Operationalizing Assessment

Eric J. Warm M.D., F.A.C.P.
Program Director, Internal Medicine
Richard W. and Sue P. Vilter Professor of Medicine
University of Cincinnati College of Medicine
Twitter @Cincy1M
1. We are in the assessment business

2. We are not very good at assessment
Part 2

1. Describe our work in assessment

2. Assess our assessment system

3. Propose a way forward
Assessment

Hitting for average

Hitting for power

Base running skills and speed

Throwing ability

Fielding ability
Sabermetrics: the search for **objective knowledge** about baseball

- Base runs (BsR)
- Batting average on balls in play (BABIP)
- Defense independent pitching statistics (DIPS)
- Defense-Independent ERA
- Defense-Independent Component ERA
- Fielding independent pitching (FIP)
- Expected FIP (xFIP)
- Defensive Runs Saved (DRS)
- Equivalent average (EQA) Late-inning pressure situations (LIPS)
- On-base plus slugging (OPS)
- PECOTA (Player empirical comparison and optimization test algorithm)
- Peripheral ERA (PERA)

A's have compiled a 1484-1268 (.539) record over the **last 17 seasons**, which is the **fourth-best record** in the American League.

- Pythagorean expectation
- Range factor
- Runs created
- Secondary average
- Similarity score
- Speed Score
- Total player rating, or Batter-Fielder Wins (TPR, BFW)
- Total Pitcher Index, or Pitcher Wins (TPI, PW)
- Ultimate zone rating (UZR)
- Value over replacement player (VORP)
- Win shares
- wOBA
- Wins above replacement (WAR)

0 World Series Titles
Hercules

Percentiles vs. All

NFL Combine: the search for objective knowledge about football?
We Are In The Assessment Business

Smart
( enough)

Synthesis

Patient Engagement
( communication/kinesthetic)

Internal Locus of Control
( curiosity)

Reflective

Resilience/Grit

Medical Knowledge

Patient Care

Communication Skills

Systems-Based Practice

Practice-Based Learning

Professionalism
We Are In The Assessment Business

Smart (enough)

Synthesis

Patient Engagement (communication/kinesthetic)

Internal Locus of Control (curiosity)

Reflective

Resilience/Grit

Cognitive

Non-Cognitive

National Longitudinal Survey of Youth 1979

Armed Services Vocational Aptitude Battery
- arithmetic reasoning
- word knowledge
- paragraph comprehension
- mathematical knowledge
- coding speed

Cognitive

Rotter Locus of Control Scale
- measures the degree of control individuals feel they possess over their life

Non-Cognitive

Rosenberg Self-Esteem Scale
- measures perceptions of self worth
We Are In The Assessment Business

Patient/System Outcomes?

Medical Knowledge

Professionalism

Practice-Based Learning

Systems-Based Practice

Communication Skills

Patient Care

Patient Engagement

Internal Locus of Control

Reflective

Resilience/Grit

Smart

Synthesis

Engagement

Locus of Control

Reflection

Resilience/Grit

Professionalism

Medical Knowledge
We Are In The Assessment Business
We Are In The Assessment Business

Assess Recruits

Little Leagues

High School

Pre-Med

Medical School

Assess Players

College/Minor Leagues

Residency

Fellowship

MLB/NFL

Practice

Assess Applicants

Assess Trainees
We Are In The Assessment Business

Assess Applicants

Assess Recruits

Assess Players

Shared Incentives

Little Leagues

High School

College/Minor Leagues

MLB/NFL

Pre-Med

Medical School

Residency

Fellowship

Practice

Weak/No Incentives

Assess Practitioners
We Are In The Assessment Business

Assess Recruits
- Game Stats
- Video Tape
- Scouting Visits
- Interviews

Assess Players
- Game Stats
- Wins/Losses
- Video Tape
- Analysis (press)

Little Leagues | High School | College/Minor Leagues | MLB/NFL
---|---|---|---
Pre-Med | Medical School | Residency | Fellowship | Practice
We Are In The Assessment Business

Assess Recruits
- Game Stats
- Video Tape
- Scouting Visits
- Interviews

Assess Players
- Game Stats
- Wins/Losses
- Video Tape
- Analysis (press)

Little Leagues
High School
College/Minor Leagues
MLB/NFL

Pre-Med
Medical School
Residency
Fellowship
Practice

Assess Applicants
- USMLE I
- Clerkship Grades
- Letters of Rec
- MSPE
- Interviews

Assess Trainees
- Faulty Framing
- Inference
- Halo Effect
- Range Restriction
- “Gut Feeling”
We Are In The Assessment Business

Assess Recruits
Direct Observation
Measurement of Outcomes

Assess Players
Direct Observation
Measurement of Outcomes

Little Leagues
High School
College/Minor Leagues
MLB/NFL

Pre-Med
Medical School
Residency
Fellowship
Practice

Assess Applicants
Indirect Observation of Deeply Flawed, Obscured and Subjective Data

Assess Trainees
Faulty Framing
Inference
Halo Effect
Range Restriction
“Gut Feeling”
We Are In The Assessment Business

- **Assess Recruits**
  - Wins/Losses
  - Video Tape
  - Analysis (press)
  - Combine Results

- **Assess Players**
  - Game Stats
  - Wins/Losses
  - Video Tape
  - Analysis (press)

---

- **Little Leagues**
- **High School**
- **College/Minor Leagues**
- **MLB/NFL**

- **Pre-Med**
- **Medical School**
- **Residency**
- **Fellowship**
- **Practice**
We Are In The Assessment Business

Assess Recruits
- Game Stats
- Wins/Losses
- Video Tape
- Analysis (press)

Assess Players
- Game Stats
- Wins/Losses
- Video Tape
- Analysis (press)

Little Leagues
High School
College/Minor Leagues
MLB/NFL
Pre-Med
Medical School
Residency
Fellowship
Practice

Indirect Observation of Deeply Flawed, Obscured and Subjective Data

Assess Applicants

Assess Practitioners

OPPE/FPPE
We Are In The Assessment Business

What in the training files predicts performance in practice?

- Disciplinary action by medical boards
  - strongly associated with unprofessional behavior in medical school
  - Especially if:
    - described as irresponsible
    - having diminished ability to improve their behavior

- Poor performance on behavioral and cognitive measures during residency
  - associated with greater risk for state licensing board at every point on a performance continuum.


We Are In The Assessment Business

What in the medical school file predicts performance in residency?

**Pediatrics:** “We conclude that **no objective or subjective selection factors** can reliably predict the level of residency performance.”


**Neurology:** There was **no significant relationship** between Step I scores and resident quality, though Step I scores correlated significantly with ITE scores. Resident quality was not correlated with ranking of their medical school.

What in the medical school file predicts performance in residency?

**Otolaryngology:** Many of the application factors typically used during otolaryngology residency candidate selection may not be predictive of future capabilities as a clinician. Prior excellence in a team sport may suggest continued success in the health care team.


**Obstetrics and Gynecology:** In a program designed to train academic obstetrician-gynecologists, objective data from medical students' applications did not correlate with successful resident performance in our obstetrics-gynecology residency program.

What in the medical school file predicts performance in residency?

**Radiology:** Dean's letters, letters of recommendation, AOA selection during the senior year, and medical school prestige do not appear to predict future resident performance.


**Dermatology:** No one medical student factor can be used to predict performance in residency. There is a need for a more consistent and systematic approach to determining predictors of success in residency.

Pre-Med | Medical School | Residency | Fellowship | Practice

What in the medical school file predicts performance in residency?

**Internal Medicine:** Higher USMLE scores were associated with higher than US median ITE scores. **None of the other characteristics** including visa status were associated with the outcomes.


**Internal Medicine:** Professionalism scores were positively associated with mean strength of comparative statements in recommendation letters (0 = no comparative statement, 1 = equal to peers, 2 = top 20%, 3 = top 10% or "best"). **No other associations** between ERAS application variables and professionalism scores were found.

What in the medical school file predicts performance in residency?

**AAMC: The Medical Student Performance Evaluation**
- include a summative assessment, based upon the school’s evaluation system, of the student’s comparative performance in medical school, relative to his/her peers

We Are In The Assessment Business

What in the medical school file predicts performance in residency?

Little in the medical school file predicts performance in residency.

**Commission?**  
10%

**Omission?**  
90%
Manning
#1 Draft Pick 1998

Bowie
#2 Draft Pick 1984

Brown
#1 Draft Pick 2000

Leaf
#2 Draft Pick 1998

Jordan
#3 Draft Pick 1984

Brady
#199 Draft Pick 2000
Traditionally, most assessment is external assessment of learning.
Ideal assessment: balanced assessment FOR learning.
Feedback is ignored in assessment regimes with a summative orientation.

Learners will primarily orient themselves on passing the test instead of feedback.


External assessment

- We use different **frames of reference**
- We use high levels of **inference** during the observation process
- We struggle with **converting judgments** into **numerical** ratings
- We let factors **external** to resident performance influence ratings
- We have **poor inter-rater** reliability


Faculty Ratings of Trainees

The End Result…

Most traditional evaluations are marked by:

- Bias
- Leniency
- Range Restriction
- Halo Effect


Ratings of Trainees

- 9.00
- 8.00
- 7.00
- 5.49
- 9.58
What is my job as Program Director?

Determine who is best?

Determine who is good enough?

Ensure trainees are improving and graduates are safe for practice...

...and producing good outcomes.
What do we know about self assessment?
The Dunning–Kruger Effect: On Being Ignorant of One’s Own Ignorance

David Dunning
1. Two Assertions About Ignorance
   1.1. Ignorance is prevalent in everyday life
   1.2. Ignorance is often invisible to those to suffer from it

3. The Dunning–Kruger Effect
   3.2. The double burden of incompetence
   3.3. Expertise and metacognitive judgment

“The ability to monitor one’s approach to problem solving—to be metacognitive—is an important aspect of the expert’s competence. Experts step back from their first, oversimplistic interpretation of a problem or situation and question their own knowledge.”

Advances in Experimental Social Psychology, Volume 44
Humor Reasoning Ability Grammar

Social Psychology: Dunning Kruger Effect

Cognitive Psychology: Illusion of Explanatory Depth

Organizational Psychology: Normalization of Deviance

Metacognitive deficit = Part of the Human Experience


The preponderance of evidence suggests that physicians have a limited ability to accurately self-assess.
We Are In The Assessment Business

Pieter Bruegel the Elder, 1568
5.3. The paradox of gaining expertise

“One way to train incompetent people to recognize their incompetence is to rid them of that incompetence.”
Part 2
High Quality Assessment

1. Rating scales should be aligned with clinician-assessor priorities

2. Global judgments of a given activity should be sought as these are often better at assessing performance than objective observation checklists

3. Assessment should focus on competencies central to the activity observed

4. Chosen assessors should be the ones best placed to observe performance

1. Develop a master plan for assessment
2. Develop examination regulations that promote feedback orientation
3. Adopt a robust system for collecting information
4. Assure that every low-stakes assessment provides meaningful feedback for learning
5. Provide mentoring to learners
6. Ensure trustworthy decision-making
7. Organize intermediate decision-making assessments
8. Encourage and facilitate personalized remediation
9. Monitor and evaluate the learning effect of the program and adapt
10. Use the assessment process information for curriculum evaluation
11. Promote continuous interaction between the stakeholders
12. Develop a strategy for implementation
High Quality Assessment

• “High-quality feedback should be the prime purpose of any individual data point.”


• “Every data point should be optimized for learning.”

2. van der Vleuten CP, A Programmatic View on Assessment, ACGME meeting, 2015
Ambulatory Long Block
Medical student attitudes towards caring for elderly and chronically ill decline during training.


Griffith CH III, Wilson J F. *The loss of student idealism in the 3rd-year clinical clerkships.* Eval Health Prof. 2001;24(1):61-71
Long Block
Master the continuous healing relationship

• The paradox:
  – Most students want continuity of care
  – But... not in the current practice environment


Master the continuous healing relationship

Reduce conflict between inpatient and outpatient medicine
## Long Block

**Master the continuous healing relationship**

### Example Weekly Schedule

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00 am</td>
<td>Elective</td>
<td>Elective</td>
<td>AME</td>
<td>Elective</td>
<td>Elective</td>
</tr>
<tr>
<td>8:30 am</td>
<td>Practice</td>
<td>Practice</td>
<td>Elective</td>
<td>Elective</td>
<td>Elective</td>
</tr>
<tr>
<td>1:00 pm</td>
<td>Elective</td>
<td>Elective</td>
<td>Elective</td>
<td>Practice</td>
<td>Elective</td>
</tr>
</tbody>
</table>

- Residents see patients 3 half days per week on average
- Part of one morning per week reserved for education (Ambulatory topics)
- The rest of the time is elective
Patient Focus
Weekly team meetings focused on improving patient care

Team = most transformative feature of the practice.

Long Block
Master the continuous healing relationship

Therapeutic Dyad

Set an Agenda
Ask – Tell – Ask
Close the Loop

Teach Back Technique

Education and Training
Residents and nursing trained together in quality improvement techniques

- Train the residents and staff together in quality improvement concepts

Jumpstart “In the water”

Long Block
Master the continuous healing relationship

<table>
<thead>
<tr>
<th></th>
<th>Time for Learning</th>
<th>Ability to Focus in Clinic without Interruption</th>
<th>Ability to Balance Ward/Inpatient Duties on Clinic Days</th>
<th>Personal Reward from Work</th>
<th>Relationship with Patients</th>
<th>Ownership/Personal Responsibility for Patient Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before LB (n = 18/21)</td>
<td>2.94</td>
<td>3.44</td>
<td>3.00</td>
<td>3.33</td>
<td>4.06</td>
<td>3.72</td>
</tr>
<tr>
<td>After LB (n = 21/21)</td>
<td>4.44</td>
<td>4.56</td>
<td>4.59</td>
<td>4.44</td>
<td>4.72</td>
<td>4.78</td>
</tr>
<tr>
<td>paired t test</td>
<td>0.0004</td>
<td>0.0057</td>
<td>0.0018</td>
<td>0.0042</td>
<td>0.0001</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

A Multiple Choice Testing Program Coupled with a Year-long Elective Experience Is Associated with Improved Performance on the Internal Medicine In-Training Examination. J Gen Intern Med 26(11):1253–7
4-year experience

Highest Scores:

• *Demonstrate empathetic and caring behavior to the patient and patient’s family*

• *Respect, compassion, and integrity for my patients*
### 360 Evaluation Attending/Peer/Staff

<table>
<thead>
<tr>
<th>Patient Care</th>
<th>Team Work</th>
<th>Professionalism</th>
<th>Efficiency</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Score</td>
<td>4.42</td>
<td>4.85</td>
<td>4.82</td>
<td>4.17</td>
</tr>
<tr>
<td>Rank in Class</td>
<td>13</td>
<td>1</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Class Average</td>
<td>4.40</td>
<td>4.46</td>
<td>4.54</td>
<td>4.20</td>
</tr>
<tr>
<td>Class Range</td>
<td>3.89-4.74</td>
<td>3.65-4.85</td>
<td>3.80-4.82</td>
<td>3.34-4.67</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.23</td>
<td>0.29</td>
<td>0.24</td>
<td>0.32</td>
</tr>
</tbody>
</table>

### Patient Evaluations

<table>
<thead>
<tr>
<th>Your Score</th>
<th>Your Rank</th>
<th>Class Average</th>
<th>Range</th>
<th>Stand. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Evaluations Total</td>
<td>29</td>
<td>5</td>
<td>26.87</td>
<td>16-49</td>
</tr>
<tr>
<td>Average Visits per Patient</td>
<td>2.72</td>
<td>17</td>
<td>2.97</td>
<td>2.75-3.875</td>
</tr>
<tr>
<td>Physician Explains</td>
<td>5.55</td>
<td>11</td>
<td>5.52</td>
<td>4.75-5.96</td>
</tr>
<tr>
<td>Physician Listens</td>
<td>5.72</td>
<td>9</td>
<td>5.62</td>
<td>5.00-5.92</td>
</tr>
<tr>
<td>Physician Gives Instructions</td>
<td>3.61</td>
<td>11</td>
<td>5.58</td>
<td>5.06-6.00</td>
</tr>
<tr>
<td>Physician Knows History</td>
<td>5.52</td>
<td>8</td>
<td>5.44</td>
<td>4.90-5.89</td>
</tr>
<tr>
<td>Physician Respects Patient</td>
<td>5.79</td>
<td>8</td>
<td>5.67</td>
<td>5.20-6.00</td>
</tr>
<tr>
<td>Physician Spends Enough Time</td>
<td>5.72</td>
<td>4</td>
<td>5.53</td>
<td>4.81-5.93</td>
</tr>
<tr>
<td>Physician Calls With Results</td>
<td>5.19</td>
<td>9</td>
<td>5.04</td>
<td>4.14-5.71</td>
</tr>
<tr>
<td>*Rate on Scale of 0-10</td>
<td>8.93</td>
<td>16</td>
<td>8.99</td>
<td>7.89-9.70</td>
</tr>
<tr>
<td>Would Recommend To Others (scale 1-4)</td>
<td>3.76</td>
<td>12</td>
<td>3.75</td>
<td>3.42-3.96</td>
</tr>
</tbody>
</table>

### Competency Based Self Evaluation (your rating of yourself compared to how your peers self-rated)

<table>
<thead>
<tr>
<th>Patient Care</th>
<th>Medical Knowledge</th>
<th>Problem Based Learning</th>
<th>Communication</th>
<th>Professionalism</th>
<th>System Based Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your November Score Self Score</td>
<td>3.27</td>
<td>2.40</td>
<td>2.60</td>
<td>3.00</td>
<td>3.27</td>
</tr>
<tr>
<td>Rank</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Class Average</td>
<td>3.45</td>
<td>2.91</td>
<td>3.00</td>
<td>2.98</td>
<td>3.54</td>
</tr>
<tr>
<td>Range (Scale 1-5)</td>
<td>2.82-4.18</td>
<td>2.4-3.8</td>
<td>1.00-4.00</td>
<td>2.52-4.14</td>
<td>2.73-3.30</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.35</td>
<td>0.47</td>
<td>0.64</td>
<td>0.51</td>
<td>0.47</td>
</tr>
<tr>
<td>Your April Score Self Score</td>
<td>2.82</td>
<td>2.00</td>
<td>2.17</td>
<td>2.17</td>
<td>2.55</td>
</tr>
<tr>
<td>Rank</td>
<td>19</td>
<td>21</td>
<td>18</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Class Average</td>
<td>3.39</td>
<td>2.97</td>
<td>2.78</td>
<td>3.45</td>
<td>3.57</td>
</tr>
<tr>
<td>Range (Scale 1-5)</td>
<td>2.73-4.18</td>
<td>1.64-6.00</td>
<td>1.50-3.67</td>
<td>2.50-4.50</td>
<td>3.00-4.82</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.45</td>
<td>0.64</td>
<td>0.58</td>
<td>0.64</td>
<td>0.57</td>
</tr>
</tbody>
</table>

### Testing

<table>
<thead>
<tr>
<th>Score</th>
<th>Rank</th>
<th>Average</th>
<th>Range</th>
<th>Stand. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITE1</td>
<td>60</td>
<td>8</td>
<td>58.59</td>
<td>51.75</td>
</tr>
<tr>
<td>ITE2</td>
<td>61</td>
<td>14</td>
<td>65.57</td>
<td>56.80</td>
</tr>
<tr>
<td>GIM 1</td>
<td>53.33</td>
<td>12</td>
<td>56.91</td>
<td>57.78-91.11</td>
</tr>
<tr>
<td>Cardiology</td>
<td>68.89</td>
<td>6</td>
<td>63.57</td>
<td>40.00-82.22</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>64.44</td>
<td>10</td>
<td>64.75</td>
<td>44.44-82.22</td>
</tr>
<tr>
<td>GI Liver</td>
<td>68.89</td>
<td>11</td>
<td>68.99</td>
<td>55.56-84.44</td>
</tr>
<tr>
<td>Mathis Testing Overall</td>
<td>63.89</td>
<td>12</td>
<td>62.85</td>
<td>55.00-82.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent</th>
<th>Rank</th>
<th>Average</th>
<th>Range</th>
<th>Stand. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAP Attendance % Expected Per Schedule</td>
<td>100%</td>
<td>1</td>
<td>83%</td>
<td>44-100%</td>
</tr>
<tr>
<td>Open Notes January to April 2015</td>
<td>7</td>
<td>11</td>
<td>13</td>
<td>1-43</td>
</tr>
</tbody>
</table>

### Resident Assessment

**Part 1: Numerical Data**

---

<table>
<thead>
<tr>
<th>360 Evaluation Attending/Peer/Staff</th>
<th>Patient Care</th>
<th>Team Work</th>
<th>Professionalism</th>
<th>Efficiency</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Score</td>
<td>4.73</td>
<td>4.88</td>
<td>4.80</td>
<td>3.69</td>
<td>4.52</td>
</tr>
<tr>
<td>Rank in Class</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>Class Average</td>
<td>4.36</td>
<td>4.35</td>
<td>4.44</td>
<td>4.20</td>
<td>4.34</td>
</tr>
<tr>
<td>Class Range</td>
<td>3.57-4.73</td>
<td>3.47-4.88</td>
<td>3.64-4.90</td>
<td>3.47-4.55</td>
<td>3.61-4.67</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.29</td>
<td>0.32</td>
<td>0.30</td>
<td>0.30</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient Evaluations</th>
<th>April</th>
<th>October</th>
<th>October Rank</th>
<th>Class Average</th>
<th>Range</th>
<th>Stand. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Evaluations Total</td>
<td>24</td>
<td>27</td>
<td>4</td>
<td>19.78</td>
<td>4.00-40</td>
<td>8.87</td>
</tr>
<tr>
<td>Average Visits per Patient</td>
<td>3.08</td>
<td>4.33</td>
<td>4</td>
<td>3.80</td>
<td>2.68-4.70</td>
<td>0.54</td>
</tr>
<tr>
<td>Physician Explains</td>
<td>5.92</td>
<td>5.85</td>
<td>5</td>
<td>5.74</td>
<td>5.35-5.97</td>
<td>0.20</td>
</tr>
<tr>
<td>Physician Listens</td>
<td>5.84</td>
<td>5.89</td>
<td>9</td>
<td>5.81</td>
<td>5.41-6.00</td>
<td>0.18</td>
</tr>
<tr>
<td>Physician Gives Instructions</td>
<td>6.00</td>
<td>5.89</td>
<td>6</td>
<td>5.81</td>
<td>5.40-5.95</td>
<td>0.15</td>
</tr>
<tr>
<td>Physician Knows History</td>
<td>5.72</td>
<td>5.70</td>
<td>8</td>
<td>5.69</td>
<td>5.24-5.91</td>
<td>0.22</td>
</tr>
<tr>
<td>Physician Respects Patient</td>
<td>5.92</td>
<td>5.89</td>
<td>7</td>
<td>5.82</td>
<td>5.24-6.00</td>
<td>0.17</td>
</tr>
<tr>
<td>Physician Spends Enough Time</td>
<td>5.80</td>
<td>5.93</td>
<td>4</td>
<td>5.76</td>
<td>5.32-6.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Physician Calls With Results</td>
<td>5.30</td>
<td>5.56</td>
<td>6</td>
<td>5.30</td>
<td>4.6-5.71</td>
<td>0.53</td>
</tr>
<tr>
<td>*Rate on Scale of 0-10</td>
<td>9.44</td>
<td>9.70</td>
<td>3</td>
<td>9.46</td>
<td>8.71-9.82</td>
<td>0.34</td>
</tr>
<tr>
<td>Would Recommend To Others (scale 1-4)</td>
<td>3.84</td>
<td>3.93</td>
<td>3</td>
<td>3.86</td>
<td>3.56-4.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Average/Overall Rank</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Self Evaluation</th>
<th>Patient Care</th>
<th>Medical Knowledge</th>
<th>Problem Based Learning</th>
<th>Communication</th>
<th>Professionalism</th>
<th>Systems Based Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Your November Score</strong></td>
<td>3.36</td>
<td>2.40</td>
<td>2.20</td>
<td>3.29</td>
<td>3.09</td>
<td>2.17</td>
</tr>
<tr>
<td>Rank</td>
<td>12</td>
<td>17</td>
<td>21</td>
<td>12</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>Class Average</td>
<td>3.45</td>
<td>2.91</td>
<td>3.00</td>
<td>3.28</td>
<td>3.54</td>
<td>2.92</td>
</tr>
<tr>
<td>Range (Scale 1-5)</td>
<td>2.82-4.18</td>
<td>2.4-3.8</td>
<td>1.00-4.00</td>
<td>2.57-4.14</td>
<td>2.73-4.30</td>
<td>2.33-3.67</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.35</td>
<td>0.47</td>
<td>0.64</td>
<td>0.51</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>Your April Score</strong></td>
<td>3.91</td>
<td>3.20</td>
<td>3.00</td>
<td>4.00</td>
<td>3.64</td>
<td>3.00</td>
</tr>
<tr>
<td>Rank</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Class Average</td>
<td>3.39</td>
<td>2.97</td>
<td>2.78</td>
<td>3.45</td>
<td>3.57</td>
<td>2.82</td>
</tr>
<tr>
<td>Range (Scale 1-5)</td>
<td>2.73-4.18</td>
<td>1.6-4.00</td>
<td>1.50-3.67</td>
<td>2.50-4.50</td>
<td>3.00-4.82</td>
<td>1.83-3.83</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.45</td>
<td>0.64</td>
<td>0.58</td>
<td>0.64</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Your October Score</strong></td>
<td>3.82</td>
<td>3.40</td>
<td>3.33</td>
<td>4.67</td>
<td>4.18</td>
<td>3.50</td>
</tr>
<tr>
<td>Rank</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Class Average</td>
<td>3.53</td>
<td>3.12</td>
<td>3.03</td>
<td>3.50</td>
<td>3.52</td>
<td>2.98</td>
</tr>
<tr>
<td>Range (Scale 1-5)</td>
<td>3.00-4.55</td>
<td>2.40-4.00</td>
<td>1.67-4.17</td>
<td>2.83-4.67</td>
<td>2.91-3.00</td>
<td>2.17-4.00</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.38</td>
<td>0.47</td>
<td>0.67</td>
<td>0.55</td>
<td>0.52</td>
<td>0.51</td>
</tr>
</tbody>
</table>

1 = Area where I know that I need improvement
2 = Area where I think that I need improvement
3 = Area where I think that I perform adequately
4 = Area where I think that I am above average
5 = Area where I think that I am very skilled
## Testing

<table>
<thead>
<tr>
<th>Testing</th>
<th>Score</th>
<th>Rank</th>
<th>Average</th>
<th>Range</th>
<th>Stand. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITE1</td>
<td>57</td>
<td>11</td>
<td>58.59</td>
<td>51-75</td>
<td>5.88</td>
</tr>
<tr>
<td>ITE2</td>
<td>66</td>
<td>12</td>
<td>65.57</td>
<td>56-80</td>
<td>7.56</td>
</tr>
<tr>
<td>GIM 1</td>
<td>53.33</td>
<td>12</td>
<td>56.91</td>
<td>37.78-91.11</td>
<td>12.04</td>
</tr>
<tr>
<td>Cardiology</td>
<td>48.89</td>
<td>20</td>
<td>63.57</td>
<td>40.00-82.22</td>
<td>10.71</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>64.44</td>
<td>10</td>
<td>64.75</td>
<td>44.44-82.22</td>
<td>9.28</td>
</tr>
<tr>
<td>GI Liver</td>
<td>68.89</td>
<td>11</td>
<td>68.99</td>
<td>55.56-84.44</td>
<td>7.80</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>71.11</td>
<td>3</td>
<td>62.12</td>
<td>42.22-84.44</td>
<td>8.46</td>
</tr>
<tr>
<td>Hem/Onc</td>
<td>66.67</td>
<td>5</td>
<td>61.85</td>
<td>46.47-80.00</td>
<td>8.80</td>
</tr>
<tr>
<td>Nephrology</td>
<td>60.00</td>
<td>13</td>
<td>66.76</td>
<td>45.00-90.00</td>
<td>11.04</td>
</tr>
<tr>
<td>ID</td>
<td>55.56</td>
<td>16</td>
<td>68.50</td>
<td>48.89-93.33</td>
<td>10.00</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>64.44</td>
<td>13</td>
<td>71.35</td>
<td>57.58-86.87</td>
<td>8.15</td>
</tr>
<tr>
<td>First Half Testing Average</td>
<td>58.89</td>
<td>15</td>
<td>63.55</td>
<td>51.67-82.78</td>
<td>7.48</td>
</tr>
<tr>
<td>Second Half Testing Average</td>
<td>63.56</td>
<td>13</td>
<td>66.12</td>
<td>54.06-86.94</td>
<td>7.01</td>
</tr>
<tr>
<td>Overall Testing Average</td>
<td>61.48</td>
<td>14</td>
<td>64.45</td>
<td>53.27-84.09</td>
<td>6.53</td>
</tr>
</tbody>
</table>

## Citizenship

<table>
<thead>
<tr>
<th>Citizenship</th>
<th>Open Encounters</th>
<th>Tests Taken</th>
<th>Missing Patient Evaluations</th>
<th>Percent Refils in 48</th>
<th>Evaluation On Time</th>
<th>Overall Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Score</td>
<td>15</td>
<td>9</td>
<td>0</td>
<td>92.64%</td>
<td>Y</td>
<td>10</td>
</tr>
<tr>
<td>Average</td>
<td>10.22</td>
<td>8.00</td>
<td>6.57</td>
<td>98.18%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0-108</td>
<td>5-9</td>
<td>0-21</td>
<td>87%-124%</td>
<td>18 Yes, 5 No</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>23.25</td>
<td>0.95</td>
<td>7.01</td>
<td>7.68%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mathis BR, Warm EJ, Schauer DP, Holmboe E, Rouan, GW, **A Multiple Choice Testing Program Coupled with a Year-long Elective Experience Is Associated with Improved Performance on the Internal Medicine In-Training Examination.** J Gen Intern Med 26(11):1253–7

### Work Intensity (4/1/14 to 10/1/15)

<table>
<thead>
<tr>
<th>Work Units</th>
<th>April</th>
<th>October</th>
<th>October Rank</th>
<th>Average</th>
<th>Range</th>
<th>Stand Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>157</td>
<td>187</td>
<td>5</td>
<td>68-241</td>
<td>135</td>
<td>48</td>
</tr>
<tr>
<td>Prescription Requests</td>
<td>104</td>
<td>99</td>
<td>22</td>
<td>75-185</td>
<td>138.74</td>
<td>29</td>
</tr>
<tr>
<td>Patient Calls</td>
<td>131</td>
<td>170</td>
<td>1</td>
<td>80-170</td>
<td>116.61</td>
<td>22</td>
</tr>
<tr>
<td>Staff Messages</td>
<td>67</td>
<td>72</td>
<td>4</td>
<td>21-149</td>
<td>53.57</td>
<td>27</td>
</tr>
<tr>
<td>MyChart Messages</td>
<td>15</td>
<td>25</td>
<td>2</td>
<td>0-29</td>
<td>9.48</td>
<td>8</td>
</tr>
<tr>
<td>Total Work Units</td>
<td>474</td>
<td>553</td>
<td>6</td>
<td>449-524</td>
<td>453.39</td>
<td>82</td>
</tr>
<tr>
<td>Patients Number</td>
<td>157</td>
<td>179</td>
<td>1</td>
<td>120-179</td>
<td>151.78</td>
<td>13</td>
</tr>
</tbody>
</table>

#### Ratio of Work Units to Number of Patients

<table>
<thead>
<tr>
<th>Work Units</th>
<th>April</th>
<th>October</th>
<th>October Rank</th>
<th>Average</th>
<th>Range</th>
<th>Stand Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>1.00</td>
<td>1.04</td>
<td>5</td>
<td>0.47-1.52</td>
<td>0.89</td>
<td>0.29</td>
</tr>
<tr>
<td>Prescription Requests</td>
<td>0.66</td>
<td>0.55</td>
<td>22</td>
<td>0.48-1.29</td>
<td>0.92</td>
<td>0.21</td>
</tr>
<tr>
<td>Patient Calls</td>
<td>0.83</td>
<td>0.95</td>
<td>3</td>
<td>0.50-1.03</td>
<td>0.77</td>
<td>0.14</td>
</tr>
<tr>
<td>Staff Messages</td>
<td>0.43</td>
<td>0.40</td>
<td>7</td>
<td>0.16-0.93</td>
<td>0.35</td>
<td>0.17</td>
</tr>
<tr>
<td>MyChart Messages</td>
<td>0.10</td>
<td>0.14</td>
<td>2</td>
<td>0.0-0.18</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Total Work Units</td>
<td>3.02</td>
<td>3.09</td>
<td>11</td>
<td>2.08-3.83</td>
<td>2.99</td>
<td>0.50</td>
</tr>
</tbody>
</table>

### Quality (see also detailed report)

<table>
<thead>
<tr>
<th></th>
<th>Prevention</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>Change</td>
</tr>
<tr>
<td>Team Average Prevention</td>
<td>77.10%</td>
<td>5.50%</td>
</tr>
<tr>
<td></td>
<td>Overall Rank</td>
<td>Change Rank</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Diabetes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>Change</td>
</tr>
<tr>
<td>Team Average Process</td>
<td>80.40%</td>
<td>0.10%</td>
</tr>
<tr>
<td>Team Average Outcome</td>
<td>70.90%</td>
<td>14.80%</td>
</tr>
<tr>
<td>Team Average Overall</td>
<td>70.90%</td>
<td>3.20%</td>
</tr>
<tr>
<td></td>
<td>Overall Rank</td>
<td>Change Rank</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Median Discharge Followup</td>
<td>19.4</td>
<td>19.8</td>
</tr>
<tr>
<td>Blood Pressure 150/90 Age 60+</td>
<td>80.20%</td>
<td>81.30%</td>
</tr>
<tr>
<td>Diastolic BP 90 Age &lt; 60</td>
<td>78.90%</td>
<td>86.70%</td>
</tr>
<tr>
<td>Diastolic BP 90 CKD III</td>
<td>68%</td>
<td>72.70%</td>
</tr>
<tr>
<td>Mammogram</td>
<td>72.40%</td>
<td>81.30%</td>
</tr>
<tr>
<td>Dexascan</td>
<td>80%</td>
<td>82.90%</td>
</tr>
<tr>
<td>Pap Smear</td>
<td>59.60%</td>
<td>68.70%</td>
</tr>
<tr>
<td>Colonoscopy AF4Q</td>
<td>76.80%</td>
<td>80.80%</td>
</tr>
<tr>
<td>HIV</td>
<td>63.50%</td>
<td>74.50%</td>
</tr>
<tr>
<td>Pneumovax</td>
<td>95%</td>
<td>96.30%</td>
</tr>
<tr>
<td>Influenza</td>
<td>40.80%</td>
<td>61.20%</td>
</tr>
<tr>
<td>Tetanus</td>
<td>87.10%</td>
<td>92.50%</td>
</tr>
<tr>
<td>Zoster</td>
<td>85.10%</td>
<td>91.20%</td>
</tr>
<tr>
<td>Hep C</td>
<td>46.50%</td>
<td>56.30%</td>
</tr>
<tr>
<td>Dexascan</td>
<td>80%</td>
<td>82.90%</td>
</tr>
<tr>
<td>Hospital Discharge F/U</td>
<td>59.30%</td>
<td>59.70%</td>
</tr>
<tr>
<td>ACE Inhibitor/CHF</td>
<td>66.70%</td>
<td>69.20%</td>
</tr>
<tr>
<td>URI w/o Antibiotic Usage</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>AAA Screening 65-75 Male Smokers</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Prevention Summary</td>
<td>75.60%</td>
<td>81.50%</td>
</tr>
<tr>
<td>Rank Change</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DM Process Summary</td>
<td>82.90%</td>
<td>83.80%</td>
</tr>
<tr>
<td>DM Outcome Summary</td>
<td>58.40%</td>
<td>58.10%</td>
</tr>
<tr>
<td>DM Overall Summary</td>
<td>69%</td>
<td>69.70%</td>
</tr>
<tr>
<td>Change Rank</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

**Personal Reports**
<table>
<thead>
<tr>
<th>Seligman Scale</th>
<th>Positive Emotion</th>
<th>Engagement in work</th>
<th>Positive relationships with patients and colleagues</th>
<th>Meaning in your work</th>
<th>Sense of accomplishment</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>November</td>
<td>75</td>
<td>75</td>
<td>85</td>
<td>80</td>
<td>80</td>
<td>11</td>
</tr>
<tr>
<td>November Class Average</td>
<td>71</td>
<td>74</td>
<td>82</td>
<td>80</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>40-100</td>
<td>40-100</td>
<td>50-100</td>
<td>41-100</td>
<td>24-100</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>19.52</td>
<td>16.21</td>
<td>16.23</td>
<td>17.33</td>
<td>21.04</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>85</td>
<td>90</td>
<td>85</td>
<td>85</td>
<td>89</td>
<td>10</td>
</tr>
<tr>
<td>April Class Average</td>
<td>84</td>
<td>86</td>
<td>87</td>
<td>85</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>70-100</td>
<td>75-100</td>
<td>65-100</td>
<td>80-100</td>
<td>60-90</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.99</td>
<td>10.11</td>
<td>10.53</td>
<td>7.18</td>
<td>10.41</td>
<td></td>
</tr>
<tr>
<td>October</td>
<td>90</td>
<td>93</td>
<td>88</td>
<td>90</td>
<td>95</td>
<td>9</td>
</tr>
<tr>
<td>October Class Average</td>
<td>87</td>
<td>85</td>
<td>92</td>
<td>88</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>65-100</td>
<td>65-95</td>
<td>70-100</td>
<td>70-100</td>
<td>75-100</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>9.69</td>
<td>8.23</td>
<td>7.18</td>
<td>7.96</td>
<td>7.69</td>
<td></td>
</tr>
</tbody>
</table>
360: Strengths

- Consummate professional and humanist; this is the person you want caring for you because he cares
- Great responses time for questions with patients.
- I can tell he does a good chart review. Seems to genuinely care for his patients and their outcomes.
- Always a positive attitude and very helpful
- Very thorough
- Dr. X is incredibly professional. He puts others above himself at all times. He is receptive to feedback. He really thinks about a patient's problem prior to speaking.
- Calm presence in clinic
- Chart prep, huddles, professional
- Willing to go the extra step for his patients
- All around good doctor
- Through, complete, thoughtful
- Always willing to help others
- Incredible influence amongst clinic
- Great with faculty, peers and staff. Always willing to lend a helping hand
- Often see him on clinic off days catching up on mailbox work and returning patient calls
- ...many more
360: Opportunities for Improvement

- Improve knowledge base
- Dr. X sometimes lacks confidence. You know more than you think you do!
- Efficiency
- Develop plan for assessments based on medical knowledge/evidence and take a stand on decision making. Dr. X can be passive in making decision to titrate insulin or adjust blood pressure regimen, especially when the patient sees consultants like endocrine for diabetes or heart failure clinic.
- Just being efficient and getting as much done as possible on clinic days
- Do not have suggestions for improvements
- Keep working hard.
- Efficiency, setting goals and keeping on task, preping charts ahead of time huddle with nursing before clinic
- Efficiency in outpatient setting, although sometimes this is out of our control with our difficult patients.
- Share more of your ideas, you have a lot of good ones
- Could be more confident in his abilities, he is a much better doctor than he thinks he is
- …many more

Resident Assessment
Part 2: Narrative
I think he is a very good Dr. I haven't known him very long, but I really care for him a lot, thank you.
He is a really good doctor.
Thank-you wholeheartedly!
Very understanding, easy to talk to, always comes up with a solution or at least a trial of some sort. Wish he could stay longer than a year. If given the chance, would follow him to a private practice.
Dr. X is honestly the first Dr. that I have felt completely comfortable with. As the majority of doctors I have dealt with have treated me in a condescending manner. Dr. X has been patient, genuinely concerned, and respectful.
Dr. X is a concerned doctor that is easy to talk to. I love him.
He is very good!
This was only second visit. I am not comfortable giving higher marks until I get to know him better.
Dr. X seems very conscientious and I like his practice of making notes each day of each patient’s issues and what he wants to ask them about.
Dr. X listens to me really well and he is kind and warm. Also, good eye contact.
...many more
Program Director Review

Assessment for Learning
<table>
<thead>
<tr>
<th>Patient Evaluations</th>
<th>April</th>
<th>October</th>
<th>October Rank</th>
<th>Class Average</th>
<th>Range</th>
<th>Stand. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Evaluations Total</td>
<td>21</td>
<td>20</td>
<td>11</td>
<td>19.78</td>
<td>4.00-40</td>
<td>8.87</td>
</tr>
<tr>
<td>Average Visits per Patient</td>
<td>2.29</td>
<td>3.75</td>
<td>12</td>
<td>3.80</td>
<td>2.68-4.70</td>
<td>0.54</td>
</tr>
<tr>
<td>Physician Explains</td>
<td>4.70</td>
<td>5.80</td>
<td>7</td>
<td>5.74</td>
<td>5.35-5.97</td>
<td>0.20</td>
</tr>
<tr>
<td>Physician Listens</td>
<td>5.00</td>
<td>5.95</td>
<td>2</td>
<td>5.81</td>
<td>5.41-6.00</td>
<td>0.18</td>
</tr>
<tr>
<td>Physician Gives Instructions</td>
<td>5.05</td>
<td>5.95</td>
<td>1</td>
<td>5.81</td>
<td>5.40-5.95</td>
<td>0.15</td>
</tr>
<tr>
<td>Physician Knows History</td>
<td>4.90</td>
<td>5.90</td>
<td>2</td>
<td>5.69</td>
<td>5.24-5.91</td>
<td>0.22</td>
</tr>
<tr>
<td>Physician Respects Patient</td>
<td>5.20</td>
<td>5.90</td>
<td>6</td>
<td>5.82</td>
<td>5.24-6.00</td>
<td>0.17</td>
</tr>
<tr>
<td>Physician Spends Enough Time</td>
<td>4.81</td>
<td>5.95</td>
<td>2</td>
<td>5.76</td>
<td>5.32-6.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Physician Calls With Results</td>
<td>4.63</td>
<td>5.26</td>
<td>13</td>
<td>5.30</td>
<td>4.6-5.71</td>
<td>0.53</td>
</tr>
<tr>
<td>*Rate on Scale of 0-10</td>
<td>7.89</td>
<td>9.80</td>
<td>2</td>
<td>9.46</td>
<td>8.71-9.82</td>
<td>0.34</td>
</tr>
<tr>
<td>Would Recommend To Others (scale 1-4)</td>
<td>3.42</td>
<td>3.80</td>
<td>2</td>
<td>3.86</td>
<td>3.56-4.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Average/Overall Rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Long Block Assessment Data

Pros:
• Multiple data points
• Multiple data sources
• Shows trainee relative strengths and weaknesses
• Connected to patient outcomes

Cons:
• Labor intensive
• Normative data used
• No clear threshold of performance
Assessing our Assessment System

“Warm”

√ 1. Smart enough
√ 2. Synthesis skills
√ 3. Patient Engagement Skills
√ 4. Internal Locus of Control
√ 5. Reflection
√ 6. Resilience/Grit
? 7. Outcomes

Crossley and Jolly

? 1. Scales should be aligned with clinician-assessor priorities
? 2. Global judgments of a given activity sought
? 3. Focus on competencies central to the activity observed
√ 4. Chosen assessors should be the ones best placed to observe performance

van der Vleuten

X 1. Develop a master plan for assessment
√ 2. Develop examination regulations that promote feedback orientation
√ 3. Adopt a robust system for collecting information
? 4. Assure that every low-stakes assessment provides meaningful feedback for learning
√ 5. Provide mentoring to learners
√ 6. Ensure trustworthy decision-making
√ 7. Organize intermediate decision-making assessments
√ 8. Encourage and facilitate personalized remediation
√ 9. Monitor and evaluate the learning effect of the program and adapt
√ 10. Use the assessment process information for curriculum evaluation
? 11. Promote continuous interaction between the stakeholders
X 12. Develop a strategy for implementation
Can Milestones and EPAs help?
Operationalizing Ideas

• We found broad Milestones and EPAs difficult to use in direct assessment

Can you drive?

What can you drive?

How well can you drive it?
Observable Practice Activities
What residents DO.

Content-Based Observable Practice Activities (COPAs)
1. Initiate basal bolus insulin therapy
2. Manage elevated blood pressure
3. Manage COPD

Process-Based Observable Practice Activities (POPAs)
1. Minimize unfamiliar terms during patient encounters.
2. Use teach-back method

Our Rating Scale: Entrustment

1. Resident **not trusted** to perform skill even with supervision (**critical deficiencies**)
2. Resident trusted to perform skill with **direct supervision**
3. Resident trusted to perform skill with **indirect supervision**
4. Resident trusted to perform skill **independently**
5. Resident trusted to perform skill at **aspirational level**
6. Skill was **not observed** on this rotation (produces **no score**)

![Miller's Pyramid of Learner Assessment](image.png)
Rotation: Digestive Diseases

PGY-1 Content-Based OPAs (selected)

1. Write initial admission orders for gastrointestinal bleeding
6. Initiate enteral and parenteral nutrition
9. Perform paracentesis

PGY-2 Content-Based OPAs (selected)

1. Manage gastrointestinal bleeding
2. Manage pancreatitis
9. Manage complications of immunosuppressive therapy

Rotation: Initiative on Poverty Justice and Health

Content-Based OPAs (selected)

1. Identify a patient at risk for vulnerability in a clinical setting.
3. Recognize a cultural-bounded syndrome or belief
5. Manage a chronic medical condition in a resource-limited setting

Multisource: Interns and Residents

**Allied Health Professionals** Process-Based OPAs

1. Respond to pages in timely and courteous manner
2. Demonstrate commitment to relieve pain and suffering

**Residents and Students** Process-Based OPAs

1. Provide regular feedback to other members of the team.
7. Assist colleagues in the provision of duties.

How many OPAs do we have?

- >400 – always evolving

https://www.youtube.com/watch?v=t5JVRZrXBBu
Consider each skill individually and discriminate between relative strengths and weaknesses for each resident.
• **Narrative feedback** has more impact on complex skills than scores

1. van der Vleuten CP, *A Programmatic View on Assessment*. 
Structured Specific Feedback

Mid-rotation  End-rotation

Assessment for Learning
Program Director Review

Assessment for Learning
Observable Practice Activities (OPAs)

**Content-Based (COPA)**
- Reporting Milestones
- External Reporting NAS report

**Process-Based (POPA)**
- Mid/End of Rotation Feedback (formative)
- Entrustment (1-5)

**Curricular Milestones**
- Longitudinal Assessment
  - Formative > Summative

**End-of-training EPAs**
- Mapping

---

Initial Outcomes

• Annualized over 3 years each resident receives an average of:

• 3703 milestone assessments
  – 81% are from attending physicians, and 19% are from peers and allied health

• 4325 words of narrative assessment

Warm EJ, Mathis BR, Ashbrook L, Held JD, Hellmann M, Kelleher M, Kinnear B, Sall D, Tolentino J, Mueller C, Schauer D, Entrustment of Milestones in an Internal Medicine Residency, manuscript submitted for publication
By Competency

By Milestone

N = 303428

Warm EJ, Mathis BR, Ashbrook L, Held JD, Hellmann M, Kelleher M., Kinnear B, Sall D, Tolentino J, Mueller C, Schauer D, Entrustment of Milestones in an Internal Medicine Residency, manuscript submitted for publication
## Least entrusted OPAs PGY-1 Level

### 2014 End of Year Averages

<table>
<thead>
<tr>
<th>OPA -- PGY-1 LEVEL</th>
<th>Entrustment</th>
<th>#Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrate basic EKG reading skills</td>
<td>2.19</td>
<td>178</td>
</tr>
<tr>
<td>Distinguish between hepatitic, cholestatic, and infiltrative patterns of liver disease</td>
<td>2.22</td>
<td>95</td>
</tr>
<tr>
<td>Begin initial management plan for basic arrhythmias</td>
<td>2.23</td>
<td>432</td>
</tr>
<tr>
<td>Manage decompensated liver disease</td>
<td>2.24</td>
<td>184</td>
</tr>
<tr>
<td>Diagnose the cause of loss of consciousness and differentiate syncope from other etiologies.</td>
<td>2.25</td>
<td>183</td>
</tr>
<tr>
<td>Refer patients for appropriate cardiac imaging</td>
<td>2.26</td>
<td>310</td>
</tr>
<tr>
<td>Diagnose acute coronary syndrome (unstable Angina, NSTEMI, STEMI)</td>
<td>2.26</td>
<td>267</td>
</tr>
<tr>
<td>Manage heart failure (acute, chronic, systolic and diastolic)</td>
<td>2.27</td>
<td>360</td>
</tr>
<tr>
<td>Evaluate complex medical patient in a timely manner</td>
<td>2.29</td>
<td>575</td>
</tr>
<tr>
<td>Discuss indications for GI procedures</td>
<td>2.29</td>
<td>82</td>
</tr>
<tr>
<td>Manage and write ventilator orders</td>
<td>2.30</td>
<td>212</td>
</tr>
<tr>
<td>Integrate clinical evidence into decision making</td>
<td>2.33</td>
<td>896</td>
</tr>
<tr>
<td>Manage parathyroid abnormalities in the renal patient</td>
<td>2.33</td>
<td>156</td>
</tr>
<tr>
<td>Recommend acute dialysis for selected patients</td>
<td>2.33</td>
<td>72</td>
</tr>
</tbody>
</table>
Entrustment of Milestones in an Internal Medicine Residency, *manuscript submitted for publication*
Single Resident

\[ r^2 = 0.7642 \]

\[ n = 3508 \]

Warm EJ, Mathis BR, Ashbrook L, Held JD, Hellmann M, Kelleher M., Kinnear B, Sall D, Tolentino J, Mueller C, Schauer D, **Entrustment of Milestones in an Internal Medicine Residency**, *manuscript submitted for publication*
Cardiology

Gastroenterology

Nephrology

Hematology-Oncology

$R^2 = 0.0001$

$n = 6952$

$R^2 = 0.1388$

$n = 5075$

$R^2 = 0.2527$

$n = 4858$

$R^2 = 0.3651$

$n = 6300$
Average Score PGY-2 Year

- Cohort 2
- Cohort 3
- Cohort 4
PGY-1 Assessments of Senior Residents

N = 9926
Assessment Bias

- Some people are more trusting than others
- Some rotations are more trusting than others
- Context effects
  - type and amount of direct observation
  - type of physician (even within specialty)
  - types of OPAs
  - personal perspective
- Rotation order effects
- Time of year effects
- Cohort effects

- All of this has implications with 6 month reporting requirements
We needed a guidepost

- Evaluator
- Service
- Type of evaluation
- Actual question

- PGY level
- Year
- Month within the year

- Created a regression model that predicts what the score should be on average irrespective of resident

P.S. = \beta_0 + \beta_1 \times \text{Evaluator} + \beta_2 \times \text{Service} + \beta_3 \times \text{Question} + \beta_4 \times \text{Month}....
Manages patients with progressive responsibility and independence.
PC-2
Develops and achieves comprehensive management plan
Kinnear, B., Choe J., Knight C., Steinberg, K., Warm, E. Efficient reporting of milestones: Experience from the Clinical Competency Committees of two Internal Medicine residency programs. *manuscript in preparation*
Clinical Competency Committee Dashboard
## Multisource Interns and Residents Evaluated by Critical Care Nurses

Please complete the following evaluation using entrustment as your frame of reference.

At what level do you TRUST the resident to do the particular skill?

<table>
<thead>
<tr>
<th>Skill Description</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>N/A</th>
<th>Avg (Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Create a collaborative learning/teaching environment that incorporates other staff members including nursing?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
<td>3.00 (1.6)</td>
</tr>
<tr>
<td>2. Demonstrate communication regarding patient care to other healthcare team members in a timely manner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.33 (1.2)</td>
</tr>
<tr>
<td>3. Manage patients care in a safe manner.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.33 (1.2)</td>
</tr>
<tr>
<td>4. Escalate care in a critical or unstable patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.33 (1.2)</td>
</tr>
<tr>
<td>5. Comments about this resident? Please be descriptive and consider using stories to make your point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Approaches most skills and decisions with being timid. Still requires a lot of backing from more senior physicians

Not only is this resident difficult to approach from a nursing aspect she doesn't appear to want to include nursing in rounds. She doesn't take suggestions/correction well when it comes to changes in the patients status from her exam early in the day as compared to when the team rounds. Responding to a critical situation doesn't happen - critical lab values aren't responded to until another resident/fellow/attending shows concern.
Assessments (peer/allied health) n = 721
22. If you assessed the resident with a 4 (working at the attending level) or a 5 (demonstrating authoritative expertise) for any of the skills above please document why here.

If you do not feel the resident is working at these levels please correct your assessment above.

23. Overall Knowledge Assessment*

24. Standard of Work Assessment*

25. Autonomy Assessment*

26. Coping with Complexity Assessment*

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>23.</td>
</tr>
<tr>
<td>24.</td>
</tr>
<tr>
<td>25.</td>
</tr>
<tr>
<td>26.</td>
</tr>
</tbody>
</table>

AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points*

Able to calmly handle complex patient care; able to handle complications and new info without losing calm and logical approach to pt care. Took care of sick pts and incorporated new learning (eg. Swan parameters) into her care. Communication skills are a major asset.

OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to

Continued exposure to sick patients will increase her confidence and ability to handle complex situations in hospital.

Attending assessment completed at the same time as the previous nurse assessments
28. If you assessed the resident with a 4 (working at the attending level) or a 5 (demonstrating authoritative expertise) for any of the skills above please document why here.

If you do not feel the resident is working at these levels please correct your assessment above.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>29. Overall Knowledge Assessment*</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>30. Standard of Work Assessment*</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>31. Autonomy Assessment*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Coping with Complexity Assessment*</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points *

Dr. [Name] is a pleasure to work with. She is so empathetic and humanistic, always patient-centered. She is even-keeled and does not let anything phase her. She is an excellent team leader and inspires confidence in the whole team including RNs and other staff. Above all else, medicine is still exciting to her and her excitement is infectious.

OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to demonstrate your points. What is your action plan?

Dr. [Name] biggest flaw is her speed. She is deliberate and thoughtful, but sometimes this does not mesh well with a busy service. Keep working on weeding through the less important things quicker in order to improve efficiency.

Example Attending Evaluation
**Example Peer Evaluations**

**AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points**

Phenomenal senior. Just gets it. She knows when to step in and when to hold back and let you come to your own conclusion. Great senior to work with not only because of the teaching that she does daily with articles but also the amount of fun that she is to be around.

**OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to demonstrate your points. What is your action plan?**

Nothing. Great overall.

**AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points**

was always mindful of trying to make my rotation educational. She often helped me with busywork so I could focus on learning from each patient I admitted, pointed out interesting findings/clinical pearls in the workup of a patient, and made it a point to be very approachable. I felt comfortable asking her questions on things I didn't understand about my patients. She also brought food. Brain food.

**OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to demonstrate your points. What is your action plan?**

is going to be a great chief! I hope I can still seek her out for advice on being a newbie senior when she's chiefing (is that a word?) next year.
### Multisource Interns and Residents Evaluated by Ward Nurses

Please complete the following evaluation using entrustment as your frame of reference.

At what level do you TRUST the resident to do the particular skill?

<table>
<thead>
<tr>
<th></th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>N/A</th>
<th>Avg (Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Respond to pages in timely and courteous manner</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
<td>4.67 (0.5)</td>
</tr>
<tr>
<td>2. Demonstrate empathy, compassion, and a commitment to relieve pain and suffering</td>
<td>33.33%</td>
<td>66.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Was this resident receptive to the unique information/learning/skills development that nurses have to offer?</td>
<td>33.33%</td>
<td>66.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Demonstrates respectful behavior to all members of the health care team.</td>
<td>33.33%</td>
<td>66.67%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Comments about this resident? Please be descriptive and consider using stories to make your point.</td>
<td>WDE</td>
<td>is an excellent doctor! She works great with the nursing staff and responds to all pages in a timely manner! She is very thorough and has an awesome bedside manner!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Pitfalls...
Single resident assessments (faculty only)  n = 1775
17. Identify the appropriate clinical question for consultative services

18. Identify clinical questions as they emerge in patient care activities and access medical information resources

19. Perform bedside presentations that engage the patient and focus the discussion around the patient's central concerns

20. Minimize unfamiliar terms during patient encounters

21. Demonstrate shared decision-making with the patient

22. Use teach-back method with patients regarding medications and plan

23. Communicate with primary care physicians

24. Recognize the scope of his/her abilities and ask for supervision and assistance appropriately

25. Minimize unnecessary care including tests

26. Use feedback to improve performance

27. Demonstrate empathy, compassion, and a commitment to relieve pain and suffering

28. If you assessed this intern with a 3 (working at the senior resident level) or a 4 (working at the attending level) for any of the skills above please document why here.

If you do not feel this intern is working at these levels please correct your assessment above.

29. Overall Knowledge Assessment*  
30. Standard of Work Assessment*  
31. Autonomy Assessment*  
32. Coping with Complexity Assessment*  
33. AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points  
34. OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to demonstrate your points. What is your action plan?  

Good presentations

Management and plan
<table>
<thead>
<tr>
<th>Same Resident -- 1 month Prior (Gen Med)</th>
</tr>
</thead>
</table>

| 27. Demonstrate shared decision-making with the patient | 66.67% | 33.33% |
| 28. Use teach-back method with patients regarding medications and plan | 33.33% | 33.33% |
| 29. Communicate with primary care physicians | 33.33% | 33.33% |
| 30. Recognize the scope of his/her abilities and ask for supervision and assistance appropriately | 66.67% | 33.33% |
| 31. Minimize unnecessary care including tests | 33.33% | 33.33% |
| 32. Use feedback to improve performance | 33.33% | 66.67% |
| 33. Demonstrate empathy, compassion, and a commitment to relieve pain and suffering | 33.33% | 66.67% |
| 34. If you assessed this intern with a 3 (working at the senior resident level) or a 4 (working at the attending level) for any of the skills above please document why here. Performing in these skills as I would expect at her level of training Dr. is very professional and takes ownership of her patients. Her medical knowledge is excellent and she is ready to be a senior. During the 11 days that I supervised her, Dr. did a good job at the level of an intern because we did patient-centered rounding and we were quite busy, she did not have the opportunity to do any teaching for the team while I was on service, which I think she would have liked to do. |
| 36. Overall Knowledge Assessment* | 100.00% |
| 36. Standard of Work Assessment* | 33.33% | 33.33% | 33.33% |
| 37. Autonomy Assessment* | 33.33% | 33.33% | 33.33% |
| 38. Coping with Complexity Assessment* | 66.67% | 33.33% |
| 39. AREAS OF STRENGTH -- please comment on areas of clinical strength -- use vignettes and stories to demonstrate your points she was meticulous in her patient care and tried to find time to read about her patients' problems. She was quite detail oriented. She seemed to have good rapport with patients. Works hard to gain a complete understanding of her patients' problems through thorough review of past records and present history. Professional excellent documentation and very involved in patient care. Strong history taking skills and formulation of plan. |
| 39. OPPORTUNITIES FOR IMPROVEMENT -- please comment on opportunities for improvement -- use vignettes and stories to demonstrate your points. What is your action plan? Made good progress in developing independence in assessments and plans not only on admission but through the course of her patients' admission as Dr. and I discussed today we were too busy for her to be able to cope with complex situations through deliberate analysis and planning, as she would have liked to do. She appropriately asked for guidance when dosing of pain meds was unusually high in a patient with Huntington's disease, and she had many appropriate questions about her patient with PHH and her patient with HIV disease and cryptococcal meningitis. She has made a very good start in her first year of residency and if she keeps working hard learning the details, she will develop more confidence in her decision-making in the future. |
2. Did this attending discriminate between entrustment levels of different observable practice activities, or describe the discrimination level choices in the narrative?

3. Did this attending’s narrative language justify the numerical entrustment?

4. Did this attending’s narrative language describe specific strengths and/or weakness of resident performance?

5. Would this feedback be helpful to the resident?

6. Comments
OPA Assessment System

Pros:
• Multiple data points
• Multiple data sources

Cons:
• Labor intensive
• Not directly related to patient outcomes (yet)
• Passes smell test, but lacking formal validity testing at this point
### Assessing our Assessment System

#### “Warm”

<table>
<thead>
<tr>
<th></th>
<th>Crossley and Jolly</th>
<th></th>
<th>van der Vleuten</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔️</td>
<td>1. Smart enough</td>
<td>✔️</td>
<td>1. Develop a master plan for assessment</td>
</tr>
<tr>
<td>✔️</td>
<td>2. Synthesis skills</td>
<td>✔️</td>
<td>2. Develop examination regulations that promote feedback orientation</td>
</tr>
<tr>
<td>✔️</td>
<td>3. Patient Engagement Skills</td>
<td>✔️</td>
<td>3. Adopt a robust system for collecting information</td>
</tr>
<tr>
<td>✔️</td>
<td>4. Internal Locus of Control</td>
<td>✔️</td>
<td>4. Assure that every low-stakes assessment provides meaningful feedback for learning</td>
</tr>
<tr>
<td></td>
<td>5. Reflection</td>
<td>✔️</td>
<td>5. Provide mentoring to learners</td>
</tr>
<tr>
<td>X</td>
<td>7. Outcomes</td>
<td>✔️</td>
<td>7. Organize intermediate decision-making assessments</td>
</tr>
</tbody>
</table>

#### van der Vleuten

| ✔️ | 1. Develop a master plan for assessment |
| ✔️ | 2. Develop examination regulations that promote feedback orientation |
| ✔️ | 3. Adopt a robust system for collecting information |
| ✔️ | 4. Assure that every low-stakes assessment provides meaningful feedback for learning |
| ✔️ | 5. Provide mentoring to learners |
| ✔️ | 6. Ensure trustworthy decision-making |
| ✔️ | 7. Organize intermediate decision-making assessments |
| ✔️ | 8. Encourage and facilitate personalized remediation |
| ✔️ | 9. Monitor and evaluate the learning effect of the program and adapt |
| ✔️ | 10. Use the assessment process information for curriculum evaluation |
| X | 11. Promote continuous interaction between the stakeholders |
| ✔️ | 12. Develop a strategy for implementation |
Way Forward

• Kelvin: “If you cannot measure it, you cannot improve it.”
  – Develop measurement strategies
• van der Vleuten: “High-quality feedback should be the prime purpose of any individual data point.”
  – Connect assessment to feedback (for learning)
• Berra: “If you don’t know where you are going, you will end up someplace else.
  – Connect education and assessment to patient and system outcomes
• Buddha: “Three things cannot be long hidden: the sun, the moon, and the truth.”
  – Seek the truth (validity) and tell the truth (together)
What Questions Do You Have?

- Mapping
- Faculty Development
- Use in Remediation
- Paying it Forward
- Validity
- Complexity Theory
- Argument
- Theory
- Timely Completion