Validation of “Time to Task” Performance Assessment Method in Simulation

• Mary Ann Shinnick, PhD, ACNP, CCNS, CHSE
• Mary A. Woo, PhD
International Nursing Association for Clinical Simulation & Learning is accredited as a provider of continuing nursing education by the American Nurses Credentialing Center’s Commission on Accreditation.
DISCLOSURES

Conflict of Interest
- Mary Ann Shinnick: No Disclosures
- Mary Woo: No Disclosures

Successful Completion
- Attend 100% of session
- Complete online evaluation
Learning Outcomes

Upon completion of this educational activity, participants will be able to:

1. List common methods of assessment in nursing and simulation

2. Describe research efforts to test three different assessment methods

3. Report recommended next steps for competency assessment using simulation
Background

- Patient safety relies on health care educators to train and certify safe, competent practitioners.
- There is a lack of objective, standardized, valid and reliable measures of nursing competence in both simulation and the clinical setting.
- **Current methods to assess clinical performance:**
  - Subjective evaluation by clinical preceptors/instructors “Pass/Fail”
    - are known to be often invalid and unreliable.
  - Simulation Assessments
  - Time to task completion, “time to task”-used in CPR, Acute Coronary Syndrome and Stroke
Study

Compared three assessment methods in order to validate one as being best

1. **Objective assessment of clinical competence:** “Time to Task”
   a. ability to perform specific, critical nursing care activities within 5 minutes

2. **Objective (?) assessment used in simulation based on Tanner’s Clinical Judgment Theory -- The Lasater Clinical Judgement Rubric (LCJR)**

3. **Subjective assessment, the common “pass/fail” assessment**
   a. subjective measure of whether the subject would “pass” or “fail” as a senior nursing student
Methods: Design

- Prospective, “Known Groups” (expert vs. novice nurses) design
  - Group 1: Expert nurses (ICU nurses with > 5 years of clinical experience)
  - Group 2: Novice nurses (senior prelicensure nursing students)
  - Both groups participated individually in a simulation of a patient in decompensated heart failure.
- Twenty eight participants completed the simulation (Experts [n = 15]; Novices [n = 13]).
- Video and audio capture by eye tracking camera system
Eye Tracking Glasses
Methods (cont.)

• All 28 videos were reviewed by 14 reviewers
• Each reviewer scored the participant on the LCJR, subjective pass/fail
  • There were basic instructions given to the Reviewers but no formal training for either of the assessments.
• The researcher coded all videos for time to task in 5 areas based upon expert faculty recommendation:
  1. “elevated the head of the bed”
  2. “applied the pulse oximeter”
  3. “administered oxygen”
  4. “auscultated lungs”
  5. “chose correct medication“ [i.e., diuretic]).
Assessments
The Simulation

- Simulation of patients with HF is important to developing nurses who can rapidly identify and treat patients who have decompensated.
- As the most common hospital discharge diagnosis in the United States in patients aged 65 years and older this is a common situation such that all levels of nurses are likely to encounter.
Clinical Knowledge Questionnaire

- A 5-item, multiple choice format HF Clinical Knowledge Questionnaire
  - administered to each subject following the simulation.
  - It addressed nursing interventions applicable to care of a patient who was dyspneic that would apply to the HF patient but HF was not explicitly identified.
Demographic Questionnaire

- Participant's age, gender, history of personal or family experience with HF, prior simulation exposure, years and type of employment and previous employment of the nursing students as a nurse helper (i.e., nurse's aide, care partner, etc.)
- Collected after the simulation so the topic of the simulation would not be divulged.
Assessment: Lasater Clinical Judgment Rubric (LCJR)

- Designed to evaluate a subject in a single episode using clinical judgment.
- It has been used in several simulation studies which reported validity and in which faculty were able to consistently and accurately recognize the correct level of student performance.\(^3\)
- Reliability from three different studies is reported as 57%-100%.\(^4\)
- The LCJR measures proficiency in four categories from Exemplary to Novice.\(^5\)
- Within each category are 11 domains
- For the purpose of this study, 4 domains were eliminated that did not pertain to the subject matter of the simulation ("information seeking", "clear communication", "evaluation/self-analysis" and "commitment to improvement")
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Exemplary</th>
<th>Accomplished</th>
<th>Developing</th>
<th>Beginning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focused Observation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focuses observation appropriately; regularly observes and monitors a wide variety of objective and subjective data to uncover any useful information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regularly observes and monitors a variety of data, including both subjective and objective; most useful information is noticed; may miss the most subtle signs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recognizing deviations from expected patterns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognizes subtle patterns and deviations from expected patterns in data and uses these to guide the assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognizes most obvious patterns and deviations in data and uses these to continually assess</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effective Interpreting involves:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prioritizing Data</td>
<td>Generally focuses on the most important data and seeks further relevant information but also may try to attend to less pertinent data</td>
<td>Makes an effort to prioritize and focus on the most important, but also attends to less relevant or useful data</td>
<td>Has difficulty focusing and appears not to know which data are most important to the diagnostic attempts to attend to all available data</td>
<td></td>
</tr>
<tr>
<td>Making sense of data</td>
<td>Even when facing complex, conflicting, or confusing data, is able to notice and make sense of patterns in the patient’s data</td>
<td>Is able to compare the patient’s data patterns and compare with known patterns to develop an intervention plan and accompanying rationale</td>
<td>Has difficulty interpreting or making sense of data requires assistance both in diagnosing the problem and in developing an intervention</td>
<td></td>
</tr>
<tr>
<td><strong>Effective Responding involves:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calm, confident manner</td>
<td>Assumes responsibility, involves patient and measures them</td>
<td>Generally displays leadership and confidence and is able to control or calm most situations; may show stress in difficult or complex situation</td>
<td>Is tentative in leader role; measures patient in simple situation but becomes stressed and disorganized easily</td>
<td>Is stressed and disorganized, lacks control, makes patient anxious</td>
</tr>
<tr>
<td>Well planned intervention; flexibility</td>
<td>Interventions are tailored for the individual patient; monitors patient progress closely and is able to adjust treatment as indicated by patient response</td>
<td>Develops interventions on the basis of the most obvious data; monitors patient progress regularly but does not expect to have to change treatments</td>
<td>Develops interventions on the basis of the most obvious data; monitors patient progress but is unable to make adjustments as indicated by patient’s response</td>
<td>Focuses on developing a single intervention, addressing a single problem, but it may be vague, confusing, and incomplete; some monitoring or skill may not be observed</td>
</tr>
<tr>
<td>Being skillful</td>
<td>Shows mastery of necessary nursing skills</td>
<td>Exemplifies proficiency in the use of most nursing skills; could improve speed or accuracy</td>
<td>Is hortassed or ineffective in using nursing skills</td>
<td>Is unable to select and/or perform nursing skills</td>
</tr>
</tbody>
</table>

**Research will Score:** Exemplary = 4 pts; Accomplished = 3 pts; Developing = 2 pts; Beginning = 1 pt. Total points possible = 28
Subjective “Pass/Fail”

• Replicating the common practice of clinical faculty or Preceptors assessing whether a student passes or fails a clinical rotation based on subjective recall of performance.

• Would the subject “pass” if they were a senior nursing student. This was a “Yes” or “No” response.
  • This same assessment was recently used in another study exploring the use of simulation for high stakes testing which asked raters if the simulation video participant was competent.
Time to Task

- The time needed to achieve each task was calculated from each video by the researcher.
- Elements to be completed in the first 5 minutes as deemed critical to this specific case by an expert clinical nursing panel and included:
  1. “elevated the head of the bed”
  2. “applied the pulse oximeter”
  3. “administered oxygen”
  4. “auscultated lungs”
  5. “chose correct medication” [i.e., diuretic].
Statistical Analysis

- Unpaired, independent T-Tests for demographic data and the HF Clinical Knowledge Questionnaire
- Non-parametric Median’s test and Chi² for Time to Task data.
- Statistical significance was set at $p < 0.05$.
- To determine the ability of each assessment to best differentiate between the Expert and Novice nurses, sensitivity and specificity analysis was also done.
Sensitivity & Specificity

• Sensitivity
  • Ability to pick out persons with the condition (i.e., positive)

• Specificity
  • Ability to pick out persons who do not have the condition (are negative)

• True Positives: those correctly identified as experts (nurses)

• True Negatives: those correctly identified as novices (students)
# Demographics
(Experts $n=15$; Novices $n=13$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean/SD</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>38.9 ± 10.1</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td>Novice</td>
<td>25.9 ± 6.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.9 ± 6.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>38.9 ± 10.1</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of Prior Simulations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>4.2 ± 2.3</td>
<td>0.51</td>
</tr>
<tr>
<td>Novice</td>
<td>4.7 ± 1.4</td>
<td></td>
</tr>
<tr>
<td><strong>HF Clinical Knowledge Questionnaire</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>4.2 ± 2.3</td>
<td>0.24</td>
</tr>
<tr>
<td>Novice</td>
<td>3.8 ± 1.1</td>
<td></td>
</tr>
</tbody>
</table>
Video Reviewers

- Video review of all 28 videos was completed by prelicensure nursing faculty and hospital clinical preceptors (n=14)
- Reviewers: all reported experience evaluating nurses in the clinical setting
  - 93% reported experience evaluating experienced nurses in the clinical setting as well.
  - 43% considered themselves expert in this area
  - 50% had experience as a hospital clinical preceptor to nursing students.
- Employment areas of the Reviewers:
  - Academia (57%), the hospital setting (27%), or both (14%)
Nursing Background of Reviewers

- Critical Care: 43%
- Peds/Msurg: 29%
- PQCU/OR: 7%
- Telemetry: 7%
- Mental Health: 7%
- Education: 7%
Results: LCJR

- While the score agreement among reviewers on individual videos was inconsistent, statistical significance was found between the mean scores among all reviewers ($p < 0.01$) except for one ($p = 0.38$).
- The overall LCJR mean scores were of Expert ($22.58 \pm 8.14$) and Novice ($14.9 \pm 14.6$) out of 28 possible points.
- Sensitivity for this instrument was good (indicating a fair ability to identify “Expert” nurses; 0.72),
  - but it had a low specificity (inability to identify “Novice” nurses; 0.40).
<table>
<thead>
<tr>
<th>Reviewer</th>
<th>Employment Area</th>
<th>Years as RN</th>
<th>Experience Evaluating Students</th>
<th>Mean Scores/SD</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewer 1</td>
<td>Hospital</td>
<td>26</td>
<td>Yes</td>
<td>Expert Novice 22.87 ± 6.31</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14.31 ± 7.79</td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>Academia 2 year RN program</td>
<td>26</td>
<td>Yes</td>
<td>Expert Novice 24.67 ± 4.22</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17.08 ± 7.48</td>
<td></td>
</tr>
<tr>
<td>Reviewer 3</td>
<td>Academia BSN program</td>
<td>49</td>
<td>Yes</td>
<td>Expert Novice 23.87 ± 5.33</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15.69 ± 4.64</td>
<td></td>
</tr>
<tr>
<td>Reviewer 4</td>
<td>Hospital</td>
<td>30</td>
<td>Yes</td>
<td>Expert Novice 16.40 ± 3.53</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.54 ± 3.82</td>
<td></td>
</tr>
<tr>
<td>Reviewer 5</td>
<td>Novice Academia</td>
<td>47</td>
<td>Yes</td>
<td>Expert Novice 25.40 ± 5.03</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18.62 ± 5.62</td>
<td></td>
</tr>
<tr>
<td>Reviewer 6</td>
<td>Academia BSN program</td>
<td>38</td>
<td>Yes</td>
<td>Expert Novice 23.40 ± 5.04</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15.08 ± 5.63</td>
<td></td>
</tr>
<tr>
<td>Reviewer 7</td>
<td>Academia and Hospital</td>
<td>40</td>
<td>Yes</td>
<td>Expert Novice 19.27 ± 7.00</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.54 ± 3.57</td>
<td></td>
</tr>
<tr>
<td>Reviewer 8</td>
<td>Hospital novice</td>
<td>6</td>
<td>Yes</td>
<td>Expert Novice 25.87 ± 2.64</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20.38 ± 4.27</td>
<td></td>
</tr>
<tr>
<td>Reviewer 9</td>
<td>Hospital</td>
<td>14</td>
<td>NO</td>
<td>Expert Novice 23.93 ± 4.33</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14.15 ± 5.59</td>
<td></td>
</tr>
<tr>
<td>Reviewer 10</td>
<td>Academia 2 year RN program</td>
<td>32</td>
<td>Yes</td>
<td>Expert Novice 16.67 ± 4.92</td>
<td>0.376</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15.23 ± 3.17</td>
<td></td>
</tr>
<tr>
<td>Reviewer 11</td>
<td>Academia BSN program</td>
<td>37</td>
<td>Yes</td>
<td>Expert Novice 25.87 ± 3.40</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17.23 ± 6.89</td>
<td></td>
</tr>
<tr>
<td>Reviewer 12</td>
<td>Academia BSN program</td>
<td>38</td>
<td>Yes</td>
<td>Expert Novice 26.53 ± 1.19</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18.38 ± 8.28</td>
<td></td>
</tr>
<tr>
<td>Reviewer 13</td>
<td>Novice Academia</td>
<td>34</td>
<td>Yes</td>
<td>Expert Novice 19.60 ± 7.15</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.62 ± 4.01</td>
<td></td>
</tr>
<tr>
<td>Reviewer 14</td>
<td>Academia 2 year RN program</td>
<td>25</td>
<td>Yes</td>
<td>Expert Novice 22.13 ± 3.94</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14.92 ± 6.55</td>
<td></td>
</tr>
</tbody>
</table>
“Pass/Fail” Subjective Assessment

- The Reviewers each recorded a subjective decision whether the simulation participant would “pass” or “fail” if they were a senior nursing student.
- The reviewers reported a “failing” vote for 3 of the 15 Experts (20%) and 2 of the 13 Novices (15%).
- Sensitivity was high (0.90)
- Specificity was low (0.47)
  - indicating an ability to identify “Expert” nurses but inability to identify “Novice” nurses.
- 13 of 14 Reviewers failed one or two Experts
  - one Reviewer failed 3 Experts. This Reviewer had more years as a nurse (49), was a retired Army nurse, had extensive experience in both the clinical and academic settings.
Time to Task

- There was statistical significance between the groups in the Time to Task assessment ($p < 0.001$).
- Sensitivity and Specificity for this assessment was good (ability to correctly identify “Expert” nurses: 0.80 and ability to correctly identify “Novice” nurses: 0.85)
- 3 Experts did not complete the expected tasks in 5 minutes yet 2 Novices did.
Time to Task (cont.)

- Of the three Experts who did poorly on the Time to Task assessment, most of the Reviewers subjectively “passed” them
  - Expert participant #1: passed by 50% of Reviewers
  - Expert participant #2: passed by 100% of Reviewers
  - Expert participant #3 passed by 93% of Reviewers
  - For these same Expert participants, the LCJR scores were 16/28, 23/28 and 21/28, respectively.
- For the two Novices who achieved the key tasks in 5 minutes, one was “passed” by all of the Reviewers and was one was “failed” by 57% of the Reviewers.
- The LCJR scores for these two subjects were 23/28 and 13/28, respectively.
## Summary Analysis between Assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasater Clinical Judgment Rubric</td>
<td>.72</td>
<td>.40</td>
</tr>
<tr>
<td>Time to Task</td>
<td>.80</td>
<td>.85</td>
</tr>
<tr>
<td>Subjective “Pass or Fail”</td>
<td>.90</td>
<td>.47</td>
</tr>
</tbody>
</table>
Conclusions

• This pilot study of three different assessments reveals significant weaknesses with two and promising results of the third.
• We found the LCJR to have good sensitivity but poor specificity with a wide range of variability of scores between raters, especially among the Novice participants.
• While the instrument may be helpful as a onetime assessment, it probably should not be relied upon for high stakes assessment.
• These findings on the LCJR were congruent with the recent findings of other researchers’ who found large inconsistencies among Reviewers of simulation videos using this instrument.
Conclusions (cont.)

- The subjective assessment “pass/fail” had high sensitivity but poor specificity.
- The Reviewers were able to pick out the Experts far more clearly than the Novices though they “failed” up to three of them.
  - This speaks to the subjectivity of this form of assessment.
  - The lack of specificity is concerning as the Novice needs to be recognized as such and assisted in further development in order to ensure patient safety.
Conclusions (cont.)

• The recommendation to exclude subjective measures of assessment is not new.
  • Awad and colleagues⁸, in studying assessment of surgical knowledge among medical students, found a weak correlation between subjective and objective measures evaluation.
  • These authors questioned the value of even including subjective assessments of knowledge as part of the students’ evaluations.

• Kardong-Edgren and colleagues⁷ add the recommendation to educators to be aware of evaluators who consistently have outlier scores, positive or negative, and suggest these faculty not be raters for high stakes assessment.
Conclusions (cont.)

- The objective measure, Time to Task performed much better than the LCJR or subjective “pass/fail” as a measure of clinical competency.
- Sensitivity and specificity were good and were superior to the other two assessments.
- Improvements in both sensitivity and specificity for Time to Task could possibly be achieved with simulation training focused on this outcome.
  - especially for high risk clinical events such as a patient with dyspnea that require basic tasks to be done in a timely manner in order to prevent patient deterioration.
Conclusions (cont.)

- While one single assessment instrument is not appropriate for all types of nursing assessment in simulation, an objective measure such as Time to Task warrants further study.

- Evaluators in the clinical and the simulation setting need to be clear on what is being assessed and choose the most appropriate instrument or instruments to determine competency, ensuring the use of valid and reliable tools with good sensitivity and specificity.

- A single instrument will not fit all types of simulation or clinical events so further study is warranted before instruments are adopted for formalized assessment.
Limitations

• Past experience with simulation nearly reached statistical significance between the groups and could also have been a limitation as this study only involved a single simulation event.
  • Next study a “warm up” simulation could be added for those with little to no experience with manikins.

• The sample size of Reviewers though each Reviewer viewed all 28 videos.

• Another potential limitation is the absence of Reviewer training in the use of the LCJR. Other studies have reported faculty training of 1-7 hours in the use of this tool.
  • The authors wished to explore the use of the instrument without any formal training as it represents common use of instruments in the literature and may further reinforce the use or addition of assessment methods that are more objective.
References


References (cont.)


Questions?
Contact Information

Mary Ann Shinnick, PhD, ACNP, CHSE
UCLA School of Nursing, Director of Simulation
mshinnic@sonnet.ucla.edu
310-794-4219