The Role of Descriptive Evaluation in Health Sciences Education: RIME, Competencies, and Milestones

Paul A. Hemmer, MD, MPH
Professor and Vice Chairman for Educational Programs
Department of Medicine
Uniformed Services University
Bethesda, MD
Acknowledgements

• Louis Pangaro, MD
• Eric Holmboe, MD
Disclaimer

• Views expressed in this talk are those of the author and do not reflect the official views of the Uniformed Services University, the United States Air Force, the Department of Defense, or other federal agencies.
• Faculty want to do well
• It Matters What Faculty Have to Say
  – Help them to say what matters
• Give them a usable, portable framework
• Train them use it: “We need to talk”
Descriptive Evaluation

• Words instructors use in their assessment of [a trainee’s] demonstrated competence…usually based on their observations over a given period of time.

• Conveying one’s ideas, thoughts, observations, and a synthesized judgment with words

Beliefs about Evaluation

• Descriptive (words) = “Subjective”

• Quantified (numerically scored) = “Objective”
Why is Descriptive Evaluation Important?
Disciplinary Action by Medical Boards

Papadakis M. et.al. NEJM. 2005;353:2673-82.

• Cases: OR 3.0: Unprof behavior in med school
  – Explained 24% of variance in disciplinary action
  – 19% of controls had something in record

• Most significant behaviors
  – Irresponsibility: OR 8.5 if cited ≥ 3 times
  – Diminished capacity for self improvement (e.g., argumentative w/Fb): OR 3.1 if ≥ 3 times

• Other variables: failure to pass course on first attempt, low MCAT (1% of variance), low preclinical GPA (7% of variance)
Need for Expert Judgment

• “Information gathering for the assessment of such general competencies will increasingly be based on qualitative, descriptive and narrative information…”
  – “…we will come to rely more on professional judgment as a basis for decision making about the quality and the implications of that information. The challenge…making [this] as rigorous as possible without trivializing the content for ‘objectivity’ reasons.”

Evaluation Forms = Rating Scale + Narrative Comments

How Do We Do Descriptive Evaluation?
Rating Scales

• Lack consensus on what they are evaluating
  – Lack understanding of program goals/objectives
• Poor rater agreement about performance
  – “Doves” and “Hawks”
• Rater errors are common (Halo, compensation)
• Changes to evaluation form will account for less than 10% of the variance

Comments on Forms: Who Has Written the Following?

- “Read more”
- “Enthusiastic”, “Engaged”
- “Will do well in whatever he/she chooses”
- “Too bad he/she wants to do XXX”
- “Who?”
- “A pleasure to work with”
- Or, are people even writing their own words?
A Pleasure to Work With: Analysis of Comments on Student Evaluations

• Peds clerkship, 1017 comments
  – 14% had been eliminated as useless
• Mean: 4 comments/form [1-14]
• Learner (26%), Personal (25%) characteristics most common
• “Pleasure to work with”: most common, 8%
• 34% had some specificity (“focused write-up”)
• Most comments don’t comply with good Fb
Message?

• Evaluation forms don’t evaluate
• Evaluation forms can communicate goals
Broadening Perspectives on Clinical Performance Assessment: Rethinking the Nature of In-training Assessment


- “...performance assessment is a judgment and decision making process, in which rating outcomes are influenced by interactions between individuals and the social context in which assessment occurs.”
- “…focusing on the context of performance assessment may be more effective in improving ITA practices than focusing strictly on raters and rating instruments.”

Does your descriptive evaluation process do this?
Challenges to Improving Descriptive Evaluation

• Provide a usable evaluation framework
• Embed the evaluation and training processes into usual activities
Frameworks for Goals

Three useful models of expressing expectations:

1. Analytic
2. Developmental
3. Synthetic
Analytic expression of Goals

- “ana-lytic”: takes the learner “apart”
- into domains, categories “attitude”, “skills”, “knowledge”
- domains = generic terms
- useful for discrete assessments
Use of analytic to encompass complex tasks

managing Cardio-Pulmonary Resuscitation

Skills .......... Placing central line
Knowledge.. Knowing the right drug
Attitude....... Confidence to “run” code
2. Developmental Dreyfus and Dreyfus

- Novice
- Advanced beginner
- Competent performance
- Proficient performance
- Intuitive expert
- Master

Mind Over Machine (1986)
Dreyfus Definitions

- **Novice**
  - Applies rules to facts, features—“ignoring” context

- **Advanced Beginner**
  - Begins to recognize “situations” and connecting rules to situations, past experience

- **Competent**
  - Considers both context-free and situational elements
  - Organizes and reduces elements to reach decision
  - Has sense of “ownership” in process and outcomes

Charting the Road to Competence: Developmental Milestones for Internal Medicine Residency Training. JGME. 2009;Sep:5-20.
Competence

• “Competence is the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and the community being served.” (Epstein)

• The ability to give to each situation all that belongs to that situation, and no more. (Pangaro)

• A personal quality, not an action (ten Cate)

• Key: Competence is Synthetic, contextual
Competency

• An observable ability of a health professional, integrating multiple components such as knowledge, skills, values, and attitudes. They can be measured and assessed to ensure their acquisition.

Competent

• Possessing the required abilities in all domains in a certain context at a defined stage of medical education or practice.

Competence

• How I feel when we talk about these definitions!

Hemmer 2012
Common Frameworks

ACGME Core Competencies

- Patient Care
- Medical Knowledge
- Interpersonal Skills
- Professionalism
- PBLI
- SBP

MSOP

- Knowledgeable
- Skillful
- Altruistic
- Dutiful

ACGME = Accreditation Council for Graduate Medical Education
MSOP = Medical Schools Objectives Project
Translating Competencies: Milestones

www.abim.org
### APPENDIX 2.1

**DEVELOPMENTAL MILESTONES FOR INTERNAL MEDICINE TRAINING—PATIENT CARE**

<table>
<thead>
<tr>
<th>ACME Competency</th>
<th>Developmental Milestones Informing ACME Competencies</th>
<th>Approximate Time Frame Trainee Should Achieve Stage (months)</th>
<th>General Evaluation Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical skills and reasoning</td>
<td><strong>Historical data gathering</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Manage patients using clinical skills of interviewing and physical examination</td>
<td>1. Acquire accurate and relevant history from the patient in an efficiently customized, prioritized, and hypothesis driven fashion</td>
<td>6</td>
<td>Standardized patient Direct observation</td>
</tr>
<tr>
<td>▪ Demonstrate competence in the performance of procedures mandated by the ABIM</td>
<td>2. Seek and obtain appropriate, verified, and prioritized data from secondary sources (eg, family, records, pharmacy)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>▪ Appropriately use laboratory and imaging techniques</td>
<td>3. Obtain relevant historical subtleties that inform and prioritize both differential diagnoses and diagnostic plans, including sensitive, complicated, and detailed information that may not often be volunteered by the patient</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Role model gathering subtle and reliable information from the patient for junior members of the health care team</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Performing a physical examination</td>
<td><strong>Performing a physical examination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Perform an accurate physical examination that is appropriately targeted to the patient’s complaints and medical conditions. Identify pertinent abnormalities using common maneuvers</td>
<td>6</td>
<td>Standardized patient Direct observation Simulation</td>
</tr>
</tbody>
</table>

Risk of Competency Models

• Reduces trainee “success” to LCD
• Trivializing what it means to be competent
  • “atomizing competencies, increasing bureaucracy, and moving away from expert opinion and from what really matters in day-to-day clinical practice”*
• Challenge: “produce expert professionals... in a culture that emphasizes competency rather than expertise.”**

* ten Cate O, et.al. Med Teacher. 2010;32:669-75
**BMJ. 2006;333:99.
Third, alternative model:

- Framework that is
  - synthetic
  - developmental
  - behavioral
  - can visualize progress
Synthetic Model
“R.I.M.E.”

Reporter
Interpreter
Manager-Educator

Reporter

- Takes ownership of working in patient care and monitoring own patients
- Answers “What” questions
- Accurately, reliably assesses and communicates on one’s own
  - Complete, Honest
- Takes: knowledge, responsibility, hard-work, trust
The standard

• more than simple attendance (“Observer”)
• more than repeater or reciter of others’ work.
• consistent, reliable data gathering is essential and must be directly observed, documented (DOC)
Interpreter

- Ownership of the “Why” questions
- Prioritizes, analyzes, synthesizes
  - Reasonable, not “right” (student)
- Takes: more knowledge, confidence, greater independence
- “Explain your ideas, reasoning for me”
Manager

- Ownership of the “How” questions
- Proposes actions and options applied to their patient
- Has maturity, skill, and knowledge to negotiate with patients/team on plans
Educator

- Owning the growth toward expertise
- Poses questions, independently seeks answers
- Shares new knowledge, teaches others, becomes a leader
The Rhythm of RIME

**SOAP**

Symptoms
("Subjective")

Observations
("Objective")

**Observation**

**Reflection**

**Action**

**Observer**

**Interpreter**

**Assessment**

**Manager**

**Plan**

**Educator**
3. the “Synthetic” framework

- “synthetic” - putting the learner back together
- “K S A” are all required, integrated
- useful for complex tasks (functioning in patient care)
3. the “Synthetic" framework

- developed for descriptive evaluation for clinicians
- terms are a bit less generic
- Embraces that growth toward independence does not leave behind prior “steps”
  - “Competent” is no longer “Novice”, but Manager/Educator must still Report
Matrix: Transition to Higher Expectations

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>PGY1</th>
<th>PGY 2...n</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUCATOR</td>
<td>I</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>P</td>
<td>M</td>
</tr>
<tr>
<td>MANAGER</td>
<td>I</td>
<td>R</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>INTERPRETER</td>
<td>I</td>
<td>R</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>REPORTER</td>
<td>I</td>
<td>R</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
</tbody>
</table>

I = introduced in the curriculum,  
R = repetition, practice,  
P = sufficient proficiency for the next level of independence  
M = mastery in practice
### Matrix: Transition to Higher Expectations

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>PGY1</th>
<th>PGY2/3</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUCATOR</td>
<td></td>
<td></td>
<td></td>
<td>Reason</td>
<td>“Right”</td>
</tr>
<tr>
<td>MANAGER</td>
<td></td>
<td></td>
<td>Reason</td>
<td>“Right”</td>
<td>Right</td>
</tr>
<tr>
<td>INTERPRETER</td>
<td>Reason</td>
<td>“Right”</td>
<td>Right</td>
<td>Right</td>
<td>Right</td>
</tr>
<tr>
<td>REPORTER</td>
<td>Right</td>
<td>Right</td>
<td>Right</td>
<td>Right</td>
<td>Right</td>
</tr>
</tbody>
</table>

Right = Correct  [What is minimum?]
“Right” = Usually/often correct
Reason = Reasonable
Models

**Analytic**
- Takes a learner “apart”
- Knowledge, Skills, Attitudes
- Reductionist (“whole is sum of its parts”)
- Difficult to embrace complex tasks
- Helpful diagnostically

**Synthetic**
- Puts a learner together
- RIME, “Medical Expert”
- “Whole may be greater than sum of its parts”
- Embraces complex tasks
- May not localize problem
Complimentary methods

• Synthetic first (observation)
  – At what RIME level is this TRAINEE?

• Analytic next (reflection)
  – what are the barriers?
  – honesty? handling stress?

• Feedback follows (action)
  – discussion of values
  – counseling
Helping teachers:

- How do we get them to play from the same sheet of music?
- How do we get them to play at all??
- How do we get them to observe and listen?
- SIMPLICITY AND FAMILIARITY
How to Improve Descriptive Evaluation?

• Provide a usable framework

• **Talk with one another:** Embed the evaluation and training processes into usual activities
Evaluation Sessions

• Meeting with teachers during rotations
• Attended by all working with trainee

Goals
  • To learn about and evaluate trainees
  • To generate and provide feedback
  • To develop the teachers

Hemmer PA, Pangaro L. Acad Med. 2000;75:1216-21
Formal Evaluation Sessions

Format

- 15 minutes per trainee
- Leader sets goals
- Teachers respond, in turn, to open-ended and directed questions
- Evaluators recommend evaluation
- Feedback given to teachers
- Develop action plan: “Next Step”
Eval Sessions Engage Teachers

• **Frame of Reference Training**
  – Teach faculty how to use evaluation tool
  – Agree on relative importance of different components being assessed

• **Rater Error Training**
  – Identify, discuss common sources of error to improve self-identification
  – Halo effect, Compensation fallacy

**Practical Guide to the Evaluation of Clinical Competence.**
Formal Evaluation Sessions
Capitalize On:

• Low-Tech

• Teachers will tell you what they won’t write down
Evaluation Sessions can reflect our professional commitment

- To society (Evaluation)
- To students (Feedback)
- To teachers (Faculty Development)
Do People Really Come to the Sessions?

If You Feed Them They Will Come
Strategies

• Simplicity, portability of goals

  Synthetic, then Analytic

• Sit down with Teachers

• Then talk with trainees
What We Know So Far

Some Selections
RIME and Evaluation Sessions
Clerkship Directors in IM Annual Survey: 2005

• RIME
  – Used by 42% of US IM clerkships
    • 4.5 yrs (± 2.3, 0-19 yrs)

• Evaluation Sessions
  – Used by 45% of US IM clerkships
    • 2.4 meetings during clerkship

Using RIME in Evaluations

- Ambulatory teachers who come to evaluation sessions use RIME more frequently

<table>
<thead>
<tr>
<th></th>
<th>Attended Eval Session (n= 190)</th>
<th>Did Not Attend Eval Session (n=68)</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIME Utterances per narrative</td>
<td>1.9</td>
<td>0.9</td>
<td>&lt;0.0001</td>
<td>0.55</td>
</tr>
<tr>
<td>Narratives containing RIME utterances</td>
<td>69.8%</td>
<td>40.4%</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Were Teachers’ Grade Recommendations Consistent with Their Narrative Comments?

- Comparing teacher and coder grade recommendations based on blinded review of narratives

<table>
<thead>
<tr>
<th>Attended Eval Session</th>
<th>r = 0.72*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not attend eval session</td>
<td>r = 0.47*</td>
</tr>
</tbody>
</table>

*p < 0.001

Sensitivity of Evaluation Methods
Identifying Students with Poor Fund of Knowledge

% failed NBME

Check List | Written Forms | Eval Sessions
---|---|---
19 | 25 | 44

class of 93
124 students

Medicine
USUHS
Detecting Deficiencies in Professionalism


25% of comments made only at Eval Session

* DI = % of professionalism domains rated less than acceptable by evaluators
# p < 0.04, Eval vs Check, OR 1.8; + p < 0.001, Eval vs Check or Writ, OR 1.7
Effect of RIME: Medicine Clerkship Grade Distributions Univ. of Utah


O = Observer, R = Reporter, I = Interpreter, M = Manager, E = Educator
0 = Poor, 4 = Excellent
Sensitivity of Third Year Grades in Predicting Internship Problems

Acad Med. 1998 Sep;73(9):998-1002.

USU Medicine Evaluation System, classes of 86 - 93
Using PGY1 director ratings

• Students identified during the clerkship as needing remediation were \textbf{10 times} more likely to receive low rating or negative comments from internship directors.
Reasons to try RIME

Bloomfield L., et.al. Med Educ. 2007;41:1083

• Univ New South Wales, Sydney, Aust
  – Implemented with jr and sr students
  – RIME and eval sessions

• RIME motivated students to see and present more patients

• Teachers more motivated to be involved in assessment and reflect on teaching
BLUF
(Bottom Line Up Front)

• Faculty want to do well
• It Matters What Faculty Have to Say
  – Help them to say what matters
• Give them a usable, portable framework
• Train them use it: “We need to talk”
Disciplinary Action by Medical Boards and Prior Behavior in Medical School.

Papadakis M, et.al. NEJM. 2005;353:2673-82

• Retrospective, case control, 3 med schools
• Reviewed/Collected data
  – Narratives (admission interviews, course evaluations including check marks), Dean’s letter, any other documentation
    • Rated: Concern, Problem, Extreme=Unprofessional
    • Severity=frequency of occurrence
  – Undergrad GPA, MCAT, NBME/USMLE 1, preclinical and clinical grade
Authenticity of Assessment

Assessing Professional Competence: from methods to programmes.

Descriptive Evaluation

- Unannounced SPs, Chart review
- Simulations, OSCEs, mCEX
- Simulations, Simulators, Prosthetics
- NBME/USLME vignettes

Miller (Acad Med. 1990)
Faculty Comments about Residents

- 1,770 evaluations, 180 IM residents, U Toronto
- 5.8 codable comments per form
- Many comments mapped to CANMEDS
  - Knowledge, Professionalism common
  - Often mapped to more than one role
- Others comments
  - Global comments most common of all types
  - Impact of resident on faculty
  - Trajectory of resident
- Comments revealed synthetic judgment

The Quality of Written Comments on Professional Behaviors in a Developmental Multisource Feedback Program


- 8 GME training programs piloting NBME’s APB
- 970 surveys done; 282 (29%) had comments for 1,019 feedback phrases:
  - 75% general, self-directed comment ("great doctor")
  - 90% of comments were positive
  - 11% referenced specific instance of behavior
  - 7%--specific behavioral strategy for improvement
- “Comments often lack effective feedback characteristics.”
• “Trust in and acceptance of the assessment system by raters and ratees is a crucial factor.”

• “The underlying performance theories should be explicated and communicated to all parties involved in the assessment.”

• “Rater training should focus not only on rater ability, but also (and perhaps even more so) on rater motivation.”
Competence is Contextual

• “...the definition of medical competence is bound to local political, social, and economic circumstances, to health needs, to the availability of resources, and to the structure of the health care system. Thus any effort to find a universal definition of competence will inevitably fail.”

Milestones Project

• make explicit the profession’s expectations
• promote competency-based resident education
• track the progress of residents
• inform decisions regarding promotion and readiness for independent practice.
• may guide curriculum development
• suggest specific assessment strategies
• provide benchmarks for resident self-assessment
• assist remediation, identification of specific deficits

American Board of Internal Medicine.  www.abim.org
The heart of the matter:

- Competence
- All frameworks, lists of attributes, etc. are ways of trying to express this concept of what “success” looks like.
When to Entrust Residents with Unsupervised Tasks?
Acad. Med. 2010;85:1406-1411
## Link RIME With ACGME

### Practice Based Learning and Improvement

<table>
<thead>
<tr>
<th></th>
<th>Reporter (PGY-1)</th>
<th>Interpreter (PGY-2)</th>
<th>Manager/Educ (PGY-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poses the question to self/others</td>
<td>Shows how to gather information</td>
<td>Presents analysis to others</td>
<td></td>
</tr>
<tr>
<td>Knows where to get information</td>
<td>Explains limits of question, data</td>
<td>Creates plan for how to change</td>
<td></td>
</tr>
<tr>
<td>Familiar with data retrieval/analysis</td>
<td>Understands how to evaluate data</td>
<td>Defines success for future care</td>
<td></td>
</tr>
<tr>
<td>Reliably meets deadlines</td>
<td>Poses reasonable explanations for findings</td>
<td>Open to comments, focused on needs of patients</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation forms don’t evaluate!

• Faculty evaluate
• Forms can help Communicate goals
• There is no “Holy Grail” of forms
• Basic tenets of “good” forms
  – Short, clear
  – Behavioral anchors
  – Reinforce goals, expectations
Eval Sessions Engage Teachers

- **Performance Dimension Training**
  - Standardize observation of behavior of interest
  - Consensus on terms for desired expectations

- **Frame of Reference Training**
  - Teach faculty how to use evaluation tool
  - Agree on relative importance of different components being assessed

- **Rater Error Training**
  - Identify, discuss common sources of error to improve self-identification
  - Halo effect, Compensation fallacy

How Has Training Been Influenced?
Hodges BD. Acad Med. 2010;85(9):S34-44

**Tea-Steeping Model**
- Enough time and care— _voila_!
- Change the tea leaves (trainees)
- Change the water (environment)

**i-Doc**
- Competency move at extreme
- Mechanistic, engineering model
- Focus on efficiencies, best way to produce product

It’s probably a combination of both but we’re moving away from dwell time model and is the pendulum swinging too far?
The goal: progressive independence of the learner after SFDP
## RIME and ACGME

### Patient Care Competency

<table>
<thead>
<tr>
<th>ACGME</th>
<th>RIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication with patient and family</td>
<td>Reporter</td>
</tr>
<tr>
<td>Able to collect data from patient or family</td>
<td>Reporter</td>
</tr>
<tr>
<td>Appropriate diagnostic and therapeutic interventions</td>
<td>Manager</td>
</tr>
<tr>
<td>Use of information technology</td>
<td>Reporter and Manager</td>
</tr>
</tbody>
</table>
## Linking RIME With ACGME

### Practice Based Learning and Improvement

<table>
<thead>
<tr>
<th>Reporter (PGY-1)</th>
<th>Interpreter (PGY-2)</th>
<th>Manager/Educ (PGY-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poses the question to self/others</td>
<td>Shows how to gather information</td>
<td>Presents analysis to others</td>
</tr>
<tr>
<td>Knows where to get information</td>
<td>Explains limits of question, data</td>
<td>Creates plan for how to change</td>
</tr>
<tr>
<td>Familiar with data retrieval/analysis</td>
<td>Understands how to evaluate data</td>
<td>Defines success for future care</td>
</tr>
<tr>
<td>Reliably meets deadlines</td>
<td>Poses reasonable explanations for findings</td>
<td>Open to colleagues’ comments, focused on needs of patients</td>
</tr>
</tbody>
</table>
Describing (minimal) success

• Finishing clerkship students:
  – every day owns
    • how patient feels
    • important findings (about patient and underlying disease)
    • reasonable understanding when asked
  ➢ reliable reporter; moving to interpreter
For most core clerkship students, curriculum should be a requirement to report and an invitation to interpret.
Describing (minimal) success

- finishing interns can also:
  - pro-actively explain new findings,
  - give a “differential”,
  - prioritize urgency
  - implement diagnostic plan
  - suggest therapy

- interpreters and early managers for common, acute problems
For most interns, curriculum should be an invitation to interpret and manage.
Describing (minimal) success

- finishing residents can also:
  - on ward, clinic and consultation
  - work with patients on plans
  - able to give to all usual, even complex, situations all that belongs to those situations
  - are self-correcting, learn quickly what is required, can help others grow.

- manger-educators
For finishing residents, curriculum should be a requirement to manage and educate.