Annual Review of the Medical Education Literature

Amy Shaheen, MD, MSc
University of North Carolina
Chapel Hill, NC
March 20th, 2018
Alliance for Academic Medicine Annual Meeting
Disclosures:

None
With the assistance of my friendly neighborhood health sciences librarian...

The previous award winner....

My work wives...
And my family

Yes, those are under-aged kids with beer!

26 years together (I was a child bride...)

But these people are sober...
Search Strategy and Article Choice

For the first time, scientists were close to determining how difficult it is to actually find a needle in a haystack.
KeyLIME – Key Literature in Medical Education
Ultimately... From March 2017-Feb 2018

Whittled down to about 1000 based on interest/quality/my ability to read/ less visible articles

24 Key LIME podcasts
Dr. MERL reviews

Three sections:
• Bias, perceptions, feedback, evaluations
• Clinical reasoning
• Teaching and testing in the classroom and within our systems
Role of a medical student: patient perspectives

David Evans, Stephanie Owen and John Green
School of Medicine, Cardiff University, UK
Background

• Medical students are part of teams
• Do patients recognize students?
• How they feel about the role of students?

**AIM:**
Explore patient perspectives on medical students and their engagement in patient care activities
Results

- 33% male and 67% female patients, Mean age 63.4 (21-100)
- 74% believed they could recognize a student
- Usually by age and attitude
- 87% believed students should wear a uniform
- Only 10/75 mentioned the only reliable method to tell-name tags
Skills

50% or more agreed they would let a student do the following on them:
• Draw blood
• ABG
• Catheterization
• Preparing drugs for injection or giving injections
• Transfuse blood
• Suture
• Get consent for surgery
What is the role of medical students?

• “to understand the science and communication required to be a competent doctor”
• “to be taught what they need to know and if it weren’t for students we’d have no doctors”
• “to be sober and pleasant and to back up the opinion of the senior’s position”
Impact/limitations

• Patients understand trade-off between safety and student learning
• ID learners better
• Competencies aligned well with what patients thought was acceptable-not exactly the same as US competencies

• Limitations: May have given socially acceptable answers to students who were conducting survey and those who agree may view students more positively
• Surgery culture is unique
• Culture affects how a learner may want to be perceived
• That perception may be important to further learning experiences
Methods:

• Qualitative study of general surgery trainees to understand how they control their image while trying to learn and how their efforts effect judgment and decision-making

• Using semi-structured interviews, interviewed fifteen surgical residents (8 male and 7 female), eight of them PGY 1-2 and Seven PGY 3

• Themes were identified inductively and applied deductively to known concepts of impression management by Goffman
Results:

• Themes were similar for women and men
• Feel pressure to be “all-knowing”, “quick”, “decisive”, and “confident”
• Strategies to exhibit competence, especially when they sensed that they were not meeting expectations
  1. Fabrication
  2. Silence
  3. Avoiding calling for help
  4. Mis-ordering/over-ordering of tests to try to sort things out
Resident Consequences-Intended and Otherwise

• Building a reputation of “competent” resident—More opportunities
• Avoid branding as “incompetent”—Less opportunities

• Withholding questions so as not to appear incompetent
• Affects resident wellness and causes shame for weak residents
• A lot of mental energy to look perfect for high performing residents
Patient care effects

Describe being so nervous about being watched, that surgery was rushed—“to the point that we ended up putting the kidney in upside down”

Limitations

• Single institution, culture varies from place to place, specialty to specialty
• Canadian Institution
Many factors play into whether or not a trainee participates in procedures or “technical skills”

- trainee (personal interest, intentionality, and engagement)
- teacher trust (trainee ability, integrity, keenness and benevolence)
- environmental factors (availability of procedures, competing learners)
Results

- 419 hours of observation/18 interviews/4 focus groups
- 26 residents/5 fellows/17 faculty/9 respiratory therapists/44 nurses
- 17 different procedures for a total of 287 times over 3 months
- 27.5% performed by residents
- Most common were suturing, fracture reduction, casting, and lumbar puncture
Results-qualitative

• **Learner factors-Desire** to learn, understanding of own needs, competing priorities

• **Teacher factors**-Desire to teach, understanding of learner needs, competing priorities, estimation of learner trustworthiness and competency and keenness

• **Environmental factors**-family willingness, team trust, patient appropriateness, competing learners, and environmental constraints
Summary/limitations

• Trust, motivation/keenness, curriculum and logistics are themes for all participants
• Many opportunities lost
• Team members (like nurses) may interfere with learner involvement due to “trust issues”
• Enthusiasm is rewarded with opportunities (or penalized)
• Site may be a limitation (peds ER) for generalizability
• Almost half of the procedures were IV access (is this important?)
• Errors make us all feel bad
• External feedback must track with our own self-assessment to be useful
• Some physicians, no matter how competent they are, constantly doubt their abilities and feel like “an imposter”
• The “Imposter syndrome” is an extreme form of self doubt
• This feeling of inadequacy can lead to burn out and depression in competent physicians
Methods

• Qualitative study that looked at how self-doubt unfolds over time
• Sent invites to 1000 Physicians at one large Canadian Hospital (unnamed)
• 28 agreed to be interviewed about their experiences with perceived or real underperformance or failure
• Asked how those experiences impacted teaching
Results

• Ten females and 18 male
• Mostly mid or late career physicians (23/28)
• Only one IM physician
• Pediatrics, Neurology, EM, Radiology, and Surgery each had 3 or more interviewed
• Some shared examples of underperformance or failures
• Most perceived underperformance
• Those at the extreme end of this anxiety, they identified as “Imposters”
Results

• The imposter syndrome (pretended confidence) protected them from arrogance—important in a risk averse profession
• Transitions (new jobs or new responsibilities) are high risk times for adequately or even high performing faculty
• Perceived competence may not ever equal actual competence
Limitations

• One institution
• Only one IM physician
• Self nomination for physicians willing to speak candidly about their own perceived failures may self-select for certain level of anxiety
• Fewer women may have skewed results
• Study about perceptions not reality
Impact/future directions

• Do frequent transitions exacerbate this?
• Is there a way to ameliorate this with mentoring or coaching or discussion?
• Direct observation may exacerbate this “showmanship” because the feedback is not in an authentic setting
• Feedback may not dismissed by both the competent and incompetent
• The very insecure may excel but at significant cost to their own well-being
• Widening Access applicants=Lower SES applicants in this country
• Leadership effects tone
• Certain terms may “pathologise” applicants and students, make others view them as different or deficient
• Does this perhaps perpetuate a stereotype of what an applicant or leader should look like?
Methods

• Qualitative study of the language about WA applicants in telephone interviews with 26 admissions deans representing 24/32 UK medical schools

• Used the construct “othering” from Coupland as the lens with which to evaluate the interviews—using objectifying, differences

Table 1 The analysis process

| Step 1: Corpus linguistic analysis—“large pattern” identification |
| Step 2: Categories of representation (pronominal use, naming strategies) |
| Step 3: Data tagging [as per Coupland’s (2000) discursive manifestations] |
| Step 4: Critical discourse analysis (CDA) |
Data Tagging as “Others”

• **Homogenization**: specific cultural traits/behaviors to a group
• **Perjoration**: assigning negative attributes to a group
• **Suppression/Minoritization**: omission, selective representation or positioning group in specific roles
• **Displaying liberalism**: Hedging offensive remarks by claiming non-prejudicial intention to make oneself look better
• **Subverting tolerance**: ridiculing political correctness
Results

Deans use a variety of terms to describe WA candidates that use all of the ways to “other”

**Homogenizing:** “I think we do realize that, from some backgrounds, it is the case of people don’t think of medicine in the first place. They think it’s not...a suitable career for them”

**Perjoration:** “That they feel, is in our area, total lack of aspiration. It’s not that they don’t have gifted children who could do it, that they’ve got no aspiration.”
Results

• **Suppression/minoritisation:** “another bit of work needs to be done on making sure that they are supported so that they are performing at the same level quickly enough once they get to medical school and it’s not that...they come to medical school with all of these hopes and then they’re struggling to keep up with their peers”

• **Displaying liberalism:** “We do work very actively to try and encourage widening participation. Um, it you know, XXX is a fairly down to earth sort of place and, and we are keen to, to, you know, widen the net if we can.”
Impact

• Adds to the literature about implicit bias and “othering” which is subconscious at a leadership level
• Suggests that leadership and culture are one and the same
• Serves to reinforce existing stereotypes and a significant “us” and “them” rhetoric exists in medical student admissions
• Limitations: English medical schools, socioeconomic differences
Medical School Applicant Characteristics Associated With Performance in Multiple Mini-Interviews Versus Traditional Interviews: A Multi-Institutional Study

Mark C. Henderson; Carolyn J. Kelly; Erin Griffin; Theodore R. Hall; Anthony Jerant; Ellena M. Peterson; Julie A. Rainwater; Francis J. Sousa; David Wofsy; Peter Franks

Academic Medicine. Publish Ahead of Print(); OCT 2017
AIM/Methods

• Examine the applicant characteristics associated with Multi mini interview and Traditional interview performance.

• Applicants to 5 California medical schools from 2011-2013

• Retrospective study of 5 California medical schools from 2011-2013, 2 using Multi-Mini interviews (MMI) and 3 using Traditional Interviews (TI).

• Looked at how MMI and TI were scored for a variety of applicant characteristics
Results:

• 4993 applicants/ 7516 interviews.
• 18.6%(931) included UIM and 962 (19.3%) self described socially disadvantaged students.
• Females, older, and number of prior interviews predicted better performance on the MMIs and the TIs.
• Socially disadvantaged students better traditional interview skills but worse MMI scores.
Results

• UIM status no impact on interview performance
• Higher GPA was associated with lower MMI scores but no change in the TI score
• Practice, i.e. more interviews, also had a statistically significant positive impact on performance
• GPA negatively impacted MMI but extraversion positively impacted MMI.
Limitations and Impact

• Mean MCAT was 94th percentile so may not be generalizable to other medical schools
• How were interviewers and interviewees matched?
• Consider strongly the impact on your pool when switching interviewing methods
REAL QUEENS
FIX EACH OTHER'S
CROWNS.
• Clinical Performance Evaluations account for 50-70% of grades
• Known: OSCE female graders grade women higher than men
• In general, preclinical grades, MCATs, Step 1 are not different for men and women who are accepted

**AIM:** To determine whether student or evaluator gender is associated with the assigned grade at one large academic medical school
Methods:

- Retrospective study at Alpert medical school, one year of CPEs (4462 by 829 evaluators) from all specialties
- Excluded from analysis “below expectations” (16), duplicates (38), lacking analyzable information (136).
- Total 4272 analyzed for
  - Gender of evaluator and student
  - Length of exposure to evaluator
  - Evaluation department
  - Age of student and age of evaluator
Results

• Women had lower USMLE 1 scores 221 than men 231. (p=0.0083)
• Median age of evaluators is 33. (29-45)
• 57% of the evaluations came from Medicine (30%) and Peds (27%)
• Median number of evaluations per student 27 (IQR 6-39)
• Median number of evaluations per evaluator 3 (IQR, 1-7)
<table>
<thead>
<tr>
<th>Odds of Better grades</th>
<th>Odds of Lower grades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female student</td>
<td>Female evaluator of any age</td>
</tr>
<tr>
<td>Younger evaluator</td>
<td>Older male</td>
</tr>
<tr>
<td>Internal Medicine rotation</td>
<td>Ob/Gyne rotation</td>
</tr>
<tr>
<td>Older Student</td>
<td>Younger student</td>
</tr>
<tr>
<td>Longer time with evaluator</td>
<td>Shorter time with evaluator</td>
</tr>
</tbody>
</table>

- Female evaluators gave lower grades to male students
- Older male evaluators gave lower grades than younger male evaluators
- Younger male evaluators give the best grades of all
Impact-female students have higher CPEs driven mostly by female grading patterns

• Do female physicians detect a true difference in student performance between men and women?
• Do women have a better learning attitude?
• Or do patients respond differently to the female learners that result in better grades?
• Differences from course to course may be confusing to students
• Duration of rotation and evaluation is significant
Comparison of Male vs Female Resident Milestone Evaluations by Faculty During Emergency Medicine Residency Training

Arjun Dayal, BS; Daniel M. O'Connor, BA; Usama Qadri, BA; Vineet M. Arora, MD, MAPP

Gender Differences in Attending Physicians’ Feedback to Residents: A Qualitative Analysis

Anna S. Mueller, MA, PhD
Tania M. Jenkins, MA, PhD
Melissa Osborne, MA

Arjun Dayal, MD
Daniel M. O’Connor, MD
Vineet M. Arora, MD, MAPP

Journal of Graduate Medical Education, October 2017
Introduction

• Does gender bias impact current competency evaluations?
• Emergency Medicine has fully implemented the NAS Competency based evaluation system with milestones that have been tested and implemented across 100% of programs

Aim

To compare male and female direct observation milestone evaluations using a longitudinal, multi-institution data set.
Methods:

• Retrospective Cohort study of Individual Milestone Evaluations from 2013-5 from 8 EM programs that used an App to record evaluations of direct observations (1-3/shift)
• Residents were evaluated on a scale from 1-5 with anchors and optional comments
• Compared male and female evaluations
• Looked for confounders: Time off, community programs, type of competency (procedural or otherwise), gender of evaluator
Results

- 33,456 evaluations of 359 residents (237 men, 122 women) by 285 faculty (194 men, 91 women) over 2 years
- 8 training programs (6 academic and 2 community) over four geographic areas
- Mean number of evaluations: 96 for female residents and 87 for males residents
Results

• 23 sub-competencies
• No difference between males and females in PGY1
• By PGY 3, Women were evaluated lower on all 23 sub-competencies by both male and female evaluators
• Men attained competency, on average, about 0.07 milestone levels faster than women
• No differences in periods without an evaluation between males and females. (This is a marker for extended periods of absence that could account for pregnancies, illness, etc.)
Impact

• Is this gender evaluation bias?
• Are the evaluation forms biased against females?
• Also, when expectations are not met, then feedback, teaching and support may not be the same?
• Or is there less mentoring for women and women end up less proficient (the system performs exactly as it was designed to perform)
• 15,000 comments accompanied evaluations could help understand what led to this difference between males and females
Aim:
To understand the lagging performance of females in EM milestone evaluations

Methods:

• Semi-Analytic Qualitative study of the EM PGY3 evaluations in one hospital for over 2 years (47 PGY 3’s evals and 63 faculty’s evals)
• 35 residents had more than 15 comments and those were the ones included in the analysis
Results

• 1317 comments were evaluated for characteristics that embody the ideal resident
• Many of these were consistent with stereotypical masculine traits
• Feedback was gendered with expectations for females to show typical masculine traits
• Multiple women were both criticized and praised for receptiveness to feedback
• Only one male was criticized for being sensitive to feedback
• One male was praised for it
Results

• Half of males received no negative comments versus only one female resident
• 62% female versus 36% of male residents were criticized multiple times for not having valued EM personality traits
• Males received consistent feedback on what they needed to improve
• Women received more discordant feedback-frequent tension between autonomy and receptiveness to feedback-does this impact progression?
• Limitations: Single institution, One specialty
British culture is deeply impregnated with 'hinting', getting the message across, without ever 'coming out with it'.
Influence of feedback characteristics on perceived learning value of feedback in clerkships: does culture matter?

<table>
<thead>
<tr>
<th>Impact on Feedback</th>
<th>Indonesia-Strong hierarchies And MODESTY is valued</th>
<th>Netherlands-Weak hierarchies And INDIVIDUALISM valued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attending versus Resident giving feedback</td>
<td>Strong impact</td>
<td>Weak Impact</td>
</tr>
<tr>
<td>Direct Observation</td>
<td>Weak impact</td>
<td>Strong impact</td>
</tr>
</tbody>
</table>
Methods

• One medical school in Indonesia
• Neurology and Internal Medicine rotations with evaluators trained in using the “mini-CEX”
• Five attributes of high quality feedback were rated for perceived value
• Controlled for by gender and department giving feedback
• Attributes for feedback that were evaluated included:
  • Trainees’ strengths
  • Trainees’ weaknesses
  • Compared performance to a gold standard or expert
  • Explanation of the correct performance
  • Action plan for improvement
Results

• 889 surveys about feedback from 250 medical students-131 on IM and 119 on neurology
• Neurology feedback=higher learning value
• Gender of trainee did not matter as to perceived value
• Feedback that mentioned their weaknesses and compared to a standard perceived as valuable (p<0.01)
• Explaining or demonstrating correct performance and preparing an action plan perceived as most valuable (p<0.001)
• Mentioning strengths did not influence the value of the feedback
Limitations

• 1 Medical school
• Impact on evaluator gender not mentioned
• Positive feedback may have been countered by a culture of modesty and not wanting to stand out
• Systematic training in the use of the mini-CEX for all evaluators may limit generalizability
RAISE YOUR HANDS-
“Solid resident, performing at expected level”
I'm not sure, but I think 'F' stands for 'Effort.'
Mixed Messages or Miscommunication? Investigating the Relationship Between Assessors’ Workplace-Based Assessment Scores and Written Comments
Stefanie S. Sebok-Syer, PhD, Don A. Klinger, PhD, Jonathan Sherbino, MD, MEd, and Teresa M. Chan, MD, MHPE
Acad Med. 2017;92:1774-1779

Purpose: Assess relationships between checklist scores, task ratings, global ratings, and written comments
Assessment Tools from the EM McMAP project-standardized across institutions for Emergency Medicine
• Task checklist
• Task rating-behaviorally anchored
• Global rating-behaviorally anchored
• Comments
Minor Task | Discharge Instructions
Today’s focus is on discharge instructions.

<table>
<thead>
<tr>
<th>Checklist</th>
<th>Done</th>
<th>Done but needs attention</th>
<th>Not done</th>
<th>N/A for case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informs patient of results of any investigations in simple language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informs patient of diagnosis (if possible), other possible diagnoses, and describes prognosis (if possible)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informs patient of care plan (overall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) explains any prescriptions (rationale for use, potential side effects)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) logistics of follow-up (confirm phone number, give consultant contact info, explains how to return for next-day testing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) contingency plan (return instructions, symptoms of serious diagnosis or complication)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensures patient understands diagnosis and care plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs assistance</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>Resident displays any of the below:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>• Significant gaps in discharge instruction (see checklist)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Overly medicalized jargon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Confusing to patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Conflict arose and escalated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resident displays most of the below:</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Inefficient with time</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Used complicated concepts or jargon at times</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Patient’s questions were answered most of the time</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Eventually, arrived at a plan that was amenable to all parties</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Resident displays ALL of the below:</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Time efficient</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Catered to patient’s level of understanding and needs</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Answered questions from patient and/or family</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>• Arrived at a plan that was amenable to all parties easily</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**The evidence:** Please provide an example with an explanation that supports your rating (mandatory):

**The next step:** Based on the above evidence, please give one specific suggestion (educational prescription) for the resident to attempt during his/her next shift. (You do not need to record this.)
**Methods:** Used Logistic regression to determine if other variables

- Task Rating
- Global Rating
- Task Strengths comments-using Ginsberg positive red flag words*
- Task Weaknesses comments-using Ginsberg negative red flag words*
- Task Comment length

Were associated with the choice of DONE, NEEDS ATTENTION on task checklist

---

Results: If “ Needs Attention”

• More likely to comment on weaknesses and conversely, if they chose needs attention, less likely to provide specific strengths in comments.

• They were also more likely to defend and explain judgments by elaborating on what was done well and what needed improving.

• The task rating was the most reliable variable associated with use of the “done, needs attention”

• Comments about what needed attention almost always included “Good” “needs” “consider” “works hard” “remember” and “hardworking”

“Good knowledge. Hardworking. Work on charting legible notes.”
Results for “DONE”

• Comments usually used exclamation marks, capital letters, and comparisons to others.
• Feedback included next steps to achieve higher global level of achievement
Impact

• Task ratings, not global ratings, are associated with the “Done, needs attention”.

• Task comments correlate with “Done, needs attention” if you understand how to interpret them.

• Comments are associated with task ratings not global ratings.

• Tasks, and the comments, are more helpful evaluation tool than global assessments.
24 y.o. female with headache, fever, chills and sweats for 2 days

• Returned from a month long trip to Thailand 3 weeks ago
• Lots of bug bites
• No travel prophylaxis
• Played with Elephants and ate as a local
• Three loose stools the day prior to admission
• History of recurrent UTIs
• Exam shows temp 39.7 HR 116 otherwise normal
Testing

- Chemistries are normal except sodium 130*
- CBC normal, differential normal, no eosinophils
- Flu swab, urinalysis normal
- AST/ALT slightly elevated*
- Thick/thin Smear negative*
- Chest x-ray normal
The most likely diagnosis is:

A. UTI
B. Malaria
C. Enteric fever
D. Melioidosis
E. Leptospirosis
“The integration of knowledge and contextual information (gathered through history, physical, labs, etc) to form a cognitive representation of the problem, with the goal of adjusting that representation with additional information and understanding to reach a judgment or decision (about a diagnosis, risk, or therapy)” OR

The cognitive and physical processes by which a health care professional consciously and unconsciously interacts with a patient and the environment to collect and interpret patient data, weigh the benefits and risks of actions, and understand patient preferences to determine a working diagnostic and therapeutic management plan whose purpose is to improve a patient’s well-being”
What are the components of Clinical Reasoning?

• “Hypothesis generation and refinement, problem representation, data acquisition, interpretation, diagnostic verification, and therapeutic decision making. As part of these steps are the cognitive processes of probabilistic reasoning, causal reasoning, and incorporating EBM”

• “Interplay between individual health care provider, patient, and environmental components”
Introduction: Physician diagnostic errors are a common cause of errors in medicine. Little research in medical students.

Errors are in reasoning are classified as
• ‘System-related’ (technical failures or organizational problems)
• ‘No-fault’ (unusual presentation or patient-related such as deception or poor cooperation)
• ‘Cognitive’ (faulty knowledge, data gathering or synthesis)
Cognitive Errors in Physicians

The majority are Cognitive Errors or Cognitive PLUS Systems errors and are classified as

- Faulty Knowledge
- Faulty Data Gathering
- Faulty Information Processing
- Faulty Verification

Aim:

• To qualitatively understand the cause of diagnostic errors in medical students

Methods:

• 88 German 3-5th year medical students each did eight on-line internal medicine cases (CASUS) with chief complaint of SOB
• Diagnostic knowledge, clinical reasoning and skills (i.e. x-rays, ecg) were assessed by computer navigation and written diagnosis and justifications
Results

• Misdiagnosed 304/704 cases
• All cases demonstrated some faulty data gathering
• Always looked at Chest xray, ecg, and labs but frequently skipped spirometry, micro, and echos.
### Table 2 Diagnostic errors of medical students

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge base inadequate</td>
<td>16% (49/304)</td>
</tr>
<tr>
<td>Skills inadequate</td>
<td>24% (75/304)</td>
</tr>
<tr>
<td>Faulty context generation</td>
<td>15% (47/304)</td>
</tr>
<tr>
<td>Overestimating/underestimating</td>
<td>9% (28/304)</td>
</tr>
<tr>
<td>Faulty triggering</td>
<td>12% (35/304)</td>
</tr>
<tr>
<td>Misidentification</td>
<td>10% (30/304)</td>
</tr>
<tr>
<td>Premature closure</td>
<td>10% (29/304)</td>
</tr>
<tr>
<td>Cluelessness</td>
<td>3% (9/304)</td>
</tr>
</tbody>
</table>

### Table 3 Most common diagnostic errors with respect to the different cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Total number of errors (of 88)</th>
<th>Most common error</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuberculosis</td>
<td>51</td>
<td>Faulty knowledge</td>
<td>45% (23/51)</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>45</td>
<td>Faulty skills</td>
<td>42% (19/45)</td>
</tr>
<tr>
<td>Myocarditis</td>
<td>28</td>
<td>Faulty context generation, Misidentification</td>
<td>36% each (10/28)</td>
</tr>
<tr>
<td>Valvular aortic stenosis</td>
<td>3</td>
<td>Faulty context generation</td>
<td>66% (2/3)</td>
</tr>
<tr>
<td>COPD</td>
<td>32</td>
<td>Faulty skills</td>
<td>47% (15/32)</td>
</tr>
<tr>
<td>Uremia</td>
<td>33</td>
<td>Faulty context generation, Overestimating</td>
<td>30% each (10/33)</td>
</tr>
<tr>
<td>Hyperventilation</td>
<td>43</td>
<td>Misidentification</td>
<td>30% (13/43)</td>
</tr>
<tr>
<td>AVNRT</td>
<td>72</td>
<td>Faulty skills</td>
<td>38% (27/72)</td>
</tr>
</tbody>
</table>
Compared to physicians

• Knowledge is a bigger problem in students
• Improved clinical skills such as ECG, chest x-ray interpretation may help diagnostic accuracy
• Premature closure happens more when expertise increases OR with poorer knowledge
• In students premature closure happens with non-specific information (i.e. OCPs, Smoking, etc)

  • Faulty knowledge $\rightarrow$ conceals premature closure
  • But since premature closure continues in experts, it isn’t clear if this is due to knowledge deficits or another reason (heuristic biases)
Impact

• Knowledge plays greater role in clinical reasoning in students
• More obscure cases lead to more errors and they are more knowledge based
• More practice with clinical skills may improve diagnostic accuracy
• Faulty synthesis is usually due to faulty context generation in students—Don’t realize something is important in this particular clinical scenario
• Ultimately, context generation and premature closure in students may be a function of lack of knowledge
Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians

Eleixs McBee, Temple Ratcliffe, Katherine Picho, Lambert Schuwirth, Anthony R. Artino Jr., Ana Monica Yepes-Rios, Jennifer Masel, Cees van der Vleuten and Steven J. Durning
Physician factors include both cognitive and contextual factors.

Cognitive load, knowledge, experience, dual-process theory, self-regulation, motivation and emotion, scripts, sleepiness, well-being

Physician factors

Patient factors

Clinical reasoning

Practice factors

Appointment length, ambulatory or inpatient setting, staffing

Authenticity, acuity of illness, common versus rare presentation

Clarifying Assumptions to Enhance Our Understanding and Assessment of Clinical Reasoning

Steven J. Durning, MD, PhD, Anthony R. Artino, Jr, PhD, Lambert Schuwirth, MD, PhD, and Cees van der Vleuten, PhD

Academic Medicine, Vol. 88, No. 4 / April 2013
AIM: To understand

1) The relationship between diagnostic and therapeutic reasoning
2) The impact of patient and physician contextual factors on performance
3) Are more experienced physicians more or less impacted by contextual factors?
Methods

• Board certified internist and internal medicine residents
• Physician factors such as Sleepiness and burnout assessed with survey
• Three videos (3-5min) in length same actors with straightforward disease presentations but varied the context
  1. ESL in a straightforward HIV case
  2. Emotional lability in patient with signs and symptoms of colon cancer
  3. Emotional lability and low English proficiency in patient with Diabetes
• 15 Internists (12 males, 3 females)
• 10 residents (5 males, 5 females)
• IRR for the Post Encounter form was 0.93
• Residents were sleepier and more burned out
• Despite this there were no differences in scores for differential diagnosis, leading diagnosis, supporting data, or treatment between residents and internists
• No correlation between sleepiness or burnout and lead diagnosis or treatment score

• Supporting data was better for internists than for residents for both treatment and lead diagnosis

<table>
<thead>
<tr>
<th>Case</th>
<th>Role</th>
<th>Correct Diagnosis (%)</th>
<th>Correct Therapy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV Case</td>
<td>Resident</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>(Spanish Speaking)</td>
<td>Internist</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>Colon Cancer case</td>
<td>Resident</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>(Emotional labile)</td>
<td>Internist</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Diabetes case</td>
<td>Resident</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>(Emotional labile AND</td>
<td>Internist</td>
<td>53</td>
<td>67</td>
</tr>
<tr>
<td>low English proficiency)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Impact

• Weak correlation between lead diagnosis and therapeutic plan
• Board certified Internists are not any better than residents when contextual factors are in place
• Impact of contextual factors is important
• Contextual factors have significant impact on safety and so this may also be important to understand how it impacts diagnosis
• Contextual factors may impact therapeutic plans but differently than diagnostic plans
Back to the 24 year old with flea bites from country with endemic typhus-

A. UTI-urine negative and no flank pain
B. Malaria-thin/thick negative, no anemia, less likely
C. **Enteric (typhoid) fever-transmitted fecal oral with contaminated water**
D. Melioidosis-usually cellulitis or pneumonia
E. Leptospirosis-unlikely without rash, may have conjunctival suffusion
ARE MULTIPLE CHOICE EXAMS AN ACCURATE MEASURE OF ONE'S KNOWLEDGE?

A. YES  
B. A AND C  
C. A AND B  
D. ALL OF THE ABOVE
HPI: 45 year old male body-builder with increasing thirst, urination, and headache comes to clinic. He has had to purchase larger shoes and enlarge his wedding ring. Facial features are much coarser now than on driver’s license picture taken 3 years prior.

Physical exam: blood pressure is 130/80
- Hypertrophic bone over the frontal area and eye brows, spacing of the teeth, and large lower jaw
- Vision is grossly normal but there appears to be a visual field deficit on the left lower visual field
- Hands and feet are both large
The most likely diagnosis is:
What if I gave you choices?

A. Diabetes type 2
B. Hypothyroid
C. Acromegaly
D. Pituitary necrosis
E. Exogenous use of growth hormone
Negative and Positive Cueing

• You might not have remembered the word acromegaly and seeing and picking it is “Positive Cueing”

• Or you might be distracted by the exogenous growth hormones in this body builder and picking it is “Negative Cueing”

• Single best answer(multiple choice) does both and may not feel authentic since doctors don’t get to do multiple choice with patients

**AIM:** Compare performance of medical students on multiple choice (MCT) and very short answer questions
Results

• 155 medical students took 60 questions Very Short Answer (VSAT) and then MCT

• 144 medical students took 60 questions MCT and then very short answer test (VSAT)

• The two tests had positive correlation (r=0.83, p<0.001)

• VSAT had high reliability (0.91) across groups

• Lower Standard Error Mean than MCT’s (p<0.001)

• VSAT Higher mean item-total point bi-serial correlation (p<0.001)

• The VSATs had lower scores (p<0.001)
Results/Student perceptions

• **Positive Cueing:**
  • VSAT first and then MCT second, 19.2% of questions
  • If MCT first and then VSAQ, 7.5% of the questions

• **Negative Cueing:** 3.5% in both
  • For 1/5 of the items, over 30% of the students who did VSAT first could only get the answers when given them in multiple choice

• MCTs are easier
• VSATs are more authentic
• Learning strategies may need to change and spelling needs to improve
Limitations and Impact

- Natural language processing software
- Teach the system the right answers
- The learners were all one stage (year 3) and VSAQ may not be as acceptable to all stages of learners
- Not sure how it translates to performance on boards
- Needs to be studied in different stage learners to see if reliability, discrimination, and acceptability are similar
“Rapid pulse, sweating, shallow breathing ... According to the computer, you’ve got gallstones.”
**Background:** Geographic barriers and teaching methods like the flipped classroom make the use of podcasts or video-podcasts useful. (VOD-CAST)

- Software can be interactive and include quizzes that allow the learner to interact (Articulate)
- Software can be purely Power point slides with audio and no interaction

**Aim:** Determine medical students’ preferences for Vodcast features and use those findings to guide improvements
Results

- 99% response rate MS1, 64% MS2, 47% MS3
- Male and female, age response rates similar
- Most highly valued attributes are:
  - Clear explanations*
  - High-yield for boards
  - Ability to control speed*
  - Well organized
  - Concise*
  - Relevant to clinical application
  - Practice questions

*well described features of high quality on-line learning in other environments

Preferred length is 27-28 minutes
Use breaks if longer!
Unhelpful

• Music
• Objects moving on the screen
• Table of Contents
• Suggested Readings

Pet Peeves:
• Poor sound quality
• Subtheme ambient noise
• Poor image quality
• Can’t control speed
• Too long
• Overly cluttered slides

• Interactive questions more valuable to MS1s than older students
• Practice questions highly valuable to all but essential to MS1s
Summary and Impact-Weed, Signal, and Segment

- Weed
- Signal
- Segment

- Use non-flash software, without interactive questions, for advanced learners so that they can speed up videos/sound

- Use flash software, with interactive questions for earlier learners
Panel management is a CQI strategy to improve the health of patients.

AIM: To assess the trainee reported outcomes of a novel PPM curriculum including:

- confidence in using PPM tools (mainly registries and EHR)
- ability to work in inter-professional teams to overcome care gaps
- understanding of system resources for chronic disease management
Methods

• EBM curriculum in PPM that involved the following:
  • Used existing clinical teams
  • Disease registries and identified care gaps
  • With resident leading group, developed and implemented patient focused action plans during quarterly small group sessions

• Measure team members’ perceptions of curricular content, confidence in PPM, and likelihood of using techniques in the future on Likert scale on post intervention surveys
Curriculum

• 168 participants
• All learners rated the curriculum highly (over 4/5 for all learners)
• Highly interactive, resident leading team
• Took about 2 hours and 15 minutes each quarter
Outcomes/Results

• Confidence rated as a 3.7 (0, not at all confident, 5 Very much confidence)

• Major take home lessons from learners on qualitative review:
  • PPM good practice. “Good to routinely look at list of patients to see who has fallen through the cracks”
  • Use of systems resources considered valuable
    ➢ Registries
    ➢ Consultative services
    ➢ Team members for outreach and care coordination
  • Learned something new clinically, i.e. preferred treatment algorithms.
Impact

• Curriculum seems transferable and doable
• The use of panels and registries is becoming more accessible
• Could medical students use these same techniques to improve care or could students be part of the team?
• Would be nice to see impact on patient care or physician metrics...
• And could empowering residents with these skills improve the numbers of residents choosing primary care
“Turn the pain into power.”
SUBJECT: E/M Service Documentation Provided by Students (Manual Update)

I. SUMMARY OF CHANGES: This revision to Pub. 100-04, Medicare Claims Processing Manual, Chapter 12, Section 100.1.1, B. E/M Documentation Provided by Students, allows the teaching physician to verify in the medical record any student documentation of components of E/M services, rather than re-documenting the work. (Manual Update Only)

EFFECTIVE DATE: January 1, 2018
*Unless otherwise specified, the effective date is the date of service.
IMPLEMENTATION DATE: March 5, 2018
WELL DUH!!
• Interns lack the EHR skills necessary to provide high quality, safe care
• The use of the HER is a critical skill for medical students
• Access to EHRs is inconsistent, frequently read-only and varies by student placement
• Therefore, the lack of real access to the EHR impacts skill development for all.
Lack of Real Access is due to

• An overly rigid interpretation of HIPPA
• And perceived legal liability related to the risk of student documentation.
• So even if student notes are endorsed by CMS will our compliance departments endorse them?
• Will they endorse other, perhaps more meaningful access?
• Finally, FIVE EHRs account for the majority of market share in academic centers.
• Are there best practices that could be adopted by these major players in the HERs that would benefit student engagement and learning?
Future Directions

- National organizations should develop language for hospital compliance departments that addresses:
  - The legal ramifications of students in the EMR
  - The benefits to systems such as
    - Safer care by residents who have had accelerated training as students
    - More thorough documentation
    - Less holes in the swiss cheese for patient error
    - More team members doing population health management
Thank you!

Happy Birthday, Marty!
**Background:** The use of virtual patients before or after classroom lectures may improve knowledge retention. Not clear which is better

**Inductive learning:** Use virtual patient before to stimulate inquisitiveness from a practical situation. Then learn pathology, disease presentations in a lecture

**Deductive learning:** Learn in a lecture and be able to transfer that knowledge to a patient.
Results

• This study, set in a Saudi Arabian Dental school, showed that it doesn’t really matter.
• No difference for immediate, retained, or transfer of knowledge to a virtual patient
• And there is no difference between men and women.
• For me, I think it is clear that letting a learner choose which way to use the SIMPLE or CLIPP cases is fine.
• Either before or after our didactics will likely work the same!
An annual survey of family medicine clerkship directors was done to quantify the requirements for ambulatory/community based faculty development, delivery methods, and barriers.

AIM: Examine the relationship between the requirement for faculty development and payment of preceptors, financial support for CD, and the number and length of face-to-face faculty development.
Results

- 125 U.S. and 16 Canadian Medical Schools family medicine clerkship directors with a 79% response rate
- 89% of Schools of medicine offer FD, but only 16% require it.
- The types of faculty development include:
  - Personalized Feedback
  - Face to face sessions
  - On-line curriculum
  - Targeted articles on teaching
  - AV materials for independent study
  - Listservs
  - Podcasts
Results

• Of those with face-to-face option, 47% reported spending 30-60min
• 28% of U.S. Medical Schools pay community preceptors versus 100% of Canadian Medical schools.
• Paying preceptors (p<0.05) and face-to-face options(p<0.01)were correlated with requirements to participate in FD
• Barriers to faculty development included time, preceptor availability and time, geographic distribution, financial resources, competing learners, faculty with skill to teach the teachers
Impact

- Need to explore whether requirements for faculty development should be linked to payment
- 3% of clerkship directors reported tax incentives for teaching
- More robust needs assessments for what faculty desire