(3):308-319. doi:10.1097/ACM.0b013e318244bc71

**6.** Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care*. 2009;18(2):104-108. doi:10.1136/qshc.2007.024760

**7.** Chen C, White L, Kowalewski T, et al. Crowd-sourced assessment of technical skills: a novel method to evaluate surgical performance. *J Surg Res*. 2014;187(1):65-71. doi:10.1016/j.jss.2013.09.024

**8.** Holst D, Kowalewski TM, White LW, et al. Crowd-sourced assessment of technical skills: an adjunct to urology resident surgical simulation training. *J Endourol*. 2015;29(5):604-609. doi:10.1089/end.2014.0616

**9.** Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. *JAMA Surg*. 2015;150(11):1086-1087. doi:10.1001/jamasurg.2015.2405

**10.** Powers MK, Boonjindasup A, Pinsky M, et al. Crowdsourcing assessment of surgeon dissection of renal artery and vein during robotic partial nephrectomy: a novel approach for quantitative assessment of surgical performance. *J Endourol*. 2016;30(4):447-452. doi:10.1089/end.2015.0665

**11.** Liepert AE, Velic AJ, Rademacher B, et al. Proficiency development for graduating medical students, using skills-level-appropriate mastery learning versus traditional learning for chest tube placement: Assessing anxiety, confidence, and performance. *Surgery*. 2019;165(6):1075-1081. doi:10.1016/j.surg.2019.01.015

**12.** Blumenfeld A, Velic A, Bingman EK, et al. A Mastery Learning Module on Sterile Technique to Prepare Graduating Medical Students for Internship. *MedEdPORTAL*. 2020; Jun 29;16:10914. doi:10.15766/mep_2374-8265.10914

**13.** Lopreiato JO. *Healthcare Simulation Dictionary*. Agency for Healthcare Research and Quality; 2016. AHRQ Publication No. 16(17)-0043. Accessed Nov 30, 2021. https://www.ahrq.gov/sites/default/files/publications/files/sim-dictionary_0.pdf

**14.** Piedmont RL. Inter-item correlations. In: Michalos AC, ed. *Encyclopedia of Quality of Life and Well-being Research*. Springer; 2014:3303-3304.

**15.** Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

**16.** Krogh CL, Ringsted C, Kromann CB, et al. Effect of engaging trainees by assessing peer performance: a randomised controlled trial using simulated patient scenarios. *Biomed Res Int*. 2014;2014:610591. doi:10.1155/2014/610591

advancing the art & science of medicine in the midwest

WMJ

*WMJ* (ISSN 1098-1861) is published through a collaboration between The Medical College of Wisconsin and The University of Wisconsin School of Medicine and Public Health. The mission of *WMJ* is to provide an opportunity to publish original research, case reports, review articles, and essays about current medical and public health issues.

© 2022 Board of Regents of the University of Wisconsin System and The Medical College of Wisconsin, Inc.

**Visit www.wmjonline.org to learn more.**