The Seventh Annual MHPE Summer Conference

RESEARCH THAT MOVES THE FIELD FORWARD: BROADENING THE SCOPE OF HEALTH PROFESSIONS EDUCATION SCHOLARSHIP

Master of Health Professions Education
2006 Summer Conference
July 27-28, 2006

STEVEN M. DOWNING, PHD - CONFERENCE ORGANIZER
UNIVERSITY OF ILLINOIS AT CHICAGO

GEOFF NORMAN, PHD - CONFERENCE KEYNOTE SPEAKER
MCMASTER UNIVERSITY

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

The Seventh Annual MHPE Summer Conference

July 27-28, 2006
Room 106 CMW & Faculty Alumni Center

Sponsored by:
Department of Medical Education
Steven M. Downing, PhD -
Conference Organizer

MHPE Vision Statement
Better Healthcare through Effective and Innovative Health Professions Education Leaders and Scholars

Purpose: The Master of Health Professions Education (MHPE) program offers a degree in education for health professionals. Most students are health professionals (especially physicians) in practice whose career goals include educational leadership. The purpose of this activity is to provide a venue for MHPE students, alumni, and faculty to present and discuss innovations in health professions education, and for physicians in the audience to learn about ongoing educational research and program development topics.

Intended Audience: Physicians in any discipline and other health professions educators.

Program Objectives:
At the conclusion of this program, participants should be able to:

1. Describe newer methods of performance assessment in the clinical teaching setting
2. Evaluate the effectiveness (in medical education) of newer test development and psychometric models
3. Discuss specialty choice issues in light of training issues in medical education
4. Apply newer concepts of test validation to contemporary measurement issues in medical education
5. Describe innovative methods and issues in contemporary health professions education curricula

The University of Illinois at Chicago (UIC) College of Medicine is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

The University of Illinois at Chicago (UIC) College of Medicine designates this educational activity for a maximum of 11 AMA PRA Category 1 Credits™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

The Seventh Annual MHPE Summer Conference

July 27-28, 2006

Steven M. Downing, PhD
Conference Organizer
Department of Medical Education, College of Medicine/University of Illinois at Chicago

RESEARCH THAT MOVES THE FIELD FORWARD: BROADENING THE SCOPE OF HEALTH PROFESSIONS EDUCATION SCHOLARSHIP

Meals, Breaks, and Reception will be in the Faculty Alumni Lounge (College of Medicine West).
All Paper and Works in Progress Sessions will be in Room 106 (College of Medicine West).

THURSDAY, JULY 27, 2006

8:30 – 9:00 **CONTINENTAL BREAKFAST**

9:00 – 9:05 **WELCOME**
Leslie J. Sandlow, MD
Head, Department of Medical Education
Senior Associate Dean, Educational Affairs/College of Medicine, University of Illinois at Chicago/Chicago, Illinois

9:05 – 10:20 **KEYNOTE ADDRESS**
"Research in Clinical Reasoning: Three Decades of Progress"
Geoff Norman, PhD
Professor, Department of Clinical Epidemiology and Biostatistics
Assistant Dean, Programme for Educational Research and Development/McMaster University/Hamilton, Ontario, Canada

10:20 – 10:35 **BREAK**

10:35 – 11:50 **SESSION 1: PAPERS**

CURRICULUM EVOLUTION TO IMPROVE CLINICAL OUTCOME

CHAIR: SANDRA SUFIAN, PHD; DISCUSSANT: MARCIA EDISON, PHD

Short-Term Impact of a Laparoscopic "Mini-Residency" Experience on Postgraduate Urologists Practice Patterns

Elspeth McDougall, MD

Department of Urology/University of California, Irvine/Irvine, California

Assessing the Development of a Military Pre-deployment Combat Injury Care Educational Program

Brian Eastridge, MD

Brooke Army Medical Center/Fort Sam Houston, Texas ; University of Texas Health Science Center/San Antonio, Texas

Designing, Implementing and Assessing a Multidisciplinary Course in Patient Safety and Quality Outcomes

David Mayer, MD

Department of Anesthesiology, College of Medicine/University of Illinois at Chicago/Chicago, Illinois

11:50 – 12:50 **SESSION 2: PAPERS**

MEASURING THE IMPACT OF INSTRUCTIONAL INTERVENTIONS

CHAIR: MARK GELULA, PHD; DISCUSSANT: STEVEN DOWNING, PHD

Teaching Residents How to Teach in the Clinical Setting: A Controlled Prospective Randomized Study

Michael Giuliano, MD, MEd

Department of Neonatology/Hackensack University Medical Center/Hackensack, New Jersey; Department of Pediatrics/State University of New York Downstate Medical Center/Brooklyn, New York

Construct Validity Testing of the LAPMentor™ Laparoscopic Surgical Simulator

Elspeth McDougall, MD

Department of Urology/University of California, Irvine/Irvine, California

12:50 – 1:50 **LUNCH AND POSTERS**

1:50 – 2:50 **SESSION 3: PAPERS**

INFLUENCES ON DECISION MAKING & PROFESSIONAL OUTCOMES

CHAIR: TIMOTHY MURPHY, PHD; DISCUSSANT: ALAN SCHWARTZ. PHD

Does Delaying the Third Year to Pursue a Complementary Degree or Research Training Adversely Impact Subsequent Academic Performance?

Lotte Dyrbye, MD
Mayo Clinic/Rochester, Minnesota

Gender or Generation? Influences on Medical Student Career Choice

Hilary Sanfey, M.B. Ch.B.
Department of Surgery/University of Virginia Health System/Charlottesville, Virginia

2:50 – 4:05 **SESSION 4: WORKS IN PROGRESS**

CLINICAL EDUCATION: SITUATIONAL LEARNING THAT TRANSCENDS THE SITUATION

MODERATOR: RACHEL YUDKOWSKY, MD, MHPE

Further Validity Evidence for a Third-Year Surgery Clerkship Evaluation System

Julia Corcoran, MD
Department of Surgery, Plastic Surgery/Northwestern University Feinberg School of Medicine ; Children's Memorial Hospital/Chicago, Illinois

A Learner-Centered Method for Outpatient Case Presentations

Terry Wolpaw, MD
Case Western Reserve University School of Medicine/Cleveland, Ohio

Using a Virtual Clinic to Supplement Clinical Experiences in the Third-Year of Medical School: A Pilot Study

Lotte Dyrbye, MD
Mayo Clinic/Rochester, Minnesota

4:05 – 4:20 **BREAK**

4:20 – 5:20 **SESSION 5: PAPERS**

ITEM ANALYSIS IN ASSESSMENT: THE DEVIL IS IN THE DETAILS

CHAIR: PHILIP BASHOOK, EDD; DISCUSSANT: DORTHEA JUUL, PHD

Scale Reliability of Examinations Composed of Testlets

Suja Mathew, MD
Department of Medicine/John H. Stroger Hospital of Cook County/Chicago, Illinois; Rush Medical College/Chicago, Illinois

Comparison of the Classical Test Theory and the One Parametric Logistic Item Response Theory (1PL/Rasch) Models to Evaluate Item Quality of a Pharmacy Law Examination

Hsiang-Wen (Margaret) Lin, M.Sc., Doctoral Student
Department of Pharmacy Administration, College of Pharmacy/University of Illinois at Chicago/Chicago, Illinois

5:30 – 7:00 **MHPE GRADUATE RECOGNITION CEREMONY, 2006 BEST MHPE THESIS AWARD, AND RECEPTION**

7:00 **CONFERENCE ADJOURNS FOR THE EVENING**

FRIDAY, JULY 28, 2006

7:30 – 8:00 **CONTINENTAL BREAKFAST**

8:00 – 9:00 **SCHOLARSHIP PANEL: *MHPE STUDENT AND ALUMNI PERSPECTIVES***

MODERATOR: ILENE HARRIS, PHD

How to Use Your Professional and Class Work as Grist for Scholarship

Bob Sedlack, MD *(Student)*
Mayo Clinic/Rochester, Minnesota

Hilary Haftel, MD, MHPE *(Alumna)*
University of Michigan/Ann Arbor, Michigan

Barbara Barzansky, PhD, MHPE *(Alumna)*
American Medical Association/Chicago, Illinois

9:00 – 10:30 **SESSION 6: WORKS IN PROGRESS**

SO DOES IT WORK? EVALUATION OF COURSES & PROGRAMS

MODERATOR: GEORGES BORDAGE, MD, PHD

Proposed Course in Narrative Medicine to Promote Self and Group Reflection on Communication: What and How to Assess?

Susan Arjmand, MD

Department of Family Medicine/John H. Stroger Hospital of Cook County/Chicago, Illinois

Assessment of an Evidence-Based Medicine Course at the Mexican Army Medical School

Melchor Sanchez-Mendiola, MD

Postgraduate Studies Division/National Autonomous University of Mexico/Mexico City, Mexico

Distinctive Features of Masters Programs of Health Professions Education in the World

Masami Tagawa, MD, PhD

Chiba University Hospital/Department of General Medicine/Health Professional Education Center/Chiba, Japan

10:30 – 10:45 **BREAK**

10:45 – 12:00 **SESSION 7: PAPERS**

HIGH-STAKES EXAMS: IMPROVING VALIDITY AND STANDARD SETTING IN CLINICAL ASSESSMENT

CHAIR: JORGE GIROTTI, PHD; DISCUSSANT: ARA TEKIAN, PHD, MHPE

A Modified Ebel Standard Setting Method for Assessment of Medical Students' Clinical Skills

Saman Aziz, MD, MHPE

Department of Medical Education, College of Medicine/University of Illinois at Chicago/Chicago, Illinois

Assessing Rater Errors in an Objective Structured Clinical Examination

Cherdsak Iramaneerat, MD, MHPE

Department of Educational Psychology, College of Education/University of Illinois at Chicago/Chicago, Illinois

Item Dependency in an Objective Structured Clinical Examination

Cherdsak Iramaneerat, MD, MHPE
Department of Educational Psychology, College of Education/University of Illinois at Chicago/Chicago, Illinois

12:00 – 12:15 **CLOSING REMARKS**

12:15 – 12:30 **HOW WELL DID WE PRESENT?**
Best Paper Presentation Award
Best Work in Progress Presentation Award

12:30 **CONFERENCE ENDS**

1:30 **IRB TRAINING SESSION** *(optional)*

ABSTRACTS

Short-Term Impact of a Laparoscopic "Mini-Residency" Experience on Postgraduate Urologists' Practice Patterns

Elspeth McDougall, MD
Department of Urology/University of California, Irvine/Irvine, California

[Federico A. Corica, John R. Boker, David S. Chou, Shannon M. White, Corollos S. Abdelshehid, Gabriella Stoliar, Leandro G. Sala, Allan M. Shanberg, and Ralph V. Clayman/Department of Urology, University of California, Irvine]

Background & Objective: To assist urologists in acquiring and incorporating laparoscopic techniques in their practice, a five-day mini-residency (M-R) program was established at the University of California, Irvine. We present the follow-up practice patterns for the first 32 participants.

Methods: Between September 2003 and September 2004, 32 urologists underwent either the laparoscopic renal ablative (n=17) or the laparoscopic renal reconstructive (n=15) training module. The 5-day course consists of didactic lectures, pelvic trainer and virtual reality simulator practice, animal laboratory sessions, and observation of operative procedures. In addition, trainees are offered a proctoring experience at their hospital. A questionnaire assessing practice patterns was mailed to participants 1 to 15 months (mean=8 months) after completing the M-R program. Statistical analysis was performed using the Wilcoxon Signed Rank Exact Test.

Results: A 100% response rate was achieved for the questionnaire. Mean participant age was 49 years (31 – 70 years). Although no participant had fellowship training in laparoscopy, the majority (72%) had performed between 5 and 15 laparoscopic cases prior to attending the M-R course. Eight months after completing the program, 26 participants (81%) were practicing laparoscopic surgery. Procedures performed before and after the M-R are compared in the table below.

Type of Laparoscopic Surgery	Before M-R (#)	After M-R (#)	p-value
Hand-Assisted Nephrectomy	41% (13/32)	16% (5/32)	0.008
Radical Nephrectomy	28% (19/32)	53% (17/32)	0.008
Radical Nephroureterectomy	6% (2/32)	44% (14/32)	<0.0005
Adrenalectomy	16% (5/32)	31% (10/32)	0.063
Pyeloplasty	0	25% (8/32)	0.008
Partial Nephrectomy	3% (1/32)	19% (6/32)	0.063
Radical Prostatectomy	6% (2/32)	16% (5/32)	0.250
Donor Nephrectomy	0	6% (2/32)	0.500

Discussion: Our intensive 5-day M-R program encourages urologists with some degree of laparoscopic experience to perform more complex operations. Indeed, the number of participants performing radical nephrectomy and adrenalectomy doubled, while there was a seven-fold increase in the number of course participants performing laparoscopic nephroureterectomy. In addition, there was a decrease in the use of the hand-assisted techniques. Continued follow-up will determine the long-term effectiveness of this training program on urologic practice patterns.

Source of funding: Department of Urology, UCI (Research and Education Fund)

Assessing the Development of a Military Pre-deployment Combat Injury Care Educational Program

Brian Eastridge, MD
Brooke Army Medical Center/Fort Sam Houston, Texas; University of Texas Health Science Center/San Antonio, Texas

Background: In response to a perceived need for advanced medical provider pre-deployment trauma training, a collective triservice effort spearheaded through the Brooke Army Medical Center and U.S. Army Institute of Surgical Research, resulted in the development of the first Joint Forces Combat Trauma Management Course (JFCTMC). The aim of this course expanded upon previous training paradigms often favoring surgical assets with advanced trauma training. The originality of this new module was not only its more broad training aims, but more importantly, its incorporation of contemporary lessons learned into the pre-deployment trauma training package. The basic training outline was developed as a core lecture series with break-out modules for surgery, anesthesia, and emergency medicine/primary care taught over a four-day cycle immediately prior to deployment to Iraq.

Objectives: The objective of this analysis was to evaluate the degree to which participation in the JFCTMC provided appropriate preparation and knowledge of injury care on the battlefield.

Methods: Thirty physicians from the 47th Combat Support Hospital were blindly surveyed at the end of the course of instruction and 8 weeks after deployment to the combat zone. The survey consisted of basic demographic information including medical specialty and prior exposure to trauma care in the scope of routine practice. The survey itself consisted of 15 questions on a Likert scale of 1-7 (strongly disagree – strongly agree), which subjectively assessed the success of the educational program with respect to its impact on relevant knowledge of injury care and subsequent impact upon patient outcomes.

Results: Thirty surveys were completed pre-deployment and 27 (90%) follow-up surveys were completed and returned during the deployment interval. Overall, all providers rated the trauma training as appropriate and timely with an average increase in score from 4.8 to 6.2 between surveys. Stratifying the data by surgical versus non-surgical specialty, non-surgical providers rated a greater overall increased assessment scores starting at a lower baseline rating and demonstrating a higher rating average at the second survey interval.

Discussion: This pilot educational program to train deploying physicians proved subjectively successful in providing appropriate and adequate knowledge to care for injury in the combat zone. Future studies will be required to validate the training paradigm and develop a structured curriculum to build upon the successes of the current educational program in order to improve injury care on the battlefield.

Designing, Implementing, and Assessing a Multidisciplinary Course in Patient Safety and Quality Outcomes

David Mayer, MD
Department of Anesthesiology, College of Medicine/University of Illinois at Chicago/Chicago, Illinois

Background: The ultimate purpose of a curriculum in medical education is to address problems that affect the health of the public. Over the last few years, medical errors, patient safety, and quality health care have emerged as central, public concerns. Since the release of the Institute of Medicine's (IOM) report in 1999 entitled "To Err is Human; Building a Safer Health System," there has been considerable discussion in both the public and private sectors regarding ways to modify the current medical education system to address these growing concerns.

Objectives: The goal of this MHPE 503 project was to design, implement, and assess a two-week course that would introduce multi-disciplinary health science students to patient safety and quality care outcomes. Students participating in the course would begin to learn about, experience, appreciate, and practice many of the non-traditional processes and skills required for optimal patient safety and quality care outcomes in today's rapidly changing healthcare environment.

Methods: (1) A list of the relevant stakeholders was identified to participate in deliberative inquiry aimed at designing a student curriculum in safety and quality. The stakeholders were brought together to:

a. Define what patient safety and quality outcomes knowledge, values, and skills medical students should demonstrate by graduation.
b. Review what is currently being taught in undergraduate medical education.
c. Define the current educational gap that exists in undergraduate medical education.
d. Identify when different aspects of patient safety and quality outcomes education should optimally be introduced into undergraduate medical education.
e. Define educational methodologies and faculty development strategies that would be most appropriate for successful implementation.
f. Identify early barriers to implementation and formulating plans to overcome them (the hidden curriculum).

(2) The outcomes of the deliberative inquiry process were used by the stakeholders to design and implement a two-week, multidisciplinary patient safety course. (3)Assessment tools were also developed by the stakeholders to do formative and summative evaluation of the course. (4)The course was offered in February, 2006. Eighteen students from five different health science backgrounds participated.

Results: We will share the results of our course (which just finished) with meeting participants for interactive discussion and guidance on continued curricular improvement. Areas of assessment that will be presented include pre/post knowledge skills, pre/post motor performance skills, pre/post attitude surveys, student standardized patient roll-play videos, and webcast/podcast modules developed as tool kits for the development of a distance learning patient safety curriculum.

Teaching Residents How to Teach in the Clinical Setting: A Controlled Prospective Randomized Study

Michael Giuliano, MD, MEd
Department of Neonatology/Hackensack University Medical Center/Hackensack, New Jersey; Department of Pediatrics/State University of New York Downstate Medical Center/Brooklyn, New York

Background: Few studies utilizing rigorous research design have demonstrated that we can teach residents to become better clinical teachers.

Objectives: To develop a month-long "teaching rotation" and demonstrate its effectiveness in improving resident teaching ability utilizing a rigorous research design.

Methods: Ninety-three Pediatric residents were randomized, with the treatment group (48) rotated to the intervention site for a month long "teaching rotation." Control residents were assigned to the other sites. All residents were evaluated throughout the academic year for teaching ability on all inpatient services by medical students.

Results: (Tentative) Significant gains in teaching ability in a number of domains were noted after completion of the month however these gains deteriorated quickly over time.

Discussion: Teaching improvement programs need to be prolonged and repeated if we wish to maintain gains in teaching ability.

Construct Validity Testing of the LAPMentor™ Laparoscopic Surgical Simulator

Elspeth McDougall, MD
Department of Urology/University of California, Irvine/Irvine, California

[Tadashi Matsuda, Peter D. Vlaovic, Federico A. Corica, John R. Boker, Leandro G. Sala, James F. Borin, Shannon M. White, Yoshinari Ono, and Ralph V. Clayman/Departments of Urology, University of California, Irvine, USA and Kansai Medical University, Osaka, Japan]

Background & Objective: Construct validation of laparoscopic surgical simulation is a mandatory step prior to its incorporation into the surgical education curricula. Herein we present data on construct validity testing (i.e., ability to differentiate based on surgeon experience) of the Lapmentor™ (Simbionix, Cleveland, OH) laparoscopic surgical simulator.

Methods: The LAPMentor is a computer-based virtual reality simulator for learning basic laparoscopic skills; it features 2 virtual working instruments and a camera. Camera and instrument movements are translated into a virtual surgical environment, which is displayed on a LCD monitor. Medical students (MS), residents/fellows (R/F), experienced surgeons with < 30 laparoscopic cases/year (ES<30), and experienced surgeons with >30 laparoscopic cases/year (ES>30) were tested on 9 skill tasks (SK). Following a single practice trial, participants' performance was recorded. Group scores were compared and correlations of measures comprising each SK were examined.

Results: Mean MS (n=23), R/F (n=24), ES<30 (n=26) and ES>30 (n=30) ages were 26 (range 21 – 32), 31 (range 27 – 39), 49 (range 31 – 70) and 47 years (range 34 – 69), respectively.

Skill Task (SK)	*Comparison of Study Group Scores*
SK1 - Manipulation of 0° camera	ES>30 = ES<30 = R/F = MS (p = 0.7)
SK2 – Manipulation of 30° camera	ES>30 = ES<30 = R/F = MS (p = 0.30)
SK3 – Eye-hand coordination	ES>30 = ES<30 = RF > MS (p < 0.001)
SK4 – Clipping	ES>30 = ES<30 = R/F > MS (p< 0.006)
SK5 – Grasping and clipping	ES>30 = ES<30 = R/F > MS (p < 0.01)
SK6 – Two-handed maneuvers	ES>30 = ES<30 = R/F > MS (p < 0.009)
SK7 – Cutting	ES>30 = R/F > ES<30 = MS (p< 0.01)
SK8 – Fulguration	ES>30 > R/F (p < 0.01) = ES<30 > MS (p< 0.001)
SK9 – Object tranlocation	ES>30 = R/F > ES<30 = MS (p < 0.001)
OVERALL Performance Score	ES>30 = R/F (p = 0.95) > ES<30 (p = 0.0001) > MS (p = 0.0001)

In the higher level skill tasks (SK7, SK8 & SK9), the ES>30 scores tended to be better than the R/F and ES<30, which tended to be similar, and these in turn were significantly better than the MS scores.

Discussion: The non-camera skills (SK3-9) of the LAPMentor laparoscopic simulator can distinguish between the laparoscopically naïve and the laparoscopic, experienced surgeon. Among the 9 tasks, SK8 showed the highest level of construct validity in that it accurately differentiated among the MS, R/F, ES<30 and ES>30.

Source of funding: Department of Urology, UCI (Research and Education Fund)

Does Delaying the Third Year to Pursue a Complementary Degree or Research Training Adversely Impact Subsequent Academic Performance?

Lotte Dyrbye, MD
Mayo Clinic/Rochester, Minnesota

[Matthew R. Thomas, M.D., Neena Natt, M.D., Charles Rohren, M.D, Mayo Clinic, Rochester, MN]

Purpose: An increasing number of medical students break from the traditional curriculum to pursue a complementary degree or obtain training in research. The effect of such a break on subsequent academic performance is unknown. The study examines the impact of delaying the third year on performance after re-entry into the MD degree program.

Method: Graduates of Mayo Medical School (1997 – 2004) were categorized as having a delay of one year (12 –18 months), 2 years (19–30 months) or 3 years or more (31 or more months) between the second and third year of medical school. Comparisons between groups were made using undergraduate grade point average (UGPA), scores on the MCAT, NBME Introduction to Clinical Diagnosis (ICD) exam, NBME Medicine Subject Test (Med Shelf), USMLE Step 1, USMLE Step 2, and end-of-third-year Clinical Skills Examination (CSE). The association between length of time away and academic performance was assessed using univariate analysis and multivariable linear regression models adjusting for UGPA and USMLE Step 1 score.

Results: Forty-four students delayed entry into the third year to pursue other studies: 13 by one year, none by 2 years, and 31 by three or more years. UGPA, MCAT scores, and ICD exam scores were similar between groups. A delay of one year had no significant effect on academic performance parameters. In contrast, a delay of three or more years was associated with lower Med Shelf exam percentiles ($p < .0001$) and CSE scores ($p = .04$) than peers who had no delay; USMLE Step 2 scores were not adversely affected by the delay. Among students with a three or more year delay between the second and third year, the magnitude of the effect on the Med shelf exam ($r = -.38$, $p = .04$), CSE ($r = -.44$, $p = .04$), and USMLE Step 2 ($r = -.44$, $p = .01$) correlated with the time away. Significance remained for the shelf exam and USMLE Step 2 after adjusting for USMLE Step 1 scores, but differences dissolved after UGPA was considered.

Conclusions: A one-year break between the second and third year of medical school does not appear to adversely impact academic performance. In contrast, breaks greater than three years appear to be associated with lower third year NBME Medicine shelf examination score and worse performance on the CSE. Further research is needed to determine the optimal timing for research training and to develop effective interventions to facilitate re-entry into the medical school curriculum.

Gender or Generation? Influences on Medical Student Career Choice

Hilary Sanfey, M.B. Ch.B.
Department of Surgery/University of Virginia Health System/Charlottesville, Virginia

Background: Much has been written about the lifestyle priorities of the current generation of medical students who value protected time for family and friends in their working lives. Some educators attribute the declining interest in general surgery at least in part to the increased number of women in medical schools. Women now constitute nearly half of medical students, and female sex has been demonstrated to be a strong negative predictor of pursuing a career in general surgery, with women only representing 24% of those registered in U.S. general surgery training programs. Since a substantial portion of female students enter medical fields that do not offer a controllable lifestyle; in particular obstetrics and gynecology, a specialty with

lifestyle demands similar to surgery – it seemed more reasonable to hypothesize that both gender and generational differences would be important influences on medical student career choices.

Objectives: To examine and stratify by gender the influences on medical student career choice.

Methods: A web-based survey was emailed to medical students in nine medical schools. Students scored statements on surgical careers on five-point Likert scales regarding agreement and whether these statements encouraged them to pursue surgery. Data were analyzed using the Mann-Whitney U test. Qualitative comments were iteratively coded using a constant comparative method.

Results: 680/1300(52%) respondents were male. Males and females disagreed surgeons lead well balanced lives (68% and 77%), and saw this as a deterrent. 35% of women (3% men $p<0.0001$) were discouraged by a lack of female role models. Compared with students unlikely to do surgery, lower percentages of male (74%vs.65%) and female students (85%vs.58%) likely to do surgery agreed career choice was influenced by their decision to have a family ($p<0.01$ males, $p<0.0001$ females). Of medical students who agreed their skill sets were compatible with surgical careers, similar percentages were likely (30% men vs. 24% women) and unlikely (49% men vs. 54% women) to do surgery. All gender differences were less apparent when students likely to do surgery were compared with students unlikely to do surgery.

Discussion: Today both male and female medical students want flexible and protected time for family and friends in their working careers. Regardless of career choice, the written comments from both men and women bear testament to the importance that all students today attach to family and lifestyle, further supporting our hypothesis. Although many studies have examined the factors influencing medical student career choice, our study is unique in terms of the large number of medical students surveyed, the breadth of student years, and the insightful comments from the medical students.

Further Validity Evidence for a Third-Year Surgery Clerkship Evaluation System

Julia Corcoran, MD
Department of Surgery, Plastic Surgery/Northwestern University Feinberg School of Medicine/Children's Memorial Hospital/Chicago, Illinois

Background: In 2003, the Feinberg School of Medicine of Northwestern University (FSM-NU) introduced a new multidisciplinary, interdepartmental surgery clerkship for third-year students in place of its previous General Surgery clerkship. The clerkship is in its third year at this time. All students have experiences in General Surgery, Genitourinary Surgery, Orthopedic Surgery and Otolaryngology. Group sessions for all students and selective clinical experiences are provided by Neurosurgery, Ophthalmology, Vascular Surgery and Anesthesia. The clerkship has traditional in-patient team ward experiences, apprenticeship experiences with attending surgeons and ambulatory clinic experiences. The goals and objectives for this course were determined by multidisciplinary panel using a needs assessment from FSM-NU graduates practicing in primary care and from the literature from individual disciplines.

A new student evaluation system was developed for this clerkship to reflect both objective assessments of student knowledge and clinical performance (faculty-developed multiple choice midterm examination, NMBE Surgery subject examination and faculty-developed 12 -station summative OSCE), and subjective assessment of performance (Faculty Appraisals Global Performance Ratings). The faculty chose a compensatory system, grading pass-fail at the end of the clerkship based on a composite score rather than set a pass-fail mark for each component of the composite score. The scores from the assessments are standardized to z-scores, scaled to a mean of 100 points and standard deviation of 15 and then weighted (Midterm 15%, NBME

Subject 20%, OSCE 25%, Faculty PA 30% and Participation 10%). The weighted scores are then summed. The Faculty chose a pass-fail cut point of 85 points by using the Hofstee process.

Scores have been remarkably reliable since the clerkship's inception with a mean total score of 100 +/- 1 point and standard deviation of 10.2 +/- 0.8. Validity data for the grades has been examined internally but has not been compared to student performance outside of this clerkship. The goal of this study is to see whether performance in the surgery clerkship is predictive of performance in the other required third-year clerkships, performance on the USMLE Step 2 Examinations, performance in traditional markers of scholastic achievements such as AOA status and quartile rank on the MSPE, and finally on post-graduate training choices.

Methods: Demographic data from the clerkship was provided to the Office of Medical Education in a spreadsheet format. Data requested included the following: Clerkship grades and NBME subject examination scores of required third year clerkships (Medicine, Surgery, Primary Care, OB-Gyn, Pediatrics, Neurology and Psychiatry), USMLE Step 2 Examination scores, AOA status, match status, appearances before student promotions committee, and residency matched. The data were returned de-identified for statistical analysis using SPSS software (version 10, Macintosh).

Results: Results are pending at this time.

Discussion: There are two local points of importance for this study. This clerkship's evaluation structure differs significantly from most of the other clerkships at FSM-NU. In general, faculty performance appraisals contribute a larger amount to clerkship grades and may be supplemented by a NBME shelf test, locally developed test, or a project. The sense of other clerkship directors is that clinical performance should drive the grades in the clinical years. The surgery clerkship faculty maintains that it counts clinical performance for 65% of the grade (OSCE, PA and Participation points), and knowledge assessments (MCQ tests) count 35%, very similar to other clerkships. Anecdotally, we also note that students who have trouble in this clerkship are often the same individuals with difficulties in other clerkships that cannot be easily quantified. This study will provide the opportunity to statistically analyze the situation and see what inferences can be drawn about the grading system in this clerkship.

A Learner-Centered Method for Outpatient Case Presentations

Terry Wolpaw, MD
Case Western Reserve University School of Medicine/Cleveland, Ohio

Background: The fast-paced ambulatory teaching environment often results in passive learners who depend on preceptors to drive learning encounters. Encounters focus on factual information, frequently without promotion or expression of thinking and reasoning behaviors. Can a learner-centered method for case presentations to outpatient preceptors promote expression of clinical reasoning by learners?

Objectives: 1) to change the focus of case presentations from predominant reporting of factual information to expression of thinking behaviors and expression of uncertainties; 2) to teach SNAPPS, a learner centered method for case presentations, to third-year medical students. Research question: Do third-year medical students who use the SNAPPS method for case presentations to family medicine preceptors more often 1) summarize the clinical presentation concisely and thoroughly, 2) provide a narrow but appropriate differential diagnosis, 3) analyze the possibilities in the differential diagnosis adequately, 4) express uncertainties and obtain clarification from preceptors, 5) initiate a discussion of patient management, and 6) identify case-related topics to read about, more often than students in either a comparison or control group setting?

Methods: A randomized posttest only design was used with three levels for the intervention variable: SNAPPS intervention, comparison group, and control group. The SNAPPS intervention consisted of 1) one-hour orientation workshop for students to teach the SNAPPS method for case presentations, 2) two weekly 30-minute follow-up meetings, and 3) preceptor orientation. The comparison group consisted of 1) one-hour outpatient workshop about feedback in the office, and 2) two weekly 30-minute sessions to answer general questions about feedback. The comparison group was used to study the role of increased attention to students. A control group that follows usual clerkship routines was used to study content learning and maturation during the four-week intervention period. Students audiotaped all case presentations during the final week of the four-week clerkship block, unless done in the presence of patients. The audiotapes were coded and analyzed to measure eleven dependent variables organized in six categories of outcomes that correspond to the SNAPPS method. The results for the SNAPPS intervention will be compared to comparison and control groups.

Results: Three variables were evaluated in a preliminary analysis. 1) Summary conciseness: For students in the SNAPPS intervention group, the H&P occupied about 40% of presentations, for comparison group 56%, and for controls 62%. 2) The number of diagnoses in the differential: SNAPPS students averaged 2.6 diagnoses. The comparison and control groups averaged one or less. 3) Identifying topics to read about for self-directed learning: SNAPPS students identified reading selections approximately half the time. Comparison and control students did not identify readings.

Discussion: SNAPPS represents a shift in ambulatory education from a more passive, preceptor-driven approach to a collaborative learning conversation in the context of patient care. The time spent reporting data can be shortened to allow the expression of clinical reasoning; an expectation can be set for giving a differential diagnosis; patient-based reading selections can be facilitated.

Questions For Consultation: Are there other ways of analyzing this data? What is the role of the preceptor in a learner-centered method?

Using a Virtual Clinic to Supplement Clinical Experiences in the Third - Year of Medical School: A Pilot Study

Lotte Dyrbye, MD
Mayo Clinic/Rochester, Minnesota

[M. Thomas, N. Natt/Mayo Clinic, Rochester, Minnesota]

Computer simulation in medical education is gaining popularity. Medical educators are placing increased reliance on simulation to supplement the current apprenticeship-style approach to clinical education and fulfill requirements for exposure to specific problems, such as the LCME ED-2 requirement in the U.S.

We investigated whether a simulated experience would enhance clinical competence as assessed by a performance-based examination.

Third-year students were assigned in blocks to complete a computer simulated (CS) case on chest pain or HIV during their Medicine Clerkship. The CS case consisted of interactive, multimedia, web-based learning modules where the learner obtained a history, conducted an examination, engaged in clinical reasoning and patient management decisions, and received feedback. Students interacted with the site through free-text entry. Two-sample t-tests were used for between group comparison of USMLE Step I scores, NBME Medicine Shelf Examination percentile scores (Med Shelf), and performance on the end-of-third-year clinical skills examination (CSE) chest pain station and HIV station.. Scoring of the CSE station was done by checklist using criterion-referenced pass-fail. Logbooks were analyzed for documentation of clinical

encounters with patients who had chest pain (CP Exp.) or HIV (HIV Exp.). The Institutional Review Board considered the study exempt.

Fourteen students completed the CP CS case, and 17 completed the HIV CS case. There were no differences in USMLE Step I scores or NBME Med Shelf scores between the groups. The mean number of months between the end of the clerkship and the CSE exam was 3.7 (SD 2.86). Although there was no significant difference between the groups in CSE score on the CP or HIV station, students who had completed the CP CS tended to have higher mean checklist scores on the CP CSE station (Mean 13.4 vs. 12.1, $p = .18$) and greater likelihood of passing the CP CSE station (93% vs. 71%, $p = .22$). The impact of the CP CS case on CP CSE station performance became more apparent when CP Exp was taken into account. Thirteen students recorded CP Exp. Among students without CP Exp., those who had completed the CP CS case were more likely at a trend level to pass the CP CSE station than peers who did not complete the CP CS case (100% vs. 67%, $p = 0.06$). In contrast, students with CP Exp. were equally as likely to pass the CSE regardless of whether or not they had completed the CS (75% vs. 78%, $p = 1.0$).

Our pilot study suggests that computer simulations may have utility in filling gaps in clinical experience, while they may have little value for problems encountered in the clinical realm. Whether the failure of the computer simulation case on HIV to impact performance on the HIV CSE case reflects a poor case, a problem with the CSE station, or consequences of lack of clinical reinforcement needs to be explored in larger studies.

Questions:

1) Are there other ways of analyzing this data? 2) How can I better deal with a small sample size?
3) How can I find out why one computer simulation case did not impact performance on a CSE?

Scale Reliability of Examinations Composed of Testlets

Suja Mathew, MD
Department of Medicine/John H. Stroger Hospital of Cook County/Chicago, Illinois;
Rush Medical College/Chicago, Illinois

[Steven M. Downing/Department of Medical Education, University of Illinois at Chicago]

Background: Testlets consist of some introductory or stimulus information, generally in one or two short paragraphs, followed by two or more test items. Testlet-based examinations often violate principles of local independence required for typical internal-consistency reliability estimates. Therefore, reliability coefficients for testlet examinations are likely to be overestimated if individual items, rather than the testlet, are used as the unit of reliability analysis.

Objective: This study aims to assess the practical consequences of ignoring the testlet design in undergraduate medical education examinations.

Method: Reliability estimates were calculated for two pre-clinical selected-response tests, comparing items and testlets as units of reliability analysis. The Spearman-Brown prophesy formula was used to estimate the change in test length needed to compensate for decreased reliability.

Results: Test A, a short MCQ test comprised of 12 testlets containing 37 total scored items, had a reliability decrease of 18% when analyzed according to testlet structure. Test B, comprised of 75 individual items in 38 testlets, had a corresponding 3% decrease in estimated reliability. While modest decreases in reliability were

observed for both analyzed tests, the changes in test length needed to compensate for lower reliability were meaningful. Test A must be increased to about 18 testlets of 3 items each (55 total items) to equal the reliability of a 37-item stand-alone exam. Test B must be increased in length by 4 testlets, for a total of 83 items, to equal the reliability of a 75-item stand-alone test.

Discussion: The practical consequence of ignoring testlet design was greater in Test A than Test B; the greater impact of the testlet structure on reliability in Test A may be explained by its shorter length, fewer total testlets, and the greater number of items per testlet. While testlet-based examinations may increase construct validity evidence for tests in medical education, decrements in score reliability may necessitate longer tests to achieve appropriate reliability indices. Thus, the use of testlet-based exams may yield large impact on item writers and the test development process.

Comparison of the Classical Test Theory and the One Parametric Logistic Item Response Theory (1PL/Rasch) Models to Evaluate Item Quality of a Pharmacy Law Examination

Hsiang-Wen (Margaret) Lin, M.Sc., Doctoral Student
Department of Pharmacy Administration, College of Pharmacy/University of Illinois at Chicago/Chicago, Illinois

[Nicholas G. Popovich/Department of Pharmacy Administration, College of Pharmacy; Steven M. Downing/Department of Medial Education, University of Illinois at Chicago]

Background: To facilitate the determination of examination question performance and score adjustment, an item analysis based on the classical test theory (CTT) and the item response theory (IRT) (i.e., one parametric logistic IRT model/Rasch model) can be employed either for a standardized professional board examination or a classroom achievement assessment. Given the attractive features of these two theories, it is likely that an instructor might be willing to adopt either or both methods for these purposes. However, the question remains as to the consistency and inconsistencies between these two models in terms of examination question quality assessment. And, if inconsistencies do exist, what are the possible explanations for them.

Objectives: The goal of this study was to ascertain the level of consistency and inconsistency associated with examination question quality assessment obtained from the model of classical test theory and the Rasch model for a pharmacy law examination.

Methods: According to the responses to 41 multiple-choice questions, those items without sufficient discrimination (i.e., the point biserial correlation, rpbis, less than 0.15) and appropriate difficulty (i.e., the proportion correct, p value, outside of the range of 0.4 and 0.9) were classified as "flag items" in the CTT model. Those items that violated basic assumptions (i.e., inter-item correlation demonstrated significant negative correlation) and which misfitted the Rasch model (i.e., the unweighted and weighted, INFIT and OUFIT, standardized mean square values, ZSTD, were out of the range of -2 and +2) were classified as problematic items. The discrepancies between the two models were further assessed according to the examination question written quality assessment. .

Results: Twenty-six items (63.41%) were recognized as qualified items in both models. Three qualified items in the CTT model did not fit the Rasch model well, whereas six qualified items in the Rasch model were recognized as "flag" items in the CTT model. The examination question content assessment demonstrated that the CTT model was better able to detect problematic items due to obvious "cues" of keyed options where response patterns were still predictable. The Rasch model detected problematic items due to "cues" from the examination question distractors better.

Conclusion: There was over a 60% agreement on the item quality assessment for the CTT and 1PL/Rasch model in this study. Thus, the study demonstrated that the use of either model would suffice for the instructor. However, the choice of which model to utilize would depend upon the purpose of the statistical analysis for the examination.

Proposed Course in Narrative Medicine to Promote Self and Group Reflection on Communication: What and How to Assess?

Susan Arjmand, MD
Department of Family Medicine/John H. Stroger Hospital of Cook County/Chicago, Illinois

Background: The reading and writing of narrative (short stories, plays, essays, and novels) is proposed as a means to promote reflection among medical students about self as physician, and about physician and patients and physician and colleagues. The process of reading, writing, and reflecting on these issues may enhance the capacity for communication.

Objectives: Qualitative and quantitative assessment instruments might be used to assess a student's capacity for self-reflection and the ability to apply acquired insights to professional communication as a result of narrative medicine training. Positive findings would support the use of this instructional method in a medical school curriculum.

Format: During a six week clinical clerkship in family medicine, a narrative medicine course is proposed in which students meets weekly to discuss reading and writing assignments, and to reflect on gained insights. Students will discuss whether/how these insights might affect their view of themselves as physicians, their relationships with patients and with colleagues.
The author will share what is known about assessment tools being used in eight existing narrative medicine courses in US medical schools and hospitals.

Discussion: Multiple levels of assessment at various stages in the process are needed to evaluate process and outcomes.

Questions for Consideration:

1. What is being assessed in a course like this?
2. What are the most appropriate evaluation tools?
3. What are the relevant conceptual frameworks?

Assessment of an Evidence-Based Medicine Course at the Mexican Army Medical School

Melchor Sanchez-Mendiola, MD
Postgraduate Studies Division/National Autonomous University of Mexico/Mexico City, Mexico

Background: Evidence-based medicine (EBM) is currently thought to be necessary for quality medical practice as a core competency for all healthcare professionals. The Mexican Army Medical School has recently introduced an EBM program for its fifth-year medical students, and is in the process of assessing its

effect. We will assess attitudes, knowledge and skills about EBM, comparing groups with and without the EBM course.

Objectives: To assess the effect of the EBM course in attitudes, knowledge, and skills in the Mexican Army Medical School students.

Methods: The study is designed to compare a group of fifth-year medical students at the end of the course, with a similar control group (the other half of fifth-year) without the educational intervention. We will also compare the intervention group with control groups from the fourth-year and the sixth-year of the medical school, to measure the EBM outcomes in a more junior and a more senior group (the sixth-year group had the course one year previously). The study will use a quasi-experimental design with static-groups comparison for the four-and six-year groups, while the fifth-year group students are randomly allocated to the educational intervention. The EBM outcomes will be measured comparing the fifth-year half-group that just took the course with the fifth-year half without the course, the fourth-year students without the course, and the sixth-year students that had the course one year before. The assessment of attitudes, knowledge, and skills related to EBM will be done with the following instruments: Taylor's questionnaire, a published, validated instrument designed to evaluate the effectiveness of evidence-based medicine teaching, that includes items to assess self-reported skills, knowledge and attitudes regarding evidence-based clinical practice. The other instrument is the final summative test of the EBM Course at our school, which was the end product of the Master in Health Professions Education elective Course on "Test Construction – Developing Objective Achievement Tests in the Health Professions" by a group of MHPE students. The instrument was developed, administered, scored, and analyzed following the 12 steps for effective test development described by Downing. We will compare the attitude scores with the Kruskal-Wallis test, and the knowledge test scores with one-way analysis of variance. A p value of less than 0.05 will be considered statistically significant.

Results: We expect to find a significant difference in attitude and knowledge scores between the groups that had the EBM course when compared to those that haven't taken the course. The sixth-year students, which had the course one year before, probably will have higher scores than the fourth-year and fifth-year students without the course, but lower than the fifth-year students that just took the course.

Discussion: The consultation questions are: what is the proper classification of this study? What would be the best ways of comparing the groups and analyzing the data? What issues of validity of the instruments are more relevant for the study?

Distinctive Features of Masters Programs of Health Professions Education in the World

Masami Tagawa, MD, PhD
Chiba University Hospital/Department of General Medicine/Health Professional Education Center/Chiba, Japan

Background: Recently, new Masters programs for health profession educators are planned in several countries and started in Germany and Australia last year. Distinctive features of each program are very important for the target people. Current and future program providers need to know characteristics of their own and other programs, change of needs, and key issues to maintain the quality and competitiveness of the program.

Objectives: Different types of existing programs for health professions education are analyzed to 1) clarify the characteristics and difficulties of each program with different requirements and restrictions, and 2) elucidate common issues for a masters program to be considered as qualified.

Methods: Masters programs for health professions educators provided by University of Illinois at Chicago, Universiteit Maastricht, University of Dundee, University of Southern California, University of Houston (Houston Medical Center) and The University of New South Wales are qualitatively analyzed. Information related to target and goals, student selection, educational strategies, curricula, staffs, maintenance are obtained by interview of program directors and by written materials from October to November in 2005.

Results: Targets of the programs broadly range from clinical educators to education leaders. Program goals are competent educators, and some programs aim to produce scholars and leaders. Coursework is delivered by part-time learning or short intensive courses, and it provides opportunities for practice and reflection. Coursework design is completely different among programs, because of university, state, and national regulation; requirements of postgraduate education; history of the program; different philosophy of education; and limitation of resources. Program ownership and marketing difficulties may be the problems for maintenance of the qualified program.

Discussion: All programs tried to improve and adapt to a changing environment and market. Because of the different curriculum, outcomes of students might be varied among programs. The difference

A Modified Ebel Standard Setting Method for Assessment of Medical Students' Clinical Skills

Saman Aziz, MD, MHPE
Department of Medical Education, College of Medicine/University of Illinois at Chicago/Chicago, Illinois

Background: In the context of assessment, standards are an expression of values and a systematic way of gathering value judgments, reaching consensus and expressing that consensus as a single score on a test. Standard setting for assessment of clinical knowledge and skills is a critical step in the design of medical school examinations. Despite considerable advances in the design of assessment methods, clinical skills assessment is still fraught with issues of practicality. The Ebel method has been successfully used in written examinations. However, its use with respect to performance-based examination like SPs is not very common.

Objectives: The main aims of this study were to explore the psychometric properties and feasibility of using a modified Ebel method to establish absolute passing standards for standardized patient (SP)-based examinations for medical students.

Methods: This research was conducted using a retrospective, descriptive study design. Five faculty judges were trained to use a modified Ebel method to establish the standards for three SP-based cases, which were part of an examination administered to fourth-year medical students (N=157). Passing scores, passing rates, intraclass reliability coefficients, and Meskauskas indices were computed. Judges completed a post-judgment questionnaire assessing their views about the feasibility of the standard-setting process.

Results: The results indicate wide variation in passing scores and passing rates, as well as reliability coefficients. Passing scores ranged from 40 percent to 52 percent. Pass rates ranged from 50 percent to 91 percent. Intraclass reliability coefficients ranged from 0 to 0.90 for multiple judges and from 0 to 0.64 for a single judge. Meskauskas indices were lower than optimal, with mean SIS1= 1.43 and SIS2 = 0.72. Three out of five judges found the Ebel a feasible method.

Conclusion: Findings suggest that the modified Ebel approach to standard setting requires further refining to be acceptable for assessment of clinical skills. The low reliability coefficients highlight the need for enhanced rater training.

Assessing Rater Errors in an Objective Structured Clinical Examination

Cherdsak Iramaneerat, MD, MHPE
Department of Educational Psychology, College of Education/University of Illinois at Chicago/Chicago, Illinois

[Rachel Yudkowsky/Department of Medical Education, University of Illinois at Chicago]

Background: Objective Structured Clinical Examinations (OSCEs) are widely used for performance assessment of medical students and physicians in various medical specialties worldwide. However, OSCEs are subject to various types of potential measurement errors introduced by standardized patient raters. The sources of variance in performance ratings associated with raters and not with the actual performance of examinees are called rater errors. To ensure the objectivity in OSCEs, we should monitor and control these rater errors. A multi-faceted Rasch measurement (MFRM) model is an extension of the basic one-parameter item response theory (IRT) model that allows the estimation of rater severity in addition to examinee's ability and item difficulty. A MFRM model is a useful psychometric approach to assess rater errors in OSCEs.

Objectives: The purpose of this study is to demonstrate how to assess various types of rater errors, including leniency/severity (difference in the levels of severity between raters), rater inconsistency (instability of the level of severity within each rater), halo (rater's tendency to let the rating of one trait influence his/her ratings on other traits), and restriction of range (clustering of ratings around a particular point on the rating scale, failing to use rating categories on the rest of the scale), using a multi-faceted Rasch measurement approach.

Methods: We used a MFRM model to study rater errors in a clinical skills OSCE of 173 fourth-year medical students in a Midwestern medical school. Each student performed six clinical tasks with six Standardized Patients (SPs), who were selected from a pool of 17 SPs. The performance of students in each task was assessed in six domains: history taking, physical examination, interpersonal skills, communication technique, counseling skills, and the manner in conducting physical examination. The ratings of each domain were given on a five-point rating scale, ranging from 1(unacceptable) to 5 (excellent).

Results: The severity levels of SPs varied from -0.76 to 0.55, indicating leniency/severity rater errors, $c2\ (16) = 203.4$, $p < 0.05$. Four SPs (24%) have high infit mean-square statistics, indicating their rating inconsistency. None of the SPs have low infit mean-square statistics, suggesting that these SPs were free from halo and restriction of range rating errors.

Discussion: Like other types of judge-mediated performance assessment, OSCEs are also subject to rater errors. Leniency/severity error and rating inconsistency were major validity threats in this OSCE. Considering students encountered different sets of SPs in this OSCE, leniency/severity rater errors could introduce construct-irrelevant variance into students' scores. Rating inconsistency could lead to construct-irrelevant variance in the scores of students who were rated by the four inconsistent SPs. Nevertheless, all the SPs did a good job in avoiding halo and restriction of range rating errors. We can use the diagnosis information obtained from a MFRM analysis to develop an effective rater training program that addresses the major rating problems and targets specific SPs who clearly exhibited these errors.

Item Dependency in an Objective Structured Clinical Examination

Cherdsak Iramaneerat, MD, MHPE
Department of Educational Psychology, College of Education/University of Illinois at Chicago/Chicago, Illinois

[Carol M Myford/College of Education, University of Illinois at Chicago; Rachel Yudkowsky/ Department of Medical Education, University of Illinois at Chicago]

Background: Unlike traditional multiple-choice test items in which each item is developed independently, items used to evaluate the performance of residents in each OSCE station are linked to the same clinical situation and are rated by the same standardized patient (SP). The level of clinical performance on one item may be dependent on the level of performance on other items for the same OSCE station, violating the assumption of conditional item independence that underlies the multi-faceted Rasch measurement (MFRM) model.

Objectives: The purposes of this study were (1) to check for the existence of item dependency in an OSCE, (2) to outline an alternative approach for analyzing rating data with the MFRM model to ameliorate the problem of item dependency, and (3) to compare reliability estimates, parameter estimates, and fit statistics obtained from MFRM analyses when the assumption of local independence was violated and when the assumption was met.

Methods: We used a MFRM model to analyze ratings from a communication skills assessment of 79 internal medicine and family medicine residents, using six OSCE stations, each scored on 18 five-point rating scale items. We first used the resident's rating on each item as a scoring unit and checked for the amount of item dependency using Fisher's Z statistic, a modification of Yen's Q3 index of local item dependence. We then combined the ratings of all items for each OSCE station to produce a station score, which we considered as a scoring unit in another MFRM analysis. We checked the amount of item dependency resulting from this scoring approach. We then compared reliability estimates, parameter estimates, and fit statistics obtained from the two approaches.

Results: When treating the rating on each item as a scoring unit, the MFRM analysis revealed item dependency in 65% of the item pairs within an OSCE station. This resulted in overestimation of resident separation reliability and inaccurate parameter estimation. When using station scores as scoring units, item dependency was reduced to 27%. This helped improve the standardized fit statistics for resident competency, as well as the standardized and unstandardized fit statistics for station difficulty. However, the loss of information from combining item scores into station scores reduced resident separation reliability from 0.94 to 0.74.

Discussion: Conducting MFRM analysis using dependent items as scoring units is a violation of a basic psychometric assumption of the model that may result in overestimation of reliability estimates as well as inaccurate resident competency measures and station difficulty measures. We recommend routine evaluation of item dependency in the analysis of OSCE data when using a MFRM model. If significant item dependency

is present, combining item scores into station scores should help alleviate the problem, resulting in more realistic reliability estimates, and producing more accurate measures of residents' competence and station difficulty.

MHPE Graduate Recognition Ceremony

Thursday, July 27, 2006 at 5:30 pm

Welcome	***Ilene B. Harris, PhD****; Professor and Director of Graduate Studies, Department of Medical Education, University of Illinois at Chicago*
Remarks	***Leslie J. Sandlow, MD****; Head, Department of Medical Education; Senior Associate Dean, Educational Affairs, College of Medicine, University of Illinois at Chicago*
	Joseph A. Haherty, MD*; Dean, College of Medicine, University of Illinois at Chicago*
	Jonathan J. Art, PhD*; Associate Dean, Graduate College, University of Illinois at Chicago*
Presentation of the Graduates and their Advisors	***Ilene B. Harris, PhD***
Presentation of the 2006 Best MHPE Thesis Award	***Marcia I. Edison, PhD****; Research Assistant Professor, Department of Medical Education, University of Illinois at Chicago*
Closing Remarks	***Ilene B. Harris, PhD***

Graduates and Advisors

Albert A. DePolo, Jr. *(Fall 2005): Needs and Skill-Level Assessment of Osteopathic Surgical Program Directors*
✧*Advisors: M. Seefeldt, S. Downing, M. Garg*

Medha Joshi *(Fall 2005): Design of a Professionalism Curriculum for Residents at the Rajiv Gandhi University of Health Sciences*
✧*Advisors: I. Harris, M. Gelula, L. Sandlow, M. Seefeldt, R. Yudkowsky*

Sheryl Koski *(Fall 2005): Development of an Instrument for the Evaluation of Clinical Teaching in the Medical Oncology Program*
✧ *Advisors: M. Seefeldt, A. Schwartz, C. Butts*

Masami Tagawa *(Summer 2006): Development of a Master's Program of Health Professions Education Leadership in Japan*
✧ *Advisors: M. Gelula, I. Harris, M. Seefeldt*

John Frederick Watkins *(Fall 2005): The Factors Influencing Image Quality on the Canadian Qualifying Examination in Plastic Surgery*
✧ *Advisors: A. Schwartz, S. Downing, J. Davidson*

A Note on Academic Regalia and the Presentation of Stoles

Academic regalia originated in the twelfth century medieval European universities of Bologna, Oxford, Cambridge, and Paris. The academic costume that we have today developed from the long robe and hood garments worn by scholars who were primarily monks and friars. Their dress met practical needs and incorporated church and state ceremonial traditions. Later, the beautiful robes of Roman Popes and the garments of church prelates set the tradition followed by bishops and vice chancellors as they became heads of universities. Universities developed regulations dictating costume styles to distinguish their officials from doctors, from lesser clerics, and from townspeople.

The use of academic costumes in this country was limited and sporadic before the Civil War. Subsequently, a renewed interest was spurred by the growth of American universities and their graduate programs and by increased contact with European universities. Also, students wanted to wear garments that would distinguish them as graduates at their graduation ceremonies.

Academic regalia include a gown, a cap, a hood, and a stole worn over the hood. Each of these components of the regalia contributes to the identification of the degree, the institution, and the field studied. At this ceremony, we are presenting each graduate with a stole. The sky blue identifies the specialty training area of the graduate as Health Professions Education.

Moved? New email address?

If so and you would like to receive communications from the program, please take a few minutes and fill out the form below. Please place this in the box marked **"Mailing List"**. You may also at any time, email your contact information to Laura Schaaf at lpschaaf@uic.edu .

Name: _____

Last **First** **Credentials**

Preferred Mailing Address: _____ ☐ **Home** ☐ **Work**

_____

Preferred Email Address: _____ ☐ **Home** ☐ **Work**