Clinical Skills Assessment
& USMLE Step2 cs experience

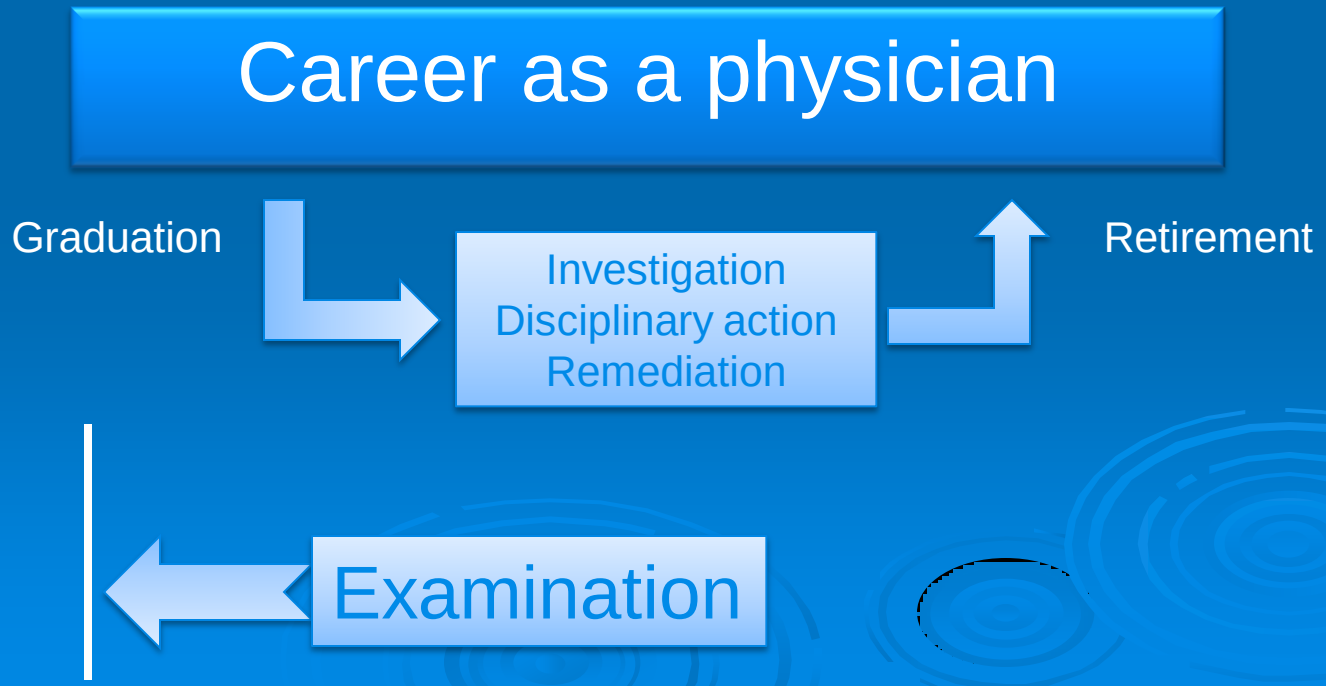
Graham McMahon MD MMSc
Assistant Professor of Medicine,
Diabetes, Endocrinology, and Hypertension
Brigham & Women's Hospital
Harvard Medical School

Overview

- Overview of USMLE Step 2 CS
  - History and Strife
  - Case Development
  - Scoring and Reporting
  - Results
  - Quality Assurance
- Implications for Japan

Professional Self-Regulation

- Physicians are trusted by patients and government to self-regulate



U.S. support for an independent assessment program

- Woven into culture for credentialing professionals
  - Education in an accredited program is necessary but not sufficient
  - Licensing, like many processes, best served by many independent checks
- Integrates well with the importance given to the rights and responsibilities of the individual state
  - State will decide on: method of assessment and standards to be applied
  - Public expectation that the state is directly involved in the demonstration and maintenance of knowledge and skills of physicians practicing within its borders

The National Board of Medical Examiners® (NBME®)

- Founded 1915 - 95 years old
- Non-profit - 501(c)(3) - organization
- Based in Philadelphia
- Staff of ~400
- Supported by examination fees
- Approximately $95 million annual revenues
- ~300,000 tests annually

NBME Mission –

To protect the health of the public through state of the art assessment of health professionals.

While centered on the assessment of physicians, this mission encompasses the spectrum of health professionals along the continuum of education, training and practice and includes research in evaluation in well as development of assessment instruments.

Examination performance predicts competence

- Physicians who have more training in a discipline are more knowledgeable and achieve higher scores in their respective discipline on recertification examinations.
- More knowledgeable physicians are more likely to adhere to evidence-based guidelines in the delivery of care and achieve better patient outcomes.
- Certification status, which represents pass/fail status on certification examinations, is an important predictor of quality of care

USMLE: Purpose

- Three step examination system primarily designed to support the medical licensing process
- **Overall purpose**
  - **Assess physician's ability to apply knowledge, concepts, and principles, and to demonstrate fundamental patient-centered skills that constitute the basis of safe and effective patient care**

WHY CLINICAL ASSESSMENT?

Rationale

- Clinical skills are
  - Essential to the practice of medicine
  - In decline
  - Erratically assessed
  - Assessed using unreliable approaches
  - A leading source of complaints and litigation
- Those who license physicians have a duty to protect the public and to demonstrate the requisite quality of medicine.

Arguments against…

- Inconveniences many to find few
- High cost (test + travel)
  - many students already have large debt
- Medical schools are already doing this
- Medical schools are better suited to this responsibility
- Interrupt key clinical rotations
  - Timing problematic

Controversy

- Support
  - Federation of state medical boards
  - American Association of Medical Colleges
  - Patient advocacy organizations
- Oppose
  - American medical students association
  - American Medical Association
  - Council of Medical School Deans

Graham McMahon appointed as AMA's representative on the NBME board to oppose the CS examination


The NEW ENGLAND JOURNAL of MEDICINE
amednews.com
American Medical News
Published by the American Medical Association
MOBILE
MULTIMEDIA
TOPICS
PAST ISSUES
PROFESSION
Delegates oppose testing of clinical skills for licensure
Supporters of the NBME exam say a national standard is needed, but opponents say medical schools are better for skill assessment.
By BONNIE BOOTH, amednews staff. Jan. 6, 2003.
Follow
PRINT | E-MAIL | RESPOND | REPRINTS | SHARE
New Orleans -- Medical students who have been opposed to the National Board of Medical Examiners' plan to test their clinical skills as a condition of licensure won long-sought backing from the American Medical Association House of Delegates during their meeting here in December 2002.
National Board of Medical Examiners (NBME) heard candidates as they took a history and watched them perform a physical examination. Three years later, however, the board analyzed the results of that examination and determined that it failed to produce the degree of statistical reliability that had been anticipated; thus, the examination was discontinued.1 The hope was to replace the unreliable
D-295.968 Proposed Implementation of Clinical Skills Assessment Exam
(1) Our AMA shall urgently contact the National Board of Medical E
(NBME), all organizations represented on the
ard, and the Federation of State Medical
spension of the implementation of the
Assessment Examination (CSAE) until such
ation has been demonstrated to be
practical, and evidence-based;(b)
en published in peer review journals
or US medical students and
that the fiscal and societal
stify the costs; and (c) Testing
onable geographic locations than
E. (2) Our AMA and state
state medical licensing boards
dical licensure requirements
(3) Our AMA shall continue
Federation of State Medical
of the CSAE. (4) Our
Liaison Committee on
l students' clinical skills
training. (Sub. Res.

USMLE CLINICAL SKILLS EXAMINATION

USMLE Step 2 CS

- Enhancement to USMLE Step 2
- Testing using standardized patients has been shown to be a reliable, valid, and feasible means of assessing clinical skills
- The implementation of the Step 2 clinical-skills examination emphasizes the importance of these skills and drives curricular reform

USMLE: Purpose

- Patient-centered Skills
  - History taking / Physical examination
  - Communication and Interpersonal Skills
  - Medical Record Documentation

- Important for safe and effective patient care
  - History/PE → Diagnosis and Management
  - Communication → Patient Health Outcomes
  - Medical Record → Errors and Patient Safety

CASE DEVELOPMENT

USMLE Step 2 CS: Content

- **Step 2 CS Blueprint**
  - **Defines content categories**
  - **In meeting blueprint specifications, each test form provides:**
    - **Adequate sampling of content domain**
    - **Comparable content between test forms**

Blueprint for each case

Case Acuity

Acute
Subacute / Chronic

Case Content

Cardiovascular
Respiratory
Gastrointestinal
Musculoskeletal
Constitutional
Neurological
Psychiatric
Genitourinary
Women's health
Unclassified / multi-system

Test Form

Patient age

Age less than 18
Age 18 – 44
Age 45 – 64
Age 65 +

Patient Gender

Male
Female

"common and important cases"

Case Creation

- Prompted by blueprint gaps
- Two panels of 5 physicians
- Meet over 3 days in Philadelphia
- With trainer and case-writer
  - Develop the testing objectives
  - Write the scenario, HPI, etc.
  - Distill the SP checklist
  - Create the scoring key for the patient note
  - Test the case

Case Refinement and Training

- Cases are refined by the case-writer
  - May consult with committee members
- Single SP is trained
- SP encounter with test students is recorded for training with other SPs
  - SPs receive extensive training in the proper portrayal of appropriate signs and symptoms and in the rating of examinee performance

SP and SP-Trainer Training

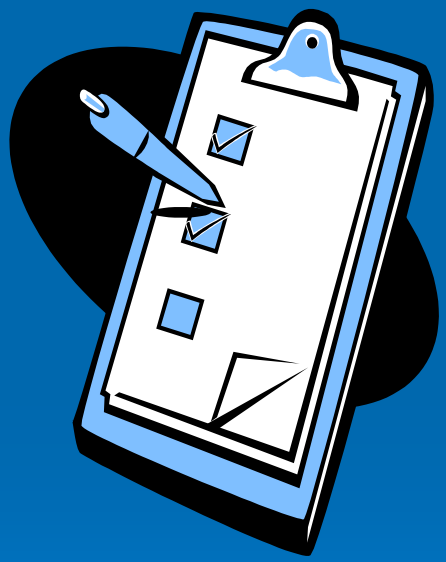
Standardized Patients

- Rigorous training – generic and case specific
- Electronic delivery ("Ecase") of case materials
- Sign off process for SPs – Criteria : # portrayals, tests

SP Trainers

- Adherence to training protocols
- "Training academy" for SP trainers

Case Development



- Iterative Process
  - Involves test committees
  - Focus on clinical presentation
  - Checklists “limited”
  - Encounters with SPs
  - Appropriateness of content and difficulty
- Stepwise progression through pilot and calibration stages
  - Validation / refinement via review of examinee performance data

ADMINISTRATION

Step 2 CS: Structure & Tasks

- **12 patient encounters**
- **15 min. for encounter / 10 min. for patient note**
- **Each encounter:**
  - **Elicit pertinent history,**
    Perform appropriate physical examination,
    Communicate effectively
  - **Document:**
    - **Findings from the history and physical**
    - **Diagnostic impression / Further work-up**

USMLE Step 2 CS: Logistics

- Regional delivery model
  -
- Five regional test centers across US
- Projected examinee volume – 30,000+ / year
- Individual center capacity
  - 3 examinations / day (33 examinees); up to 7 days/week

Structure 1

- Exam is delivered at five regional centers
  - Centers are a joint venture of NBME and the ECFMG
  - Locations: Atlanta, Chicago, Houston, Los Angeles, Philadelphia
  - Optimal combination of convenience, cost-efficiency and standardization
- Delivery nearly every day, year round

Exam structure 2

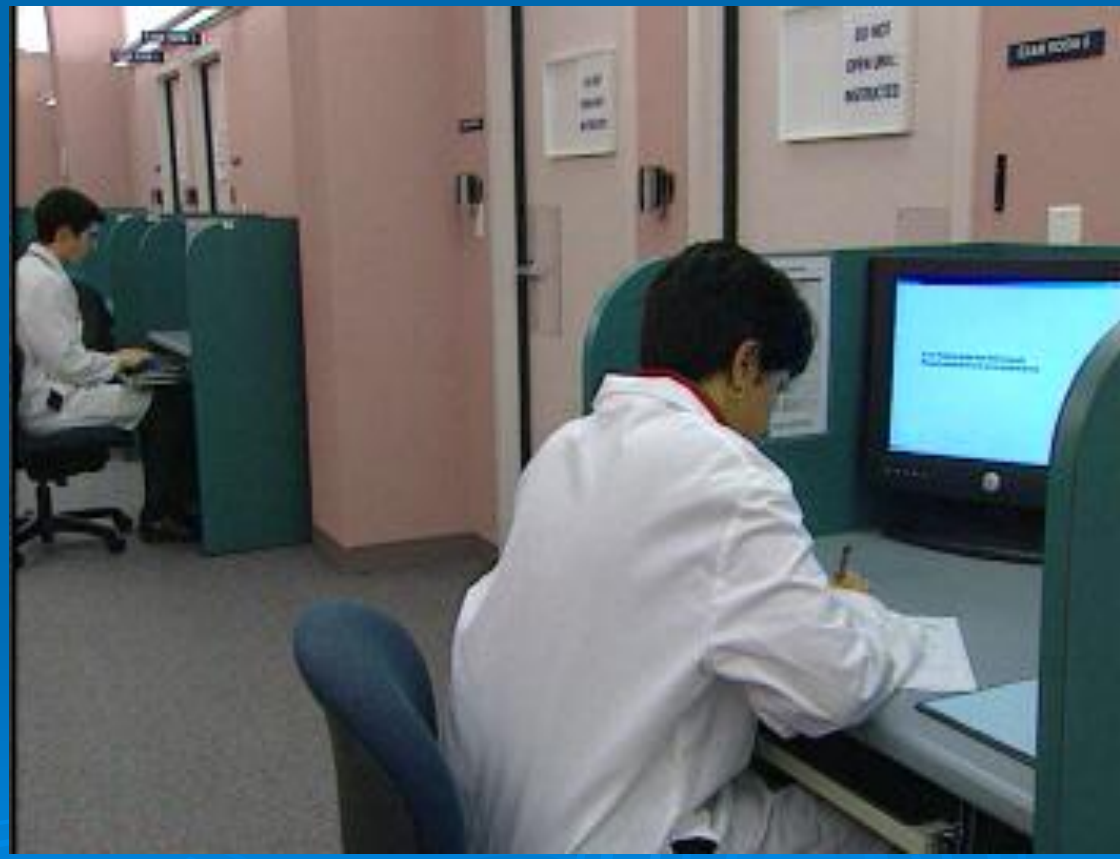
- Examinees go through 12 encounters , interaction with SPs (10 are scored)
- They are given basic vitals and chief complaint on the door
- They interview the SP, perform relevant examination, and discuss their findings with the SP

Structure 3

- Examinees are assessed by the SP on:
  - ability to take a history and do a focused physical exam
  - communicate effectively
  - proficiency in spoken English
- SPs use checklists and rating scales

Structure 4

- Examinees then type a note
- Physician raters score each note for
  - The accuracy of the recorded history
  - The differential diagnosis accuracy and importance
  - The appropriateness of work-up plan



SCORING

Components of USMLE 2 cs

Scored	Components	Evaluated by	Reported as
Data gathering	History elements Examination elements	SP	Clinical Encounter score (ICE)
Documentation	History correct Examination correct Differential diagnosis Investigation	Physician Rater	Clinical Encounter score (ICE)
Communication	Communication skills Spoken English	SP	Communication English Proficiency

ICE Scoring 1: Data Gathering

- Physician observers not feasible
- SP checklist completion reliable and valid
- Checklists are developed by clinical faculty
  - 14-20 checklist items per SP
  - Include essential history and physical examination elements for specific clinical encounters.
  - Very Specific:
    - Not ‘asked about my headache’,
    - Rather ‘asked if my headache was worse in the morning’
  - Some uninformative checklists are removed after initial testing

ICE Scoring 2: Documentation

- The ability to document in the patient note the findings from the patient encounter, diagnostic impression, and initial patient work-up is rated by physician raters.
- Same rater rates single note
- Weights and suggested content determined by committee
- The raters evaluate
  - The inclusion of important positive and negative findings from the history and physical examination,
  - The listed differential diagnoses and
  - The diagnostic assessment plans.
  - A holistic rating of overall quality

HISTORY: Include significant positives and negatives from history of present illness, past medical history, review of system(s), social history and family history.

48 YO ♀ - CHEST PAIN X 90 MINS

HPI
- BURNING
- NO RADIATION
- sl SOB
- sl NAUSEA & DIAPHORESIS
- RESOLVED SPONTANEOUSLY
- SIMILAR EPISODES 2-3 MOS, AFTER HEAVY MEAL OR EXERTION
- SOME RELIEF ☡ ANTACIDS

PMH
- ↑ CHOLESTEROL, NO FOLLOW-UP OR TREATMENT
- TENNIS WKLY
- SMOKED 30 PK YRS, STOPPED 3 YRS AGO
- NO UNUSUAL STRESS
- MOTHER ☡ NIDDM, BROTHER ☡ UNKNOWN HEART
- NO hx HTN, HAS NOT SEEN MD X 2 YRS

PHYSICAL EXAMINATION: Indicate only pertinent positive and negative findings related to patient's chief complaint.

BP 160/80 NO OBVIOUS DISTRESS, ANXIOUS TO LEAVE

CHEST NON-TENDER, CLEAR BS BILAT, NO WHEEZES CRACKLES OR RALES

HEART PMI NON-DISPLACED, REG RHYTHM, NO (M) OR RUBS

ABD BS(+) NONDISTENDED, NO MASSES OR ORGANOMEGALY TENDERNESS IN EPIGASTRUM W/O REBOUND

DIFFERENTIAL DIAGNOSES: In order of likelihood (with 1 being the most likely), list up to 5 potential or possible diagnoses for this patient's presentation (in many cases, fewer than 5 diagnoses are likely).

1. ESOPHAGEAL REFLUX DISEASE
2. PEPTIC ~~ULCER~~ ULCER
3. CORONARY ARTERY DISEASE
4. CHOLECYSTITIS
5. MUSCULOSKELETAL CHEST PAIN

DIAGNOSTIC WORK UP: List immediate plans (up to 5) for further diagnostic workup.

1. STOOL FOR OB
2. EKG
3. CXR
4. UPPER GI ENDOSCOPY
5.

Patient Notes

Room Number: 15

Candidate Name: Doe, John USMLE ID: 1-234-567-8

Candidate Number: 20

HISTORY – Include significant positives and negatives from history of present illness, past medical history, review of system(s), social history and family history.

PHYSICAL EXAMINATION - Indicate only pertinent positive and negative findings related to the patient's chief complaint.

DIFFERENTIAL DIAGNOSIS - **In order of likelihood** (with 1 being the most likely), list up to 5 potential or possible diagnoses for this patient's presentation (in many cases, fewer than 5 diagnoses are likely):

1
2
3
4
5

DIAGNOSTIC WORKUP - List immediate plans (up to 5) for further diagnostic workup:

1
2
3
4
5

Submit

Coming Soon

- Justification for each differential diagnosis and test

Communication Skills

- The ability to communicate effectively with patients, demonstrating appropriate interpersonal skills, is essential to safe and effective patient care.
- Step 2 CS is intended to determine whether physicians seeking an initial license to practice medicine in the United States, regardless of country of origin, can communicate effectively with patients.

Communication Subscales

- **Data gathering skills**
  - **Open-ended questions, transitional statements, not interrupting the patient**
- **Information sharing skills**
  - **Responsiveness to patient questions/concerns, provision of counseling when appropriate, avoidance of jargon**
- **Personal manner and rapport**
  - **Expression of interest in the impact of the illness, concern for patient comfort and modesty**

SP ratings of Communication and English Language

- Candidates are rated by the standardized patients through
  - Communication of initial diagnostic impression and work-up plan.
  - Answers to scripted questions

REPORTING

Step 2 CS Reporting

- Reporting is Pass or Fail only
- Examinees must pass all three subcomponents
  - Failure of one means failing the examination
- Feedback to examinees
  - Performance report – overall and subcomponent outcomes
  - Failing examinees only – graphical profiles
    - Intended to show relative strengths and weaknesses

INFORMATION PROVIDED FOR EXAMINEE USE ONLY

The Performance Profile below is provided solely for the benefit of the examinee.

These profiles are developed as self-assessment tools for examinees only and will not be reported or verified to any third party.

USMLE STEP 2 CS PERFORMANCE PROFILES

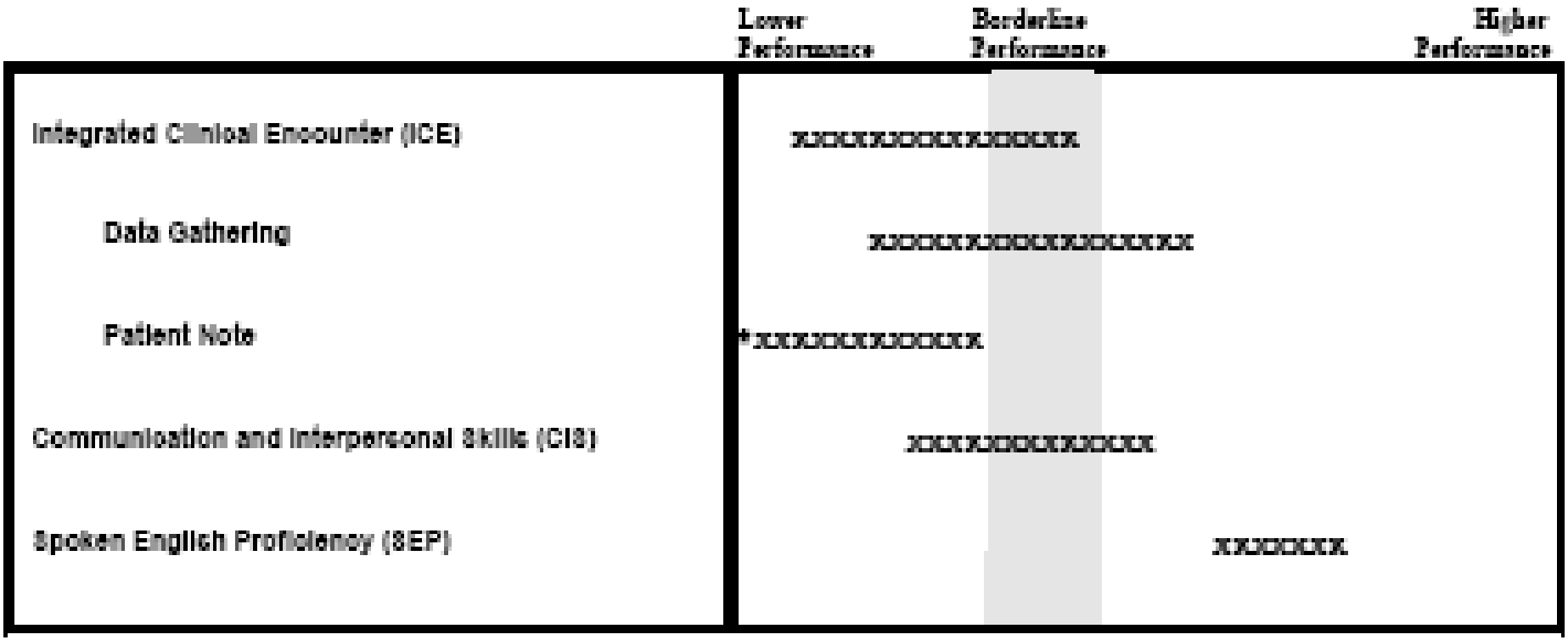
The above Performance Profile is provided to aid in self-assessment. The shaded area defines a borderline level of performance for each subcomponent (ICE, CIS, SEP); borderline performance is comparable to a HIGH FAIL/LOW PASS on the subcomponent.

For the ICE subcomponent, additional information is provided for Data Gathering and for the Patient Note. Data Gathering represents performance on history-taking and physical examination tasks. Patient Note represents performance on completion of the post-encounter summaries.

Performance bands indicate areas of relative strength and weakness. Some bands are wider than others. The width of a performance band reflects the precision of measurement: narrower bands indicate greater precision. An asterisk indicates that your performance band extends beyond the displayed portion of the scale. Small differences in the location of bands should not be over interpreted. If two bands overlap, performance in the associated areas should be interpreted as similar.

Additional information concerning the Step 2 CS subcomponents can be found in the *USMLE Step 2 CS Content Description and General Information*.

Sample Score Report



Scoring and reporting - 3

- US medical schools
  - Receive a yearly report on their group of students taking CS
  - Group performance on all of the subcomponents

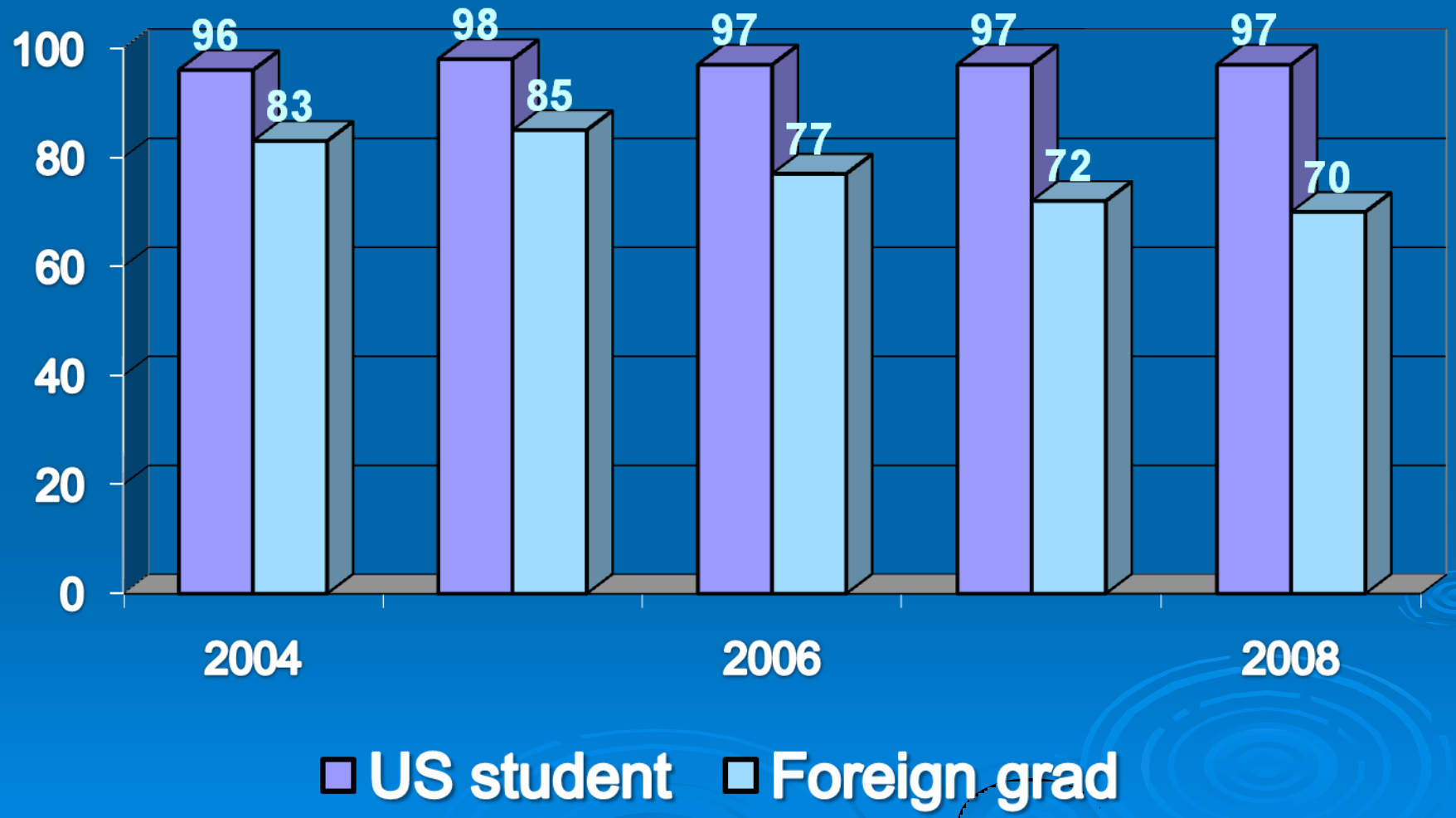
RESULTS

Results by Examinee Group

	2006-2007*		2007-2008*	
	#Tested	%Passing	#Tested	%Passing
Examinees from US/Canadian Schools that Grant				
MD Degree	17,256	97%	17,302	97%
1st Takers	16,769	97%	16,715	97%
Repeaters**	487	93%	587	92%
Examinees from Non-US/Canadian Schools				
1st Takers	14,439	77%	13,787	72%
Repeaters**	2,379	68%	3,436	64%
Total non-US/Canadian	16,818	76%	17,223	70%

Overall 84% passed of 34,525 examinations per year,

Performance over Time



Results by Subscale

	2006-2007		
	ICE	CIS	SEP
All US/Canadian Schools	98	99	>99
All Non-US/Canadian Schools	86	81	92

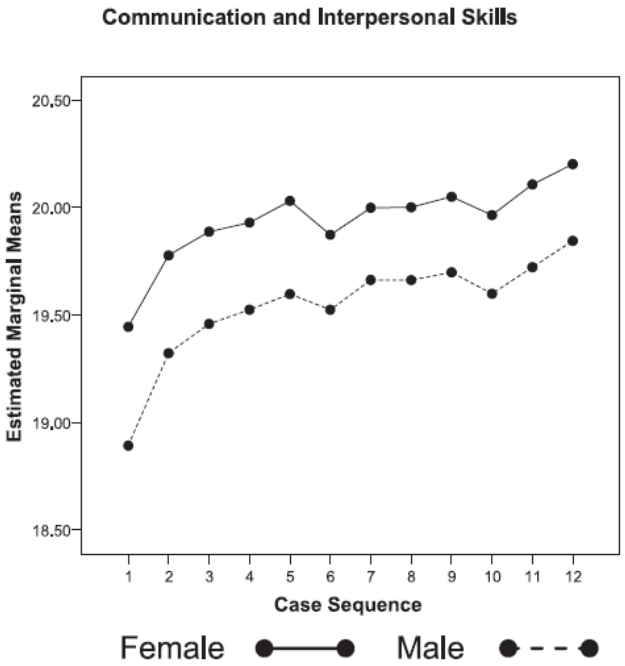
Data Comm English

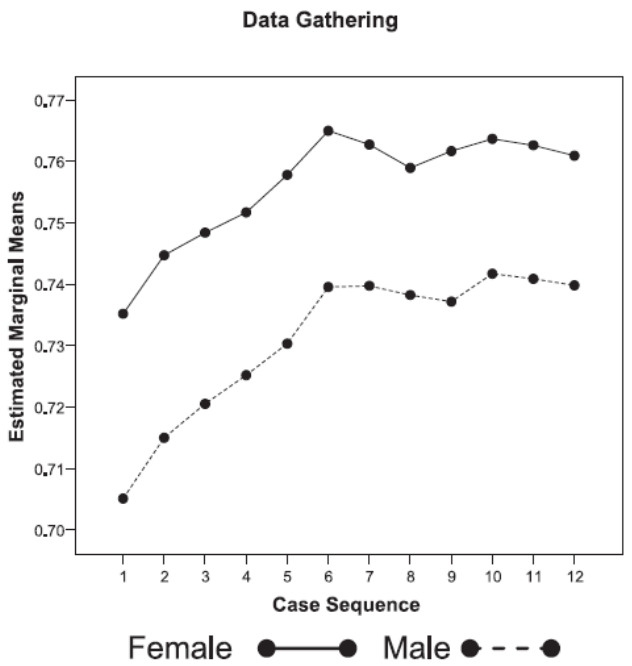
Any interesting patterns?

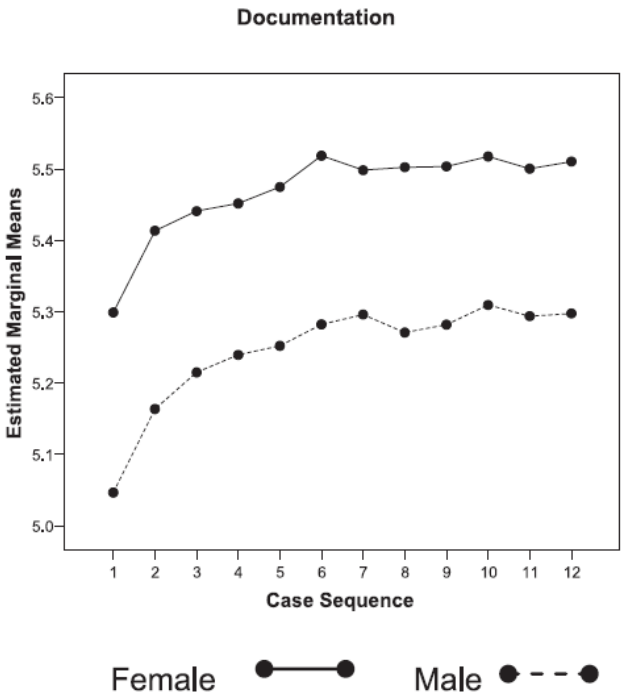
- US Students
  - Difficulty from most to practically none: data→communication→English
  - Rare to be unsuccessful on both data and communication
- International Students
  - Difficulty from most to least: communication→data→English
  - Rare to be unsuccessful on both communication and English
  - Extremely rare to fail all subcomponents

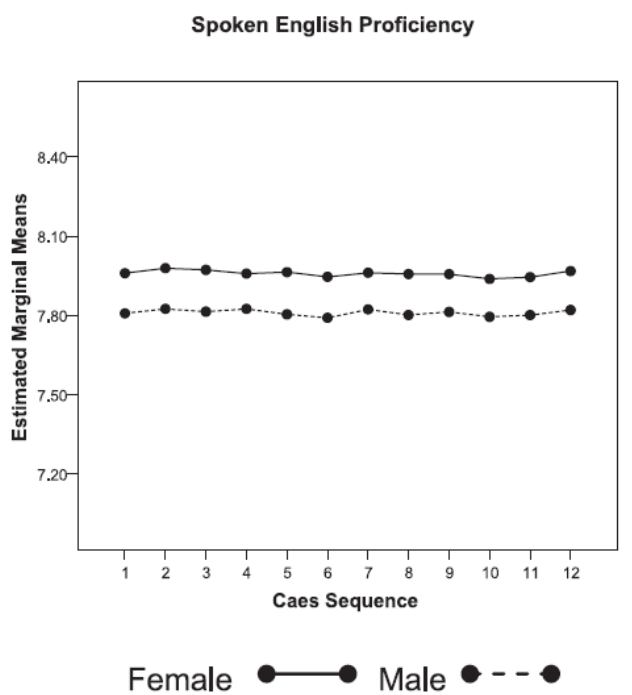
Sequence effects

Students do better as exam progresses









Acad Med. 2007;82(10 Suppl):S101–S104.

QUALITY ASSURANCE

Threats to Validity and Reliability

- Content and/or tasks not relevant or realistic
- Individual test forms vary in content coverage
- Scoring methods not appropriate for skills tested
- Inconsistency in SP portrayal and scoring
  - Between cases, across sites, over time
- Level of difficulty of cases / exams inconsistent
- Standard setting approach inappropriate

Quality Assurance

- Content quality
- Careful SP sign-off process
- Monitoring procedures and analyses:
  - Qualitative (portrayal accuracy):
    - Live and video review of SP performances
  - Quantitative (scoring accuracy)
    - Score-based analyses
    - Case level and item level comparisons
- Unscored calibration encounters
- Only marginal score gains for exposed materials
- More than 200 cases

Equating Procedures

- Within site
  - SP-case combination
- Between site
  - Patient note rater – case combination
  - Central video review
    - Data gathering
    - Communication and Interpersonal Skills

Standard Setting

- Step 2 Committee selects minimum passing performance levels for each subscale
- Decision informed by data from:
  - Independent review of videos, checklists, ratings
    - Exercises intended to identify minimally acceptable performance (video review)
  - Survey of constituents: faculty, program directors, licensing boards:
    - Experience with/opinions about clinical skills of recent grads
  - Data on performance levels and score reliability

CHALLENGES

Challenges

- Systems
  - Technical and staffing requirements are complex.
  - Support for registration, scheduling, delivery, scoring, reporting, quality assurance, record-keeping,…
- Security
  - Enormous amount of energy and resources dedicated to deter and detect attempts of the very FEW to subvert the system
  - Elaborate identification systems, test center monitoring, statistical analyses

What goes well?

- Despite the technical challenges, program seems to be efficient and effective - “on” every day
- In the testing community, the USMLE program is held in reasonably high regard
- Support of the volunteer army of faculty members and practicing physicians is remarkable.
- No successful challenges to the overall program from a legal or psychometric perspective.

LESSONS FOR JAPAN

Lessons from USMLE

- Start with a pilot to gain experience
- Need political allies to achieve what is educationally justified
- The passing standard has to be tolerable to schools and licensing authorities
- Need a cadre of professional SPs
- Quality assurance must be rigorous

Questions for Japan

- Do you need a clinical skills examination?
- Does it need to be a national examination?
- Will the licensing authorities deny a license to those who fail?
- How many centers do you need?
- Who will pay?
- Who will oppose its introduction?

Conclusions

- Educators and licensing authorities have a duty to confirm the clinical skill of their graduates and licensees
- Introducing additional high-stakes summative assessment can be controversial
- Exam fairness and consistency enhanced by:
  - Sound test and case development practices
  - Intensive SP training
  - Rigorous quality assurance, equating and standard setting procedures
- Clinical skills assessment can be achieved by overcoming logistical and psychometric challenges
- Assessment positively drives behavior

Thank You

Communication Skills

- **Questioning skills:**
  - *use of* open-ended questions, transitional statements, facilitating remarks
  - *avoidance of* leading or multiple questions, repeat questions unless for clarification, medical terms/jargon unless immediately defined, interruptions when the patient is talking
  - *accurately summarizing* information from the patient
- **Information-sharing skills:**
  - *acknowledging* patient issues/concerns and clearly responding with information
  - *avoidance of* medical terms/jargon unless immediately defined
  - *clearly providing*
    - counseling when appropriate
    - closure, including statements about what happens next

Communication Skills

- **Professional manner and rapport:**
  - *asking about*
    - expectations, feelings, and concerns of the patient
    - support systems and impact of illness, with attempts to explore these areas
  - *showing*
    - consideration for patient comfort during the physical examination
    - attention to cleanliness through hand washing or use of gloves
  - *providing* opportunity for the patient to express feelings/concerns
  - *encouraging* additional questions or discussion
  - *making*
    - empathetic remarks concerning patient issues/concerns
    - patient feel comfortable and respected during the encounter