THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Validity Studies on the Assessment of Clinical Reasoning Skills Using Patient Notes

University of Tokyo

Medical Education Lecture

Yoon Soo Park, PhD

September 21, 2017

Overview

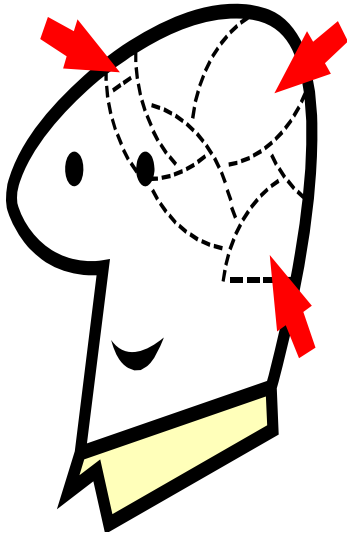
1. Overview of Clinical Reasoning: Diagnostic Justification

2. USMLE Step 2 Clinical Skills Examination

3. Validity Studies of Patient Note Scoring Rubric
   - Five Projects (Years 2012 to 2017)
   - Gather Validity Evidence

4. Implications for Validity Research

Clinical Reasoning (1)



Data Gathering

History

Physical Examination

Lab Results

Differential Diagnosis

Working Diagnosis

Plan

Diagnoses listed –
Not always relate to actual findings!

Finding 1

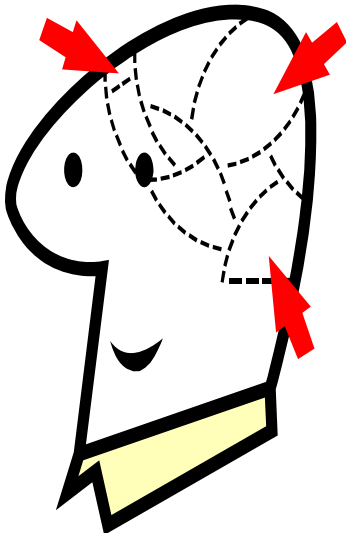
Finding 2

Finding 3

Finding 4

Finding 5

?????

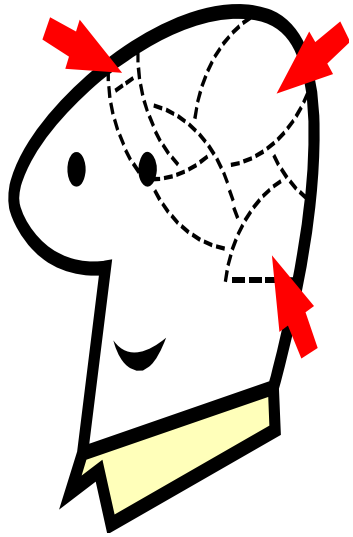


Diagnosis A

Diagnosis B

Diagnosis C

Clinical Reasoning (2)

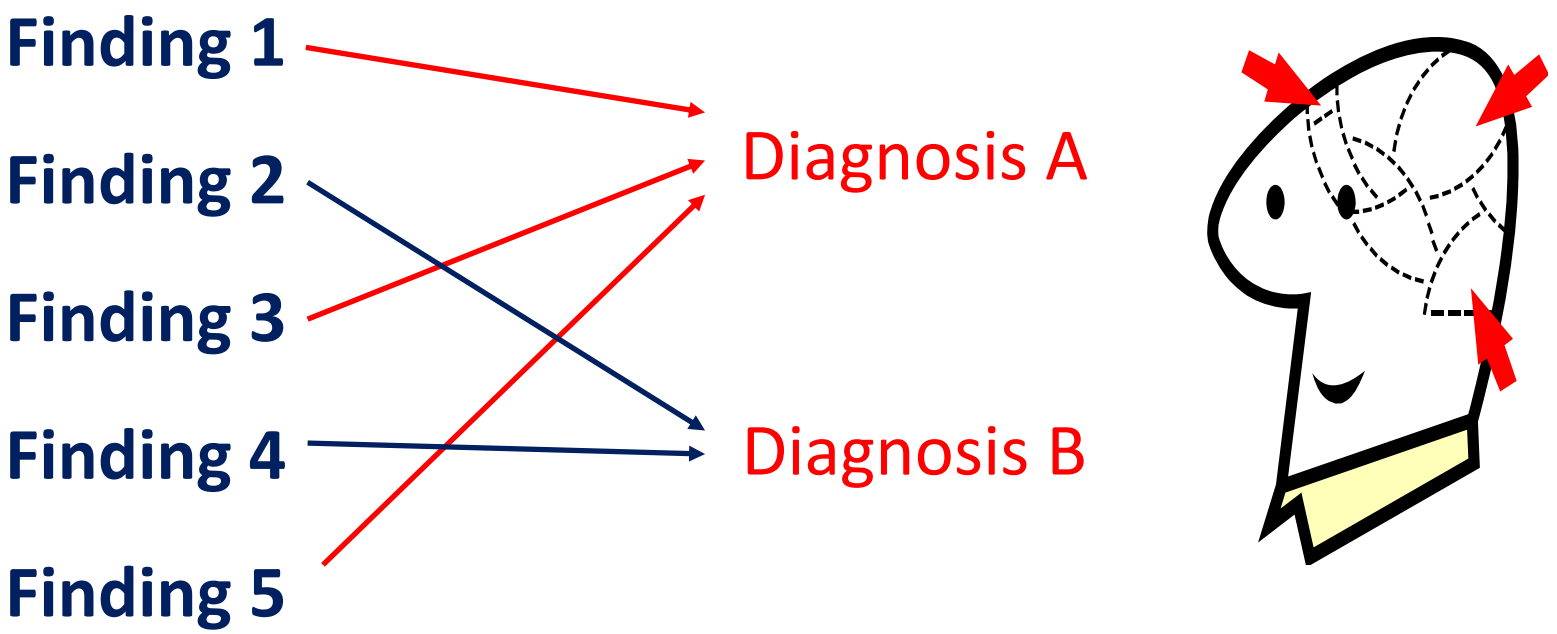


- Organize and use knowledge
- Collect pertinent data about patients
- Generate appropriate DDx given chief complaint
- Make good diagnostic decisions using data

Diagnostic Justification

1. How to assess?
2. Validity?

Clinical Reasoning –
Diagnostic Justification



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Diagnostic Justification Study – 4th year Medical Students

Examining the Diagnostic Justification Abilities of Fourth-Year Medical Students

Reed G. Williams, PhD, and Debra L. Klamen, MD, MHPE

Abstract

Purpose
Fostering ability to organize and use medical knowledge to guide data collection, make diagnostic decisions, and defend those decisions is at the heart of medical training. However, these abilities are not systematically examined prior to graduation. This study examined diagnostic justification (DXJ) ability of medical students shortly before graduation.

Method
All senior medical students in the Classes of 2011 (n = 67) and 2012 (n = 70) at Southern Illinois University were required to take and pass a 14-case, standardized patient examination prior to graduation. For nine cases, students were required to write a free-text response indicating how they used patient data to move from their differential to their final diagnosis. Two physicians graded each DXJ response. DXJ scores were compared with traditional standardized patient examination (SCCX) scores.

Results
The average intraclass correlation between raters' rankings of DXJ responses was 0.75 and 0.64 for the Classes of 2011 and 2012, respectively. Student DXJ scores were consistent across the nine cases. Using SCCX and DXJ scores led to the same pass–fail decision in a majority of cases. However, there were many cases where discrepancies occurred. In a majority of those cases, students would fail using the DXJ score but pass using the SCCX score. Common DXJ errors are described.

Conclusions
Commonly used standardized patient examination component scores (history/physical examination checklist score, findings, differential diagnosis, diagnosis) are not direct, comprehensive measures of DXJ ability. Critical deficiencies in DXJ abilities may thus go undiscovered.

Williams and Klamen, *Academic Medicine*, 2012

Williams and Klamen Study

- Case specificity
  - Justification scores variable across cases
- Task specificity
- >20% of students had borderline or poor Diagnostic justification on > 50% of cases

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

USMLE Step 2 Clinical Skills

Patient Note Format

Old Format: **<u>Before</u> June 2012**	**New Format:** **<u>After</u> June 2012**
1. Document H & PE	1. Document H & PE
2. DDX →	2. **Ranked and Justified** DDX
3. Plan for immediate workup	3. Plan for immediate workup

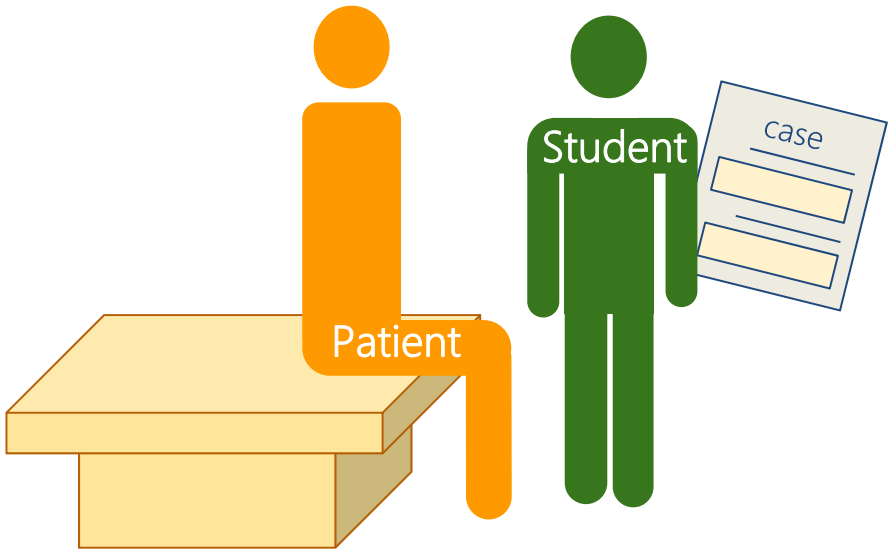
THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

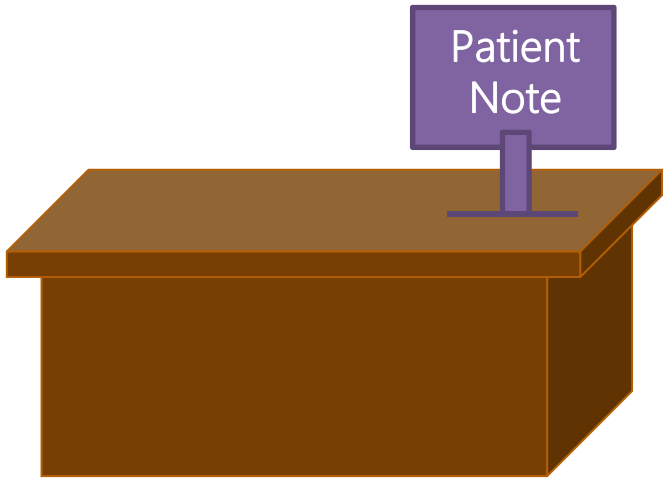
USMLE Step 2 Clinical Skills (1)

Repeat for
12 Stations

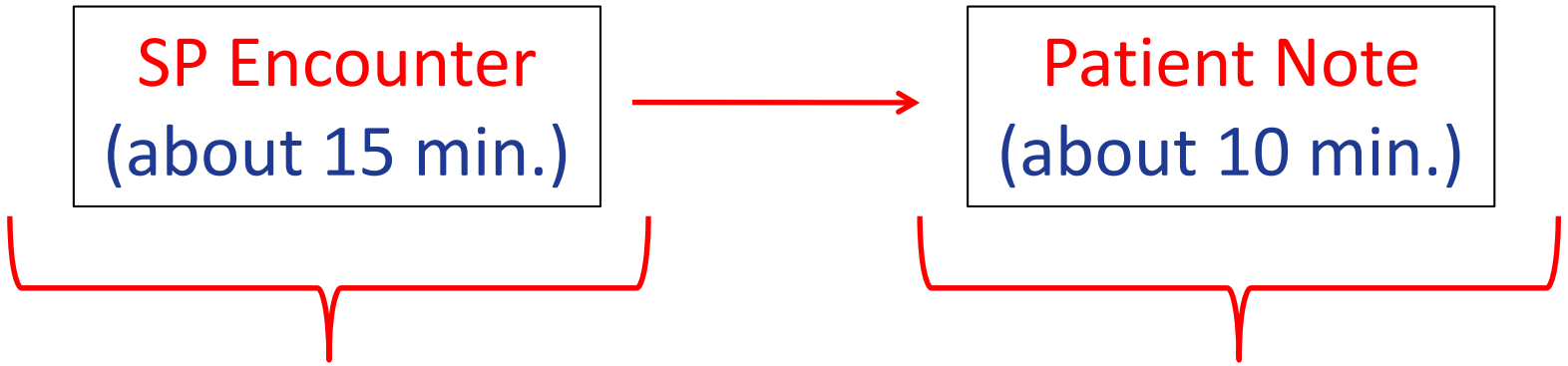
Standardized Patient
Encounter
(15 minutes)

Patient Note
Writing
(10 minutes)





USMLE Step 2 Clinical Skills (2)



Scoring: Standardized Patients

- Physical Examination

Scoring: Physician Raters

1. Documentation
2. Differential Diagnosis
3. Diagnostic Justification
4. Workup Plan

- Medical schools → align curriculum and assessment
  - Develop scoring rubrics
  - Train faculty raters to score PNs

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

USMLE Patient Note Form (1)

CLINICAL SKILLS EVALUATION
PATIENT NOTE

HISTORY: Describe the history you just obtained from this patient. Include only information (pertinent positives and negatives) relevant to this patient's problem(s).

PHYSICAL EXAMINATION: Describe any positive and negative findings relevant to this patient's problem(s). Be careful to include *only* those parts of examination you performed in *this* encounter.

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

USMLE Patient Note Form (2)

DIAGNOSIS #1:

HISTORY FINDING(S)	PHYSICAL EXAM FINDING(S)

(+) Click to add row(s)

DIAGNOSIS #2:

HISTORY FINDING(S)	PHYSICAL EXAM FINDING(S)

(+) Click to add row(s)

DIAGNOSIS #3:

HISTORY FINDING(S)	PHYSICAL EXAM FINDING(S)

(+) Click to add row(s)

DIAGNOSTIC STUDIES

(+) Click to add row(s)

Examinee Name	Jane Doe		
USMLE ID	0-123-456-7	Badge ID	2
Room Number	2	Encounter	1

HISTORY: Describe the history you just obtained from this patient. Include only information (pertinent positives and negatives) relevant to this patient's problem(s).

- 75-yo woman, sudden onset back pain while lifting turkey from oven.
- Pain in midline, mid-thoracic
- PMH: fracture of foot bone last year
- Meds: Acetaminophen for occasional headache …

Lines: 1 / 15 **Characters:** 0 / 950

PHYSICAL EXAMINATION: Describe any positive and negative findings relevant to this patient's problem(s). Be careful to include *only* those parts of examination you performed in *this* encounter.

- Well-nourished woman looks her age, alert and lucid;
- Appears in pain, especially with movements;
- Point tenderness over T8;
- Straight leg test negative; sensation intact …

Lines: 1 / 15 **Characters:** 0 / 950

DATA INTERPRETATION: *Based on what you have learned from the history and the physical examination*, list up to 3 diagnoses that might explain this patient's complaint(s). List your diagnoses from most to least likely. For some cases, fewer than 3 diagnoses will be appropriate. Then, enter the positive or negative findings from the history and the physical examination (if present) that support each diagnosis. Lastly, list initial *diagnostic* studies (if any) you would order for each listed diagnosis (e.g. restricted physical exam maneuvers, laboratory tests, imaging, ECG, etc.)

Diagnosis #1

Pulled Back Muscle

History Finding(s)	Physical Exam Finding(s)
Sudden onset back pain with lifting	Range of motion limited by pain
Pain worse with movement	

Add a Row

Diagnostic Study/Studies

Imaging of back/spine

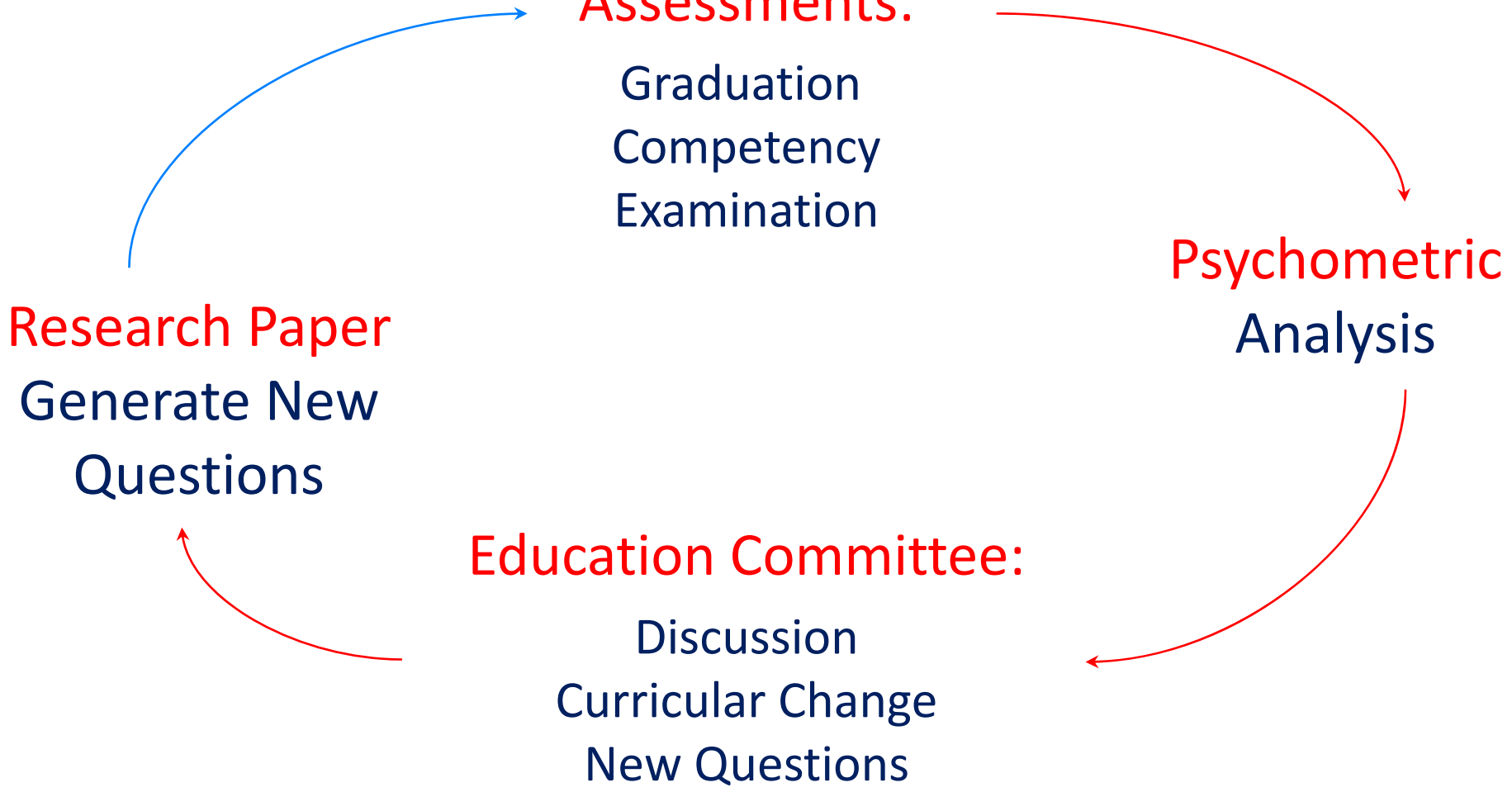
Add a Row

Five Sources of Validity Evidence (AERA, APA, & NCME, 2014)

- Content
  - Content sampled, blueprint
- Response process
  - Response, Rater Issues
- Internal structure
  - Reliability, Psychometrics
- Relations to other variables
  - Correlations between assessment score to other measures
- Consequences
  - Impact on pass-fail rates, curriculum

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Education and Research Cycle:
University of Illinois – College of Medicine



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Quality Improvement
Accreditation Agency

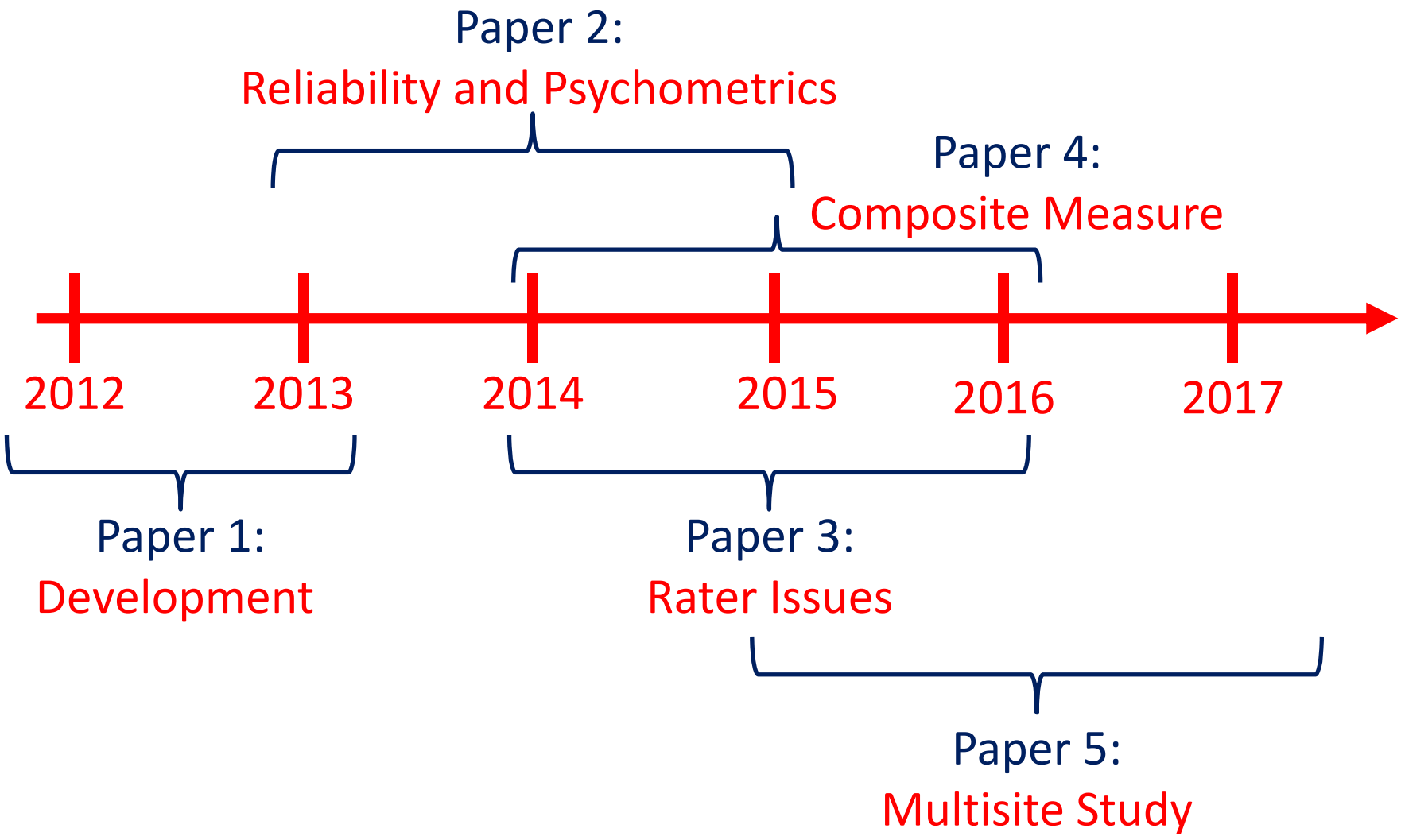
- 1.1 Continuous Quality Improvement
  - *A medical school engages in ongoing planning and continuous improvement processes … achievement of measureable outcomes …*

- 8.4 Program Evaluation
  - *A medical school collects and uses a variety of outcome data, including national norms of accomplishment, to demonstrate the extent to which medical students are achieving medical education program objectives and to enhance medical education program quality.*

Standards for Accreditation of Medical Education Programs Leading to the MD Degree,
Liaison Committee on Medical Education (LCME), 2016

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Validity Research: Patient Note Rubric



THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Validity Research Agenda

(Sireci 2013; Kane 2013; Messick 1995)

Project	Sources of Validity Evidence				
	Content	Internal structure	Relations with other variables	Response Process	Consequences
Paper 1 (2012-2013)	X	X	X	X	X
Paper 2 (2013-2015)		X			
Paper 3 (2014-2016)		X		X	
Paper 4 (2015-2016)		X			X
Paper 5 (2015-2017)			X		X

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Paper 1: Initial Development

(Project Dates: 2012-2013)

Validity Evidence for a Patient Note Scoring Rubric Based on the New Patient Note Format of the United States Medical Licensing Examination

Yoon Soo Park, PhD, Matthew Lineberry, PhD, Abbas Hyderi, MD, MPH, Georges Bordage, MD, PhD, Janet Riddle, MD, and Rachel Yudkowsky, MD, MHPE

Abstract

Purpose
This study examines validity evidence for the Patient Note Scoring Rubric, which was developed for a local graduation competency exam (GCE) to assess patient notes written in the new United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills format. The rubric was designed to measure three dimensions: Documentation, justified differential diagnosis (DDX), and Workup.

Method
Analyses used GCE data from 170 fourth-year medical students who completed five standardized patient (SP) cases in May 2012. Five physician raters each scored all responses for one case. Internal structure was examined using correlations between dimensions and between cases; a generalizability study was also conducted. Relationship to other variables was examined by correlating patient note scores with SP encounter scores. Consequence was assessed by comparing pass–fail rates between the rubric and the previous global rating. Response process was examined using rater feedback.

Results
Correlations between scores from different dimensions ranged between 0.33 and 0.44. Reliability of scores based on the phi coefficient was 0.43; 15 cases were required to reach a phi coefficient of 0.70. Evidence of case specificity was found. Documentation scores were moderately correlated with SP scores for data gathering ($r = 0.47$, $P < .001$). There was no meaningful change in pass–fail rates. Raters' feedback indicated that they required more training for scoring the DDX and Workup dimensions.

Conclusions
There is initial validity evidence for use of this rubric to score local clinical exams that are based on the new USMLE patient note format.

Park et al, *Academic Medicine*, 2013

Task with Description	Score and Anchor
1. Documentation: Documentation of findings in history (Hx) and physical examination (PE) (30 points)	1. Key Hx and PE findings are missing or incorrect 2. Some key positive findings present but poorly documented or disorganized or missing pertinent negatives 3. Most key positive findings well documented and organized, may miss a few pertinent negatives 4. All key information present, concise and well organized with little irrelevant information
2. DDX: Differential Diagnosis (30 points)	1. [0-1 of 3] or [0 of 2] of the correct diagnoses listed 2. [2 of 3] or [1 of 2] of the correct diagnoses listed, in any order 3. All diagnoses listed, incorrect rank order 4. All diagnoses listed and correctly rank ordered
3. Justification: Justification of Differential Diagnosis (30 points)	1. No justification provided OR many missing or incorrect links between findings and Dx 2. Some missing or incorrect links between findings and Dx 3. Only a few missing or incorrect attributions, which would not impact Dx 4. Links to diagnoses are correct and complete
4. Workup: Plan for Immediate Diagnostic Workup (10 points)	1. Diagnostic workup places patient in unnecessary risk or danger 2. Ineffective plan for diagnostic workup – essential tests missed, irrelevant tests included 3. Reasonable plan for diagnostic workup, may have some unnecessary tests or missing few essential tests 4. Plan for diagnostic workup is effective and efficient, includes all essential tests, and few or no unnecessary tests

Literature Review and Interview/Focus Groups

Literature Review

- Determine if construct already exist
- Alignment between construct and theory

Interview / Focus Group

Conceptualization of Construct → match? → Prospective Learners

- As described in the literature?
- What language do learners use?

Conduct Cognitive Interviews

Read and paraphrase it –

"tell me what you think it is asking, in your own words"

Ask more focused questions to evaluate if respondents agree

- Observable?
- Relevant?
- Wording clear?
- Responses clear?
- Other?

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Pilot Study – Task Score: Documentation

1. Documentation of Key Hx and PE (30 points)

7%	1. Key Hx and PE findings are missing or incorrect
27%	2. Some key positive findings present but poorly documented or disorganized or missing pertinent negatives
57%	3. Most key positive findings well documented and organized, may miss a few pertinent negatives
9%	4. All key information present, concise and well organized with little irrelevant info

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Pilot Study – Task Score: Differential Diagnosis

2. Differential Diagnosis (30 points)

11%	1. [0-1 of 3] or [0 of 2] of the correct diagnoses listed
34%	2. [2 of 3] or [1 of 2] of the correct diagnoses listed, in any order
29%	3. All diagnoses listed, incorrect rank order
25%	4. All diagnoses listed and correctly rank ordered

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Pilot Study – Task Score: Justification

3. Justification (30 points)

6%	1. No justification provided OR many missing or incorrect links between findings and Dx
34%	2. Some missing or incorrect links between findings and Dx
55%	3. Only a few missing or incorrect attributions, which would not impact Dx
5%	4. Links to diagnoses are correct and complete

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Pilot Study – Task Score: Workup

4. Plan for Immediate Dx Workup (10 points)

7%	1. Diagnostic workup places patient in unnecessary risk or danger
54%	2. Ineffective plan for diagnostic workup – essential tests missed, irrelevant tests included
36%	3. Reasonable plan for diagnostic workup, may have some unnecessary tests or missing few essential tests
3%	4. Plan for diagnostic workup is effective and efficient, includes all essential tests, and few or no unnecessary tests.

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

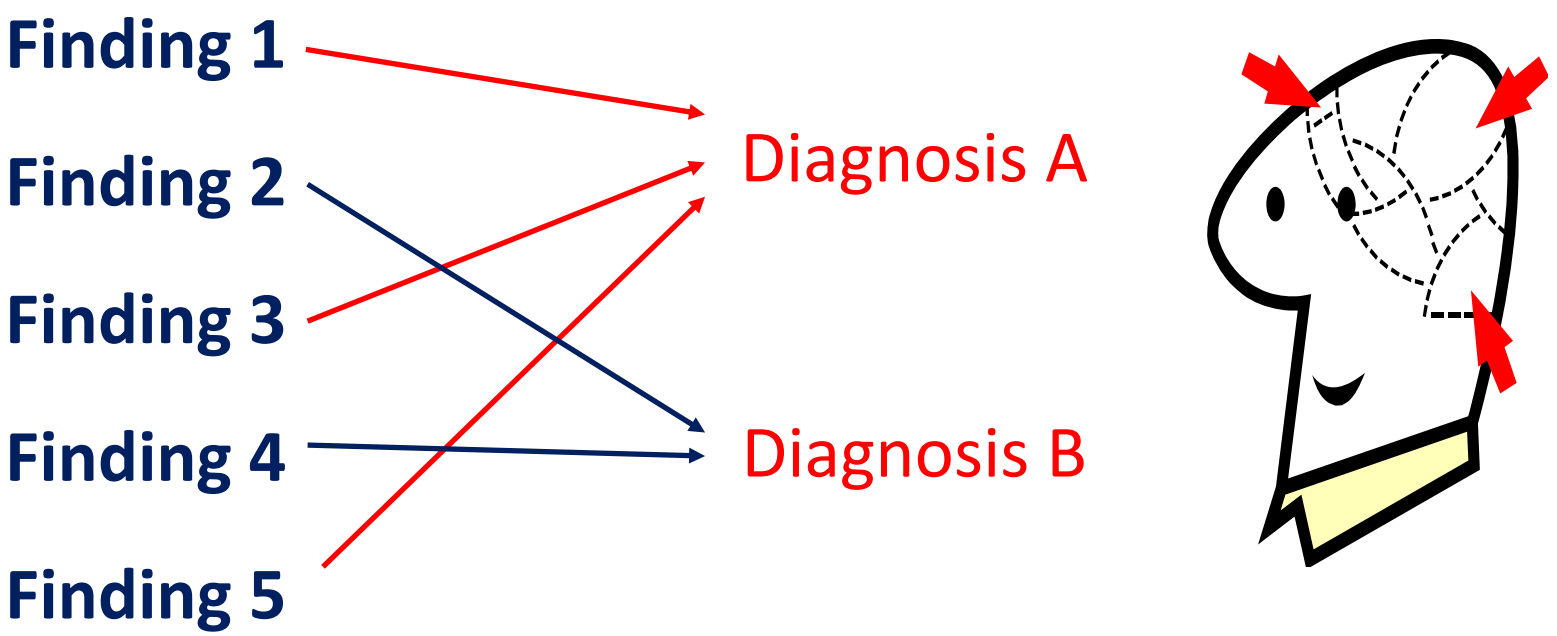
Students who received scores of "Poor" or "Borderline" on Justification

# Cases	%
0	9
1	31
2	30
3	21
4	8
5	1

Consistent with findings from Williams Study –

>20% of students had **borderline or poor Diagnostic justification** on >50% of cases.

Students need more practice linking findings and diagnoses!



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Paper 2: Reliability and Psychometrics

(Project Years: 2013-2015)

Characteristics and Implications of Diagnostic Justification Scores Based on the New Patient Note Format of the USMLE Step 2 CS Exam

Rachel Yudkowsky, MD, MHPE, Yoon Soo Park, PhD, Abbas Hyderi, MD, MPH, and Georges Bordage, MD, PhD

Abstract

Background
To determine the psychometric characteristics of diagnostic justification scores based on the patient note format of the United States Medical Licensing Examination Step 2 Clinical Skills exam, which requires students to document history and physical findings, differential diagnoses, diagnostic justification, and plan for immediate workup.

Method
End-of-third-year medical students at one institution wrote notes for five standardized patient cases in May 2013 (n = 180) and 2014 (n = 177). Each case was scored using a four-point rubric to rate each of the four note components. Descriptive statistics and item analyses were computed and a generalizability study done.

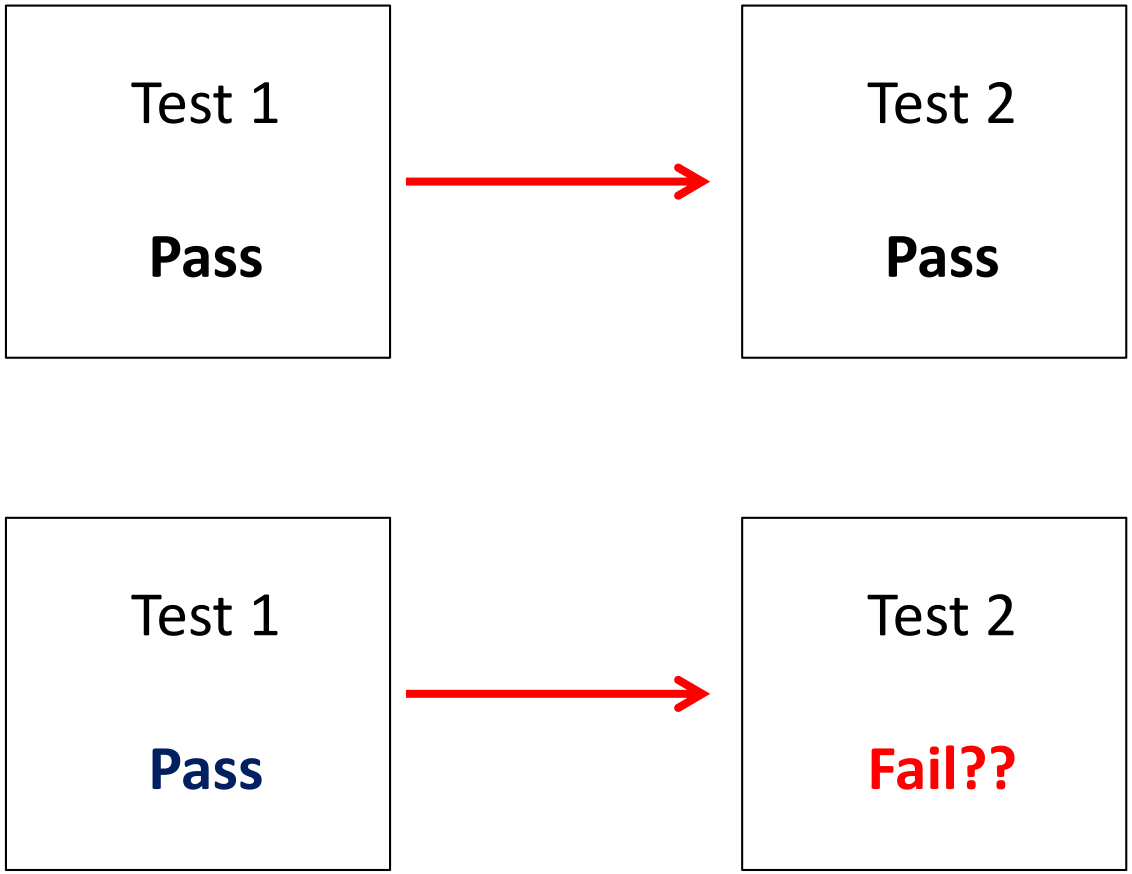
Results
Across cases, 10% to 48% provided no diagnostic justification or had several missing or incorrect links between history and physical findings and diagnoses. The average intercase correlation for justification scores ranged from 0.06 to 0.16; internal consistency reliability of justification scores (coefficient alpha across cases) was 0.38. Overall, justification scores had the highest mean item discrimination across cases. The generalizability study showed that person–case interaction (12%) and task–case interaction (13%) had the largest variance components, indicating substantial case specificity.

Conclusions
The diagnostic justification task provides unique information about student achievement and curricular gaps. Students struggled to correctly justify their diagnoses; performance was highly case specific. Diagnostic justification was the most discriminating element of the patient note and had the greatest variability in student performance across cases. The curriculum should provide a wide range of clinical cases and emphasize recognition and interpretation of clinically discriminating findings to promote the development of clinical reasoning skills.

Yudkowsky et al, *Academic Medicine*, 2015

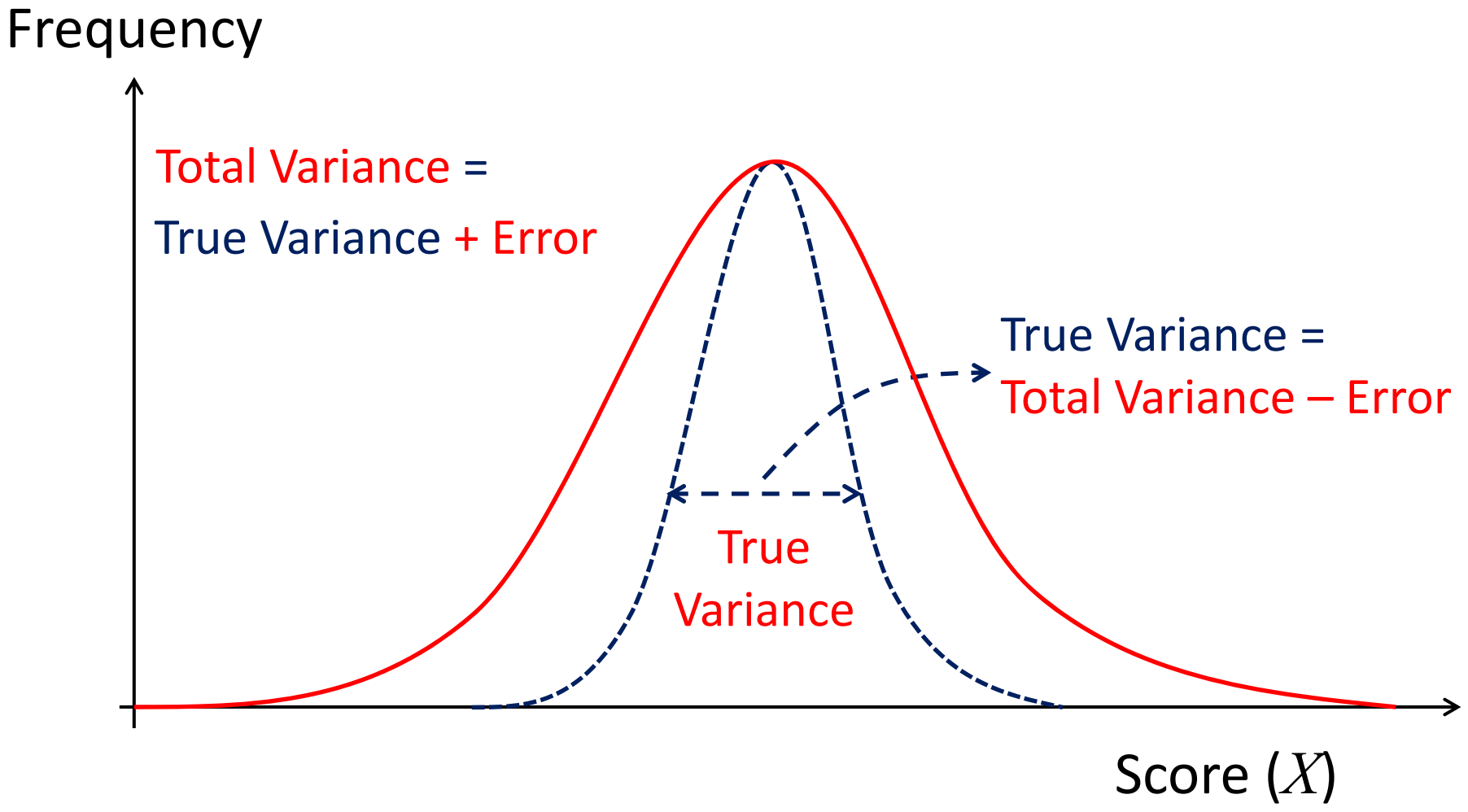
THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Reliability as Consistency



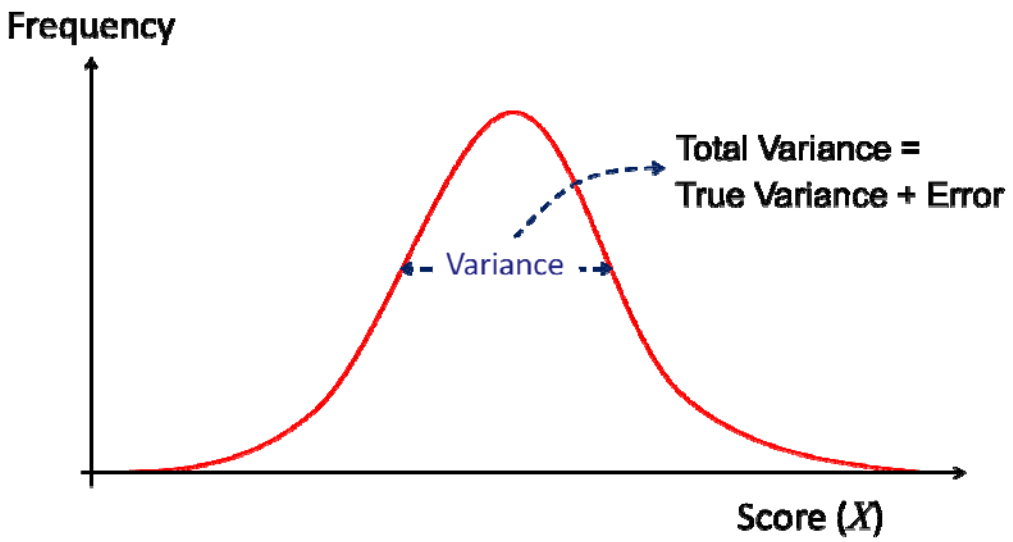
THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

$$\text{Reliability} = \frac{\text{"true" variance}}{\text{"true" variance + error}}$$



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Reliability: Example



$$reliability = \frac{TrueVariance}{TotalVariance} = \frac{True}{True + Error}$$

Example (Note: Hypothetical!)

Test Mean = 65
Standard Deviation = 5
Total Variance = 25

Total (True + Error) = 25
→ True = 15, Error = 10

Reliability =

$$\frac{15}{25} = \frac{15}{15 + 10} = .60$$

Internal Structure: Reliability

- Reliability with 5 cases
  - G-coefficient: .47
  - Φ-coefficient: .43

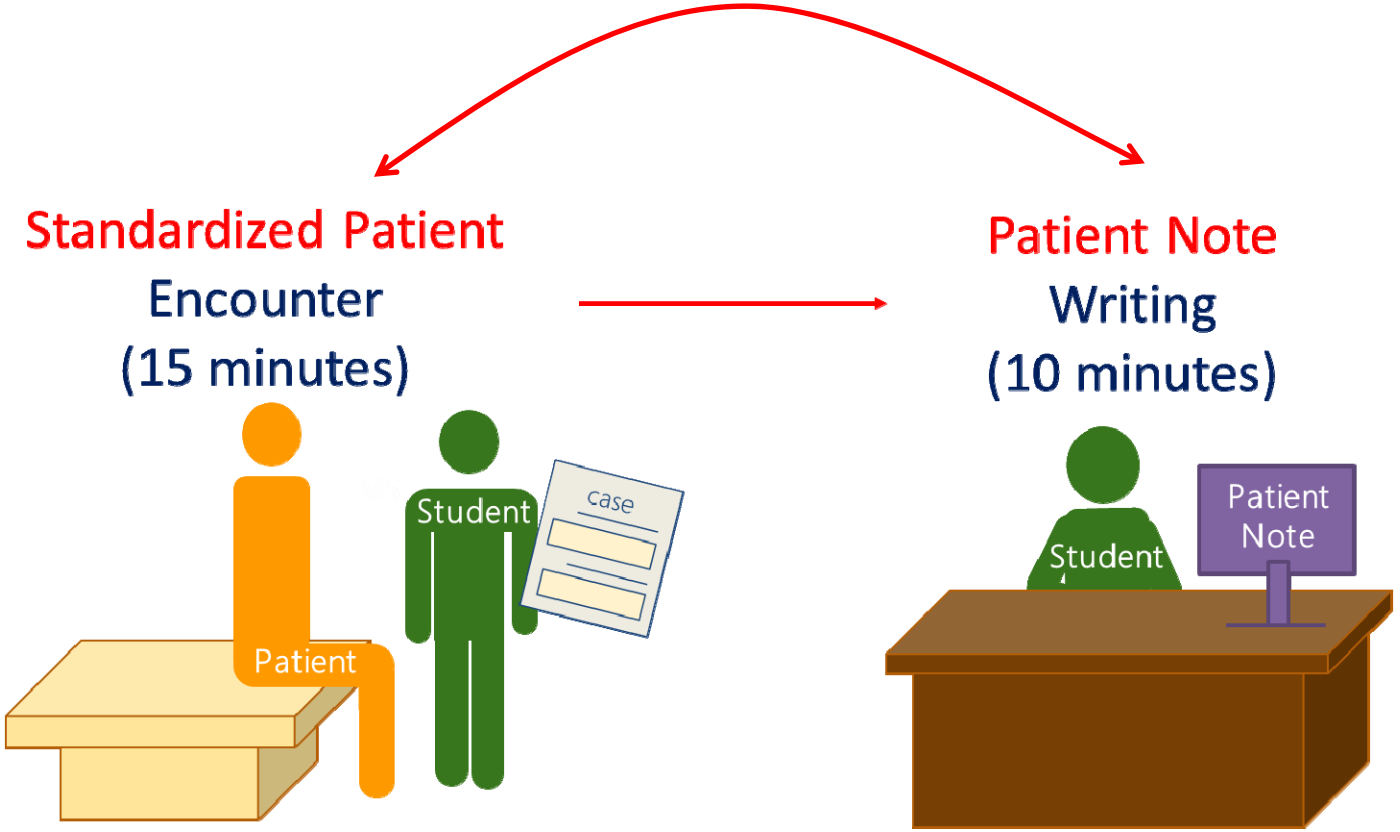
- Variance components
  - Students: 6%
  - Student-case interaction (case specificity): 20%

- Need 15 cases to reach Φ-coefficient of .70

Relations to other variables

Data Gathering: r = .47

Total Score: r = .38



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Paper 3: Rater Issues
(Project Years: 2014-2016)

Inter-rater reliability and generalizability of patient note scores using a scoring rubric based on the USMLE Step-2 CS format

Yoon Soo Park[1] · Abbas Hyderi[2] · Georges Bordage[1] · Kuan Xing[1] · Rachel Yudkowsky[1]

Abstract Recent changes to the patient note (PN) format of the United States Medical Licensing Examination have challenged medical schools to improve the instruction and assessment of students taking the Step-2 clinical skills examination. The purpose of this study was to gather validity evidence regarding response process and internal structure, focusing on inter-rater reliability and generalizability, to determine whether a locally-developed PN scoring rubric and scoring guidelines could yield reproducible PN scores. A randomly selected subsample of historical data (post-encounter PN from 55 of 177 medical

Park et al, *Advances in Health Sciences Education*, 2016

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Rater Agreement

- How well do raters agree?
  - Exact: % exact match
  - Adjacent: % match within ±1 adjacent point
  - Discrepant : % discrepant by ±2 or more points

Rater 1	Rater 2	
3	2	"Adjacent"
4	4	"EXACT"
1	3	"Discrepant"
2	4	"Discrepant"
1	4	"Discrepant"
3	4	"Adjacent"
1	1	"EXACT"
2	2	"EXACT"
4	2	"Discrepant"
1	1	"EXACT"

Agreement

% Exact = 40%

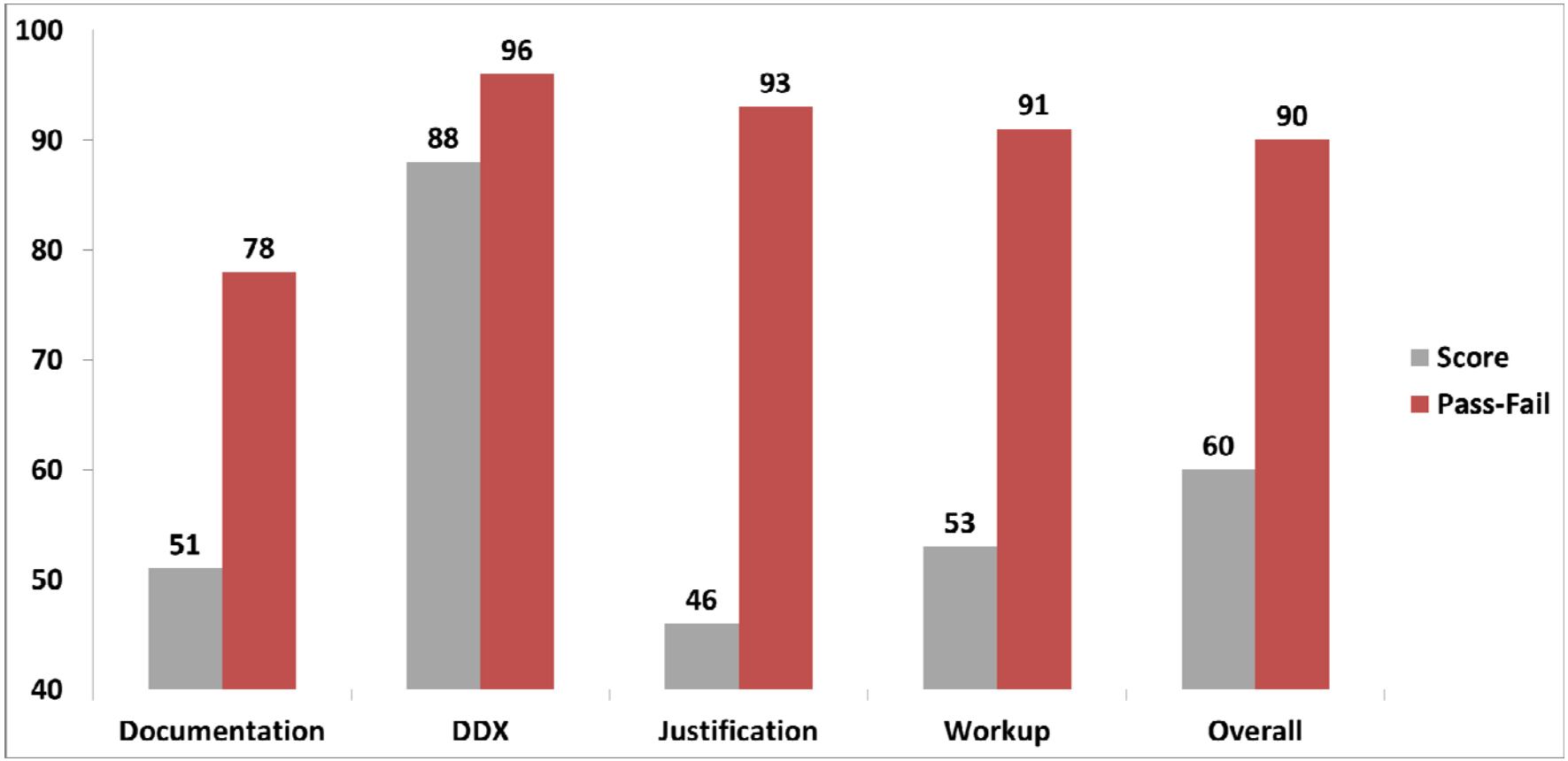
% Adjacent = 20%

% Discrepant = 40%

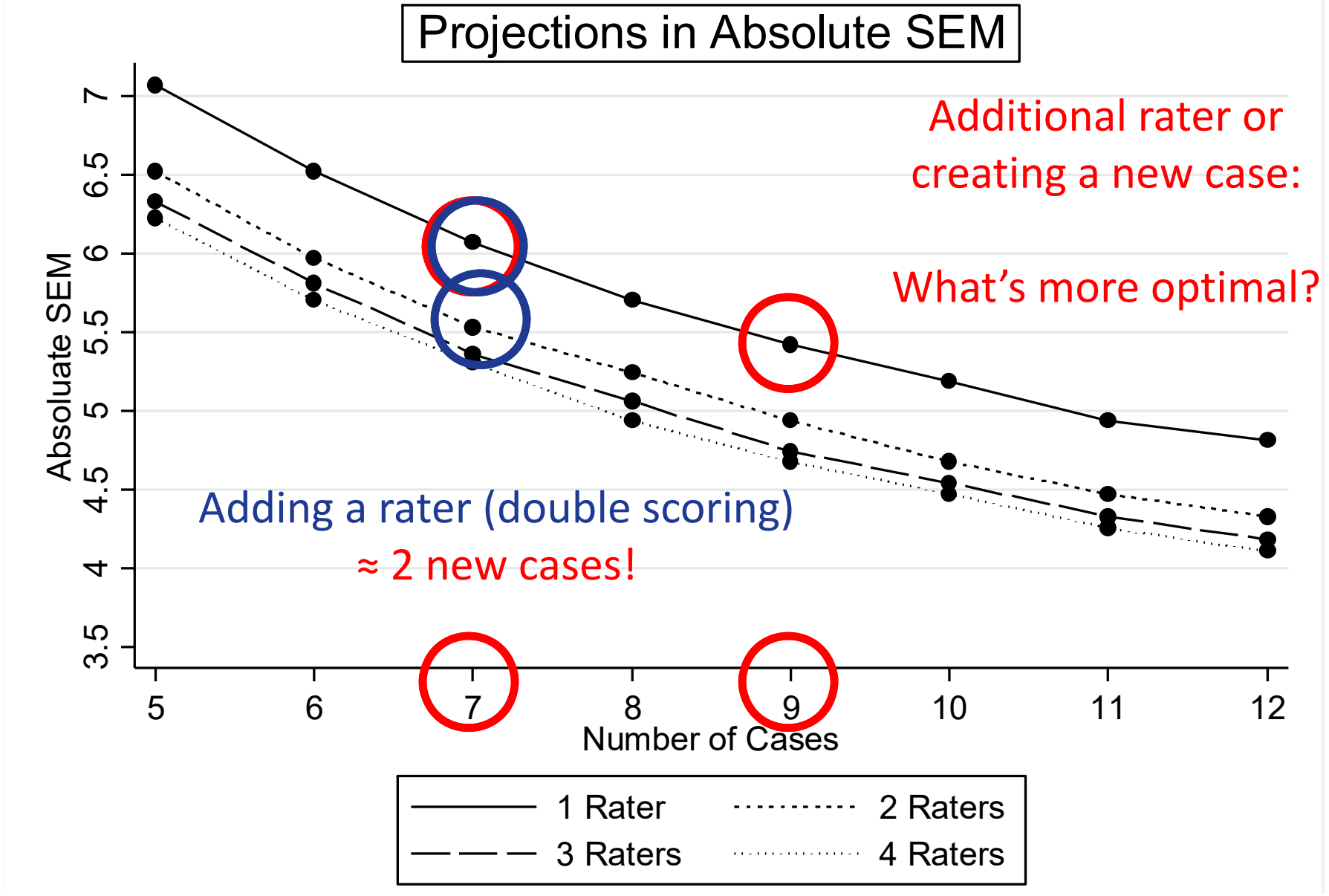
- Kappa
- Intraclass Correlations

% Agreement and ICC

- Intraclass Correlation (ICC) = .79



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Rater Training

Training Materials

1. Detailed case information
2. "Gold Standard" exemplar note
3. Case-specific scoring guidelines
4. 3 sample notes (actual student notes)
5. 10 extra notes (optional cases to review)

Training (2 sessions: 3 hours total)

1. Moderator presented summary of case
2. Discussion of 3 sample notes
3. Discrepancies in scoring discussed
4. Update scoring guidelines and "Gold Standard" exemplar note

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Paper 4: Composite Measures

(Project Years: 2015-2016)

Differential Weighting for Subcomponent Measures of Integrated Clinical Encounter Scores Based on the USMLE Step 2 CS Examination: Effects on Composite Score Reliability and Pass–Fail Decisions

Yoon Soo Park, PhD, Matthew Lineberry, PhD, Abbas Hyderi, MD, MPH, Georges Bordage, MD, PhD, Kuan Xing, MEd, and Rachel Yudkowsky, MD, MHPE

Abstract

Purpose
Medical schools administer locally developed graduation competency examinations (GCEs) following the structure of the United States Medical Licensing Examination Step 2 Clinical Skills that combine standardized patient (SP)-based physical examination and the patient note (PN) to create integrated clinical encounter (ICE) scores. This study examines how different subcomponent scoring weights in a locally developed GCE affect composite score reliability and pass–fail decisions for ICE scores, contributing to internal structure and consequential validity evidence.

Method
Data from two M4 cohorts (2014: n = 177; 2015: n = 182) were used. The reliability of SP encounter (history taking and physical examination), PN, and communication and interpersonal skills scores were estimated with generalizability studies. Composite score reliability was estimated for varying weight combinations. Faculty were surveyed for preferred weights on the SP encounter and PN scores. Composite scores based on Kane's method were compared with weighted mean scores.

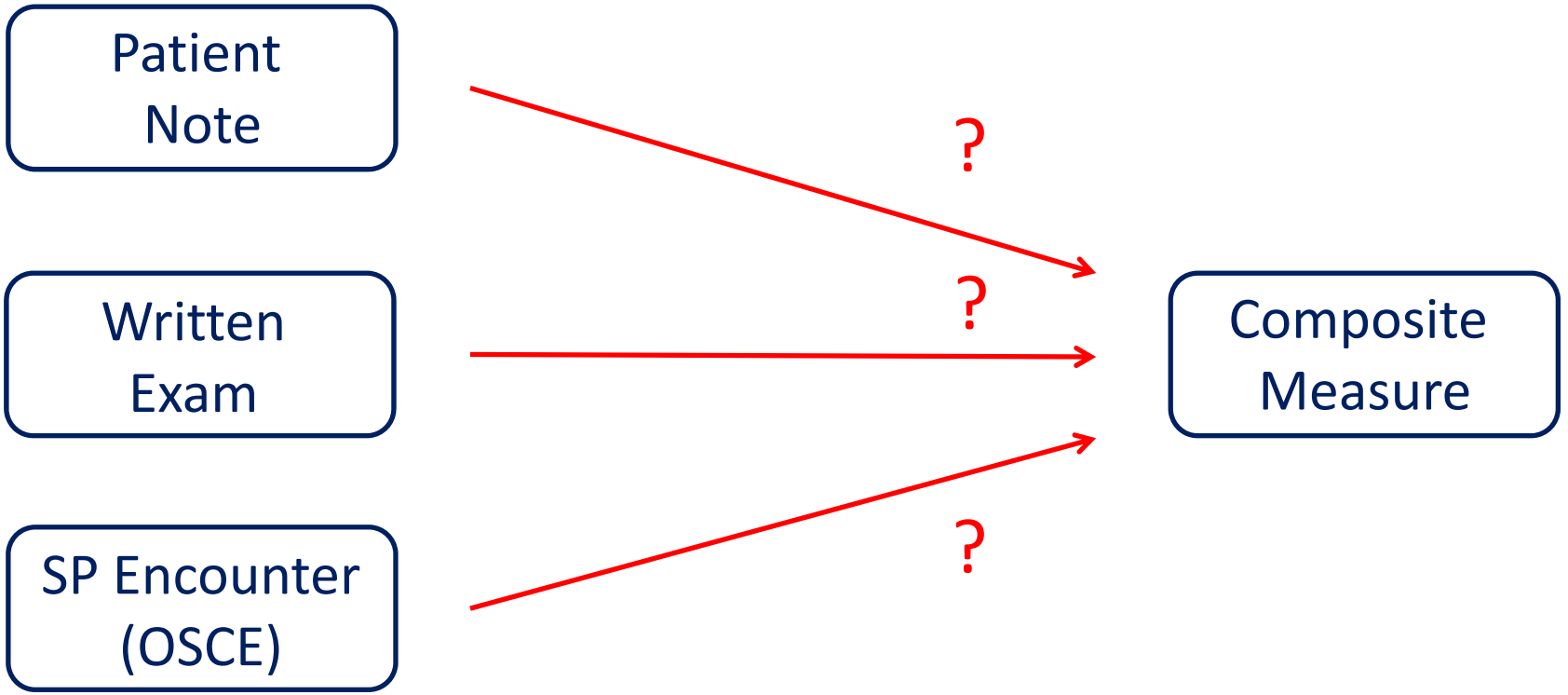
Results
Faculty suggested weighting PNs higher (60%–70%) than the SP encounter scores (30%–40%). Statistically, composite score reliability was maximized when PN scores were weighted at 40% to 50%. Composite score reliability of ICE scores increased by up to 0.20 points when SP–history taking (SP-Hx) scores were included; excluding SP-Hx only increased composite score reliability by 0.09 points. Classification accuracy for pass–fail decisions between composite and weighted mean scores was 0.77; misclassification was < 5%.

Conclusions
Medical schools and certification agencies should consider implications of assigning weights with respect to composite score reliability and consequences on pass–fail decisions.

Park et al, *Academic Medicine*, 2016

How to combine different educational experiences?

- Example: How can weights be specified to optimize signal?



THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Composite Scores:
Weighted Mean vs Kane Method

Weighted Mean Score

Weight = .30
SP Physical Exam
(SP-PE)
Score = 80%

+

Weight = .70
Patient Note
(PN)
Score = 65%

→ Weighted Mean Score
69.5%
(= 80% x .30 + 65% x .70)

Composite Score: Kane Method

SP Physical Exam
(SP-PE)
↓
Standardize

+

Patient Note
(PN)
↓
Standardize

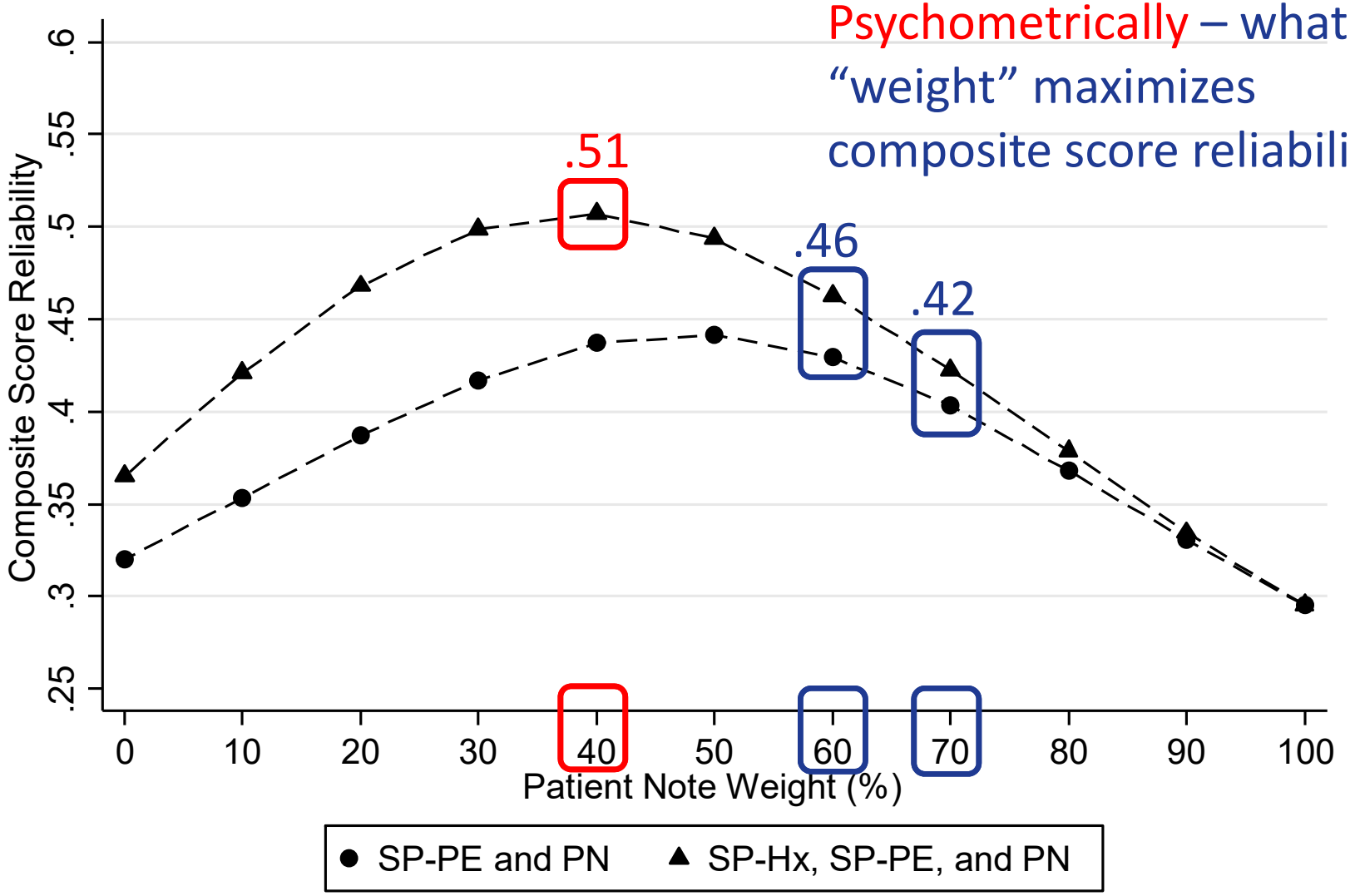
Weighted Mean → Adjust Composite Score Variance

→ Kane Composite Score

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Weights and Composite Score Reliability

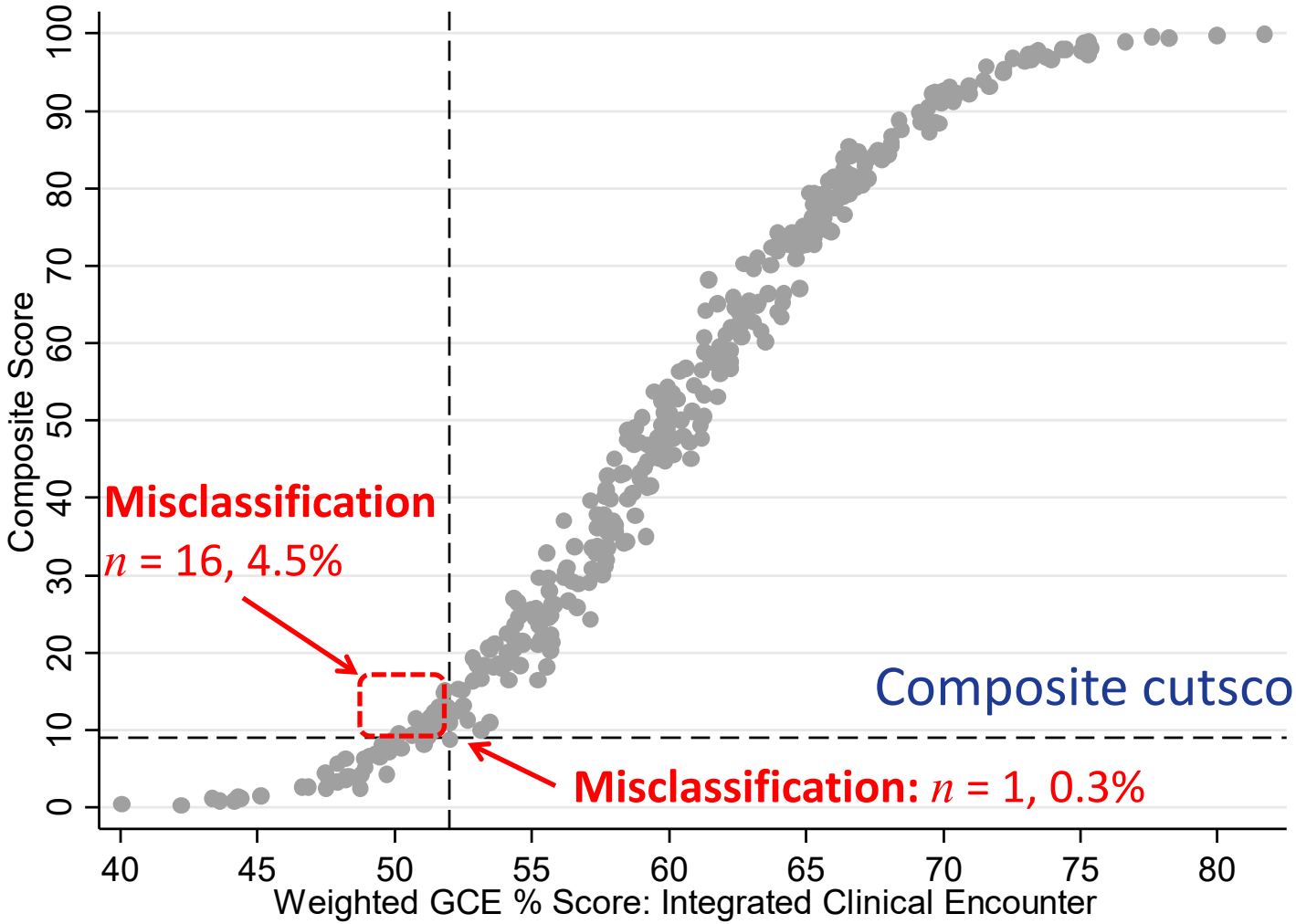
Psychometrically – what "weight" maximizes composite score reliability?



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Pass-Fail Classification

Weighted GCE % cutscore = 52%



Classification accuracy (κ) = .77

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Paper 5: Multisite Study

(Project Years: 2015-2016)

Validity Evidence and Scoring Guidelines for Standardized Patient Encounters and Patient Notes from a Multisite Study of Clinical Performance Examinations in Seven Medical Schools

Yoon Soo Park, PhD, Abbas Hyderi, MD, MPH, Nancy Heine MEd, RN, ANP, Win May MD, PhD, Andrew Nevins, MD, Ming Lee, PhD, Georges Bordage, MD, PhD, and Rachel Yudkowsky, MD, MHPE

Park et al, *Academic Medicine*, in press

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Why multisite validity study?

- *Enhance* psychometric rigor → Validity

Psychometric

- Greater heterogeneity of learner samples
- Reliability and validity inferences

Educational

- Broaden mix of clinical cases
- Shared case development
- Develop common rater training protocols
- Evaluate program
- Many more!

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Multisite Participants

Validity Evidence:

UI-College of Medicine
(n = 160)

Generalize Inferences?

California Consortium for the Assessment of Clinical Competence (CCACC)

(n = 830)

1. Loma Linda University
2. UC Davis
3. UCLA
4. UC San Diego
5. University of Southern California
6. Stanford University

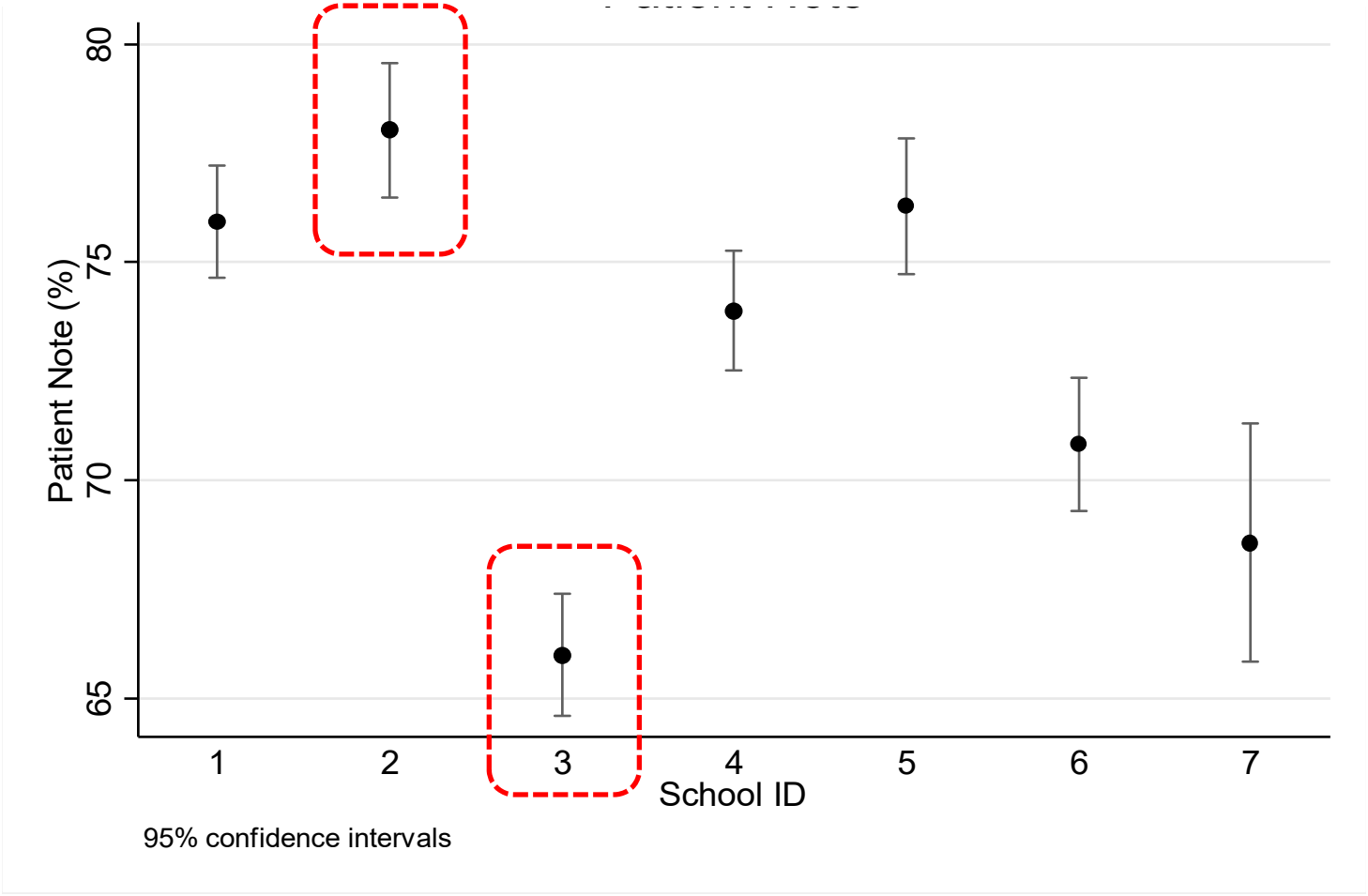
Cross-Institutional Consequential Evidence

Multisite Validity Evidence

- Consistent with previous findings
- Notable school-effect differences on patient note scores
  - Evidence from variance components
  - Differences in effective teaching of clinical reasoning
- Ability to "justify" diagnoses → most discriminating
- Develop case-specific scoring guidelines
  - More accessible

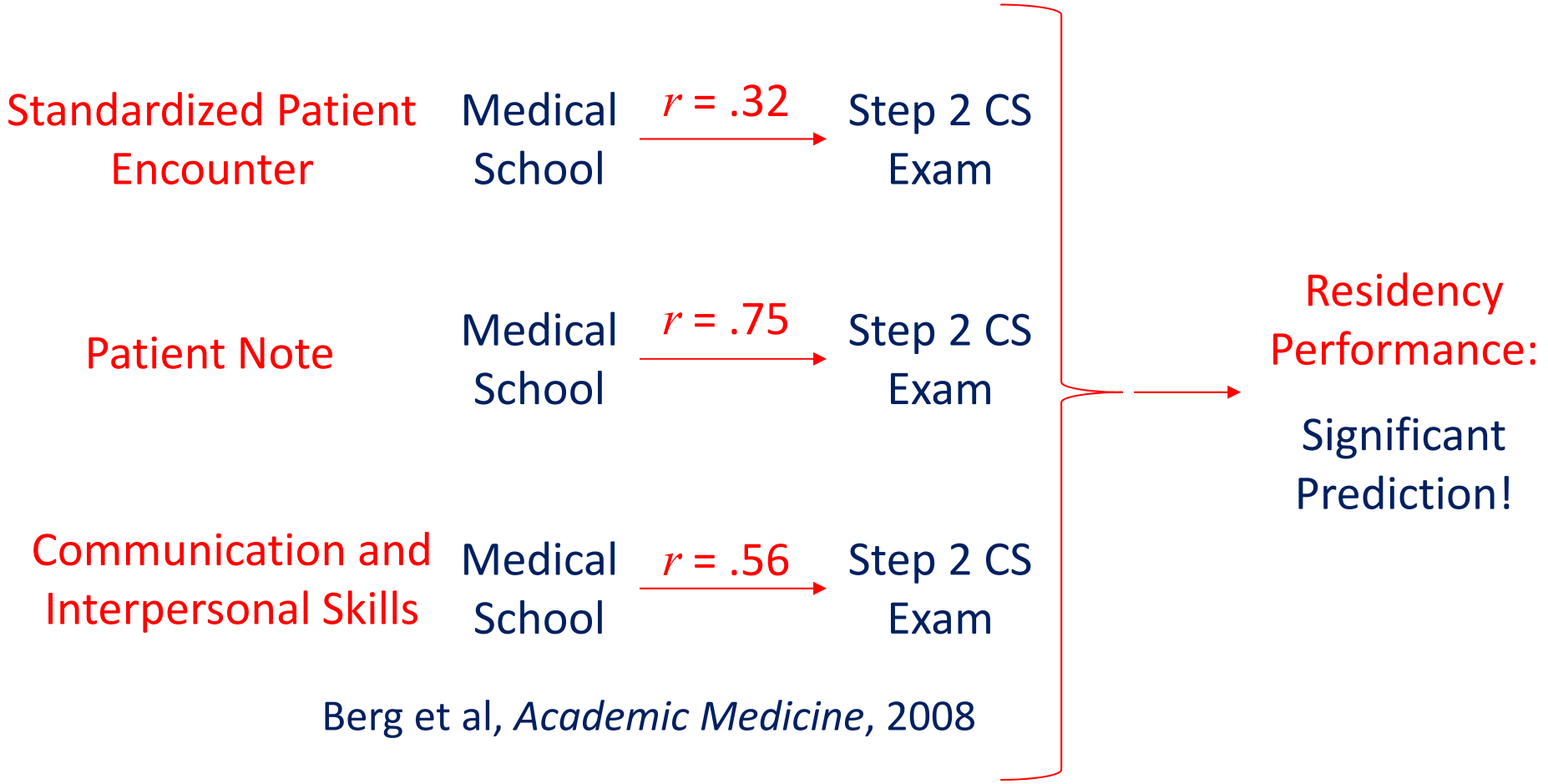
THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Program Evaluation



THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Why is performance on Medical School Clinical Skills Exam important?

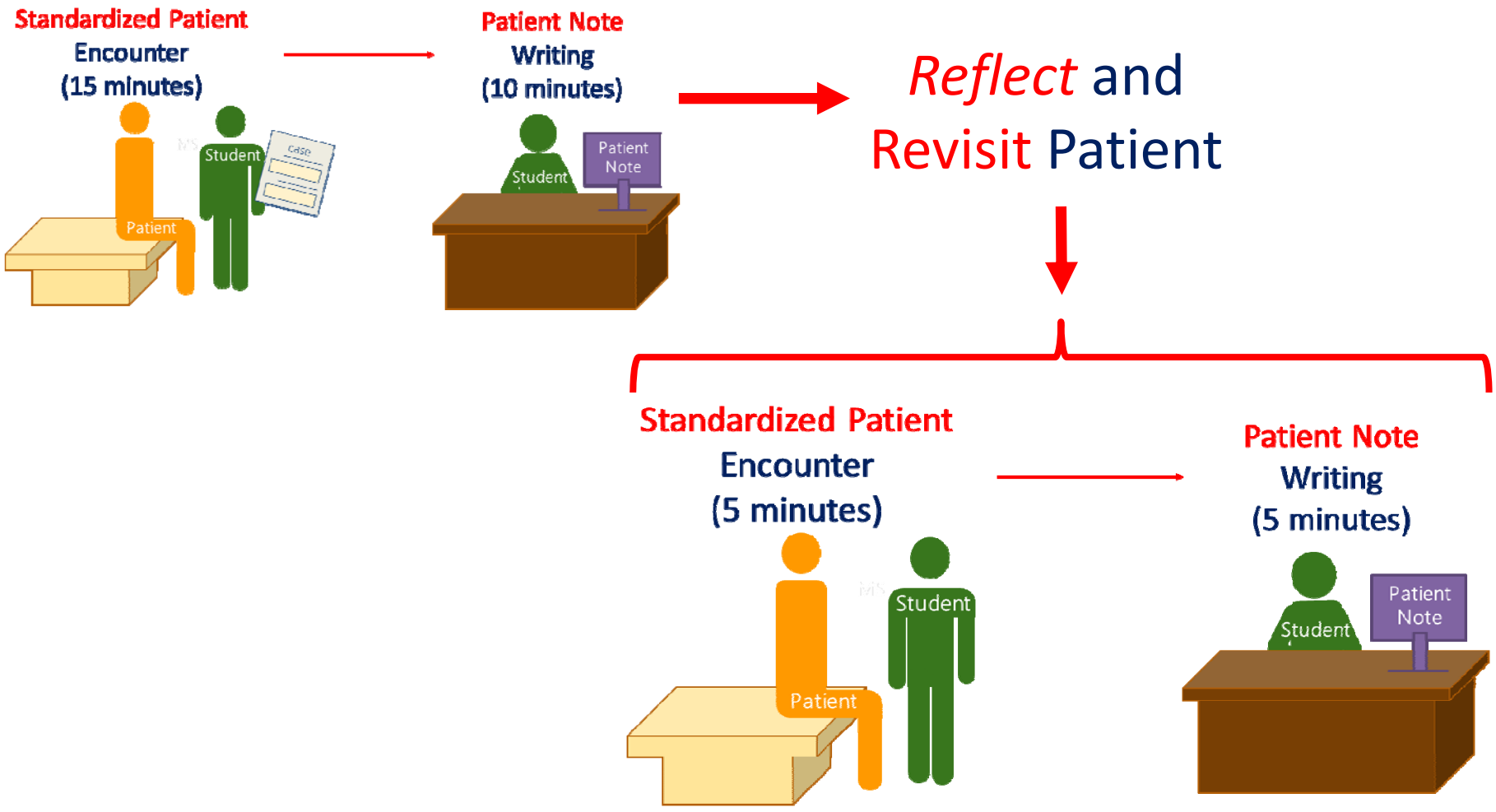


Berg et al, *Academic Medicine*, 2008

Cuddy et al, *Academic Medicine*, 2016

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Current Studies on the Patient Note (1) – Revisit Project (8 medical schools)



THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Reflection Study –
Standardized Patient Encounter

Acting on Reflection: the Effect of Reflection on Students' Clinical Performance on a Standardized Patient Examination

Benjamin Blatt, MD[1], Margaret Plack, PT, EdD[2], Joyce Maring, PT, EdD[3], Matthew Mintz, MD[4], and Samuel J. Simmens, PhD[5]

[1]Clinical Learning and Simulation Skills (CLASS) Center, Department of Medicine, The George Washington University School of Medicine and Health Sciences, George Washington University Hospital, 6th floor, 900 23rd Street NW, Washington, DC 20037, USA; [2]Department of Health Care Sciences, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA; [3]Program in Physical Therapy, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA; [4]Department of Medicine, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA; [5]Department of Epidemiology and Biostatistics, The George Washington University School of Public Health and Health Services, Washington, DC, USA.

BACKGROUND: Little evidence exists to support the value of reflection in the clinical setting.

OBJECTIVE: To determine whether reflecting and revisiting the "patient" during a standardized patient (SP) examination improves junior medical students' performance and to analyze students' perceptions of its value.

KEY WORDS: reflection; standardized patients; medical student.
DOI: 10.1007/s11606-007-0110-y
© 2007 Society of General Internal Medicine

Blatt et al, *Journal of General Internal Medicine*, 2007

Current Studies on the Patient Note (2) – Non-Physician Rater (15 medical schools)

- Scoring Patient Note – Expensive
- Cost of training and scoring notes – UIC estimates:
  - Physician: $40,869 (excluding case development costs)
  - Non-Physicians: $9,990
- California pilot study (4 medical schools)
  - Physician and Non-Physician correlation: r = .64

THE UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Translational Science in Education

K-12 Education (example)

Teacher training, development → Teacher knowledge and skills → Classroom teaching → Student achievement

Allen et al, *Science*, 2011

Medical Education

Medical Education → Trainee Clinical Skills and knowledge → Clinics and offices for better health care → Patient or public health outcome

T1 (education setting)

T2 (clinic)

T3 (clinic and community)

McGaghie WC, *Science Translational Medicine*, 2010

THE UNIVERSITY OF ILLINOIS
COLLEGE OF MEDICINE
CHICAGO PEORIA ROCKFORD URBANA

Yoon Soo Park, PhD
email: yspark2@uic.edu