



ความรู้พื้นฐาน สำหรับครูแพทย์มือใหม่

ทักษะที่ครูแพทย์มือใหม่ต้องรู้
เพื่อให้สอนได้อย่างมีประสิทธิภาพ



SHEE Live Training
อบรมออนไลน์ผ่านระบบ



เนื้อหาการอบรม

ไม่ว่าคุณจะทำอยู่ที่ไหนก็สามารถเข้าอบรมออนไลน์ และ
ร่วมอภิปรายกับวิทยากรได้ทันที

Part 1 : Basic principles and teaching techniques

- Teaching a large class
- Clinical supervision
- Teaching on the run
- Bedside teaching
- Professionalism of teachers
- Teaching in laboratory
- Small group teaching
- Ambulatory teaching
- Simulation

Part 2 : Assessment and evaluation

- Basic principles of assessment
- MCQ item analysis
- Essential skills for thesis advisors
- OSCE item development
- EPA and WPBA
- Performance assessment

Jul Part 2 : พฤ. 23 - ศ. 24 ก.ค. 63
เวลา 8.00 - 16.00 น.

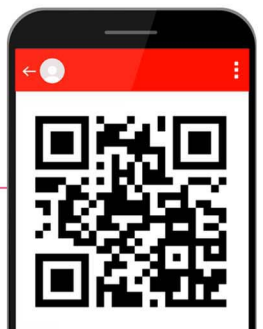
เอกสารประกอบการอบรม



เนื้อหาครอบคลุมตามเกณฑ์
การฝึกอบรมครูแพทย์ของ กสพท.

ติดต่อสอบถาม

Go to SHEE Website



สารบัญ

	หน้า
กำหนดการ	1
รายชื่อผู้ร่วมอบรม.....	3
เอกสารประกอบการอบรม (วันที่ 23 กรกฎาคม 2563).....	5
หัวข้อ : Basic principles of assessment	7
วิทยากร : รศ.ดร. นพ.เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : MCQ.....	15
วิทยากร : รศ.ดร. นพ.เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : MCQ item analysis.....	33
วิทยากร : รศ.ดร. นพ.เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : Constructed response item exam	51
วิทยากร : ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์	
เอกสารประกอบการอบรม (วันที่ 24 กรกฎาคม 2563).....	77
หัวข้อ : Standard setting.....	79
วิทยากร : รศ.ดร. นพ.เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : Essential skills for thesis advisors.....	99
วิทยากร : รศ. ดร. นพ. ถาวรชัย ลิ้มจินดาพร	
ศ. ดร. เพทาย เย็นจิตโสมนัส	
อ. ดร. นพ.บดีรินทร์ ทรัพย์สมบูรณ์	
หัวข้อ : Long case exam.....	105
วิทยากร : รศ.ดร. นพ.เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : OSCE item development	115
วิทยากร : รศ. ดร. นพ. เชิดศักดิ์ ไอรรมณีรัตน์	
รศ. พญ. กษณา รักษมณี	
ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์	
รศ. นพ. สุพจน์ พงศ์ประสพชัย	
รศ. พญ. รัชวรรณ จิระติวานนท์)	
หัวข้อ : EPA and WPBA.....	127
วิทยากร : รศ. ดร. นพ. เชิดศักดิ์ ไอรรมณีรัตน์	
หัวข้อ : Performance assessment.....	185
วิทยากร ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์	
หัวข้อ : Portfolio.....	187
วิทยากร : รศ. ดร. นพ. เชิดศักดิ์ ไอรรมณีรัตน์	
ช่องทางการติดต่อสื่อสาร	251



กำหนดการอบรม

โครงการอบรมเชิงปฏิบัติ เรื่อง ความรู้พื้นฐานสำหรับครูแพทย์มือใหม่

วันที่ 23-24 กรกฎาคม 2563

รูปแบบ SHEE live training ผ่านระบบ ZOOM meeting

Part II: Assessment and evaluation		วิทยากร
วันพฤหัสบดีที่ 23 กรกฎาคม 2563		
08.00-08.30 น.	ลงทะเบียน	
08.30-09.30 น.	Basic principles of assessment	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์
09.30-10.30 น.	MCQ	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์
10.30-10.45 น.	รับประทานอาหารว่าง	
10.45-12.00 น.	MCQ item review	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์ รศ. นพ. ตรีภพ เลิศบรรณพงษ์ รศ. นพ. สุพจน์ พงศ์ประสพชัย ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์
12.00-12.45 น.	รับประทานอาหารกลางวัน	
12.45-13.45 น.	MCQ item analysis	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์
13.45-14.30 น.	Constructed response item exam (Review part 1)	ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์
14.30-14.45 น.	รับประทานอาหารว่าง	
14.45-16.00 น.	Constructed response item exam (Review part 2)	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์ ผศ. นพ. ทศ หายรุ่งโรจน์ รศ. นพ. สุพจน์ พงศ์ประสพชัย ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์
วันศุกร์ที่ 24 กรกฎาคม 2563		
08.00-08.30 น.	ลงทะเบียน	
08.30-09.15 น.	Standard setting	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์
09.15-10.00 น.	ห้อง 1 Essential skills for thesis advisors	รศ. ดร. นพ. ถาวรชัย ลีมีจินดาพร ศ. ดร. เพทาย เย็นจิตโสมนัส อ. ดร. นพ.บดินทร์ ทรัพย์สมบูรณ์
	ห้อง 2 Long case exam	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์
10.00-10.15 น.	รับประทานอาหารว่าง	
10.15-12.00 น.	ห้อง 1 Essential skills for thesis advisors (Continued)	รศ. ดร. นพ. ถาวรชัย ลีมีจินดาพร ศ. ดร. เพทาย เย็นจิตโสมนัส อ. ดร. นพ.บดินทร์ ทรัพย์สมบูรณ์
	ห้อง 2 OSCE item development	รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์ รศ. พญ. กษณา รักษมณี ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์ รศ. นพ. สุพจน์ พงศ์ประสพชัย รศ. พญ. ชัชววรรณ จิระดิวานนท์
12.00-12.45 น.	รับประทานอาหารกลางวัน	
12.45-13.45 น.	ห้อง 1 Essential skills for thesis advisors (Continued)	รศ. ดร. นพ. ถาวรชัย ลีมีจินดาพร ศ. ดร. เพทาย เย็นจิตโสมนัส อ. ดร. นพ.บดินทร์ ทรัพย์สมบูรณ์

	ห้อง 2 EPA and WPBA	รศ. ดร. นพ. เชิดศักดิ์ ไอรณณรัตน์
13.45-14.30 น.	Performance assessment	ผศ. นพ. สุประพัฒน์ สนใจพาณิชย์
14.30-14.45 น.	รับประทานอาหารว่าง	
14.45-15.30 น.	Portfolio	รศ. นพ. ตริภพ เลิศบรรณพงษ์
15.30-16.00 น.	Summary	รศ. ดร. นพ. เชิดศักดิ์ ไอรณณรัตน์

หมายเหตุ กำหนดการอาจมีการเปลี่ยนแปลงตามความเหมาะสม

รายชื่อผู้เข้าร่วมอบรม

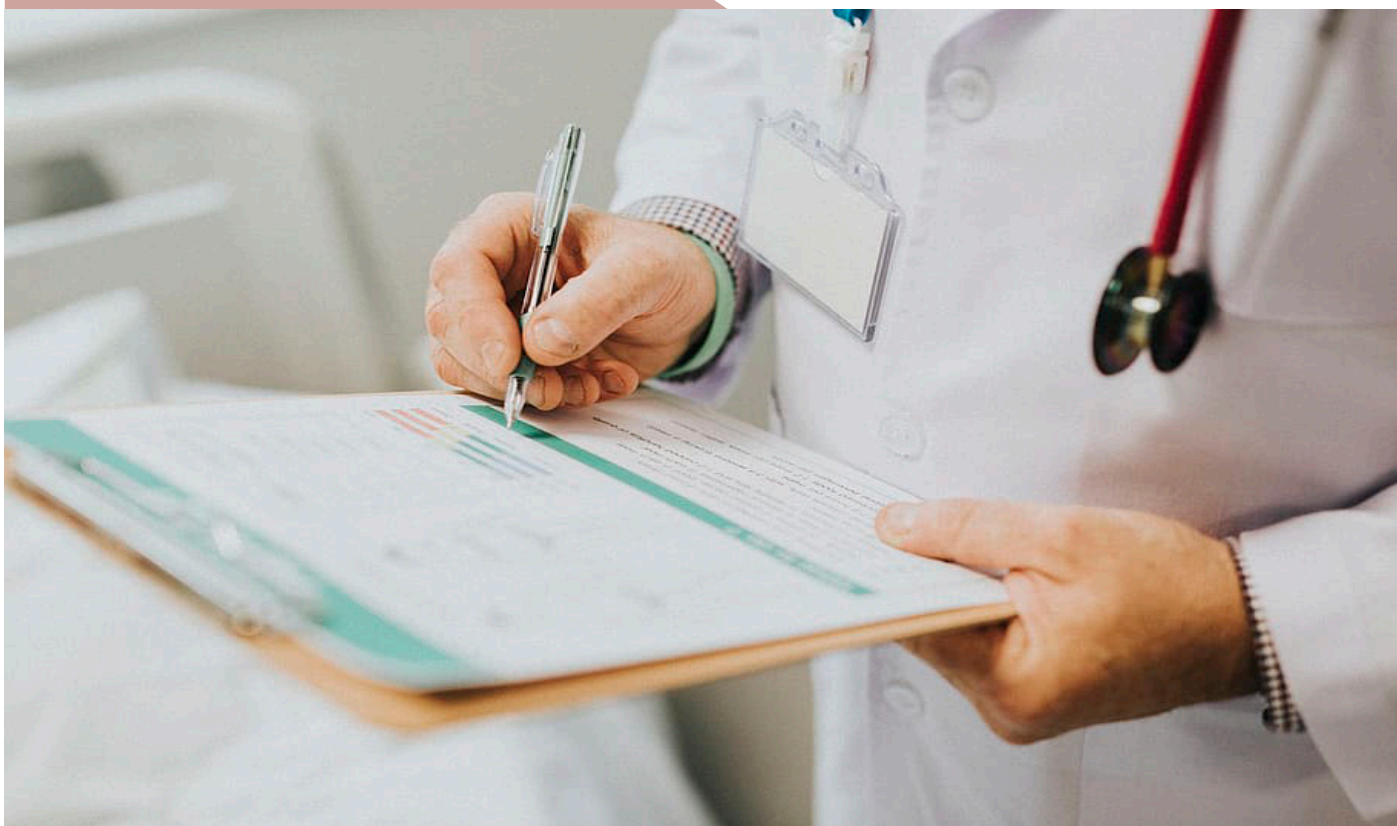
รายชื่อผู้เข้าอบรม Part 2 วันที่ 23-24 กรกฎาคม 2563

ลำดับที่	ชื่อ-สกุล	ตำแหน่ง	หน่วยงาน/ภาควิชา	สังกัด
1	ทพญ. ธิญญาพัทธ์ อิงบุญมีสกุล	แพทย์	ศัลยศาสตร์ช่องปาก	คณะทันตแพทยศาสตร์ มหาวิทยาลัยนครสวรรค์
2	นพ. ธนชัย แสนลิ่ง	แพทย์	ศูนย์ความเป็นเลิศด้านโรคมะเร็ง	คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
3	นพ. อาทิตย์ บุญรอด	แพทย์	ออร์โธปิดิกส์	คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
4	นพ. กัทวัฒน์ เปล่งพานิช	แพทย์	ภ.วิสัญญีวิทยา	คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
5	ทพญ. ศิริญา ปัญญา	แพทย์	ภ.ศัลยศาสตร์	คณะแพทยศาสตร์ มหาวิทยาลัยบูรพา
6	นพ. ภาษา สุขสอน	แพทย์	อายุรศาสตร์	คณะแพทยศาสตร์ มหาวิทยาลัยพะเยา
7	ทพญ. อภิญญา ชิวะพันธ์	แพทย์	ภ.วิทยาศาสตร์การแพทย์พื้นฐาน	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยนวมินทราธิราช
8	ทพญ. จุฑามาศ วันเพชร	แพทย์	ภ.วิทยาศาสตร์การแพทย์พื้นฐาน	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยนวมินทราธิราช
9	ทพญ. ฐาปนี เรืองฤทธิ์	แพทย์	ภ.วิทยาศาสตร์การแพทย์พื้นฐาน	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยนวมินทราธิราช
10	ทพญ. สุธิชา กฤตยารักษ์สกุล	แพทย์	ภ.วิทยาศาสตร์การแพทย์พื้นฐาน	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยนวมินทราธิราช
11	นาย วิศิษฐ์ สมบัติถาวรกุล	ผู้ช่วยอาจารย์	สถานเทคโนโลยีการศึกษาแพทยศาสตร์	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
12	รศ. นพ. พิพัฒน์ เขียววิทย์	อาจารย์	ภ.รังสีวิทยา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
13	ผศ. นพ. พรพจน์ เปรมโยธิน	อาจารย์	ภ.อายุรศาสตร์	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
14	อ. นพ. ภพ กนกโรจน์	อาจารย์	ภ.ศัลยศาสตร์ออร์โธปิดิกส์และกายภาพบำบัด	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
15	อ. นพ. ศิรส จิตประไพ	อาจารย์	ภ.ศัลยศาสตร์	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
16	ทพญ. ณิชชยาณ์ เรชะธีระโรจน์	แพทย์	ภ.วิสัญญีวิทยา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
17	ทพญ. นवलพร กิตติสิน	แพทย์	ภ.วิสัญญีวิทยา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
18	ทพญ. ณิชชิตา สุนทรสิงห์	แพทย์	ภ.เวชศาสตร์การธนาคารเลือด	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
19	ผศ. ทพญ. วีรวดี จันทร์นิภาพงศ์	อาจารย์	ภ.เภสัชวิทยา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
20	อ. ทพญ. ณิชชวดี เลากุลรัตน์	อาจารย์	ภ.พยาธิวิทยา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
21	น.ส. เกษมณี จงเจียมดี	แพทย์แผนไทยประยุกต์	สถานการแพทย์แผนไทยประยุกต์	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
22	ผศ. ดร. วรวรรณ วาณิชย์เจริญชัย	อาจารย์	ฝ่ายการศึกษา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
23	ผศ. ดร. ทศนีย์ยา รัตนฤทัย นพรัตน์แจ่มจรัส	อาจารย์	ฝ่ายการศึกษา	คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
24	ทพญ. สลิล ศิรินาม	แพทย์	ภ.กุมารเวชศาสตร์	คณะเวชศาสตร์เขตร้อน มหาวิทยาลัยมหิดล
25	นพ. เอกรัชต์ แก่นลา	แพทย์	ศัลยกรรม	โรงพยาบาลกาฬสินธุ์
26	ทพญ. วราพร อนุภาพอุกฤษฏ์	แพทย์	สูติศาสตร์และนรีเวชวิทยา	โรงพยาบาลเจริญกรุงประชารักษ์
27	ทพญ. นิชาภา ไถ่บ้านกวย	แพทย์	อายุรกรรม	โรงพยาบาลเจริญกรุงประชารักษ์
28	ทพญ. บุณชรีกา อาจนาศีโย	แพทย์	ภ.วิสัญญีวิทยา	โรงพยาบาลชัยภูมิ
29	นพ. ชีรภัทร อุดยธรรม	แพทย์	แผนกศัลยกรรมออร์โธปิดิกส์	โรงพยาบาลนพรัตนราชธานี
30	ทพญ. อรุณี แสงसानนท์	แพทย์	วิสัญญีวิทยา	โรงพยาบาลนพรัตนราชธานี
31	ทพญ. เสาวนีย์ นาคแก้ว	แพทย์	กุมารเวชกรรม	โรงพยาบาลมหารajahนครศรีธรรมราช
32	ทพญ. กานต์ชนก สุทธิผล	แพทย์	กลุ่มงานเวชกรรมสังคม	โรงพยาบาลราชบุรี
33	นพ. อธิป แสงรุ่งเรือง	แพทย์	รังสีวินิจฉัย	โรงพยาบาลราชบุรี
34	ทพญ. อรพิสาส์ พรหมโสด	แพทย์	กุมารเวชกรรม	โรงพยาบาลราชบุรี
35	ทพญ. มัลลิกา รัตนวิจิตร	แพทย์	กุมารเวชกรรม	โรงพยาบาลลำปาง
36	ทพญ. พัชรธาดา อ่ำสำอางค์	แพทย์	สูตินรีเวช	โรงพยาบาลลำปาง
37	ทพญ. อลิษา จิกรพงษ์	แพทย์	อายุรศาสตร์	โรงพยาบาลลำปาง
38	นพ. อภิศักดิ์ สุนทรพันธ์	แพทย์	ศัลยศาสตร์ทางเดินปัสสาวะ	โรงพยาบาลลำปาง
39	ทพญ. จารุภา เขียวละออง	แพทย์	เวชศาสตร์ฉุกเฉิน	โรงพยาบาลลำปาง
40	ทพญ. ปิยะกาญจน์ วัชรนทร์วงศ์	แพทย์	อายุรศาสตร์	โรงพยาบาลศรีนครินทร์
41	นพ. กัทร เหล่าภัทรเกษม	แพทย์	ออร์โธปิดิกส์	โรงพยาบาลศรีนครินทร์
42	นพ. รชฎ วงศ์จิรัฎฐิติกาล	แพทย์	ภ.อายุรศาสตร์	โรงพยาบาลศรีนครินทร์
43	นพ. จักรทิพย์ สุทธินรากร	แพทย์	ภ.วิสัญญีวิทยา	โรงพยาบาลศรีนครินทร์
44	ทพญ. สุมลีน ชุมคช	แพทย์	ภ.เวชศาสตร์ฉุกเฉิน	โรงพยาบาลสงขลา
45	ทพญ. ภูมรินทร์ ภูมิศิริโพบูลย์	แพทย์	สูตินรีเวช	โรงพยาบาลสงขลา

รายชื่อผู้เข้าอบรม Part 2 วันที่ 23-24 กรกฎาคม 2563

ลำดับที่	ชื่อ-สกุล	ตำแหน่ง	หน่วยงาน/ภาควิชา	สังกัด
46	พ.ญ. ศรดา ไม้สุพร	แพทย์	เวชศาสตร์ฉุกเฉิน	โรงพยาบาลสระบุรี
47	นพ. ประธาน บุรณกุลกิจการ	แพทย์	อายุรกรรม	โรงพยาบาลสิรินธร
48	นพ. ณัฐพงษ์ อังศ์ธาสาวิตย์	แพทย์	โสต ศอ นาสิก	โรงพยาบาลสิรินธร
49	พญ. จิราภา คชวัฒน์	แพทย์	เวชศาสตร์ครอบครัว	โรงพยาบาลสิรินธร
50	พญ. ประทุมทิพย์ พงศ์พรทรัพย์	แพทย์	อายุรกรรม	โรงพยาบาลสิรินธร
51	พญ. สาธิตา ชัยเลิศวัฒนกุล	แพทย์	อายุรกรรม	โรงพยาบาลสิรินธร
52	นพ. ธนกฤต โรจน์ดำรงการ	แพทย์	อายุรกรรม	โรงพยาบาลสิรินธร
53	นพ. คเชนทร์ ผลสว่าง	แพทย์	อายุรกรรม	โรงพยาบาลสิรินธร
54	พญ. จันทร์เพ็ญ เพียรมานะกิจ	แพทย์	วิสัญญีวิทยา	โรงพยาบาลสิรินธร
55	พญ. กัญจนา ดวงจำปา	แพทย์	กุมารเวช	โรงพยาบาลสิรินธร
56	พญ. อูติพรรณ ชยวงศ์รุ่งเรือง	แพทย์	สูติ-นรีเวชกรรม	โรงพยาบาลสิรินธร
57	พญ. วิภาวี ภัทรจินดานวงศ์	แพทย์	สูติ-นรีเวชกรรม	โรงพยาบาลสิรินธร
58	นพ. ภัทรพล ชนชนนทกานต์	แพทย์	ศัลยกรรม	โรงพยาบาลสิรินธร
59	พญ. อลิสรดา ศรีนิลทา	แพทย์	ศัลยกรรม	โรงพยาบาลสิรินธร
60	นพ. ร่มเย็น จิตมุงงาน	แพทย์	ภ.ศัลยศาสตร์	ศูนย์การแพทย์กาญจนาภิเษก

เอกสารประกอบการอบรม



23 July 2020

รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์

หัวข้อ : Basic principles of assessment

Basic Principles of Assessment

นพ. เชิดศักดิ์ ไอรมณีรัตน์
ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล
มหาวิทยาลัยมหิดล

Assessment

- The process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs.

Assessment drives instruction.

*"Purposeful assessment
drives instruction and affects
learning."*

Wisconsin's guiding principles for teaching and learning

Outline

- Assessment and instruction
- Basic considerations in planning an assessment
- Guidelines for effective assessment
- Assessment pyramid

Assessment and Instructional Process

- **Placement**
 - Aims at determining the readiness of students for the planned instruction
- **Formative**
 - Aims at providing feedback to students and teachers concerning learning successes and failures
- **Summative**
 - Aims at determining the extent to which instructional goals have been achieved; used primarily for assigning grades

Activity

- ให้อาจารย์แต่ละกลุ่ม ช่วยกันระดมสมอง หาวิธีการประเมินการเรียนรู้ในวัตถุประสงค์ต่อไปนี้
 1. พฤตินิสัย คุณธรรม จริยธรรม
 2. ทักษะในการสื่อสาร
 3. ทักษะการทำหัตถการเพื่อตรวจรักษาโรค
 4. ทักษะการพัฒนาศักยภาพทางวิชาชีพ
 5. ทักษะการตรวจร่างกายผู้ป่วย
 6. ความสามารถในการส่งเสริมสุขภาพและป้องกันโรค (เวลา 5 นาที)

Question

- อาจารย์รู้ได้อย่างไรว่าวิธีประเมินที่เสนอมานี้เป็นวิธีที่ดี

Criteria for Good Assessment

- Validity
- Reliability (Reproducibility)
- Equivalence
- Feasibility
- Educational Effect
- Catalytic Effect
- Acceptability

Norcini J, et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 conference. Med Teach 2011; 33 (3) 206-14.

1. Validity

- The extent to which an assessment instrument measures what it intends to measure
- The degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests

Validity Threats

- **Construct Underrepresentation**

The degree to which a test fails to capture important aspects of the construct. The test does not adequately sample some parts of the content

- **Construct-Irrelevant Variance**

The degree to which test scores are affected by processes that are extraneous to its intended construct

2. Reliability

- Consistency of test scores
 - If we test the students/residents again, will they get the same scores?
- Range: 0 – 1
- High values: highly consistent test scores

How Much is Enough?

- Depends on test scores uses
 - High-stakes exam: 0.9 or higher
 - Medium-stakes exam: 0.80 – 0.89
 - Low-stakes exam: 0.70 – 0.79

3. Equivalence

- การทดสอบหัวข้อเดียวกันกับนักศึกษาระดับชั้นเรียนเดียวกัน ที่จัดสอบกันต่างเวลา ได้คะแนนที่เทียบเคียงกันได้

4. Feasibility

ความเป็นไปได้ของการจัดสอบ

The assessment is practical, realistic, and sensible, given appropriate contexts:

- Time
- Money
- Expertise
- Administration

5. Educational Effect

- การประเมินผลนั้นกระตุ้นให้ผู้เรียนมีการเรียนรู้ในเรื่องที่ควรเรียนรู้ ... educational benefit

6. Catalytic Effect

- การประเมินผลก่อให้เกิดการนำผลของการสอบไปใช้ให้ feedback เพื่อสร้าง หรือส่งเสริม หรือสนับสนุนการเรียนรู้ของนักศึกษา

7. Acceptability

- ผู้เกี่ยวข้อง (stakeholders) ทั้งหมดเชื่อถือผลการประเมิน

Guidelines for Effective Assessment (1)

1. Effective assessment requires a clear conception of all intended learning outcomes.
2. Effective assessment requires that a variety of assessment procedures be used.
3. Effective assessment requires that the instructional relevance of the procedures be considered.

Guidelines for Effective Assessment (2)

4. Effective assessment requires an adequate sample of student performance.
5. Effective assessment requires that the procedures be fair to everyone.
6. Effective assessment requires the specifications of criteria for judging successful performance.

Guidelines for Effective Assessment (3)

7. Effective assessment requires feedback to students that emphasizes strengths of performance and weaknesses to be corrected.
8. Effective assessment must be supported by a comprehensive grading and reporting system

Assessment Approaches



21 21

Activity

- ให้อาจารย์ระดมสมอง รวบรวมวิธีการประเมินในแต่ละระดับของ Miller's pyramid
- (เวลา 2 นาที)

Questions & Comments

Cherdsaklramaneerat@gmail.com

Summary

- Assessment and instruction
- Basic considerations in planning an assessment
- Guidelines for effective assessment
- Miller's pyramid

Iramaneerat C. Validity threats [Thai]. Medical Education Pamphlet 2006; 2(9): 1.

สิ่งไม่พึงประสงค์ในการสอบ เชิดศักดิ์ ไอรมนิวัตน์

ในบทความนี้ผมจะขอกล่าวถึงสิ่งอันไม่พึงประสงค์ในการสอบ (Validity threats) ที่เราต้องคำนึงถึงในการจัดสอบ ดังที่ได้กล่าวในบทความก่อนหน้านี้แล้วว่า Validity นั้นคือการประเมินคุณค่าของการแปลผลและการนำผลสอบไปใช้ ดังนั้น สิ่งอันไม่พึงประสงค์ในการสอบ หรือ validity threats ก็คือสิ่งใดก็ตามที่เข้ามาบรบกวนการแปลผลสอบ สิ่งรบกวนเหล่านี้แยกได้เป็น 2 ปัจจัยหลัก คือ construct underrepresentation และ construct-irrelevant variance

Construct underrepresentation หมายถึงการประเมินผลที่ไม่ครอบคลุมสิ่งที่ต้องการวัดอย่างเพียงพอ ทำให้ผลการสอบไม่สามารถบ่งบอกถึงความสามารถของนักเรียนผู้สอบในเรื่องที่ต้องการวัดผลอย่างครบถ้วน ตัวอย่างเช่นในการสอบ OSCE เพื่อวัดความสามารถของแพทย์ประจำบ้านในการให้คำแนะนำปรึกษาแก่ผู้ป่วย หากเกณฑ์การให้คะแนนมีเพียงหัวข้อที่เกี่ยวกับการพูดกับผู้ป่วย แต่ไม่มีหัวข้อที่เกี่ยวกับการใช้ วิชาเฉพาะ เช่น การใช้ท่าทาง น้ำเสียง การรับฟังปัญหา เป็นต้น ก็จัดว่า ทำการประเมินไม่ครอบคลุมเนื้อหา ผลการประเมินก็นำไปใช้บอกได้เพียงว่าแพทย์ประจำบ้านให้ข้อมูลผู้ป่วยครบถ้วน แต่ไม่สามารถบอกได้ว่าแพทย์ประจำบ้านทำการสื่อสารกับผู้ป่วยได้ดีในทุกด้าน ในการสอบข้อเขียนสำหรับวัดความรู้ของนักเรียน หากใช้ข้อสอบที่สั้นเกินไป มีจำนวนข้อสอบไม่กี่ข้อ ก็จะมีปัญหาที่ไม่สามารถวัดความรู้ของนักเรียนได้ครอบคลุมเนื้อหาที่ต้องการวัดผล

Construct-irrelevant variance หมายถึง ปัจจัยอื่นที่นอกเหนือไปจากความรู้ความสามารถของนักเรียนที่สามารถส่งผลต่อคะแนนสอบของนักเรียนได้ ปัจจัยที่อาจรบกวนคะแนนสอบ multiple-choice examination ได้แก่

- ข้อสอบที่ไม่มีคุณภาพ โจทย์คำถามกำกวม มีตัวเลือกที่ถูกมากกว่า 1 ตัวเลือก ทำให้นักเรียนที่มีความรู้ตอบผิด หรือโจทย์คำถามบอกใบ้ให้นักเรียนตอบถูกโดยไม่ต้องใช้ความรู้ ข้อสอบเก่าที่รั่วไหลออกจากคลังข้อสอบทำให้นักเรียนที่รู้ข้อสอบมาก่อนสามารถตอบได้โดยไม่ต้องคิด
 - นักเรียนที่ทุจริตในการสอบ ลอกข้อสอบของเพื่อน หรือใช้วิธีการอื่นในการได้มาซึ่งคำตอบโดยที่ไม่ได้ใช้ความรู้ในเรื่องที่ทำการสอบ
 - อาจารย์ที่บอกข้อสอบให้นักเรียนในการสอน ทำให้นักเรียนที่ท่องคำตอบเข้าไปสอบ ทำข้อสอบได้โดยไม่ต้องคิด สำหรับการสอบในรูปแบบอื่นที่ต้องใช้กรรมการให้คะแนน เช่น OSCE การสอบข้อสอบบรรยาย หรือการสอบปากเปล่า นั้นจะมีปัจจัยที่เกี่ยวข้องกับกรรมการผู้ให้คะแนนเข้ามาบรบกวนการแปลผลคะแนนสอบได้ด้วย เช่น
 - ความไม่เสมอภาคของอาจารย์ในเกณฑ์การให้คะแนน นักเรียนที่สอบกับอาจารย์ที่กดคะแนน เสียเปรียบนักเรียนที่สอบกับอาจารย์ที่ใจดี และปล่อยคะแนน
 - ความไม่สม่ำเสมอของอาจารย์ในการให้คะแนน อาจารย์บางท่านมีแนวโน้มจะให้คะแนนต่ำลงในกลุ่มนักเรียนที่สอบตอนท้าย เนื่องด้วยความเหนื่อยล้า ในขณะที่อาจารย์บางท่านมีแนวโน้มจะให้คะแนนสูงขึ้นในตอนท้ายของการสอบ เนื่องจากได้เห็นความสามารถของนักเรียนจำนวนหนึ่งแล้วพบว่าเกณฑ์ที่ตั้งเป้าไว้นั้นสูงเกินความสามารถของนักเรียนส่วนใหญ่จึงปรับเกณฑ์การให้คะแนนให้ง่ายลง ทำให้นักเรียนในกลุ่มหลังได้คะแนนง่ายขึ้น
 - การจำกัดช่วงของคะแนน ที่พบบ่อยคืออาจารย์บางท่านนิยมเดินสายกลาง ไม่ว่าจะนักเรียนจะทำดีมากหรือน้อยเพียงใด ก็มักจะให้คะแนนอยู่ในเกณฑ์ปานกลาง ไม่กล้าให้คะแนน 0 ในรายที่ทำไม่ดี แต่ก็ไม่กล้าให้คะแนนเต็มในนักเรียนที่ทำได้ดี
- ปัจจัยต่างๆ เหล่านี้ เป็นสิ่งที่ผู้จัดสอบต้องคำนึงถึงเสมอในการจัดสอบและตั้งมาตรฐานเพื่อควบคุมและกำจัดปัจจัยรบกวนเหล่านี้จากการสอบ เพื่อให้ได้ผลการสอบที่มีความเที่ยงตรง เป็นธรรม และสามารถใช้อธิบายความรู้ ความสามารถของนักเรียนได้ตามที่ต้องการ

Iramaneerat C. Reliability: Part I [Thai]. Medical Education Pamphlet 2006; 2(10): 4.

Iramaneerat C. Reliability: Part II [Thai]. Medical Education Pamphlet 2006; 2(11): 4.

ความแม่นยำของคะแนนสอบ (Reliability)

เชิดศักดิ์ ไอรรมณีรัตน์

ในบทความนี้ผมจะขอกล่าวถึงการประเมินความแม่นยำของคะแนนสอบ (Reliability) การตรวจสอบความแม่นยำของคะแนนสอบเป็นการตอบคำถามว่า หากทำการสอบซ้ำนักเรียนจะได้คะแนนเท่าเดิมหรือไม่ ในการสอบทั่วไปมักรายงานความแม่นยำของคะแนนสอบด้วยค่า reliability coefficient ซึ่งมีค่าได้ตั้งแต่ 0 ถึง 1 โดยค่ายิ่งสูงบ่งบอกว่าผลสอบมีความน่าเชื่อถือมาก ค่า reliability coefficient = 0 บอกถึงคะแนนสอบที่ขาดความแม่นยำโดยสิ้นเชิง เทียบได้กับการให้คะแนนนักเรียนโดยการสุ่มตัวเลขให้ ส่วนค่า reliability coefficient = 1 บอกถึงคะแนนสอบที่มีความแม่นยำมาก หากให้นักเรียนสอบซ้ำก็จะได้คะแนนเท่าเดิม เพื่อขยายความเข้าใจผมจะขอกล่าวถึงคุณลักษณะที่สำคัญของ reliability ได้แก่

1. Reliability เป็นคุณสมบัติของคะแนนสอบ ไม่ใช่ตัวข้อสอบ ข้อสอบชุดหนึ่งทำการสอบกับนักเรียนกลุ่มหนึ่งพบว่ามี ความแม่นยำสูง แต่เมื่อเอาข้อสอบชุดเดียวกันไปทำการสอบนักเรียนอีกกลุ่มหนึ่ง อาจมีความแม่นยำต่ำได้
2. Reliability มีด้วยกันหลายชนิด และค่า reliability coefficient ที่ได้จากการประเมินความแม่นยำแต่ละชนิดก็แปลผลแตกต่างกัน ดังได้กล่าวแล้วว่า การประเมินความแม่นยำของคะแนนสอบ เป็นการตรวจสอบว่าหากทำการสอบซ้ำจะได้คะแนนเท่าเดิมหรือไม่ ประเด็นสำคัญคือเราจะทำการสอบซ้ำอย่างไร จะสอบซ้ำด้วยข้อสอบชุดเดิม หรือ ข้อสอบชุดใหม่ที่ยกแบบให้เปรียบเทียบได้กับข้อสอบชุดเดิม, สอบซ้ำ ณ เวลาเดียวกัน หรือใกล้เคียงกัน หรือเวลาห่างกันเป็นสัปดาห์, สอบซ้ำโดยใช้กรรมการให้คะแนนคนเดิม หรือสอบซ้ำโดยเปลี่ยนกรรมการให้คะแนน จะเห็นได้ว่า วิธีการสอบซ้ำต่างกันก็บอกความแม่นยำของคะแนนในสถานการณ์ต่างกัน (ความแม่นยำเมื่อเปลี่ยนชุดข้อสอบ หรือความแม่นยำเมื่อเปลี่ยนเวลา หรือ ความแม่นยำเมื่อเปลี่ยนกรรมการให้คะแนน) ดังนั้นการแปลผลของค่า reliability coefficient ต้องทำความเข้าใจว่าค่าดังกล่าวบ่งบอกถึงความแม่นยำชนิดใด โดยทั่วไปในการวัดความแม่นยำของคะแนนสอบ multiple-choice examination จากการสอบครั้งเดียว มักเป็นการประเมิน internal consistency reliability ซึ่งบ่งบอกว่าข้อสอบทุกข้อที่ใช้ในการสอบนักเรียนกลุ่มหนึ่งๆทำการวัดความรู้ในเรื่องเดียวกันหรือไม่

3. Reliability เป็นปัจจัยที่สำคัญเพียงปัจจัยหนึ่งในการประเมินคุณค่าของผลสอบ ผลสอบที่ไม่มีความแม่นยำนั้นเป็นผลสอบที่มีคุณค่าต่ำไม่สามารถให้ข้อมูลที่เป็นประโยชน์เกี่ยวกับนักเรียนผู้สอบได้ แต่ผลสอบที่มีความแม่นยำสูงนั้นก็มักจำเป็นที่จะเป็นผลสอบที่เราสามารถนำไปใช้ประโยชน์ได้เสมอไป จำเป็นต้องพิจารณาปัจจัยร่วมอื่นๆ อีกหลายอย่าง เช่น หากมีนักเรียนทุจริตในการสอบ คะแนนสอบที่ได้ก็อาจมีค่า reliability coefficient สูง แต่ผลสอบนั้นก็จะเป็นผลสอบที่บิดเบือน ไม่สามารถบอกได้ว่านักเรียนที่ได้คะแนนสูงเป็นนักเรียนที่มีความรู้ หรือเป็นนักเรียนที่ไม่มีความรู้แต่ลอกข้อสอบเพื่อน

ประเด็นที่ได้รับความสนใจกันมากคือ ค่า reliability coefficient ต้องสูงแค่ไหนจึงจะเพียงพอที่จะนำผลสอบไปใช้ได้ โดยทั่วไปนั้นจำเป็นต้องพิจารณาควบคู่ไปกับการนำผลสอบไปใช้ หากผลสอบนั้นนำไปใช้ในการตัดสินใจที่สำคัญ เมื่อตัดสินใจไปแล้วผลเป็นที่สุดไม่สามารถเปลี่ยนแปลงได้ และส่งผลยาวนาน โดยเฉพาะการตัดสินใจที่ส่งผลกระทบต่อตัวบุคคล มักต้องการคะแนนสอบที่มีค่า reliability coefficient สูงมาก ในทางกลับกัน หากผลสอบนั้นใช้ในการตัดสินใจที่ไม่ค่อยสำคัญ มีผลระยะสั้น และการตัดสินใจอาจเปลี่ยนแปลงได้หลังจากการสอบนี้โดยพิจารณาจากการสอบอื่นที่จะจัดตามมามากมายหลัง โดยเฉพาะการตัดสินใจที่มีผลต่อนักเรียนเป็นกลุ่ม ไม่ส่งผลกระทบต่อตัวบุคคล มักไม่ต้องการค่า reliability coefficient ที่สูงมาก โดยทั่วไปสำหรับการสอบย่อยๆ ใน

ชั้นเรียน ควรให้ค่า reliability coefficient สูงกว่า 0.7 สำหรับการสอบลงของนักศึกษาแพทย์ การสอบปลายภาค หรือการสอบใหญ่ต่างๆ ในโรงเรียนแพทย์ ควรให้ค่า reliability coefficient สูงกว่า 0.8 สำหรับการสอบที่มีความสำคัญมาก เช่น การสอบคัดเลือกเข้าเรียนมหาวิทยาลัย การสอบใบอนุญาตประกอบวิชาชีพเวชกรรม การสอบวุฒิปับตรผู้เชี่ยวชาญเฉพาะทาง มักต้องให้ reliability coefficient สูงกว่า 0.9

อีกประเด็นหนึ่งที่มีความสำคัญคือ มีปัจจัยใดบ้างที่ส่งผลต่อค่า reliability coefficient สิ่งเหล่านี้มีความสำคัญมากเมื่อเราต้องการอธิบายว่าเหตุใดคะแนนสอบที่ได้จึงไม่แม่นยำ และเราต้องทำอย่างไรจึงจะทำให้คะแนนสอบมีความแม่นยำมากขึ้น โดยทั่วไปปัจจัยที่สำคัญที่ส่งผลต่อความแม่นยำของคะแนนสอบมีด้วยกัน 4 ปัจจัย คือ

1. จำนวนข้อสอบ ถ้าทำการสอบด้วยข้อสอบที่สั้น ประกอบด้วยคำถามไม่กี่ข้อ คะแนนสอบที่ได้มักไม่แม่นยำ วิธีเพิ่มความแม่นยำของคะแนนสอบที่ง่ายที่สุดคือการเพิ่มจำนวนข้อสอบ
2. การกระจายตัวของคะแนนสอบ ถ้าคะแนนสอบมีความแตกต่างกันมาก มีทั้งนักเรียนที่ทำคะแนนได้สูง และนักเรียนที่ทำคะแนนได้ต่ำ คะแนนสอบมักมีความแม่นยำสูง ในทางตรงข้ามหากนักเรียนทำคะแนนใกล้เคียงกัน คะแนนเกาะกลุ่มกันมาก คะแนนสอบมักมีความแม่นยำต่ำ วิธีการเพิ่มความแม่นยำของคะแนนสอบโดยการเพิ่มการกระจายตัวของคะแนนของนักเรียนทำได้โดยใช้ข้อสอบที่มีความยากมากขึ้น
3. ปัจจัยรบกวนการสอบของนักเรียน หากทำการจัดสอบไม่ดี มีสิ่งมารบกวนนักเรียนในขณะที่ทำการสอบ (เช่น มีเสียงดังรบกวน ห้องสอบร้อนอบอ้าวจนนักเรียนไม่มีสมาธิ) คะแนนสอบมักมีความแม่นยำต่ำ ดังนั้นผู้คุมสอบต้องจัดสถานที่สอบให้ดี เพื่อให้ให้นักเรียนมีสมาธิในการทำข้อสอบ ซึ่งจะนำไปสู่คะแนนสอบมีความแม่นยำสูง
4. ลักษณะการให้คะแนนของข้อสอบ ข้อสอบที่ไม่ต้องใช้กรรมการตรวจ เช่น multiple-choice examination มักให้คะแนนที่มีความแม่นยำสูง ในทางตรงข้ามข้อสอบที่ต้องใช้กรรมการให้คะแนน เช่น ข้อสอบบรรยาย ข้อสอบ OSCE คะแนนที่ได้มักมีความแม่นยำไม่สูงนักเนื่องจากมีปัจจัยที่นอกเหนือไปจากความสามารถของนักเรียน (เช่น ความเหนื่อยล้าของกรรมการ ความไม่สม่ำเสมอของการใช้เกณฑ์ให้คะแนน หรือ อารมณ์ของกรรมการตรวจข้อสอบ) เข้ามาส่งผลต่อคะแนนสอบ

รศ. ดร. นพ. เชิดศักดิ์ ไอรมนิรัตน์

หัวข้อ : MCQ

MCQ Item Development

รศ. นพ. เชิดศักดิ์ ไอรมนิรัตน์
ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล
มหาวิทยาลัยมหิดล

Multiple-Choice Questions

- Selected Response Exam
 - True/False
 - Simple True/False items
 - Multiple true/false items (K-type)
 - One best response
 - Standard MCQ
 - Extended matching items

Multiple-Choice Questions

- Advantages
 - Objective scoring
 - High internal consistency reliability
 - Strong research evidence to support its validity
 - Efficiency in testing and scoring

3

Multiple-Choice Questions

- Limitations
 - Cueing of correct answer
 - Random guessing
 - Testing of trivial knowledge
 - Difficulty of development of good MCQ items
 - Focus only in cognitive abilities, not good for assessing psychomotor skills or attitudes

4

MCQ in Thai Medical Education

- Medical school admission
- Classroom tests
- Comprehensive exam
- National licensing exam steps 1, 2
- Postgraduate exam
 - Basic science exam
 - Board exam

Activity

- Open a web browser
- Go to <http://socrative.com>
- Select [Student login]
- In Room name, type in: IRAMANEERAT
- Click [Join]
- Type in your own name

A Good MCQ Item

1. Content
2. Structure

Multiple-Choice Questions

1. Acute intermittent porphyria is the result of a defect in the biosynthetic pathway for
- A. Thyroxine
 B. Glucose
 C. Fatty acid
 D. Collagen
 E. Heme
- Stem**
- Options**
 - Distractors
 - Correct option

Multiple-Choice Questions

2. A 32-year-old man has progressive weakness of extremities for 4 days. Physical exam revealed T 37.8 c, BP 130/80 mmHg, P 94/min, R 42/min. He has symmetrical weakness of both sides of face and extremities. Sensation is intact. No deep tendon reflex can be detected.
- Which of the following is the most likely diagnosis?
- A. Polymyositis
 B. Poliomyelitis
 C. Guillain-Barre syndrome
 D. Myasthenia gravis
 E. Acute encephalomyelitis
- Stem**
- Lead-in**
- Options**
 - Distractors
 - Correct option

Guidelines for MCQ items

- Content guidelines
- Format guidelines
- Stem guidelines
- Option guidelines

Content Guidelines

- Focus on a single idea for each item
- Avoid trivial content
- Avoid opinion-based items
- Avoid direct quotes from textbooks
- Keep item content independent from one another

Format Guidelines

- Simplify vocabulary and sentence structures
- Avoid presenting unrelated information, minimize reading time
- Proofread each item for correct grammar, punctuation, and spelling

Stem Guidelines

- Make the question as clear as possible
- Avoid using negative words (not, except)
- Place the main idea of an item in the stem, not in options

13

Option Guidelines

- Develop as many effective options as you can
- Vary the location of the correct answers
- Keep options independent
- Keep options homogeneous
- Keep the length of options about the same
- Avoid "none of above" or "all of above"
- Avoid giving clues

14

Activity

- Open a web browser
- Go to <http://socrative.com>
- Select [Student login]
- In Room name, type in: IRAMANEERAT
- Click [Join]
- Type in your own name

Common Pitfalls

- Grammatical cues
- Logical cues
- Absolute terms
- Long correct option
- Repetition
- Convergence
- Suggestion by other item

การสร้างข้อสอบปรนัย

ผู้ช่วยศาสตราจารย์ นายแพทย์เชิดศักดิ์ โอรมนิรัตน์

ภาควิชาพยาธิศาสตร์, คณะแพทยศาสตร์ศิริราชพยาบาล, มหาวิทยาลัยมหิดล, กรุงเทพมหานคร ๑๐๓๑๐.

ข้อสอบปรนัย (multiple-choice question) เป็นรูปแบบการประเมินผลที่นิยมใช้กันอย่างแพร่หลายในวงการแพทยศาสตรศึกษาเนื่องด้วยคุณสมบัติที่หลายประการด้วยกัน ได้แก่ ประสิทธิภาพในการประเมินความรู้ปริมาณมากในเวลาอันสั้น ผลการประเมินที่ไม่มีผลกระทบจากความรู้สึกส่วนตัวของผู้ตรวจให้คะแนน คะแนนที่มีความเที่ยงสูง รวมถึงผลการวิจัยจำนวนมากที่สนับสนุนความถูกต้องของผลการประเมินด้วยข้อสอบปรนัย^{๑-๖} ข้อสอบปรนัยที่พัฒนาขึ้นอย่างดีนั้นสามารถวัดความรู้ได้ทั้งระดับการจดจำ การทำความเข้าใจ และการประยุกต์ความรู้ไปใช้ในการดูแลคนไข้^{๗-๙} อย่างไรก็ตาม การประยุกต์ความรู้ไปใช้ในการดูแลคนไข้^{๗-๙} อย่างไรก็ดี ผลการศึกษาวิจัยเกี่ยวกับคุณภาพของข้อสอบปรนัยที่พัฒนาขึ้นใช้ในโรงเรียนแพทย์หลายแห่งพบว่าข้อสอบจำนวนไม่น้อยมีลักษณะที่ไม่เหมาะสม^{๕-๖} ข้อสอบปรนัยที่ถูกพัฒนาขึ้นอย่างไม่ถูกหลักการนั้นส่งผลเสียหลายอย่าง เช่น ทำให้ข้อสอบยากขึ้นโดยไม่จำเป็น ทำให้ผู้สอบเกิดความสับสน ทำให้ผู้สอบบางกลุ่มเสียเปรียบผู้สอบคนอื่น ทำให้การตัดสินผลสอบผิดพลาด เป็นต้น^{๖-๗} ดังนั้นการออกข้อสอบปรนัยที่ดี วางอยู่บนหลักการที่ต้องจึงมีความสำคัญมากในการควบคุมคุณภาพการศึกษาในโรงเรียนแพทย์ บทความนี้จึงถูกเขียนขึ้นเพื่อเป็นการรวบรวมหลักการพื้นฐานในการออกข้อสอบปรนัยที่ได้รับการยอมรับกันทั่วไปในวงการวัดและประเมินผล ผู้นิพนธ์หวังว่าข้อแนะนำต่าง ๆ ที่ได้นำเสนอในบทความนี้จะเป็นแนวทางที่เป็นประโยชน์ในการพัฒนาข้อสอบปรนัยที่มีคุณภาพให้ผู้อ่านไม่มากนักน้อย

รูปแบบพื้นฐานของข้อสอบปรนัย

ข้อสอบปรนัยคือข้อสอบชนิดที่มีคำถามแล้วมีตัวเลือกให้ผู้สอบเลือกตัวเลือกที่เหมาะสมเพื่อตอบคำถามดังกล่าว ข้อสอบปรนัยสามารถแบ่งออกได้เป็น ๒ รูปแบบ^๖ ได้แก่

๑. ข้อสอบถูกผิด (True/false item)

ในข้อสอบประเภทนี้จะมีข้อความให้ผู้สอบพิจารณาว่าถูกหรือผิด ในยุคแรกข้อสอบเหล่านี้แต่ละข้อจะแยกเป็นอิสระจากกัน ผู้สอบตัดสินใจว่าข้อความแต่ละข้อถูกหรือผิดโดยไม่เกี่ยวข้องกับข้อความในข้ออื่น ต่อมาเมื่อผู้พัฒนาข้อสอบเป็นชุดของข้อความ (multiple true/false หรือ K-type item) โดยในแต่ละข้อจะมีสี่ข้อความ ผู้สอบต้องพิจารณาว่าแต่ละข้อความถูกหรือผิด แล้วทำการเลือกตัวเลือกที่บรรยายจำนวนข้อความที่ถูกต้องได้อย่างเหมาะสม (เช่น ตอบ ก. เมื่อข้อความที่ ๑, ๒, และ ๓ ถูกต้อง, ตอบ ข. เมื่อข้อความที่ ๑ และ ๓ ถูกต้อง ฯลฯ)

ข้อสอบชนิดถูกผิดนี้เคยเป็นที่นิยมมากในวงการแพทยศาสตรศึกษาอยู่ระยะหนึ่งเนื่องจากสามารถทดสอบความรู้ได้ปริมาณมาก แต่ข้อสอบชนิดนี้มีข้อจำกัดที่สำคัญคือสามารถใช้ได้เฉพาะกับเนื้อหาที่มีความถูกต้องชัดเจนเท่านั้น ซึ่งการตัดสินใจทางการแพทย์ส่วนมากไม่เป็นเช่นนั้น การตัดสินใจในการวินิจฉัย การตรวจค้นเพิ่มเติม หรือการรักษาผู้ป่วยส่วนใหญ่นั้นแพทย์ตัดสินใจเลือกกระหว่างทางเลือกที่แตกต่างกันสามสี่อย่างซึ่งทุกทางเลือกมีความเป็นไปได้ มีส่วนถูก หรือมีความเหมาะสมในบางด้าน

แต่ก็มีความไม่เหมาะสมในด้านอื่นด้วย เช่นการเลือกใช้ยาในผู้ป่วยที่มีการติดเชื้อ นักศึกษาแพทย์มักรู้ว่าควรใช้ยาปฏิชีวนะ ซึ่งยาปฏิชีวนะหลายชนิดก็รักษาการติดเชื้อชนิดนั้น ๆ ได้ แต่นักศึกษาต้องเลือกระหว่างยาที่ลืมนำมาใช้ได้ในการรักษานั้นว่ายาใดที่มีประสิทธิภาพสูงสุด เหมาะสมที่สุดกับชนิดของเชื้อก่อโรคที่พบบ่อยในการติดเชื้อนั้น มีผลข้างเคียงน้อยที่สุด และราคาเหมาะสมด้วย ซึ่งในสถานการณ์นี้ข้อสอบชนิดถูกผิดจะนำมาใช้ได้ยาก ด้วยเหตุนี้ทำให้ข้อสอบชนิดถูกผิดไม่เป็นที่นิยมกันมากนักในปัจจุบัน

๒. ข้อสอบเลือกคำตอบที่ถูกต้องที่สุด (one best response item)

ในข้อสอบประเภทนี้จะมีคำถามแล้วตามด้วยตัวเลือกจำนวนหนึ่งให้ผู้สอบเลือกตัวเลือกที่เหมาะสมที่สุดเป็นคำตอบ ข้อสอบประเภทนี้ที่เป็นที่นิยมกันมากที่สุดคือข้อสอบที่มีตัวเลือก ๔-๕ ตัวเลือก (A-type) แต่นอกจากข้อสอบมาตรฐานนี้แล้วก็มีผู้ใช้ข้อสอบประเภทที่มีลักษณะเป็นการจับคู่ (extended matching item) โดยให้ผู้สอบเลือกตัวเลือกที่เหมาะสม (จากตัวเลือกจำนวนมาก ๔-๒๐ ตัวเลือก) ไปจับคู่กับโจทย์ (stem) ซึ่งมีหลายข้อ เช่นจับคู่ระหว่างคำบรรยายอาการของผู้ป่วยจำนวน ๕-๑๐ ราย กับการวินิจฉัยโรคที่เหมาะสม จำนวน ๑๕ โรค เป็นต้น

เนื่องจากข้อสอบชนิดที่มีใช้กันแพร่หลายในวงการแพทยศาสตรศึกษาในประเทศไทยในปัจจุบันคือข้อสอบประเภทที่มีตัวเลือก ๔-๕ ตัวเลือก (A-type) ผู้นิพนธ์จะขอเน้นหลักการสำหรับการออกข้อสอบประเภทนี้ที่สำคัญ

องค์ประกอบของข้อสอบปรนัยชนิดเลือกคำตอบที่ถูกต้องที่สุด

ข้อสอบปรนัยแต่ละข้อมีส่วนประกอบสำคัญ ๒ ส่วนด้วยกันคือ

๑. โจทย์ (stem) เป็นข้อมูลของโรค หรือภาวะหรือผู้ป่วยตามด้วยคำถาม หรือเว้นช่องว่างสำหรับเติมคำหรือข้อความที่เหมาะสมลงไป

๒. ตัวเลือก (options) คือคำ หรือข้อความที่

ผู้ออกข้อสอบนำเสนอตามหลังจากโจทย์เพื่อให้ผู้สอบเลือกไปใช้ตอบคำถาม หรือเติมลงในช่องว่างในโจทย์

๒.๑ ตัวเลือกที่ถูกต้อง (correct option) เป็นคำตอบที่ถูกต้องมีเพียงตัวเลือกเดียวต่อข้อสอบข้อหนึ่ง

๒.๒ ตัวลวง (distractors) เป็นคำตอบที่ผิด หรือไม่เหมาะสม มีไว้ลวงให้ผู้สอบที่ไม่มีความรู้ หรือมีความเข้าใจไม่ถูกต้องในเนื้อหาที่นำมาออกข้อสอบเลือกตอบ ตัวลวงไม่จำเป็นต้องเป็นคำตอบที่ผิดชัดเจนเสมอไป ตัวลวงที่ดีมักมีส่วนถูกบ้าง แต่มีระดับของความถูกต้องเหมาะสมน้อยกว่าคำตอบที่ถูกต้อง

ข้อแนะนำพื้นฐานของการเขียนข้อสอบปรนัย

มีผู้เชี่ยวชาญทางการประเมินผลให้ข้อแนะนำจำนวนมากในการเขียนข้อสอบปรนัย เคยมีผู้รวบรวมไว้ถึง ๔๓ ข้อ^{๒,๔} ในที่นี้ผู้นิพนธ์ขอนำเสนอเฉพาะข้อแนะนำที่ได้รับการยอมรับอย่างกว้างขวางและสามารถประยุกต์ใช้ได้ชัดเจนในการพัฒนาข้อสอบทางการแพทย์ โดยจะทำการจัดหมวดหมู่ของข้อแนะนำเหล่านี้ออกเป็น ๔ กลุ่มด้วยกัน ได้แก่ (๑) เนื้อหาข้อสอบ, (๒) การจัดรูปแบบข้อสอบ, (๓) การเขียนโจทย์, และ (๔) การเขียนตัวเลือก

๑. เนื้อหาข้อสอบ

๑.๑ ข้อสอบหนึ่งข้อควรมุ่งเน้นประเมินความรู้เพียงเรื่องเดียว

ก่อนเริ่มเขียนข้อสอบอาจารย์ผู้ออกข้อสอบควรตั้งวัตถุประสงค์ให้ชัดเจนว่าต้องการประเมินความรู้ของผู้สอบในเรื่องใด และเขียนโจทย์เพื่อตอบสนองวัตถุประสงค์ดังกล่าวเท่านั้น เนื่องจากเนื้อหาวิชาทางการแพทย์มีมาก อาจารย์แต่ละท่านเมื่อทำการสอนไปแล้วจึงอยากทดสอบความรู้ในหลายเรื่องที่ตนได้สอนไป แต่กลับมีโควตาจำกัดในการออกข้อสอบ ทำให้อาจารย์จำนวนไม่น้อยเขียนข้อสอบหนึ่งข้อถามทั้งเรื่องการวินิจฉัยโรค การตรวจค้นเพิ่มเติม การรักษาโรค และภาวะแทรกซ้อนของโรคไปพร้อมกัน ลักษณะข้อสอบเช่นนี้ไม่ควรใช้ เพราะมักซับซ้อนเกินไป เมื่อผู้สอบตอบข้อสอบผิด ก็ไม่สามารถวินิจฉัยได้ว่าผู้สอบขาดความรู้ ความเข้าใจในเรื่องใด

๑.๒ หลีกเลี่ยงการถามความรู้ในรายละเอียดปลีกย่อยที่ไม่มีที่ชี้ทางคลินิก (trivial content)

ฉบับที่ ๓๓๓๓

บทความทั่วไป

องค์ความรู้ทางการแพทย์นั้นมีปริมาณมาก ไม่มีผู้ใดที่จดจำเนื้อหาที่มีในตำรา หรือวารสารทางการแพทย์ได้ทั้งหมด แม้ว่าองค์ความรู้หลายเรื่องมีความน่าสนใจ แต่มีประโยชน์ในการประยุกต์ใช้ทางคลินิกค่อนข้างน้อย องค์ความรู้ดังกล่าวจัดเป็นรายละเอียดปลีกย่อย (trivial content) ซึ่งไม่แนะนำให้ทำการทดสอบ สิ่งที่ควรทำการประเมินคือความสามารถในการประยุกต์ใช้ความรู้ในทางคลินิก (application of knowledge) ไม่แนะนำการทดสอบวัดความสามารถในการจดจำเป็นหลัก อย่างไรก็ตามการที่แนะนำให้ออกข้อสอบที่เน้นการประยุกต์ใช้ความรู้ ไม่ได้หมายความว่า การแก้ปัญหาผู้ป่วยนั้นไม่ต้องใช้ความจำเลย ตรงกันข้ามการจดจำเนื้อหาเป็นพื้นฐานที่สำคัญในการแก้ปัญหาทางคลินิก ผู้สอบย่อมต้องจำเนื้อหาได้บ้าง จึงจะประยุกต์องค์ความรู้ดังกล่าวไปแก้โจทย์ปัญหาที่นำเสนอได้

๑.๓ หลีกเลียงการถามความรู้ในเรื่องที่ยังมีความขัดแย้งกันในแนวทางปฏิบัติ (controversy)

ความรู้ทางการแพทย์ในหลายหัวข้อยังเป็นเรื่องที่ยังมีข้อถกเถียงกันอยู่ ผู้ป่วยรายเดียวกันไปพบแพทย์สองคนอาจได้รับการรักษาที่แตกต่างกันซึ่งวิธีการรักษาทั้งสองวิธีก็มีความจำเป็นสนับสนุนด้วยกันทั้งคู่ อย่างไรก็ตามยังคงมีความขัดแย้ง (controversy) ในเรื่องดังกล่าวอยู่ เนื้อหาในลักษณะนี้ไม่ควรนำมาออกสอบด้วยข้อสอบปรนัย เนื่องจากในขณะที่ทำข้อสอบอยู่นั้น ผู้สอบไม่มีทางรู้ได้เลยว่าอาจารย์ผู้ออกข้อสอบอ้างอิงจากตำราหรือบทความวิชาการใด เนื้อหาที่ยังมีความขัดแย้ง ที่ผู้เชี่ยวชาญจากต่างสถาบันมีแนวทางในการปฏิบัติที่ต่างกันอย่างนี้แนะนำให้ใช้ข้อสอบในรูปแบบอื่นในการทดสอบเช่นข้อสอบอัตนัย เป็นต้น

๑.๔ หลีกเลียงการลอกประโยคหรือข้อความจากตำราโดยตรง

ดังได้กล่าวแล้วว่าข้อสอบที่ดีควรมุ่งเน้นการประเมินความเข้าใจ หรือ การประยุกต์ใช้ความรู้ ไม่ควรออกข้อสอบที่ประเมินความสามารถในการจำรายละเอียดปลีกย่อย การออกข้อสอบโดยวิธีการเปิดตำราแล้วคัดลอกประโยคจากตำราโดยตรงมักจะลงเอยด้วยข้อสอบที่ทดสอบความจำว่าผู้สอบท่องเนื้อหาในตำราตรงส่วนนั้นได้หรือไม่

ข้อสอบที่ดีควรได้จากการดูผู้ป่วย โจทย์ที่ดีควรเป็นปัญหาของผู้ป่วยที่พบในการทำงานนั่นเอง ตัวเลือกก็ได้จากข้อผิดพลาดที่นักศึกษาหรือแพทย์ประจำบ้านมักปฏิบัติกับผู้ป่วยแล้วทำให้ผลการรักษาไม่ดีขึ้นเอง

๑.๕ หลีกเลียงการนำเสนอข้อสอบที่ประเมินความรู้ในเรื่องเดียวกันสองข้อในข้อสอบชุดเดียวกัน

เนื่องจากเนื้อหาวิชาที่ต้องทำการประเมินในการสอบแต่ละครั้งนั้นมีมาก ดังนั้นองค์ความรู้ในแต่ละเรื่องแต่ละโรคจึงมักมีสัดส่วนของข้อสอบที่จะออกได้เพียงหนึ่งหรือสองข้อเท่านั้น การที่อาจารย์ออกข้อสอบในเรื่องหรือโรคเดียวกันซ้ำสองข้อในชุดข้อสอบเดียวกันจึงมักเป็นการลดโอกาสในการประเมินความรู้เรื่องอื่นซึ่งก็มีความสำคัญเช่นกัน การออกข้อสอบที่ดีนั้นควรต้องครอบคลุมวัตถุประสงค์การเรียนรู้ตามที่กำหนดในหลักสูตร หรือในเกณฑ์มาตรฐานผู้ประกอบวิชาชีพเวชกรรมอย่างสมดุล การที่จะบรรลุเป้าหมายดังกล่าวได้นั้นต้องเริ่มต้นจากการกำหนดสัดส่วนข้อสอบสร้างเป็นตารางกำหนดจำนวนข้อสอบ (table of specification) เมื่ออาจารย์ได้รับมอบหมายให้ออกข้อสอบควรต้องตรวจสอบให้ชัดเจนว่าเนื้อหาที่ต้องออกข้อสอบนั้นอยู่ในส่วนใดของตารางดังกล่าว การออกข้อสอบซ้ำซ้อนในเนื้อหาเรื่องเดียวกันเป็นสัญญาณบ่งบอกว่าอาจไม่ได้สร้างข้อสอบตามข้อกำหนดในตาราง นอกจากนี้การมีโจทย์สองข้อประเมินความรู้เรื่องเดียวกันมีความเป็นไปได้สูงที่เนื้อหาในข้อสอบข้อหนึ่งอาจบอกคำตอบในข้อสอบอีกข้อหนึ่งได้

๒. การจัดรูปแบบข้อสอบ

๒.๑ เลือกใช้คำศัพท์หรือรูปประโยคที่ง่ายต่อการทำความเข้าใจ

อาจารย์ผู้ออกข้อสอบต้องระลึกไว้เสมอว่าข้อสอบที่อาจารย์ออกเพื่อใช้ในการประเมินผลนักศึกษาแพทย์หรือแพทย์ประจำบ้านนั้นมีวัตถุประสงค์เพื่อทดสอบความรู้ทางการแพทย์เป็นสำคัญ มิใช่การประเมินความรู้ทางภาษาศาสตร์ ดังนั้นการเขียนข้อสอบของอาจารย์ควรเลือกใช้รูปแบบประโยคที่ง่ายต่อการทำความเข้าใจ อย่าเขียนประโยคซับซ้อนที่มีความยาวประโยคหลายบรรทัด มุ่งเน้นให้ภาษาเป็นสื่อในการนำเสนอความคิดของอาจารย์ผู้ออกข้อสอบไปยังผู้สอบ อย่าให้

เวบบทกวีศรีราช

บทความทั่วไป

ภาษาเป็นอุปสรรคในการสื่อสาร การจะเลือกใช้ภาษาใดในการเขียนข้อสอบนั้นให้พิจารณาตามข้อกำหนดขององค์กรหรือหน่วยงานที่ควบคุมการสอบที่อาจารย์ส่งข้อสอบไปให้ใช้ ข้อสอบที่ใช้ในระดับการศึกษาหลักสูตรแพทยศาสตรบัณฑิตทั้งในระดับคณะ หรือข้อสอบที่ใช้ในการสอบระดับประเทศในปัจจุบันยังนิยมใช้ข้อสอบที่เขียนด้วยภาษาไทยโดยมีการใช้ศัพท์เทคนิคเป็นภาษาอังกฤษเหมือนดังภาษาที่แพทย์ใช้สื่อสารกันในการทำงานปกติ ส่วนข้อสอบในระดับหลังปริญญาตรีมีหลายการสอบที่ภาควิชา หรือราชวิทยาลัยที่เกี่ยวข้องกำหนดให้ใช้ภาษาอังกฤษทั้งหมด ก่อนที่อาจารย์จะสร้างข้อสอบต้องมีการศึกษาข้อกำหนดของแต่ละการสอบให้ดี

๒.๒ หลีกเลี่ยงการนำเสนอข้อมูลที่ไม่เกี่ยวข้องกับการแก้ปัญหาของโจทย์ข้อนั้น

โจทย์แต่ละข้อควรเขียนให้กระชับ ไม่ยาวเยิ่นเย้อโดยไม่จำเป็น นำเสนอเฉพาะข้อมูลที่จำเป็นในการแก้ปัญหาโจทย์ดังกล่าว อาจารย์บางท่านนำเสนอข้อมูลเยอะมากในโจทย์หนึ่งข้อ บางครั้งข้อสอบข้อหนึ่งมีความยาวถึงครึ่งหน้า โดยให้เหตุผลว่าเป็นเหมือนสถานการณ์จริงที่แพทย์ต้องตัดสินใจบนข้อมูลทางคลินิกปริมาณมาก แพทย์ต้องพิจารณาเองว่าข้อมูลใดสำคัญกับการแก้ปัญหาโจทย์ข้อนั้น ๆ แต่อาจารย์ก็ต้องไม่ลืมว่าเวลาที่ผู้สอบมีในการทำข้อสอบแต่ละข้อนั้นมีจำกัด ในการสอบทางการแพทย์ในประเทศไทยส่วนใหญ่ผู้สอบจะมีเวลาราว ๑ นาทีในการทำข้อสอบ ๑ ข้อ หากเนื้อหาโจทย์ข้อใดมีความยาวมาก ผู้สอบจำนวนไม่น้อยจะเลือกที่จะข้ามข้อสอบข้อนั้นไปก่อนด้วยเกรงว่าจะเสียเวลาอ่านและคิดแก้ปัญหาในข้อนั้นนานเกินไปทำให้ทำข้อสอบไม่ทัน ดังนั้นหากอาจารย์ต้องการให้ข้อสอบที่อาจารย์เขียนขึ้นมานั้นได้ถูกใช้จริง และผู้เข้าสอบได้คิดแก้ปัญหาจริงในการสอบ ไม่ถูกอ่านข้ามไป อาจารย์ควรเขียนข้อสอบให้กระชับ ไม่นำเสนอข้อมูลที่ไม่เกี่ยวข้องกับการแก้ปัญหา

๒.๓ จัดให้มีการตรวจสอบเนื้อหา คำศัพท์ และรูปประโยคที่ใช้ในข้อสอบแต่ละข้อก่อนนำไปใช้

ถึงแม้ว่าอาจารย์ผู้เขียนข้อสอบจะได้มีการอ่านทวนสิ่งที่ตนเองเขียนแล้วเข้าใจเนื้อหาได้ดีและคิดว่าข้อสอบอยู่ในรูปแบบที่สามารถนำไปใช้ได้แล้ว ก็ไม่ควร

นำข้อสอบข้อนั้นไปใช้สอบเลย ควรให้มีคณะกรรมการข้อสอบซึ่งประกอบไปด้วยอาจารย์หลายท่านช่วยกันตรวจสอบและพิจารณาปรับแก้ข้อสอบทุกข้อก่อนนำไปใช้จริงเสมอ เนื่องจากผู้เขียนข้อสอบย่อมเข้าใจสิ่งที่ตนเขียนเสมอ แต่เมื่อผู้อื่นอ่านแล้วอาจพบว่ามีเนื้อหาที่กำกวมหรือเข้าใจโจทย์ต่างออกไปได้ การปรับแก้เนื้อหาที่มีความกำกวม หรือเฉลยซึ่งอาจารย์บางท่านอาจไม่เห็นด้วยให้ได้ข้อสอบที่มีความชัดเจน และอาจารย์ทุกท่านยอมรับในค่าเฉลยได้ก่อนจะนำข้อสอบไปทำการสอบจริงย่อมเป็นสิ่งที่ดีกว่าการตรวจพบปัญหาหลังจากสอบเสร็จแล้วซึ่งต้องมาตัดสินใจกันอีกว่าจะทำอย่างไรกับกรณีคิดคะแนนของข้อสอบข้อดังกล่าว

๓. การเขียนโจทย์

๓.๑ เขียนโจทย์ให้มีความชัดเจน ผู้สอบทุกคนอ่านแล้วมีความเข้าใจตรงกัน

ข้อแนะนำนี้อาจดูเหมือนตรงไปตรงมา แต่กลับเป็นปัญหาที่พบบ่อยมากในการพัฒนาข้อสอบปรนัยประเด็นสำคัญคือโจทย์ที่ตีนั้นต้องมีความสมบูรณ์ในตัวเองโดยไม่ต้องอาศัยตัวเลือก โจทย์ข้อสอบที่ตีนั้นเมื่ออ่านโจทย์เสร็จแล้ว หากผู้สอบมีความรู้ในเรื่องที่ทำการประเมินนั้น เขาจะบอกคำตอบได้โดยไม่ต้องอ่านตัวเลือกเลย ดังนั้นเมื่ออาจารย์เขียนข้อสอบเสร็จแล้วแนะนำให้ลองปิดตัวเลือกแล้วอ่านเฉพาะโจทย์ดู หากอาจารย์อ่านแล้วบอกได้ว่าโจทย์ถามอะไรและบอกได้ว่าควรตอบอะไรโดยไม่ต้องอ่านตัวเลือกจัดว่าข้อสอบข้อดังกล่าวมีโจทย์ที่มีความชัดเจน

๓.๒ เรียบเรียงเนื้อหาให้ใจความสำคัญของข้อสอบอยู่ในโจทย์

เนื่องจากข้อสอบปรนัยมีตัวเลือกที่อาจารย์ต้องสร้างขึ้นหลายตัวเลือก บางครั้งอาจารย์ผู้พัฒนาข้อสอบอาจเผลอเรอเอาใจความสำคัญไปใส่ไว้ในตัวเลือกซึ่งทำให้เนื้อหาในโจทย์ขาดสาระสำคัญ อ่านโจทย์แล้วไม่เข้าใจว่าผู้ออกข้อสอบต้องการถามความรู้เรื่องอะไร ตัวอย่างข้อสอบที่ไม่เป็นไปตามข้อแนะนำนี้คือข้อสอบที่ถามว่า ข้อใดต่อไปนี้เป็นไปต้อ หรือข้อใดต่อไปนี้เป็นไปต้อแล้วเขียนรายละเอียดเกี่ยวกับโรค หรือการรักษาบางอย่างในตัวเลือกแต่ละข้อ ข้อสอบในลักษณะนี้มักทำให้

เวบบิ้นทีกีธีรธา

บทความทั่วไป

ผู้สอบต้องอ่านข้อสอบย้อนไปมาหลายรอบกว่าจะเข้าใจจุดประสงค์ของข้อสอบ แล้วจึงตัดสินใจเลือกคำตอบโดยทั่วไปและแนะนำให้อาจารย์นำเสนอรายละเอียดต่าง ๆ ไว้ในตัวโจทย์ให้มากที่สุด ส่วนตัวเลือกเขียนเป็นคำหรือข้อความสั้น ๆ

๓.๓ หลักเบื้องการเขียนโจทย์ที่มีรูปประโยคเป็นเชิงปฏิเสธ

โจทย์ที่ดีไม่ควรอยู่ในประโยคเชิงปฏิเสธ เช่นถามถึงสิ่งที่เป็นข้อยกเว้น สิ่งที่ไม่ควรปฏิบัติ สิ่งที่มีน้อยที่สุด หรือสิ่งที่ไม่น่านึกถึง เป็นต้น งานวิจัยส่วนใหญ่พบว่าข้อสอบที่มีโจทย์ในรูปแบบปฏิเสธเหล่านี้มีระดับความยากง่ายไม่ต่างจากข้อสอบอื่น ๆ แต่งานวิจัยบางชิ้นพบว่าข้อสอบที่มีโจทย์ในรูปแบบปฏิเสธมีความยากมากกว่าข้อสอบอื่นชัดเจนโดยเฉพาะในข้อสอบวัดความรู้ระดับสูง^{๑๑-๑๒} แต่ผู้เชี่ยวชาญในการประเมินผลส่วนใหญ่มีความเห็นพ้องกันว่าข้อสอบประเภทนี้สามารถสร้างความสับสนให้กับผู้สอบได้ จึงไม่แนะนำให้ใช้ แต่หากอาจารย์ผู้ออกข้อสอบมีความจำเป็นต้องใช้ข้อสอบที่มีการใช้คำปฏิเสธในโจทย์ แนะนำให้พิมพ์คำปฏิเสธให้เด่นชัด โดยใช้ตัวหนาและขีดเส้นใต้เพื่อให้ผู้สอบเห็นชัด^{๑๑}

๔. การเขียนตัวเลือก

๔.๑ เขียนตัวเลือกที่มีประสิทธิภาพให้มีจำนวนมากที่สุดเท่าที่เหมาะสมกับบริบท

เรื่องจำนวนตัวเลือกที่เหมาะสมนี้เป็นเรื่องผู้เชี่ยวชาญด้านการประเมินผลจำนวนมากสนใจ มีงานวิจัยเกี่ยวกับเรื่องจำนวนตัวเลือกที่เหมาะสมในข้อสอบปรนัยอยู่มากมาย^{๑๓} อาจารย์ผู้ออกข้อสอบส่วนมากจะคุ้นเคยกับข้อสอบปรนัยชนิดที่มีห้าตัวเลือก บ่อยครั้งที่อาจารย์ออกข้อสอบแล้วนึกตัวเลือกได้เพียงสามหรือสี่ตัว จึงเกิดคำถามว่าจำเป็นต้องมีตัวเลือกครบห้าตัวเลือกหรือไม่ งานวิจัยบางชิ้นพบว่าการลดจำนวนตัวเลือกลงทำให้ข้อสอบง่ายขึ้น^{๑๓-๑๔} แต่งานวิจัยบางชิ้นพบว่าการลดจำนวนตัวเลือกลงทำให้ได้ข้อสอบยากขึ้น^{๑๕-๑๖} ผู้เชี่ยวชาญในการประเมินผลเสนอว่าข้อสอบปรนัยที่มีตัวเลือกเพียงสามตัวเลือกก็สามารถทดสอบความรู้ได้อย่างมีประสิทธิภาพ^{๑๖-๑๗, ๑๗} แต่มีอาจารย์จำนวนไม่น้อยที่ไม่สบายใจที่มีตัวเลือกในข้อสอบแต่ละข้อน้อยกว่าห้าตัว

เลือกด้วยกังวลว่าจะทำให้มีโอกาสสูงที่ผู้สอบที่ไม่มีความรู้จะเดาสุ่มได้คำตอบที่ถูกต้อง แต่จากข้อมูลที่ปรากฏในปัจจุบันพบว่าผู้สอบในการสอบในระดับสูงนั้นพฤติกรรมการเดาสุ่มโดยที่ผู้สอบปราศจากความรู้นั้นน่าจะมีบทบาทน้อยมาก ผู้สอบส่วนใหญ่มักพอมีความรู้บ้างและสามารถตัดตัวเลือกที่ไม่สมเหตุสมผลอย่างชัดเจนได้^{๑๖} ในการศึกษาข้อสอบปรนัยส่วนใหญ่พบตัวเลือกที่ไม่ทำงานเป็นจำนวนไม่น้อย^{๑๖} ข้อมูลที่ได้จากการวิเคราะห์ข้อสอบปรนัยที่ใช้ในทางแพทยศาสตรศึกษาในประเทศไทยหลายครั้งก็สอดคล้องกับงานวิจัยในต่างประเทศที่พบว่าข้อสอบส่วนใหญ่มักมีตัวเลือกที่ทำงานจริงราวสามหรือสี่ตัวเลือก มีข้อสอบน้อยข้อมากที่ตัวเลือกทั้งห้าตัวเลือกทำงานอย่างมีประสิทธิภาพ

ด้วยข้อมูลจากการศึกษาต่าง ๆ ข้อแนะนำในการออกข้อสอบปรนัยในปัจจุบันคือให้อาจารย์เขียนจำนวนตัวเลือกมากที่สุดที่มีความเหมาะสมกับเนื้อหาโจทย์ ไม่จำเป็นต้องเขียนตัวเลือก ๕ ตัวเลือกเสมอไป เนื่องจากตัวเลือกที่ห้าที่เขียนขึ้นเพื่อเติมเต็มโดยไม่สมเหตุสมผลนั้นมักไม่ค่อยมีคนเลือก หากเนื้อหาที่อาจารย์นำมาสอบมีตัวเลือกที่เหมาะสมเพียงสามหรือสี่ตัวเลือกก็เขียนจำนวนตัวเลือกเพียงสามหรือสี่ตัวเลือก^{๑๖} แต่อย่างไรก็ตามให้อาจารย์ศึกษาข้อกำหนดของแต่ละการสอบที่อาจารย์เกี่ยวข้องด้วย เนื่องจากนโยบายของแต่ละการสอบแตกต่างกันไป องค์กรที่จัดสอบทางแพทยศาสตรศึกษาจำนวนไม่น้อยยังคงตั้งข้อกำหนดให้ใช้ข้อสอบ ๕ ตัวเลือกเสมอ ซึ่งหากอาจารย์ไม่ทำตามข้อกำหนดดังกล่าวข้อสอบที่ออกไปอาจไม่ได้รับการพิจารณาได้

๔.๒ จัดให้ตัวเลือกที่ถูกต้องมีการกระจายตำแหน่งไปให้มีจำนวนพอ ๆ กันในทุกตัวเลือก

ข้อแนะนำนี้มีวัตถุประสงค์เพื่อป้องกันไม่ให้ผู้สอบที่ตอบแบบเดาสุ่มแบบเลือกตัวเลือกเดียวกันทั้งหมดสอบผ่านได้ด้วยความบังเอิญ หากอาจารย์สร้างข้อสอบที่มีสี่ตัวเลือก เป็น ก ข ค ง อาจารย์ก็ต้องกระจายให้ตัวเลือกที่ถูกมีทั้งข้อ ก ข ค และ ง ในสัดส่วนที่ใกล้เคียงกัน

๔.๓ เขียนตัวเลือกแต่ละข้อให้เป็นอิสระ ไม่ขึ้นต่อกัน

เวบบ์ทีกีรราช

บทความทั่วไป

ในการเขียนตัวเลือกของข้อสอบแต่ละข้อ อาจารย์ต้องระมัดระวังให้ตัวเลือกแต่ละตัวเลือกไม่มีความซ้ำซ้อนกัน เช่นตัวเลือก ก เป็นยากลุ่มย่อยของตัวเลือก ข ตัวเลือก ก เป็นช่วงอายุ ๒ - ๑๐ ปี ตัวเลือก ข เป็นช่วงอายุ ๕ - ๑๑ ปี เป็นต้น การเขียนตัวเลือกที่ซ้ำซ้อนกันนี้ หากเกี่ยวเนื่องกับตัวเลือกที่ถูกต้องอาจมีผู้สอบแย้งว่ามีตัวเลือกที่ถูกต้องมากกว่าหนึ่งตัวเลือก หากตัวเลือกที่ซ้ำซ้อนกันนี้ไม่เกี่ยวกับคำตอบที่ถูก ก็จะทำให้ผู้สอบบางส่วนสามารถตัดตัวเลือกบางตัวเลือกได้โดยไม่ต้องมีความรู้ทางการแพทย์ในเรื่องดังกล่าวได้

๔.๔ เขียนตัวเลือกให้ทุกตัวเลือกมีความเป็นเนื้อเดียวกัน (homogeneous)

การเขียนตัวเลือกให้มีความเป็นเนื้อเดียวกันนั้นหมายถึง ตัวเลือกแต่ละตัวมีรูปร่างหน้าตาและรายละเอียดไปในทิศทางหรือเรื่องราวเดียวกัน หรือเป็นของกลุ่มเดียวกัน การเป็นเนื้อเดียวกันนี้ครอบคลุมตั้งแต่รูปร่างหน้าตา (ตัวเลือกทุกตัวเป็นภาษาแบบเดียวกัน หากตัวเลือกตัวหนึ่งเป็นคำ ตัวเลือกอื่น ๆ ก็ควรเป็นคำ ไม่ใช่วลี หรือประโยค, ตัวเลือกหนึ่งเป็นคำนาม ตัวเลือกอื่นก็เป็นคำนามเหมือนกัน ไม่ใช่กิริยา หรือคำคุณศัพท์) และเนื้อหา (โจทย์ถามการรักษา ตัวเลือกทุกตัวก็เป็นการรักษา ไม่ใช่บางตัวเป็นการตรวจค้นเพิ่มเติม, ตัวเลือกหนึ่งเป็นยาปฏิชีวนะ ตัวเลือกอื่น ๆ ก็น่าจะเป็นยาปฏิชีวนะเช่นกันไม่ใช่ยาเคมีบำบัด หรือยาต้านเชื้อรา) การที่มีตัวเลือกที่ไม่เข้าพวก ไม่มีความเป็นเนื้อเดียวกันกับตัวเลือกอื่นเป็นคำบอกใบ้ในการตัดตัวเลือกที่ผู้สอบนิยมใช้มาก ดังนั้นอาจารย์ผู้ออกข้อสอบควรหลีกเลี่ยง

ในบางบริบทของการดูแลรักษาผู้ป่วย สิ่งที่ต้องตัดสินใจเลือกอาจมีทั้งการเลือกที่จะให้การรักษาเลยหรือจะส่งตรวจค้นเพิ่มเติมก่อน ในกรณีนี้อาจารย์สามารถเขียนตัวเลือกที่มีการรักษาและการตรวจเพิ่มเติมปะปนกันได้ แต่การเขียนรูปประโยคคำถามต้องไม่เป็นการบอกใบ้ว่าจะไปทิศทางใด แต่ต้องเลือกใช้คำถามที่เป็นกลาง เช่น ท่านจะปฏิบัติต่อผู้ป่วยอย่างไร, ท่านจะดำเนินการอย่างไรต่อไป เป็นต้น

๔.๕ เขียนตัวเลือกแต่ละข้อให้มีความยาวพอ ๆ กัน

จากการสังเกตข้อสอบปรนัยจำนวนมากจะพบว่าตัวเลือกที่ถูกต้องมักมีความยาวมากกว่าตัวเลือกอื่น ซึ่งข้อสังเกตนี้ผู้สอบจำนวนมากไม่น้อยก็ทราบดี และผู้สอบส่วนมากเมื่อไม่ทราบคำตอบก็มักเลือกตัวเลือกที่มีความยาวมากที่สุด ดังนั้นอาจารย์ผู้ออกข้อสอบควรระมัดระวังไม่ให้ตัวเลือกตัวใดตัวหนึ่งมีความยาวแตกต่างไปจากตัวเลือกอื่นชัดเจน เพราะจะทำให้ผู้สอบเดาคำตอบที่ถูกต้องได้ง่าย

๔.๖ หลีกเลี่ยงการใช้ตัวเลือก “ถูกทุกข้อ” หรือ “ไม่มีข้อใดถูก”

ตัวเลือก “ถูกทุกข้อ” เป็นตัวเลือกที่ผู้เชี่ยวชาญในการประเมินผลส่วนใหญ่เห็นสอดคล้องกันว่าไม่ควรใช้เนื่องจากมักช่วยไปตัวเลือกที่ถูกต้องให้กับผู้สอบ ทำให้ผู้สอบส่วนหนึ่งตอบถูกโดยไม่ต้องอาศัยองค์ความรู้ที่สมบูรณ์ในเรื่องที่ทดสอบ งานวิจัยพบว่าข้อสอบที่มีตัวเลือกชนิดนี้จะมีผลให้ค่าความเที่ยงของคะแนนสอบลดลง^{๑๐} จึงแนะนำให้หลีกเลี่ยงการใช้

ตัวเลือก “ไม่มีข้อใดถูก” เป็นประเด็นที่ผู้เชี่ยวชาญในการประเมินผลยังคงถกเถียงกันอยู่บ้าง ผู้เชี่ยวชาญบางส่วนเห็นว่าไม่ควรใช้ตัวเลือกประเภทนี้ แต่ผู้เชี่ยวชาญบางส่วนให้ความเห็นว่าสามารถใช้ได้ในบางกรณี^{๑๑} เหตุผลที่ตัวเลือกชนิดนี้เป็นปัญหาคือการใช้ตัวเลือกนี้มักสร้างความลำบากใจให้กับผู้สอบในการเลือกคำตอบที่ถูกในกรณีที่ตัวเลือกแต่ละตัวเลือกไม่ถูกหรือผิดชัดเจน เพราะผู้สอบจะต้องทำการเปรียบเทียบตัวเลือกที่น่าเสนอในข้อสอบกับทางเลือกอื่น ๆ ที่เขานึกได้^{๑๒} หากโจทย์ถามว่า ยาใดที่ควรให้แก่ผู้ป่วย แล้วมีชื่อยาสี่ชนิด และมีตัวเลือก “ไม่มีข้อใดถูก” นอกจากที่ผู้สอบต้องนึกว่าในบรรดา ยาที่ปรากฏในตัวเลือกนั้นเหมาะสมหรือไม่แล้วเขายังนึกต่อไปอีกว่ามียาอื่นใดที่สามารถให้ในผู้ป่วยรายนี้ได้อีก หากเขานึกออกว่ามียาอื่นที่น่าจะเหมาะสมกับผู้ป่วยมากกว่ายาในตัวเลือก (ด้วยเหตุผลที่อาจแตกต่างไปจากที่อาจารย์ผู้ออกข้อสอบคิด) เขาก็จะเลือก “ไม่มีข้อใดถูก”

การใช้ตัวเลือก “ไม่มีข้อใดถูก” จะยังเป็นปัญหามากขึ้นในข้อสอบที่ถามถึงสิ่งที่ไม่ควรทำ เช่นยาใดไม่ควรใช้ในผู้ป่วย ซึ่งนอกจากยาที่น่าเสนอในตัวเลือกแล้วย่อมมียาชนิดอื่นอีกมากมายในบัญชียาที่ไม่เหมาะสม ซึ่งไม่มี

ทางที่ใครจะรู้ว่าการที่ผู้สอบเลือกตอบ “ไม่มีข้อใดถูก” นั้นเขาคิดถึงยาใด และยานั้นไม่เหมาะสมมากไปกว่ายาที่มีอยู่ในตัวเลือกหรือไม่ งานวิจัยทั้งหมดที่ศึกษาถึงตัวเลือกชนิดนี้ได้ข้อสรุปที่ตรงกันว่าข้อสอบที่ใช้ตัวเลือกประเภทนี้เพิ่มระดับความยากให้ข้อสอบ^{๑๖} โดยทั่วไปแล้วจึงไม่แนะนำให้ใช้ตัวเลือกประเภทนี้ในการสอบทางแพทยศาสตรศึกษาซึ่งทางเลือกสำหรับสถานการณ์ที่นำเสนอมีได้มากและการตัดสินใจเลือกคำตอบต้องอาศัยการเปรียบเทียบข้อดีข้อเสียของแต่ละตัวเลือก

สรุป

ในบทความนี้ผู้นิพนธ์ได้กล่าวถึงข้อแนะนำขั้นพื้นฐานในการพัฒนาข้อสอบปรนัยชนิดเลือกคำตอบที่ถูกที่สุดโดยสรุปข้อแนะนำเหล่านี้ออกเป็นสี่กลุ่มด้วยกัน ได้แก่ (๑) เนื้อหาข้อสอบ, (๒) การจัดรูปแบบข้อสอบ, (๓) การเขียนโจทย์, และ (๔) การเขียนตัวเลือก ผู้นิพนธ์หวังว่าข้อแนะนำเหล่านี้คงพอเป็นแนวทางสำหรับอาจารย์แพทย์ในการพัฒนาข้อสอบปรนัยที่มีคุณภาพเพื่อใช้ในการประเมินนักศึกษาแพทย์และแพทย์ประจำบ้านได้บ้าง อย่างไรก็ตามบทความนี้เป็นกรกล่าวถึงข้อแนะนำเบื้องต้นเท่านั้น ยังมีข้อแนะนำอื่นๆ ที่ผู้นิพนธ์ไม่ได้นำมารวบรวมไว้ในบทความนี้เพื่อต้องการทำให้เนื้อหากระชับโดยข้อแนะนำอื่นๆ ที่ผู้นิพนธ์ไม่ได้กล่าวถึงนี้พบว่าเป็นปัญหาน้อยในการออกข้อสอบทางการแพทย์ หรือเป็นข้อแนะนำที่ไม่ได้รับการสนับสนุนอย่างกว้างขวางจากผู้เชี่ยวชาญทางการวัดและประเมินผล หากผู้อ่านสนใจรายละเอียดของข้อแนะนำอื่นๆ ที่มีผู้กล่าวไว้สามารถศึกษาเพิ่มเติมได้จากเอกสารอ้างอิงที่แสดงไว้ท้ายบทความ

มีข้อควรพิจารณาในการประยุกต์ใช้ข้อแนะนำเหล่านี้ในการพัฒนาข้อสอบที่ผู้นิพนธ์ขอกล่าวถึงประการหนึ่งคือ แม้ว่าข้อแนะนำที่กล่าวถึงเหล่านี้หลายข้อมีการศึกษาวิจัยสนับสนุนที่ชัดเจน แต่สิ่งเหล่านี้ก็เป็นเพียงข้อแนะนำว่าผู้ออกข้อสอบควรปฏิบัติ ไม่ใช่กฎเกณฑ์ตายตัว การเขียนข้อสอบปรนัยนั้นเป็นงานที่ต้องอาศัยทั้งศาสตร์และศิลป์ผสมผสานกันอย่างเหมาะสม

หาใช้สูตรคณิตศาสตร์ที่ไม่มีข้อยกเว้น ผู้นิพนธ์ไม่คาดหวังให้อาจารย์ผู้พัฒนาข้อสอบยึดข้อแนะนำเหล่านี้เสมือนกฎเกณฑ์ตายตัวที่ต้องทำตามในทุกกรณี หากแต่ต้องการให้อาจารย์ใช้เป็นแนวทางในการสร้างข้อสอบ ในบางบริบทผู้ออกข้อสอบอาจเลือกที่จะไม่ปฏิบัติตามข้อแนะนำบางประการได้บ้าง แต่การที่จะไม่ปฏิบัติตามข้อแนะนำเหล่านี้จำเป็นต้องมีเหตุผลที่เหมาะสม และควรทำไม่บ่อยนัก ยกตัวอย่างเช่นข้อแนะนำว่า โจทย์ไม่ควรเขียนถามข้อยกเว้น จะพบได้ว่ามีบางบริบทที่การรู้ข้อยกเว้น หรือข้อห้ามปฏิบัติก็เป็นองค์ความรู้ที่สำคัญในการดูแลรักษาผู้ป่วย ดังนั้นในบริบทที่เหมาะสมผู้นิพนธ์เองก็เห็นด้วยว่าอาจเขียนโจทย์ที่ถามข้อยกเว้นได้ แต่อย่างไรก็ตามการจะไม่ปฏิบัติตามข้อแนะนำนี้ต้องไม่ทำบ่อยจนเกินจำเป็น หากออกข้อสอบ ๑๐๐ ข้อ จะมีข้อสอบที่ถามข้อยกเว้น ประมาณ ๒-๓ ข้อ ย่อมเป็นสิ่งที่ยอมรับได้ แต่หากในชุดข้อสอบมีข้อสอบถึงร้อยละ ๒๐-๓๐ ที่โจทย์เขียนในรูปประโยคปฏิเสธ ถามสิ่งที่ไม่ควรปฏิบัติ หรือสิ่งที่ไม่ถูกต้อง อย่างนี้ย่อมจัดว่าละเลยแนวทางในการพัฒนาข้อสอบอย่างไม่เหมาะสม ซึ่งย่อมส่งผลให้คุณภาพของข้อสอบด้อยลงอย่างชัดเจน

เอกสารอ้างอิง

1. Downing SM. Assessment of knowledge with written test forms. In: Norman GR, van der Vleuten C, Newble DI, editors. International handbook of research in medical education. Dordrecht: Kluwer Academic Publishers, 2002:647 - 72.
2. Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. Appl Meas Educ 1989;2:37-50.
3. Haladyna TM. Developing and validating multiple-choice test items, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2004.
4. Maatsch JL, Huang RR, Downing SM, Munger BS. The predictive validity of test formats and a psychometric theory of clinical competence. The 23rd Conference on Research in Medical Education. Washington, DC: Association of American Medical Colleges, 1984.
5. Jozefowicz RF, Koeppe BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med 2002;77(2):156-61.
6. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008;42:198-206.

7. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Adv Health Sci Educ Theory Pract* 2005;10:133-43.
8. Case SM, Swanson D. *Constructing written test questions for the basic and clinical sciences*, 3rd ed. Philadelphia, PA: National Board of Medical Examiners, 2002.
9. Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. *Appl Meas Educ* 1989;2(1):51-78.
10. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ* 2002;15:309-34.
11. Downing SM, Dawson-Saunders B, Case SM, Powell RD. The psychometric effects of negative stems, unfocused questions, and heterogeneous options on NBME Part I and Part II item characteristics. the annual meeting of the National Council on Measurement in Education. Chicago, IL, 1991.
12. Tamir P. Positive and negative multiple choice items: How different are they? *Stud Educ Eval* 1993;19:311-25.
13. Rogers WT, Harley D. An empirical comparison of three- and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Educ Psychol Meas* 1999;59:234-47.
14. Sidick JT, Barrett GV, Doverspike D. Three-alternative multiple choices tests: An attractive option. *Pers Psychol* 1994;47:829-35.
15. Cizek GJ, Rachor RE. Nonfunctioning options: A closer look. The annual meeting of the American Educational Research Association. San Francisco, CA, 1995.
16. Crehan KD, Haladyna TM, Brewer BW. Use of an inclusive option and the optimal number of options for multiple-choice items. *Educ Psychol Meas* 1993;53:241-7.
17. Lord FM. Optimal number of choices per item. *J Educ Meas* 1977; 14:33-8.
18. Haladyna TM, Downing SM. How many options is enough for a multiple-choice item? *Educ Psychol Meas* 1993;53:999-1010.

ข้อผิดพลาดที่ควรระวังในการสร้าง ข้อสอบปรนัย

ผู้ช่วยศาสตราจารย์ นายแพทย์เชิดศักดิ์ โสมนิรัตน์

ภาควิชาศัลยศาสตร์, คณะแพทยศาสตร์ศิริราชพยาบาล, มหาวิทยาลัยมหิดล, กรุงเทพมหานคร ๑๐๓๐๐.

ข้อผิดพลาดที่ควรระวังในการสร้างข้อสอบปรนัย

ข้อสอบปรนัย (multiple-choice question) เป็นรูปแบบการประเมินผลที่นิยมใช้กันอย่างแพร่หลายในวงการแพทยศาสตรศึกษา ข้อสอบชนิดนี้เป็นที่ชื่นชอบของนักศึกษาผู้เข้าสอบจำนวนมากเนื่องจากมีคำตอบให้เลือก หากไม่มีความรู้ก็สามารถเดาได้ ซึ่งต่างไปจากข้อสอบประเภทอัตนัยซึ่งผู้สอบต้องเขียนคำตอบจากความคิดของตนเอง^๑ ดังนั้นข้อสอบปรนัยจึงเป็นข้อสอบที่ผู้สอบทำได้ง่าย แต่ในทางตรงข้ามข้อสอบปรนัยเป็นข้อสอบที่สร้างปัญหาให้กับอาจารย์ผู้สร้างข้อสอบไม่น้อย เนื่องจากในกระบวนการเขียนข้อสอบปรนัยแต่ละข้อนั้นต้องใช้ทักษะอย่างมาก ต้องใช้ทั้งศาสตร์และศิลป์ และบ่อยครั้งอาจารย์ผู้สร้างข้อสอบก็ถูกขอให้ทำการปรับแก้ข้อสอบเนื่องจากคณะกรรมการพิจารณาข้อสอบมีความเห็นว่ารายละเอียดในข้อสอบไม่เหมาะสม มีการศึกษาวิจัยพบว่าคุณภาพของข้อสอบปรนัยที่พัฒนาขึ้นในโรงเรียนแพทย์หลายแห่งนั้นไม่สู้ดีนัก มีข้อสอบที่มีลักษณะไม่เหมาะสมอยู่จำนวนไม่น้อย^{๒-๓} ข้อสอบปรนัยที่มีลักษณะไม่เหมาะสมเหล่านี้ส่งผลเสียต่อการสอบได้หลายประการ เช่น ทำให้ข้อสอบยากขึ้น สร้างความสับสนให้ผู้สอบ ทำให้ผู้สอบบางกลุ่มเสียเปรียบ และทำให้การตัดสินผลสอบผิดพลาด เป็นต้น^{๓-๕} ดังนั้นการออกข้อสอบปรนัยที่มีคุณภาพดีจึงเป็นงานที่มีความสำคัญและท้าทายความสามารถ

การสร้างข้อสอบปรนัยที่มีคุณภาพดีนั้นควรเริ่มต้นจากการมีองค์ความรู้พื้นฐานในการสร้างข้อสอบแล้ว เกิดการฝึกฝนทักษะ สังเกตประสบการณ์ในการออกข้อสอบ จนเกิดความชำนาญ ปัญหาที่พบบ่อยในโรงเรียนแพทย์หลายแห่งคือมีอาจารย์จำนวนไม่น้อยที่ได้รับมอบหมายให้ออกข้อสอบปรนัย โดยไม่ได้มีการพัฒนาองค์ความรู้พื้นฐานที่เหมาะสมก่อน ซึ่งเป็นเหตุให้มีข้อสอบปรนัยที่มีลักษณะไม่เหมาะสมตามหลักการออกข้อสอบปะปนมาในข้อสอบที่ให้นักศึกษาแพทย์และแพทย์ประจำบ้านทำอยู่บ้าง ผู้นิพนธ์จึงเห็นความสำคัญของการเผยแพร่องค์ความรู้พื้นฐานของการออกข้อสอบปรนัย องค์ความรู้พื้นฐานในการสร้างข้อสอบปรนัยนั้นมีสองส่วน ส่วนแรกเป็นหลัก การของการสร้างข้อสอบทั่วไปซึ่งได้มีผู้รวบรวมเป็นข้อแนะนำตีพิมพ์ในตำราและวารสารทางวิชาการอยู่บ้าง^{๑,๕-๗} ส่วนที่สองเป็นข้อผิดพลาดในการสร้างข้อสอบที่อาจารย์ผู้ออกข้อสอบพึงหลีกเลี่ยง ในบทความนี้ผู้นิพนธ์จะมุ่งเน้นในส่วนที่สองนี้ โดยจะรวบรวมข้อผิดพลาดในการสร้างข้อสอบปรนัย ที่อาจเป็นตัวบ่งชี้ให้ผู้สอบที่ไม่มีความรู้ในเรื่องที่ทำการทดสอบสามารถเลือกคำตอบที่ถูกต้องได้ ดังนั้นการที่อาจารย์ผู้ออกข้อสอบทราบถึงสิ่งเหล่านี้และหลีกเลี่ยงเสียจะส่งผลให้ข้อสอบปรนัยที่สร้างขึ้นสามารถใช้วัดองค์ความรู้ทางการแพทย์ได้จริง โดยปราศจากปัจจัยรบกวนจากการสังเกตพบสิ่งบ่งชี้คำตอบ

ข้อสอบปรนัยที่กล่าวถึงในบทความนี้มุ่งประเด็นไปที่ข้อสอบปรนัยชนิดเลือกคำตอบที่ถูกที่สุด (one best response) เป็นสำคัญ เนื่องจากเป็นข้อสอบที่ใช้กันแพร่หลายมากที่สุดในการวัดผลการศึกษาในโรงเรียนแพทย์ไทยปัจจุบัน ในข้อสอบชนิดนี้แต่ละข้อจะมีโจทย์ (stem) ตามด้วยตัวเลือก (options) จำนวน ๔-๕ ตัวเลือก ผู้สอบต้องเลือกคำตอบที่ถูกที่สุดเพียงคำตอบเดียวจากตัวเลือกเหล่านี้ ตัวเลือกอื่น ๆ ที่ไม่ใช่คำตอบเรียกว่าตัวลวง (distractors)

ในบทความนี้ผู้นิพนธ์ขอนำเสนอข้อผิดพลาดในการออกข้อสอบ ๗ กลุ่มด้วยกัน ได้แก่ (๑) ข้อผิดพลาดในไวยากรณ์, (๒) การไปคำตอบด้วยหลักตรรกะ, (๓) การใช้คำคุณศัพท์บอกระดับของความแน่ชัด, (๔) ความยาวของตัวเลือก, (๕) การใช้คำซ้ำในโจทย์และตัวเลือก, (๖) การเข้าพวกของคำ หรือข้อความที่ปรากฏในตัวเลือก, และ (๗) การบอกไปคำตอบโดยโจทย์ข้ออื่น

๑. ข้อผิดพลาดในไวยากรณ์

ตัวเลือกทุกตัวต้องสามารถตอบโจทย์ได้อย่างถูกต้องตามหลักไวยากรณ์ บ่อยครั้งอาจารย์ผู้ออกข้อสอบมุ่งความสนใจไปที่คำตอบที่ถูก และให้ความสนใจกับตัวลวงน้อยไปจนทำให้ตัวลวงผิดหลักไวยากรณ์^๑ โดยมักพบบ่อยในข้อสอบที่เป็นภาษาอังกฤษ ข้อผิดพลาดที่พบได้บ่อยเช่น ความไม่เข้ากันของ article (A, An, The) กับคำนามที่ตามหลัง, คำนามกับกริยาที่ไม่เข้ากันในเชิงเอกพจน์หรือพหูพจน์, การเติมคำในประโยคที่เว้นว่างไว้สำหรับเติมคำนามแต่ตัวลวงเป็นกริยาหรือเป็นคำนามในลักษณะที่ไม่เข้ากับรูปประโยค เป็นต้น

ตัวอย่างที่ ๑. A 70-year-old woman was brought in an emergency room with alteration of consciousness. Her vital signs were stable, but her Glasgow coma score was E1V1M3. After endotracheal intubation, the next step is to provide intravenous administration of ...

- A. lumbar puncture
- B. computerized scan of the brain
- C. glucose with Thiamine
- D. Sodium bicarbonate

ในตัวอย่างที่ ๑ นี้โจทย์ให้ผู้สอบเลือกตัวเลือกไปเติมในช่องว่าง ซึ่งสิ่งที่เติมลงในช่องว่างได้นั้นต้องเป็นยาที่สามารถให้ทางหลอดเลือดดำได้ ผู้สอบที่มีทักษะการทำข้อสอบดีจะตัดตัวเลือก A และ B ได้โดยไม่ต้องใช้ความรู้ทางการแพทย์

ตัวอย่างที่ ๒. Which organism is the cause of syphilis?

- A. *Neisseria gonorrhoeae*
- B. *Chlamydia trachomatis* and *Giardia lamblia*
- C. *Treponema pallidum*
- D. *Ureaplasma urealyticum* and *Mycoplasma genitalium*

ในตัวอย่างที่ ๒ นี้โจทย์ถามหาเชื้อก่อโรค โดยใช้รูปประโยคถามหาคำตอบที่เป็นเอกพจน์ ดังนั้นคำตอบที่ถูกต้องย่อมมีเชื้อก่อโรคตัวเดียว ผู้สอบที่มีทักษะการทำข้อสอบดีจะตัดตัวเลือก B และ D ได้โดยไม่ต้องใช้ความรู้ทางการแพทย์

๒. การไปคำตอบด้วยหลักตรรกะ

ในการเขียนตัวเลือก อาจารย์ผู้ออกข้อสอบต้องระมัดระวังไม่ให้ผู้สอบสามารถตัดตัวเลือกได้ด้วยหลักตรรกศาสตร์ เนื่องจากผู้สอบที่มีทักษะการทำข้อสอบดีจะสามารถพิจารณาความเป็นไปได้ของตัวเลือกต่าง ๆ และตัดตัวลวงที่ไม่มีทางเป็นไปได้ตามหลักของเหตุและผลออกไปได้โดยไม่ต้องอาศัยความรู้เรื่องที่อาจารย์ตั้งเป้าหมายว่าจะทดสอบ

ตัวอย่างที่ ๓. กวาระไส้เลื่อนบริเวณขาหนีบ (inguinal hernia)

- A. พบในผู้ชายบ่อยกว่าผู้หญิง
- B. พบในผู้หญิงบ่อยกว่าผู้ชาย
- C. พบเกิดขึ้นในผู้หญิงและผู้ชายในอัตราเท่ากัน
- D. พบบ่อยในผู้ที่มีเศรษฐกิจฐานะยากจน
- E. พบในผู้ที่มีภูมิลำเนาในทวีปเอเชีย มากกว่าผู้ที่มีภูมิลำเนาในทวีปยุโรป

ในตัวอย่างที่ ๓ นี้อาจารย์ผู้ออกข้อสอบต้องการวัดความรู้เรื่องอุบัติการณ์ของไส้เลื่อนขาหนีบ แต่หาก

เวบบันทึทศึรึรึรึ

บทความทึวไป

พิจารณาตามหลักตรรกศาสตร์แล้ว ตัวเลือก A, B, และ C เพียงสามตัวเลือกก็ครอบคลุมสิ่งที่เป็นไปได้ทั้งหมดแล้ว (เนื่องจากมนุษย์มีสองเพศ ภาวะไส้เลื่อนนี้หากไม่มีอัตราการเกิดเท่ากันในสองเพศแล้วก็ต้องมีเพศใดเป็นมากกว่าอีกเพศหนึ่ง) ดังนั้นผู้สอบที่มีทักษะการทำข้อสอบดีสามารถตัดตัวเลือก D และ E ได้โดยไม่ต้องมีความรู้เรื่องไส้เลื่อนเลย

๓. การใช้คำคุณศัพท์บอกระดับของความแน่ชัด

อาจารย์ผู้ออกข้อสอบพึงระมัดระวังการใช้คำคุณศัพท์ที่บ่งบอกถึงความแน่ชัดของข้อความ ซึ่งจะมีหลายระดับ โดยทั่วไปแล้วคำคุณศัพท์ที่แสดงความแน่ชัดมาก แสดงความมั่นใจมาก (เช่น always, never) มักไม่ถูกต้อง เนื่องจากในทางการแพทย์นั้นมีความไม่แน่นอนเกิดขึ้นเป็นประจำ ข้อความที่บอกเล่าถึงสิ่งที่เป็นไปได้โดยไม่ชี้ชัดลงไปว่าต้องเกิดขึ้นแน่นอน (เช่น may, might, can, could) มักเป็นข้อความที่ถูก

ตัวอย่างที่ ๔. Which of the following statements is true regarding the etiology of an inguinal hernia?

- A. Some connective tissue diseases may increase the incidence of inguinal hernia.
- B. Patients with Marfan syndrome always developed inguinal hernia.
- C. MRI scan of pelvis is the only reliable investigation for detection of groin hernia.
- D. Persistent lifting of heavy weights inevitably leads to the development of groin hernia.

ในตัวอย่างที่ ๔ นี้ผู้สอบต้องเลือกข้อความเกี่ยวกับไส้เลื่อนขาหนีบที่ถูกต้องหนึ่งข้อความ หากสังเกตดูทั้งสี่ข้อความมีการใช้คำคุณศัพท์บอกความแน่ชัดของข้อความ ได้แก่ may (ตัวเลือก A), always (ตัวเลือก B), the only (ตัวเลือก C), inevitably (ตัวเลือก D) ซึ่งจะเห็นว่าตัวเลือก B, C, และ D เป็นข้อความที่แสดงความแน่ชัดว่าต้องเป็นแน่ ต้องใช้แน่นอน ไม่มีทางเลี่ยงได้ ข้อความทำนองนี้มีโอกาสสูงที่จะผิด ในทางตรงข้ามตัวเลือก A เป็นข้อความบอกว่ามีโอกาสเป็นไปได้โดยไม่ชี้ชัดว่าต้องเกิด

ผู้สอบที่มีทักษะการทำข้อสอบดีจะตัดตัวเลือก B, C, และ D ได้โดยไม่ต้องอาศัยความรู้ทางการแพทย์เลย

๔. ความยาวของตัวเลือก

มีการตั้งข้อสังเกตว่าอาจารย์แพทย์มักชอบสอนและอธิบายแม้กระทั่งในการสอบอาจารย์แพทย์หลายท่านก็ติดนิสัยรักการสอนนี้มาด้วย ทำให้อาจารย์มักเขียนตัวเลือกที่ถูกต้องที่มีคำอธิบายประกอบอย่างครบถ้วนทำให้ตัวเลือกที่ถูกมักมีความยาวมากกว่าตัวลวง^๑ นักศึกษาผู้เข้าสอบจำนวนไม่น้อยรู้ถึงความจริงข้อนี้และมักเลือกตัวเลือกที่มีความยาวมากที่สุด หากเขาไม่สามารถหาคำตอบได้ด้วยความรู้ทางการแพทย์ที่เขา

ตัวอย่างที่ ๕. ผู้หญิงอายุ ๒๘ ปี แต่งงานมานาน ๑ ปี ยังไม่มีบุตร คุณกำเนิดโดยการกินยาคุมเป็นประจำ สังเกตว่าตนเองน้ำหนักตัวเพิ่มขึ้นหลังจากกินยาคุมมาขอคำแนะนำเรื่องการคุมกำเนิด ท่านจะแนะนำอย่างไร

- A. ให้เปลี่ยนไปใช้การใส่ห่วงอนามัย
- B. ให้ใช้ถุงยางอนามัย
- C. ให้กินยาคุมกำเนิดต่อได้เนื่องจากการมีศึกษาแล้วว่ายาคุมกำเนิดชนิดกินไม่ส่งผลให้เกิดการเพิ่มขึ้นของน้ำหนักตัว
- D. ให้รับประทานยาลดความอ้วน

ในตัวอย่างที่ ๕ นี้จะสังเกตเห็นว่าตัวเลือก C มีการอธิบายเหตุผลประกอบส่งผลให้มีความยาวมากกว่าตัวเลือกอื่นชัดเจน ลักษณะเช่นนี้จะเป็นการบอกใบ้ให้นักศึกษาเลือกตัวเลือกนี้

๕. การใช้คำซ้ำในโจทย์และตัวเลือก

การใช้คำเดียวกัน หรือคำที่มีความหมายเหมือนกันในโจทย์และตัวเลือก มักเป็นการบอกใบ้ว่าตัวเลือกลงกล่าวเป็นตัวเลือกที่ถูกต้อง^๑

ตัวอย่างที่ ๖. Which of the following statements is true regarding saccular theory of indirect inguinal hernia formation?

- A. An increased intra-abdominal pressure is the cause of inguinal hernia.
- B. A developmental diverticulum associated with a patent processus vaginalis is the cause of inguinal hernia.

C. All persons with a persistent processus vaginalis will develop an inguinal hernia.

D. A direct inguinal hernia is caused by the weakness of the posterior inguinal wall.

ในตัวอย่างที่ ๖ นี้โจทย์ถามถึง saccular theory ซึ่งหากแปลความหมายก็น่าจะเป็นเรื่องที่เกี่ยวข้องกับถุง (sac) ผู้สอบที่มีทักษะการทำข้อสอบดีจะหาตัวเลือกที่มีคำที่มีความหมายเกี่ยวกับถุง แล้วเลือกตัวเลือกดังกล่าวทันที ซึ่งในที่นี่จะพบคำว่า diverticulum ซึ่งมีความหมายว่าถุงในข้อ B การที่มีคำที่มีความหมายซ้ำกันเช่นนี้เป็นตัวบอกใบ้คำตอบที่อาจารย์ผู้ออกข้อสอบต้องตรวจตราให้ดีก่อนนำข้อสอบไปใช้

๖. การเข้าพวคของคำ หรือข้อความที่ปรากฏในตัวเลือก

ข้อสอบจำนวนไม่น้อยนำเสนอรายการของหลายอย่างในตัวเลือก (เช่น ชื่อการตรวจค้นเพิ่มเติม ชื่อโรค ชื่อยา ฯลฯ) มีผู้เชี่ยวชาญในการประเมินผลตั้งข้อสังเกตว่าในข้อสอบเหล่านี้ตัวเลือกที่ถูกต้องมักมีลักษณะเข้าพวคกับตัวเลือกอื่นมากที่สุด หากเป็นรายการของตัวเลือกที่ถูกก็คือข้อที่มีจำนวนรายการซ้ำกับตัวเลือกอื่นมากที่สุด ดังนั้นในการนำเสนอตัวเลือกอาจารย์ผู้ออกข้อสอบพึงระมัดระวังอย่าให้ตัวเลือกที่ถูกต้องมีลักษณะที่เข้าพวคได้อย่างชัดเจน พยายามทำตัวลวงอื่นให้มีลักษณะเข้าพวคให้ใกล้เคียงกับตัวเลือกที่ถูกต้อง

ตัวอย่างที่ ๗. โรคที่แพทย์วินิจฉัยผิดว่าเป็นไส้ติ่งอักเสบบ่อยที่สุดเรียงลำดับจากมากไปน้อยคือ

A. acute mesenteric lymphadenitis, pelvic inflammatory disease, twisted ovarian cyst

B. acute mesenteric lymphadenitis, Meckel diverticulitis, acute cholecystitis

C. Meckel diverticulitis, twisted ovarian cyst, sigmoid diverticulitis

D. pelvic inflammatory disease, acute gastroenteritis, right ureteric calculi

ในตัวอย่างที่ ๗ นี้โจทย์ถามชื่อโรค ตัวเลือกแสดงรายการชื่อโรค ตัวเลือกละสามโรค หากนับจำนวนของคำซ้ำจะพบว่าโรคที่กล่าวถึงบ่อยที่สุดคือ acute

mesenteric lymphadenitis, pelvic inflammatory disease, twisted ovarian cyst, และ Meckel diverticulitis (กล่าวถึงโรคละ ๒ ครั้ง) ส่วนโรคที่เหลือกล่าวถึงโรคละครั้งเดียว ดังนั้นตัวเลือกที่มีพวคมากที่สุดคือตัวเลือก A ซึ่งเป็นคำตอบที่ถูกต้อง

การเข้าพวคของตัวเลือกที่ถูกต้องนั้น ไม่จำเป็นต้องเป็นลักษณะของการมีจำนวน หรือความถี่ของคำมากที่สุดเพียงเท่านั้น อาจหมายรวมถึงการมีรูปร่างลักษณะ หรือความหมายคล้ายคลึงกันได้ด้วย

ตัวอย่างที่ ๘. ชายอายุ ๕๕ ปีเป็นมะเร็งเม็ดเลือดขาว หลังได้รับยาเคมีบำบัด ๑๔ วันมีไข้สูง ได้รับการวินิจฉัยเป็น febrile neutropenia การรักษาในข้อใดเหมาะสมที่สุด

A. Amoxycillin PO

B. Ceftazidime IV + Amikacin IV

C. Amphotericin B IV + Ceftazidime IV

D. Cloxacillin IV + Metronidazole IV

ในตัวอย่างที่ ๘ นี้โจทย์ถามถึงยาที่ควรให้กับผู้ป่วย ในตัวเลือกสี่ตัวเลือกนี้มียาเกินเพียงข้อเดียว (A) ที่เหลือเป็นยาฉีดสองขนานควบกัน ดังนั้นตัวเลือกข้อ A ไม่เข้าพวค จะถูกตัดทิ้งได้โดยง่าย ในบรรดายาฉีดจะเห็นว่ามียาต้านเชื้อราที่ไม่เข้าพวค (ตัวเลือก C) ดังนั้นจะเหลือตัวเลือกที่นักศึกษาต้องคิดเลือกจริง ๆ เพียงตัวเลือก B กับ D ซึ่งหากดูกลุ่มยาก็จะพบว่ายากลุ่ม Cephalosporin เข้าพวคมากที่สุด ทำให้ผู้สอบที่มีทักษะการทำข้อสอบดีสามารถเลือกคำตอบที่ถูกต้อง (ตัวเลือก B) ได้โดยไม่ต้องมีความรู้เรื่องการรักษาผู้ป่วย febrile neutropenia

๗. การบอกใบ้คำตอบโดยโจทย์ข้ออื่น

ข้อผิดพลาดนี้เป็นข้อผิดพลาดที่ตัวผู้เขียนข้อสอบไม่ค่อยรู้ แต่ผู้ที่ตรวจพบข้อผิดพลาดนี้คืออาจารย์ผู้เลือกข้อสอบไปใช้ เนื่องจากในการสอบแต่ละครั้งใช้ข้อสอบจำนวนมาก หากเลือกข้อสอบโดยไม่ระมัดระวังอาจมีข้อสอบสองข้อที่ถามเกี่ยวกับโรคหรือกลุ่มอาการเดียวกัน ซึ่งข้อมูลจากโจทย์ในข้อหนึ่งอาจเป็นตัวบอกใบ้คำตอบของข้อสอบอีกข้อได้ ดังนั้นเมื่อทำการเลือกข้อสอบเสร็จแล้วจัดหน้ากระดาษเข้ารูปเล่มข้อสอบแล้วอาจารย์ควรอ่านข้อสอบฉบับสมบูรณ์อีกครั้งหนึ่งหรือสองรอบก่อนส่ง

เวบบันทึทศึรราษ

บทความท้อไป

ไปพิมพ์ ซึ่งการอ่านทวนในขั้นตอนนี้อาจทำให้ตรวจพบข้อสอบที่มีเนื้อหาซ้ำซ้อนกันได้

ตัวอย่างที่ ๙. ผู้ป่วย febrile neutropenia มักมีไข้ขึ้นหลังจากได้รับยาเคมีบำบัดเป็นเวลากี่วัน

- A. 2 - 4 วัน
- B. 3 - 5 วัน
- C. 5 - 7 วัน
- D. 10 - 14 วัน

ในตัวอย่างที่ ๙ นี้อาจารย์ผู้ออกข้อสอบต้องการวัดความรู้ของผู้สอบเรื่อง febrile neutropenia ซึ่งเนื้อหาไปซ้ำซ้อนกับโจทย์ในตัวอย่างที่ ๘ ซึ่งผู้สอบที่มีทักษะการทำข้อสอบดีสามารถย้อนกลับไปอ่านโจทย์ในข้อก่อนหน้านี้อแล้วได้ข้อมูลว่าผู้ป่วยที่น่าเสนอว่าเป็น febrile neutropenia มีไข้ขึ้น ๑๔ วันหลังได้ยาเคมีบำบัด ก็สามารถตอบข้อสอบข้อนี้ถูกต้องได้ง่าย

สรุป

ผู้นิพนธ์ได้รวบรวมข้อผิดพลาดในการสร้างข้อสอบปรนัยที่ผู้สอบอาจใช้เป็นแนวทางในการเลือกคำตอบที่ถูกต้องได้โดยไม่ต้องอาศัยความรู้ทางการแพทย์ที่อาจารย์ต้องการประเมินผล โดยเรียบเรียงเป็นเจ็ดกลุ่มข้อผิดพลาดด้วยกัน ผู้อ่านทุกท่านพึงตระหนักว่าสิ่งเหล่านี้ไม่ใช่หลักการทางวิทยาศาสตร์ที่ชัดเจนดังกฎทางคณิตศาสตร์หรือฟิสิกส์ หากแต่เป็นการรวบรวมข้อสังเกต

และคำแนะนำของผู้เชี่ยวชาญทางการวัดและประเมินผล จึงเป็นเพียงแนวทางเบื้องต้นในการพิจารณาตรวจสอบเนื้อหาของข้อสอบเท่านั้น การประยุกต์ใช้องค์ความรู้นี้คงต้องอาศัยศิลปะพอสมควรเพื่อที่จะได้ข้อสอบที่ดีสามารถวัดองค์ความรู้ทางการแพทย์ของนักศึกษาหรือแพทย์ประจำบ้านที่เข้าสอบได้ตามวัตถุประสงค์ของการสอบ

เอกสารอ้างอิง

1. Haladyna TM. Developing and validating multiple-choice test items, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2004.
2. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med. 2002;77:156-61.
3. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ. 2008;42:198-206.
4. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract. 2005;10:133-43.
5. Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. Appl Meas Educ. 1989;2:37-50.
6. Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. Appl Meas Educ. 1989;2:51-78.
7. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ. 2002;15:309-34.
8. Case SM, Swanson D. Constructing written test questions for the basic and clinical sciences, 3rd ed. Philadelphia, PA: National Board of Medical Examiners, 2002.

รศ. ดร. นพ. เชิดศักดิ์ ไอรมนิรัตน์

หัวข้อ : MCQ item analysis

MCQ Item Analysis

Cherdsak Iramaneerat
Department of Surgery
Faculty of Medicine Siriraj Hospital
Mahidol University

Item Analysis

- A group of statistical analyses having two characteristics:
 - The data consist of actual responses of test takers to individual test items
 - The primary purpose is to gain information about the items (rather than about test takers)

Livingston SA. Item analysis. In: Downing SM, Haladyna TM. Handbook of test development. Mahwah, NJ: LEA, 2006, p. 421-444.

MCQ item analysis

Objectives

- เมื่อสิ้นสุดการอบรมแล้ว อาจารย์ผู้เข้าอบรมสามารถ
 - อธิบายผลการวิเคราะห์ข้อสอบ MCQ ที่ใช้บ่อยทางแพทยศาสตรศึกษาได้อย่างถูกต้อง
 - นำผลการวิเคราะห์ข้อสอบไปเป็นแนวทางในการพัฒนาคุณภาพของข้อสอบ MCQ ในภาควิชาของตนได้
 - บอกถึงข้อควรระวัง และข้อจำกัดในการวิเคราะห์ผลการสอบ MCQ

MCQ item analysis

Outline

- Item statistics
- Test statistics
- Applications
- Limitations

MCQ item analysis

Two Parts of Item Analysis

- Item statistics
 - Item difficulty
 - Item discrimination
 - Distractor functionality
- Test statistics
 - Internal consistency reliability
 - Standard deviation and mean
 - Average difficulty
 - Average discrimination

MCQ item analysis

Item Statistics

Looking at individual test items

MCQ item analysis

Item Difficulty

- Proportion of examinees answering an item correctly (p)

$$p = \frac{C}{C+I}$$

C = number of examinees with a correct answer
 I = number of examinees with incorrect answers

- Ideal: 0.45 - 0.75
- Good: 0.76 - 0.91
- Acceptable: 0.25 - 0.44
- Problematic: < 0.24 or > 0.91

MCQ Item analysis

Item Discrimination

- The ability of an item to discriminate high scorers from low scorers
- Point-biserial correlation (r)

$$r = \frac{M_p - M_q}{SD} \sqrt{pq}$$

- M_p = Mean score of examinees with a correct answer
- M_q = Mean score of examinees with incorrect answers
- SD = Standard deviation of test scores
- p = Proportion of examinees with a correct answer
- q = Proportion of examinees with incorrect answers

MCQ Item analysis

Point-Biserial Correlation

The correlation between an item score with the total score

- Range: -1.0 - 1.0
- Point-biserial of an item should be positive
 - Ideal: 0.20 or higher
 - Acceptable: 0.1 - 0.19
 - Problematic: < 0

MCQ Item analysis

Distractor Functionality

A functioning distractor is an incorrect option that:

- Is chosen by at least 5 percent of examinees
- Has a negative point-biserial correlation with the total score

MCQ Item analysis

Example 1

Number 148	Correct answer = 2					
P-VALUE = 0.65	PT BISERIAL = 0.1					Total number of examinees
DISTRACTOR	1	2	3	4	5	
N OF PEOPLE	4	158	17	58	5	242
MEAN SCORE	77.25	84.81	81.35	83.86	76.6	
P-VALUE	0.02	0.65	0.07	0.24	0.02	
PT BISERIAL	-0.09	0.1	-0.07	-0.01	-0.11	

MCQ Item analysis

Example 2

Number 145	Correct answer = 3					
P-VALUE = 0.79	PT BISERIAL = 0.34					Total number of examinees
DISTRACTOR	1	2	3	4	5	
N OF PEOPLE	7	27	190	9	9	242
MEAN SCORE	77	78.11	85.81	78.22	75.89	
P-VALUE	0.03	0.11	0.79	0.04	0.04	
PT BISERIAL	-0.12	-0.21	0.34	-0.11	-0.16	

MCQ Item analysis

Example 3

Number 124		Correct answer = 2							
P-VALUE =	0.14	PT BISERIAL = -0.14							Total number of examinees
DISTRACTOR		1	2	3	4	5			
N OF PEOPLE		8	33	22	133	46		242	
MEAN SCORE		87	87.52	78.05	84.3	83.17			
P-VALUE		0.03	0.14	0.09	0.55	0.19			
PT BISERIAL		0.05	0.14	-0.19	0.03	-0.04			

MCQ Item analysis

Example 4

Number 112		Correct answer = 3							
P-VALUE =	0.73	PT BISERIAL = -0.05							Total number of examinees
DISTRACTOR		1	2	3	4	5			
N OF PEOPLE		0	1	177	1	63		242	
MEAN SCORE		0	84	83.74	83	84.92			
PVALUE		0	0	0.73	0	0.26			
PT BISERIAL		0	0	-0.05	-0.01	0.05			

MCQ Item analysis

Siriraj Hospital's IA report

No. : 1		p Value : 0.64					rpb1 : 0.23				
A	B	C	* D	E							
rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%
0.02	6.98	-0.18	5.08	-0.17	8.57	0.23	63.81	-0.07	15.56		

MCQ Item analysis

Item Analysis and Option Analysis

Faculty of Medicine Siriraj Hospital
Mahidol University

No. : 1		p Value : 0.64					rpb1 : 0.23					No. : 2		p Value : 0.34					rpb1 : 0.19				
A	B	* C	D	E								A	B	* C	D	E							
rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%
0.02	6.98	-0.18	5.08	-0.17	8.57	0.23	63.81	-0.07	15.56			0.01	4.76	-0.02	25.40	-0.19	15.79	-0.06	24.70	0.19	33.91		
No. : 3		p Value : 0.56					rpb1 : 0.35					No. : 4		p Value : 0.50					rpb1 : 0.33				
A	B	* C	D	E								A	B	* C	D	E							
rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%
-0.03	9.38	-0.20	23.17	0.35	55.87	-0.05	3.17	-0.16	8.88			-0.15	1.90	0.33	50.48	-0.15	4.13	-0.18	10.43	-0.13	33.02		
No. : 5		p Value : 0.24					rpb1 : 0.06					No. : 6		p Value : 0.53					rpb1 : 0.20				
A	B	* C	D	E								A	B	* C	D	E							
rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%	rpb1	%
-0.06	3.49	-0.06	53.02	0.05	12.06	0.06	23.81	0.02	7.42			-0.16	23.17	-0.11	3.81	0.20	53.33	-0.02	5.40	-0.02	14.29		

MCQ Item analysis

Exercise

- แบ่งกลุ่มย่อย 6 กลุ่ม
- ให้อาจารย์พิจารณาผลการวิเคราะห์ข้อสอบ 72 ข้อที่ได้รับ
- ข้อสอบที่มีปัญหามากที่สุด 3 อันดับแรกคือ...

MCQ Item analysis

Test Statistics

Looking at the whole test

MCQ Item analysis

Reliability

- Consistency of test scores
 - If we test the students again, will they get the same scores?
 - Range: 0 – 1
 - High values: highly consistent test scores

20

Internal Consistency Reliability

- Consistency of test scores: If we test the students again, will they get the same scores?
- In MCQ exam, one commonly reported index of reliability is Cronbach's Alpha

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

- n = number of testlets
- σ_t^2 = score variance of total scores
- σ_i^2 = score variance of the i^{th} testlet

MCQ Item analysis

How Much is Enough?

- Depends on test scores uses
 - High-stakes exam: 0.9 or higher
 - Medium-stakes exam: 0.80 – 0.89
 - Low-stakes exam: 0.70 – 0.79

22

Improving Reliability

- Increase the number of test items
- Adjust item difficulty to obtain larger spread of test scores
- Adjust testing conditions to eliminate interruptions, noise, and other disrupting factors
- Eliminate subjectivity in scoring

23

Mean and Standard Deviation

- Effective instruction => All students can do the test well.
 - High mean scores
 - Low standard deviation
- High standard deviation: Wide range of students' scores
 - Some students can solve the problems in the tests, while some students cannot do.
- Too difficult test => Most students fail to get correct answers.
 - Low mean scores
 - Low standard deviation

MCQ Item analysis

Average Difficulty

- Average of p values of all items on the test
- Small group of students:
 - Difficult to interpret
 - Depends on the ability distribution of students
- Large group of students:
 - Assume a fair sampling of students
 - Indicates the average difficulty of the whole test

MCQ Item analysis

Average Discrimination

- Average point-biserial correlation of the whole test
- Indicates how good the items on the test can differentiate high scorers from low scorers.
- High values generally indicate a good test.
- Effective instruction: All students can do well on the test.
 - A low value does not necessarily indicate bad items.

MCQ Item analysis

Applications

1. Posttest score adjustment
2. Item revision
3. Item pool management
4. Improvement of instruction

MCQ Item analysis

Limitations

1. Sample dependency
2. Reliability is the property of test scores, not test items.
3. Numbers are there to serve us, not the other way around.

MCQ Item analysis

Summary

- Item statistics
- Test statistics
- Applications
- Limitations

MCQ Item analysis

Questions and Comments

Cherdsakiramaneerat@gmail.com

MCQ Item analysis

"We all need people who will give us feedback. That's how we improve."

Bill Gates

การวิเคราะห์ข้อสอบปรนัย

อาจารย์ นายแพทย์เชดศักดิ์ โสมณรัตน์

ภาควิชาสรีรศาสตร์, คณะแพทยศาสตร์ศิริราชพยาบาล, มหาวิทยาลัยมหิดล, กรุงเทพมหานคร ๑๐๓๑๐.

การวิเคราะห์ข้อสอบปรนัย (Item analysis) เป็นการใช้วิธีการทางสถิติเพื่อวิเคราะห์คำตอบที่ผู้สอบตอบข้อสอบปรนัยในการสอบครั้งหนึ่ง เพื่อประเมินว่าข้อสอบที่นำมาใช้ในการสอบครั้งนั้นมีคุณสมบัติอย่างไร ทำงานได้ตามที่ต้องการหรือไม่ มีระดับความยากง่ายของข้อสอบเหมาะสมหรือไม่ มีข้อบกพร่องหรือไม่ และควรได้รับการปรับปรุงแก้ไขอย่างไร การวิเคราะห์ข้อสอบเป็นศาสตร์ที่ได้รับการพัฒนาอย่างต่อเนื่องมาเป็นเวลานาน มีเทคนิคและวิธีการต่าง ๆ มากมายที่ผู้วิเคราะห์สามารถใช้เพื่อบอกคุณสมบัติของข้อสอบแต่ละข้อ ตั้งแต่วิธีการง่าย ๆ ไปจนถึงวิธีการที่มีความซับซ้อนมาก โดยแต่ละเทคนิคการวิเคราะห์ก็มีจุดประสงค์แตกต่างกันไป ตั้งแต่การบอกระดับความยากง่าย การบอกถึงความสามารถในการแยกผู้สอบที่เก่งออกจากผู้สอบที่ไม่เก่ง ไปจนถึงเทคนิคขั้นสูงที่สามารถบอกได้ว่าข้อสอบมีความลำเอียงต่อผู้สอบเพศใดเพศหนึ่ง หรือผู้สอบจากสถาบันใดสถาบันหนึ่งเป็นพิเศษหรือไม่ มีการเดาข้อสอบมากน้อยเพียงใด ผู้สอบรู้ข้อสอบมาก่อนเข้าสอบหรือไม่ หรือมีความน่าจะเป็นมากน้อยเพียงใดที่ผู้สอบลอกคำตอบ ในบทความนี้ผู้เขียนไม่ได้ตั้งเป้าประสงค์ที่จะรวบรวมและอธิบายเทคนิคการวิเคราะห์ข้อสอบทุกวิธีที่มีใช้อยู่ในปัจจุบัน แต่ต้องการเพียงนำเสนอความรู้พื้นฐานที่เกี่ยวกับการวิเคราะห์ข้อสอบและอธิบายถึงวิธีการวิเคราะห์ข้อสอบที่นิยมใช้กันในทางแพทยศาสตรศึกษา โดยเฉพาะในประเทศไทย โดยประสงค์ให้อาจารย์ผู้อ่านสามารถนำเอาความรู้ที่ได้จากบทความนี้ไปใช้แปลผลการวิเคราะห์ข้อสอบที่ตน

เกี่ยวข้อง และดำเนินการปรับปรุงคุณภาพของข้อสอบได้อย่างเหมาะสม

ความรู้พื้นฐานเกี่ยวกับข้อสอบปรนัย

ก่อนที่จะกล่าวถึงรายละเอียดในการวิเคราะห์ข้อสอบ ผู้นิพนธ์ก็จะขอทบทวนความรู้พื้นฐานเกี่ยวกับข้อสอบปรนัยก่อน โดยทั่วไปข้อสอบปรนัยแต่ละข้อมีส่วนประกอบสำคัญ ๒ ส่วนด้วยกันคือ

๑. โจทย์ (stem) เป็นข้อมูลของโรค หรือภาวะหรือผู้ป่วยตามด้วยคำถาม หรือเว้นช่องว่างสำหรับเติมคำหรือข้อความที่เหมาะสมลงไป

๒. ตัวเลือก (options) คือคำ หรือข้อความที่ผู้ออกข้อสอบนำเสนอตามหลังจากโจทย์เพื่อให้ผู้สอบเลือกไปใช้ตอบคำถาม หรือเติมลงในช่องว่างในโจทย์

๒.๑ ตัวเลือกที่ถูกต้อง (correct option) เป็นคำตอบที่ถูกต้องมีเพียงตัวเลือกเดียวต่อข้อสอบข้อหนึ่ง

๒.๒ ตัวลวง (distractors) เป็นคำตอบที่ผิด มีไว้ลวงให้ผู้สอบที่ไม่มีความรู้ หรือมีความเข้าใจไม่ถูกต้องในเนื้อหาที่นำมาออกข้อสอบเลือกตอบ ข้อสอบที่ใช้ในคณะแพทยศาสตร์ศิริราชพยาบาล และที่ใช้ทั่วไปในการสอบของนักศึกษาแพทย์ และแพทย์ประจำบ้านในประเทศไทย นิยมจัดให้มีตัวลวง ๔ ตัวต่อข้อสอบ ๑ ข้อ

ทฤษฎีที่ใช้ในการวิเคราะห์ข้อสอบ

ทฤษฎีที่ใช้ในการวิเคราะห์ข้อสอบในปัจจุบันนั้นมี ๒ ทฤษฎีด้วยกัน ได้แก่ทฤษฎีการสอบแบบดั้งเดิม

(classical test theory) และทฤษฎีการตอบสนองต่อข้อสอบ (item response theory) ทฤษฎีการสอบแบบดั้งเดิมนั้นเป็นทฤษฎีที่ได้ถูกพัฒนาขึ้นตั้งแต่ตอนต้นของศตวรรษที่ ๒๐ โดยมีการรวบรวมเป็นตำราในครั้งแรกตั้งแต่ปี ค.ศ. ๑๙๒๑ โดย William Brown และ Godfrey H Thomson^๒ หลังจากนั้นทฤษฎีนี้ก็ได้รับการใช้อย่างแพร่หลายในการวิเคราะห์ข้อสอบและได้รับการพัฒนาอย่างต่อเนื่อง ทฤษฎีการสอบแบบดั้งเดิมนี้อาศัยอยู่บนสมมติฐานว่าคะแนนสอบที่ได้มานั้นประกอบไปด้วยคะแนนที่แท้จริง (true score) กับความผิดพลาดจากการวัด (error) ซึ่งสมมติฐานดังกล่าวต่อมาพบว่ามีข้อจำกัดหลายประการด้วยกัน ในราว ค.ศ. ๑๙๗๐ จึงได้มีความพยายามพัฒนาทฤษฎีที่ใช้ในการวิเคราะห์ข้อสอบแบบใหม่ขึ้นซึ่งใช้หลักการของความน่าจะเป็นมาวิเคราะห์ข้อสอบ ทำให้สามารถแยกผลการวิเคราะห์ข้อสอบแต่ละข้อเป็นอิสระจากข้อสอบข้ออื่นในการสอบเดียวกัน ทฤษฎีใหม่นี้เรียกว่าทฤษฎีการตอบสนองต่อข้อสอบ (item response theory) ทฤษฎีใหม่นี้มีข้อได้เปรียบกว่าทฤษฎีเดิมหลายประการด้วยกัน ได้แก่ ความสามารถในการปรับตัวเข้ากับสถานการณ์ต่าง ๆ (flexibility) ความมีประสิทธิภาพในการใช้ข้อมูล (efficiency) และความสามารถในการวิเคราะห์ถึงคุณภาพของข้อสอบ และผู้สอบโดยละเอียด (in-depth analysis)^๓ จึงเป็นเหตุให้ทฤษฎีการตอบสนองต่อข้อสอบนี้ได้รับความนิยมอย่างกว้างขวางตั้งแต่ในค.ศ. ๑๙๘๐ ในปัจจุบันการสอบต่าง ๆ ได้ถูกวิเคราะห์ด้วยทฤษฎีการตอบสนองต่อข้อสอบนี้มากขึ้นเรื่อย ๆ

เนื่องจากการวิเคราะห์ข้อสอบในวงการแพทยศาสตร์ศึกษาในประเทศไทยทั้งหมดในปัจจุบันยังใช้เทคนิคต่าง ๆ ตามทฤษฎีการสอบแบบดั้งเดิมอยู่ ดังนั้นผู้นิพนธ์จะขอกล่าวถึงเทคนิคการวิเคราะห์ข้อสอบตามทฤษฎีการสอบแบบดั้งเดิมเท่านั้น เพราะจะเป็นสิ่งที่อาจารย์แพทย์ทุกท่านจะได้พบและใช้งานเป็นประจำ

การวิเคราะห์ข้อสอบตามทฤษฎีการสอบแบบดั้งเดิม

การวิเคราะห์ข้อสอบตามทฤษฎีการสอบแบบดั้งเดิมนั้นประกอบไปด้วย ๒ ส่วนใหญ่ ๆ คือ (๑) การ

วิเคราะห์ข้อสอบรายข้อ (item analysis) และ (๒) การวิเคราะห์ข้อสอบโดยรวม (test analysis)

๑. การวิเคราะห์ข้อสอบรายข้อ (item analysis)

การวิเคราะห์ข้อสอบแต่ละข้อให้อาจารย์พิจารณา ๓ ปัจจัย คือ

๑.๑ ความยากง่ายของข้อสอบ (item difficulty, p)

ความยากง่ายของข้อสอบวัดโดยใช้ค่า p ซึ่งย่อมาจาก proportion of examinees answering items correctly (สัดส่วนของผู้สอบที่ตอบข้อสอบข้อนั้นถูก) ซึ่งหาได้จากการนำจำนวนผู้สอบที่ตอบข้อสอบข้อนั้นถูกต้องหารด้วยจำนวนผู้สอบที่ตอบข้อสอบข้อนั้นทั้งหมด หากข้อสอบข้อนั้นเป็นข้อสอบที่ง่ายผู้สอบทุกคนตอบถูกค่า p ก็จะเป็น ๑ หากไม่มีผู้สอบคนใดตอบถูกเลยข้อสอบข้อนั้นก็จะมีค่า p เป็น ๐ หากมีคนตอบถูก ๗๐% ข้อสอบข้อนั้นก็จะมีค่า p เท่ากับ ๐.๗ ข้อสอบที่ดีมากจะมีค่า p อยู่ในช่วง ๐.๔๕ - ๐.๗๕, ข้อสอบที่ดีจะมีค่า p อยู่ในช่วง ๐.๗๖ - ๐.๙๑, ข้อสอบที่พอใช้ได้มีค่า p อยู่ในช่วง ๐.๒๕ - ๐.๔๔, ข้อสอบที่มีค่า p ต่ำกว่า ๐.๒๕ เป็นข้อสอบที่ยากเกินไป และข้อสอบที่มีค่า p สูงกว่า ๐.๙๑ เป็นข้อสอบที่ง่ายเกินไป^๔

๑.๒ ความสามารถในการจำแนกผู้สอบตามระดับความสามารถ (item discrimination, r)

ความสามารถในการจำแนกผู้สอบ หมายถึงความสามารถของข้อสอบข้อหนึ่ง ๆ ในการแยกผู้สอบที่ทำคะแนนได้ดี ออกจากผู้สอบที่ทำคะแนนได้ไม่ดี ข้อสอบที่มีความสามารถในการแยกแยะได้ดีนั้นผู้สอบที่ตอบข้อสอบข้อนั้นถูกมักจะได้คะแนนสูง และผู้สอบที่ตอบข้อสอบข้อนั้นผิดมักจะได้คะแนนต่ำ ดัชนีที่ใช้วัดความสามารถในการจำแนกผู้สอบที่ใช้กันมากที่สุดในปัจจุบันคือค่า point-biserial correlation ซึ่งนิยมใช้อักษรย่อเป็น $r^{๐.๔}$ ซึ่งสามารถคำนวณได้จากสูตรต่อไปนี้^๕

$$r = \frac{M_p - M_q}{SD} \sqrt{pq}$$

เมื่อ Mp = คะแนนรวมเฉลี่ยของผู้สอบที่ตอบข้อสอบถูก
 Mq = คะแนนรวมเฉลี่ยของผู้สอบที่ตอบข้อสอบผิด
 SD = ค่าเบี่ยงเบนมาตรฐาน (standard deviation) ของคะแนนสอบ
 p = สัดส่วนของผู้สอบที่ตอบข้อสอบถูกต้องต่อผู้สอบทั้งหมด
 q = สัดส่วนของผู้สอบที่ตอบข้อสอบผิดต่อผู้สอบทั้งหมด

ค่า point-biserial correlation ที่คำนวณได้นี้มีค่าอยู่ในช่วง -1 ถึง 1 โดยค่าที่ติดลบหมายถึง ข้อสอบข้อนั้นผู้ที่ตอบถูกมักสอบได้คะแนนรวมต่ำ แต่ผู้ที่ตอบผิดมักสอบได้คะแนนรวมสูง ในทางตรงข้าม หากค่า point-biserial ยิ่งสูง แสดงถึงข้อสอบที่มีความสามารถในการแยกแยะดี ผู้ที่ตอบข้อสอบข้อนั้นถูกมักทำคะแนนรวมได้สูง ข้อสอบที่ดีควรมีค่า point-biserial สูงกว่า 0.20, ข้อสอบที่พอใช้ได้ควรมีค่า point-biserial อยู่ในช่วง 0.1 - 0.19, ข้อสอบที่มีค่า point-biserial ต่ำกว่า 0.1 เป็นข้อสอบที่ไม่สู้ดีนัก โดยเฉพาะอย่างยิ่งข้อสอบที่มีค่า point-biserial ต่ำกว่า 0 ไม่ควรนำมาคิดคะแนน^{๕๖} (โดยทั่วไปแล้วข้อสอบที่มีค่า point-biserial ติดลบ ให้สงสัยว่าจะเฉลยผิด)

๑.๓ ประสิทธิภาพของตัวลวง (distractor functionality)

ตัวลวงที่มีประสิทธิภาพนั้นมีคุณสมบัติ ๒ ประการคือ^{๕๗}

(๑) มีผู้สอบเลือกตัวลวงนั้นไม่ต่ำกว่าร้อยละ ๕ ของจำนวนผู้สอบทั้งหมด

(๒) มีค่า point-biserial correlation ของตัวลวงนั้นเป็นลบ กล่าวคือตัวลวงที่ดีจะลวงให้ผู้สอบที่มีความรู้ไม่ดี (มีคะแนนต่ำ) มาเลือก แต่ไม่ลวงให้ผู้สอบที่มีความรู้ดี (มีคะแนนสูง) มาเลือก หากตัวลวงใดมีค่า point-biserial correlation เป็นบวก ให้ทบทวนข้อสอบข้อนั้นดูว่าอาจจะเฉลยผิดหรือมีคำตอบที่ถูกต้องมากกว่า ๑ ตัวเลือก

ตัวลวงใดที่มีผู้สอบเลือกน้อย หรือลวงให้ผู้ที่มี

ความรู้ดีมาเลือกจัดเป็นตัวลวงที่ไม่ดี สมควรพิจารณาตัดทิ้งหรือปรับเปลี่ยน

๒. การวิเคราะห์ข้อสอบโดยรวม (test analysis)

การวิเคราะห์ข้อสอบโดยรวมเป็นการพิจารณาว่าเมื่อข้อสอบทั้งหมดทำงานร่วมกันแล้วผลสอบที่ได้ออกมาเป็นอย่างไร มีระดับความยากง่ายเป็นอย่างไร มีการกระจายตัวของคะแนนเป็นอย่างไร มีความน่าเชื่อถือของคะแนนสอบมากน้อยเพียงใด ดัชนีต่าง ๆ ที่ต้องพิจารณาได้แก่

๒.๑ ความเที่ยงตรงของคะแนนสอบ (internal consistency reliability)

การประเมินความเที่ยงตรงของคะแนนสอบเป็นการตรวจสอบว่าคะแนนที่ได้ออกมานั้นมีความน่าเชื่อถือเพียงใด เป็นการตอบคำถามว่าหากนำผู้สอบมาสอบใหม่ในสภาวะการเดิม ด้วยข้อสอบที่มีระดับความยากง่ายเท่าเดิม และผู้สอบมีความรู้เท่าเดิมไม่ได้ไปศึกษาหาความรู้เพิ่มเติม จะได้คะแนนสอบเท่าเดิมหรือไม่^{๕๘}

ดัชนีชี้วัดความเที่ยงตรงของคะแนนสอบที่นิยมใช้ในการรายงานผลสอบด้วยข้อสอบปรนัยคือค่าสัมประสิทธิ์ อัลฟา (Coefficient Alpha) ซึ่งสามารถคำนวณได้จากสูตร^{๕๙}

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_{x_i}^2}{\sigma_x^2} \right)$$

เมื่อ α = สัมประสิทธิ์ อัลฟา (Coefficient Alpha)

n = จำนวนชุดย่อยของข้อสอบที่ทำการแบ่งออกเพื่อหาความเที่ยง

σ_x^2 = การกระจายตัว (variance) ของคะแนนรวม

$\sigma_{x_i}^2$ = การกระจายตัว (variance) ของคะแนนข้อสอบย่อยชุดที่ i

ค่าสัมประสิทธิ์อัลฟานี้มีค่าอยู่ในช่วง 0 - 1 ค่าต่ำแสดงว่าคะแนนที่ได้มีความเชื่อถือได้น้อย ไม่แตกต่างไปจากการเดาสุ่ม ค่าสูงแสดงว่าคะแนนที่ได้มีความน่าเชื่อถือมาก หากทำการทดสอบซ้ำคะแนนที่ได้ก็จะใกล้เคียงเดิม โดยทั่วไประดับของความเที่ยงตรง

ของคะแนนสอบที่ยอมรับได้นั้นขึ้นอยู่กับว่าต้องการนำเอาคะแนนสอบไปใช้ทำอะไร หากการตัดสินผลสอบนั้นมีความสำคัญมาก (high-stakes examination) เช่น การตัดสินผลสอบขอรับใบประกอบวิชาชีพเวชกรรม หรือประกาศนียบัตรแพทย์ผู้เชี่ยวชาญเฉพาะสาขา มักต้องการคะแนนสอบที่มีค่าสัมประสิทธิ์ อัลฟา ไม่ต่ำกว่า ๐.๙ หากการตัดสินผลสอบนั้นมีความสำคัญปานกลาง (medium-stakes examination) เช่นการสอบลงกอง การสอบเลื่อนชั้นเรียน มักต้องการคะแนนสอบที่มีค่าสัมประสิทธิ์ อัลฟา อยู่ในช่วง ๐.๘ - ๐.๘๙ หากการตัดสินผลสอบนั้นมีความสำคัญน้อย (low-stakes examination) เช่นการสอบย่อยในชั้นเรียน การสอบแบบ formative assessment มักต้องการคะแนนสอบที่มีค่าสัมประสิทธิ์ อัลฟา อยู่ในช่วง ๐.๗ - ๐.๗๙^{๑๒}

ประเด็นสำคัญที่ต้องพิจารณาคือเมื่อได้คะแนนสอบที่มีค่าสัมประสิทธิ์ อัลฟาต่ำ จะต้องดำเนินการอย่างไรเพื่อพัฒนาให้การสอบครั้งต่อไปไม่ประสบปัญหาเรื่องความไม่น่าเชื่อถือของคะแนนสอบอีก ปัจจัยหลักที่จะช่วยเพิ่มความเที่ยงตรงของคะแนนสอบปรนัยมี ๓ ปัจจัยด้วยกัน^{๑๓} คือ

(๑) เพิ่มจำนวนข้อสอบให้มากขึ้น ยังมีข้อสอบมากข้อคะแนนที่ได้ก็จะมีค่าสัมประสิทธิ์เพิ่มขึ้น

(๒) ปรับให้ข้อสอบมีการคละกันของข้อสอบที่ยากและง่ายอย่างเหมาะสม เพื่อปรับให้คะแนนมีการกระจายตัวมากขึ้น หากข้อสอบทั้งหมดประกอบไปด้วยข้อสอบที่ง่ายหมด ผู้สอบเกือบทั้งหมดได้คะแนนสูงมาก จะทำให้มีความแตกต่างของคะแนนน้อย โอกาสที่จะแยกแยะผู้สอบที่มีความรู้ดีออกจากผู้ที่มีความรู้ปานกลาง หรือไม่ผู้ดีได้อย่างมั่นใจก็เป็นไปได้น้อย ดังนั้นหากอาจารย์ปรับให้มีการคละกันของข้อสอบยากและง่ายอย่างเหมาะสม ก็จะทำให้ผู้สอบมีระดับคะแนนแตกต่างกันมาก ค่าสัมประสิทธิ์อัลฟาก็จะสูงขึ้นด้วย

(๓) ปรับสภาวะแวดล้อมของการสอบให้เหมาะสม กำจัดสิ่งรบกวนสมาธิของผู้สอบให้มากที่สุด เช่น เสียงรบกวน แสงไฟที่ไม่เพียงพอ หรือไฟที่ติด ๆ ดับ ๆ เป็นต้น

๒.๒ การกระจายตัวของคะแนน และคะแนน

เฉลี่ย (standard deviation and mean score)

การตรวจดูลักษณะพื้นฐานของคะแนนสอบนี้จะช่วยบอกได้คร่าว ๆ ว่าการเรียนการสอนมีประสิทธิภาพเพียงใด หากอาจารย์สอนได้ดี นักเรียนทั้งชั้นเรียนเข้าใจเนื้อหาดี คะแนนสอบที่ได้ออกมา ก็ควรจะกระจายตัวมากนัก (คะแนนเกาะกลุ่มกัน) และคะแนนเฉลี่ยก็ควรจะค่อนข้างสูงเมื่อเทียบกับนักเรียนรุ่นอื่น ๆ หากคะแนนสอบของนักเรียนมีการกระจายตัวมากเกินไป แสดงว่าอาจมีปัญหาบางประการในการเรียนการสอนทำให้นักเรียนบางคนมีความรู้ความเข้าใจดี แต่มีนักเรียนบางกลุ่มที่ไม่ค่อยรู้เรื่อง^{๑๔}

๒.๓ ค่าความยากง่ายเฉลี่ยของข้อสอบ (average difficulty)

จากการวิเคราะห์ข้อสอบรายข้อ เราได้ค่าความยากง่ายของข้อสอบแต่ละข้อ (p) เมื่อนำค่า p ของข้อสอบทุกข้อมาหาค่าเฉลี่ย เราก็จะได้ค่าความยากง่ายของข้อสอบทั้งหมด ค่าที่ได้มานี้ใช้เป็นดัชนีชี้วัดว่าข้อสอบทั้งหมดโดยรวมแล้วมีระดับความยากง่ายเป็นอย่างไร หากผู้สอบเป็นนักศึกษาในกลุ่มใหญ่พอที่เราจะตั้งสมมติฐานว่าระดับความสามารถมีการกระจายตัวอย่างเหมาะสมและไม่ต่างจากระดับความสามารถเฉลี่ยของกลุ่มผู้สอบปีก่อน ๆ เราก็สามารถนำค่าความยากง่ายของข้อสอบทั้งหมดนี้มาเทียบได้ว่าข้อสอบที่นำมาใช้ในปีนี้อาจง่ายกว่าข้อสอบปีก่อน ๆ ซึ่งอาจารย์อาจนำข้อมูลนี้มาใช้พิจารณาปรับเกณฑ์การตัดเกรดด้วยว่าต้องมีการปรับระดับคะแนนที่ได้เกรดต่าง ๆ หรือไม่ อย่างไร

๒.๔ ค่าความสามารถในการแยกแยะผู้สอบเฉลี่ย (average discrimination)

การนำค่า point-biserial correlation ของข้อสอบทั้งหมดมาหาค่าเฉลี่ย เป็นการบอกคร่าว ๆ ว่าโดยรวมแล้วข้อสอบชุดนี้มีความสามารถในการแยกแยะผู้สอบตามระดับความสามารถเพียงใด ยิ่งได้ค่าสูงก็ยิ่งดี แต่มีข้อควรระวังในการแปลผลในกรณีที่มีการเรียนการสอนเป็นไปได้ดี และผู้สอบทั้งหมด หรือเกือบทั้งหมดทำคะแนนได้สูง ค่า point-biserial correlation เฉลี่ยของข้อสอบทั้งหมดจะไม่สูงแต่ไม่ได้แปลว่าข้อสอบที่ใช้มีคุณภาพไม่ดี^{๑๕}

เวบบันทึกศิริราช

บทความทั่วไป

การนำผลการวิเคราะห์ข้อสอบไปใช้

ผลการวิเคราะห์ข้อสอบด้วยดัชนีชี้วัดต่าง ๆ ดังกล่าวข้างต้นสามารถนำไปใช้ประโยชน์ได้หลายประการ เช่น

๑. ใช้เป็นประโยชน์ในการปรับแก้คะแนนสอบ

จากผลการวิเคราะห์ข้อสอบจะช่วยชี้แนะให้เราทราบว่าข้อสอบข้อใดน่าจะมีเฉลยผิด ข้อสอบข้อใดน่าจะมีคำตอบที่ถูกมากกว่า ๑ ตัวเลือก ข้อสอบข้อใดน่าจะมีปัญหา เช่น มีความคลุมเครือในคำถาม หรือตัวเลือกมีความซ้ำซ้อนกัน หรือเนื้อหาของข้อสอบอยู่นอกเหนือไปจากสิ่งที่สอนนักเรียน เป็นต้น ข้อสอบที่มีปัญหาเหล่านี้ต้องได้รับการประเมินโดยคณะกรรมการตรวจข้อสอบซึ่งประกอบไปด้วยอาจารย์ผู้มีความรู้ความชำนาญในเนื้อหาวิชาที่ทำการสอบว่าจะดำเนินการอย่างไรกับการคิดคะแนน หากปัญหาที่พบมีความรุนแรงไม่มากจนทำให้การตัดสินใจเลือกคำตอบที่ถูกต้องเปลี่ยนไป คณะกรรมการอาจพิจารณาคิดคะแนนของข้อสอบข้อนั้นตามปกติ หากข้อสอบเฉลยผิดคณะกรรมการสามารถพิจารณาแก้คำตอบแล้วทำการตรวจให้คะแนนข้อสอบข้อนั้นใหม่ หากข้อสอบข้อใดมีคำตอบที่เหมาะสม ๒ ข้อ คณะกรรมการอาจพิจารณาให้ผู้สอบที่ตอบข้อใดข้อหนึ่งใน ๒ ข้อดังกล่าวได้คะแนนในข้อนั้น หากข้อสอบนั้นมีความคลุมเครือมากจนไม่สามารถตัดสินใจเลือกคำตอบที่เหมาะสมได้ คณะกรรมการสามารถตัดข้อสอบข้อนั้นออกจากการคิดคะแนน และปรับคะแนนเกณฑ์ผ่านลดลงตามความเหมาะสม

๒. ใช้เป็นประโยชน์ในการปรับปรุงคุณภาพข้อสอบ

ภายหลังจากการรายงานคะแนนสอบเป็นที่เรียบร้อยแล้ว คณะกรรมการสอบสามารถนำผลการวิเคราะห์ข้อสอบแต่ละข้อมาพิจารณาโดยละเอียดเพื่อดูว่าข้อสอบข้อใดสมควรได้รับการปรับปรุงแก้ไข ข้อสอบที่พบว่ายากเกินไปอาจเกิดจากโจทย์คำถามมีความคลุมเครือ ต้องทำการปรับแก้ให้โจทย์ชัดเจนขึ้น หรือเพิ่มเติมข้อมูลบางประการเข้าไปเพื่อให้การวินิจฉัย

ชัดเจนขึ้น ข้อสอบที่พบว่าง่ายเกินไปอาจพิจารณาปรับให้ยากขึ้นโดยการแก้โจทย์หรือตัวเลือก ข้อสอบที่มีค่า point-biserial ต่ำมักเกิดจากโจทย์ที่คลุมเครือ สร้างความสับสนให้ผู้สอบ สมควรได้รับการปรับโจทย์คำถามใหม่

นอกจากนี้อาจารย์ยังต้องพิจารณาถึงการทำงานของตัวเลือกด้วย ปัญหาที่พบบ่อยมากในการวิเคราะห์ข้อสอบปรนัยคือมีตัวลวงจำนวนมากที่ไม่ทำงาน (มีผู้สอบเลือกน้อยมาก หรือลวงเฉพาะผู้ที่มีความรู้ดีให้มาเลือก) จากการศึกษาวิจัยข้อสอบปรนัยจำนวนมากพบว่าข้อสอบส่วนใหญ่มักมีตัวเลือกที่ทำงานจริงเพียง ๓ ตัวเลือกเท่านั้น^๖ ตัวเลือกที่เหลือเป็นตัวเลือกที่ไม่มีประโยชน์ พิมพ์ลงมาในข้อสอบก็เป็นการเปลืองเนื้อที่หน้ากระดาษ และเสียเวลาอ่านโดยใช้เหตุ อาจารย์ควรพิจารณาตัดตัวลวงที่ไม่ทำงานออกเสียหรือเปลี่ยนเป็นตัวลวงอื่นที่น่าจะมีประสิทธิภาพมากขึ้น

๓. ใช้เป็นประโยชน์ในการบริหารคลังข้อสอบ

ข้อสอบแต่ละข้อนั้นได้มาด้วยความยากลำบาก อาจารย์แต่ละท่านต้องใช้เวลาและความคิดอย่างมากเพื่อพัฒนาข้อสอบที่ดีขึ้นมาใช้ ดังนั้นเมื่อนำข้อสอบมาใช้แล้วผลการวิเคราะห์ข้อสอบแสดงว่าข้อสอบข้อใดเป็นข้อสอบที่ดี มีระดับความยากง่ายเหมาะสม มีความสามารถในการจำแนกผู้สอบที่ดีก็ควรพิจารณาเลือกเก็บข้อสอบดังกล่าวไว้ในคลังข้อสอบเพื่อที่จะได้นำกลับมาใช้ใหม่ในอนาคต ในการเก็บข้อสอบเข้าในคลังข้อสอบก็ต้องมีการแนบข้อมูลเกี่ยวกับประวัติการใช้งานและผลการวิเคราะห์ข้อสอบในแต่ละครั้งไว้คู่กันด้วย เพื่อที่จะได้เป็นประโยชน์ในการเลือกข้อสอบมาใช้งาน หากอาจารย์ต้องการข้อสอบที่มีระดับความยากง่าย หรือความสามารถในการจำแนกผู้สอบมากน้อยเพียงใดจะได้ดึงเอาข้อสอบที่มีคุณลักษณะตามต้องการออกมาใช้ได้ตามต้องการ

๔. ใช้เป็นประโยชน์ในการพัฒนาคุณภาพการสอน

การพิจารณาผลการวิเคราะห์ข้อสอบโดยละเอียดในหัวข้อที่อาจารย์ท่านใดท่านหนึ่งรับผิดชอบ

เวบบันทึทศึรึรึรึ

บทความทึวไป

ในการสอนนักเรียนหรือแพทย์ประจำบ้านอยู่นั้นจะทำให้ได้ข้อมูลที่เป็นประโยชน์ในการพัฒนาการเรียนการสอนได้ กล่าวคืออาจารย์สามารถตรวจสอบดูได้ว่านักเรียนหรือแพทย์ประจำบ้านมีความเข้าใจที่ถูกต้องในเรื่องดังกล่าวหรือไม่ ประเด็นใดที่มีผู้เข้าใจผิดอยู่มากก็สมควรที่อาจารย์จะทำการเน้นย้ำในบรรดาการเรียนหรือแพทย์ประจำบ้านในการสอนครั้งต่อ ๆ ไปเพื่อแก้ไขความเข้าใจผิดดังกล่าว ประเด็นใดที่นักเรียนหรือแพทย์ประจำบ้านมีความเข้าใจดีมากอยู่แล้ว อาจารย์อาจไม่ต้องใช้เวลามากนักในการสอนเรื่องดังกล่าว แต่เอาเวลามาใช้สอนในเรื่องที่นักเรียนหรือแพทย์ประจำบ้านยังไม่ค่อยเข้าใจให้มากขึ้นได้

ข้อจำกัดของการวิเคราะห์ข้อสอบ

ถึงแม้ว่าการวิเคราะห์ข้อสอบด้วยวิธีการที่ได้อธิบายมาข้างต้นจะให้ข้อมูลที่เป็นประโยชน์หลายอย่างด้วยกัน แต่เนื่องจากวิธีการวิเคราะห์เหล่านี้เป็นเทคนิคที่วางรากฐานอยู่บนทฤษฎีการสอบแบบดั้งเดิม (classical test theory) ซึ่งมีข้อจำกัดหลายประการด้วยกัน ในการนำค่าต่าง ๆ ที่ได้จากการวิเคราะห์ข้อสอบไปใช้นั้น อาจารย์ควรคำนึงถึงข้อจำกัดของผลการวิเคราะห์ด้วย ในที่นี้จะกล่าวถึงเฉพาะข้อจำกัดในการแปลผลการวิเคราะห์ขั้นพื้นฐานเท่านั้นเนื่องจากเป็นการแปลผลที่ใช้กันทั่วไปในวงการแพทยศาสตรศึกษา ข้อจำกัดในการนำผลการวิเคราะห์ไปประยุกต์ในงานวิจัยทางจิตวิทยาการศึกษายังมีอีกหลายประการที่ผู้นิพนธ์ขอไม่นำมากล่าวในที่นี้ เนื่องจากมีความซับซ้อนและไม่มีที่ใช้ในวงการแพทยศาสตรศึกษาในประเทศไทยในปัจจุบัน

พื้นฐานสำคัญที่เป็นข้อจำกัดของผลการวิเคราะห์ข้อสอบด้วยทฤษฎีการสอบแบบดั้งเดิมคือค่าต่าง ๆ ที่ได้มาจากการวิเคราะห์นั้นขึ้นอยู่กับกลุ่มตัวอย่างที่ใช้ในการเก็บข้อมูล^{๓๓,๓๔} หากได้ข้อมูลมาจากกลุ่มตัวอย่างที่มีขนาดใหญ่พอและมีการกระจายตัวของระดับความสามารถของผู้สอบที่เหมาะสม ค่าต่าง ๆ ที่ได้ (p, r, coefficient alpha) จะค่อนข้างเที่ยงตรง ปัญหาที่สำคัญในการวิเคราะห์ข้อสอบในโรงเรียนแพทย์คือการสอบจำนวนมากจัดในนักศึกษาในกลุ่มเล็ก และ

นักศึกษาแต่ละกลุ่มก็มีการกระจายตัวของระดับความสามารถแตกต่างกัน นักศึกษาบางกลุ่มมีความสามารถสูงกว่านักศึกษากลุ่มอื่น ดังนั้นผลการวิเคราะห์ข้อสอบไม่ว่าจะเป็นค่า p, r, coefficient alpha, mean, หรือ standard deviation อาจจะไม่เปลี่ยนแปลงไปในแต่ละกลุ่มของนักศึกษา ดังนั้นการนำผลการวิเคราะห์ข้อสอบไปใช้ในทางปฏิบัติจึงมีข้อควรระวังดังต่อไปนี้

การพิจารณาว่าข้อสอบยากหรือง่ายโดยใช้ค่า p นั้นเป็นค่าที่ไม่คงที่ ขึ้นอยู่กับกลุ่มผู้สอบ หากนำข้อสอบข้อหนึ่งไปไปใช้กับนักเรียนกลุ่มที่มีความรู้ดี นักเรียนส่วนใหญ่จะทำข้อสอบได้ถูกต้องทำให้ค่า p สูง แต่เมื่อนำข้อสอบข้อเดิมไปใช้กับนักเรียนกลุ่มที่ความรู้ไม่ดีนัก สัดส่วนของนักเรียนที่ทำข้อสอบข้อเดียวกันได้ถูกต้องจะลดลงทำให้ค่า p ลดลง นอกจากนี้ในข้อสอบที่เน้นการท่องจำที่เคยใช้แล้ว เมื่อนำกลับมาใช้ใหม่ในนักเรียนกลุ่มใหม่ อาจมีนักเรียนจำนวนหนึ่งที่สามารถตอบข้อสอบถูกต้องเนื่องจากรู้ข้อสอบมาก่อนก็จะทำให้ค่า p สูงขึ้นกว่าเดิมได้

การพิจารณาว่าข้อสอบมีความสามารถในการแยกแยะผู้สอบได้ดีเพียงใดโดยใช้ค่า r ก็ประสบปัญหาในลักษณะเดียวกัน กล่าวคือค่า r นั้นขึ้นกับกลุ่มตัวอย่างของผู้สอบ หากกลุ่มผู้สอบมีระดับความรู้ที่ใกล้เคียงกัน มีคะแนนค่อนข้างเกาะกลุ่มกัน เมื่อคิดค่า r ก็จะได้ต่ำ แต่หากใช้ข้อสอบข้อเดิมในกลุ่มผู้สอบที่มาจากหลายสถาบัน มีความแตกต่างกันของระดับความรู้อย่างมาก ก็จะได้ค่า r สูง

ค่าสัมประสิทธิ์อัลฟา เป็นค่าที่มีความเฉพาะเจาะจงกับการสอบของนักเรียนกลุ่มใดกลุ่มหนึ่งเท่านั้น หากใช้เป็นคุณสมบัติติดตัวข้อสอบแต่ละข้อไม่ หากข้อสอบชุดหนึ่งทำการสอบกับนักเรียนกลุ่มหนึ่งแล้วพบว่าคะแนนสอบที่ได้มานั้นมีค่าสัมประสิทธิ์อัลฟาสูงในระดับที่ต้องการก็ไม่ได้เป็นตัวรับประกันว่าหากนำข้อสอบชุดเดิมนั้นไปทำการสอบกับนักเรียนกลุ่มอื่นจะได้ค่าสัมประสิทธิ์อัลฟาที่สูงเช่นเดียวกัน นอกจากนี้ค่าสัมประสิทธิ์อัลฟาที่สูงไม่ได้เป็นตัวบอกถึงคุณภาพของข้อสอบรายข้อแต่อย่างใด

ค่าสัมประสิทธิ์อัลฟาที่สูงช่วยบอกแค่เพียงว่า

คะแนนสอบในข้อสอบข้อหนึ่งมีความผันแปรไปในทิศทางเดียวกันกับคะแนนสอบในข้อสอบข้ออื่นในการสอบชุดเดียวกัน นั่นคือในข้อสอบชุดที่มีค่าสัมประสิทธิ์อัลฟ่าสูงก็อาจประกอบไปด้วยข้อสอบที่ดี และข้อสอบที่ไม่ดีรวมกันอยู่ ต้องไปตรวจสอบดัชนีชี้วัดคุณภาพของข้อสอบตัวอื่น ๆ ในแต่ละข้ออีกครั้ง

ข้อควรจำในการวิเคราะห์ข้อสอบที่ผู้นิพนธ์ขอ ย้ำในตอนท้ายของบทความนี้ก็คือค่าดัชนีชี้วัดคุณภาพต่าง ๆ ของข้อสอบที่กล่าวมาทั้งหมดนี้เป็นเพียงตัวช่วยให้อาจารย์เข้าใจข้อสอบดีขึ้นและช่วยแนะแนวทางในการพัฒนาปรับปรุงข้อสอบให้ดีขึ้น ดัชนีเหล่านี้ไม่ใช่ค่าตัดสินหรือตัวชี้ชะตาของข้อสอบ ไม่มีดัชนีใดที่ได้จากการวิเคราะห์ข้อสอบจะมาทดแทนดุลยพินิจของอาจารย์ไปได้ ดัชนีคุณภาพของข้อสอบไม่ว่าจะคำนวณมาด้วยวิธีการที่ถูกต้องแล้วก็ตามก็เป็นเพียงตัวเลขที่สามารถเกิดความผิดพลาดในการแปลผลได้ดังเช่นการแปลผลการวิเคราะห์ทางสถิติต่าง ๆ บทบาทของอาจารย์ในการวิเคราะห์ข้อสอบคงไม่ใช่การยึดถือตัวเลขดัชนีต่าง ๆ เป็นกฎตายตัว หากแต่ใช้ดัชนีเหล่านี้ช่วยเป็นแนวทางในการพิจารณาข้อสอบ หากดัชนีตัวใดระบุว่าข้อสอบอาจมีปัญหา อาจารย์ก็นำข้อสอบนั้นมาพิจารณากัน โดยคณะกรรมการข้อสอบ หากหลังจากการพิจารณาโดยถี่ถ้วนแล้วอาจารย์คิดว่าข้อสอบข้อนั้นเหมาะสมแล้ว ไม่ควรทำการปรับแก้เนื้อหา อาจารย์ก็ยืนยันไปว่าไม่แก้ไข อาจารย์คงไม่ตัดสินการรักษาผู้ป่วยโดยใช้ผลเลือดตัวใดตัวหนึ่งเป็นเกณฑ์โดยไม่พิจารณาอาการและอาการแสดงของผู้ป่วยร่วมด้วย ฉะนั้นดัชนีนั้น อาจารย์

ไม่ควรตัดสินชะตากรรมของข้อสอบโดยใช้เพียงค่า p หรือ r โดยไม่พิจารณาความเหมาะสมของเนื้อหาโจทย์ และตัวเลือกต่าง ๆ ในข้อสอบข้อนั้น

เอกสารอ้างอิง

๑. Livingston SA. Item analysis. In: Downing SM, Haladyna TM, eds. Handbook of test development. Mahwah, NJ: Lawrence Erlbaum Associates; 2006:421-41.
๒. Brown W, Thomson GH. The essentials of mental measurement, 2nd ed. Cambridge, England: University Press; 1921.
๓. Yen WM, Fitzpatrick AR. Item response theory. In: Brennan RL, ed. Educational measurement, 4th ed. Westport, CT: Praeger Publishers; 2006:111-53.
๔. Haladyna TM. Writing test items to evaluate higher order thinking. Boston, MA: Allyn and Bacon; 1997.
๕. Haladyna TM. Writing multiple choice items. Chicago, IL: CAT Inc.; 2003.
๖. Haladyna TM. Developing and validating multiple-choice test items, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2004.
๗. Aleamoni LM, Spencer RE. A comparison of biserial discrimination, point biserial discrimination, and difficulty indices in item analysis data. Educ Psychol Meas 1969;29:353-8.
๘. Haladyna TM, Downing SM. How many options is enough for a multiple-choice test item? Educ Psychol Meas 1993;53:999-1010.
๙. Gronlund NE. Assessment of student achievement, 7th ed. Boston: Allyn & Bacon, 2003.
๑๐. Linn RL, Miller MD. Measurement and assessment in teaching, 9th ed. Upper Saddle River, NJ: Prentice Hall, 2004.
๑๑. Haertel EH. Reliability. In: Brennan RL, editor. Educational measurement, 4th ed. Westport, CT: Praeger Publishers; 2006:65-110.
๑๒. Downing SM. Reliability: On the reproducibility of assessment data. Med Educ 2004;38:1006-12.
๑๓. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
๑๔. Smith EV. Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In: Smith EV, Smith RM, eds. Introduction to Rasch measurement: Theory, models, and applications. Maple Grove, MN: JAM Press, 2004:93-112



โปรแกรมวิเคราะห์ข้อสอบ

รุ่น 2.0

การสอบ : SIID 521 (Basic Sciences)

วันที่ : 22 ธันวาคม 2555

จำนวนข้อสอบ = 120

จำนวนผู้เข้าสอบ = 244

Difficulty Index --> p-value (proportion of students answer item correctly)

$$p\text{-Value} = \frac{\text{number of students answer correctly}}{\text{total number of students answer that item}}$$

Discrimination Index --> D or r-value --> Point-biserial correlation coefficient (r_{pb_i})

=====

SCORE STATISTICS

Mean = **68.152** S.D. = **11.915**

Mode = **65** (freq = **14**)

Max = **94** Min = **28**

DIFFICULTY INDEX (p value)

Average (p-bar) = **0.566** Max p = **0.990** Min p = **0.010**

DISCRIMINATION INDEX (D or r value)

Average (D-bar) = **0.244** Max D = **0.680** Min D = **-0.180**

RELIABILITY COEFFICIENT (rtt) = **0.847**
(Kuder-Richardson formula 20)

STANDARD ERROR OF MEASUREMENT (SEM) = **4.655**
(S.D. x $\sqrt{1-rtt}$)

Item Analysis and Option Analysis

Faculty of Medicine Siriraj Hospital

Mahidol University

No. : 1									
p Value : 0.55					r _{pbi} : 0.37				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.24	21.31	-0.10	13.52	0.37	54.92	-0.16	6.15	-0.07	4.10

No. : 2									
p Value : 0.74					r _{pbi} : 0.00				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.02	5.33	0.07	11.48	-0.02	1.23	0.00	74.18	-0.09	7.79

No. : 3									
p Value : 0.84					r _{pbi} : 0.25				
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.22	14.34	0.25	84.43	0.01	0.41	0.00	0.00	-0.12	0.41

No. : 4									
p Value : 0.68					r _{pbi} : 0.43				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.26	8.20	-0.09	8.20	0.43	68.03	-0.06	1.64	-0.29	13.93

No. : 5									
p Value : 0.92					r _{pbi} : 0.26				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.16	4.10	-0.07	0.41	0.26	91.80	-0.16	2.87	-0.08	0.82

No. : 6									
p Value : 0.75					r _{pbi} : 0.30				
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.30	74.59	-0.03	13.93	-0.22	2.87	-0.24	3.69	-0.17	4.92

No. : 7									
p Value : 0.99					r _{pbi} : 0.06				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.06	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.06	99.18

No. : 8									
p Value : 0.70					r _{pbi} : 0.53				
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.53	70.49	-0.13	1.23	-0.21	5.74	-0.38	17.21	-0.17	5.33

No. : 9									
p Value : 0.63					r _{pbi} : 0.19				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	0.41	0.00	0.00	0.01	2.05	-0.19	34.43	0.19	63.11

No. : 10									
p Value : 0.90					r _{pbi} : 0.25				
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.25	90.16	-0.09	0.41	-0.22	9.02	-0.08	0.41	0.00	0.00

No. : 11									
p Value : 0.54					r _{pbi} : 0.48				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.44	31.97	-0.09	4.51	-0.05	8.61	0.48	53.69	-0.06	1.23

No. : 12									
p Value : 0.55					r _{pbi} : 0.47				
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.27	28.28	0.47	54.92	0.00	0.00	-0.24	11.07	-0.16	5.74

No. : 13									
p Value : 0.81					r _{pbi} : 0.32				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.23	5.33	-0.16	9.84	0.32	81.15	-0.13	3.28	-0.06	0.41

No. : 14									
p Value : 0.45					r _{pbi} : 0.39				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.22	34.84	-0.09	1.64	-0.17	11.89	-0.08	6.15	0.39	45.49

No. : 15									
p Value : 0.73					r _{pbi} : 0.32				
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.24	2.46	0.32	72.95	-0.17	2.05	-0.17	21.72	-0.07	0.41

No. : 16									
p Value : 0.09					r _{pbi} : -0.03				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.14	11.89	0.15	70.08	-0.18	3.28	0.08	5.74	-0.03	8.61

No. : 17									
p Value : 0.36					r _{pbi} : 0.13				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.05	4.10	0.06	22.13	0.13	35.66	-0.07	9.43	-0.12	28.69

No. : 18									
p Value : 0.83					r _{pbi} : 0.06				
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.06	82.79	0.01	0.82	-0.05	2.05	-0.10	4.92	0.01	9.43

Item Analysis and Option Analysis

Faculty of Medicine Siriraj Hospital
Mahidol University

No. : 19		p Value : 0.25				r _{pbi} : 0.04			
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.10	51.23	0.04	13.11	0.00	0.00	0.04	24.59	0.05	11.07

No. : 20		p Value : 0.36				r _{pbi} : 0.55			
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.21	22.54	0.55	35.66	-0.12	2.46	-0.25	34.43	-0.19	4.92

No. : 21		p Value : 0.81				r _{pbi} : 0.20			
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.20	80.74	-0.07	3.69	-0.13	11.89	-0.05	1.64	-0.11	2.05

No. : 22		p Value : 0.46				r _{pbi} : 0.47			
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.47	45.90	-0.14	6.15	-0.11	4.92	-0.18	17.21	-0.24	25.82

No. : 23		p Value : 0.00				r _{pbi} : -0.06			
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.03	0.41	0.00	0.41	-0.06	0.41	-0.14	4.10	0.16	94.26

No. : 24		p Value : 0.64				r _{pbi} : 0.40			
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.08	5.33	-0.16	9.43	0.40	64.34	-0.20	9.02	-0.21	11.89

No. : 25		p Value : 0.61				r _{pbi} : 0.40			
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.15	2.87	-0.10	13.11	-0.23	14.34	0.40	60.66	-0.19	9.02

No. : 26		p Value : 0.70				r _{pbi} : 0.47			
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.15	7.38	-0.22	9.84	-0.26	7.79	-0.18	5.33	0.47	69.67

No. : 27		p Value : 0.51				r _{pbi} : 0.35			
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.15	9.02	0.35	50.82	-0.26	25.82	-0.05	5.33	-0.02	9.02

No. : 28		p Value : 0.50				r _{pbi} : 0.17			
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.17	49.59	-0.17	20.49	-0.03	4.51	-0.04	15.98	0.01	9.43

No. : 29		p Value : 0.75				r _{pbi} : 0.17			
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.09	14.34	-0.16	3.28	-0.01	2.87	-0.06	4.92	0.17	74.59

No. : 30		p Value : 0.58				r _{pbi} : 0.37			
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.22	6.15	-0.30	31.15	0.37	57.79	0.05	4.92	0.00	0.00

No. : 31		p Value : 0.86				r _{pbi} : 0.28			
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.28	86.07	-0.05	2.05	-0.21	9.43	-0.10	1.23	-0.17	1.23

No. : 32		p Value : 0.88				r _{pbi} : 0.32			
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.30	8.20	-0.16	2.87	0.32	87.70	0.03	1.23	0.00	0.00

No. : 33		p Value : 0.44				r _{pbi} : 0.37			
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.09	4.92	0.37	44.26	-0.41	45.08	0.01	2.46	-0.03	3.28

No. : 34		p Value : 0.73				r _{pbi} : 0.25			
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.25	72.54	-0.22	9.02	-0.15	6.15	-0.05	1.23	-0.02	11.07

No. : 35		p Value : 0.45				r _{pbi} : 0.42			
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.06	9.02	-0.18	12.30	-0.38	18.44	-0.06	15.16	0.42	45.08

No. : 36		p Value : 0.68				r _{pbi} : 0.35			
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.15	4.51	-0.29	16.39	0.35	68.03	-0.04	6.97	-0.07	4.10

Item Analysis and Option Analysis

Faculty of Medicine Siriraj Hospital
Mahidol University

No. : 37 p Value : 0.29 r _{pbi} : -0.02									
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.05	2.05	0.22	52.05	-0.14	7.38	-0.20	9.84	-0.02	28.69

No. : 38 p Value : 0.75 r _{pbi} : 0.11									
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.11	74.59	-0.11	22.95	-0.14	0.82	0.08	0.82	0.08	0.82

No. : 39 p Value : 0.51 r _{pbi} : 0.23									
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.02	10.25	-0.21	27.46	0.23	51.23	-0.07	9.02	0.09	1.64

No. : 40 p Value : 0.21 r _{pbi} : 0.13									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	40.57	0.13	20.90	0.00	4.51	0.07	17.62	-0.21	16.39

No. : 41 p Value : 0.42 r _{pbi} : -0.03									
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.02	7.38	0.07	43.03	-0.02	0.41	-0.03	41.80	-0.10	7.38

No. : 42 p Value : 0.79 r _{pbi} : 0.33									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.22	5.33	0.33	79.10	-0.20	4.92	-0.02	2.87	-0.15	7.79

No. : 43 p Value : 0.81 r _{pbi} : 0.37									
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.37	80.74	-0.33	14.75	0.01	0.82	-0.14	2.05	-0.07	1.64

No. : 44 p Value : 0.56 r _{pbi} : 0.34									
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.14	1.64	-0.18	6.56	0.34	55.74	-0.22	20.08	-0.05	15.98

No. : 45 p Value : 0.86 r _{pbi} : 0.39									
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.16	2.05	-0.11	0.82	-0.04	1.23	-0.33	9.84	0.39	86.07

No. : 46 p Value : 0.81 r _{pbi} : 0.31									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.19	10.66	0.31	80.74	-0.09	2.87	-0.15	1.64	-0.15	4.10

No. : 47 p Value : 0.93 r _{pbi} : 0.26									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.14	2.46	0.26	93.44	-0.01	0.82	-0.17	1.64	-0.15	1.64

No. : 48 p Value : 0.07 r _{pbi} : -0.20									
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.20	12.70	-0.08	4.51	-0.18	2.87	-0.20	6.56	0.37	73.36

No. : 49 p Value : 0.95 r _{pbi} : 0.21									
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	0.00	0.00	0.00	-0.21	4.92	0.21	95.08	0.00	0.00

No. : 50 p Value : 0.83 r _{pbi} : 0.24									
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	0.00	0.00	0.00	0.24	83.20	-0.23	15.98	-0.09	0.82

No. : 51 p Value : 0.76 r _{pbi} : 0.26									
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.26	76.23	-0.14	2.87	-0.04	2.46	0.07	0.41	-0.23	18.03

No. : 52 p Value : 0.70 r _{pbi} : 0.24									
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.15	0.82	-0.21	11.89	0.01	12.70	0.25	70.08	-0.16	4.51

No. : 53 p Value : 0.51 r _{pbi} : 0.31									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.02	4.51	0.31	50.82	-0.07	2.05	-0.07	2.87	-0.28	39.75

No. : 54 p Value : 0.37 r _{pbi} : 0.28									
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.07	9.43	0.28	36.89	-0.19	13.52	-0.09	16.80	-0.04	23.36

Item Analysis and Option Analysis

Faculty of Medicine Siriraj Hospital

Mahidol University

No. : 55									
p Value : 0.71					r _{pbi} : 0.25				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.18	2.87	-0.20	14.75	-0.08	5.74	0.25	70.90	0.01	5.74

No. : 56									
p Value : 0.81					r _{pbi} : 0.29				
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.02	1.23	0.29	81.15	-0.15	7.38	-0.10	4.92	-0.22	5.33

No. : 57									
p Value : 0.26					r _{pbi} : 0.19				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.08	6.15	-0.17	29.51	-0.01	15.57	0.19	26.23	0.03	22.54

No. : 58									
p Value : 0.66					r _{pbi} : 0.29				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.16	25.00	-0.14	2.46	-0.22	0.41	0.29	65.98	-0.14	6.15

No. : 59									
p Value : 0.73					r _{pbi} : 0.36				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.13	0.82	-0.25	19.67	-0.26	5.33	0.36	73.36	0.10	0.82

No. : 60									
p Value : 0.93					r _{pbi} : 0.28				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	0.00	-0.13	4.10	-0.27	2.87	-0.03	0.41	0.28	92.62

No. : 61									
p Value : 0.89					r _{pbi} : 0.26				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.05	0.41	-0.30	2.46	-0.13	5.74	-0.06	2.46	0.26	88.93

No. : 62									
p Value : 0.89					r _{pbi} : 0.38				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.32	7.38	-0.09	0.82	-0.17	3.28	0.38	88.52	0.00	0.00

No. : 63									
p Value : 0.69					r _{pbi} : 0.05				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.00	0.00	-0.12	1.64	-0.02	29.51	0.05	68.85	0.00	0.00

No. : 64									
p Value : 0.81					r _{pbi} : 0.20				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.09	0.82	0.05	2.46	0.20	80.74	-0.16	11.89	-0.10	3.69

No. : 65									
p Value : 0.68					r _{pbi} : 0.10				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.06	9.43	-0.15	1.64	0.10	68.44	-0.04	1.23	-0.01	19.26

No. : 66									
p Value : 0.55					r _{pbi} : 0.32				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.22	23.36	-0.08	11.48	0.32	54.92	-0.11	6.15	-0.07	4.10

No. : 67									
p Value : 0.45					r _{pbi} : 0.29				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.20	26.64	-0.07	17.62	-0.05	1.23	0.29	45.49	-0.06	8.61

No. : 68									
p Value : 0.28					r _{pbi} : -0.03				
A		B		* C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.02	14.34	0.07	1.64	-0.03	27.87	0.06	10.25	-0.04	45.90

No. : 69									
p Value : 0.39					r _{pbi} : 0.37				
A		B		C		* D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.05	23.77	-0.07	13.93	-0.22	0.41	0.37	38.93	-0.28	22.95

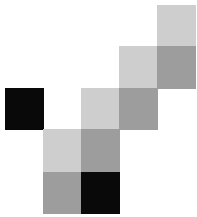
No. : 70									
p Value : 0.25					r _{pbi} : 0.13				
A		* B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.02	7.79	0.13	24.59	-0.10	1.64	0.06	10.66	-0.10	54.92

No. : 71									
p Value : 0.80					r _{pbi} : 0.09				
* A		B		C		D		E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
0.09	80.33	-0.03	1.64	-0.13	3.28	0.00	5.74	-0.03	9.02

No. : 72									
p Value : 0.65					r _{pbi} : 0.37				
A		B		C		D		* E	
r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%	r _{pbi}	%
-0.25	6.97	-0.05	6.56	-0.23	20.08	-0.05	1.23	0.37	65.16

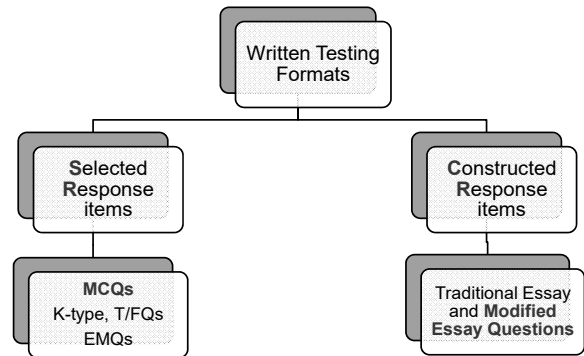
ผศ. นพ. สุประพัฒน์ สนใจพานิชย์

หัวข้อ : Constructed response item exam (Review part 1)



Constructed Response Items

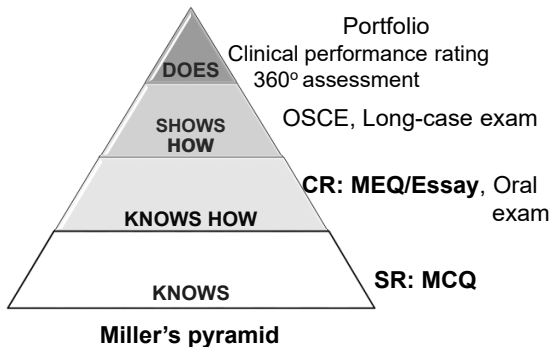
Suprath Sonjaipanich MD.
Department of Pediatrics
Faculty of Medicine Siriraj Hospital



Downing S.M. & Yudkowsky R. Written Tests: Constructed-Response and Selected-Response Formats. Assessment in Health Professions Education 2009



Assessment approach



Modified Essay Questions

Strengths

- Able to measure higher-order cognitive abilities
- Uncued written responses
- Mimic actual clinical problem solving
- Motivation for clinical learning

Modified Essay Questions

Limitations

- Construct underrepresentation
- Difficult to develop and score
- Unexpected responses
- Subjective scoring
- Low reliability

Modified Essay Questions

- Standard modified essay questions
- Key features problem (KFP)
- Patient management problem (PMP)
- Short answer questions (SAQ)

Outline

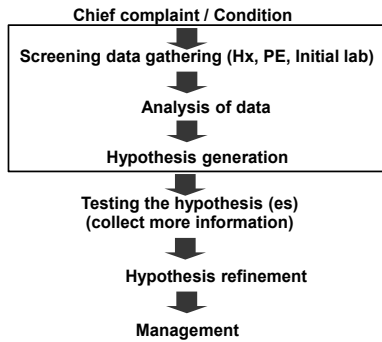
- Clinical problem solving methods and MEQ process
- Key Features Problem
- Developing an MEQ item

Clinical Problem Solving Methods

1. Pattern recognition
2. Algorithm
3. Forward reasoning (data driven process)
4. Backward reasoning (hypothesis driven process)

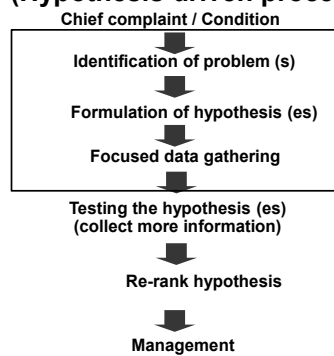
Mahidol University Faculty of Medicine Siriraj Hospital SHEE

Forward Reasoning (Data driven process)



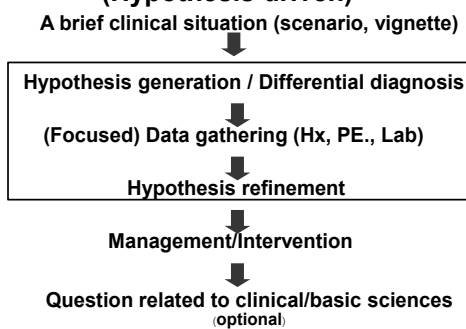
Mahidol University Faculty of Medicine Siriraj Hospital SHEE

Backward reasoning (Hypothesis driven process)



Mahidol University Faculty of Medicine Siriraj Hospital SHEE

MEQ Process (Hypothesis driven)



Mahidol University Faculty of Medicine Siriraj Hospital SHEE

MEQ

- เสมือนการแก้ปัญหาผู้ป่วยในชีวิตจริง
- การแก้ปัญหของผู้ป่วยรายหนึ่งๆ ประกอบด้วยหลายขั้นตอน
 - มีข้อมูลผู้ป่วยบางส่วนในช่วงแรก
 - ต้องสืบค้นหาข้อมูลเพิ่มเติมและวิเคราะห์ ตัดสินใจแก้ปัญหาที่ละขั้นตอน
 - เมื่อทำแต่ละขั้นตอนแล้ว ไม่สามารถย้อนกลับไปแก้ไขสิ่งที่ทำไปก่อนหน้านี้ได้

Physician tasks / Competencies

- Data Gathering (Hx, PE, Lab)
- Hypothesis Generation (Differential Dx)
- Hypothesis Refinement (Dx)
- Management (Emergency, Acute, Long-term)
- Health promotion and maintenance
- Counseling education
- Medical ethics
- Evidence-based
- Mechanism of diseases

Key Features Problem (KFP)

- Key features
 - critical steps in the resolution of each problem
- Focus on
 - a step in which examinees are most likely to make errors
 - a difficult aspect of the identification and management of the problem in clinical practice

1. Page G, Bordage G. The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Acad Med* 1995
2. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005

Key Features Problem (KFP)

- Allow for more cases, items for testing a broader content domain

*"In any clinical case, there are **a few essential elements** in decision making which are the **critical steps** in the resolution of the clinical problem."*

- Reliability of 0.8 in 4 hours of testing had been demonstrated

Page G, Bordage G. The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Acad Med* 1995; 70: 104-10.

Key features: Example

Topic Anaphylaxis

Key features / Critical steps

- I. Diagnosis
- II. Emergency management
- III. Prevention and counseling

Developing an MEQ / KFP

- Assembling problem-writing groups
- Selecting a problem
- Defining the key features
- Writing the questions
- Selecting question formats
- Specifying the number of required answers
- Preparing scoring keys
- Validation and references

Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005, 39: 1188 – 1194.

Assembling Problem-Writing Groups

- Item writers
 - Clinical expertise
 - Multidisciplinary approach / combined expertise
- The written problem
 - well grounded in practice
 - represent a wide range of real-life practice
- Review the content by a group of writers

Select A Problem

- Refer to test specification table
- Select an appropriate clinical problem
 1. พบบ่อยในเวชปฏิบัติ
 2. ความยากง่ายเหมาะสมกับระดับของผู้เรียน
 3. ประเมินทักษะการแก้ปัญหาและการตัดสินใจ
 4. เกี่ยวข้องกับหลายระบบ
 5. มีการบูรณาการของสาขาวิชา
 6. แพทย์มักตัดสินใจผิดพลาด

Defining Key Features

- ปรึกษาในกลุ่มผู้เขียนโจทย์จนได้ consensus
- Critical steps
 - ประเด็นสำคัญในการตัดสินใจ/จัดการกับปัญหาของผู้ป่วย
 - ขั้นตอนที่ขาดไม่ได้ในการดูแลรักษาผู้ป่วย
 - อาจเป็นประเด็นเกี่ยวกับ medical ethics, medico-legal

Defining Key Features (cont.)

- Typical KFs
 - ประวัติเพิ่มเติมที่สำคัญ
 - การตรวจร่างกายที่ต้องมองหาหรือตรวจเพิ่มเติม
 - การสืบค้นเพิ่มเติมเพื่อ confirm หรือ exclude การวินิจฉัย
 - การรักษาที่เฉพาะเจาะจงกับโรค

ไม่จำเป็นต้องเริ่มต้นด้วยการถามประวัติ หรือ ตรวจร่างกาย

From A Problem to A Case

- Select a case scenario
 - age, gender
 - setting of the encounter: OPD, IPD, ER
 - brief case: KFP on diagnosis
 - longer case: KFP on management

Writing the Questions

- Number of questions
 - Most case scenario: 2 – 4 questions
 - Each question test one key feature
- Number of answers for each question
 - Vary: 1 – 10
 - Typical: 3 – 5 answers

KFP on diagnosis

ชายอายุ 60 ปี ถ่ายเป็นเลือดสดปนลิ่มเลือด 3 แก้วน้ำในระยะเวลา 4 ชั่วโมง และมีอาการหน้ามืดใจสั่น มาพบท่านที่ห้องแพทย์เวร

Q1: โรคที่เป็นสาเหตุ 3 อันดับแรก

Q2: ประวัติที่สำคัญในการช่วยวินิจฉัยแยกโรค 5 ข้อ

Q3: ตรวจร่างกายที่สำคัญที่ช่วยวินิจฉัยแยกโรคและประเมินความรุนแรง 5 ข้อ

Specify the Number of Required Answers

ระบุคำถามให้ชัดเจนว่าจะให้ทำอะไร อย่างไร เช่น

- บอกชื่อโรคที่ผู้ป่วยรายนี้น่าจะเป็นมากที่สุด 1 โรค
- บอกสิ่งตรวจพบจากการตรวจร่างกายที่สำคัญที่จะช่วยในการยืนยันการวินิจฉัยโรค มา 3 ประการ
- เขียนคำสั่งการรักษาสำหรับผู้ป่วยรายนี้ในใบคำสั่งการรักษา

Preparing Scoring Keys (1/4)

- List of correct and incorrect responses
- Scores to be assigned to each response
 - Multiple acceptable answers

Key answer	Score
Viral / Rotavirus gastroenteritis	5
Acute gastroenteritis / Infectious diarrhea	3
Acute diarrhea	0

- Only one acceptable answer

Key answer	Score
Acute post-streptococcal glomerulonephritis / Post-infectious glomerulonephritis	10
Glomerulonephritis	0

Preparing Scoring Keys (2/4)

- Partial credit system
- Rubric score

Complete score	คำตอบถูกต้องและสมบูรณ์
Partial score	คำตอบถูกต้องและสมบูรณ์ เพียงบางส่วน
No score	คำตอบไม่ถูกต้อง

Preparing Scoring Keys (3/4)

- Rubric score
- e.g. Investigation

Complete score (5)	AST, ALT
Partial score (3)	LFT

Treatment

Complete score (10)	IV Ceftriaxone
Partial score (5)	IV 3 rd generation cephalosporin
No score (0)	IV antibiotic

Preparing Scoring Keys (4/4)

Penalty

- Absence of “must have” answers
 - score of “0” despite the presence of other less important answers
- Presence of “unnecessary” investigations or treatment
 - no score
 - negative score (but not cross items)
- Harmful treatment
 - negative score (but not cross items)

Time

- แต่ละคำถาม ควรกำหนดเวลาให้เพียงพอสำหรับ
 1. อ่านข้อมูลเพิ่มเติมในแต่ละหน้า ที่อาจมีเนื้อหาหมาก
 2. วิเคราะห์คำถาม
 3. เขียนคำตอบ
- เวลาที่นักศึกษาใช้ในการตอบคำถามนั้นๆ จะมากกว่าเวลาที่อาจารย์ใช้ 30 – 50 %
 - ทดลองตอบคำถามด้วยตนเองและจับเวลา หรือ ให้เพื่อนอาจารย์ทดลองทำ

Validation and References

- Validation
 - pilot the problem with colleagues new to the problem: discussion, revision
- References
 - especially in the field of rapidly developing intervention and discovery



การสร้างข้อสอบอัตนัยประยุกต์

รองศาสตราจารย์ นายแพทย์เชิดศักดิ์ โอรมนิรัตน์ พ.บ., ป.ชั้นสูง (ศัลยศาสตร์), ว.จ. ศัลยศาสตร์, MHPE, Ph.D.
ภาควิชาศัลยศาสตร์, คณะแพทยศาสตร์ศิริราชพยาบาล, มหาวิทยาลัยมหิดล, กรุงเทพมหานคร 10700.

ข้อสอบอัตนัยประยุกต์ (modified essay question, MEQ) เป็นรูปแบบการประเมินผลที่นิยมใช้กับนักศึกษาแพทย์ระดับคลินิกเพื่อประเมินความสามารถในการแก้ปัญหา และตัดสินใจเลือกการตรวจรักษาที่เหมาะสมสำหรับผู้ป่วย ในปัจจุบันมีการใช้ข้อสอบอัตนัยประยุกต์ในการสอบของนักศึกษาแพทย์ในหลายภาควิชา รวมทั้งใช้ในการสอบขั้นตอนที่สามของการประเมินความรู้ความสามารถในการประกอบวิชาชีพเวชกรรม ของแพทยสภาด้วย อย่างไรก็ตาม จากการติดตามเนื้อหาของโจทย์ข้อสอบอัตนัยประยุกต์ ร่วมกับการพิจารณาเกณฑ์การให้คะแนนของข้อสอบเหล่านี้ที่ใช้กับการสอบของนักศึกษาแพทย์ในหลายการสอบ ผู้นิพนธ์ยังคงพบเห็นปัญหาในการสร้างข้อสอบชนิดนี้อยู่พอสมควร บทความนี้จึงได้รับการเขียนขึ้นเพื่อสร้างความเข้าใจในหลักการพื้นฐาน และแนวปฏิบัติที่เหมาะสมในการสร้างข้อสอบอัตนัยประยุกต์สำหรับการประเมินความรู้ทางการแพทย์

ลักษณะพื้นฐานของข้อสอบอัตนัยประยุกต์

ข้อสอบอัตนัยประยุกต์เป็นรูปแบบหนึ่งของข้อสอบอัตนัย (Essay question) ซึ่งในรูปแบบดั้งเดิม (traditional essay) นั้นผู้ออกข้อสอบจะเขียนโจทย์คำถามแล้วให้ผู้สอบเขียนคำตอบด้วยตนเองในขั้นตอนเดียว โดยไม่มีตัวเลือกให้ ในการเขียนคำตอบอาจเขียนตอบเป็นคำ หรือวลีสั้น ๆ (Short essay) หรือ ตอบเป็นบทความที่มีความยาวเป็นย่อหน้า หรือ หลายย่อหน้า (Long essay) ซึ่งผู้ออกข้อสอบคาดหวังการสอบในลักษณะที่ผู้สอบไม่มี

ตัวเลือก แต่ต้องคิดคำตอบด้วยตนเองนี้จะสามารถวัดความรู้ขั้นสูงในระดับการวิเคราะห์ สังเคราะห์ หรือประเมินคุณค่าได้^{1,2}

อย่างไรก็ตามข้อสอบในรูปแบบอัตนัยแบบดั้งเดิมนั้นประสบปัญหาในการใช้ประเมินความรู้ทางการแพทย์อยู่หลายประการ ทั้งความยากในการตรวจให้คะแนน ความจำกัดในปริมาณเนื้อหาที่สามารถสอบได้ในเวลาที่มี ความเห็นที่แตกต่างกันของผู้ตรวจให้คะแนน ความไม่เที่ยงของคะแนนสอบ เป็นต้น^{1,2} ปัญหาที่สำคัญยิ่งที่ทำให้การสอบอัตนัยแบบดั้งเดิมไม่ได้รับความนิยมในการประเมินความรู้ในระดับคลินิกคือ การที่ข้อสอบอัตนัยแบบดั้งเดิมนั้นมักวัดความรู้ในระดับการท่องจำ หรือความเข้าใจพื้นฐานเท่านั้น และรูปแบบการคิดวิเคราะห์เพื่อตอบโจทย์ข้อสอบอัตนัยแบบดั้งเดิมนั้นมีลักษณะแตกต่างไปจากกระบวนการแก้ปัญหาในระดับคลินิกที่แพทย์ปฏิบัติจริง

ข้อสอบอัตนัยแบบดั้งเดิมที่ดีนั้นผู้ออกข้อสอบสามารถประเมินทักษะการคิดวิเคราะห์ขั้นสูงได้ แต่อุปสรรคสำคัญที่ทำให้ไม่สามารถบรรลุวัตถุประสงค์ดังกล่าวได้คือการสร้างข้อสอบที่ผู้สอบตั้งใจให้ตรวจให้คะแนนได้ง่ายเป็นสำคัญ ทำให้ข้อสอบอัตนัยแบบดั้งเดิมส่วนใหญ่ทำการประเมินเพียงความรู้ระดับความจำหรือความเข้าใจพื้นฐานเท่านั้น

สมมติฐานพื้นฐานในการตอบข้อสอบอัตนัยแบบดั้งเดิมคือการวิเคราะห์และหาแนวทางแก้ปัญหาเป็นกระบวนการที่ทำในขั้นตอนเดียว ดังนั้นข้อสอบจึง

นำเสนอข้อมูลทั้งหมดในขั้นตอนเดียวแล้วให้ผู้เข้าสอบ แสดงการวิเคราะห์และแก้ปัญหา ซึ่งเป็นกระบวนการ แก้ปัญหาทางคลินิกที่แพทย์ใช้ในการวินิจฉัยผู้ป่วยที่ไม่ซับซ้อนที่ไม่ต้องการกระบวนการคิดวิเคราะห์ที่ซับซ้อนมากนัก อย่างไรก็ตามปัญหาผู้ป่วยที่มีความซับซ้อนและต้องการ วิเคราะห์มากมักต้องการกระบวนการแก้ปัญหาหลายขั้นตอน แพทย์จะต้องทำการประเมินข้อมูลพื้นฐานที่ได้จากผู้ป่วย แล้วซักประวัติ หรือตรวจร่างกายเพื่อเก็บข้อมูลเพิ่มเติมอย่างเหมาะสม เมื่อได้ข้อมูลพื้นฐานมาแล้ว แพทย์ ต้องทำการตั้งสมมติฐานถึงโรคที่ผู้ป่วยน่าจะเป็น แล้ว ทำการสืบค้นเพิ่มเติมด้วยการตรวจทางห้องปฏิบัติการ หรือใช้ภาพถ่ายรังสี ในบางกรณีแพทย์จำเป็นต้องให้การ รักษาเบื้องต้นก่อน พร้อมกับทำการสืบค้นเพิ่มเติม ซึ่ง เมื่อเวลาผ่านไปแพทย์จะได้รับข้อมูลของผู้ป่วยมากขึ้นเรื่อยๆ จากผลตรวจทางห้องปฏิบัติการ หรือการตอบสนองต่อการรักษาที่ให้ เมื่อได้ข้อมูลมากขึ้นแพทย์จะต้อง ทำการประเมินสถานการณ์ใหม่ ข้อมูลที่เพิ่มขึ้นอาจทำให้ แพทย์สามารถให้การวินิจฉัยที่แน่ชัด และวางแผนการ รักษาที่เหมาะสมได้ จะเห็นได้ว่ากระบวนการแก้ปัญหา ของแพทย์มักทำเป็นหลายขั้นหลายตอน แต่ละขั้นตอน จะได้ข้อมูลเพิ่มเติมเรื่อยๆ การตัดสินใจในแต่ละขั้นเมื่อ ได้เลือกที่จะตรวจหรือให้การรักษาใดแก่ผู้ป่วยแล้ว ไม่สามารถย้อนเวลากลับไปแก้ไขการตัดสินใจที่ทำผิดพลาด ไปก่อนหน้านี้ได้

จากข้อจำกัดของข้อสอบอัตนัยแบบดั้งเดิม ที่กล่าวมาข้างต้น ทำให้มีการพัฒนารูปแบบการสอบ เป็นข้อสอบอัตนัยประยุกต์ (modified essay question, MEQ) ซึ่งเป็นข้อสอบที่เริ่มจากการให้สถานการณ์ของ ผู้ป่วย แล้วมีโจทย์ถามให้ผู้สอบตอบคำถามที่เกี่ยวกับการ แก้ปัญหาผู้ป่วยในสถานการณ์นั้นโดยไม่มีตัวเลือกให้ เมื่อผู้สอบตอบคำถามแล้วจะมีการเปิดเผยข้อมูลเพิ่มเติม เกี่ยวกับผู้ป่วยมากขึ้นทีละน้อย และมีโจทย์ถามคำถาม เพิ่มเติมเป็นลำดับ โดยที่ผู้สอบไม่มีโอกาสย้อนกลับไป แก้ไขคำตอบของตนเองที่ได้ตอบไปในขั้นตอนก่อนหน้านั้น^{1,3} รูปแบบของข้อสอบอัตนัยประยุกต์ที่นิยมใช้กัน มากในยุคแรกๆ มีลักษณะเป็นการสอบถามกระบวนการ ดูแลผู้ป่วยตั้งแต่ต้นจนจบในรูปแบบที่เรียกว่าการจัดการ

ปัญหาของผู้ป่วย (Patient management problem, PMP)^{1, 4, 5}

เนื่องจากข้อสอบอัตนัยประยุกต์ที่ใช้ในทาง การแพทย์มักมุ่งเน้นการประเมินทักษะการวินิจฉัยโรค ผู้นิพนธ์จึงขอทบทวนทฤษฎีเกี่ยวกับกระบวนการวินิจฉัย โรคเล็กน้อยก่อนนำเข้าสู่หลักการสร้างข้อสอบ โดย ทัวไปแล้ววิธีการที่แพทย์ใช้ในการวินิจฉัยโรคมีสามวิธี หลักได้แก่ (1) วิธีจำได้จากแบบแผนของความผิดปกติที่ พบ (pattern recognition), (2) วิธีปฏิบัติตามขั้นตอนวิธี ที่มีแบบแผน (algorithm), และ (3) วิธีทดสอบสมมติฐาน (hypothesis testing)⁶ ซึ่งในวิธีทดสอบสมมติฐานนี้ สามารถแบ่งออกเป็นวิธีการย่อยได้สองวิธีคือ (3.1) การ แก้ปัญหาด้วยวิธีอุปนัย (inductive reasoning) ซึ่งแพทย์ จะรวบรวมข้อมูลอย่างครบถ้วนตามแบบแผนก่อนจึง ตั้งสมมติฐาน และ (3.2) การแก้ปัญหาด้วยวิธีนิรนัย (deductive reasoning) ซึ่งแพทย์จะเริ่มตั้งสมมติฐาน ตั้งแต่เมื่อเริ่มเก็บข้อมูลจากผู้ป่วยเพียงเล็กน้อย แล้วใช้ สมมติฐานที่ได้มานั้นเป็นแนวทางในการซักประวัติ และ ตรวจร่างกายอย่างมีจุดหมายเพื่อทดสอบสมมติฐานที่ตั้ง ขึ้นจนค่อยๆ ตัดโรคที่ไม่สอดคล้องกับข้อมูลที่ได้รับออก ไปเรื่อยๆ โดยทั่วไปแล้ววิธีอุปนัยเป็นวิธีที่มีประสิทธิภาพ น้อยกว่าวิธีนิรนัย เนื่องจากการเก็บข้อมูลเป็นไปอย่าง ขาดจุดหมายทำให้เสียเวลาและอาจพลาดการเก็บข้อมูล ที่สำคัญไป⁶

การสร้างข้อสอบอัตนัยประยุกต์ที่มีคุณภาพ ดีควรเริ่มจากความเข้าใจในปรัชญาพื้นฐานของการ ประเมินผลว่าข้อสอบอัตนัยประยุกต์นั้นได้รับการพัฒนา ขึ้นเพื่อประเมินทักษะการแก้ปัญหาด้วยวิธีนิรนัยเป็น สำคัญ ข้อผิดพลาดที่พบบ่อยของการสร้างข้อสอบอัตนัย ประยุกต์ประการหนึ่งคือการสร้างข้อสอบที่ให้ข้อมูล ผู้ป่วยสั้นมาก (จนไม่มีทางตั้งสมมติฐานที่ชัดเจนได้) แล้ว ตั้งโจทย์ให้ผู้เข้าสอบเขียนรายการประวัติที่จะสอบถาม หรือการตรวจร่างกายที่จะดำเนินการในผู้ป่วยดังกล่าว เช่น ให้สถานการณ์เป็นหญิงอายุ 45 ปี ปวดท้อง 1 วัน แล้วตั้งโจทย์ว่า จงทำการซักประวัติที่เหมาะสม ซึ่งการ ให้สถานการณ์ในลักษณะนี้มีโรคที่สามารถเป็นไปได้ มากมาย ในหลายระบบ สิ่งที่จะประเมินได้จากการตอบ

เวบบันทึกศิริราช

บทความทั่วไป

คำถามลักษณะนี้คือความจำขึ้นพื้นฐาน (simple recall) ว่าแบบแผนการชักประวัติผู้ป่วยต้องเขียนปล้นมืออะไรบ้าง ซึ่งผู้เข้าสอบเขียนอะไรมาก็น่าจะถูกหมด ไม่มีการชักประวัติที่ไม่เข้าประเด็น เนื่องจากข้อมูลจากโจทย์ไม่มีรายละเอียดมากพอที่จะจำกัดโรคที่ควรนึกถึง ข้อสอบอัตนัยประยุกต์ที่ดีควรเริ่มจากข้อมูลที่สามารถสร้างสมมติฐานที่ชัดเจนพอได้ เช่น หญิงอายุ 50 ปี จุดแน่นลิ้นปี่และใต้ชายโครงขวาเป็น ๆ หาย ๆ 4 เดือน มีอาการปวดท้องใต้ชายโครงขวามาก ร่วมกับมีไข้ต่ำ ๆ 7 ชั่วโมง การให้ข้อมูลที่มีรายละเอียดพอสมควรนี้ผู้สอบที่มีความรู้จะตั้งสมมติฐานได้ว่าผู้ป่วยน่าจะเป็นโรคใด หากโจทย์กำหนดให้ชักประวัติเพิ่มเติม ผู้สอบที่มีความรู้จะสามารถสอบถามอาการที่สอดคล้องกับการวินิจฉัยที่เหมาะสมได้ ในกรณีนี้คำตอบที่ไม่สอดคล้อง (เช่นสมมติฐานที่เหมาะสมคือภาวะถุงน้ำดีอักเสบเฉียบพลัน แต่ผู้สอบชักประวัติประจำเดือน ประวัติเพศสัมพันธ์) ไม่ควรได้คะแนน

พัฒนาการของข้อสอบอัตนัยประยุกต์

หลังจากที่มีรายงานการใช้ข้อสอบอัตนัยประยุกต์ในการประเมินผลทางแพทยศาสตรศึกษาตั้งแต่ปี พ.ศ. 2514 โดยราชวิทยาลัยแพทยเวชปฏิบัติทั่วไปเพื่อประเมินทักษะการแก้ปัญหาทางคลินิกแล้ว^{3,7,8} ข้อสอบอัตนัยประยุกต์ก็ได้ถูกใช้ในการประเมินทางการแพทย์และสาธารณสุขในหลากหลายบริบท⁹⁻¹² โดยรูปแบบที่เป็นที่นิยมกันมากเป็นการสอบถามกระบวนการดูแลผู้ป่วยตั้งแต่ต้นจนจบในรูปแบบ การจัดการปัญหาของผู้ป่วย (Patient management problem, PMP) ซึ่งการแก้ปัญหาผู้ป่วยแต่ละรายมักใช้เวลาานานมาก ทำให้การสอบแต่ละครั้งมักมีจำนวนสถานการณ์ผู้ป่วยที่นำมาสอบไม่มากนัก¹³

จากการใช้ข้อสอบอัตนัยประยุกต์ในรูปแบบการจัดการปัญหาของผู้ป่วย พบว่ามีข้อจำกัดบางประการ กล่าวคือ ข้อสอบส่วนใหญ่มุ่งเน้นวัดความครบถ้วนสมบูรณ์ของคำตอบมากกว่าการตัดสินใจแก้ปัญหา จำนวนสถานการณ์ผู้ป่วยที่มีจำนวนน้อยทำให้ไม่สามารถครอบคลุมองค์ความรู้ที่ต้องการประเมินได้ครบ และความ

เที่ยงของคะแนนสอบที่ต่ำ^{4,13,14} ปัญหาที่สำคัญยิ่งในการสอบด้วยสถานการณ์ผู้ป่วยจำนวนน้อยคือ ทักษะในการแก้ปัญหาทางคลินิกมีความจำเพาะต่อบริบทของผู้ป่วยแต่ละราย (case specificity)¹⁵⁻¹⁸ การที่ผู้เข้าสอบสามารถแก้ปัญหาผู้ป่วยที่มีอาการเจ็บหน้าอกได้ดีนั้นไม่สามารถจะบอกได้ว่าผู้เข้าสอบคนดังกล่าวจะสามารถแก้ปัญหาผู้ป่วยที่มีอาการปวดศีรษะได้ดีด้วยหรือไม่ ดังนั้นหลักการที่สำคัญประการหนึ่งในการสร้างข้อสอบอัตนัยประยุกต์ก็คือการจัดทำข้อสอบให้มีหลากหลายสถานการณ์ เพื่อให้สามารถประเมินการแก้ปัญหาของผู้เข้าสอบได้ในหลากหลายบริบท ในหลายระบบอวัยวะ จากปัญหาในการใช้ข้อสอบอัตนัยประยุกต์ต่าง ๆ เหล่านี้ ทำให้นักการศึกษาได้มีการพัฒนารูปแบบข้อสอบอัตนัยประยุกต์ให้ต่างไปจากรูปแบบดั้งเดิม รูปแบบข้อสอบที่ผู้เชี่ยวชาญในการประเมินผลแนะนำในปัจจุบันคือ การแก้ปัญหาสำคัญ (key features problems, KFP)

ข้อสอบอัตนัยประยุกต์แบบแก้ปัญหาสำคัญนี้ได้รับการพัฒนามนหลักสำคัญคือในการแก้ปัญหาผู้ป่วยแต่ละรายมีประเด็นปัญหาที่เป็นหัวใจสำคัญเพียงไม่กี่ประเด็นเท่านั้น ซึ่งประเด็นปัญหาเหล่านี้เรียกว่า ปัญหาสำคัญ (key features)¹⁹ ซึ่งในผู้ป่วยแต่ละรายจะมีปัญหาสำคัญที่แพทย์ต้องให้ความสนใจต่างกันไป บางรายเป็นเรื่องการชักประวัติ บางรายเป็นการเลือกการส่งตรวจทางห้องปฏิบัติการ ในขณะที่บางรายเป็นการตัดสินใจเลือกวิธีการรักษาที่เหมาะสม เป็นต้น ในข้อสอบอัตนัยประยุกต์รูปแบบการแก้ปัญหาสำคัญจะมุ่งเน้นตั้งโจทย์ถามเฉพาะประเด็นปัญหาสำคัญเหล่านี้เท่านั้น ไม่จำเป็นต้องถามกระบวนการดูแลผู้ป่วยตั้งแต่ต้นจนจบในผู้ป่วยทุกราย การสร้างข้อสอบอัตนัยประยุกต์ในลักษณะนี้ทำให้ผู้สอบใช้เวลาในการแก้ปัญหาผู้ป่วยแต่ละรายไม่มากนัก และสามารถประเมินทักษะการแก้ปัญหาได้ในหลากหลายสถานการณ์ คะแนนสอบที่ได้จึงมีความเที่ยงสูง มีรายงานค่าความเที่ยงของคะแนนสอบถึง 0.8 ในการสอบอัตนัยประยุกต์แบบแก้ปัญหาสำคัญเป็นเวลาสี่ชั่วโมง¹⁴

ตัวอย่างข้อสอบอัตนัยประยุกต์แบบแก้ปัญหาสำคัญ
ตอนที่ 1 ชาย 36 ปี น้ำหนักตัว 55 กิโลกรัม ท้องร่วงถ่ายเป็นน้ำ 20 ครั้งในเวลา 1 วัน ตรวจร่างกายพบ อุณหภูมิ 36.9 องศาเซลเซียส ชีพจร 112 ครั้งต่อนาที ตรวจความดันโลหิตท่านอน 104/56 มิลลิเมตรปรอท ความดันโลหิตท่านั่ง 90/50 มิลลิเมตรปรอท

คำถามที่ 1.1 ให้ผู้สอบเขียนปัญหาสำคัญที่สุดของผู้ป่วยรายนี้ 1 อย่าง

ตอนที่ 2 ผู้ป่วยได้รับการประเมินว่ามีภาวะขาดสารน้ำปานกลางถึงรุนแรง ท่านต้องการให้สารน้ำทางหลอดเลือดดำแก่ผู้ป่วย

คำถามที่ 2.1 จงเขียนคำสั่งการรักษาเพื่อให้สารน้ำที่เหมาะสมแก่ผู้ป่วย

คำถามที่ 2.2 จงส่งตรวจเพิ่มเติมทางห้องปฏิบัติการเพื่อช่วยวินิจฉัยผู้ป่วยรายนี้ 2 การตรวจ

จากตัวอย่างข้างต้นจะเห็นว่าผู้ออกข้อสอบไม่ได้เริ่มจากการถามว่าจะซักประวัติ หรือตรวจร่างกายอะไรในผู้ป่วยที่มีภาวะท้องร่วงรุนแรง เนื่องจากผู้ออกข้อสอบเห็นว่าปัญหาสำคัญในการดูแลผู้ป่วยในภาวะนี้เป็นเรื่องการประเมินความรุนแรงของการขาดสารน้ำและการให้น้ำเกลือทดแทนในปริมาณที่เหมาะสมร่วมกับการสืบค้นหาสาเหตุของท้องร่วง ดังนั้นโจทย์ข้อนี้จึงมีเพียงสองตอนและใช้เวลาสอบไม่เกินสิบนาที

ขั้นตอนการสร้างข้อสอบอัตนัยประยุกต์

การสร้างข้อสอบอัตนัยประยุกต์ที่มีคุณภาพดีควรมีการดำเนินการเป็นขั้นตอน ดังนี้^{4,20}

1. ตั้งกลุ่มพัฒนาข้อสอบ

ข้อสอบอัตนัยประยุกต์ที่ดีควรเป็นการแก้ปัญหาที่อาศัยความรู้จากหลากหลายวิชา การมีทีมคณาจารย์ที่มีประสบการณ์และความชำนาญแตกต่างกันมาช่วยกันสร้างข้อสอบจะได้สถานการณ์ผู้ป่วยที่เหมือนจริงในเวชปฏิบัติและสามารถประเมินความรู้ของผู้เข้าสอบได้ครอบคลุมสหสาขาวิชา และมั่นใจได้ว่าการเฉลยคำตอบทำได้อย่างรอบคอบ

2. เลือกปัญหาทางคลินิกที่จะทำการประเมินผู้สอบ

ขั้นตอนนี้เป็นขั้นตอนที่สำคัญมาก เนื่องจากโดยลักษณะข้อสอบอัตนัยประยุกต์จะทำให้ทำการสอบได้จำนวนข้อไม่มากนัก จึงเป็นไปได้ที่จะทำให้สถานการณ์ที่เป็นปัญหาทางคลินิกทุกอย่างจะมาปรากฏอยู่ในชุดข้อสอบ ดังนั้นการเลือกปัญหาทางคลินิกที่จะทำการสอบจึงต้องทำอย่างเป็นระบบ ควรมีการจัดทำตารางกำหนดลักษณะข้อสอบที่ชัดเจนว่าในการสอบครั้งหนึ่ง ๆ จะมีข้อสอบกี่ข้อ จะประเมินความรู้ในระบอบวัยวะใด และจัดสรรให้ข้อสอบไม่ซ้ำซ้อนกัน (ไม่ควรมีข้อสอบสองข้อถามความรู้ในระบอบวัยวะเดียวกัน ในขณะที่บางระบบวัยวะไม่มีข้อสอบเลย)

ลักษณะปัญหาทางคลินิกที่ควรเลือกมาสอบด้วยข้อสอบอัตนัยประยุกต์ ได้แก่

- ปัญหาที่พบได้บ่อยในเวชปฏิบัติ
- ปัญหาที่แพทย์เกิดความผิดพลาดในการดูแลผู้ป่วยค่อนข้างบ่อย
- ปัญหาที่ยังไม่สามารถวินิจฉัยสาเหตุได้ชัดเจน
- ปัญหาที่มีความเกี่ยวข้องกับหลายระบบ

เมื่อทีมคณาจารย์กำหนดปัญหาทางคลินิกที่จะทำการประเมินได้ชัดเจนแล้ว (เช่น ปัญหาตัวเหลือง, น้ำหนักลด เป็นต้น) สิ่งที่ต้องดำเนินการต่อคือการสร้างสถานการณ์ผู้ป่วยที่แสดงถึงปัญหาดังกล่าวขึ้น โดยกำหนดรายละเอียดต่าง ๆ ให้ผู้เข้าสอบอ่านแล้วนึกภาพผู้ป่วยได้ ในสถานการณ์ควรมีรายละเอียดเกี่ยวกับอายุ เพศ อาการสำคัญ บริบทของการดูแลผู้ป่วย (เช่น ห้องฉุกเฉินของโรงพยาบาลชุมชน หรือ หอผู้ป่วยในโรงพยาบาลมหาวิทยาลัย เป็นต้น)

3. กำหนดปัญหาสำคัญ

เมื่อทีมคณาจารย์เลือกปัญหาทางคลินิกที่จะทำการสอบแล้ว คณาจารย์ต้องตั้งคำถามว่าขั้นตอนใดในการดูแลผู้ป่วยที่มีปัญหาดังกล่าวจัดเป็นขั้นตอนสำคัญที่สุดในการจัดการปัญหานั้น ซึ่งขั้นตอนดังกล่าวจะได้รับการกำหนดให้เป็น ปัญหาสำคัญของสถานการณ์ผู้ป่วยที่จะใช้สอบ ในบางกรณีที่มีทีมคณาจารย์ไม่สามารถเลือกขั้นตอนสำคัญในปัญหาทางคลินิกนั้น ๆ จากวิธีดังกล่าวได้

อาจใช้คำถามว่าขั้นตอนใดในการดูแลผู้ป่วยที่มีปัญหา ดังกล่าวเป็นขั้นตอนที่นักศึกษาแพทย์หรือแพทย์ประจำบ้านทำผิดพลาดมากที่สุด⁴

มีข้อเสนอแนะสองประการสำหรับการกำหนดปัญหาสำคัญในแต่ละสถานการณ์ได้แก่

- สิ่งที่ต้องตัดสินใจในผู้ป่วยแม้เป็นสิ่งที่ถูกต้อง และควรปฏิบัติอาจไม่ได้เป็นขั้นตอนสำคัญที่จะต้องนำมาสอบเสมอไป การปฏิบัติต่อผู้ป่วยหลายอย่างที่มักกันเป็นปกติ โดยไม่ต้องคิดวิเคราะห์ เป็นขั้นตอนที่ไม่ค่อยทำผิดพลาด มักไม่ใช่ปัญหาสำคัญในสถานการณ์นั้น

- ปัญหาสำคัญไม่จำกัดอยู่เฉพาะประเด็นปัญหาทาง ชีววิทยาการแพทย์ (biomedical) เท่านั้น ในบางสถานการณ์ปัญหาสำคัญอาจเป็นประเด็นทางจริยธรรม กฎหมาย หรือ การส่งเสริมสุขภาพและป้องกันโรคก็ได้

4. เขียนโจทย์คำถาม

เมื่อมีสถานการณ์ผู้ป่วยและขั้นตอนที่เป็นปัญหาสำคัญในสถานการณ์นั้นแล้ว ทีมคณาจารย์ต้องเขียนโจทย์คำถามที่มีความชัดเจน เพื่อประเมินว่าผู้เข้าสอบมีความสามารถในการตัดสินใจในการแก้ปัญหาสำคัญในสถานการณ์ดังกล่าวหรือไม่ โดยทั่วไปแล้วลักษณะโจทย์คำถามที่ใช้บ่อยในข้อสอบอัตนัยประยุกต์ได้แก่

- จงสอบถามประวัติที่สำคัญเพิ่มเติม
- จงบอกการตรวจร่างกายที่สำคัญที่ต้องมองหา (หรือตรวจเพิ่มเติม) ในผู้ป่วย
- จงให้การวินิจฉัย (หรือ การวินิจฉัยแยกโรค)
- จงสั่งการตรวจค้นเพิ่มเติมเพื่อให้การวินิจฉัยโรค
- จงสั่งการรักษาที่เหมาะสมให้ผู้ป่วย

โดยทั่วไปแล้วสถานการณ์ผู้ป่วยหนึ่ง ๆ ควรมีคำถามราว 2 – 3 ข้อ แต่ละข้อประเมินความสามารถในการจัดการกับปัญหาสำคัญ 1 ประเด็น^{4,21} ในการเขียนโจทย์คำถามแต่ละข้อนั้นแนะนำให้มีการกำหนดจำนวนคำตอบที่สามารถตอบได้ไว้ด้วย เช่น

- จงบอกชื่อโรคที่ผู้ป่วยรายนี้น่าจะเป็นมากที่สุด 1 โรค
- จงบอกผลการตรวจร่างกายที่สำคัญที่จะช่วยยืนยันการวินิจฉัยโรคมา 3 ประการ

- จงระบุการตรวจเพิ่มเติมทางห้องปฏิบัติการที่จะช่วยในการวินิจฉัยโรค 1 การตรวจ

การกำหนดจำนวนคำตอบนี้จะทำให้ผู้เข้าสอบต้องเลือกสิ่งที่ถูกต้องเหมาะสมที่สุดเท่านั้นมาเขียนตอบ หากผู้เข้าสอบเขียนคำตอบเกินจำนวนที่กำหนด อาจารย์ผู้ตรวจข้อสอบจะไม่อ่านคำตอบที่เกินมา การปฏิบัติเช่นนี้จะช่วยกำจัดปัญหาการตรวจกระดาษคำตอบที่ผู้เข้าสอบเขียนคำตอบแบบหว่านแห ให้ครอบคลุมทุกอย่างโดยที่ผู้เข้าสอบเองไม่มีความรู้ ความเข้าใจว่าสิ่งใดเป็นประเด็นสำคัญในการดูแลผู้ป่วยในขั้นตอนนั้น ๆ

เมื่อทำการเขียนโจทย์คำถามและจำนวนคำตอบที่ต้องการแล้ว ให้อาจารย์ระบุเวลาที่ใช้ในการตอบคำถามตอนนั้นด้วย เนื่องจากข้อสอบอัตนัยประยุกต์มีการดำเนินการของสถานการณ์ผู้ป่วยที่กำหนดให้โดยมีการให้ข้อมูลที่ละส่วน ผู้เข้าสอบจำเป็นต้องรู้เวลาที่มีในการทำข้อสอบแต่ละตอนก่อนที่จะต้องส่งคำตอบและสถานการณ์ผู้ป่วยดำเนินต่อไป ในการกำหนดเวลาในการทำข้อสอบแต่ละตอนให้อาจารย์ผู้ออกข้อสอบพิจารณาจากทั้งเวลาที่ต้องใช้ในการอ่าน และเวลาที่ต้องใช้ในการเขียนคำตอบในข้อสอบตอนที่ต้องอ่านเนื้อหาโจทย์มาก หรือต้องเขียนคำตอบหลายบรรทัด ควรต้องมีการให้เวลาในการทำข้อสอบมากพอ หากเป็นไปได้ควรได้มีการลองทำการอ่านโจทย์และเขียนคำตอบโดยตัวอาจารย์ผู้ออกข้อสอบเองหรือเพื่อนอาจารย์แล้วลองจับเวลาที่อาจารย์ใช้ในการทำข้อสอบตอนนั้น ๆ เวลาที่ได้จะเป็นเวลาที่ผู้เชี่ยวชาญใช้แก้ปัญหาผู้ป่วยในสถานการณ์ดังกล่าว หากให้นักศึกษาทำ ควรเพิ่มเวลาให้ร้อยละ 30 – 50 ของเวลาที่อาจารย์ใช้

5. กำหนดเกณฑ์การให้คะแนน

ขั้นตอนสุดท้ายในการสร้างข้อสอบอัตนัยประยุกต์คือการกำหนดเกณฑ์การให้คะแนน ซึ่งเป็นขั้นตอนที่มีความท้าทาย และสร้างความลำบากใจให้แก่อาจารย์ผู้ออกข้อสอบหลายท่าน เนื่องด้วยเกรงว่าจะเฉลยคำตอบไม่ครอบคลุมสิ่งที่ผู้เข้าสอบจะเขียนตอบมา หรือเกิดความไม่เป็นธรรมขึ้น ในที่นี้ผู้นิพนธ์ขอเสนอแนะแนวทางในการกำหนดเกณฑ์ให้คะแนนดังนี้

- แนะนำให้กำหนดคะแนนเต็มในการแก้ปัญหา

เวบบันทึทศึรึรึรึ

บทความทึวไป

สถานการณ์หนึ่ง ๆ เป็น 100 คะแนน เท่ากันในทุกสถานการณ์ เพื่อให้ไม่ต้องทำการปรับคะแนนสอบหลังการตรวจข้อสอบ

- กรณีที่มีคำตอบที่ถูกต้องยอมรับได้เพียงคำตอบเดียว เช่นข้อมูลจากโจทย์มีความชัดเจนว่าผู้ป่วยเป็นโรคอะไร แล้วโจทย์ให้ผู้เข้าสอบตอบชื่อโรค หากผู้เข้าสอบตอบตรงตามเฉลยที่ตั้งไว้ให้ได้คะแนนเต็ม หากตอบอื่นนอกจากนั้นไม่ได้คะแนน

- ในกรณีที่มีคำตอบที่เป็นไปได้หลายคำตอบ เช่นถามการวินิจฉัยแยกโรค 3 โรค ในกรณีนี้ผู้ออกข้อสอบควรเตรียมเฉลยไว้หลายคำตอบ (มากกว่าที่กำหนดให้ตอบ) โดยแต่ละคำตอบสามารถมีน้ำหนักคะแนนไม่เท่ากันได้ โดยคำตอบที่ถูกต้องมาก สอดคล้องกับสิ่งที่ควรคิดถึงหรือปฏิบัติในขั้นตอนดังกล่าว จะได้คะแนนสูง ในขณะที่สิ่งที่สามารถเป็นไปได้หรือควรปฏิบัติน้อยกว่าจะได้คะแนนลดลงไป แต่เมื่อรวมคะแนนจากทุกคำตอบที่ผู้เข้าสอบตอบมาแล้วคะแนนสูงสุดที่ผู้เข้าสอบจะได้ต้องไม่สูงเกินคะแนนที่กำหนดไว้เป็นคะแนนเต็มของข้อสอบตอนนั้น

- คำตอบบางลักษณะมีการเขียนเนื้อหาที่มีความครบถ้วนสมบูรณ์แตกต่างกันได้ การกำหนดเกณฑ์สามารถกำหนดให้คำตอบที่มีความสมบูรณ์ได้คะแนนเต็ม ส่วนคำตอบที่ไม่สมบูรณ์จะได้คะแนนลดหลั่นลงไปตามความเหมาะสม (เช่น โจทย์ถามเรื่องการให้สารน้ำทางหลอดเลือดดำ คำตอบ Normal saline solution 1000 ml IV drip 200 ml/hr จะได้คะแนนเต็ม 4 คะแนน แต่หากเขียนตอบ Normal saline solution โดยไม่บอกอัตราเร็วของการให้ ได้เพียง 2 คะแนน หากบอกอัตราการให้ถูกต้องให้ 2 คะแนน)

- คำตอบที่ไม่ถูกต้อง ไม่สมควรปฏิบัติแก่ผู้ป่วยโดยทั่วไปแล้วพิจารณาไม่ให้เป็นคะแนน ซึ่งก็จัดเป็นการทำโทษในระดับหนึ่งแล้ว เพราะผู้สอบมีสิทธิเขียนคำตอบได้จำนวนจำกัด การที่ไม่ให้คะแนนในคำตอบที่ไม่เหมาะสม ก็จะทำให้คะแนนสูงสุดที่ผู้สอบจะทำได้ลดลงไปแล้ว การปฏิบัติที่ไม่ถูกต้องที่มีผลเสียรุนแรงต่อผู้ป่วยเท่านี้ที่ควรพิจารณาให้คะแนนติดลบ และแม้มีการให้คะแนนติดลบก็ไม่ควรมีการติดลบข้ามไปถึงข้อสอบข้ออื่นในชุดข้อสอบนั้น

- การกำหนดเกณฑ์การให้คะแนน ไม่ควรใช้อาจารย์ท่านเดียวในการกำหนด เพราะมักได้คำตอบที่ไม่ครอบคลุม ควรใช้ทีมคณาจารย์หลายท่านช่วยกันคิดคำตอบที่ผู้เข้าสอบอาจจะตอบได้ในสถานการณ์ดังกล่าว ซึ่งจะได้เกณฑ์การให้คะแนนที่สมบูรณ์กว่า อย่างไรก็ตาม ถึงแม้ว่าจะใช้คณาจารย์หลายท่านช่วยกันคิดคำตอบแล้วก็ตาม จะพบว่าในการตรวจข้อสอบอัตโนมัติประยุกต์หลายครั้ง จะพบคำตอบที่ผู้เข้าสอบตอบมาที่ น่าจะได้คะแนนแต่อาจารย์ผู้ออกข้อสอบไม่ได้กำหนดเกณฑ์คะแนนไว้ล่วงหน้าอยู่ประปราย ดังนั้นในการนำข้อสอบอัตโนมัติประยุกต์ที่สร้างขึ้นใหม่มาใช้ในการสอบ 2-3 รอบแรกแนะนำให้อาจารย์ผู้ออกข้อสอบและมีความเชี่ยวชาญชำนาญในการดูแลผู้ป่วยในสถานการณ์นั้น ๆ เป็นผู้ทำการตรวจข้อสอบ เพื่อให้สามารถพิจารณาได้ว่าคำตอบใดที่น่าจะเพิ่มเข้าไปในเกณฑ์การให้คะแนนด้วย ซึ่งเมื่อทำไป 2-3 รอบการสอบแล้วมักจะได้เกณฑ์การให้คะแนนที่มีความครอบคลุมคำตอบที่ผู้สอบจะตอบมาได้ทั้งหมด แล้วจึงมอบหมายให้อาจารย์ท่านอื่นช่วยตรวจให้คะแนนข้อสอบต่อไป

เมื่อทำการกำหนดเกณฑ์การให้คะแนนในข้อสอบเสร็จทุกข้อย่อยแล้วกระบวนการขั้นตอนสุดท้ายในการสร้างข้อสอบอัตโนมัติประยุกต์คือการกำหนดเกณฑ์ผ่านของโจทย์สถานการณ์นั้น กล่าวคือจากคะแนนเต็ม 100 คะแนน ผู้สอบต้องทำคะแนนได้อย่างน้อยที่สุดกี่คะแนนจึงจะจัดว่าสอบผ่านในการแก้ปัญหาสถานการณ์นั้น ๆ วิธีการตั้งเกณฑ์ผ่านทำได้หลายวิธี แต่วิธีที่เป็นที่นิยมมากที่สุดสำหรับข้อสอบอัตโนมัติประยุกต์ และเป็นวิธีที่คณะแพทยศาสตร์ศิริราชพยาบาลใช้เป็นประจำในการตัดสินผลสอบอัตโนมัติประยุกต์คือวิธี Modified Angoff ซึ่งมีขั้นตอนที่สำคัญสามขั้นตอนคือ

(1) กำหนดลักษณะของผู้ที่มีความรู้ ความสามารถคาบเส้น (borderline examinee) ว่าในความเห็นของคณาจารย์แล้วผู้ที่มีความรู้เทียบเท่าระดับต่ำสุดของเกณฑ์มาตรฐานการทำงานในการแก้ปัญหาเรื่องนั้น ๆ น่าจะทำอะไรได้ ทำอะไรไม่ได้

(2) ไล่ดูโจทย์คำถามทีละข้อพร้อมเฉลย แล้วทำสัญลักษณ์ * ไว้ในคำตอบที่คาดว่าผู้ที่มีความรู้ ความสามารถคาบเส้นจะตอบในข้อสอบแต่ละตอน

เวบบันทึกศิริราช

บทความทั่วไป

(3) ทำการรวมค่าคะแนนที่ได้รับการทำสัญลักษณ์ * ไว้ตั้งแต่ข้อแรกจนถึงข้อสุดท้าย จะได้คะแนนเกณฑ์ผ่านในการแก้ปัญหสถานการณ์นั้น ๆ²²

แนวทางการพัฒนาข้อสอบอัตนัยประยุกต์ในคณะแพทยศาสตร์ศิริราชพยาบาล

คณะแพทยศาสตร์ศิริราชพยาบาลมีการใช้ข้อสอบอัตนัยประยุกต์ในการประเมินความรู้ของนักศึกษาแพทย์ชั้นคลินิกมานานแล้ว โดยเริ่มต้นจากการสอบของแต่ละภาควิชา และต่อมาเมื่อศูนย์ประเมินและรับรองความรู้ความสามารถในการประกอบวิชาชีพเวชกรรมกำหนดให้การสอบอัตนัยประยุกต์เป็นส่วนหนึ่งของการประเมินขั้นตอนที่ 3 ในการขอใบประกอบวิชาชีพเวชกรรมตั้งแต่ปีการศึกษา 2550 ทางคณะแพทยศาสตร์ศิริราชพยาบาลก็ได้มีการจัดสอบประมวลความรู้ทางการแพทย์สหสาขาวิชา ด้วยข้อสอบอัตนัยประยุกต์ (comprehensive MEQ examination) ในนักศึกษาแพทย์ปีที่ 6 อย่างต่อเนื่อง ตลอดช่วงเวลาที่มีการใช้ข้อสอบอัตนัยประยุกต์ในคณะฯได้มีการพัฒนาข้อสอบประเภทนี้อย่างต่อเนื่อง จากเดิมเคยจัดสอบข้อสอบอัตนัยประยุกต์ในรูปแบบข้อสอบกระดาษ จนพัฒนาให้จัดสอบอัตนัยประยุกต์ด้วยการนำเสนอข้อมูลผู้ป่วยบนจอภาพคอมพิวเตอร์ ร่วมกับการเขียนคำตอบในกระดาษคำตอบ ตั้งแต่ปีการศึกษา 2552 จนถึงปัจจุบัน แต่ถึงแม้ว่าฝ่ายการศึกษาจะมีการพัฒนาระบบจัดสอบข้อสอบอัตนัยประยุกต์ให้มีประสิทธิภาพมากขึ้น อำนาจความสะดวกให้ผู้เข้าสอบมากขึ้น และเพิ่มความพึงพอใจในประสบการณ์การสอบขึ้นอย่างต่อเนื่อง จากการเก็บรวบรวมข้อมูลการวิเคราะห์ข้อสอบ วิเคราะห์คะแนน และแบบสำรวจความพึงพอใจของผู้สอบที่ผ่านมาผู้นิพนธ์มีความเห็นว่าการจัดสอบประมวลความรู้ทางการแพทย์ด้วยข้อสอบอัตนัยประยุกต์ของนักศึกษาแพทย์ยังสามารถพัฒนาให้มีคุณภาพดีขึ้นได้อีกในหลายด้าน ดังนี้

(1) เนื้อหาข้อสอบ

ข้อสอบอัตนัยประยุกต์ที่ใช้ในการสอบประมวลความรู้ทางการแพทย์ของคณะแพทยศาสตร์ศิริราชพยาบาลที่ผ่านมามีหลายข้อเป็นเนื้อหาวิชาที่ยากและเป็นความรู้ลึกในระดับผู้เชี่ยวชาญเฉพาะทาง แนวทางการ

พัฒนาการสอบอัตนัยประยุกต์อันดับแรกคือการพัฒนาเนื้อหาให้เหมาะสมกับการประเมินความรู้ของแพทย์เวชปฏิบัติทั่วไป

เนื้อหาข้อสอบอัตนัยประยุกต์สำหรับการสอบประมวลความรู้ที่เน้นเนื้อหาที่เป็นสหสาขาวิชา กล่าวคือต้องอาศัยองค์ความรู้ที่นักศึกษาได้ศึกษาจากหลายภาควิชามาช่วยกันแก้ปัญหาผู้ป่วย ข้อสอบอัตนัยประยุกต์ที่นำมาสอบนักศึกษาแพทย์ทุกข้อในปัจจุบันล้วนมีความเป็นสหสาขาวิชาทั้งสิ้น มีอาจารย์จากหลากหลายภาควิชามาร่วมกันออกข้อสอบ แต่อย่างไรก็ตามข้อสอบบางข้ออาจมีลักษณะการใช้ความรู้สหสาขาวิชาแบบแยกเป็นส่วน ๆ กล่าวคืออาจารย์ต่างภาควิชากันใช้การแบ่งงานออกเป็นส่วน ๆ อาจารย์ภาควิชาที่หนึ่งออกข้อสอบในตอนหนึ่งกับสอง อาจารย์ภาควิชาที่สองออกข้อสอบในตอนที่สามกับสี่ และอาจารย์ภาควิชาที่สามออกข้อสอบในตอนที่ยี่ห้ากับหก ข้อสอบลักษณะนี้มักจะยากมาก เนื่องจากเป็นการใช้ความรู้เชิงลึกของแต่ละภาควิชาที่ละเรื่อง เช่น ชักประวัติ ตรวจร่างกายแล้วก็ไม่สามารถวินิจฉัยโรคได้ ต้องส่งต่อไปทำการตรวจเพิ่มเติมในอีกภาควิชาหนึ่ง ซึ่งผลการตรวจเพิ่มเติมก็แปลผลได้ยาก เมื่อได้ข้อสรุปแล้วก็ต้องส่งต่อไปให้แพทย์อีกสาขาวิชาหนึ่งทำการรักษา เมื่อรักษาแล้วก็มีการตรวจซ้ำต้องส่งต่อให้แพทย์อีกสาขาวิชาหนึ่งทำการแก้ไขภาวะแทรกซ้อนให้ เป็นต้น โดยทั่วไปแล้วข้อสอบอัตนัยประยุกต์ที่ใช้ความรู้สหสาขาวิชาที่เป็นที่ต้องการในการสอบประมวลความรู้ที่นั้นไม่ควรเป็นการประเมินความรู้ในเชิงลึกที่ละวิชาในข้อสอบแต่ละตอน แต่ควรเป็นการผสมผสานความรู้จากหลากหลายสาขาวิชาในทุกขั้นตอน เช่น หญิงอายุ 30 ปี ปวดท้องน้อยคือ ๆ ตลอดเวลา 6 ชั่วโมง มีไข้ต่ำ ๆ คลื่นไส้เล็กน้อย โจทย์ให้ผู้สอบซักประวัติเพื่อการวินิจฉัยโรคซึ่งผู้สอบที่จะตอบคำถามได้ดีต้องอาศัยความรู้ทั้งโรคในระบบทางเดินอาหาร ทางเดินปัสสาวะ อวัยวะสืบพันธุ์สตรี กระดูกและกล้ามเนื้อ เป็นต้น

ข้อแนะนำในเรื่องเนื้อหาที่สำคัญคืออาจารย์ผู้ออกข้อสอบต้องตระหนักว่าการสอบนี้เป็นการประเมินความรู้เวชปฏิบัติทั่วไป มิใช่การประเมินความรู้เชิงลึกในศาสตร์ของแต่ละสาขาวิชา โรคหรือภาวะที่นำมาออก

เวชบันทึกศิริราช

บทความทั่วไป

ข้อสอบส่วนใหญ่ควรวอยู่ในเกณฑ์มาตรฐานผู้ประกอบวิชาชีพเวชกรรมในกลุ่มที่ 1 หรือ 2 (โรคหรือภาวะที่แพทย์เวชปฏิบัติทั่วไปสามารถให้การดูแลด้วยตนเองได้ และพิจารณาส่งต่อในกรณีโรครุนแรงหรือซับซ้อน) โรคหรือภาวะที่อยู่ในเกณฑ์มาตรฐานฯ กลุ่มที่ 3 (โรคหรือภาวะที่แพทย์เวชปฏิบัติทำการดูแลเบื้องต้นแล้วให้ส่งต่อไปยังผู้เชี่ยวชาญ) ควรนำมาออกข้อสอบไม่มากนัก หากจะนำโรคหรือภาวะในเกณฑ์มาตรฐานฯ กลุ่มที่ 3 มาออกสอบ ต้องมุ่งเน้นการดูแลรักษาเบื้องต้นที่แพทย์เวชปฏิบัติทั่วไปพึงทำได้ ไม่ควรมุ่งประเด็นไปที่การรักษาโดยผู้เชี่ยวชาญ เฉพาะสาขามากจนเกินไป

(2) รูปแบบคำถาม

หลักการสำคัญของการวัดและประเมินผลคือการเลือกใช้เครื่องมือที่เหมาะสมในการวัดผลการเรียนรู้ ข้อสอบอัตนัยประยุกต์ได้รับการพัฒนาขึ้นเพื่อประเมินทักษะในการตัดสินใจทางคลินิกเป็นสำคัญ สิ่งที่ยังเป็นปัญหาในข้อสอบอัตนัยประยุกต์บางข้อคือการเลือกถามคำถามในรูปแบบที่ไม่ตรงตามเป้าประสงค์ของการสอบอัตนัยประยุกต์ เช่นถามความจำขึ้นพื้นฐานโดยไม่ต้องคิดวิเคราะห์และตัดสินใจว่าจะทำหรือไม่ทำอะไรกับผู้ป่วย รูปแบบคำถามที่ไม่เหมาะสมเหล่านี้เช่น ผู้ชายอายุ 40 ปี มีไข้สองเดือน จงถามประวัติ การใช้รูปแบบคำถามลักษณะนี้จะวัดเพียงว่าผู้เข้าสอบจดจำหัวข้อทั้งหมดของการซักประวัติในผู้ป่วยที่มีไข้เรื้อรังได้หรือไม่ และผู้สอบคนใดเขียนได้เร็วและครบถ้วนกว่ากัน ซึ่งอาจารย์สามารถใช้เครื่องมือประเมินผลชนิดอื่นในการวัดความจำขึ้นพื้นฐานได้ดีกว่าการใช้ข้อสอบอัตนัยประยุกต์ การใช้ข้อสอบอัตนัยประยุกต์ควรมุ่งเน้นคำถามประเมินความสามารถในการวิเคราะห์ปัญหาผู้ป่วย และตัดสินใจสั่งการตรวจ หรือรักษาผู้ป่วยอย่างเหมาะสม

(3) จำนวนสถานการณ์ผู้ป่วยที่ใช้สอบ

ในการสอบประมวลความรู้ด้วยข้อสอบอัตนัยประยุกต์ของคณะแพทยศาสตร์ศิริราชพยาบาลที่ผ่านมามีการใช้สถานการณ์ผู้ป่วยในข้อสอบตั้งแต่ 5 ถึง 8 ราย ถึงแม้ว่าจำนวนสถานการณ์ในการสอบระยะหลังมี

แนวโน้มเพิ่มขึ้น แต่หากพิจารณาในแง่ของความจำเพาะต่อบริบทของผู้ป่วย (case specificity) ที่ได้อภิปรายไปก่อนหน้านี้แล้วจะเห็นได้ว่าการที่ผู้สอบแก้ปัญหาผู้ป่วยได้ 5 ถึง 8 รายนี้น่าจะยังครอบคลุมประเด็นปัญหาทางคลินิกได้ไม่มากเพียงพอ และคะแนนสอบที่ได้มาน่าจะพัฒนาให้มีความเที่ยงสูงขึ้นไปได้อีกหากในการสอบมีจำนวนสถานการณ์มากขึ้น เนื่องด้วยรูปแบบข้อสอบอัตนัยประยุกต์ที่ใช้ในการสอบของคณะฯยังเน้นการสอบถามการจัดการปัญหาของผู้ป่วยตลอดตั้งแต่ต้นจนจบ (Patient management problem, PMP) จึงทำให้เวลาที่ใช้ในการสอบในแต่ละสถานการณ์ค่อนข้างนาน (แต่ละสถานการณ์มีคำถามย่อย 4 – 8 ข้อ ใช้เวลา 15 ถึง 30 นาทีต่อสถานการณ์) จึงทำให้ไม่สามารถสอบได้หลายสถานการณ์

หากพิจารณาจากข้อเสนอแนะของผู้เชี่ยวชาญในการประเมินผลที่ได้อภิปรายไปก่อนหน้านี้ที่แนะนำให้ใช้ข้อสอบอัตนัยประยุกต์แบบแก้ปัญหาสำคัญ แนวทางการพัฒนาข้อสอบอัตนัยประยุกต์ของคณะฯ ให้มีความครอบคลุมสถานการณ์ผู้ป่วยที่มากขึ้น และ มีความเที่ยงของคะแนนสอบมากขึ้นคือการใช้ข้อสอบแบบแก้ปัญหาสำคัญมาแทนการจัดการปัญหาของผู้ป่วยตั้งแต่ต้นจนจบ กล่าวคือในแต่ละสถานการณ์ผู้ป่วย ข้อสอบควรมุ่งถามคำถามสำคัญเพียงสองหรือสามข้อ และเพิ่มจำนวนสถานการณ์ผู้ป่วยให้มากขึ้นนั่นเอง

(4) การนำเสนอข้อสอบ

การทำข้อสอบอัตนัยประยุกต์ ผู้สอบต้องทำงานภายใต้ข้อจำกัดด้านเวลา เวลาที่ใช้ในการตอบข้อสอบอัตนัยประยุกต์เป็นผลรวมของเวลาที่ใช้อ่านโจทย์ คิดวิเคราะห์ และเขียนคำตอบ ปัญหาสำคัญประการหนึ่งที่สร้างความลำบากให้กับผู้สอบคือปริมาณข้อมูลที่นำเสนอให้ผู้สอบอ่านในสถานการณ์ผู้ป่วยแต่ละรายนั้นมีมาก ทำให้ผู้สอบต้องใช้เวลาในการอ่านมากและเหลือเวลาสำหรับเขียนคำตอบน้อย ถึงแม้ว่าในการนำเสนอข้อมูลของข้อสอบอัตนัยประยุกต์จะได้มีการแยกข้อมูลเดิมที่เคยนำเสนอไปก่อนหน้านี้ ออกจากข้อมูลใหม่ที่เพิ่มเติมขึ้นมาในการนำเสนอข้อสอบแต่ละตอนแล้วก็ตาม ด้วย

เวบบันทึทศึรึรึรึ

บทความท่วไ้

รายละเอียดที่นำเสนอมีมาก ผู้สอบก็ยังคงมีความจำเป็นต้องประมวลผลข้อมูลปริมาณมากอยู่ดี จากการทบทวนเนื้อหาของข้อสอบอัตนัยประยุกต์ที่ได้จัดสอบไปหลายครั้งพบว่าข้อสอบหลายข้อใช้ข้อมูลเพียงส่วนน้อยของที่นำเสนอเท่านั้นก็สามารถนำไปสู่การแก้ปัญหาและการตัดสินใจเลือกการส่งตรวจหรือให้การรักษาผู้ป่วยได้อย่างถูกต้อง ดังนั้นแนวทางในการพัฒนาคุณภาพของข้อสอบอัตนัยประยุกต์อีกทางหนึ่งคือการที่อาจารย์ผู้ออกข้อสอบพึงตระหนักถึงข้อจำกัดเรื่องเวลาในการทำข้อสอบของนักศึกษาและเขียนสถานการณ์ผู้ป่วยให้มีความกระชับ นำเสนอเฉพาะข้อมูลที่มีความจำเป็นในการตัดสินใจให้การดูแลรักษาผู้ป่วยเท่านั้น ในการนำเสนอข้อมูลแต่ละตอนควรต้องทบทวนว่าข้อมูลเก่าที่เคยให้ในขั้นตอนก่อนหน้านั้นมีความจำเป็นต้องนำเสนอซ้ำทั้งหมดหรือไม่ หากทำได้ควรทำการสรุปข้อมูลให้ผู้เข้าสอบ และตัดทอนข้อมูลที่ไม่จำเป็นในการแก้ปัญหาขั้นตอนนั้น ๆ ออกไป ตัวอย่างเช่น ในข้อสอบตอนที่หนึ่งมีการนำเสนอประวัติผู้ป่วยสั้น ๆ แล้วมีโจทย์ถามถึงประวัติที่จะชักเพิ่มเติม และการตรวจร่างกายที่จะทำเพื่อนำไปสู่การวินิจฉัยโรค ในข้อสอบตอนที่สองอาจารย์นำเสนอประวัติและผลการตรวจร่างกายเพิ่มเติมให้ แล้วมีโจทย์ถามถึงการวินิจฉัยโรค และการส่งตรวจทางห้องปฏิบัติการที่เหมาะสม ในข้อสอบตอนที่สามอาจารย์นำเสนอข้อมูลการวินิจฉัยโรคของผู้ป่วยพร้อมผลการตรวจทางห้องปฏิบัติการ แล้วถามแนวทางการรักษา การนำเสนอข้อสอบในลักษณะนี้ในข้อสอบหลายข้อมีการนำเสนอข้อมูลของโจทย์ซ้ำเดิมและค่อย ๆ เพิ่มข้อมูลขึ้นในทุกขั้นตอน ในข้อสอบตอนที่สองก็นำเสนอข้อมูลที่เสนอในตอนหนึ่งกับสอง ในข้อสอบตอนที่สามก็นำเสนอข้อมูลที่เสนอในตอนหนึ่ง สอง และ สาม ซึ่งเมื่อผ่านการสอบไปหลายตอนจะมีข้อมูลสะสมจำนวนมากที่ผู้สอบต้องอ่าน การนำเสนอข้อสอบที่มีประสิทธิภาพมากกว่าควรมีการสรุปข้อมูลอย่างเหมาะสม ในข้อสอบตอนที่สาม หากได้ข้อสรุปการวินิจฉัยโรคแล้ว จะถามแนวทางการรักษาโรค อาจารย์ควรพิจารณาตัดข้อมูลประวัติและการตรวจร่างกายออก หากการสั่งการรักษาจำเป็นต้องทราบข้อมูลจากประวัติ หรือการตรวจร่างกายบางอย่าง เช่น น้ำหนักตัว หรือ โรคร่วมที่ส่งผลต่อการ

วางแผนการรักษา ก็ให้นำเสนอเฉพาะข้อมูลที่ส่งผลต่อการตัดสินใจในขั้นตอนนั้นเท่านั้น

การนำเสนอข้อสอบอัตนัยประยุกต์ด้วยระบบคอมพิวเตอร์ก็เป็นอีกแนวทางหนึ่งที่คณะแพทยศาสตร์ศิริราชพยาบาลเห็นความสำคัญ และได้ดำเนินการพัฒนาอย่างต่อเนื่อง คณะแพทยศาสตร์ศิริราชพยาบาลมีความพร้อมในการพัฒนาด้านนี้มากพอสมควร เนื่องด้วยมีห้องคอมพิวเตอร์ที่มีจำนวนคอมพิวเตอร์มากพอที่จะจัดให้ผู้เข้าสอบทุกคนมีจอคอมพิวเตอร์ส่วนตัว มีการวางระบบเครือข่ายให้มีการส่งผ่านข้อมูลระหว่างเครื่องคอมพิวเตอร์ได้ดี และมีความเสถียรของระบบพอสมควร มีการวางมาตรการรักษาความปลอดภัยของข้อมูลในระบบที่ดี สามารถควบคุมการเข้าออกของข้อมูลจากระบบเครือข่ายคอมพิวเตอร์ได้ จึงส่งผลให้คณะได้ปรับปรุงแบบการจัดสอบอัตนัยประยุกต์จากระบบสอบด้วยข้อสอบกระดาษมาเป็นการนำเสนอข้อสอบบนจอคอมพิวเตอร์ ตั้งแต่ปีการศึกษา 2552 ซึ่งจากการสำรวจความเห็นของนักศึกษาผู้เข้าสอบได้รับการตอบรับดีมาก นักศึกษาพึงพอใจกับการสอบในระบบนี้ในระดับมากถึงมากที่สุด อย่างไรก็ตามระบบการสอบนี้ยังมีโอกาสที่จะพัฒนาให้ดีขึ้นได้อีก ในระบบการจัดสอบปัจจุบันของคณะฯ ยังคงเป็นรูปแบบที่ไม่ได้ใช้คอมพิวเตอร์อย่างเต็มรูปแบบ ยังคงให้ผู้สอบเขียนคำตอบลงในกระดาษคำตอบและเก็บกระดาษในตอนท้ายของการสอบในแต่ละสถานการณ์ผู้ป่วย การใช้ประโยชน์ของคอมพิวเตอร์ในการสอบปัจจุบันเน้นไปในการนำเสนอข้อมูลที่ทำให้ผู้สอบสามารถเห็นภาพถ่ายรังสี ภาพการตรวจทางห้องปฏิบัติการ แผนภาพ ตาราง รวมถึงรูปของผู้ป่วยได้ โดยผู้สอบทุกคนเห็นภาพที่มีความละเอียดสูงเท่าเทียมกัน และทำให้การบริหารการสอบทำได้มีประสิทธิภาพมากขึ้น ตัดปัญหาผู้สอบลักลอบเปิดดูข้อสอบในตอนต่อไปล่วงหน้า หรือทำข้อสอบในบางตอนเกินเวลา การแสดงเวลาที่เหลือในการทำข้อสอบแต่ละตอนบนหน้าจอทำให้ผู้สอบบริหารเวลาในการทำข้อสอบได้ดีขึ้น

ระบบจัดสอบอัตนัยประยุกต์ด้วยคอมพิวเตอร์อย่างเต็มรูปแบบที่ไม่ต้องมีการเขียนตอบในกระดาษเลยนั้นมีการจัดทำในต่างประเทศ^{12,23} แต่ต้องยอมรับว่าการ

สร้างระบบการทดสอบอัตโนมัติประยุกต์ด้วยคอมพิวเตอร์อย่างเต็มรูปแบบนั้นเป็นงานที่ซับซ้อนและมีความท้าทายหลายอย่าง ทั้งในด้านผู้จัดสอบ ระบบเครือข่ายคอมพิวเตอร์ และผู้เข้าสอบ ในอนาคตอันใกล้นี้ทางฝ่ายการศึกษาฯยังไม่มีแนวทางที่จะพัฒนาการสอบอัตโนมัติประยุกต์เป็นระบบคอมพิวเตอร์อย่างเต็มรูปแบบ ด้วยข้อจำกัดสำคัญสามประการคือ ความพร้อมของผู้เข้าสอบ ความพร้อมของผู้ตรวจข้อสอบ และความพร้อมของระบบการสื่อสารระหว่างผู้ใช้กับคอมพิวเตอร์ กล่าวคือ ผู้เข้าสอบจำนวนไม่น้อยยังไม่คุ้นเคยกับการพิมพ์คำตอบที่มีทั้งภาษาไทยและภาษาอังกฤษผสมกันภายในเวลาที่จำกัด อาจารย์ผู้ตรวจข้อสอบจำนวนไม่น้อยยังไม่สะดวกที่จะทำการตรวจข้อสอบและกรอกคะแนนบนหน้าจอคอมพิวเตอร์ในสถานที่และเวลาที่กำหนด และการสร้างระบบการสื่อสารระหว่างคอมพิวเตอร์กับผู้ใช้ให้ทั้งนำเสนอข้อมูลผู้ป่วยที่มีรายละเอียดมาก พร้อมกับตอบรับคำตอบที่มีทั้งอักษร ตัวเลข และสัญลักษณ์พิเศษ ที่ผู้เข้าสอบจะพิมพ์เข้าเครื่องพร้อม ๆ กันหลายร้อยคนโดยมีการควบคุมเวลาอย่างรวดเร็วด้วย ยังเป็นสิ่งที่ทำได้ยากในระบบเครือข่ายคอมพิวเตอร์ในปัจจุบัน ดังนั้นในอนาคตอันใกล้นี้ทิศทางการพัฒนาระบบการทดสอบข้อสอบอัตโนมัติประยุกต์คงยังมุ่งเน้นไปในรูปแบบการนำเสนอข้อสอบผ่านจอภาพคอมพิวเตอร์ ร่วมกับการเขียนตอบในกระดาษคำตอบอยู่

แต่ถึงแม้ว่าจะคงการทดสอบอัตโนมัติในรูปแบบผสมผสานเช่นนี้ ผู้นิพนธ์ก็ยังเห็นว่ามีสิ่งที่ระบบการนำเสนอข้อมูลผ่านจอคอมพิวเตอร์สามารถทำให้ดีขึ้นได้ เช่นการทำให้ภาพมีรายละเอียดสูงขึ้น การเปิดโอกาสให้ผู้เข้าสอบสามารถขยายภาพเพื่อดูรายละเอียดในบางส่วน การปรับรูปแบบการนำเสนออักษร และพื้นหลังของจอภาพให้ผู้เข้าสอบอ่านข้อมูลได้ง่ายขึ้น เป็นต้น ซึ่งสิ่งเหล่านี้จะได้รับการศึกษาหาแนวทางในการพัฒนาในการสอบอัตโนมัติประยุกต์ครั้งต่อไป แต่อย่างไรก็ตามด้วยศักยภาพของระบบการทดสอบในปัจจุบัน ผู้นิพนธ์ยังมีความเห็นว่าอาจารย์ผู้ออกข้อสอบก็ยังไม่สามารถใช้ศักยภาพของระบบอย่างเต็มที่ ยังมีข้อสอบหลายข้อที่ใช้การบรรยายสิ่งตรงจหน้าที่สามารถมองเห็นเป็นภาพได้

แต่นำมาเขียนเป็นอักษรบรรยายสิ่งตรงจหน้าดังกล่าวซึ่งทำให้ผู้เข้าสอบไม่ได้คิดวิเคราะห์และแปลผลการตรวจด้วยตนเอง แนวทางการพัฒนาข้อสอบอัตโนมัติประยุกต์ที่สมควรได้รับการส่งเสริมในระบบการทดสอบปัจจุบันคือการใช้สื่อที่เป็นรูปภาพในข้อสอบให้มากขึ้น ไม่ว่าจะเป็นการตรวจร่างกายจากการดู การดูภาพรังสี การดูคลื่นไฟฟ้าหัวใจ การดูสิ่งส่งตรวจด้วยกล้องจุลทรรศน์ ล้วนแล้วแต่ควรนำเสนอเป็นรูปภาพทั้งสิ้น

บทสรุป

ในบทความนี้ผู้นิพนธ์ได้กล่าวถึงความรู้พื้นฐานในการสร้างข้อสอบอัตโนมัติประยุกต์ โดยได้สรุปลักษณะพื้นฐานของข้อสอบอัตโนมัติประยุกต์ พัฒนาการของข้อสอบประเภทนี้จากรูปแบบการจัดการปัญหาผู้ป่วยเป็นการแก้ปัญหาสำคัญ มีการสรุปขั้นตอนสำคัญในการสร้างข้อสอบอัตโนมัติประยุกต์ห้าขั้นตอน ได้แก่ (1) ตั้งกลุ่มพัฒนาข้อสอบ, (2) เลือกรูปแบบทางคลินิก, (3) กำหนดปัญหาสำคัญ, (4) เขียนโจทย์คำถาม, และ (5) กำหนดเกณฑ์การให้คะแนน และในตอนท้ายได้มีการนำหลักการพัฒนารูปแบบข้อสอบต่าง ๆ ที่กล่าวมาแล้วมาวิเคราะห์สถานการณ์การทดสอบอัตโนมัติประยุกต์สำหรับนักศึกษาแพทย์คณะแพทยศาสตร์ศิริราชพยาบาลและเสนอแนะแนวทางในการพัฒนาคุณภาพการสอบอัตโนมัติประยุกต์สี่แนวทาง ได้แก่ (1) เนื้อหาข้อสอบ, (2) รูปแบบคำถาม, (3) จำนวนสถานการณ์ผู้ป่วย, และ (4) การนำเสนอข้อสอบ ผู้นิพนธ์เชื่อมั่นว่าหากการทดสอบอัตโนมัติประยุกต์ได้รับการพัฒนาอย่างเหมาะสมจะนำไปสู่การประเมินความรู้ และทักษะการตัดสินใจดูแลผู้ป่วยในระดับคลินิกที่มีประสิทธิภาพ

เอกสารอ้างอิง

- Downing SM. Assessment of knowledge with written test forms. In: Norman GR, van der Vleuten C, Newble DI, editors. International handbook of research in medical education. Dordrecht: Kluwer Academic Publishers, 2002:647 - 72.
- Epstein RM. Assessment in medical education. *New Engl J Med* 2007;356:387-96.
- The Board of Censors of the Royal College of General Practitioners. The modified essay question. *J Roy Coll Gen Practit* 1971;21:373-6.
- Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005;39: 1188 -94.

5. McGuire CH, Babbott D. Simulation technique in the measurement of problem solving skills. *J Educ Meas* 1967;4:1-10.
6. จินตนา ศิรินาวิน, สาธิต วรรณแสง. ทักษะทางคลินิก, พิมพ์ครั้งที่ 2. กรุงเทพฯ: หมอชาวบ้าน, 2549.
7. Hodgkin K, Knox JDE. Problem centered learning. London, United Kingdom: Churchill Livingstone, 1975.
8. Stratford P, Pierce-Fenn H. Modified essay question. *Phys Ther* 1985; 65(1075-9).
9. Feletti GI, Smith EK. Modified essay questions: Are they worth the effort? *Med Educ* 1986;20:126 - 32.
10. Rabinowitz HK. The modified essay question: An evaluation of its use in a family medicine clerkship. *Med Educ* 1987;21:114-8.
11. Wallerstedt S, Erickson G, Wallerstedt SM. Short answer questions or modified essay questions - More than a technical issue. *Int J Clin Med* 2012;3:28-30.
12. Lim EC, Seet RC, Oh VMS, Chia B, Aw M, S Q, et al. Computer-based testing of the modified essay question: The Singapore experience. *Med Teach* 2007;29:e261-8.
13. Norman G, Bordage G, Curry L, et al. Review of recent innovations in assessment. In: Wakeford R, editor. Directions in clinical assessment: Report of the Cambridge conference on the Assessment of Clinical competence. Cambridge: Office of the Regius Professor of Physic, Cambridge University School of clinical Medicine, 1985:8-27.
14. Page G, Bordage G. The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Acad Med* 1995;70:104-10.
15. Neufeld VR, Norman GR, Barrows HS, Feightner JW. Clinical problem solving by medical students: A longitudinal and cross-sectional analysis. *Med Educ* 1981;15:315-22.
16. Perkins DN, Salomon G. Are cognitive skills context-bound? *Educ Researcher* 1989;18:16-25.
17. van der Vleuten CPM, Swanson DB. Assessing clinical skills with standardized patients: The state of the art. *Teach Learn Med* 1990;2 (58-76).
18. Eva KW. On the generality of specificity. *Med Educ* 2003;37(7): 587-88.
19. Bordage G, Page G. An alternate approach to PMPs, the key feature concept. In: Hart I, Harden R, editors. Further developments in assessing clinical competence. Montreal: Can-Heal Publications, 1987:57-75.
20. Page G, Bordage G, Allen T. Developing key features problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;70:194-201.
21. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ* 2006;40:618-23.
22. Hambleton RK, Pitoniak MJ. Setting performance standards. In: Brennan RL, editor. Educational measurement, 4th ed. Westport, CT: Praeger publishers, 2006:433-70.
23. Federation of State Medical Boards of the United States, National Board of Medical Examiners. USMLE Step 3: Content description and general information, Available from http://www.usmle.org/pdfs/step-3/2014content_Step3.pdf. June 2014.

ตามปกหน้าเวชบันทึกศิริราช ปีที่ 7 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2557 หน้า 74-83 เรื่อง
“หน้ากากครอบกล่องเสียง Laryngeal Mask Airway (LMA)” โดย อรุณทัตย์ ศิริอัศวกุล

ขอแก้ไขเป็น

เวชบันทึกศิริราช

ปีที่ 7 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2557 หน้า 74-83 เรื่อง

“หน้ากากครอบกล่องเสียง Laryngeal Mask Airway (LMA)” โดย อังศุมาศ หวังดี

และได้ทำการแก้ไข pdf เรียบร้อยแล้ว

the metric of medical education

A practical guide to assessing clinical decision-making skills using the key features approachELIZABETH A FARMER¹ & GORDON PAGE²

AIM This paper in the series on professional assessment provides a practical guide to writing key features problems (KFPs). Key features problems test clinical decision-making skills in written or computer-based formats. They are based on the concept of critical steps or 'key features' in decision making and represent an advance on the older, less reliable patient management problem (PMP) formats.

METHOD The practical steps in writing these problems are discussed and illustrated by examples. Steps include assembling problem-writing groups, selecting a suitable clinical scenario or problem and defining its key features, writing the questions, selecting question response formats, preparing scoring keys, reviewing item quality and item banking.

CONCLUSION The KFP format provides educators with a flexible approach to testing clinical decision-making skills with demonstrated validity and reliability when constructed according to the guidelines provided.

KEYWORDS *decision making; clinical competence/*standards; educational measurement/*methods/standards; problem-based learning; *education, medical; questionnaires; Canada.

Medical Education 2005; 39: 1188–1194

doi:10.1111/j.1365-2929.2005.02339.x

¹Royal Australian College of General Practitioners, Melbourne, Victoria, Australia

²Department of Medicine, Division of Educational Support and Development, College of Health Disciplines, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence: Associate Professor Elizabeth A Farmer BSc, MBBS, PhD, FRACGP, Department of General Practice, Level 7, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia.
Tel: 00 61 88 204 5606; Fax: 00 61 88 276 3305;
E-mail: liz.farmer@flinders.edu.au

INTRODUCTION

In this article, we introduce the concept of a key feature, which is the cornerstone of a problem format known as the key features problem used in written examinations of clinical decision-making skills.¹ We then focus on practical guidance in creating key features problems to test clinical decision-making skills at both undergraduate and postgraduate levels.

Bordage and Page² first introduced the term 'key feature' in 1987, following a critical analysis of research on the nature and assessment of clinical decision-making skills published in 1985.³ At that time, most assessments of these skills used small numbers of lengthy clinical problems (sometimes only 1), on the premise that the skills were generic and largely independent of the factual knowledge and procedural skills demanded in any particular problem.⁴ The most popular such assessment format was the patient management problem (PMP), a written problem which consisted of a clinical scenario, followed by sections of items which elicited candidates' responses in relation to history taking, physical examination, investigations and diagnosis. One PMP could take up to 90 minutes to complete.⁵

Although its high authenticity and face validity made it popular, it became clear that the PMP format had serious drawbacks. First, the reliability of the test was very low³ and it was evident that content specificity was just as much a factor in testing clinical decision-making skills as in all other areas of clinical competence. In practical terms, this required many hours of testing in order to obtain a reliable result. In addition, the scoring of PMPs often rewarded thoroughness of data gathering, rather than ability to make appropriate decisions. Moreover, the expected differences in performance between junior and experienced doctors were not found. Finally, scores

Overview

What is already known on this subject

The value of testing clinical decision-making skills using the key features problem format has been increasingly recognised over the last decade. The approach is feasible and offers high reliability and support for face and content validity if items are well constructed.

What this study adds

The key features approach is gaining interest amongst educators in health sciences curricula; however, few have practical experience in writing high quality problems. In this paper we present a practical guide to writing and scoring key features problems in health sciences. Various attributes of the approach are highlighted, including the flexibility of the format in testing decision-making skills in a wide variety of domains.

Suggestions for further research

Further examination of predictive validity and effects on candidates' preparation for testing would be valuable.

on PMP tests correlated highly with scores on knowledge tests, suggesting that they added little additional measurement information.^{4,6}

A NEW APPROACH

In order to overcome these difficulties, Page and Bordage⁶ suggested that, in any clinical case, there are a few unique, essential elements in decision making which, alone or in combination, are the critical steps in the successful resolution of the clinical problem. They labelled these elements 'key features'.² This concept led to the creation of a new test of clinical decision-making skills, which elicited candidates' responses concerning only the critical steps in the resolution of each problem – the problem's key features. Testing only critical steps enabled candidates to be tested on a much larger number of clinical problems than was the case with the PMP format. The new test format was called the

'key features problem' (KFP) and was shown to have a potential reliability of 0.8 in 4 hours of testing.⁶

The KFP format proposed by Page and Bordage⁶ also added to other written test formats in that it allowed more than 1 correct answer as required by the question. These involved either 1 or more very brief written answers, or 1 or more items selected from a long list. The flexibility in allowing for more than 1 correct answer often mirrors real-life practice more closely than is possible in single answer written formats, such as multiple-choice questions (MCQs) or extended matching questions. In addition, the KFP format also maintained the advantages of the longitudinal nature of the PMP format in that following a problem through various stages enabled testing of candidates' clinical decisions over the course of a clinical scenario. This is similar to other sequential formats, such as the modified essay question format, and again mirrors real-life clinical practice more closely than is possible in more basic test constructions such as MCQs. Key features problem test formats may be presented in either paper-based or computer-based formats. The latter suits high volume, high stakes testing, and allows for low cost incorporation of pictures into the problems, but overall is more expensive to deliver.

Key features problems are now used in a variety of testing situations. While the reliability of the format is good, in high stakes testing the format is presented as part of a suite of assessment approaches. For example, the Medical Council of Canada uses a 4-hour KFP format test in the Part 1 Qualifying Examination for licensure, together with a 3.5-hour MCQ test. Candidates for the Royal Australian College of General Practitioners (RACGP) Fellowship Examination for certification sit a 3-hour KFP paper, together with a 4-hour written test and a 3-hour objective structured clinical examination (OSCE). Key features problem formats are also employed by the University of Toronto as part of its internal examinations for medical students and by the American College of Physicians in the Medical Knowledge Self-Assessment Program (MKSAP) for continuing medical education purposes.

SAMPLE KEY FEATURES PROBLEM:

—DIARRHOEA

The following problem (Fig. 1) has been reproduced from a guide to writing KFPs prepared for the

A 35-year-old mother of 3 presents to your office at 17.00 hours with complaints of severe, watery diarrhoea. On questioning, she indicates that she has been ill for about 24 hours. She has had 15 watery bowel movements in the past 24 hours, has been nauseated, but not vomited. She works during the day as a cook in a longterm care facility but left work to come to your office. On her chart, your office nurse notes a resting blood pressure of 105/50 mmHg supine (a pulse of 110/minute), 90/40 standing, and an oral temperature of 36.8 °. On physical examination, you find she has dry mucous membranes and active bowel sounds. A urinalysis (urine microscopy) was normal, with a specific gravity of 1.030.

1 What clinical problems would you focus on in your immediate management of this patient? List up to 3

2 How should you treat this patient at this time? Select up to 3

- 1 Antidiarrhoeal medication
- 2 Antiemetic medication
- 3 Intravenous 0.9% NaCl
- 4 Intravenous 2/3–1/3
- 5 Intravenous gentamicin
- 6 Intravenous metronidazole
- 7 Intravenous Ringer lactate
- 8 Nasogastric tube and suction
- 9 Nothing by mouth
- 10 Oral ampicillin
- 11 Oral chloramphenicol
- 12 Oral fluids
- 13 Rectal tube
- 14 Send home with close follow-up
- 15 Surgical consultation
- 16 Transfer to hospital

3 After management of the patient's acute condition, what additional measures, if any, would you take? Select up to 4 or select #11, none, if none are indicated

- 1 Avoid dairy products
- 2 Colonoscopy
- 3 Enteric precautions
- 4 Gastroenterology consultation
- 5 Give immune serum globulin to patients at longterm care facility
- 6 Infectious disease consultation
- 7 Notify Public Health Authority
- 8 Stool cultures
- 9 Strict isolation of patient
- 10 Temporary absence from work
- 11 None

Figure 1 A sample key features problem.

Medical Council of Canada.⁷ The key features tested by the questions are:

- 1 recognise dehydration (tested) and its level of severity (not tested);

- 2 manage dehydration appropriately, and
- 3 evaluate the possible communicability of the underlying disease (family or hospital spread, possible common source).

Each question directly tests 1 of these key features, and each challenges the candidate to apply his or her knowledge in making clinical decisions.

DEVELOPING KEY FEATURES PROBLEMS

The first section of this article highlighted the rationale, nature and main advantages of the key features approach. The sections that follow outline a practical guide to the steps involved in developing KFPs, which build upon the guidelines for writing KFPs presented by Page and Bordage.¹

Assembling problem-writing groups

Both face validity and content validity require the use of problem writers whose backgrounds and clinical expertise are pertinent to the context of the examination. In Australia, for example, the RACGP employs general practitioners from diverse metropolitan, rural and remote practices across the country, who work in small guided groups to create draft KFPs for use in part of the fellowship examination.⁸ This ensures that the problems written are well grounded in practice and experience and represent a wide range of real-life Australian general practice contexts. Using the writing process outlined below, problems are written so that they do not represent mere abstractions or generalisations from textbooks.⁹ This is an important step in supporting the content validity of the format and applicability to real-life practice, as perceived by the candidate group.¹⁰

Selecting a problem, defining its key features

First, problem writers are asked to select a clinical problem (e.g. diarrhoea), usually selected from a blueprint for a key features examination. They are asked to think of several instances (real cases) of the problem in practice. Relative to these cases, they are then asked to address the most important question they face as a problem writer: 'What are the essential steps in the resolution of this problem?'⁷ This fundamental question prepares writers to concentrate on only the most critical decisions within each case – the problem's key features. It is essential to differentiate between decisions or steps that are appropriate, but not critical, and those that *must* be present. Coming to grips with this distinction is the

single biggest issue for novice writers. This step usually requires discussion amongst a small group or panel of writers to clarify which steps are critical and achieve consensus. Secondary considerations which can guide the identification of a problem's key features involve asking problem writers to also identify the elements or steps most likely to result in errors by candidates at particular levels of training (e.g. graduating medical students), and to identify the difficult aspects of the identification and management of the problem in clinical practice.

Key features are unique for each clinical problem, and may pertain to any component of the work-up and management of a case; for example, in initial data gathering and diagnostic steps, in longterm management, or in prevention of complications. Key features focus on clinical decisions (e.g. 'include depression in a differential diagnosis') or clinical actions (e.g. 'elicit risk factors', 'order a mammogram') where the clinical action is an expression of a clinical decision. Figure 2 illustrates typical decisions or actions tested in KFPs.

- Elicit history or reasons for patient request
- Interpret symptoms
- Seek critical physical findings
- Interpret physical findings
- Make a diagnosis or differential
- Order investigations to confirm or deny differential diagnoses
- Specify management goals or decisions
- Prescribe drugs
- Specify follow-up

Figure 2 Critical clinical decisions or actions tested in KFPs.

A final component of a key feature is a qualifier that may reflect such issues as the urgency of a decision (e.g. 'What *initial* action...?'), or a decision-making priority (e.g. 'What are the *most important*...?'). Figure 3 presents some common qualifiers.

- Immediate
- Initial
- Longterm
- Definitive
- Urgent
- Most important
- Most likely
- Must not miss

Figure 3 Common qualifiers in key features.

It is important to note that key features may pertain to a broad range of clinical decisions in addition to the biomedical. Key features problems can be constructed to assess ethical, medico-legal, population, preventive and organisational decisions, and in a range of health care settings. This flexibility is a useful attribute of KFP formats in contrast to the more limited multiple-choice and extended matching approaches.

Following their discussion of key features, the problem writers select 1 case for development into a problem scenario and related questions. The clinical scenario for the problem usually begins by stating a patient's age, gender and setting for the encounter. If the key features for that problem focus on the diagnostic component of the problem, the case scenario is often brief (e.g. patient demographics, presenting complaint and limited clinical information). Where the KFP focuses on the management of the problem, the case scenario is typically longer and includes laboratory and diagnostic information. The KFP format is flexible in that additional clinical information can be inserted between questions. This sequential format enables the problem to be followed longitudinally. This attribute allows writers to produce realistic scenarios that evolve over time as required. In this respect, the format is similar to the flexibility found in other sequential formats, such as the modified essay question. Figure 4 gives some examples of the kinds of clinical scenarios that lend themselves to the KFP approach.

- A reason for attendance (e.g. chest pain, check-up, follow-up)
- A request (e.g. sick note, preventive care)
- Symptoms (e.g. cough)
- Signs (e.g. abdominal tenderness)
- Results (e.g. biochemistry, imaging, haematology, audiology, ECG, spirometry)
- Photographs (e.g. clinical signs, rashes)
- Complications of therapy or management

Figure 4 Typical elements in KFP clinical scenarios.

Writing the questions

With the key features defined and the case scenario written, the next step in KFP development is to write the questions that test those key features. Most KFPs consist of a case scenario, typically followed by 2 or 3 questions, each question testing 1 or more key

features. The questions request that candidates record their clinical decisions, which, depending upon the problem's key features, can relate to data gathering (e.g. 'What investigations would you order at this consultation?'), diagnosis ('What are the most likely differential diagnoses?'), management ('What are your longterm management steps?'), etc. Most questions have several answers, which comprise the critical steps in resolving this specific problem. The number of answers may vary from 1 to 10; typically there are 3 to 5.

Selecting question formats

Two question formats are used in KFPs. These are the write-in (WI) format, where candidates supply their responses in very short note form (e.g. they write in 'insulin-dependent diabetes', or 'prescribe penicillin'), and the short menu (SM) format, where candidates select responses from a list of prepared options. The length of the options list varies and may contain up to 25 items. To reduce guessing effects, the list must contain all correct responses plus common misconceptions or likely mistakes. In practice, to reduce cueing, this requires at least 4 or 5 incorrect options for each correct item.

Write-in questions must be marked by hand, whereas SM questions may be marked by computer. The WI question is strictly limited to very short notes or single words, in contrast to the modified essay or short answer question formats, thereby reducing marking time to the minimum. While the feasibility of WI questions could be a problem, data from the Medical Council of Canada and the RACGP suggest that WI formats are more effective in identifying weaker candidates and are more discriminating.¹¹ In addition, it is often harder to write sequential questions purely in SM formats because of backward cueing of candidates to correct answers. Therefore, most KFPs continue to contain both formats.

Specifying the number of required answers

Each question must contain an instruction that stipulates the number of responses to select or supply. Common instructions are:

- write, in note form only, one (1)...
- select up to 'x'...
- select 'x'...
- select as many as are appropriate, and
- select none if none are indicated.

PREPARING SCORING KEYS

The scoring key for a question consists of the list of correct and incorrect responses, and scores to be assigned to each response.

Some scoring keys can contain only a single required response, such as the scoring key for question 1 of the diarrhoea problem shown in Fig. 1 (Fig. 5).

Score	Response	Synonyms
1	Dehydration	Hypovolaemia fluid loss fluid depletion
0	Listing more than 3 items	

Figure 5 Scoring key for question 1 of the diarrhoea problem shown in Fig. 1.

To emphasise that candidates must not give more than the required number of responses to a question, a forfeit is applied if this occurs. In Fig. 5, up to 3 answers were specified. A candidate who provides say, 4 answers, will receive no marks for the question.

Other scoring keys contain several responses clustered on the basis of logical considerations regarding the correct clinical actions to be taken. A simple scoring key for question 3 of the diarrhoea problem is shown in Fig. 6.

This scoring key illustrates a partial credit system of scoring, where a weight is assigned to each response – in this case the same weight of 1 mark to each response.

Score	Correct responses
1 each	# 3 Enteric precautions # 8 Notify Public Health Authority # 11 Stool cultures # 13 Temporary absence from work
0	# 5 Give immune serum globulin to patients at longterm care facility # 12 Strict isolation of patient <i>or</i> Selecting more than 4 items

Figure 6 Scoring key for question 3 of the diarrhoea problem shown in Fig. 1.

Specifying different scores for responses allows for the instances where problem writers regard some correct answers as more important clinically than others. Starting with a default option of each correct answer scoring equally, (e.g. 1 point), more important answers may be weighted more highly (e.g. be awarded 2 or even 3 points). Simple weighting systems are preferable, as more complex systems do not improve reliability. Similarly, negative marking is not used because it does not contribute to reliability and may discriminate between students simply on the basis of their risk-taking behaviour.¹² However, an especially important answer can be specified as 'must be present'. In this case a penalty is applied such as 'no marks for the question if answer not present'. Similarly, a dangerous or negligent response (e.g. unnecessary invasive investigation, unnecessary or harmful treatment) may result in the candidate forfeiting the marks for the question involved, no matter what other responses the candidate makes to that question. Items 5 and 12 in the scoring key shown in Fig. 6 are examples of such actions. Such a penalty, if applied, results in the forfeit of marks only for the relevant question within a KFP. In most cases, where a problem consists of 2 or 3 questions, this penalty results in the forfeit of half or a third of the total marks for that problem. Whether or not such an approach is used depends on the views of the examining body and possibly partly on the stakes associated with the examination.

Total examination scores are simply the sum of the scores on each problem. Problem scores are the sum of the scores on the questions within the problem. Each problem is given the same weight in the calculation of the total mark. This can be easily achieved by transforming problem scores into a percentage.

VALIDATION AND REFERENCES

With questions and answer keys defined, the next step is their validation. Validation entails piloting the problem with discussion, review and editing by colleagues new to the problem, and confirmation of the correctness of answers through reference to suitable literature. Markers particularly appreciate evidence from the literature if questions test a new or rapidly developing area. This process is cited as enjoyable and challenging by writers, and the lively debate and sharing of clinical practice contributes to writers' own continuing education.

COMPUTERISED PRESENTATION OF KFP FORMATS

Presenting KFP in a computerised format offers 2 immediate benefits: ease of presentation of high quality pictorial material such as photographs and imaging, and a mechanism to prevent backward cueing if additional clinical information is given between questions. However, this approach requires additional resources.

QUALITY ASSURANCE ISSUES IN ITEM DEVELOPMENT

Problems that perform well can be maintained in an item bank where the performance of a problem in each examination in which it is used may be recorded. Similarly, question writers may receive feedback on the performance of a problem, and may be involved in review of their problems after use. Candidate feedback is another important source of quality assurance.

STANDARD SETTING OF KFP FORMATS

The issues of standard setting for high stakes KFP examinations are comparable to those in other written tests. The Medical Council of Canada uses the modified Angoff method while the RACGP currently employs a new approach, the Angoff at question level (AQL) method. These methods require multiple judges and are based on the concept of the borderline candidate as presented by Norcini in a previous article in the series *the Metric of Medical Education*.¹³

CONCLUSION

Writing key features problems is challenging and enjoyable. Following the steps in this guide will help ensure that KFP examination papers possess high levels of face and content validity and demonstrate levels of test score reliability that are acceptable for making decisions about individual candidates' clinical decision-making ability.

Contributors: EAF and GP conceived the paper. Both authors contributed substantially to writing and revisions. EAF took responsibility for finalising the manuscript.

Acknowledgement: we thank Brian Jolly for his helpful comments on earlier drafts of the manuscript.

Funding: there was no external funding for this manuscript.

Conflicts of interest: none.

Ethical approval: not required.

REFERENCES

- 1 Page G, Bordage G, Allen T. Developing key features problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;**70**:194–201.
- 2 Bordage G, Page G. An alternate approach to PMPs, the key feature concept. In: Hart I, Harden R, eds. *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal Publications 1987;57–75.
- 3 Norman G, Bordage G, Curry L *et al.* Review of recent innovations in assessment. In: Wakeford R, ed. *Directions in Clinical Assessment. Report of the Cambridge Conference on the Assessment of Clinical Competence*. Cambridge: Office of the Regius Professor of Physic, Cambridge University School of Clinical Medicine, Addenbrooks Hospital 1985;8–27.
- 4 van der Vleuten C, Newble DI. How can we test clinical reasoning? *Lancet* 1995;**345**:1032–4.
- 5 McGuire CH, Solomon LM, Bashook PG. *Construction and Use of Written Simulations*. New York: Psychological Corporation of Harcourt, Brace, Jovanovich 1976.
- 6 Page G, Bordage G. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;**70**:104–10.
- 7 Page G. *Writing Key Feature Problems for the Clinical Reasoning Skills Examination: a Guide for CRS Committee Members in their Understanding and Preparation of Key Feature Problems*. Ottawa: Medical Council of Canada 1999.
- 8 Farmer EA. *Writing key feature problems for general practice*. Melbourne: Royal Australian College of General Practitioners 1998.
- 9 Jolly B, Spencer J. Letter to the editor: reply from the authors. *Med Educ* 2003;**37**(5):472.
- 10 Farmer EA, Joske FM, Lew SR, McDonald EA, Page GG. Performance of candidates on key features problems in the certification examination for Australian general practice. [Abstract.] In: *Proceedings of the 10th International Ottawa Conference on Medical Education*. Ottawa, Canada 2002.
- 11 Page G, Farmer E, Spike N, McDonald E. The use of short answer questions in the key features problems in the Royal College of General Practitioners Fellowship examination. Combining marks, scores and grades. [Abstract.] In: *Proceedings of the 9th International Ottawa Conference on Medical Education*. Cape Town, South Africa 2000.
- 12 Fowell SL, Jolly B. Reviewing common practices reveals some bad habits. *Med Educ* 2000;**34**:785–6.
- 13 Norcini JJ. Setting standards on educational tests. The metric of medical education series. *Med Educ* 2003;**37**:464–9.

Received 12 November 2004; editorial comments to authors 7 December 2004, 24 June 2005; accepted for publication 29 July 2005

เอกสารประกอบการอบรม



24 July 2020

รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์

หัวข้อ : Standard setting

การตั้งเกณฑ์ผ่านการสอบ

นพ.เชิดศักดิ์ ไอรมณีรัตน์

ภาควิชาศัลยศาสตร์

คณะแพทยศาสตร์ศิริราชพยาบาล

Standard

- A score that is set to be a boundary between those who perform well enough on the test (pass) from those who do not (fail).
- Standard = cutpoint

Objectives

- **เมื่อสิ้นสุดการบรรยายแล้ว ผู้เข้าอบรมสามารถ**
 - บอกถึงความสำคัญของการตั้งเกณฑ์ผ่านได้ถูกต้อง
 - บอกถึงขั้นตอนของการตั้งเกณฑ์ผ่านได้ถูกต้อง
 - ยกตัวอย่างวิธีการตั้งเกณฑ์ผ่านได้อย่างน้อยสามวิธี

Outline

- **Basic concepts**
- **Steps in setting standards**
 - The type of standard
 - The method
 - Selecting judges
 - Standard setting meeting
 - Calculate the standards
 - Checking the standards

Basic Concepts

- A standard is an answer to the question, “How much is enough?”
- The classification of examinees into two groups can result in two types of wrong decisions
 - False positive: Passing an examinee who should fail the exam
 - False negative: Failing an examinee who should pass the exam

Judgment

1. Made by qualified judges
2. Meaningful to the persons who are making the decision
3. Made in a way that takes into account the purpose of the test

Steps in Setting Standards

1. Deciding on the type of standard
2. Deciding on the method for setting standards
3. Selecting judges
4. Holding the standard setting meeting
5. Calculating the standards
6. Checking the standards after test

1. Types of Standards

- Absolute standard
- Relative standard

Absolute Standard

- The standard is fixed, based on specific criteria of performance, but may undergo periodic re-evaluation of the standard
- Strengths
 - A standard is known in advance
 - A stable performance level is required to pass the examination => content-related standard
 - Provide clear feedback to examinees
 - Nobody has to fail the exam if their knowledge/skills is adequate for the purpose of the exam.
 - Promote a collaborative learning environment.

Relative Standard

- The standard is set in reference to the group of examinees. The resulting standard may be reasonable providing a representative heterogeneous group.
- Strengths
 - The failure rate is stable, which in some way is easy for curriculum management

2.Methods for Setting Standards

1. Test-centered methods
2. Examinee-centered methods
3. Compromised methods

11

Test-Centered Methods

- The judges set standards by reviewing the test items and provide judgments regarding the “just adequate” level of performance on these items.
 - Angoff’s method
 - Nedelsky’s method
 - Ebel’s method

12

Modified Angoff's Method

- **The judgment**
 - The probability that a borderline examinee would answer the test item correctly
- **The passing score**
 - The sum of all the probability of correct answers for all items on the exam

13

Modified Angoff's Method (2)

Item	Probability
1	0.8
2	0.6
3	0.4
4	0.5
5	0.5
Passing score	2.8

14

Nedelsky's Method

- **The judgment**
 - How many options a borderline examinee can eliminate from choosing in an item
- **The passing score**
 - The probability of correct answer for an item = $1/(\text{the number of options not eliminated})$
 - The passing score of the test = the sum of all the probability of correct answers of all items on the test

18

Nedelsky's Method (2)

Item	A	B	C	D	E	Not eliminated	Probability
1			X	X	X	2	$1/2 = 0.50$
2	X	X				3	$1/3 = 0.33$
3	X					4	$1/4 = 0.25$
4	X		X	X		2	$1/2 = 0.50$
5	X				X	3	$1/3 = 0.33$
Passing score							1.91

19

Ebel's Method

- **The judgment**
 - What is the level of difficulty of an item?
 - Easy/Medium/difficult
 - What is the level of importance of that content in clinical practice?
 - Essential/Important/Acceptable/Questionable
 - The probability that a borderline examinee will answer an item in each category correctly
- **The passing score**
 - The sum of all the probability of correct answers for all items on the exam

Ebel's Method (2)

	Easy	Medium	Difficult
Essential	0.95	0.85	0.80
Important	0.90	0.75	0.60
Acceptable	0.85	0.60	0.40
Questionable	0.55	0.45	0.35

Ebel's Method (3)

Item	Difficulty	Importance	Probability
1	Easy	Essential	0.95
2	Easy	Importance	0.90
3	Difficult	Essential	0.80
4	Difficult	Acceptable	0.40
5	Medium	Acceptable	0.60
Passing score			3.65

Examinee-Centered Methods

- The judges set a standard by reviewing the overall performance of examinees and determine who should pass and who should fail. The scores of examinees are reviewed and the passing score is set based on these judgments
 - Borderline-group method
 - Contrasting-groups method

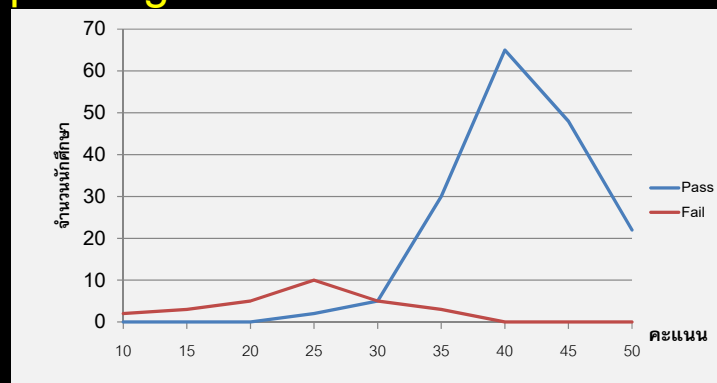
Borderline-Group Method

- **The judgment**
 - Identify examinees who are “borderline”
- **The passing score**
 - The median score of this “borderline group”

21

Contrasting-Groups Method

- **The judgment**
 - Identify examinees who should “pass” and those who should “fail”
- **The passing score**



22

Compromised Method

- **Combining relative and absolute standard setting methods**
 - Hofstee method

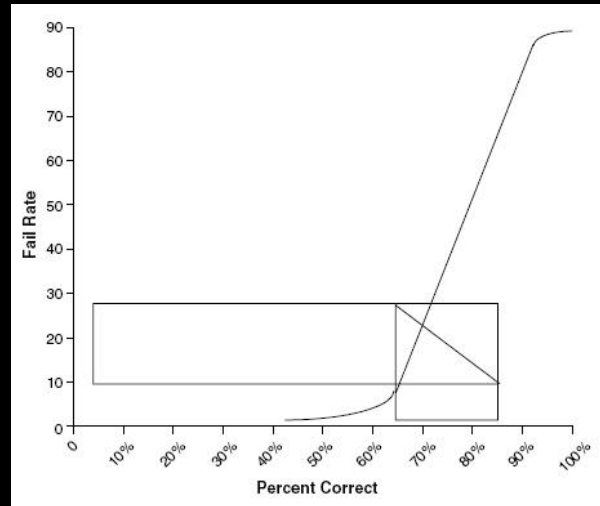
23

Hofstee Method

- **The judgment**
 - Minimum failure rate
 - Maximum failure rate
 - Minimum passing score
 - Maximum passing score
- **The passing score**
 - The intersection of test scores curve with diagonal line drawn from upper left to lower right corner

24

Hofstee Method (2)



25

3. Selecting Judges

- The number of judges
- The qualification of judges

26

4. Standard Setting Meeting

- Discussion of the purpose of the test, the characteristics of examinees, and the nature of competence.
- Explanation of the method and practice before the real standard setting procedure.

27

5. Calculating Standard

- Outliers
- Errors of the cutpoint

28

Do we have to care about error?

- **True score theory**

- Each student has a true score, a hypothetical value representing a score free of error.
- If we test a student repeatedly, the average of the obtained scores would approximate the true score, with a standard deviation of SEM.

29

SEM

$$SEM = SD\sqrt{(1-r)}$$

SD = standard deviation

r = internal consistency reliability

↑SD (more spread of score): higher SEM

↑r (more accurate measures): smaller SEM

30

What should we do with students with an SEM around cut score?

- **False positive: Passing students who should have fail the examination**
- **False negative: Failing students who should have pass the examination**

31

6. Checking Standard

- **Stakeholders' acceptance of the results**
- **Relationship with other markers of competence**
- **Prediction of future performance**

32

Summary

- Steps in setting up a standard
 1. Deciding on the type of standard
 2. Deciding on the method for setting standards
 3. Selecting judges
 4. Holding the standard setting meeting
 5. Calculating the standards
 6. Checking the standards after test

33

Questions & Comments

Cherdsak Iramaneerat
CherdsakIramaneerat@gmail.com

34

Iramaneerat C. Passing standard: Part I [Thai]. Medical Education Pamphlet 2006; 2(1): 3.

วิธีการตั้งเกณฑ์สอบผ่าน (passing standard) (ตอนที่ 1)

เชิดศักดิ์ ไชยมณีรัตน์

เกณฑ์สอบผ่าน (passing standard) คือคะแนนสอบที่น้อยที่สุดที่คณาจารย์ยินยอมให้นักเรียนสามารถสอบผ่าน นักเรียนที่สอบได้คะแนนน้อยกว่าเกณฑ์สอบผ่านจะถูกตัดสินว่าสอบตก การตั้งเกณฑ์สอบผ่านจัดเป็นขั้นตอนที่มีความสำคัญมาก ในการจัดสอบ แต่กลับไม่ได้รับความสนใจเท่าที่ควรในการวัดผลทางแพทยศาสตรศึกษาจำนวนมาก ในบทความนี้ผมขอเสนอเกร็ดความรู้เกี่ยวกับวิธีการตั้งเกณฑ์สอบผ่าน ผมหวังว่าอาจารย์ผู้อ่านจะสามารถนำเกร็ดความรู้นี้ไปใช้พัฒนาคุณภาพของการตั้งเกณฑ์สอบผ่านได้ไม่มากนักน้อยครับ

เกณฑ์สอบผ่านในทางแพทยศาสตรศึกษาจัดว่ามีความสำคัญมากเนื่องจากเกณฑ์สอบผ่านเป็นการแสดงออกถึงมาตรฐานของวิชาชีพที่อาจารย์ยอมรับ เกณฑ์สอบผ่านที่ดีต้องได้รับการตั้งขึ้นโดยใช้ดุลยพินิจของคณาจารย์ผู้เชี่ยวชาญในสาขาวิชานั้นๆ เพื่อรักษามาตรฐานการประกอบวิชาชีพเพื่อให้สังคมได้รับบริการทางการแพทย์ที่มีคุณภาพ ในขณะเดียวกันกับให้ความเป็นธรรมกับนักเรียนผู้สอบ เนื่องจากเกณฑ์สอบผ่านเป็นการแสดงออกถึง "ความยอมรับได้" ในดุลยพินิจของคณาจารย์ผู้เชี่ยวชาญ จึงไม่มีวิธีการทางวิทยาศาสตร์ใดที่จะตัดสินว่าเกณฑ์ที่ตั้งขึ้นนั้นถูกหรือผิด สิ่งที่สำคัญที่สุดในการตั้งเกณฑ์สอบผ่านหาใช่ "ตัวเลข" คะแนนที่จะใช้ตัดสินได้ตก หากแต่เป็น "กระบวนการ" ให้ได้มาซึ่งเกณฑ์ดังกล่าว เกณฑ์สอบผ่านที่ตั้งขึ้นโดยใช้อาจารย์ 1 ท่านเลือกตัวเลข 1 ตัวเลขขึ้นมาโดยไม่ได้พิจารณาถึงข้อสอบหรือนักเรียนผู้สอบ เป็นวิธีการตั้งเกณฑ์ที่ล่อแหลมต่อการถูกวิจารณ์ (และประท้วง) โดยผู้ที่ไม่พอใจในผลสอบ วิธีการตั้งเกณฑ์สอบผ่านที่ดีนั้นต้องมีหลักการและเหตุผลประกอบ และผ่านดุลยพินิจของคณาจารย์ จำนวนของอาจารย์ผู้เชี่ยวชาญที่ต้องใช้ในการตั้งเกณฑ์นั้นขึ้นกับความสำคัญของการสอบนั้นๆ ในการสอบที่มีความสำคัญสูงเช่นการสอบวุฒิปับตรแพทย์ผู้เชี่ยวชาญ แนะนำให้ใช้คณาจารย์อย่างน้อย 6 – 8 ท่าน ในการตั้งเกณฑ์ แต่หากเป็นการสอบเล็กๆ เช่น การทดสอบหลังการสอนกลุ่มย่อย อาจใช้อาจารย์เพียง 1 ท่านก็ได้

การตั้งเกณฑ์สอบผ่านมี 2 ชนิดคือ การตัดสินแบบอิงเกณฑ์ (criterion-referenced standard, absolute standard) และการตัดสินแบบอิงกลุ่ม (norm-referenced standard, relative standard) การตัดสินแบบอิงเกณฑ์ เป็นการตั้งว่า คะแนน เท่าไร จึงจัดว่าผ่านการสอบ ในทางตรงข้าม การตัดสินแบบอิงกลุ่ม เป็นการตั้งว่า จะให้ นักเรียนจำนวนเท่าไร ผ่านการสอบ การตัดสินแบบอิงเกณฑ์นั้นเหมาะกับการสอบเพื่อวัดว่าผู้สอบมีความรู้ความสามารถในด้านใดด้านหนึ่งเพียงพอหรือไม่ ส่วนการสอบแบบอิงกลุ่มนั้นเหมาะสำหรับการสอบแข่งขันเพื่อเข้าศึกษาต่อ หรือ ทำงาน ในสถาบันที่มีตำแหน่งที่จะรับได้จำกัด เช่น การสอบเข้าโรงเรียนแพทย์ หรือ การสอบคัดเลือกแพทย์ประจำบ้าน การสอบส่วนใหญ่ในทางแพทยศาสตรศึกษานั้นเหมาะกับการตัดสินแบบอิงเกณฑ์ หากผู้สอบทุกคนมีความสามารถเพียงพอก็ไม่จำเป็นต้องมีผู้สอบตก การใช้การตัดสินแบบอิงกลุ่มเพื่อวัดความรู้ความสามารถในสถานการณ์อื่นนอกจากการสอบคัดเลือกนั้นเป็นการส่งเสริมให้นักเรียนเกิดความแข่งขันกัน (แทนที่จะช่วยกันเรียน) โดยไม่จำเป็น

เนื่องจากการสอบทางแพทยศาสตรศึกษาแทบทั้งหมดเหมาะกับการตั้งเกณฑ์สอบผ่านแบบอิงเกณฑ์ ผมจะขอขยายความวิธีการตั้งเกณฑ์สอบผ่านแบบอิงเกณฑ์ที่สำคัญและใช้บ่อย 2 วิธีใหญ่ๆ คือ 1. การตั้งเกณฑ์โดยพิจารณาข้อสอบ และ 2. การตั้งเกณฑ์โดยพิจารณาจากผู้สอบ ในบทความตอนต่อไปครับ

Iramaneerat C. Passing standard: Part II [Thai]. Medical Education Pamphlet 2006; 2(2): 2.

วิธีการตั้งเกณฑ์สอบผ่าน (passing standard) (ตอนที่ 2)

เชิดศักดิ์ ไชยมณีรัตน์

ในบทความนี้ผมจะขอแนะนำวิธีการตั้งเกณฑ์สอบผ่านโดยพิจารณาตัวข้อสอบที่ใช้สอบ วิธีการตั้งเกณฑ์ผ่านแบบนี้เหมาะสำหรับการสอบ multiple-choice questions ซึ่งอาจารย์ผู้ตั้งเกณฑ์ผ่านสามารถประเมินความน่าจะเป็นของการตอบข้อสอบแต่ละข้อถูกต้อง การตั้งเกณฑ์ผ่านแบบนี้ประกอบด้วย 3 ขั้นตอนหลักคือ

1. ระบุลักษณะของนักเรียน”คาบเส้น” (borderline examinees): นักเรียนในกลุ่มคาบเส้นนี้คือนักเรียนที่มีความรู้ความสามารถอยู่ระหว่าง “ยอมรับได้” กับ “ยอมรับไม่ได้” นักเรียนกลุ่มนี้มีความรู้ไม่มากพอที่อาจารย์จะตัดสินใจให้สอบผ่านได้อย่างสบายใจ แต่ก็มีความรู้ไม่น้อยจนอาจารย์จะตัดสินใจให้สอบตกได้โดยไม่มีข้อสงสัย คณะกรรมการตั้งเกณฑ์สอบผ่านต้องระบุลักษณะของนักเรียนในกลุ่มคาบเส้นนี้อย่างชัดเจนว่า ในเนื้อหาวิชาที่ทำการสอบ นักเรียนกลุ่มนี้ควรมีความรู้ในเรื่องใด และไม่มีความรู้ในเรื่องใด ขั้นตอนนี้อาจทำได้ง่ายขึ้นหากอาจารย์แต่ละท่านนึกภาพของนักเรียนจริงที่อาจารย์เคยรู้จักที่สมควรถูกจัดให้อยู่ในกลุ่มนักเรียนคาบเส้น แล้วบรรยายลักษณะของนักเรียนคนนั้นๆ ว่าทำอะไรได้ และทำอะไรไม่ได้ รู้เรื่องอะไรบ้าง ไม่รู้เรื่องอะไรบ้าง
2. ให้กรรมการแต่ละท่านพิจารณาข้อสอบแต่ละข้อ และตัดสินใจว่านักเรียนคาบเส้นน่าจะมีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด ขั้นตอนนี้สามารถทำได้หลายวิธีด้วยกัน ผมขอยกตัวอย่างวิธีที่เป็นที่แพร่หลายมาก 2 วิธีด้วยกัน คือ
 - 2.1. Angoff's method: ให้อาจารย์ระบุว่าหากนักเรียนคาบเส้น 100 คนทำข้อสอบข้อนั้น จะมีนักเรียนกี่คนที่ตอบข้อสอบข้อนั้นถูก (หรือความน่าจะเป็นที่นักเรียนคาบเส้นตอบข้อสอบข้อนั้นถูก)
 - 2.2. Ebel's method: ให้อาจารย์สร้างตารางแยกประเภทข้อสอบตามความสำคัญของเนื้อหาและตามความยากง่ายของข้อสอบและระบุว่าในข้อสอบแต่ละกลุ่ม หากนักเรียนคาบเส้น 100 คนทำข้อสอบจะมีนักเรียนกี่คนที่ตอบถูก หลังจากนั้นให้อาจารย์พิจารณาข้อสอบแต่ละข้อแล้วจัดประเภทเข้าในกลุ่ม ตัวอย่างเช่น

	ความยากง่าย	ง่าย	ปานกลาง	ยาก
ความสำคัญ				
สำคัญมาก		95%	85%	80%
สำคัญพอควร		90%	75%	60%
สำคัญน้อย		80%	55%	35%
สำคัญน้อยมาก		50%	30%	20%

3. ทำการคิดเกณฑ์สอบผ่านสำหรับข้อสอบนั้น
 - 3.1. Angoff's method เกณฑ์ผ่านคือผลรวมของความน่าจะเป็นของการตอบข้อสอบแต่ละข้อถูก

Item	1	2	3	4	5	Passing score
Probability	0.95	0.85	0.30	0.40	0.70	3.20

- 3.2. Ebel's method เกณฑ์ผ่านคือผลรวมของ (จำนวนข้อสอบในแต่ละกลุ่ม x ความน่าจะเป็นของการตอบข้อสอบถูก สำหรับข้อสอบในกลุ่มนั้น) จากข้อสอบทั้ง 12 กลุ่ม

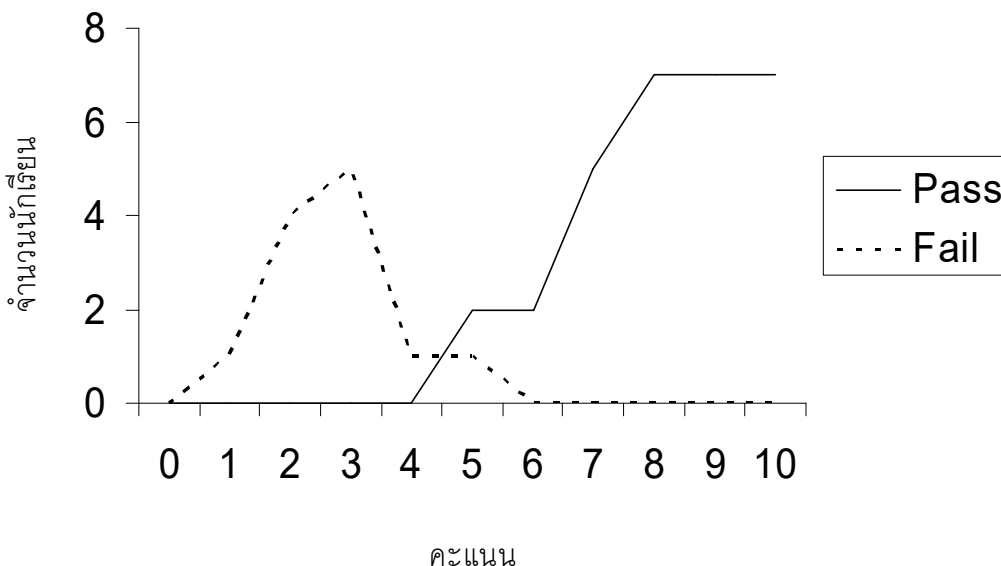
	ความยากง่าย	ง่าย	ปานกลาง	ยาก
ความสำคัญ		(24 ข้อ)	(15 ข้อ)	(11 ข้อ)
สำคัญมาก	(15 ข้อ)	95% x 5	85% x 5	80% x 5
สำคัญพอควร	(20 ข้อ)	90% x 10	75% x 7	60% x 3
สำคัญน้อย	(10 ข้อ)	80% x 5	55% x 3	35% x 2
สำคัญน้อยมาก	(5 ข้อ)	50% x 4	30% x 0	20% x 1
Passing score		37.6		

Iramaneerat C. Passing standard: Part III [Thai]. Medical Education Pamphlet 2006; 2(3): 1.

วิธีการตั้งเกณฑ์สอบผ่าน (passing standard) (ตอนที่ 3)
เชิดศักดิ์ ไอรมนีรัตน์

ในบทความนี้ผมจะขอแนะนำวิธีการตั้งเกณฑ์สอบผ่านโดยพิจารณานักเรียนผู้สอบ วิธีการตั้งเกณฑ์ผ่านแบบนี้เหมาะสำหรับการสอบวัดทักษะ การสอบสัมภาษณ์ หรือการประเมินการปฏิบัติงาน ซึ่งมักตัดสินการสอบผ่านโดยดูจากความสามารถของผู้สอบโดยรวมได้ง่ายกว่าดูจากคะแนนที่ได้ในหัวข้อประเมินแต่ละข้อ วิธีการตั้งเกณฑ์ผ่านลักษณะนี้ที่ใช้อยู่มีด้วยกัน 2 วิธีคือ

1. Borderline-group method: การตั้งเกณฑ์ผ่านวิธีนี้เริ่มจากให้คณะกรรมการสอบประชุมตกลงกันก่อนถึงลักษณะของผู้สอบที่อยู่ในกลุ่มคาบเส้น (ผู้สอบที่มีความรู้ไม่มากพอที่อาจารย์จะให้สอบผ่านได้อย่างสบายใจ แต่ก็มีความรู้ไม่น้อยจนอาจารย์สามารถตัดสินให้สอบตกได้โดยไม่มีข้อสงสัย) หลังจากนั้นอาจารย์พิจารณาความสามารถโดยรวมของผู้สอบแต่ละคน (โดยไม่ทราบคะแนนที่ผู้สอบคนนั้นได้รับ) แล้วระบุว่าผู้สอบคนใดจัดว่ามีความสามารถอยู่ในเกณฑ์ "คาบเส้น" เมื่อระบุว่าผู้สอบคนใดบ้างจัดว่ามีความสามารถคาบเส้นแล้วให้ตั้งเกณฑ์สอบผ่านที่คะแนน median ของผู้สอบกลุ่มนี้ (ไม่แนะนำให้ใช้ค่าเฉลี่ย (mean) เนื่องจากเกณฑ์ผ่านจะเบี่ยงเบนได้มากหากมีคะแนนที่สูงหรือต่ำมากเข้ามาร่วมในการคำนวณ)
2. Contrasting groups method: การตั้งเกณฑ์ผ่านวิธีนี้เริ่มจากการระบุลักษณะของผู้สอบที่ควรสอบผ่าน และ ผู้ที่ควรสอบตก หลังจากนั้นให้อาจารย์พิจารณาความสามารถของผู้สอบทีละคน (โดยไม่ทราบคะแนนที่ผู้สอบคนนั้นได้รับ) แล้วระบุว่าผู้สอบคนนั้นควรอยู่ในกลุ่ม "สอบผ่าน" หรือ "สอบตก" หลังจากนั้นให้ทำการวาดกราฟแสดงความสัมพันธ์ระหว่างจำนวนนักเรียนที่ถูกจัดให้สอบผ่าน และ สอบตก กับคะแนนที่นักเรียนได้รับ ดังตัวอย่างข้างล่าง



เกณฑ์ผ่านคือคะแนน ณ จุดที่ false positive และ false negative passing เท่ากัน (ในกรณีตัวอย่างนี้คือ 5 คะแนน) (คณะกรรมการตั้งเกณฑ์ผ่านอาจปรับเกณฑ์ผ่านได้เพื่อปรับอัตรา false positive และ false negative passing ได้ตามวัตถุประสงค์ของการสอบ)

รศ. ดร. นพ. ทาวรชัย ลิ้มจินดาพร
 ศ. ดร. เพทาย เย็นจิตโสมนัส
 อ. ดร. นพ.บดินทร์ ทรัพย์สมบูรณ์

ห้อง 1 Essential skills for thesis advisors

Essential Skills for Thesis Advisor 2020

Thawornchai Limjindaporn, MD, PhD
 Pa-thai Yenchitsomanus, PhD
 Bordin Sapsomboon, MD, PhD

As an advisor

1. How do you educate and evaluate the students?
2. How do you communicate to the students?
3. How do you build a team and create leadership skill for the students?

As an advisor

1. How do you educate and evaluate the students?
 - Observing the performance
 - Documenting the performance
 - Evaluating the performance

Observing performance

- Understand the caliber of your student.
- Give assignment from simplicity to difficulty.
- Watch over students not only to assess work quality but also to evaluate conduct, appearance, vitality, attitude and eagerness to learn.
- Check the timing always.

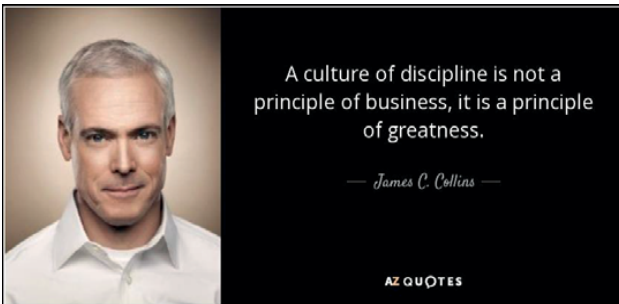
Documenting performance

- Treat documentation as a tool to gather facts thereby removing ambiguities.

Evaluating performance

- Give student ongoing feedback on performance both good and bad so that they know what they are doing right and what they need to improve as it has a lasting impact on the student.

As an advisor



2. How do you communicate to the students?

- **Inspire**
- **Available**
- **Constructive confrontation**

Inspire

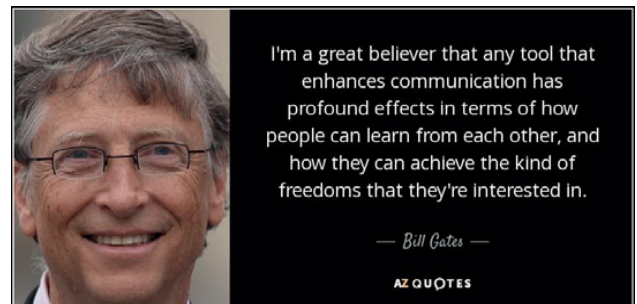
Take every opportunity to appreciate their contribution and to urge them to excellence.

Available

Maintain a proper timing in office and lab.

Constructive confrontation

- **Speak with clarity and purpose.**
- **Investigate complaints properly.**
- **Take an unbiased fact-based approach when investigation of complaints.**



As an advisor

3. How do you build the team and create leadership skill for the students?

- ***Building your team***
- ***Resolving the team conflicts***
- ***Managing violence***

Resolving the team conflicts

- ***Focus on a shared goal to referee disputes***
- ***Listen to suggestions, ideas, and feedback from other people, and find the common ground.***

Building your team

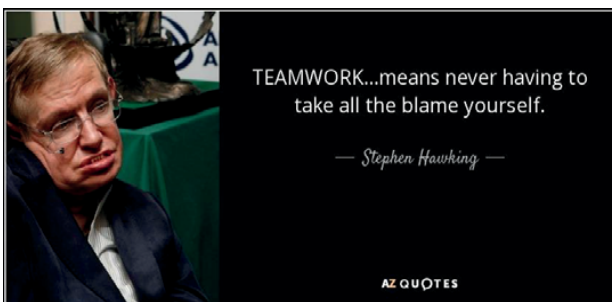
- ***Select students and get them into a shared goal.***
- ***Year plan for a shared goal for not only your group but also for a student in your group.***
- ***A true leader has no problem yielding control to another person when appropriate.***
- ***Always keep learning new things.***

Managing violence

Know the conditions that breed violence and protect your workplace.

References

- <https://www.wrike.com/blog/9-ways-develop-leadership-skills/>
- <https://blog.azendoo.com/8-ways-to-improve-team-communication/>



โจทย์ข้อที่ 1. Educate & Evaluate Students

ท่านรับเป็นอาจารย์ที่ปรึกษาของนักศึกษาปริญญาโทซึ่งเป็นนักศึกษาที่เรียนเก่ง จบปริญญาตรีจากมหาวิทยาลัยที่มีชื่อเสียงในประเทศไทย ด้วยคะแนนเกียรตินิยมอันดับหนึ่งด้วยเกรดเฉลี่ย ๓.๕๐ นักศึกษาเรียนผ่านในปีที่ ๑ ด้วยเกรดเฉลี่ย ๔.๐๐ ด้วยความเก่งจึงมีความมั่นใจในตัวเองสูงมาก เมื่อเริ่มปฏิบัติงานวิจัยเพื่อทำวิทยานิพนธ์ภายใต้การดูแลของท่านจึงเกิดปัญหาขึ้นท่านจะแก้ปัญหาที่เกิดขึ้นเหล่านี้ได้อย่างไร

1. ท่านไม่สามารถจะสอนเรื่องใหม่ๆให้กับนักศึกษาได้เพราะนักศึกษาคิดว่าตนเองรู้หมดทุกอย่างแล้วและสามารถทำทุกอย่างได้ด้วยตนเอง โดยไม่ต้องให้ท่านสอน แต่ท่านเห็นว่านักศึกษายังไม่รู้และยังไม่เข้าใจงานที่ตัวเองจะต้องทำอย่างแท้จริง และ นักศึกษาก็ทำงานแบบผัดบังถูกบ้างแต่มักจะทำผิดเป็นส่วนใหญ่
2. นักศึกษาไม่สามารถจะทำงานร่วมกับทีมงานของท่านได้และมีปัญหาขัดแย้งกับหลายคนที่เกี่ยวข้องโดยนักศึกษาไม่ได้คิดว่าตนเองมีปัญหากับใครแต่คนอื่นมามีปัญหากับเขาเองและเขาไม่รู้รู้สึกผิดหรือเดือดร้อนอะไร
3. นักศึกษาอยู่กับท่านมาเรื่อยๆ โดยใช้เงินส่วนตัวจากที่บ้าน สิ่งที่ท่านแนะนำและสอนส่วนใหญ่ไม่ค่อยปฏิบัติตามจึงทำให้งานวิจัยล่าช้าและได้ผลงานน้อยแต่นักศึกษาก็ออยู่กับท่านมา ๔ ปีแล้ว ท่านเองก็ได้รับความกดดันจากประธานหลักสูตรและภาควิชาฯจากนโยบายของมหาวิทยาลัยที่ต้องการเร่งให้นักศึกษาจบทันรอบวงที่กำหนดไว้
4. เนื่องจากนักศึกษาเห็นว่าท่านเป็นอาจารย์ที่เก่งและใจดี นักศึกษาจึงอยากจะขอทำปริญญาเอกกับท่านต่อไปเลยแม้จะยังไม่จบปริญญาโท โดยอยากจะขอข้ามไปเรียนในระดับเอกต่อเลยทันทีเนื่องจากนักศึกษามีคุณสมบัติครบตามที่หลักสูตรกำหนดไว้ทุกข้อและขอให้ท่านช่วยคุยกับประธานหลักสูตรในเรื่องนี้

โจทย์ข้อที่ 2. Communicate to Students

ท่านเป็นอาจารย์ซึ่งปฏิบัติงานในภาควิชามาระยะเวลาหนึ่ง เริ่มมีโครงการวิจัยของตนเองและมีทุนวิจัย และต้องการรับนักศึกษาปริญญาโทมาช่วยทำงานวิจัย กอปรกับมหาวิทยาลัยก็มีนโยบายให้อาจารย์รับนักศึกษาระดับบัณฑิตศึกษามาดูแลด้วย ท่านจึงรับนักศึกษาปริญญาโทซึ่งผ่านการเรียนรายวิชาในชั้นปีที่ 1 เสร็จแล้ว และจะต้องทำงานวิจัยเป็นวิทยานิพนธ์ในชั้นปีที่ ๒ โดยนักศึกษาได้รับทุนการศึกษาให้จบภายในระยะเวลา 2 ปี เนื่องจากท่านมีโครงการวิจัยที่ได้รับทุนมาอยู่แล้ว ท่านจึงให้นักศึกษาทำงานวิจัยโครงการดังกล่าวซึ่งท่านได้คิดวางแผน ตั้งสมมุติฐานและทบทวนวรรณกรรมมาแล้วเป็นอย่างดี นักศึกษาคนนี้ก็เป็นคนขยัน รับผิดชอบ มีทักษะและมีฝีมือในการทำวิจัยเป็นอย่างดี แต่การทดลองที่นักศึกษาทำซึ่งถูกต้องและเชื่อถือได้ทุกการทดลอง ให้ผลเป็นลบ ไม่เป็นไปตามที่คาด มีความเป็นไปได้ค่อนข้างสูงว่าสมมุติฐานของท่านน่าจะผิด แต่นักศึกษาได้ใช้เวลามาก่อนทางแล้ว จึงเหลือเวลาที่จะทำวิทยานิพนธ์อีกไม่มาก ท่านจะแก้ปัญหาในกรณีต่อไปนี้อย่างไร (ขอให้แก้ปัญหาแยกเป็นทีละกรณี โดยไม่เกี่ยวข้องกัน)

1. ท่านมีสมมุติฐานใหม่และได้วางแผนการทดลองขึ้นใหม่ที่อยากจะให้นักศึกษาทำ แต่ระยะเวลาที่นักศึกษามีเหลือโดยยังได้รับทุนจากคณะฯ ไม่เพียงพอต่อการทำงานใหม่นี้ และนักศึกษาก็ไม่อยากจะทำงานใหม่นี้ เนื่องจากจะต้องใช้เวลามาก และถ้าหากท่านให้นักศึกษาสำเร็จการศึกษาไป ท่านก็จะได้แต่ผลการทดลองที่เป็นลบ ไม่สามารถจะนำมาเขียนผลงานเพื่อตีพิมพ์ได้และนักศึกษาก็ยังไม่แน่ว่าจะสอบผ่านวิทยานิพนธ์หรือไม่เพราะกรรมการสอบบางท่านอาจจะไม่ยอมให้นักศึกษาสอบผ่าน หากได้ผลการทดลองที่เป็นลบ
2. นักศึกษามายกท่านว่าจะขอจบปริญญาโทภายในระยะเวลา 2 ปีและไม่สามารถจะอยู่ต่อได้เพราะไปสมัครงานไว้แล้วและที่ทำงานใหม่พร้อมรับเข้าทำงานหลังนักศึกษาสำเร็จการศึกษาภายในระยะเวลาที่กำหนด
3. หากท่านตัดสินใจใช้เงินทุนที่ท่านได้มาให้แก่นักศึกษาเพื่อเป็นเงินเดือน เมื่อทุนที่นักศึกษาได้รับหมดลงภายในระยะเวลา 2 ปี (โดยทุนที่ท่านได้รับมาเปิดโอกาสให้ท่านสามารถทำได้) ท่านจะต้องใช้เงินทุนครั้งหนึ่งทำให้เงินทุนวิจัยของท่านซึ่งได้มาแค่พอดีที่จะทำโครงการวิจัยไม่เพียงพอหากท่านใช้เงินทุนวิจัยเป็นเงินเดือน และท่านพูดคุยดีๆ นักศึกษาอาจจะอยู่ทำงานวิจัยกับท่านต่อไปได้อีกช่วงระยะเวลาหนึ่ง แต่ก็ยังไม่แน่ว่าสมมุติฐานใหม่ของท่านจะถูกต้องหรือไม่ ท่านควรทำอย่างไร
4. นักศึกษาคิดว่าหากทำงานวิจัยเรื่องนี้กับท่านต่อไปคงจะไม่จบการศึกษา จึงมาคุยกับท่านว่าอยากจะขอเปลี่ยนอาจารย์ที่ปรึกษาใหม่เป็นอาจารย์ท่านอื่นแทนท่าน

โจทย์ข้อที่ 3. Building Team for Students

ท่านเป็นอาจารย์ซึ่งปฏิบัติงานในภาควิชาที่เปิดหลักสูตรบัณฑิตศึกษาระดับนานาชาติ ท่านจึงต้องรับนักศึกษาปริญญาเอกซึ่งเป็นชาวต่างชาติ หลังจากนักศึกษาผ่านการเรียนรายวิชาในชั้นปีที่ ๑ แล้ว ได้เกรดเฉลี่ย 3.0 ซึ่งไม่มากนักสำหรับนักศึกษาในระดับปริญญาเอก นักศึกษาต้องเตรียมตัวสำหรับการสอบ qualify examination ตามคำฟัง หลังจากพลาดในการสอบครั้งที่ 1 รู้สึกเครียดมากอยากกลับไปประเทศตัวเอง ประธานหลักสูตรเรียกท่านไปพบและอยากให้ช่วยดูแลนักศึกษาให้อยู่ในหลักสูตรต่อไป ท่านจึงเรียกนักศึกษาไปพบ นักศึกษาบอกว่าขอกลับไปประเทศก่อนประมาณ 1 เดือนเพื่อไปอยู่กับพ่อแม่แล้วจะกลับมาเรียนใหม่ ท่านอนุญาตให้นักศึกษากลับไป ผ่านไป 2 เดือนนักศึกษากลับมาใหม่และอาการดีขึ้นหลังจากกลับไปประเทศและขอผู้ต่อจนสอบ qualify examination ผ่านได้ในสอบครั้งที่ ๒ เนื่องจากท่านมีโครงการวิจัยที่ได้รับทุนมาอยู่แล้ว ท่านจึงให้นักศึกษาทำงานวิจัยโครงการดังกล่าวซึ่งท่านได้คิดวางแผน ตั้งสมมุติฐาน และทบทวนวรรณกรรมมาแล้วเป็นอย่างดีทำให้นักศึกษาได้ผลงานไปตามที่คาดและใกล้จบการศึกษาก่อนเวลาที่กำหนด นักศึกษาคนไทยในห้องปฏิบัติการเดียวกันรู้สึกว่าคุณทำตามใจนักศึกษาชาวต่างชาติมากและอนุญาตให้นักศึกษาต่างชาติจบการศึกษาได้เร็วกว่านักศึกษาชาวไทย รู้สึกว่าไม่ยุติธรรม ท่านจะแก้ปัญหาที่เกิดขึ้นเหล่านี้กับนักศึกษานิเทศศาสตร์คนนี้อย่างไร

1. นักศึกษาชาวต่างชาติมาหาท่านเพราะอยากให้ใช้พูดภาษาอังกฤษในการสื่อสาร ในช่วง lab meeting เพราะเป็นหลักสูตรนานาชาติ ทำให้ไม่ถึงการประชุม lab meeting เป็นภาษาไทย ท่านเสนอเปลี่ยน lab meeting เป็นภาษาอังกฤษ แต่นักศึกษาไทยและเจ้าหน้าที่ในห้องปฏิบัติการที่ไม่ใช่เป็นนักศึกษามาประจำเพราะว่าคนส่วนใหญ่เป็นคนไทยและหากพูดภาษาอังกฤษทั้งหมดรู้สึกที่ไม่สามารถสื่อสารสิ่งที่ควรรู้จริงๆได้ ท่านควรทำอย่างไร
2. ก่อนนักศึกษาชาวต่างชาติจบการศึกษาประมาณ 3 เดือนและต้องเตรียมการสอบป้องกันวิทยานิพนธ์ นักศึกษาชาวต่างชาติป่วยอีกครั้งด้วยอาการซึมเศร้าและต้องเข้ารับการรักษาในโรงพยาบาล อย่างไรก็ตามพบว่าประกันสุขภาพของนักศึกษาชาวต่างชาติที่ซื้อเองไม่ครอบคลุมสิทธิ์ทั้งหมดและต้องหาเงินจ่ายเพิ่มเติม ผู้ปกครองของนักศึกษาจดหมายมาหาท่านเพื่อให้ท่านช่วยหาทุนช่วยเหลือเรื่องการรักษาและช่วยประคองให้จบการศึกษา ท่านควรทำอย่างไร

รศ. ดร. นพ. เชิดศักดิ์ ไอรmaniรัตน์

ห้อง 2 Long case exam

Long Case Examination

Cherdsak Iramaneerat
Department of Surgery
Faculty of Medicine Siriraj Hospital

A Long Case Exam

- While OSCE focuses on individual components of clinical competence, it is widely agreed that there is still a need for assessing students on patient care as a whole.

Long Case Examination

- The examinees spend a long period of time (usually about an hour) to explore and work up a single patient case in detail.
- An examiner assesses history taking, physical examination, communication skills, diagnostic skills, plan of investigations, management, and professionalism of the examinees

Outline

- Objectives
- Advantages and limitations
- Objective Structured Long Case Examination Record (OSLER)
- Long case exam in Thailand

Discussion

- อาจารย์คิดว่าเราต้องการใช้การสอบ long case examination ในการประเมินนักศึกษาแพทย์ ปีที่ 6 ในด้านใดบ้าง
- Group discussion 5 minutes

Assessment Objectives

- Knowledge
 - Lower order: Recall, Comprehension, Application
 - Higher order: Analysis, Synthesis, Evaluation
- Psychomotor skills
- Attitudes

Discussion

- อาจารย์สนับสนุนการใช้ long case examination ประเมินความรู้ ความสามารถทางการแพทย์ของนักศึกษาแพทย์หรือไม่ เพราะเหตุใด

(Group discussion 5 min.)

Long Case Examination

- **Advantages**
 - Comprehensive competency evaluation
 - In-depth exploration of knowledge, skills

Long Case Examination

- **Disadvantages**
 - Subjective ratings
 - Unstructured settings
 - Adequacy of observation
 - Case specificity: construct underrepresentation
 - Fairness among students: A luck of draw
 - Time commitment from medical teachers
 - Low reliability
 - Divergence of objectives: oral examination

OSLER

- **Objective Structured Long Case Examination Record (OSLER)**
 - Ten items structured record
 - History taking
 - Physical exam
 - Investigation, management, clinical acumen
 - Objectivity: prior agreement on what to be examined
 - Assess both processes and products
 - Identification of case difficulty by an examiner

Gleeson F. Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). Medical Teacher 1997, 19: 7 – 14.

OSLER's components

- **History taking**
 - Clarity of presentation, communication process, systematic approach, establishment of case facts.
- **Physical examination**
 - Systematic approach, examination technique, establishment of correct physical findings.
- **Investigations, Management, Clinical acumen**
 - Ability to identify and solve problems

The Case Difficulty

- **Standard case**
 - Single problem
- **Difficult case**
 - Up to three problem
- **Very difficult case.**
 - More than three problem

Awarding marks in the OSLER

- P+: Very good/excellent. (60-80%)
- P: Pass/ bare pass. (50-55%)
- P-: Below pass

– Each items has to be graded followed by overall grade of the complete performance

OBJECTIVE STRUCTURED LONG EXAMINATION RECORD (OSLER) DATE: _____

CANDIDATES : NAME _____	EXAMINATION NO. _____
----------------------------	-----------------------

Examiners are required to **GRADE** each of the ten items below and assign an overall **GRADE** and **MARK** concerning the candidate **PRIOR** to discussion with their co-examiner as follows:

GRADES	MARKS	
P+ = VERY GOOD/EXCELLENT	(60-80+)	See over page for specific mark details.
P = PASS/BORDERLINE PASS	(50-55)	
P- = BELOW PASS	(35-45)	

EXAMINER: _____
CO-EXAMINER: _____

	GRADE	AGREED GRADE
PRESENTATION OF HISTORY		
PACE/CLARITY →		
COMMUNICATION PROCESS (History e.g. CVS, investigation e.g. endoscopy, management e.g. patient education) →		
SYSTEMATIC PRESENTATION →		
CORRECT FACTS ESTABLISHED →		
PHYSICAL EXAMINATION		
SYSTEMATIC →		
TECHNIQUE (Including attitude to patient) →		
CORRECT FINDINGS ESTABLISHED →		
APPROPRIATE INVESTIGATIONS IN A LOGICAL SEQUENCE (Communication Process option) →		
APPROPRIATE MANAGEMENT (Communication Process option) →		
CLINICAL ACUMEN (Problem identification/Problem solving Ability) →		
ADDITIONAL COMMENTS:-		

Please Tick (✓) For CASE DIFFICULTY

	For CASE DIFFICULTY		INDIVIDUAL EXAMINER		PAIR OF EXAMINERS	
	Individual Examiner	Agreed Case Difficulty	OVERALL GRADE	MARK	AGREED GRADE	AGREED MARK
Standard	<input type="checkbox"/>	<input type="checkbox"/>				
Difficult	<input type="checkbox"/>	<input type="checkbox"/>				
Very Difficult	<input type="checkbox"/>	<input type="checkbox"/>				

Examination Time

- Examiner – candidate time must be sufficient to allow for a valid assessment.
- Identical time should be allowed for all candidates in the interest of examination reliability.
- A minimum of 20 minutes should be allowed.
- For high-stakes exam: 30 minutes is recommended.

National Medical Licensing Examination

- Step 1: MCQ in Basic medical science
- Step 2: MCQ in Clinical science
- Step 3: Clinical skills and problem solving
 1. OSCE
 2. MEQ
 3. Long case exam

Long Case Examination

- ข้อกำหนดของ ศรว. ในการสอบ long case examination
 1. จำนวนผู้ป่วยอย่างน้อย 2 ราย
 2. โรค หรือ ปัญหาสอดคล้องกับเกณฑ์มาตรฐานผู้ประกอบการวิชาชีพเวชกรรมของแพทยสภา
 3. ผู้ป่วยใน หรือ ผู้ป่วยนอก
 4. รูปแบบการสอบ 3 ขั้นตอน
 1. Patient encounter under direct observation 30 นาที
 2. Case discussion 20 – 30 นาที
 3. Patient encounter 10 นาที

Clinical Competencies

- History taking (15)
- Physical examination (15)
- Data organization and presentation (10)
- Case discussion: reasoning and analysis (15)
- Decision making and problem solving (15)
- Communication skills (15)
- Professional attitudes and etiquette (15)

Level of Competencies

- Very good
 - ความถูกต้องครบถ้วนมากกว่าร้อยละ 80
- Good
 - ความถูกต้องครบถ้วนร้อยละ 60 - 80
- Require improvement
 - ความถูกต้องครบถ้วนน้อยกว่าร้อยละ 60 (ไม่ผ่าน)

Summary

Long case exam

- Objectives
- Advantages and limitations
- Objective Structured Long Case Examination Record (OSLER)
- Long case exam in Thailand

รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์ รศ. พญ. กษณา รักขมณี
ผศ. นพ. สุพัฒน์ สนใจพานิชย์ รศ. นพ. สุพจน์ พงศ์ประสพชัย
รศ. พญ. รัชวรรณ จิระติวานนท์

ห้อง 2 OSCE item development

OSCE Item Development

เชิดศักดิ์ ไอรมณีรัตน์

ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล

มหาวิทยาลัย มหิดล

OSCE

- Objective
- Structured
- Clinical
- Examination
- มีวัตถุประสงค์ที่ชัดเจน
- มีการจัดโครงสร้างเป็นสถานีย่อย
- ประเมินทักษะทางคลินิก
- การสอบ

History

- 1975: Ronald Harden (University of Dundee) proposed a series of stations in examination of clinical skills for 5 minutes per each station.
- 1988: Faculty of Medicine, Ramathibodi hospital implemented an OSCE in M3 exam (introduction to clinical medicine)
- 1991: Medical Council of Thailand implemented an OSCE in medical licensing exam for foreign graduates.
- 2009: Center for Medical Competency Assessment and Accreditation implemented an OSCE as Step 3 medical licensing exam.

OSCE

- **Objective Structured Clinical Examination**
- **Assessment of clinical skills**
 - History taking
 - Physical examination
 - Communication skills
 - Procedural skills
 - Interpretation of medical investigations
 - Ordering of medical treatment

Components of an OSCE item

1. Scenario (ภาพรวมสถานการณ์)
2. Instruction for examinees (คำแนะนำผู้เข้าสอบ)
3. Instruction for SPs (คำแนะนำผู้ป่วยมาตรฐาน)
4. Scoring rubric (ใบให้คะแนน +/- คำแนะนำอาจารย์)

Scenario

- Title
- Objectives
- Examinees
- Clinical information
- Apparatus
- SP requirements
- Time

Scenario 1

หัวข้อ : การตรวจร่างกายผู้ป่วยที่มีอาการปวดท้อง

Objective : นักศึกษาแพทย์สามารถแสดงวิธีการตรวจร่างกายผู้ป่วยที่มีอาการปวดท้องเฉียบพลัน และให้การวินิจฉัยที่ถูกต้องได้

ผู้สอบ: นักศึกษาแพทย์ชั้นปีที่ 6

สถานการณ์: สมบูรณ์ อายุ 35 ปี มีอาการปวดท้องใต้ชายโครงด้านซ้าย 6 ชั่วโมง ปวดตื้อๆตลอดเวลา

คำสั่ง : จงแสดงวิธีการตรวจหน้าท้องผู้ป่วย บรรยายสิ่งที่ตรวจพบและให้การวินิจฉัยโรคที่คิดถึงมากที่สุด 1 โรค

เวลา : 5 นาที (ตรวจร่างกาย 4 นาทีครึ่ง บอกสิ่งที่พบและวินิจฉัยครึ่งนาที)

Scenario 1 (cont.)

Apparatus	ผู้ป่วยสมมติ (ชายอายุ 30 - 40 ปี ไม่มีแผลผ่าตัดหน้าท้อง)	1 คน
	โต๊ะหนึ่งสำหรับกรรมการ	1 ตัว
	เก้าอี้หนึ่ง	1 ตัว
	เตียงตรวจร่างกาย	1 ตัว
	ผ้าปูเตียง หมอน และผ้าห่ม	1 ชุด
	เอกสารอธิบายและแบบฟอร์มการให้คะแนน	

Instruction for Examinees

- ผู้ป่วยหญิงไทยคู่ อายุ 22 ปี มีอาการปวดท้อง 4 ชั่วโมงก่อนมาโรงพยาบาล
- คำสั่ง
 1. จงซักประวัติผู้ป่วยรายนี้ (4 ½ นาที)
 2. จงบอกการวินิจฉัยโรคที่นึกถึงมากที่สุด (1/2 นาที)

Standardized Patient (SP)

- ผู้ป่วยมาตรฐาน
 - ผู้ป่วยจริง หรือ คนปกติมาแสดงเป็นผู้ป่วย
 - ได้รับการฝึกให้นำเสนออาการ หรือ อาการแสดงที่กำหนด
 - สามารถแสดงได้เหมือนบทบาทในการแสดงทุกครั้ง
 - เพื่อใช้ในการสอน หรือ ประเมินผลนักศึกษา

Instruction for SPs

- General information about the scenario
- Information of the portrayed patient
 - Name, age, and relevant personal information (occupation, family, etc.)
 - Dress (+/- make-up)
 - Medical history/ physical findings
 - If being asked, answered ...
 - If being pressed, reacted....
 - Cue to portray or reveal special information/findings (cry, angry, guiding info., etc.)

Scoring Rubric General Format

หัวข้อการประเมิน	ปฏิบัติ		ไม่ปฏิบัติ
	สมบูรณ์	ไม่สมบูรณ์	
ตอนที่ 1. การปฏิบัติต่อผู้ป่วย	10	6	0
	ครบ	อย่างน้อย 2	1 หรือ 0 ข้อ
ตอนที่ 2. รายละเอียดอาการ/การปฏิบัติ	5	3	0
ตอนที่ 3 การวินิจฉัยแยกโรค	XXXX	10	
	YYYY	8	
	ZZZZ	5	

Scoring Rubric

ขั้นตอนการประเมิน	สมบูรณ์	ไม่สมบูรณ์	ไม่ปฏิบัติ
1. การแนะนำตัว			
1.1 การแนะนำตัวเองอย่างสุภาพ	5	3	0
1.2 การถามชื่อผู้ป่วยอย่างสุภาพ	5	3	0
2. การถามประวัติ			
2.1 ถามตำแหน่งที่ปวด	10	-	0
2.2 ถามลักษณะของการปวด	10	6	0
2.3 ถามอาการปวดร้าวไปที่อื่น	10	-	0
...			
2.8 ถามประวัติประจำเดือน	10	6	0
3. การวินิจฉัยโรค			
Ectopic pregnancy	10		
Acute appendicitis	6		

Scoring Rubric

- กระชับ ได้ใจความ สื่อความหมายตรงกัน
- กำหนดประเด็นที่สำคัญ หรือเป็นจุดที่มักทำผิดพลาด
- บรรยายพฤติกรรมที่ผู้ประเมินสังเกตได้
- กำหนดน้ำหนักคะแนนตามความสำคัญ

Questions & Comments

Cherdsak.ira@mahidol.ac.th

Activity

- แบ่งกลุ่มย่อย 4 กลุ่ม
- สร้างข้อสอบ OSCE 1 ข้อ
 - Scenario
 - Instruction for examinees
 - (Brief) Instruction for SPs
 - Scoring rubric
- เวลา 20 นาที
- เตรียมนำเสนอ เวลา 1130 น (เวลากลุ่มละ 5 นาที)

Iramaneerat C. Guidelines in developing an objective structured clinical examination: Case content [Thai]. Medical Education Pamphlet 2005; 1(8): 4.

ข้อแนะนำในการจัดสอบ OSCE (Objective Structured Clinical Examination) (ตอนที่ 1)

เชิดศักดิ์ ไอรมนวีรัตน์

Objective Structured Clinical Examination (OSCE) เป็นเทคนิคที่เป็นที่ยอมรับและได้รับการใช้มากขึ้นเรื่อยๆ ทั้งการสอนและประเมินผล ทางแพทยศาสตรศึกษาทุกระดับทั่วโลก ผมจะขอเสนอเกร็ดความรู้เกี่ยวกับการจัดสอบ OSCE โดยแบ่งออกเป็น 3 ตอนตามส่วนประกอบสำคัญของ OSCE ได้แก่ เนื้อหาของโจทย์ (content) ผู้ป่วยมาตรฐาน (standardized patient) และ อาจารย์ผู้ให้คะแนน (rater) ในบทความนี้จะขอกล่าวถึง เนื้อหาของโจทย์

1. สิ่งแรกที่ต้องคำนึงถึงคือวัตถุประสงค์ของการสอบ เนื่องจาก OSCE เป็นการสอบที่ต้องใช้ทรัพยากรมาก ควรตั้งวัตถุประสงค์การสอบเพื่อประเมินความรู้ความสามารถที่ไม่สามารถประเมินได้ด้วยวิธีอื่น เช่นทักษะในการสื่อสารกับผู้ป่วย ทักษะการให้คำแนะนำแก่ผู้ป่วย ทักษะการทำหัตถการ เป็นต้น ไม่ควรใช้ OSCE เพื่อวัดความรู้ผิวเผินที่สามารถวัดได้ด้วยข้อสอบ MCQ
2. วางแบบแปลนของเนื้อหาข้อสอบ (test blueprint) ที่ครอบคลุมเนื้อหาวิชาในทุกด้าน และทุกทักษะที่ต้องการประเมินอย่างเท่าเทียมกัน มีการระบุว่าในการสอบ OSCE นี้ทดสอบความรู้เรื่องใดบ้าง (โรคปอด โรคหัวใจ โรคไต ฯลฯ) และใช้ทักษะใดบ้าง (การซักประวัติ การตรวจร่างกาย การให้คำแนะนำ ฯลฯ) อย่างละกี่ข้อ ระวังอย่าให้เนื้อหาข้อสอบมีน้ำหนักในเรื่องใดเรื่องหนึ่งมากกว่าเรื่องอื่น
3. ในการเขียนโจทย์ OSCE แต่ละข้อ ต้องเขียนให้ครอบคลุมรายละเอียดทุกด้านของการสอบ ได้แก่ คำชี้แจงสำหรับนักเรียน สำหรับผู้ป่วยมาตรฐาน และสำหรับอาจารย์ผู้คุมสอบ สถานการณ์ผู้ป่วยจำลอง ประวัติและผลการตรวจร่างกายที่ผู้ป่วยมาตรฐานต้องแสดงออก อุปกรณ์ประกอบที่ต้องใช้ ระยะเวลาที่ต้องใช้ แบบฟอร์มให้คะแนน และเกณฑ์การให้คะแนน
4. การเขียนโจทย์ผู้ป่วยควรนำข้อมูลมาจากผู้ป่วยจริง ซึ่งจะทำให้โจทย์มีความเหมือนจริง ไม่ขาดรายละเอียดในเนื้อหาของโจทย์ และประหยัดเวลาในการแต่งโจทย์ นอกจากนี้ยังทำให้มีแฟ้มประวัติและผลการตรวจเพิ่มเติมรวมทั้งฟิล์มที่สามารถนำมาใช้เสริมโจทย์ได้ง่าย
5. โจทย์สำหรับแต่ละสถานีควรมีความยาวเหมาะสม โจทย์ที่ใช้เวลานานสามารถให้ข้อมูลเกี่ยวกับความสามารถของนักเรียนในเรื่องนั้นๆ ได้ละเอียด แต่ก็ทำให้มีโอกาสวัดความสามารถของนักเรียนได้น้อยเรื่อง เนื่องจากทักษะทางการแพทย์หลายด้านมีความเจาะจงต่อภาวะโรค (นักเรียนที่ซักประวัติโรคเลือดได้ดีอาจซักประวัติผู้ป่วยโรคซึมเศร้าไม่คล่องได้) โดยทั่วไปแนะนำให้จัดเวลาที่ใช้สอบในแต่ละสถานี ให้นักเรียนได้มีโอกาสสอบอย่างน้อย 8 – 10 สถานี (ยังมีสถานีสอบมาก ผลการสอบยังมีความแม่นยำมาก) หลายการศึกษาพบว่าเพื่อให้ได้ผลการสอบ OSCE ที่มีความแม่นยำพอยอมรับได้ จะต้องใช้เวลาในการสอบอย่างน้อย 3 – 4 ชั่วโมง
6. จัดให้มีการตอบคำถามตามหลังการสอบทักษะกับผู้ป่วย (post-encounter probe) เท่าที่จำเป็น ไม่มากเกินไป เนื่องจากคำถามเหล่านี้มักวัดความสามารถที่แตกต่างไปจากวัตถุประสงค์หลักของการสอบ OSCE (มักวัดความรู้ในทำนองเดียวกับ MCQ) จึงเป็นการเพิ่มเวลาสอบโดยไม่จำเป็นและยังลดความแม่นยำของผลการสอบอีกด้วย

Iramaneerat C. Guidelines in developing an objective structured clinical examination: Standardized patients [Thai]. Medical Education Pamphlet 2005; 1(9): 3.

ข้อแนะนำในการจัดสอบ OSCE (Objective Structured Clinical Examination) (ตอนที่ 2)
เชิดศักดิ์ ไอรมนวีรัตน์

ในบทความนี้จะขอเสนอเกร็ดความรู้เกี่ยวกับการใช้ผู้ป่วยมาตรฐาน (Standardized patients) ใน OSCE ก่อนอื่นผมขอล่าวถึงนิยามของศัพท์ที่สำคัญในการใช้ผู้ป่วยในการสอบก่อน เราเรียกคนที่ไม่มีอาการเจ็บป่วย แต่แสดงบทบาทเป็นผู้ป่วยว่า ผู้ป่วยสมมติ (simulated patient) ซึ่งผู้ป่วยสมมติเหล่านี้อาจแสดงอาการไม่สม่ำเสมอ เมื่อได้พบกับนักเรียนแต่ละคน หากเราทำการฝึกให้ผู้ป่วยสมมติ (หรือ ผู้ป่วยจริง) แสดงอาการที่อาการและอาการแสดงอย่างสม่ำเสมอ เป็นมาตรฐานเดียวกันไม่ว่าจะได้พบกับนักเรียนคนใด เราจะได้ ผู้ป่วยมาตรฐาน (standardized patient) การสอบ OSCE ให้ได้ผลการประเมินที่แม่นยำนั้นต้องใช้ผู้ป่วยมาตรฐาน (standardized patient, SP)

1. ผู้ป่วยมาตรฐานต้องได้รับการฝึกฝนอย่างจริงจังมั่นใจว่าการแสดงอาการและอาการแสดงได้มาตรฐานในทุกครั้งที่แสดงบทบาท การฝึกฝนนี้ต้องเริ่มต้นจากการมีบท (script) ที่ดี มีความละเอียดครอบคลุมข้อมูลทุกด้านที่เกี่ยวข้องกับภาวะโรคที่สนใจ และมีการฝึกซ้อมและตรวจแก้ไขโดยอาจารย์ผู้ตั้งใจเพื่อให้มั่นใจว่าความเข้าใจบทบาทของผู้ป่วยมาตรฐานถูกต้องตามความตั้งใจของผู้ตั้งใจ โดยทั่วไปเมื่อได้รับการฝึกฝนแล้วผู้ป่วยมาตรฐานสามารถแสดงอาการและอาการแสดงได้อย่างถูกต้องมากกว่า 90%
2. ในการสอบใหญ่บางครั้งมีความจำเป็นต้องใช้ผู้ป่วยมาตรฐานหลายคนเพื่อแสดงบทบาทเดียวกัน มีหลายการศึกษาแสดงว่าการใช้ผู้ป่วยมาตรฐานหลายคนในลักษณะนี้ไม่ลดความแม่นยำของผลสอบ ตราบเท่าที่เรามีสถานีสอบ OSCE มากเพียงพอ และผู้ป่วยมาตรฐานได้ถูกสุ่มกระจายตัวอยู่ตามสถานีสอบอย่างไม่ลำเอียง (randomly distributed)
3. หลายการศึกษาที่วิเคราะห์การสอบที่มีความจำเป็นต้องใช้ผู้ป่วยมาตรฐานชุดเดิมสอบนักเรียนหลายชุดต่อเนื่องกัน พบว่านักเรียนที่สอบรอบหลังไม่ได้ทำคะแนนได้ดีกว่านักเรียนที่สอบรอบแรก แสดงว่านักเรียนที่สอบก่อนไม่ให้อะไรเกี่ยวกับการสอบที่เป็นประโยชน์แก่นักเรียนที่สอบรอบหลัง หรือหากนักเรียนให้ข้อมูลแก่กัน ข้อมูลเพียงที่ได้รับเกี่ยวกับคำชี้แจงโดยไม่มีข้อมูลรายละเอียดของเกณฑ์การให้คะแนนนั้นไม่ได้ก่อให้เกิดความได้เปรียบในการสอบแก่นักเรียนรอบหลัง
4. นอกจากจะใช้ผู้ป่วยมาตรฐานเพื่อวัดทักษะของนักเรียนที่เกี่ยวข้องกับผู้ป่วยโดยตรง (เช่นการซักประวัติ ตรวจร่างกาย) แล้ว เรายังสามารถใช้ผู้ป่วยมาตรฐานประกอบกับแบบจำลองเพื่อทดสอบทักษะการทำหัตถการเพื่อทำให้การปฏิบัติหัตถการมีความสมจริงได้ด้วย เช่น การนำแบบจำลองสำหรับเย็บแผลมาติดกับแขนของผู้ป่วยจำลอง จะช่วยให้สามารถวัดทักษะในการเย็บแผลในขณะเดียวกันกับที่ต้องมีปฏิสัมพันธ์กับผู้ป่วยที่มีความเจ็บปวดจากบาดแผลด้วย

Iramaneerat C. Guidelines in developing an objective structured clinical examination: Scoring [Thai]. Medical Education Pamphlet 2005; 1(10): 1.

ข้อเสนอแนะในการจัดสอบ OSCE (Objective Structured Clinical Examination) (ตอนที่ 3)

เชิดศักดิ์ ไอรรมณีรัตน์

ในบทความนี้จะขอเสนอเกร็ดความรู้เกี่ยวกับการให้คะแนนในการสอบ OSCE

1. การให้คะแนน OSCE ทำได้ 2 วิธีใหญ่ๆ ด้วยกัน คือ checklist (ให้คะแนน 1 เมื่อทำสิ่งที่ระบุในรายการ และให้คะแนน 0 เมื่อไม่ทำรายการนั้น เช่น “นักเรียนถามประวัติประจำเดือนครั้งสุดท้าย”: 0 ทำ, 1 ไม่ทำ) และ rating scale (ให้คะแนนได้หลายระดับขึ้นกับระดับความถูกต้องของการปฏิบัติ เช่น “นักเรียนอธิบายหัตถการที่จะทำได้ชัดเจน” : 1 ไม่เห็นด้วยอย่างยิ่ง, 2 ไม่เห็นด้วย, 3 เห็นด้วย, 4 เห็นด้วยอย่างยิ่ง) การให้คะแนนด้วย checklist จะได้ผลการประเมินที่ผู้ให้คะแนน (rater) มีความเห็นพ้องกัน (inter-rater agreement) มากกว่า แต่สามารถแยกแยะความแตกต่างระหว่างนักเรียนที่มีความสามารถต่างกันได้ดีไม่เท่ากับการให้คะแนนด้วย rating scale ควรใช้ checklist สำหรับให้คะแนนโจทย์ที่ประเมินความครบถ้วนของเนื้อหาหรือขั้นตอน (เช่น ชักประวัติ ตรวจร่างกาย) แต่ควรใช้ rating scale สำหรับให้คะแนนโจทย์ที่ประเมินคุณภาพของทักษะหรือกระบวนการปฏิบัติ (เช่น ทักษะการสื่อสาร ทักษะการทำหัตถการ)
2. ไม่มีความจำเป็นต้องใช้ผู้ให้คะแนน (rater) มากกว่า 1 คน ต่อ 1 สถานี หากมีทรัพยากรบุคคลมากพอ เราควรเพิ่มจำนวนสถานีสอบ มากกว่า เพิ่มจำนวนผู้ให้คะแนนต่อสถานี การเพิ่มจำนวนสถานีสอบ ส่งผลให้คะแนนสอบ OSCE มีความแม่นยำเพิ่มขึ้นมากกว่า การเพิ่มจำนวนผู้ให้คะแนนต่อสถานี
3. นอกจากเราจะให้อาจารย์แพทย์เป็นผู้ให้คะแนนแล้ว เรายังสามารถฝึกให้ผู้ป่วยมาตรฐาน (standardized patient) ทำการให้คะแนนได้ด้วย พบว่าเมื่อได้รับการอธิบายเกณฑ์การให้คะแนนและฝึกปฏิบัติแล้ว ผู้ป่วยมาตรฐาน สามารถให้คะแนนที่มีความแม่นยำสูงไม่แพ้อาจารย์แพทย์ ข้อดีของการให้ผู้ป่วยมาตรฐานเป็นผู้ให้คะแนนคือสะดวก และประหยัด ในทางกลับกันการให้อาจารย์แพทย์เป็นผู้ให้คะแนนมีข้อได้เปรียบคืออาจารย์สามารถชี้แนะข้อบกพร่อง และแนะนำแนวทางการปรับปรุงแก้ไขทักษะและวิธีคิดของนักเรียนได้ทันที
4. ไม่ควรใช้ผลการประเมินจากสถานีใดสถานีหนึ่งเป็นตัวบ่งชี้ว่านักเรียนมีความสามารถหรือไม่มีความสามารถในด้านใด เนื่องจากผลการประเมินจากสถานีเดียวมีโอกาสผิดพลาดได้มาก การตัดสินว่านักเรียนคนใดมีความสามารถหรือไม่ให้ใช้ผลการประเมินโดยรวมซึ่งมีความแม่นยำมากกว่า
5. การรายงานคะแนน OSCE แก่นักเรียนนั้นต้องคำนึงถึงวัตถุประสงค์ของการสอบ หากทำการสอบ formative test ควรบอกข้อดี ข้อด้อย ของนักเรียนแต่ละคน และชี้แจงสิ่งที่ควรปรับปรุงอย่างละเอียด ส่วนคะแนนรวมนั้นอาจไม่ค่อยมีความสำคัญนัก ในทางกลับกัน หากทำการสอบ summative test เราต้องคำนึงถึงการรักษาความลับของข้อสอบ เนื่องจากข้อสอบ OSCE ที่ดีนั้นพัฒนาขึ้นได้ยาก และควรได้รับการเก็บไว้ในคลังข้อสอบเพื่อนำมาใช้ในอนาคต ดังนั้นเราไม่ควรแจ้งรายละเอียด ข้อถูก ข้อผิด ของนักเรียนแต่ละคนในทุกสถานี แต่แจ้งเพียงผลสอบว่าผ่านหรือไม่ผ่าน

รศ. ดร. นพ. เชิดศักดิ์ ไอรมณีรัตน์

ห้อง 2 EPA and WPBA

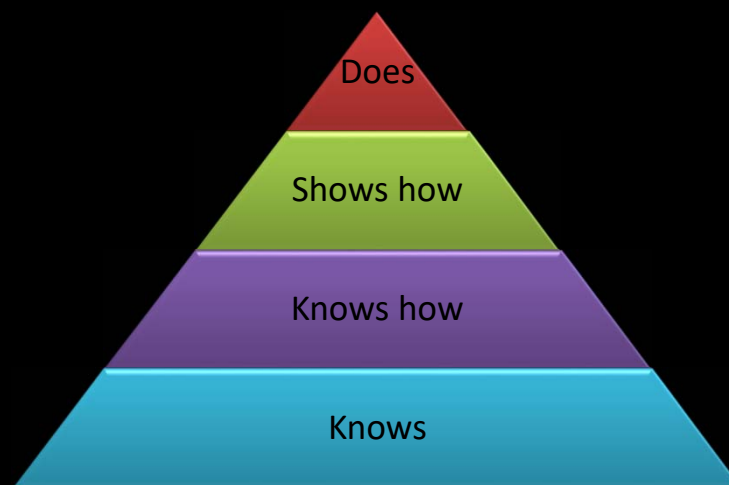
Entrustable Professional Activity (EPA) Workplace-based Assessment (WPBA)

เชิดศักดิ์ ไอรมณีรัตน์

ศูนย์ความเป็นเลิศด้านการศึกษาศาสตร์สุขภาพ

คณะแพทยศาสตร์ศิริราชพยาบาล

Assessment Approaches



Miller's Pyramid

2 2

Assessment at “Does” level

- Does => Professional task

EPA

Outline

- EPA
 - Definitions: EPA, competencies, milestones
 - Key concepts
 - How to proceed with EPA?
 - Assessment in EPA framework
- WPBA
 - Characteristics
 - Examples: Mini-CEX, DOPS, CbD, PBA
 - Guidelines for implementation

EPA

- **Entrustable Professional Activity**
 - A unit of professional practice, defined as tasks or responsibilities that trainees are entrusted to perform unsupervised once they have attained sufficient specific competence

AAMC. Core entrustable professional activities for entering residency: Faculty and learners' guide, Washington DC, 2014.

Competency

- **Competency: An observable ability of a professional, integrating multiple components such as knowledge, skills, values, and attitudes**

AAMC. Core entrustable professional activities for entering residency: Faculty and learners' guide, Washington DC, 2014.

Key Concepts

- EPAs are not an alternative for competencies, but a means to translate competencies into clinical practice.
- Competencies are descriptors of physicians.
- EPAs are descriptors of work.
- An EPA usually requires multiple competencies in an integrative, holistic nature.

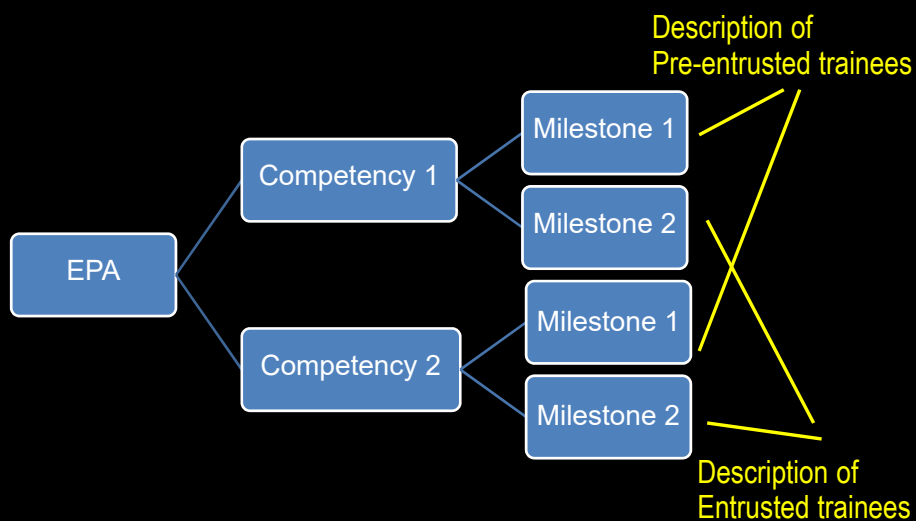
EPAs and Competencies

EPA	Med knowledge	Patient care	Interpersonal skills	Professionalism	Practice-based learning	Systems-based practice
Performing appendectomy	x	x				
Executing a patient handover	x	x	x			x
Designing therapy protocol	x				x	
Chairing multidisciplinary meeting		x	x	x		x
Request organ donation			x	x		
Manage CRF		x	x	x		x

Milestone

- Stages in the development of specific competencies
- Milestones may link to a supervisor's EPA decisions

How They Related?



Assessing Trainees

1. Observation but no execution
2. Execution with direct, proactive supervision
3. Execution with reactive supervision (i.e., on request)
4. Supervision at a distance
5. Supervision provided by the trainee to more junior colleagues

Cate OT. Nuts and bolts of entrustable professional activities. JGME 2013.

Workplace-based Assessment

เชิดศักดิ์ ไอรมนิรัตน์

ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล

มหาวิทยาลัย มหิดล

Workplace-based Assessment

- A number of assessment methods, suitable for providing feedback based on observation of trainee performance in the workplace.
 - Mini-clinical Evaluation Exercise (mini-CEX)
 - Clinical Encounter Card (CEC)
 - Blinded Patient Encounter (BPE)
 - Direct Observation of Procedural Skills (DOPS)
 - Procedure based assessment (PBA)
 - Case-based Discussion (CbD)
 - Multisource Feedback (MSF)

WPBA: Characteristics

- เป็นการประเมินที่ให้ผู้เรียนเป็นผู้เริ่มต้น
- ผู้เรียนสามารถขอให้อาจารย์ประเมินได้ตลอดช่วงเวลาปฏิบัติงาน
- เป็นการประเมินในสถานที่ปฏิบัติงานจริง
- ประเมินซ้ำได้ หากคิดคะแนนจะใช้คะแนนครั้งที่ดีที่สุด
- จุดมุ่งหมายสำคัญคือการเปิดโอกาสให้อาจารย์ได้ **feedback**

WPBA: Strengths

- Validity: assessment of “does” level
- Identify students in needs of support early
- Provide feedback
- Create a nurturing culture
- Samples widely in many workplaces
- Utilize a number of assessors

General Medical Council. Workplace based assessment: A guide for implementation, April 2010.

WPBA: Limitations

- Low reliability
- Can be opportunistic
- Trainees may delay or avoid assessment
- Learner dependent and vulnerable
- Require time and training
- Bias due to the interaction between trainers and trainees

General Medical Council. Workplace based assessment: A guide for implementation, April 2010.

Mini-Clinical Evaluation Exercise

- นักศึกษาแสดงการ **approach** ผู้ป่วยจริงในคลินิกหรือหออผู้ป่วย ขณะที่ได้รับการสังเกตการณ์โดยอาจารย์
 - Focused history taking
 - Focused physical examination
 - Making clinical diagnosis
 - Develop a management plan
- ใช้เวลาในการ **approach** ผู้ป่วย 15 นาทีต่อราย ตามด้วยการให้ **feedback** จากอาจารย์อีก 5 นาที
- อาจารย์ให้คะแนนแต่ละทักษะด้วย **rating scale 1-9**

Mini-CEX

Please refer to www.hcat.nhs.uk for guidance on this form and details of expected competencies for F1

Mini-Clinical Evaluation Exercise (CEX) - F1 Version

Please complete the questions using a cross: ☒ Please use black ink and CAPITAL LETTERS

Doctor's Surname: _____
Forename: _____
GMC Number: _____ **GMC NUMBER MUST BE COMPLETED**

Clinical setting: A&E OPD In-patient Acute Admission GP Surgery

Clinical problem category: Airway/Breathing CVS/Circulation Gastro Neuro Pain Psych/Behav Other

New or FU: New FU Focus of clinical encounter: History Diagnosis Management Explanation

Number of times patient seen before by trainee: 0 1-4 5-9 >10 Complexity of case: Low Average High

Assessor's position: Consultant GP SpR SASG SHO Other

Number of previous mini-CEXs observed by assessor with any trainee: 0 1 2 3 4 5-9 >9

Please grade the following areas using the scale below:

	Below expectations for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion	U/C*
1. History Taking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Physical Examination Skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Communication Skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Clinical Judgement	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Professionalism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Organisation/Efficiency	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Overall clinical care	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Direct Observation of Procedural Skills (DOPS)

- ประเมินทักษะการทำหัตถการในขณะทำงานกับผู้ป่วยจริง
- อาจารย์สังเกตขั้นตอนการทำหัตถการแล้วให้คะแนนด้วย rating scale 1-6 คะแนนในแต่ละมิติของการประเมิน
- แต่ละหัตถการทำการประเมินโดยอาจารย์หลายท่าน ในหลายบริบท
- แต่ละหัตถการใช้เวลาสังเกต 15 นาที และ feedback 5 นาที
- ตัวอย่างหัตถการ: endotracheal intubation, nasogastric tube insertion, IV injection, arterial blood sampling, etc.

DOPS

Please refer to www.jcat.nhs.uk for guidance on this form and details of expected competencies for F1

Direct Observation of Procedural Skills (DOPS) - F1 Version

Please complete the questions using a cross: X Please use black ink and CAPITAL LETTERS

Doctor's Surname: _____ Forename: _____

GMC Number: _____ **GMC NUMBER MUST BE COMPLETED**

Clinical setting: AGC OPD In patient Acute Admission GP Surgery

Procedure Number: _____ Other: _____

Assessor's position: Consultant GP SpR SACG AHP Nurse Specialist Nurse Other (please specify): _____

Number of previous DOPS observed by assessor with any trainee: 0 1 2 3 4 5-9 >9

Number of times procedure performed by trainee: 0 1-4 5-9 >10

Difficulty of procedure: Low Average High

Please grade the following areas using the scale below:

	Below expectations for F1 completion	Meets expectations for F1 completion	Exceeds expectations for F1 completion	UKC*
1. Demonstrates understanding of indications, relevant anatomy, technique of procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Outlets informed consent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Demonstrates appropriate preparation pre-procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Appropriate analgesia or safe sedation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Technical ability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Aseptic technique	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Seeks help where appropriate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Post procedure management	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Communication skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Consideration of patient/professionalism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Overall ability to perform procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

*UKC Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Case-based Discussion (CbD)

- นักศึกษาเลือกผู้ป่วย 2 รายที่ตนเคยดูแลนำเสนอให้อาจารย์
- อาจารย์ผู้ประเมินเลือก 1 ใน 2 ผู้ป่วยนั้นเพื่อทำการอภิปรายรายละเอียดของผู้ป่วย
 - Clinical assessment
 - Investigations
 - Treatment
 - Follow-up and future plan
- วัตถุประสงค์เพื่อประเมิน clinical reasoning skills
- การอภิปรายผู้ป่วยแต่ละรายใช้เวลาไม่เกิน 20 นาที และมี feedback 5 นาที

CbD

Please refer to curriculum at www.mmc.nhs.uk for details of expected competencies for F1 and F2

Case-based Discussion (CbD) - F2 Version

Please complete the questions using a cross (X) Please use black ink and CAPITAL LETTERS

Doctors Surname: _____ Forename: _____

GMC Number: _____ **GMC NUMBER MUST BE COMPLETED**

Clinical setting: ARE OFD In-patient Acute Admission GP Surgery

Clinical problem category: Pain Airway/Breathing CVS/Circulation Psych/Behav Neuro Gastro Other

Focus of clinical encounter: Medical Record Keeping Clinical Assessment Management Professionalism

Complexity of case: Low Average High Assessor's position: Consultant SpR GP

Please grade the following areas using the scale below:	Below expectations for F2 completion		Borderline for F2 completion	Meets expectations for F2 completion		Above expectations for F2 completion	U/C*
	1	2		4	5		
1 Medical record keeping	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Clinical assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Investigation and referrals	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 Treatment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Follow-up and future planning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 Professionalism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Overall clinical judgement	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

*U/C: Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Procedure-Based Assessment (PBA)

- A form of workplace-based assessment
- An assessor completes the form based on observation of a trainee performs a surgical procedure
- Six domains: consent, pre-operative planning, exposure and closure, intraoperative technique, postoperative management
- Two groups of items: general items, task-specific items
- Binary rating: satisfactory, unsatisfactory

Marriott J et al. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. BJS 2011; 98: 450-7.

WPBA Guidelines

- The purpose of WPBA must be clear to both trainers and trainees
 - Formative
 - Summative
- Transparent mapping of WPBA to the curriculum is essential

General Medical Council. Workplace based assessment: A guide for implementation, April 2010.

WPBA Guidelines (2)

- Setting up the WPBA
 - Environment: constructive environment, low ratings are acceptable
 - A framework to support trainees in planning WPBA
 - Multiple assessments by a range of assessors
- Roles of assessors
 - Training
 - Provide written records of feedback

WPBA Guidelines (3)

- Roles and responsibilities of trainees
 - Monitor their own progress
 - Pay attention to feedback
- Quality management
 - Constant monitoring of the implementation of WPBA

Questions & Comments

CherdsakIramaneerat@gmail.com

27

Summary

- EPA
 - Definitions: EPA, competencies, milestones
 - Key concepts
 - How to proceed with EPA?
 - Assessment in EPA framework
- WPBA
 - Characteristics
 - Examples: Mini-CEX, DOPS, CbD, PBA
 - Guidelines for implementation

*"I have failed many times,
and that's why I am a success."*

Michael Jordan



Tomorrow's Doctors, Tomorrow's Cures®



Core Entrustable Professional Activities for Entering Residency

EPA 1 Toolkit: Gather a History and Perform a Physical Examination

Learn

Serve

Lead

Association of
American Medical Colleges



EPA 1 Toolkit: Gather a History and Perform a Physical Examination

Association of American Medical Colleges
Washington, D.C.

Association of
American Medical Colleges



EPA 1 Workgroup

Beth Barron, MD, Columbia University
Philip Orlander, MD, The University of Texas Health Science Center at Houston
Michael L. Schwartz, PhD, Yale University

Senior Editors

Vivian Obeso, MD, Florida International University
David Brown, MD, Florida International University
Carrie Phillipi, MD, PhD, Oregon Health & Science University

Editors

Meenakshy Aiyer, MD, University of Illinois
Beth Barron, MD, Columbia University
Jan Bull, MA, Association of American Medical Colleges
Teresa J. Carter, EdD, Virginia Commonwealth University
Matthew Emery, MD, MSc, Michigan State University
Colleen Gillespie, PhD, New York University
Mark Hormann, MD, The University of Texas Health Science Center at Houston
Abbas Hyderi, MD, MPH, University of Illinois
Carla Lupi, MD, Florida International University
Michael L. Schwartz, PhD, Yale University
Margaret Uthman, MD, The University of Texas Health Science Center at Houston
Eduard E. Vasilevskis, MD, MPH, Vanderbilt University
Sandra Yingling, PhD, University of Illinois

AAMC Staff

Alison Whelan, MD
Chief Medical Education Officer

Chris Hanley, MBA
Project Manager

Lynn Shaull, MA
Senior Research Specialist

For inquiries and correspondence, contact Dr. Vivian Obeso at vobeso@fiu.edu, Carrie Phillipi at phillica@ohsu.edu, or Dr. Alison Whelan at awhelan@aamc.org.

This is a publication of the Association of American Medical Colleges. The AAMC serves and leads the academic medicine community to improve the health of all. aamc.org

© 2017 Association of American Medical Colleges. May be reproduced and distributed with attribution for educational or noncommercial purposes only.

Suggested Toolkit Citation:

Obeso V, Brown D, Aiyer M, Barron B, Bull J, Carter T, Emery M, Gillespie C, Hormann M, Hyderi A, Lupi C, Schwartz ML, Uthman M, Vasilevskis EE, Yingling S, Phillipi C, eds.; for Core EPAs for Entering Residency Pilot Program. *Toolkits for the 13 Core Entrustable Professional Activities for Entering Residency*. Washington, DC: Association of American Medical Colleges; 2017. aamc.org/initiatives/coreepas/publicationsandpresentations.

Suggested One-Page Schematic Citation:

Barron B, Orlander P, Schwartz ML. *Core Entrustable Professional Activities for Entering Residency—EPA 1 Schematic: Gather a History and Perform a Physical Examination*. Obeso V, Brown D, Phillipi C, eds. Washington, DC: Association of American Medical Colleges; 2017. aamc.org/initiatives/coreepas/publicationsandpresentations.



Core Entrustable Professional Activities for Entering Residency



Contents

- [User Guide](#).....2
- [One-Page Schematics](#).....3
- [Frequently Asked Questions](#).....5
- [EPA 1 Schematic](#)7
- [Appendix 1: Core EPA Pilot Supervision and Coactivity Scales](#).....8
- [Appendix 2: Resources Related to EPA 1](#)10
- [Appendix 3: Behaviors and Vignettes](#)12
- [Appendix 4: The Physician Competency Reference Set \(PCRS\)](#).....13
- [References](#).....17
 - [Publications From the Core EPA Pilot](#)17
 - [Other Related Publications](#).....17



User Guide

This toolkit is for medical schools interested in implementing the Core Entrustable Professional Activities (EPAs) for Entering Residency. Written by the AAMC Core EPA Pilot Group, the toolkit expands on the EPA framework outlined in the *EPA Developer's Guide* (AAMC 2014). The Pilot Group identified progressive sequences of student behavior that medical educators may encounter as students engage in the medical school curriculum and became proficient in integrating their clinical skills. These sequences of behavior are articulated for each of the 13 EPAs in one-page schematics to provide a framework for understanding EPAs; additional resources follow.

This toolkit includes:

- One-page schematic of each EPA
- Core EPA Pilot supervision and coactivity scales
- List of resources associated with each EPA
- Reference to EPA bulleted behaviors and vignettes from the *Core EPA Guide*
- The Physician Competency Reference Set
- Opportunities for engagement with the Core EPA Pilot



One-Page Schematics

In 2014, the AAMC launched a pilot project with 10 institutions to address the feasibility of implementing 13 EPAs for entering residency in undergraduate medical education. To standardize our approach as a pilot and promote a shared mental model, the Core EPA Pilot Group developed one-page schematics for each of the 13 EPAs.

These schematics were developed to translate the rich and detailed content within *The Core Entrustable Professional Activities for Entering Residency Curriculum Developers' Guide* published in 2014 by the AAMC into a one-page, easy-to-use format (AAMC 2014). These one-page schematics of developmental progression to entrustment provide user-friendly descriptions of each EPA. We sought fidelity to the original ideas and concepts created by the expert drafting panel that developed the *Core EPA Guide*.

We envision the one-page schematics as a resource for:

- Development of curriculum and assessment tools
- Faculty development
- Student understanding
- Entrustment committees, portfolio advisors, and others tracking longitudinal student progress

Understanding the One-Page Schematic

Performance of an EPA requires integration of multiple competencies (Englander and Carraccio 2014). Each EPA schematic begins with its list of key functions and related competencies. The functions are followed by observable behaviors of increasing ability describing a medical student's development toward readiness for indirect supervision. The column following the functions lists those behaviors requiring immediate correction or remediation. The last column lists expected behaviors of an entrustable learner.

The members of the Curriculum and Assessment Team of the Core EPA Pilot Group led this initiative. Thirteen EPA groups, each comprising representatives from four to five institutions, were tasked with creating each EPA schematic. Development of the schematics involved an explicit, standardized process to reduce variation and ensure consistency with functions, competencies, and the behaviors explicit in the *Core EPA Guide*. Behaviors listed were carefully gathered from the *Core EPA Guide* and reorganized by function and competency and listed in a developmental progression. The Curriculum and Assessment Team promoted content validity by carrying out iterative reviews by telephone conference call with the members of the Core EPA Pilot Group assigned to each EPA.

EPA Curriculum and Assessment

Multiple methods of teaching and assessing EPAs throughout the curriculum will be required to make a summative entrustment decision about residency readiness. The schematics can help to systematically identify and map curricular elements required to prepare students to perform EPAs. Specific prerequisite curricula may be needed to develop knowledge, skills, and attitudes before the learner engages in practice of the EPA.

To implement EPAs, medical schools should identify where in the curriculum EPAs will be taught, practiced, and assessed. Among other modalities, simulation, reflection, and standardized and structured experiences will all provide data about student competence. However, central to the concept of entrustment is the global performance of EPAs in authentic clinical settings, where the EPA is taught and assessed holistically, not as the sum of its parts.



Core Entrustable Professional Activities for Entering Residency



Workplace-Based Assessments: Supervision and Coactivity Scales

On a day-to-day basis, clinical supervisors make and communicate judgments about how much help (coactivity) or supervision a student or resident needs. “Will I let the student go in the room without me? How much will I let the student do versus observe? Because I wasn’t present to observe, how much do I need to double-check?” Scales for clinical supervisors to determine how much help or supervision a student needs for a specific activity have been proposed (Chen et al 2015; Rekman et al 2016). There is limited validity evidence for these scales, and no published data comparing them. Given our initial experience, the Core EPA Pilot Group has agreed on a trial using modified versions of these scales (Appendix 1).

Resources

The Pilot Group compiled a list of resources, including relevant Critical Synthesis Packages from MedEdPORTAL®, a review of current existing literature, teaching methods, and assessment tools related to each EPA (Appendix 2). This collection of products may help schools with implementation. For example, schools may find the teaching methods and assessment tools useful when considering multiple sources of data about student performance that may eventually contribute to a summative entrustment decision. The Pilot Group concluded that new teaching methods and assessment tools will be needed to complement these resources. This need is particularly relevant for workplace-based assessments where the synthetic performance of an EPA is linked to a level of supervision. We envision the one-page schematics as a resource for the development of new teaching and assessment methods.



Core Entrustable Professional Activities for Entering Residency



Frequently Asked Questions

Why are EPAs important?

In many cases, medical school graduates are perceived by residency program directors as insufficiently prepared at the beginning of their residency training for indirect supervision in clinical skills and for exhibiting professional behaviors. The EPAs define a shared set of clinical activities that residents are expected to perform on day one of residency. This is an important opportunity for undergraduate medical education to develop a new construct toward preparedness and, as an end goal, improvements in patient safety. Ideally, students will perform the Core EPAs consistently in situations of varying complexity as they practice and receive actionable feedback, formulating learning goals for future demonstrations of competence.

What does “entrustment” mean in the context of the EPAs?

Entrustment is defined as trustworthiness in applying knowledge, skills, and attitudes in performance of an EPA. To be “trustworthy,” students must consistently demonstrate attributes such as conscientiousness, knowledge of their own limits and help-seeking behavior (discernment), and truthfulness (Kennedy et al 2008). Throughout medical education, students should be assessed on trustworthiness—though this may occur implicitly or explicitly. The EPA framework makes this assessment explicit and transparent.

EPA entrustment is defined as a judgment by a supervisor or collection of supervisors signaling a student has met specific, defined expectations for needing limited supervision. The Core EPA Pilot Group recommends the formation of an entrustment committee to make evidence-based summative entrustment decisions about each student’s readiness for residency (Brown et al 2017).

What is the relationship between competencies and EPAs?

The EPA framework reorganizes competencies into observable units of clinical work by function. Each function is a subunit of work required to perform an EPA. The functions and related competencies are the parts, and the EPA is the whole. The Toolkit’s one-page schematics highlight an EPA’s specific functions with underlying competencies into observable behaviors within a developmental progression toward entrustment.

Although tracking progression within individual functions can help learners develop appropriate skills, monitoring learner progress toward entrustability for that EPA requires synthesis: At some point the learner must apply each of the functions in execution of the EPA task. *To this end, we emphasize the importance of the holistic nature of the EPA and prioritize assessment for entrustment in these activities in workplace settings as a whole, not as the sum of their parts.*

Is the one-page schematic designed as a rubric for student assessment?

No, the one-page schematics are not intended to serve as assessment tools. They can serve as guides for development of instructional, feedback, and assessment tools for EPAs. We share them as a framework for understanding the developmental progression that graduating medical students should demonstrate as a reflection of their readiness for residency.



How can I or my institution become more involved?

Medical schools in the AAMC pilot, those interested in implementing EPAs, and those wondering about the faculty resources needed to teach and assess EPAs are already part of a dynamic learning community. Opportunities for engaging with others exist through the AAMC Core EPA listserv, conference presentations, collaborative projects, and in informal medical education networks. Your contributions help shape the work of the Core EPA Pilot project and are a source of new ideas, feedback, and suggestions for implementation. We invite you to continue your conversations with us by sharing the decisions you face within the unique culture of your institution.

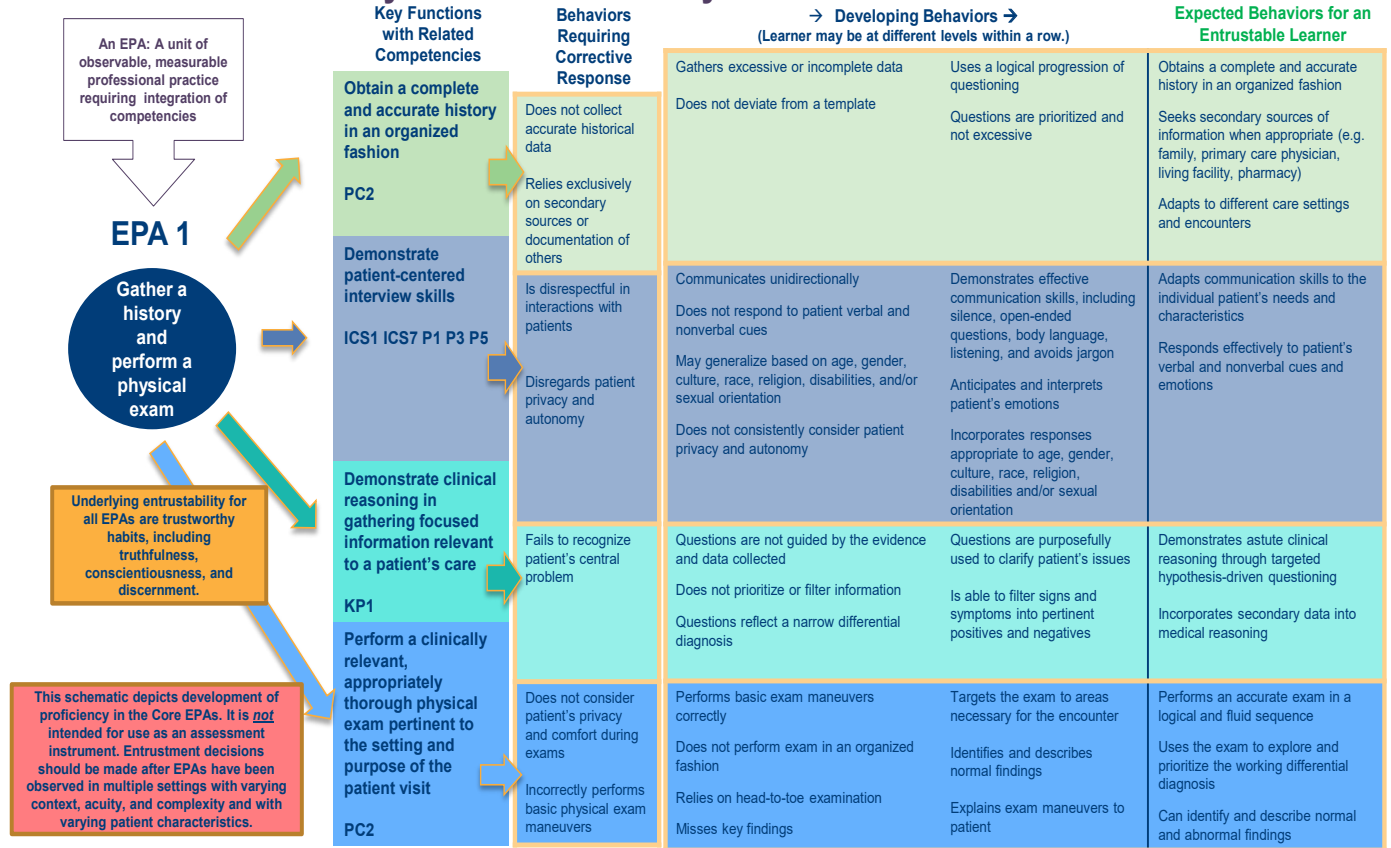
- To subscribe to the Core EPAs listserv, send a blank email to subscribe-coreepas@lists.aamc.org. To post a comment to the listserv, simply send an email to coreepas@lists.aamc.org.
- Core EPA Pilot Website: <https://www.aamc.org/initiatives/coreepas/>
- Publications from the Core EPA Pilot Group:
<https://www.aamc.org/initiatives/coreepas/publicationsandpresentations/>
- Core EPA Pilot Group email for queries and observations: coreepas@aamc.org



Core Entrustable Professional Activities for Entering Residency



EPA 1: Gather a History and Perform a Physical Examination



Barron, B, Orlander, P, Schwartz, ML, Obeso V, Brown D, Phillipi C, eds.; for Core EPAs for Entering Residency Pilot Program
 Adapted from the Association of American Medical Colleges (AAMC). Core entrustable professional activities for entering residency. 2014.



Appendix 1: Core EPA Pilot Supervision and Coactivity Scales

Scales for clinical supervisors to determine how much help (coactivity) or supervision they judge a student needs for a specific activity have been proposed—the Chen entrustment scale and the Ottawa scale (Chen et al 2015; Rekman et al 2016). There is limited validity evidence for these scales and no published data comparing them. We include these published tools here for your reference. The Core EPA Pilot Group has agreed on a trial using modified versions of these scales (described below). A description of how the pilot is working with these scales is available on the [Core EPA website](#).

<p>Modified Chen entrustment scale: If you were to supervise this student again in a similar situation, which of the following statements aligns with how you would assign the task?</p>	<p>Corresponding excerpt from original Chen entrustment scale (Chen et al 2015)</p>
<p>1b. “Watch me do this.”</p>	<p>1b. Not allowed to practice EPA; allowed to observe</p>
<p>2a. “Let’s do this together.”</p>	<p>2a. Allowed to practice EPA only under proactive, full supervision as coactivity with supervisor</p>
<p>2b. “I’ll watch you.”</p>	<p>2b. Allowed to practice EPA only under proactive, full supervision with supervisor in room ready to step in as needed</p>
<p>3a. “You go ahead, and I’ll double-check all of your findings.”</p>	<p>3a. Allowed to practice EPA only under reactive/on-demand supervision with supervisor immediately available, all findings double-checked</p>
<p>3b. “You go ahead, and I’ll double-check key findings.”</p>	<p>3b. Allowed to practice EPA only under reactive/on demand supervision with supervisor immediately available, key findings double-checked</p>



Core Entrustable Professional Activities for Entering Residency

Modified Ottawa scale: In supervising this student, how much did you participate in the task?	Original Ottawa scale (Rekman et al 2016)
1. “I did it.” Student required complete guidance or was unprepared; I had to do most of the work myself.	1. “I had to do.” (i.e., requires complete hands-on guidance, did not do, or was not given the opportunity to do)
2. “I talked them through it.” Student was able to perform some tasks but required repeated directions.	2. “I had to talk them through.” (i.e., able to perform tasks but requires constant direction)
3. “I directed them from time to time.” Student demonstrated some independence and only required intermittent prompting.	3. “I had to prompt them from time to time.” (i.e., demonstrates some independence, but requires intermittent direction)
4. “I was available just in case.” Student functioned fairly independently and only needed assistance with nuances or complex situations.	4. “I needed to be there in the room just in case.” (i.e., independence but unaware of risks and still requires supervision for safe practice)
5. (No level 5: Students are ineligible for complete independence in our systems.)	5. “I did not need to be there.” (i.e., complete independence, understands risks and performs safely, practice ready)



Appendix 2: Resources Related to EPA 1

Hypothesis-Driven Physical Examination (HDPE)

Uchida T, Heiman H. Critical synthesis package: hypothesis-driven physical examination (HDPE). MedEdPORTAL Publications. 2013;9:9435. doi.org/10.15766/mep.2374-8265.9435.

Mini-Clinical Evaluation Exercise

Perkowski L. Critical synthesis package: mini-clinical evaluation exercise (mCEX). MedEdPORTAL Publications. 2014;10:9793. doi.org/10.15766/mep.2374-8265.9793.

Faculty Observer Rating Scale (FORS)

Nadir N. Critical synthesis package: faculty observer rating scale (FORS). MedEdPORTAL Publications. 2014;10:9853. doi.org/10.15766/mep.2374-8265.9853.

Interpreter Scale (IS)

Pelts M, Albright D. Critical synthesis package: interpreter scale (IS). MedEdPORTAL Publications. 2014;10:9845. doi.org/10.15766/mep.2374-8265.9845.

Patient-Practitioner Orientation Scale (PPOS)

Trapp S, Stern M. Critical synthesis package: patient-practitioner orientation scale (PPOS). MedEdPORTAL Publications. 2013;9:9501. doi.org/10.15766/mep.2374-8265.9501.

Assessment of Professional Behaviors (APB)

Fornari A, Akbar S, Tyler S. Critical synthesis package: assessment of professional behaviors (APB). MedEdPORTAL Publications. 2014;10:9902. doi.org/10.15766/mep.2374-8265.9902.

MAAS-Global Manual 2000

Lacy N. Critical synthesis package: MAAS-global. MedEdPORTAL Publications. 2015;11:10028. dx.doi.org/10.15766/mep.2374-8265.10028.

Cross-Cultural Counseling Inventory–Revised (CCCI-R)

Young K. Critical synthesis package: cross-cultural counseling inventory–revised (CCCI-R). MedEdPORTAL Publications. 2014;10:9950. doi.org/10.15766/mep.2374-8265.9950.

CAM Health Belief Questionnaire (CHBQ)

Nicolais C, Stern M. Critical synthesis package: CAM health belief questionnaire (CHBQ). MedEdPORTAL Publications. 2014;10:9882. doi.org/10.15766/mep.2374-8265.9882.

Relational Communication Scale (RCS)



Hartmark-Hill J. Critical synthesis package: relational communication scale (RCS). MedEdPORTAL Publications. 2013;9:9454. doi.org/10.15766/mep_2374-8265.9454.

Communication Assessment Tool (CAT)

Ibrahim H. Critical synthesis package: communication assessment tool (CAT). MedEdPORTAL Publications. 2014;10:9806. dx.doi.org/10.15766/mep_2374-8265.9806.

Liverpool Communication Skills Assessment Scale (LCSAS)

Islam L, Dorflinger L. Critical synthesis package: Liverpool communication skills assessment scale (LCSAS). MedEdPORTAL Publications. 2015;11:10126. dx.doi.org/10.15766/mep_2374-8265.10126.

Communication Curriculum Package

Hofert S, Burke M, Balighian E, Serwint J. Improving provider-patient communication: a verbal and non-verbal communication skills curriculum. MedEdPORTAL Publications. 2015;11:10087. dx.doi.org/10.15766/mep_2374-8265.10087.

Professionalism Mini-Evaluation Exercise (P-MEX)

Gathright M. Critical synthesis package: professionalism mini-evaluation exercise (P-MEX). MedEdPORTAL Publications. 2014;10:9929. doi.org/10.15766/mep_2374-8265.9929.

Rochester Communication Rating Scale

Stalburg C. Critical synthesis package: Rochester communication rating scale. MedEdPORTAL Publications. 2015;11:9969. doi.org/10.15766/mep_2374-8265.9969.

Evidence in the Literature

Gowda D, Blatt B, Fink MJ, Kosowicz LY, Baecker A, Silvestri RC. A core physical exam for medical students: results of a national survey. *Acad Med*. 2014;89(3):436-442. doi: 10.1097/acm.000000000000137.



Appendix 3: Behaviors and Vignettes

The [Core EPA Guide](#) produced by the AAMC contains additional detailed information that may be useful for curriculum designers.

1. For a convenient list of behaviors for this EPA that were used to develop a developmental progression, we refer you to the [Core EPA Guide](#).
2. For exemplars of learner vignettes that highlight pre-entrustable and entrustable scenarios, please see the [Core EPA Guide](#).



Core Entrustable Professional Activities for Entering Residency

Appendix 4: The Physician Competency Reference Set (PCRS)

The Physician Competency Reference Set (Englander et al 2013) is provided for cross-referencing with the one-page schematic.

1. PATIENT CARE (PC): Provide patient-centered care that is compassionate, appropriate, and effective for the treatment of health problems and the promotion of health

- 1.1 Perform all medical, diagnostic, and surgical procedures considered essential for the area of practice
- 1.2 Gather essential and accurate information about patients and their condition through history-taking, physical examination, and the use of laboratory data, imaging, and other tests
- 1.3 Organize and prioritize responsibilities to provide care that is safe, effective, and efficient
- 1.4 Interpret laboratory data, imaging studies, and other tests required for the area of practice
- 1.5 Make informed decisions about diagnostic and therapeutic interventions based on patient information and preferences, up-to-date scientific evidence, and clinical judgment
- 1.6 Develop and carry out patient management plans
- 1.7 Counsel and educate patients and their families to empower them to participate in their care and enable shared decision making
- 1.8 Provide appropriate referral of patients, including ensuring continuity of care throughout transitions between providers or settings and following up on patient progress and outcomes
- 1.9 Provide health care services to patients, families, and communities aimed at preventing health problems or maintaining health
- 1.10 Provide appropriate role modeling
- 1.11 Perform supervisory responsibilities commensurate with one's roles, abilities, and qualifications

2. KNOWLEDGE FOR PRACTICE (KP): Demonstrate knowledge of established and evolving biomedical, clinical, epidemiological, and social-behavioral sciences, as well as the application of this knowledge to patient care

- 2.1 Demonstrate an investigatory and analytic approach to clinical situations
- 2.2 Apply established and emerging biophysical scientific principles fundamental to health care for patients and populations
- 2.3 Apply established and emerging principles of clinical sciences to diagnostic and therapeutic decision making, clinical problem solving, and other aspects of evidence-based health care
- 2.4 Apply principles of epidemiological sciences to the identification of health problems, risk factors, treatment strategies, resources, and disease prevention/health promotion efforts for patients and populations
- 2.5 Apply principles of social-behavioral sciences to provision of patient care, including assessment of the impact of psychosocial-cultural influences on health, disease, care-seeking, care compliance, and barriers to and attitudes toward care
- 2.6 Contribute to the creation, dissemination, application, and translation of new health care knowledge and practices



Core Entrustable Professional Activities for Entering Residency



3. PRACTICE-BASED LEARNING AND IMPROVEMENT (PBLI): Demonstrate the ability to investigate and evaluate their care of patients, to appraise and assimilate scientific evidence, and to continuously improve patient care based on constant self-evaluation and lifelong learning

- 3.1 Identify strengths, deficiencies, and limits in one's knowledge and expertise
- 3.2 Set learning and improvement goals
- 3.3 Identify and perform learning activities that address one's gaps in knowledge, skills, or attitudes
- 3.4 Systematically analyze practice using quality-improvement methods, and implement changes with the goal of practice improvement
- 3.5 Incorporate feedback into daily practice
- 3.6 Locate, appraise, and assimilate evidence from scientific studies related to patients' health problems
- 3.7 Use information technology to optimize learning
- 3.8 Participate in the education of patients, families, students, trainees, peers, and other health professionals
- 3.9 Obtain and utilize information about individual patients, populations of patients, or communities from which patients are drawn to improve care
- 3.10 Continually identify, analyze, and implement new knowledge, guidelines, standards, technologies, products, or services that have been demonstrated to improve outcomes

4. INTERPERSONAL AND COMMUNICATION SKILLS (ICS): Demonstrate interpersonal and communication skills that result in the effective exchange of information and collaboration with patients, their families, and health professionals

- 4.1 Communicate effectively with patients, families, and the public, as appropriate, across a broad range of socioeconomic and cultural backgrounds
- 4.2 Communicate effectively with colleagues within one's profession or specialty, other health professionals, and health-related agencies (see also interprofessional collaboration competency, IPC 7.3)
- 4.3 Work effectively with others as a member or leader of a health care team or other professional group (see also IPC 7.4)
- 4.4 Act in a consultative role to other health professionals
- 4.5 Maintain comprehensive, timely, and legible medical records
- 4.6 Demonstrate sensitivity, honesty, and compassion in difficult conversations (e.g., about issues such as death, end-of-life issues, adverse events, bad news, disclosure of errors, and other sensitive topics)
- 4.7 Demonstrate insight and understanding about emotions and human responses to emotions that allow one to develop and manage interpersonal interactions

5. PROFESSIONALISM (P): Demonstrate a commitment to carrying out professional responsibilities and an adherence to ethical principles

- 5.1 Demonstrate compassion, integrity, and respect for others
- 5.2 Demonstrate responsiveness to patient needs that supersedes self-interest
- 5.3 Demonstrate respect for patient privacy and autonomy



Core Entrustable Professional Activities for Entering Residency



- 5.4 Demonstrate accountability to patients, society, and the profession
- 5.5 Demonstrate sensitivity and responsiveness to a diverse patient population, including but not limited to diversity in gender, age, culture, race, religion, disabilities, and sexual orientation
- 5.6 Demonstrate a commitment to ethical principles pertaining to provision or withholding of care, confidentiality, informed consent, and business practices, including compliance with relevant laws, policies, and regulations
- 6. SYSTEMS-BASED PRACTICE (SBP): Demonstrate an awareness of and responsiveness to the larger context and system of health care, as well as the ability to call effectively on other resources in the system to provide optimal health care**
- 6.1 Work effectively in various health care delivery settings and systems relevant to one's clinical specialty
- 6.2 Coordinate patient care within the health care system relevant to one's clinical specialty
- 6.3 Incorporate considerations of cost awareness and risk-benefit analysis in patient and/or population-based care
- 6.4 Advocate for quality patient care and optimal patient care systems
- 6.5 Participate in identifying system errors and implementing potential systems solutions
- 6.6 Perform administrative and practice management responsibilities commensurate with one's role, abilities, and qualifications
- 7. INTERPROFESSIONAL COLLABORATION (IPC): Demonstrate the ability to engage in an interprofessional team in a manner that optimizes safe, effective patient- and population-centered care**
- 7.1 Work with other health professionals to establish and maintain a climate of mutual respect, dignity, diversity, ethical integrity, and trust
- 7.2 Use the knowledge of one's own role and those of other professions to appropriately assess and address the health care needs of the patients and populations served
- 7.3 Communicate with other health professionals in a responsive and responsible manner that supports the maintenance of health and the treatment of disease in individual patients and populations
- 7.4 Participate in different team roles to establish, develop, and continuously enhance interprofessional teams to provide patient- and population-centered care that is safe, timely, efficient, effective, and equitable
- 8. PERSONAL AND PROFESSIONAL DEVELOPMENT (PPD): Demonstrate the qualities required to sustain lifelong personal and professional growth**
- 8.1 Develop the ability to use self-awareness of knowledge, skills, and emotional limitations to engage in appropriate help-seeking behaviors
- 8.2 Demonstrate healthy coping mechanisms to respond to stress
- 8.3 Manage conflict between personal and professional responsibilities
- 8.4 Practice flexibility and maturity in adjusting to change with the capacity to alter behavior
- 8.5 Demonstrate trustworthiness that makes colleagues feel secure when one is responsible for the care of patients
- 8.6 Provide leadership skills that enhance team functioning, the learning environment, and/or the health care delivery system



Core Entrustable Professional Activities for Entering Residency



- 8.7 Demonstrate self-confidence that puts patients, families, and members of the health care team at ease
- 8.8 Recognize that ambiguity is part of clinical health care and respond by using appropriate resources in dealing with uncertainty



References

Publications From the Core EPA Pilot Group

Brown DR, Gillespie CC, Warren JB. [EPA 9—Collaborate as a member of an interprofessional team: a short communication from the AAMC Core EPAs for Entering Residency Pilot Schools](#). *Med Sci Educ*. 2016;26(3):457-461.

Brown DR, Warren JB, Hyderi A, Drusin RE, Moeller J, Rosenfeld M, et al. [Finding a path to entrustment in undergraduate medical education: a progress report from the AAMC Core Entrustable Professional Activities for Entering Residency Entrustment Concept Group](#). *Acad Med*. 2017;92(6):774-779.

Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. [Toward a common taxonomy of competency domains for the health professions and competencies for physicians](#). *Acad Med*. 2013;88(8):1088-1094.

Englander R, Carraccio C. [From theory to practice: making entrustable professional activities come to life in the context of milestones](#). *Acad Med*. 2014;89(10):1321-1323.

Favreau MA, Tewksbury L, Lupi C, Cutrer WB, Jokela JA, Yarris LM. [Constructing a shared mental model for faculty development for the Core Entrustable Professional Activities for Entering Residency](#). *Acad Med*. 2017;92(6):759-764.

Lomis KD, Ryan MS, Amiel JM, Cocks PM, Uthman MO, Esposito KF. [Core entrustable professional activities for entering residency pilot group update: considerations for medical science educators](#). *Med Sci Educ*. 2016;26(4):797-800.

Lomis K, Amiel JM, Ryan MS, et al. [Implementing an entrustable professional activities framework in undergraduate medical education: early lessons from the AAMC core entrustable professional activities for entering residency pilot](#). *Acad Med*. 2017;92(6):765-770.

Other Related Publications

Association of American Medical Colleges (AAMC). Core entrustable professional activities for entering residency. mededportal.org/icolaborative/resource/887. May 28, 2014. Accessed March 1, 2016.

Chen HC, van den Broek WS, ten Cate O. [The case for use of entrustable professional activities in undergraduate medical education](#). *Acad Med*. 2015;90(4):431-436.

Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. [Toward a common taxonomy of competency domains for the health professions and competencies for physicians](#). *Acad Med*. 2013;88(8):1088-1094.

Kennedy TJT, Regehr G, Baker GR, Lingard L. [Point-of-care assessment of medical trainee competence for independent clinical work](#). *Acad Med*. 2008;83(10):S89-S92.

Peters H, Holzhausen Y, Boscardin C, ten Cate O, Chen HC. [Twelve tips for the implementation of EPAs for assessment and entrustment decisions](#). *Med Teach*. 2017;39(8):802-807.

Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. [A new instrument for assessing resident competence in surgical clinic: the Ottawa clinic assessment tool](#). *J Surg Educ*. 2016;73(4):575-582.

RIP OUT

Nuts and Bolts of Entrustable Professional Activities

OLLE TEN CATE, PHD

The Challenge

The entrustable professional activity (EPA) concept allows faculty to make competency-based decisions on the level of supervision required by trainees. Competency-based education targets standardized levels of proficiency to guarantee that all learners have a sufficient level of proficiency at the completion of training.¹⁻⁶ Collectively, the competencies (ACGME or CanMEDS) constitute a framework that describes the qualities of professionals. Such a framework provides generalized descriptions to guide learners, their supervisors, and institutions in teaching and assessment. However, these frameworks must translate to the world of medical practice. EPAs were conceived to facilitate this translation, addressing the concern that competency frameworks would otherwise be too theoretical to be useful for training and assessment in daily practice.

What Is Known

Trust is a central concept for safe and effective health care. Patients must trust their physicians, and health care providers must trust each other in a highly interdependent health care system. In teaching settings, supervisors decide when and for what tasks they entrust trainees to assume clinical responsibilities. Building on this concept, EPAs are units of professional practice, defined as tasks or responsibilities to be entrusted to the unsupervised execution by a trainee once he or she has attained sufficient specific competence. EPAs are independently executable, observable, and measurable in their process and outcome, and therefore, suitable for entrustment decisions. Sequencing EPAs of increasing difficulty, risk, or sophistication can serve as a backbone for graduate medical education.⁶

How Do EPAs Differ From Competencies?

- EPAs are not an alternative for competencies, but a means to translate competencies into clinical practice.
- Competencies are descriptors of physicians, EPAs are descriptors of work.
- EPAs usually require multiple competencies in an integrative, holistic nature. TABLE 1 shows how different EPAs require proficiency in several competency domains.

Olle ten Cate, PhD, is Professor of Medical Education and Director of the Center for Research & Development of Education at the University Medical Center Utrecht, the Netherlands.

Corresponding author: Th J (Olle) ten Cate, PhD, PO Box 85500, 3508 GA Utrecht, the Netherlands, t.j.tencate@umcutrecht.nl

DOI: <http://dx.doi.org/10.4300/JGME-D-12-00380.1>

What Is Included in a Full EPA Description?

An EPA must be described at a sufficient level of detail to set trainee expectations and guide supervisor's assessment and entrustment decisions (see TABLE 2 for guidelines).

How Do EPAs Relate to Milestones?

Milestones, as defined by the ACGME, are stages in the development of specific competencies. Milestones may link to a supervisor's EPA decisions (eg, direct proactive supervision versus distant supervision). The Pediatrics Milestone Project provides examples of how milestones can be linked to entrustment decisions.^{7,8}

What Do Entrustment Decisions Require?

Entrustment decisions involve clinical skills and abilities as well as more general facets of competence, such as understanding one's own limitations and knowing when to ask for help. Making entrustment decisions for unsupervised practice requires observed proficiency, usually on multiple occasions.

In practice, entrustment decisions are affected by 4 groups of variables: (1) attributes of the trainee (tired, confident, level of training); (2) attributes of the supervisors (eg, lenient or strict); (3) context (eg, time of the day, facilities available); and (4) the nature of the EPA (rare, complex versus common, easy). Entrustment decisions can be further distinguished as ad hoc (eg, happening during a night shift) or structural (establishing the recognition that a trainee may do this activity at a specific level of supervision from now on). In the clinical context, many ad hoc entrustment decisions happen every day. Structural entrustment decisions formally acknowledge that a trainee has passed a threshold that allows for decreased supervision. The certificate awarded at such occasions has been called a *statement of awarded responsibility* (STAR) and should be carefully documented.²

Linking an EPA with a competency framework emphasizes essential competency domains when observing a trainee executing the EPA.

How You Can Start TODAY

Decide how many EPAs are useful for training.

While there can be many EPAs that serve to make ad hoc entrustment decisions, EPAs that lead to structural entrustment decisions (ie, certification or STARS) should involve broad-based responsibilities and be limited in number. For a graduate medical education program, no more than 20 to 30 EPAs are recommended.

RIP OUT

TABLE 1 EXAMPLES OF EPAs RELATED TO THEIR MOST IMPORTANT ACGME COMPETENCY DOMAINS						
ACGME Competencies						
Illustrative EPAs	MK	PC	ISC	P	PBLI	SBP
Performing an appendectomy	•	•				
Executing a patient handover	•	•	•			•
Designing a therapy protocol	•				•	
Chairing a multidisciplinary meeting		•	•	•		•
Requesting organ donation			•	•		
Chronic disease management		•	•	•		•

Abbreviation: EPAs, entrustable professional activities; ACGME, Accreditation Council for Graduate Medical Education; MK, Medical Knowledge; PC, Patient Care; ISC, Interpersonal Skills and Communication; P, Professionalism; PBLI, Practice-Based Learning and Improvement and SBP, Systems-Based Practice.

Use of EPAs in Assessing Trainees

EPAs can be the focus of assessment. The key question is: Can we trust this trainee to execute this EPA? The answer may be translated to 5 levels of supervision for the EPA:

1. Observation but no execution, even with direct supervision
2. Execution with direct, proactive supervision
3. Execution with reactive supervision, ie, on request and quickly available

4. Supervision at a distance and/or post hoc
5. Supervision provided by the trainee to more junior colleagues

What You Can Do LONG TERM

- Review the specialty requirements and milestones, and work with your professional organization and local colleagues to identify EPAs.
- Detail the EPAs, following TABLE 2.
- Prepare faculty to provide EPA-based assessments.
- Use structural entrustment decisions as a “license” for trainees to execute EPAs with distant supervision.

Resources

- 1 ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ.* 2005;39(12):1176–1177.
- 2 ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542–547.
- 3 Mulder H, ten Cate O, Daalder R, Berkvens J. Building a competency-based workplace curriculum around entrustable professional activities: the case of physician assistant training. *Med Teach.* 2010;32(10):e453–e459.
- 4 ten Cate O, Young JQ. The patient handover as an entrustable professional activity: adding meaning in teaching and practice. *BMJ Qual Saf.* 2012. 2012 21: i9–i12. doi: 10.1136/bmjqs-2012-001213.
- 5 Chang A, Bowen JL, Buranosky RA, Frankel RM, Ghosh N, Rosenblum MJ, et al. Transforming primary care training-patient-centered medical home entrustable professional activities for internal medicine residents [published online ahead of print September 21, 2012]. *J Gen Int Med.* DOI: 10.1007/s11606-012-2193-3
- 6 Nasca TJ. The Next Accreditation System, June 2012. <http://www.acgme-nas.org/assets/pdf/Nasca%20NAS%20June%202012%20Presentation%20Slide%20Show.pdf>. Accessed October 21, 2012.
- 7 Hicks PJ, Schumacher DJ, Benson BJ, Burke AE, Englander R, Guralnick S, et al. The pediatrics milestones: conceptual framework, guiding principles, and approach to development. *J Grad Med Educ.* 2010;2(3):410–418.
- 8 Pediatrics Milestone Project. http://www.acgme.org/acgmeweb/Portals/0/PFAAssets/ProgramResources/320_PedsMilestonesProject.pdf. Accessed October 14, 2012.

TABLE 2 GUIDELINES FOR FULL ENTRUSTABLE PROFESSIONAL ACTIVITIES DESCRIPTIONS	
1. Title	Make it short; avoid words related to proficiency or skill. Ask yourself: Can a trainee be scheduled to do this? Can an entrustment decision for unsupervised practice for this EPA be made and documented?
2. Description	To enhance universal clarity, include everything necessary to specify the following: What is included? What limitations apply? Limit the description to the actual activity. Avoid justifications of why the EPA is important, or references to knowledge and skills.
3. Required Knowledge, Skills, and Attitudes (KSAs)	Which competency domains apply? Which subcompetencies apply? Include only the most relevant ones. These links may serve to build observation and assessment methods.
4. Required KSAs	Which KSAs are necessary to execute the EPA? Formulate this in a way to set expectations. Refer to resources that reflect necessary or helpful standards (books, a skills course, etc).
5. Information to assess progress	Consider observations, products, monitoring of knowledge and skill, multisource feedback.
6. When is unsupervised practice expected?	Estimate when full entrustment for unsupervised practice is expected, acknowledging the flexible nature of this. Expectations of entrustment moments can shape an individual workplace curriculum.
7. Basis for formal entrustment decisions	How many times must the EPA be executed proficiently for unsupervised practice? Who will judge this? What does formal entrustment look like (documented, publicly announced)?

การประเมินผลในบริบทของการทำงาน (Workplace-based assessment)

รศ.นพ.เชิดศักดิ์ ไอรมนรัตน์

ภาควิชาศัลยศาสตร์

การประเมินผลมีบทบาทในการส่งเสริมการเรียนรู้ของนักศึกษาได้จากหลายกลไก ซึ่งกลไกหนึ่งที่มีความสำคัญมากและอาจารย์แพทย์ควรมีการใช้มากขึ้นคือการใช้การประเมินผลในระหว่างเรียน (formative assessment) เพื่อให้ได้ข้อมูลว่านักศึกษามีระดับความรู้ ความสามารถมากน้อยเพียงใด ยังต้องพัฒนาในด้านใดบ้าง แล้วนำข้อมูลดังกล่าวให้แก่นักศึกษา (feedback) เพื่อให้ นักศึกษาได้พัฒนาตนเองให้ดีขึ้นก่อนที่จะถูกประเมินผลในตอนสิ้นสุดการเรียน (summative assessment) งานวิจัยหลายชิ้นแสดงให้เห็นว่านักศึกษาแพทย์และแพทย์ประจำบ้านได้รับการสังเกตและประเมินผลการทำงานในลักษณะของ formative assessment นี้ไม่เพียงพอ ซึ่งส่งผลให้นักศึกษาและแพทย์ประจำบ้านเหล่านี้ขาดโอกาสที่จะพัฒนาความรู้ และทักษะพื้นฐานในการดูแลผู้ป่วย อย่างมีประสิทธิภาพ

การประเมินผลในบริบทของการทำงาน (Workplace-based assessment) เป็นกลุ่มของวิธีการประเมินผลที่ถูกพัฒนาขึ้น เพื่อให้อาจารย์แพทย์ได้มีโอกาสประเมินความรู้ และทักษะต่างๆทางคลินิกของนักศึกษาในขณะที่ทำงานกับผู้ป่วยจริง และได้นำผลการประเมินนั้นมาชี้แนะแนวทางในการพัฒนาความรู้ และทักษะของนักศึกษา วิธีการประเมินผลในกลุ่มนี้มีลักษณะสำคัญต่างๆ ดังนี้

1. เป็นการประเมินผลที่ให้นักศึกษาเป็นผู้เริ่มต้น กล่าวคือ นักศึกษาไม่ต้องรอถึงวันที่กำหนดโดยอาจารย์ว่าจะทำการสอบในวันใด เมื่อไรที่นักศึกษาพบโอกาสเหมาะได้พบผู้ป่วยที่ตนสามารถแสดงระดับความรู้ และทักษะทางคลินิกของตนให้อาจารย์แพทย์ประเมินได้ก็ทำการขอให้อาจารย์ช่วยประเมินผลได้ทันที
2. นักศึกษาสามารถขอให้อาจารย์ประเมินได้ตลอดช่วงเวลาที่ทำกรปฏิบัติงานในภาควิชา ไม่จำเป็นต้องรอถึงสิ้นสุดการปฏิบัติงาน การเปิดโอกาสให้ประเมินได้ตั้งแต่เริ่มปฏิบัติงานทำให้อาจารย์สามารถเห็นจุดบกพร่องของนักศึกษาแต่แรกในขณะที่ยังมีเวลาให้นักศึกษาได้ฝึกฝนพัฒนาตนเอง
3. นักศึกษาสามารถขอรับการประเมินทักษะเดิมซ้ำได้จนกว่านักศึกษาจะมีทักษะดังกล่าวดีเป็นที่น่าพอใจ โดยคะแนนที่จะนำไปตัดสินผลการศึกษาคือคะแนนครั้งที่นักศึกษาทำได้ดีที่สุด
4. จุดมุ่งหมายสำคัญของการประเมินผลนี้คือการเปิดโอกาสให้อาจารย์ได้สังเกตนักศึกษาปฏิบัติงานกับผู้ป่วยจริงแล้วให้ข้อมูลย้อนกลับ (feedback) แก่นักศึกษา
5. นักศึกษาเป็นผู้รับผิดชอบในการดำเนินการรวบรวมคะแนนของตนเอง และตรวจสอบว่าตนยังต้องทำการประเมินทักษะใดอีกบ้าง

วิธีการประเมินผลกลุ่มนี้มีด้วยกันหลายวิธี เช่น

1. Mini-clinical evaluation exercise (Mini-CEX) เป็นการประเมินทักษะการตรวจรักษาผู้ป่วยที่แผนกผู้ป่วยนอก หรือในหอผู้ป่วยโดยอาจารย์แพทย์ให้เวลานักศึกษาซักประวัติ ตรวจร่างกายผู้ป่วยราว 15 นาทีแล้วจึงทำการอภิปรายและทำการให้คะแนนร่วมกับให้ feedback แก่นักศึกษา

2. Direct Observation of Procedural Skills (DOPS) เป็นการประเมินทักษะการทำหัตถการพื้นฐาน โดยอาจารย์สังเกตนักศึกษาทำหัตถการดังกล่าวกับผู้ป่วย ซึ่งมักเป็นหัตถการที่ใช้เวลาทำไม่นานนัก เสร็จแล้วอาจารย์ให้คะแนนและให้ feedback แก่นักศึกษา
 3. Case-based discussion (CbD) เป็นการประเมินทักษะการคิดวิเคราะห์แก้ปัญหาทางคลินิกของนักศึกษา โดยนักศึกษาเลือกผู้ป่วยสองรายที่ตนเคยดูแลเสนอให้อาจารย์ผู้ประเมิน อาจารย์ทำการเลือกหนึ่งในสองรายของผู้ป่วยให้นักศึกษานำเสนอประวัติและแนวทางในการตรวจวินิจฉัย และการรักษา ร่วมกับการที่อาจารย์ซักถามเพื่อประเมินความเข้าใจในผู้ป่วยของนักศึกษา เมื่ออภิปรายผู้ป่วยดังกล่าวเสร็จแล้ว อาจารย์ให้คะแนนและให้ feedback แก่นักศึกษา
 4. Multisource feedback (MSF) เป็นการเก็บรวบรวมใบประเมินการปฏิบัติงานที่กรอกโดยบุคลากรที่ทำงานร่วมกับนักศึกษาที่หลากหลาย ได้แก่ อาจารย์ พยาบาล แพทย์ประจำบ้าน เพื่อนนักศึกษา แล้วนำผลการประเมินที่รวบรวมได้มาสรุปแล้วนำไปให้ feedback แก่นักศึกษาให้เห็นข้อมูลในภาพรวมว่าการทำงานของเขาในสายตาของเพื่อนร่วมงานนั้นมีประสิทธิภาพเพียงใด มีสิ่งใดที่นักศึกษาควรปรับปรุงบ้าง
- หากอาจารย์มีความต้องการจะนำผลการประเมินผลในบริบทของการทำงานมาใช้มีข้อเสนอแนะแนวปฏิบัติบางประการเพื่อช่วยให้การประเมินผลนี้บรรลุตามวัตถุประสงค์

1. ต้องทำการชี้แจงให้ทั้งนักศึกษาและอาจารย์ที่เกี่ยวข้องในการประเมินผลทุกท่านเข้าใจถึงวัตถุประสงค์ของการประเมินนี้ที่ชัดเจนว่ามุ่งเน้นที่ formative assessment
2. ต้องทำการสร้างบรรยากาศที่เอื้อให้เกิดการประเมินผลในรูปแบบนี้ โดยทำให้ทุกคนเข้าใจว่าเป็นการประเมินเพื่อให้เกิดการพัฒนาในตัวนักศึกษา ดังนั้นไม่จำเป็นต้องคาดหวังว่าจะต้องได้คะแนนสูงในการประเมินทุกครั้ง และต้องสร้างกระบวนการช่วยกระตุ้นให้นักศึกษาล้ำที่จะเชิญชวนให้อาจารย์ทำการประเมินนักศึกษาเมื่อมีโอกาส และให้นักศึกษาสามารถเข้าหาอาจารย์ได้หลากหลายท่าน
3. ต้องมีการกำหนดบทบาทที่ชัดเจนของอาจารย์ว่าในการประเมินรูปแบบนี้หน้าที่ของอาจารย์ไม่เพียงแต่ให้คะแนนตามใบประเมินแล้ว อาจารย์ยังต้องมีหน้าที่สอน และให้ข้อมูลย้อนกลับ (feedback) แก่นักศึกษาด้วย
4. ต้องสร้างความเข้าใจในกลุ่มนักศึกษาว่าตัวนักศึกษาเองมีความรับผิดชอบในการหาโอกาสที่จะรับการประเมินด้วยตนเอง สนใจรับฟัง feedback จากอาจารย์เพื่อนำไปปรับปรุงตนเองให้ดีขึ้น และตรวจสอบว่าตนเองจะต้องประเมินทักษะใดบ้าง
5. การดำเนินการให้ประสบความสำเร็จต้องอาศัยความช่วยเหลือจากเจ้าหน้าที่สายสนับสนุนทางการศึกษาในการรวบรวมข้อมูลการประเมินผล และตรวจสอบว่ามีปัญหาใดเกิดขึ้นในกระบวนการประเมินผลหรือไม่ เช่น จำนวนผู้ป่วยที่เหมาะสมสำหรับการประเมินนักศึกษามีเพียงพอหรือไม่ มีอาจารย์ท่านใดไม่มีส่วนร่วมในการประเมินนักศึกษาหรือไม่ มีอุปสรรคใดทำให้นักศึกษาไม่สามารถรับการประเมินได้หรือไม่ เป็นต้น

หากอาจารย์สามารถนำรูปแบบการประเมินผลในบริบทของการทำงานไปประยุกต์ใช้ในการพัฒนาการเรียนการสอนในภาควิชาต่างๆได้ น่าจะทำให้เพิ่ม constructive feedback ให้แก่นักศึกษาและก่อให้เกิดการพัฒนาความรู้ และทักษะพื้นฐานทางการแพทย์ในนักศึกษาแพทย์ได้อย่างมีประสิทธิภาพมากขึ้น

AMEE GUIDE

Workplace-based assessment as an educational tool: AMEE Guide No. 31

JOHN NORCINI¹ & VANESSA BURCH²¹Foundation for Advancement of International Medical Education and Research, Philadelphia, USA, ²University of Cape Town, South Africa

Abstract

Background: There has been concern that trainees are seldom observed, assessed, and given feedback during their workplace-based education. This has led to an increasing interest in a variety of formative assessment methods that require observation and offer the opportunity for feedback.

Aims: To review some of the literature on the efficacy and prevalence of formative feedback, describe the common formative assessment methods, characterize the nature of feedback, examine the effect of faculty development on its quality, and summarize the challenges still faced.

Results: The research literature on formative assessment and feedback suggests that it is a powerful means for changing the behaviour of trainees. Several methods for assessing it have been developed and there is preliminary evidence of their reliability and validity. A variety of factors enhance the efficacy of workplace-based assessment including the provision of feedback that is consistent with the needs of the learner and focused on important aspects of the performance. Faculty plays a critical role and successful implementation requires that they receive training.

Conclusions: There is a need for formative assessment which offers trainees the opportunity for feedback. Several good methods exist and feedback has been shown to have a major influence on learning. The critical role of faculty is highlighted, as is the need for strategies to enhance their participation and training.

Introduction

For just over two decades leading educationists, including medical educators, have highlighted the intimate relationship between learning and assessment. Indeed, in an educational context it is now argued that learning is the key purpose of assessment (van der Vleuten 1996; Gronlund 1998, Shepard 2000). At the same time as this important connection was being stressed in the education literature; there were increasing concerns about the workplace-based training of doctors. A study by Day et al. (1990) in the United States documented that the vast majority of first-year trainees in internal medicine were not observed more than once by a faculty member in a patient encounter where they were taking a history or doing a physical examination. Without this observation, there was no opportunity for the assessment of basic clinical skills and, more importantly, the provision of feedback to improve performance.

As one step in encouraging the observation of performance by faculty, the American Board of Internal Medicine proposed the use of the mini-Clinical Evaluation Exercise (mini-CEX) (Norcini et al. 1995). In the mini-CEX, a faculty member observes a trainee as he/she interacts with a patient around a focused clinical task. Afterwards, the faculty member assesses the performance and provides the trainee feedback. It was expected that trainees would be assessed several times throughout the year of training with different faculty and in different clinical situations.

Practice points

- The research literature on work-based formative assessment and feedback suggests that it is a powerful means for changing the behaviour of learners.
- Several formative assessment methods have been developed for use in the workplace and there is preliminary data evidence of their reliability and validity.
- The efficacy of feedback is enhanced if it is consistent with the needs of the learner, focuses on important aspects of the performance in the work-place, and has characteristics such as being timely and specific.
- Faculty development is critical to the quality and effectiveness of formative assessment.
- Strategies to encourage the participation of faculty are critical to the successful implementation of formative assessment.

An advantage of the mini-CEX and other workplace-based methods is that they fulfil the three basic requirements for assessment techniques that facilitate learning (Frederiksen 1984; Crooks 1988; Swanson et al. 1995; Shepard 2000): (1) The content of the training programme, the competencies expected as outcomes, and the assessment practices are aligned (2) Trainee feedback is provided during and/or after assessment

Correspondence: John Norcini, Foundation for Advancement of International Medical Education and Research (FAIMER) 4th Floor 3624 Market St, Philadelphia PA 19104, USA. Tel: 1 215 823 2170; fax: 1 215 386 2321; email: JNorcini@faimer.org

events;(3) Assessment events are used strategically to steer trainee learning towards the desired outcomes. Over the past several years there has been growing interest in workplace-based assessment and additional methods have been (re)introduced to the setting of clinical training (National Health Service 2007).

Previous publications have focused on the advantages and disadvantages of workplace-based methods from the perspective of assessment alone (Norcini 2007). In this role, the methods are best thought of as analogous to classroom tests and they have much strength from this perspective. However, it is difficult to assure equivalence across institutions and the observations of faculty may be influenced by the stakes and their relationships with trainees. Consequently, their use faces challenges as national high stakes assessment devices.

Perhaps more importantly, workplace-based assessment can be instrumental in the provision of feedback to trainees to improve their performance and steer their learning towards desired outcomes. This paper focuses on the use of the methods for this purpose and it is divided into five sections. The first section briefly reviews the literature on the efficacy and prevalence of formative assessment and feedback. This is followed by a section that describes some of the more common methods of work-based assessment. The third section concentrates on feedback and it is explored from the perspective of the learner, its focus, and which characteristics make it effective in the context of formative assessment. Faculty play a key role in the successful implementation of formative assessment, so the fourth section describes strategies to encourage their participation and training to improve their performance. In the closing section we draw attention to the challenges faced by medical educators implementing formative assessment strategies in routine clinical teaching practice.

Efficacy and prevalence of formative assessment and feedback

The purpose of formative assessment and feedback

Formative assessment is not merely intended to assign grades to trainee performance at designated points in the curriculum; rather it is designed to be an ongoing part of the instructional process and to support and enhance learning (Shepard 2000). Clearly, feedback is a core component of formative assessment (Sadler 1989), central to learning, and at '*the heart of medical education*' (Branch & Paranjape 2002). In fact, it is useful to consider feedback as part of an ongoing programme of assessment and instruction rather than a separate educational entity (Hattie & Timperley 2007).

Feedback promotes student learning in three ways (Gipps 1999, Shepard 2000):

- it informs trainees of their progress or lack thereof;
- it advises trainees regarding observed learning needs and resources available to facilitate their learning; and
- it motivates trainees to engage in appropriate learning activities.

Efficacy of feedback

Given these presumed benefits, it is appropriate to ask whether there is a body of research supporting the efficacy of feedback in changing trainees' behaviour. Most compelling is a synthesis of information on classroom education by Hattie which included over 500 meta-analyses involving 1,800 studies and approximately 25 million students (Hattie 1999). He demonstrated that the typical effect size (ES) of schooling on overall student achievement is about 0.40 (i.e. it increases the mean on an achievement test by 0.4 of a standard deviation). Using this as a benchmark or 'gold standard' on which to judge the various factors that affect performance, Hattie summarized the results of 12 meta-analyses that specifically included the influence of feedback. The feedback effect size was 0.79, which is certainly very powerful, and among the four biggest influences on achievement. Hattie also found considerable variability based on the type of feedback, with the largest effect being generated by the provision of information around a specific task.

Data to answer the question about the efficacy of feedback are much more limited in the domain of medical education but a recent meta-analysis by Veloski and colleagues looked at its effect on clinical performance (Veloski et al. 2006). Of the 41 studies meeting the criteria for inclusion, 74% demonstrated a positive effect for feedback alone. When combined with other educational interventions, feedback had a positive effect in 106 of the 132 (77%) studies reviewed.

A recent paper by Burch and colleagues reports on the impact of a formative assessment strategy implemented in a 4th year undergraduate medical clerkship programme (Burch et al. 2006). In this paper, students who engaged in an average of 6 directly observed clinical encounters during a 14-week clerkship reported that they more frequently undertook blinded patient encounters (McLeod & Meagher 2001) in which they did not consult the patient records before interviewing and examining the patient. Prior to implementing the formative assessment programme, students traditionally interviewed and examined patients only after consulting patient records. In addition they reported that they read more frequently on topics only relevant to patients clerked in the ward. While this paper provides information on self-reported learning behaviour changes, it does suggest that formative assessment may have the potential to strategically direct student learning by reinforcing desirable learning behaviour (Gibbs 1999).

A recent publication by Driessen and van der Vleuten (2000) support the findings reported by Burch. In their study they introduced a portfolio of learning assignments as an educational tool in a legal skills training programme comprising tutorials which were poorly attended and for which students did not adequately complete the required pre-tutorial work. The portfolio assignments, such as writing a legal contract or drafting a legislative document, were reviewed by peers and the tutor prior to being used as the teaching basis for subsequent skills training sessions. This educational intervention resulted in a twofold increase in time spent preparing for skills training sessions.

Prevalence of feedback

It is clear from these data that formative assessment and feedback have a powerful influence on trainee performance. However, there is a significant gap between what should be done and 'on the ground' practice. Lack of assessment and feedback, based on observation of performance in the workplace, is one of the most serious deficiencies in current medical education practice (Holmboe et al. 2004; Kassebaum & Eaglen 1999). Indeed, direct observation of trainee performance appears to be the exception rather than the rule.

In a survey of 97 United States medical schools, accredited between 1993 and 1998, it was found that structured, observed assessments of students' clinical abilities were done across clinical clerkships for only 7.4% to 23.1% of medical students (Kassebaum and Eaglen 1999). A more recent survey of medical graduates found that during any given core clerkship, 17% to 39% of student were not observed performing a clinical examination (Association of American Medical Colleges 2004). Likewise, Kogan & Hauer (2006) found that only 28% of Internal Medicine clerkships included an in-course formative assessment strategy involving observation of student performance in the workplace setting. Outside the US, Daelmans et al. (2004) reported that over a 6-month period, observation of trainee performance occurred in less than 35% of educational events in which observation and the provision of feedback could have taken place.

Unfortunately the situation is no better in postgraduate training programmes. In one study, 82% of residents reported that they engaged in only one directly observed clinical encounter in their first year of training; far fewer (32%) engaged in more than one encounter (Day et al. 1990). In another survey of postgraduate trainees 80% reported never or only infrequently receiving feedback based on directly observed performance (Isaacson et al. 1995).

Not only is assessment of directly observed performance infrequently done as part of routine educational practice, but the quality of feedback, when given, may be poor. Holmboe colleagues evaluated the type of feedback given to residents after mini-CEX encounters and observed that while 61% of feedback sessions included a response from the trainee to the feedback, only 34% elicited any form of self-evaluation by the trainee. Of greatest concern, however, was the finding that only 8% of mini-CEX encounters translated into a plan of action (Holmboe et al. 2004a). The paper by Holmboe and colleagues suggests that there are key reasons why clinician-educators fail to give trainees effective feedback (see Box1):

In addition to finding that trainee observation and feedback is infrequently given and often of limited value, it has also been noted that the faculties' assessment of trainee performance may be less than completely accurate. Noel and colleagues found that faculty failed to detect 68% of errors committed by postgraduate trainees when observing a videotape scripted to depict marginal competence (Noel et al. 1992). The use of checklists prompting faculty to look for specific skills increased error detection from 32% to 64%. It was, however, noted that this did not improve the accuracy of assessors. Approximately two thirds of faculty still scored the overall performance of marginal postgraduate trainees as

Box 1. Key reasons why clinician-educators fail to give trainees effective feedback.

- Current in-vivo assessment strategies such as the mini-CEX may be focusing on assessment of performance at the expense of providing adequate feedback.
- The scoring sheets currently used for in-vivo assessment events provide only limited space for recording comments thereby limiting feedback given.
- Clinician-educators do not fully appreciate the role of feedback as a fundamental clinical teaching tool.
- Clinician-educators may not be skilled in the process of providing high quality feedback.

satisfactory or superior. Similar observations attesting to the poor accuracy of faculty observations have been made elsewhere (Herbers et al. 1989; Kalet et al. 1992).

Based on the infrequency with which trainees are observed and problems with the quality of the feedback they receive, it is fair to ask whether observation of trainee performance is an outdated approach to medical training and assessment. The critical question, therefore, is whether clinical interviewing and examination skills are still relevant to clinical practice such that faculty should be trained to properly observe performance and provide effective, useful feedback.

Feedback in relation to history and physical examination

Despite major technological advances, the ability to competently interview and examine patients remains one of the mainstays of clinical practice (Holmboe et al. 2004). Data gathered over the past 30 years highlight the critical importance of these skills. In 1975 Hampton and colleagues demonstrated that a good medical history produced the final clinical diagnosis in 82% of 80 patients interviewed and examined. In only one of 80 cases did laboratory tests provide the final diagnosis not made by history or physical examination (Hampton et al. 1975).

Technological advances over the past two decades have not made the findings of this study irrelevant. In 1992 Peterson and colleagues showed that among 80 patients presenting for the first time to a primary care clinic, the patient's history provided the correct final diagnosis in 76% of cases (Peterson et al. 1992). Even more recently, an autopsy study of 400 cases showed that the combination of a history and physical examination produced the correct diagnosis in 70% of cases. Diagnostic imaging studies successfully indicated the correct diagnosis in only 35% of cases (Kirch & Schafii 1996).

Beyond diagnostic accuracy, physician-patient communication is a key component of health care. In a review of the literature, Beck et al. (2002) found that both verbal behaviours (e.g., empathy, reassurance and support) and nonverbal behaviours (e.g., nodding, forward lean) were positively associated with patient outcomes. Likewise, a study by Little et al. (2001) found that the patients of doctors who took a patient-centred approach were more satisfied, more enabled, had greater symptom relief, and had lower rates of referral.

The ability to competently interview a patient and perform a physical examination thus remains the cornerstone

of clinical practice. The ability of faculty to accurately observe trainees performing these tasks and provide effective feedback is therefore one of the most important aspects of medical training. Although methods such as standardised patients certainly provide complementary assessment and feedback information, they cannot replace the central role of observation by faculty.

Formative assessment methods

A number of assessment methods, suitable for providing feedback based on observation of trainee performance in the workplace, have been developed or regained prominence over the past decade. This section provides a brief description of the essential features of some of them including:

- Mini-Clinical Evaluation Exercise (mini-CEX);
- Clinical Encounter Cards (CEC);
- Clinical Work Sampling (CWS);
- Blinded Patient Encounters (BPE);
- Direct Observation of Procedural Skills (DOPS);
- Case-based Discussion (CbD);
- MultiSource Feedback (MSF).

Mini-clinical evaluation exercise (mini-CEX)

As described above, the mini-CEX (Figure 1, Source: www.hcat.nhs.uk) is an assessment method developed in the United States (US) that is now in use in a number of institutions around the world. It requires trainees to engage in authentic workplace-based patient encounters while being observed by faculty members (Norcini et al. 1995). Trainees perform clinical tasks, such as taking a focused history or performing relevant aspects of the physical examination, after which they provide a summary of the patient encounter along with next steps (e.g., a clinical diagnosis and a management plan).

These encounters can take place in a variety of workplace settings including inpatient, outpatient, and emergency departments. Patients presenting for the first time as well as those returning for follow up visits are suitable encounters for the mini-CEX. Not surprisingly, the method lends itself to a wide range of clinical problems including: (1) presenting complaints such as chest pain, shortness of breath, abdominal pain, cough, dizziness, low back pain; or (2) clinical problems such as arthritis, chronic obstructive airways disease, angina, hypertension and diabetes mellitus (Norcini et al. 2003).

In the original work, each aspect of the clinical encounter is scored by a faculty member using a 9-point rating scale where 1–3 is unsatisfactory, 4–6 is satisfactory and 7–9 is superior. The parameters evaluated include: interviewing skill, physical examination, professionalism, clinical judgement, counselling, organization and efficiency, and overall competence. Different scales and different parameters have been used successfully in other settings (e.g., National Health Service).

The core purpose of the assessment method is to provide structured feedback based on observed performance. Each patient encounter takes roughly 15 minutes followed by 5–10 minutes of feedback. Trainees are expected to be evaluated

several times with different patients and by different faculty members during their training period.

This assessment tool has been shown to be a reliable way of assessing postgraduate trainee performance provided there is sufficient sampling. Roughly 4 encounters are sufficient to achieve a 95% confidence interval of less than 1 (on the 9-point scale) and approximately 12–14 are required for a reliability coefficient of 0.8 (Norcini et al. 1995, 2003; Holmboe et al. 2003).

In addition to the postgraduate setting, the mini-CEX has been successfully implemented in undergraduate medical training programmes (Hauer 2000; Kogan et al. 2003; Kogan & Hauer 2006). In this context, the period of observation and feedback is often longer, ranging from 30–45 minutes (Hauer 2000; Kogan et al. 2002).

There is a growing body of evidence supporting the validity of the mini-CEX. Kogan et al. (2002, 2003) found that mini-CEX performance was correlated with other assessments collected as part of undergraduate training. Faculty ratings of videotapes of student-standardized patient encounters, using the mini-CEX forms, were correlated with the checklist scores and standardized patient ratings of communication skills (Boulet et al. 2002). In postgraduate training, mini-CEX performance was correlated with a written in-training examination and routine faculty ratings (Durning et al. 2002). Holmboe et al. (2004) found that, using the mini-CEX form, they could differentiate amongst videos, scripted to represent different levels of ability. Finally, et al. (2006) found that mini-CEX scores were correlated with the results of a Royal College oral examination.

Clinical encounter cards (CEC)

The CEC system, developed at McMaster University in Canada (Hatala & Norman 1999) and subsequently implemented in other centres (Paukert et al. 2002), is similar to the mini-CEX. The basic purpose of this assessment strategy is also to score trainee performance based on direct observation of a patient encounter. The encounter card system scores the following dimensions of observed clinical practice: history-taking, physical examination, professional behaviour, technical skill, case presentation, problem formulation (diagnosis) and problem solving (therapy). Each dimension is scored using a 6-point rating scale describing performance as 1: unsatisfactory, 2: below the expected level of student performance, 3: at the expected level of student performance, 4: above the expected level of student performance, 5: outstanding student performance, and 6: performance at the level of a medical graduate.

In addition to capturing the quality of the performance, the 4 × 6 inch score cards also provide space for assessors to record the feedback given to the trainee at the end of the encounter.

This system has been shown to be a feasible, valid, and reliable measure of clinical competence, provided that a sufficient number of encounters (approximately 8 encounters for a reliability coefficient of 0.8 or more) are collected (Hatala & Norman 1999). Moreover, introduction of the system was found to increase student satisfaction with the feedback

process (Paukert et al. 2002) and to have modest correlations with other forms of assessment (Richards et al. 2007).

Clinical work sampling (CWS)

This assessment method, developed in Canada, is also based on direct observation of clinical performance in the workplace (Turnbull et al. 2000). The method requires collection of data concerning specific patient encounters for a number of different domains either at the time of admission (admission rating form) or during the hospital stay (ward rating form). These forms are completed by faculty members directly observing trainee performance. The domains assessed by faculty include: communication skills, physical examination skills, diagnostic acumen, consultation skills, management skills, interpersonal behaviour, continued learning skills and health advocacy skills. Not all skills are evaluated on each occasion.

Trainees are also assessed by ward nursing staff (using the multidisciplinary team rating form) and the patients (using the patient rating form) who are in the care of the trainees. These rating forms, also completed on the basis of directly observed behaviour, require a global assessment and ratings of the following domains: therapeutic strategies, communications skills, consultation with other health care professionals, management of resources, discharge planning, interpersonal relations, collaboration skills, and health advocacy skills and professionalism.

All rating forms use a 5-point rating scale ranging from unsatisfactory to excellent performance. This assessment method has also been shown to be valid and reliable provided a sufficient number (approximately 7 encounters for a reliability coefficient of 0.7) of encounters are observed (Turnbull et al. 2000).

A later study found that the CWS strategy could be adapted to radiology residency using a handheld computerised device (Finlay et al. 2006). Compliance with voluntary participation was not as great as expected but this evaluation format included the opportunity to discuss performance at the time of data entry, rather than at the end of rotation. The investigators found the method less useful for summative purposes although the sample size was small ($N=14$).

Blinded patient encounters

This formative assessment method is based on the same principle as the three assessment methods already mentioned. It is unique, however, in that it forms part of undergraduate bedside teaching sessions. (Burch et al. 2006). Students, in groups of 4–5, participate in a bedside tutorial. It starts with a period of direct observation in which one of the students in the group is observed performing a focused interview or physical examination as instructed by the clinician educator conducting the teaching session. Thereafter the student is expected to provide a diagnosis, including a differential diagnosis, based on the clinical findings.

The patient is unknown to the student, hence the term 'blinded' patient encounter (McLeod & Meagher 2001). This type of patient encounter has the advantage of safely allowing the trainee to practice information gathering, hypothesis

generation, and problem solving without access to the workup by more senior doctors.

After the presentation, the session focuses on demonstrating the important clinical features of the case as well as discussing various issues, for example appropriate investigation and treatment relevant to the patient's presenting clinical problem. It concludes with a feedback session in which the student receives personal private advice about his/her performance.

Feedback is provided using a 9-point rating scale for assessment of clinical interviewing and examination skills as well as clinical reasoning skills. The rating scale ranges from 1–3 for poor performance, 4–6 for adequate performance and 7–9 for good performance. Space is provided on the score sheet to add other written comments. Students keep the score sheets which are only used for feedback purposes.

Direct observation of procedural skills (DOPS)

This assessment method (Figure 2, Source: www.hcat.nhs.uk), developed in the UK, focuses on evaluating the procedural skills of postgraduate trainees by observing them in the workplace setting (Wragg et al. 2003). Just as in CWS and the Encounter Card Assessment systems, trainees' performance is scored using a 6-point rating scale where 1–2 is below the expected level of competency, 3 reflects a borderline level of competency, 4 meets the expected level of competency and 5–6 are above the expected level of competency. The assessment procedure is generally expected to require 15 minutes of observation time and 5 minutes dedicated to feedback.

Trainees are provided with a list of commonly performed procedures for which they are expected to demonstrate competence such as endotracheal intubation, nasogastric tube insertion, administration of intravenous medication, venepuncture, peripheral venous cannulation and arterial blood sampling. They are assessed by multiple clinicians on multiple occasions throughout the training period.

This method of procedural skills assessment is not limited to postgraduate training programmes. Paukert and colleagues have included basic surgical skills to be mastered by undergraduate students in their clinical encounter card system (Paukert et al. 2002).

Although DOPS is similar to procedural skills log books, the purpose and nature of these methods differ significantly. The recording of procedures is common to both of them, but log books are usually designed to ensure that trainees have simply performed the minimum number required to be considered competent. The provision of structured feedback based on observation of a performance is not necessarily part of the log book process. Moreover, the procedure is not necessarily performed under direct observation and little feedback, if any, is expected to be given. In contrast, DOPS ensures that trainees are given specific feedback based on direct observation so as to improve their procedural skills.

Case-based discussion (CbD)

This assessment method is an anglicised version of Chart-Stimulated Recall (CSR) developed for use by the American

Please refer to www.hcat.nhs.uk for guidance on this form and details of expected competencies for F1

Direct Observation of Procedural Skills (DOPS) - F1 Version

Please complete the questions using a cross: ☒ Please use black ink and CAPITAL LETTERS

Doctor's Surname:

Forename:

GMC Number: **GMC NUMBER MUST BE COMPLETED**

Clinical setting: A&E OPD In-patient Acute Admission GP Surgery

Procedure Number: Other:

Assessor's position: Consultant GP SpR SASG AHP Nurse Specialist Nurse
 Other (please specify)

Number of previous DOPS observed by assessor with any trainee: 0 1 2 3 4 5-9 >9

Number of times procedure performed by trainee: 0 1-4 5-9 >10 Difficulty of procedure: Low Average High

Please grade the following areas using the scale below:	Difficulty of procedure:					U/C*
	Below expectations for F1 completion	Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		
1. Demonstrates understanding of indications, relevant anatomy, technique of procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Obtains informed consent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Demonstrates appropriate preparation pre-procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Appropriate analgesia or safe sedation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Technical ability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Aseptic technique	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Seeks help where appropriate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Post procedure management	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Communication skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Consideration of patient/professionalism	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Overall ability to perform procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Please use this space to record areas of strength or any suggestions for development.

Have you had training in the use of this assessment tool?: Face-to-Face Have Read Guidelines Web/CDrom

Assessor's Signature:

Date (mm/yy): /

Time taken for observation: (in minutes)

Time taken for feedback: (in minutes)

Assessor's Surname:

Assessor's registration number:

Please note: Failure of return of all completed forms to your administrator is a probity issue



Figure 2. Directly observed procedural skills form. Source: www.hcat.nhs.uk.

Board of Emergency Medicine (Maatsch et al. 1983). It is currently part of the Foundation Programme implemented for postgraduate training in the UK National Health Service. In CbD, the trainee selects two case records of patients in which they had made notes and presents them to an assessor. The assessor selects one of the two for discussion and explores one or more aspects of the case, including: clinical assessment, investigation and referral of the patient, treatment, follow-up and future planning, and professionalism (Figure 3, Source: www.mmc.nhs.uk). Since the case record is available at the time of assessment, medical record keeping can also be assessed by the examiner.

This type of performance assessment focuses on evaluating the clinical reasoning of trainees so as to understand the rationale behind decisions made in authentic clinical practice. As with other assessment methods described, each encounter is expected to last no more than 20 minutes, including 5 minutes of feedback. Trainees are expected to engage in multiple encounters with multiple different examiners during the training period.

There are several studies supporting the validity of this measure. Maatsch et al. 1983) collected several assessments for a group of practicing doctors eligible for recertification in Emergency Medicine. They found that CbD correlated with a number of the other measures, including chart audit. The score distribution and pass-fail results were consistent with scores on initial certification, ten years earlier. As importantly, CbD was considered the most valid of the measures by the practicing doctors participating in the study.

A study by Norman and colleagues compared a volunteer group of doctors to those referred for practice difficulties (Norman et al. 1989). CbD was highly correlated with a standardised patient examination and with an oral examination. More importantly, it was able to separate the volunteer group from the doctors who were referred. Likewise, Solomon et al. (1990) collected data from several different assessments on practicing doctors eligible for recertification. CbD was correlated with the oral examination as well as written and oral exams administered 10 years earlier.

MultiSource feedback (MSF)

More commonly referred to as 360-degree assessment, this method represents a systematic collection of performance data and feedback for an individual trainee, using structured questionnaires completed by a number of stakeholders. The assessments are all based on directly observed behaviour (Wragg et al. 2003) but they differ from the methods presented above in that they reflect routine performance, rather than performance during a specific patient encounter.

Although there are a number of different ways of conducting this form of assessment, the mini-peer assessment tool (mini-PAT) that has been selected for use in the Foundation Programme in the UK is a good example. Trainees nominate 8 assessors including senior consultants, junior specialists, nurses and allied health service professionals. Each of the nominated assessors receives a structured questionnaire (Figure 4) which is completed and returned to a central location for processing. Trainees also complete self-assessments, using the same

questionnaires, and submit these for processing. The categories of assessment include: good clinical care, maintaining good clinical practice, teaching and training, relationships with patients, working with colleagues and an overall assessment.

The questionnaires are collated and individual feedback is prepared for trainees. Data are provided in a graphic form which depicts the mean ratings of the assessors and the national mean rating. All comments are included verbatim, but they remain anonymous. Trainees review this feedback with their supervisor and together work on developing an action plan. This process is repeated twice yearly during the training period.

This method is widely used in industry and business, but has also been found to be useful in medicine. Applied to practicing doctors, it was able to distinguish certified from non-certified internists and the results were associated with performance on a written examination (Ramsey et al. 1989; Wenrich et al. 1993). In a follow-up study, two subscales were identified—one focused on technical/cognitive skills and the other focused on professionalism (Ramsey et al. 1993). Written examination performance was correlated with the former but not the latter.

Multisource feedback has been applied to postgraduate trainees as well as practicing doctors. The Sheffield Peer Review Assessment Tool, which is the full scale version of mini-PAT as shown in Figure 4 (Source: www.mmc.nhs.uk), was studied with paediatricians and found to be feasible and reliable (Archer et al. 2005). It also separated doctors by grade and tended to be insensitive to potential biasing factors such as the length of the working relationship. Whitehouse et al. (2002) also applied multisource feedback to postgraduate trainees with reasonable results.

Finally, this form of assessment has also been used successfully with medical students (Arnold et al. 1981, Small et al. 1993). Both positive and negative reports from peers have influenced academic actions.

Overall, reasonably reliable results can be achieved with the assessments of 8 to 12 peers.

Nature of the feedback

For the purpose of this discussion, feedback can be conceptualised as ‘*information provided by an agent (teacher, peer, self, etc.) regarding aspects of one’s performance or understanding*’ (Hattie & Timperley 2007). This information can be used by the learner to ‘*confirm, add to, overwrite, tune or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, belief about self and tasks or cognitive tactics and strategies*’ (Winnie & Butler 1994). The main purpose of feedback is, therefore, to reduce the discrepancy between current practices or understandings and desired practices or understandings (Hattie & Timperley 2007).

Perspective of the learner

In order for feedback to fulfil this purpose, it needs to address three fundamental questions for the learner:

- Where am I going?
- How am I going?
- Where to next?

Please refer to curriculum at www.mmc.nhs.uk for details of expected competencies for F1 and F2

mini-PAT (Peer Assessment Tool) - F1 Version

Please complete the questions using a cross: ☒ Please use black ink and CAPITAL LETTERS

Doctor's Surname

Forename

GMC Number:

How do you rate this Doctor in their:	Below expectations for F1 completion		Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		U/C*
	1	2	3	4	5	6	
Good Clinical Care							
1 Ability to diagnose patient problems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Ability to formulate appropriate management plans	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Awareness of their own limitations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 Ability to respond to psychosocial aspects of illness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Appropriate utilisation of resources e.g. ordering investigations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Maintaining good medical practice							
6 Ability to manage time effectively / prioritise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Technical skills (appropriate to current practice)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Teaching and Training, Appraising and Assessing							
8 Willingness and effectiveness when teaching/training colleagues	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Relationship with Patients							
9 Communication with patients	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10 Communication with carers and/or family	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11 Respect for patients and their right to confidentiality	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Working with colleagues							
12 Verbal communication with colleagues	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13 Written communication with colleagues	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14 Ability to recognise and value the contribution of others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15 Accessibility/Reliability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16 Overall, how do you rate this doctor compared to a doctor ready to complete F1 training?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Do you have any concerns about this doctor's probity or health? Yes No

If yes please state your concerns:

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment. 6927534062

Figure 4. Mini-peer assessment questionnaire. Source: www.mmc.nhs.uk.

<p>Anything especially good?</p>	<p>Please describe any behaviour that has raised concerns or should be a particular focus for development:</p>
<p>Please continue your comments on a separate sheet if required</p>	
<p>Your Gender: <input type="checkbox"/> Male <input type="checkbox"/> Female</p>	
<p>Your ethnic group:</p>	<p><input type="checkbox"/> British <input type="checkbox"/> Bangladeshi</p> <p><input type="checkbox"/> Irish <input type="checkbox"/> Other Asian Background</p> <p><input type="checkbox"/> Other White Background <input type="checkbox"/> White and Black Caribbean</p> <p><input type="checkbox"/> Caribbean <input type="checkbox"/> White and Black African</p> <p><input type="checkbox"/> African <input type="checkbox"/> White and Asian</p> <p><input type="checkbox"/> Any other Black background <input type="checkbox"/> Any other mixed background</p> <p><input type="checkbox"/> Indian <input type="checkbox"/> Chinese</p> <p><input type="checkbox"/> Pakistani <input type="checkbox"/> Any other ethnic group</p>
<p>Which environment have you primarily observed the doctor in? (Please choose one answer only)</p>	<p><input type="checkbox"/> Inpatients <input type="checkbox"/> Intensive Care</p> <p><input type="checkbox"/> Outpatients <input type="checkbox"/> Theatre</p> <p><input type="checkbox"/> Both In and Out-patients <input type="checkbox"/> General Practice</p> <p><input type="checkbox"/> A&E/Admissions <input type="checkbox"/> Other (Please specify)</p> <p><input type="checkbox"/> Community Speciality <input style="width: 100%; height: 15px;" type="text"/></p> <p><input type="checkbox"/> Laboratory/Research</p>
<p>Your position:</p>	<p><input type="checkbox"/> Consultant <input type="checkbox"/> SASG <input type="checkbox"/> SpR <input type="checkbox"/> Foundation/PRHO</p> <p><input type="checkbox"/> Nurse <input type="checkbox"/> SHO <input type="checkbox"/> Allied Health Professional</p> <p><input type="checkbox"/> GP</p> <p><input type="checkbox"/> Other (Please specify) <input style="width: 100%; height: 15px;" type="text"/></p>
<p>If you are a Nurse or AHP how long have you been qualified?:</p>	<p><input style="width: 20px; height: 15px;" type="text"/> <input style="width: 20px; height: 15px;" type="text"/> years</p>
<p>What training have you had in the use of this assessment tool?:</p>	<p><input type="checkbox"/> Face-to-Face <input type="checkbox"/> Have Read Guidelines <input type="checkbox"/> Web/CD rom</p>
<p>How long has it taken you to complete this form (in minutes)?:</p>	<p><input style="width: 20px; height: 15px;" type="text"/> <input style="width: 20px; height: 15px;" type="text"/></p>
<p>Your Signature:</p>	<p>Date: <input style="width: 20px; height: 15px;" type="text"/> / <input style="width: 20px; height: 15px;" type="text"/> / <input style="width: 20px; height: 15px;" type="text"/></p>
<p>Your Surname:</p>	<p><input style="width: 100%; height: 15px;" type="text"/></p>
<p>Your GMC Number: (Doctors only)</p>	<p><input style="width: 100%; height: 15px;" type="text"/></p>
<p>Acknowledgements: mini-PAT is derived from SPRAT (Sheffield Peer Review Assessment Tool) 5563534067</p>	

Figure 4. Continued.

To address the first question, it is critical that there be clearly defined learning goals. If the goals are not clearly articulated then *'the gap between current learning and intended learning is unlikely to be sufficiently clear for students to see a need to reduce it'* (Hattie & Timperley 2007). Goals can be wide ranging and variable, but without them students are less likely to engage in properly directed action, persist at tasks in the face of difficulties, or resume the task if disrupted (Bargh et al. 2001). The existence of goals is also more likely to lead students to seek and receive feedback, especially if they have a shared commitment to achieving them (Locke & Latham 1990). So, medical trainees need to have a clear understanding of desired practice or competence in order to seek feedback and stay focused on the task of achieving competence in the domain of interest.

The second question focuses on the provision of concrete information, derived from an assessment of the performance, relative to a task or goal. To do so well requires criteria that provide clear indicators of whether the task has been completed properly. The answer to this question addresses the traditional, restricted definition of feedback. Nonetheless, it is critical to the provision of effective feedback. Ironically, it is precisely this aspect of feedback which is usually poorly done. Clinician-educators are often reluctant to provide honest feedback, particularly in the face of poor performance. Having a set of clearly defined criteria makes it somewhat easier to provide guidance based strictly on observed performance, rather than interpretations of the trainee's intentions.

The final important question from the perspective of the trainee is what actions need to be taken in order to close the gap between actual performance and desired performance. Trainees need an action plan; specific information about how to proceed in order to achieve desired learning outcomes. As indicated previously, without honest feedback regarding actual performance, trainees are unlikely to seek advice about how to proceed in order to close the learning gap.

The interrelatedness of these questions becomes apparent when attempting to address this final question. Indeed, without clearly defined learning outcomes, including criteria which make achievement of the learning goals explicit, and honest feedback about observed performance, planning aimed at improving performance will not take place. Closing the gap between where trainees are and where they need to be is both the purpose of feedback and the source of its influence (Sadler 1989).

Focus of feedback

How effectively feedback addresses the three questions for learners is dependent in part on what aspects of the performance are addressed. Specifically, there are four foci for feedback (Hattie & Timperley 2007):

- feedback about the task;
- feedback about the process of the task;
- feedback about self-regulation;
- feedback about the self as a person.

The most basic focus of feedback addresses the quality of the task performed. Using well defined criteria, trainees are given specific information about whether they achieved the

required level of performance. This type of feedback is easiest to give, and is consequently the most frequently provided. It is most helpful when it concentrates on the performance, rather than the knowledge required for the task. The latter is best dealt with by providing direct instruction and it is not regarded as feedback (Hattie & Timperley 2007).

One of the limitations of providing feedback focused only on the task is that it is necessarily context-specific or task-specific. Consequently, it does not generalise readily to other tasks (Thompson 1998). On the other hand, providing feedback that focuses on the process can be of more value because it encourages a deeper appreciation of the performance. This involves giving feedback that enhances an understanding of relationships (the construction of meaning), cognitive processes, and transfer to different or novel situations (Marton et al. 1993). This focus for feedback is also more likely to promote deep learning (Balzer et al. 1989).

A major component of this type of feedback is the provision of strategies for error detection and correction, in other words developing the trainee's ability to provide self-feedback (Hattie & Timperley 2007). Feedback about the process underlying the task can also serve as a cueing mechanism leading to more effective information search strategies. Cueing is most useful when it assists trainees in detecting faulty hypotheses and provides direction for further searching and strategising (Harackiewicz 1979).

Feedback that focuses on self-regulation addresses the interplay between commitment, control, and confidence. It concentrates on the way trainees monitor, direct, and regulate their actions relative to the learning goal. It implies a measure of autonomy, self-control, self-direction, and self-discipline (Hattie & Timperley 2007). Effective learners are able to generate internal feedback and cognitive routines while engaged in a task (Butler & Winnie 1995).

Students who are able to self-appraise and self-manage are able to seek and receive feedback from others. At the other end of the spectrum are less effective learners who, having minimal self-regulation strategies, are more dependent on external factors, such as teachers, to provide feedback. For these learners, feedback is more effective if it directs attention back to the task and enhances feelings of self-efficacy such that trainees are likely to invest more time and become more committed to mastering the task (Kluger & DeNisi 1996).

Trainees' attributions of success and failure can have more impact than actual success or failure. Feelings of self-efficacy can be adversely affected if students are unable to relate feedback to the cause of their poor performance. In other words, feedback that does not specify the grounds on which students have achieved success or not, is likely to engender personal uncertainties and may ultimately lead to poorer performance (Thompson 1998). On the other hand, feedback that attributes performance to effort or ability is likely to increase engagement and task performance (Craven et al. 1991). Thus, when giving feedback it is critical that the assessor clearly directs the feedback to observed performance, while being aware of the impact feedback has on the self-efficacy of the trainee.

The final focus of feedback is discussed not because of its educational value but rather because it often has

adverse consequences. This feedback is typically concentrated on the personal attributes of the trainee and seldom contains task-related information, strategies to improve commitment to the task, or a better understanding of self or the task itself (Hattie & Timperley 2007). This focus for feedback is generally not effective, its impact is unpredictable, and it can have an adverse effect on learning. This is particularly true of negative feedback directed at a personal level.

Characteristics of effective feedback in the context of formative assessment

Formative assessment strategies are thought to best prompt change when they are integral to the learning process, performance assessment criteria are clearly articulated, feedback is provided immediately after the assessment event, and trainees engage in multiple assessment opportunities (Crooks 1988; Gibbs & Simpson 2004). In addition to these features, Ende (1983) suggested that specific conditions could make feedback more conducive to learning as described in Box 2.

In addition to the strategies suggested by Ende, it has also been suggested that the efficacy of feedback may be further improved by promoting trainee 'ownership' of feedback (Holmboe et al. 2004). Strategies to achieve this include:

- encouraging trainees to engage in a process of self-assessment prior to receiving external feedback;
- permitting trainees to respond to feedback;
- ensuring that feedback translates into a plan of action for the trainee.

Box 2. Specific conditions to make feedback more conducive to learning.

- Set an appropriate time and place for feedback.
- Provide feedback regarding specific behaviours, not general performance.
- Give feedback on decisions and actions, not one's interpretation of the trainees motives or intentions.
- Give feedback in small digestible quantities.
- Use language that is non-evaluative and non-judgemental.

Based on a large qualitative study, including 83 academics involved in education, Hewson & Little (1998) validated many of these literature-based recommendations. They developed a useful list of bipolar descriptors outlining feedback techniques to be adopted and avoided (Box 3).

As already mentioned, formulating an action plan at the end of a feedback session is critical to the success of formative assessment. If a plan addressing the deficiencies is not formulated, it results in failure to close the 'learning loop' and correct the identified problems (Holmboe et al. 2004). Indeed, formulation of an action plan may constitute the most critical step in providing feedback.

Beyond these actions, it is becoming increasingly recognised that ongoing coaching or mentoring improves the efficacy of feedback. This is particularly true of 360-degree feedback strategies (Luthans & Peterson 2004). Current literature in the business world reports that the role of the workplace managers has been reconceptualised such that they are seen to be facilitators of learning, creativity, and innovation rather than directors or controllers of activity. Furthermore, learning leaders or managers should foster interconnections between people and systems so as to create collective learning networks (Walker 2001). While this research has not been replicated in the medical workplace setting, the emerging success of these strategies in business suggests that similar methods merit further consideration in clinical training settings.

Faculty development

Faculty participation

From the preceding discussion it is clear that there is a need to increase the frequency of observation of trainee performance in order to provide feedback aimed at improving the quality of the services they later render in clinical practice. To this end a number of strategies have recently been implemented, but the studies of their efficacy are limited in number and they report variable success.

Holmboe and colleagues examined the impact of a scoring sheet specifically designed to remind faculty both of the dimensions of feedback and that its main purpose is to provide

Box 3. Feedback techniques to be avoided and adopted.

Feedback techniques to be avoided

Creating a disrespectful, unfriendly, closed, threatening climate
 Not eliciting thoughts or feelings before giving feedback
 Being judgemental
 Focusing on personality
 Basing feedback on hearsay
 Basing feedback on generalizations
 Giving too much/too little feedback
 Not suggesting ideas for improvement
 Basing feedback on unknown, non-negotiated goals

Feedback techniques to be adopted

Creating a respectful, open minded, non-threatening climate
 Eliciting thoughts and feelings before giving feedback
 Being non-judgemental
 Focusing on behaviours
 Basing feedback on observed facts
 Basing feedback on specifics
 Giving the right amount of feedback
 Suggesting ideas for improvement
 Basing feedback on well-defined, negotiated goals

Taken from Hewson & Little, 1998.

trainees with information about their performance aimed at improving it (Holmboe et al. 2001). In the study, the faculty control group did not receive any instruction regarding the use of the score sheet, while the intervention group received 20 minutes of instruction at the start of the clinical rotation. This information session outlined the characteristics of effective feedback and stressed the importance of direct observation of trainees to evaluate clinical competence. Results of the study indicated that while the intervention group did not provide more frequent feedback, their trainees were more satisfied with the quality of feedback they received.

Two recent studies in the Netherlands have produced similar findings. In one of the studies an undergraduate surgical clerkship was restructured in an attempt to increase the observation of trainee performance and the provision of feedback by senior faculty members (van der Hem-Stokroos et al. 2004). Restructuring of the clerkship included the introduction of a log book, a form documenting observation of skill performance, and individual appraisal by senior staff. Faculty was informed of the changes but they were not given formal instruction in trainee observation and how to provide feedback. The results indicated no significant increase in trainee observation or the provision of feedback. The authors suggest that the lack of impact of the intervention may be partly attributed to the limited input received by faculty involved in the study, particularly limited involvement in the process of restructuring the clerkship.

In the other study, Daelmans et al. (2005) introduced in-training assessment in an undergraduate medical clerkship programme. Senior clinical staff was informed about the introduction at a meeting held at the beginning of the clerkship. They also received a letter outlining the in-training assessment programme. The findings indicated that despite implementing this new programme, students were not more frequently observed performing clinical interviews and examinations in the workplace. In their discussion of the results they suggest that observation and feedback regarding student performance may have been improved if faculty members had been more frequently reminded of the programme, for example daily meetings could have been used to alert faculty to the importance and potential educational value of the programme.

In contrast to these studies, Turnbull et al. (2000) describe a strategy using clinical work sampling in which students received feedback based on directly observed patient encounters an average of eight times during a 4-week clerkship rotation. In this study, faculty members observing students in the workplace attended a 2-hour workshop outlining the assessment and feedback strategy. In addition, they received monthly communications reminding them of the project. Students were also oriented to the project before it started, and met with the research associate on a weekly basis during the clerkship rotation. Results indicated that the ongoing collection of performance data was feasible.

In another study using the clinical encounter card system, students engaged in a directly observed assessment event an average of 35 times during a 12-week surgery clerkship (Paukert et al. 2002). As in the other study, evaluators involved in the project were briefed about the project in a number of

short 15-minute meetings outlining the purpose and importance of the intervention implemented. These information sessions formed part of other meetings routinely held in the department, for example morbidity and mortality meetings. At each of these information sessions, faculty were asked to raise any issues or concerns they had regarding the project. They also received a letter explaining the assessment and feedback system prior to implementation. At the end of the clerkship, students were more satisfied with the feedback they received.

Based on these studies it is clear that a number of strategies need to be employed to successfully implement an assessment process in which trainees receive feedback based on directly observed performance in the workplace. First, it is apparent that involvement of faculty in planning an in-course formative assessment strategy is likely to enhance their engagement in the process. Second, faculty need to be thoroughly briefed about the purpose and process of the observation and feedback strategy implemented. Third, students need to be properly informed about the purpose and format of the assessment method used. In particular, it is critical that the potential learning benefits of the system are emphasized rather than the assessment aspects of the methods being used. Finally, faculty and students need to be regularly reminded of the benefit of formative assessment and the importance of keeping the assessment strategy active in the workplace.

Faculty training

While successfully implementing a formative assessment strategy in the workplace is an achievement in its own right, it is important to ensure that the quality of the observations made by attending faculty are accurate and that the feedback received by students is effective. As was highlighted earlier, faculty observations of student performance may not be sufficiently accurate to identify errors in student performance. While the use of checklists has been shown to improve the ability of assessors to detect errors in performance (Noel et al. 1992), they have not been shown to improve the overall accuracy of assessors. This is an issue that requires further research; effective strategies to address this problem clearly need to be found.

While the accuracy of examiners remains an issue needing further work, the stringency of examiners can be improved with training. A recent paper by Boulet et al. (2002) examined the stringency of examiners using the mini-CEX to evaluate directly observed trainee performance. They reported significant variability among the examiners even when they were observing the same event. Holmboe and colleagues have shown that assessor training can address this issue. In their paper, study participants engaged in a one-day video-based training session aimed at reducing variability among faculty when providing assessments and feedback on observed performance. Participants engaged in performance dimension training and frame-of-reference training (Holmboe et al. 2004). The former was accomplished by getting faculty to discuss and define key components of competence for specific clinical skills and develop criteria for satisfactory performance. The latter was addressed by giving individual faculty members the opportunity to score real-time trainee performance using

standardised patients and standardised trainees. While one faculty member scored the performance of the trainee and provided feedback, other faculty members scored the trainee's performance by watching the interview and examination on a video monitor. The encounter ended with a group discussion of how each member of the group rated the performance and reasons for the scores allocated. Finally the facilitator described what type of trainee performance the case scenario was scripted to depict.

Eight months after this faculty development effort, a set of video recordings of scripted patient encounters were again used to compare the performance of trained faculty as compared to a cohort of untrained faculty. Trained faculty were more stringent than untrained faculty members and they also reported feeling more comfortable providing trainee feedback. This study is one of the first demonstrating the beneficial impact of faculty training for the purpose of scoring performance with the intention of providing trainee feedback.

Challenges

In this closing section of the paper we wish to highlight areas where further work is needed to address some pivotal questions regarding workplace-based formative assessment and feedback. First and foremost, we need to develop strategies that will ensure successful and sustainable implementation of formative assessment in the workplace. Most of what has been done to date has been research-based, short term projects. We need studies that identify the determinants of successful, sustainable assessment and feedback strategies so that we can better understand factors that promote trainee feedback as a routine feature of training programmes rather than a unique feature of selected programmes only. Long term use may require further modification and simplification of existing methods so as to make them more user-friendly in busy clinical settings where patient care is the first priority and trainee assessment of less importance.

Based on current literature it is apparent that poor faculty participation in formative assessment and feedback strategies is probably the most significant limiting factor currently identified. Why faculty do not routinely engage in trainee assessment and feedback needs to be better understood if we wish to improve the situation. One strategy that may be of benefit would be a reward structure for busy clinicians that appropriately recognises their educational contributions and/or provides them protected time to engage in teaching activities. Another strategy would be to identify a core group of faculty whose only educational job is assessment and formative feedback. Other strategies clearly need to be identified. In any event, these realities need to be addressed before formative assessment is likely to be a routine feature of workplace-based training programmes.

Second, we need to improve the quality of the assessments and feedback given to trainees through a concerted faculty development effort. Current work indicates that feedback rarely results in the formulation of an action plan, a critical component of effective feedback, and only sometimes involves self-assessment by the trainee. Both these issues need to be addressed if feedback is to be owned by the trainee

and remedial action undertaken to improve performance. In addition, the accuracy and stringency of feedback need to be improved. Innovative strategies to address this important aspect of formative assessment need to be developed.

Finally, the impact of feedback on trainee learning behaviour and performance needs to be determined. To date there is very little information about the strategic use of formative assessment in the workplace context to drive the learning of medical trainees. The need for such data is apparent. Not only do we need to determine the impact of feedback on learning behaviour, but we also need to know what the performance-in-the-workplace benefits can be expected to be achieved by successful formative assessment strategies.

Summary

In the context of the workplace-based education of doctors, there has been concern that trainees are seldom observed, assessed, and given feedback. This has led to increasing interest in a variety of formative assessment methods that require observation and offer the opportunity for feedback, including the mini-clinical evaluation exercise, clinical encounter cards, clinical work sampling, blinded patient encounters, direct observation of procedural skills, case-based discussion, and multisource feedback. The research literature on formative assessment and feedback suggests that it is a powerful means for changing the behaviour of students and trainees.

To enhance the efficacy of the methods of workplace-based assessment, it is critical that the feedback which is provided be consistent with the needs of the learner, focus on important aspects of the performance (while avoiding personal issues), and have a series of characteristics which make it maximally effective. Since faculty play a key role in the successful implementation of formative assessment, strategies to provide training and encourage their participation are critical.

Notes on contributors

JOHN J. NORCINI, PhD has been President and CEO of the Foundation for Advancement of International Medical Education and Research (FAIMER®) since May 2002. For the 25 years before joining the Foundation, Dr. Norcini held a number of senior positions at the American Board of Internal Medicine. His principal academic interest is in the area of the assessment of physician performance.

VANESSA C. BURCH, MBChB, PhD is Associate Professor of Medicine at the University of Cape Town, South Africa. She convenes the undergraduate medical degree programme in the Faculty of Health Sciences and is also actively involved in postgraduate education in the Faculty. Her main academic interests are in the assessment of clinical competence and innovative methods of medical education in resource-constrained educational environments typical of developing countries.

References

- Arnold L, Willoughby L, Calkins V, Eberhart G. 1981. Use of peer evaluation in the assessment of medical students. *Med Educ* 56:35–41.
- Archer JC, Norcini JJ, Davies HA. 2005. Peer review of paediatricians in training using SPRAT. *Br Med J* 330:1251–1253.

- Association Of American Medical Colleges. 2004. Medical school graduation questionnaire: all schools report. Available at: URL: <http://www.aamc.org/data/gq/allschoolsreport/2004.pdf> (accessed on 11 April 2007).
- Balzer WK, Doherty ME, O'connor R Jr. 1989. Effects of cognitive feedback on performance. *Psychol Bull* 106:410-433.
- Bargh JA, Gollwitzer PM, Lee-Chai A, Barndollar K, Trötschel R. 2001. The automated will: Nonconscious activation and pursuit of behavioural goals. *J Personality Social Psychol* 81:1014-1027.
- Beck RS, Daughtridge R, Sloane PD. 2002. Physician-patient communication in the primary care office: a systematic review. *J Am Board Fam Pract* 15:25-38.
- Boulet JR, Mckinley DW, Norcini JJ, Whelan GP. 2002. Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Ad Health Sci Educ* 7:85-97.
- Branch WT, Paranjape A. 2002. Feedback and reflection: teaching methods for clinical settings. *Acad Med* 77:1185-1188.
- Burch VC, Seggie JL, Gary NE. 2006. Formative assessment promotes learning in undergraduate clinical clerkships. *S Af Med* 96:430-433.
- Butler DL, Winnie PH. 1995. Feedback and self-regulated learning: a theoretical synthesis. *Rev Educ Res* 65:245-274.
- Craven RG, Marsh HW, Debus RL. 1991. Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *J Educ Psychol* 83:17-27.
- Crooks TJ. 1988. The impact of classroom evaluation practices on students. *Rev Educ Res* 58:438-481.
- Daelmans HE, Overmeier RM, van der Hem-Stokroos HH. 2005. Reliability of the clinical teaching effectiveness instrument. *Med Educ* 39:904-910.
- Day SC, Grosso LG, Norcini JJ, Blank LL, Swanson DB, Horne MH. 1990. Residents' perceptions of evaluation procedures used by their training program. *J Gen Inter Med* 5:421-426.
- Daelmans HE, Hoogenboom RJ, Donker AJ, Scherpbier AJ, Stehouwer CD, Van Der Vleuten CP. 2004. Effectiveness of clinical rotations as a learning environment for achieving competences. *Med Teach* 26:305-312.
- Driessen E, Van Der Vleuten C. 2000. Matching student assessment to problem-based learning: lessons from experience in a law faculty. *Stud Cont Educ* 22:235-248.
- Durning SJ, Cation IJ, Markert RJ, Pangaro LN. 2002. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med* 77:900-904.
- Ende J. 1983. Feedback in medical education. *J Am Med Assoc* 250:777-781.
- Finlay K, Norman GR, Stolberg H, Weaver B, Keane DR. 2006. In-training evaluation using hand-held computerized clinical work sampling strategies in radiology residency. *J Can Ass Radiol* 57:232-237.
- Frederiksen N. 1984. The real test bias. Influences on testing and teaching and learning. *Am Psychol* 39:193-202.
- Gibbs G. 1999. Using assessment strategically to change the way students learn, in: S. Brown (Ed.) *Assessment Matters in Higher Education. Choosing and using Diverse Approaches*, (Buckingham, Society for Research into Higher Education and Open University Press).
- Gibbs G, Simpson C. 2004-2005. Conditions under which assessment supports student learning. *Learn Teach Higher Educ* 1:3-31.
- Gipps C. 1999. Socio-cultural aspect of assessment. *Rev Educ Res* 24:355-392.
- Gronlund NE. 1998. *Assessment of Student Achievement*, 6th edn (Needham Heights, MA, Allyn and Bacon).
- Hampton JR, Harrison MJG, Mitchell JRA, Prichard JS, Seymour C. 1975. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *Br Med J* 2:486-489.
- Harackiewicz JM. 1979. The effect of reward contingency and performance feedback on intrinsic motivation. *J Pers Soc Psychol* 37:1352-1363.
- Hatala R, Ainslie M, Kassen BO, Mackie I, Roberts JM. 2006. Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. *Med Educ* 40:950-956.
- Hatala R, Norman GR. 1999. In-training evaluation during an Internal Medicine clerkship. *Acad Med* 74:S118-S120.
- Hattie JA. 1999. *Influences on Student Learning*. Inaugural professorial address, University of Auckland, New Zealand. Available at: URL: <http://www.arts.auckland.ac.nz/staff/index.cfm?P=8650> (Accessed on 4 April 2007).
- Hattie J, Timperley H. 2007. The power of feedback. *Rev Educl Res* 77:81-112.
- Hauer KE. 2000. Enhancing feedback to students using the mini-CEX (clinical evaluation exercise). *Acad Med* 75:524.
- Herbers JE, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. 1992. How accurate are faculty evaluations of clinical competence? *J Gen Inter Med* 4:202-208.
- Hewson MG, Little ML. 1998. Giving feedback in medical education. Verification of recommended techniques. *J Gen Inter Med* 13:111-116.
- Holmboe ES, Yepes M, Williams F, Huot SJ. 2004a. Feedback and the mini-clinical evaluation exercise. *J Gen Inter Med* 19:558-561.
- Holmboe ES, Hawkins RE, Huot SJ. 2004b. Direct observation of competence training: a randomized controlled trial. *Ann Inter Med* 140:874-881.
- Holmboe ES, Fiebach NH, Galaty LA, Huot S. 2001. Effectiveness of a focused educational intervention on resident evaluations from faculty: a randomized controlled trial. *J Gen Intern Med* 16:427-434.
- Holmboe ES, Huot S, Chung J, Norcini JJ, Hawkins RE. 2003. Construct validity of the mini-Clinical Evaluation Exercise (MiniCEX). *Acad Med* 78:826-830.
- Isaacson JH, Posk LK, Litaker DG, Halperin AK. 1995. Residents' perceptions of the evaluation process. *J Gen Inter Med* 10(suppl.):89.
- Kalet A, Earp JA, Kowlowitz V. 1992. How well do faculty evaluate the interviewing skills of medical students? *J Gen Inter Med* 7:499-505.
- Kassebaum DG, Eaglen RH. 1999. Shortcoming in the evaluation of students' clinical skills and behaviours in medical school. *Acad Med* 74:841-849.
- Kirch W, Schaffi C. 1996. Misdiagnosis at a university hospital in 4 medical eras. *Medicine (Baltimore)* 75:29-40.
- Kluger AN, DeNisi A. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 119:254-284.
- Kogan JR, Bellin LM, Shea JA. 2002. Implementation of the mini-CEX to evaluate medical students' clinical skills. *Acad Med* 77:1156-1157.
- Kogan JR, Hauer KE. 2006. Brief report: use of the mini-clinical evaluation exercise in Internal Medicine core clerkships. *J Gen Inter Med* 21:501-502.
- Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, Ferrier K, Payne S. 2001. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. *Br Med J* 323:908-911.
- Locke EA, Latham GP. 1990. *A Theory of Goal Setting and Task Performance* (Englewood Cliffs, NJ, Prentice Hall).
- Luthans F, Peterson SJ. 2004. 360-degree feedback with systematic coaching: empirical analysis suggests a winning combination. *Hum Res Manag* 42:243-256.
- Maatsch JL, Huang R, Downing S, Barker B. 1983. Predictive validity of medical specialist examinations. *Final report for Grant HS 02038-04, National Center of Health Services Research*. Office of Medical Education Research and Development, Michigan State University, East Lansing, MI.
- Marton F, Dall'Alba G, Beaty E. 1993. Conceptions of learning. *Int J Educ Res* 19:277-300.
- McLeod PJ, Meagher TW. 2001. Educational benefits of blinding students to information acquired and management plans generated by other physicians. *Med Teach* 23:83-85.
- National Health Service. 2007. *Modernising Medical Careers: Foundation Programmes*. Available at: URL: <http://www.mmc.nhs.uk/pages/foundation> (Accessed on 7 April 2007).
- Noel GL, Herbers JE, Caplow MP, Cooper GS, Pangaro LN, Harvey J. 1992. How well do Internal Medicine faculty members evaluate the clinical skills of residents? *J Gen Inter Med* 11:757-765.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. 1995. The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Ann Inter Med* 123:795-799.
- Norcini JJ, Blank LL, Duffy FD, Fortna G. 2003. The mini-CEX: A method for assessing clinical skills. *Ann Inter Med* 138:476-481.

- Norcini JJ. 2007. Workplace-based assessment in clinical training, in: Swanwick T. (Ed.) *Understanding Medical Education series* (Edinburgh, UK: Association for the Study of Medical Education).
- Norman GR, Davis D, Painvin A, Lindsay E, Rath D, Ragbeer M. 1989. Comprehensive assessment of clinical competence of family/general physicians using multiple measures. *Proceedings of the Research in Medical Education Conference*, pp 75-79.
- Paukert JL, Richards ML, Olney C. 2002. An encounter card system for increasing feedback to students. *Am J Surg* 183:300-304.
- Peterson MC, Holbrook JH, Hales DV, Smith NL, Staker LV. 1992. Contributions of the history, physical examination and laboratory investigation in making medical diagnoses. *Wes J Med* 156:163-165.
- Ramsey P, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wenrich MD. 1989. Predictive validity of certification by the American Board of Internal Medicine. *Ann Intern Med* 110:719-726.
- Ramsey P, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. 1993. Use of peer ratings to evaluate physician performance. *J Am Med Ass* 269:1655-1660.
- Richards ML, Paukert JL, Downing SM, Bordage G. 2007. Reliability and usefulness of clinical encounter cards for a third-year surgical clerkship. *J Surg Res* 140:139-48.
- Sadler R. 1989. Formative assessment and the design of instructional systems. *Instruct Sci* 18:119-144.
- Shepard LA. 2000. The role of assessment in a learning culture. *Educ Res* 29:4-14.
- Small PA, Stevens B, Duerson MC. 1993. Issues in medical education: basic problems and potential solutions. *Acad Med* 68:S89-S98.
- Solomon DJ, Reinhart MA, Bridgham RG, Munger BS, Starnaman S. 1990. An assessment of an oral examination format for evaluating clinical competence in emergency medicine. *Acad Med* 65:S43-S44.
- Stillman PL, Haley H-L, Regan MB, Philbin MM. 1991. Positive effects of a clinical performance assessment programme. *Acad Med* 66:481-483.
- Swanson DB, Norman GR, Linn RL. 1995. Performance-based assessment: lessons from the health professions. *Educ Res* 24:5-11.
- Thomposn T. 1998. Metamemory accuracy: effects of feedback and the stability of individual differences. *Am J Psychol* 111:33-42.
- Turnbull J, MacFayden J, van Barneveld C, Norman G. 2000. Clinical works sampling. A new approach to the problem of in-training evaluation. *J Gen Inter Med* 15:556-561.
- van der Vleuten CPM. 1996. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1:41-67.
- van der Hem-Stokroos NH, Daelmans HE, van der Vleuten CP, Haarman HJ, Scherpbier AL. 2004. The impact of multi-faceted educational structuring on learning effectiveness in a surgical clerkship. *Med Educ* 38:879-886.
- Veloski J, Boex JR, Grasberger J, Evans A, Wolfson DB. 2006. Systematic review of the literature on assessment, feedback, and physicians' clinical performance: BEME Guide No 7. *Med Teach* 28:117-128.
- Walker J. 2001. The managerial mentor-leading productive learning in the workplace: an integral view. University of Technology Sydney Research Centre Vocational Education & Training. *Productive Learning Seminar Series*, November, p.3. Available at: URL: http://www.oval.uts.edu.au/working_papers/2002WP/0209walker.pdf (Accessed on 4 April 2007).
- Wenrich MD, Carline JD, Giles LM, Ramsey PG. 1993. Ratings of the performance of practicing internists by hospital-based registered nurses. *Acad Med* 68:680-687.
- Whitehouse A, Waltzman M, Wall D. 2002. Pilot study of 360° assessment of personal skills to inform record of in-training assessments for senior house officers. *Hosp Med* 63:172-175.
- Winnie PH, Butler DL., 1994. Student cognition in learning from teaching, in: T. Husen, & T. Postlewaite, (Eds.), *International Encyclopedia of Education*, pp. 5738-5745 (Oxford, UK: Pergamon).
- Wragg A, Wade W, Fuller G, Cowan G, Mills P. 2003. Assessing the performance of specialist registrars. *Clin Med* 3:131-134.

Copyright of *Medical Teacher* is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

ผศ. นพ. สุประพัฒน์ สนใจพานิชย์

หัวข้อ: Performance assessment

Performance Assessment

เชิดศักดิ์ ไอรณรัตน์
ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล
มหาวิทยาลัย มหิดล

**An ounce of performance is
Worth pounds of promises.**

Mae West

Competence and Performance

- Competence = The capacity of a person to perform a defined task (Maximal ability)
- Performance = The actual act in carrying out or execute the duty (Typical ability)

Clinical Performance Ratings

Ratings of clinical performance based on observing real-life clinical practice by attending faculty members

Objectives

- เมื่อสิ้นสุดการอบรมแล้ว อาจารย์ผู้เข้าอบรมสามารถ
 - ร่วมในการประเมินนักศึกษาแพทย์ด้วยแบบประเมิน clinical performance ratings ได้อย่างเหมาะสม
 - พัฒนาแบบประเมิน clinical performance ratings ที่มีคุณภาพดี ซึ่งนำไปสู่การประเมินที่ถูกต้อง และเที่ยงตรง

Rater Errors

- Construct-irrelevance variance in performance ratings that is associated with raters' behavior, not with the actual performance of ratees
- Valid use of clinical performance assessment requires monitoring and controlling of rater errors.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4, 386–422.

6

Reducing Rater Errors

- Improving raters
- Improving a rating instrument

Improving Raters

1. Rater training
2. Rater monitoring
3. Rater feedback

Writing Effective Items

- Remember your purpose
- Keep it simple
- Focused: include only one topic per item
- Start with easy-to-respond items
- Group items into sections, position these sections in a logical order

Characteristics of A Good Scale

1. Well-defined category
2. Appropriate number of categories
3. Proper handling of middle category
4. Ordered
5. Research-based

Key Points: Performance Ratings

- Remember what to observe
- Rate when you still remember the students
- Multiple ratings: multiple raters, time points
- Rate when you are in a stable emotional state
- Be consistent in your rating standards (within and across groups)
- Rate each item independently: avoid halo effect
- Use the full range of scores: avoid restriction of range

If my future were determined just by my performance on a standardized test, I would not be here. I guarantee you that.

Michelle Obama

รศ. นพ. ตรีกภ เลิศบรรณพงษ์

หัวข้อ: Portfolio



Portfolios for Assessment and Learning

Jan van Tartwijk
Erik W Driessen

AMEE GUIDE
Assessment

45






AMEE Guides in Medical Education

www.amee.org

Welcome to AMEE Guides Series 2

The AMEE Guides cover important topics in medical and healthcare professions education and provide information, practical advice and support. We hope that they will also stimulate your thinking and reflection on the topic. The Guides have been logically structured for ease of reading and contain useful take-home messages. Text boxes highlight key points and examples in practice. Each page in the guide provides a column for your own personal annotations, stimulated either by the text itself or the quotations. Sources of further information on the topic are provided in the reference list and bibliography.

Guides are divided into series according to subject:

-  **Teaching and Learning**
-  **Research Methods**
-  **Education Management**
-  **Curriculum Planning**
-  **Assessment**

The Guides are designed for use by individual teachers to inform their practice and can be used to support staff development programmes.

'Living Guides'

An important feature of this new Guide series is the concept of supplements, which will provide a continuing source of information on the topic. Published supplements will be available to all who have purchased the Guide.

If you would like to contribute a supplement based on your own experience, please contact the Guides Series Editor, Professor Trevor Gibbs (tjg.gibbs@gmail.com).

Supplements may comprise either a 'Viewpoint', when you communicate your views and comments on the Guide or the topic more generally, or a 'Practical Application', where you report on implementation of some aspect of the subject of the Guide in your own situation. Submissions for consideration for inclusion as a Guide supplement should be maximum 1,000 words.

Other Guides in the new series

A list of topics in this exciting new series is listed on the back inside cover.

Institution/Corresponding address:

Dr Jan van Tartwijk, ICLON – Leiden University Graduate School of Teaching, Leiden University,
PO Box 905, 2300 AX Leiden, The Netherlands

Tel: +31 71 527 3845

Fax: +31 71 527 5342

Email: jtartwijk@iclon.leidenuniv.nl

The authors:

Dr Jan van Tartwijk works at the ICLON – Leiden University Graduate School of Teaching. In his research and teaching he focuses on teacher-student communication processes in the classroom and the use of portfolios in medical education and teacher education.

Dr Erik Driessen works at the Department of Educational Development and Research at Faculty of Medicine of the University of Maastricht. He specializes in assessment and the use of portfolios in medical education.

Both have a long history with working with portfolios. Jan van Tartwijk started experimenting with portfolios in teacher education and faculty development in 1994. In 1999, he joined Erik Driessen and Cees van der Vleuten at Maastricht University, where they implemented portfolios in the undergraduate program of the Faculty of Medicine of the University of Maastricht. Since then, they have published a series of articles and books about using portfolios in higher education and have advised numerous faculties and originations in medical education and elsewhere about the use of portfolio for learning and assessment. Their corporation is not limited to the topic of portfolios; they also work together on research on how to stimulate and assess self-critical thinking and reflection.

Part of this AMEE Guide was first published in *Medical Teacher*:

Van Tartwijk J & Driessen EW (2009). Portfolios for assessment and learning. AMEE Guide No.45.

Medical Teacher, 31 (9): 790-801.

Guide Series Editor: Trevor Gibbs (tjg.gibbs@gmail.com)

Published by: Association for Medical Education in Europe (AMEE), Dundee, UK

Designed by: Lynn Thomson

© AMEE 2010

ISBN: 978-1-903934-57-9

Contents

Abstract	1
Introduction	2
Portfolio goals, content, and organization	4
Portfolios as a multipurpose instrument	4
Electronic portfolios	7
Portfolios and learning from experience	9
Theoretical background	9
Reflection and professional development	10
Using portfolios as tools for assessment	14
Factors influencing the success of the introduction of a portfolio	21
People	21
Academic leadership	23
Infrastructure	23
Concluding remarks	24
References	25

Abstract

In 1990, Miller wrote that no tools were available for assessment of what a learner does when functioning independently at the clinical workplace (Miller 1990). Since then portfolios have filled this gap and found their way into medical education, not only as tools for assessment of performance in the workplace, but also as tools to stimulate learning from experience.

We give an overview of the content and structure of various types of portfolios, describe the potential of electronic portfolios, present techniques and strategies for using portfolios as tools for stimulating learning and for assessment, and discuss factors that influence the success of the introduction. We conclude that portfolios have a lot of potential but that their introduction also often leads to disappointment, because they require a new perspective on education from mentors and learners and a significant investment of time and energy.

TAKE HOME MESSAGES

- The goals of working with a portfolio need to be clear.
- It is not problematic to use portfolios concurrently to formatively promote learning as well as for summative assessment. Summative assessment is important to ensure that portfolio learning maintains its status alongside other assessed subjects.
- The effectiveness of learning is enhanced when a mentor supports the portfolio process. Mentorship requires a substantial time investment but is crucial for the successful use of portfolios. The effectiveness of assessment can be enhanced by combining the portfolio with an interview.
- Use a flexible learner-centred portfolio format. A rigid structure in which every detail of portfolio content is prescribed will elicit negative reactions from portfolio users.
- Too much structure is a greater risk than too little structure, but learners do need clear directions and guidance to support the development and assessment of broad competencies.
- Working with a portfolio is time consuming both for learners and mentors. This is more of a problem in postgraduate training and continuous medical education than in undergraduate education.

Introduction

Today's doctors find themselves confronted not only with patients who are increasingly knowledgeable and assertive, but also with pressure to apply new findings and evidence in day-to-day practice, and with the necessity to collaborate with other health professionals in ever larger teams and communities. To deal with these complexities, doctors need generic competencies to enhance effective communication, organization, teamwork and professionalism. These generic competencies are sometimes labelled as doctors' "soft skills" in contrast to "hard clinical skills". In recent years, learning, teaching and assessment of these generic competencies has gained unexpected urgency among politicians and the general public. Headlines decrying incidents involving dysfunctional doctors and hospital departments with dramatic impact on morbidity and mortality figures catapulted generic competencies to the forefront of attention as indispensable qualities for doctors. As a result, professional associations and governments began to voice increasingly urgent demands to include these generic competencies in education and assessment (General Medical Council, 2000). At the same time, consistent with the general trend towards outcome-based education, the focus in medical education shifted from the educational process itself towards the competencies of doctors at the end of training and at important junctures during the training process (Norcini et al., 2008). The competencies described by professional organizations such as the Royal College of Physicians and Surgeons of Canada (1996) became the framework for assessment and, as a consequence, for the content and organization of programmes for medical education in many countries.

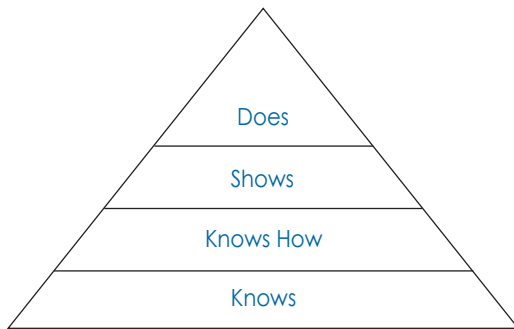
However, stimulating the development of competencies (Box 1) and the assessment of its result is complicated. Already in 1990, Miller described the challenges involved in assessing clinical competence. He presented a framework for clinical assessment, shaped like a pyramid (Figure 1), whose layers from bottom to top represent increasingly complex levels of mastery, with the lower levels providing the foundation for the higher levels (Miller, 1990).

BOX 1 Competence

The concept of competence is much used and much debated (Stoof et al., 2002; Dreyfus, 2004). Here, we define it as an integrated body of knowledge, skills, and (professional) attitudes enabling proficient performance in certain real life settings, i.e. the "Does" level in Miller's framework.

...doctors need generic competencies to enhance effective communication, organization, teamwork and professionalism.

FIGURE 1
Framework for clinical assessment: Miller's Pyramid (cf. Miller, 1990)



The bottom level is concerned with *knowledge*. This is the knowledge relating to the skills that learners must master for their future professional practice. This knowledge is best assessed by written tests. The next level represents application of the knowledge from level 1. Learners should know *how* to apply their knowledge when performing skills. For instance, at this level, learners are expected to know how to diagnose a patient and which aspects of a patient's presentation to attend to. The *knows how* level can also be assessed by written tests. One level up, at level 3, the issue of interest is that learners demonstrate their ability to use their knowledge to *take appropriate action in a simulated environment*. This level combines knowledge and action (cognition and behaviour). Not only should learners know how to diagnose a patient, they should also be able to actually perform the appropriate actions, for example a physical examination in a simulated patient (*shows how*). The top of the pyramid is concerned with *independent performance within the complex environment of day-to-day practice*. This requires integration of knowledge, skills, attitudes, and personal characteristics. Performance at the top of the pyramid is manifested when learners are working independently in professional practice. Typically, adequate performance at this level requires integrated performance of different roles; not only the role of medical expert but also that of counsellor, participant in the doctor- patient relationship, a leadership role in relation to nursing staff, etc. Good performance at the Does level (of Miller's Pyramid) implies competence.

In 1990, Miller observed that there were no instruments to evaluate performance consistent with the top of the pyramid (Miller, 1990). At the same time, scholars in the field of teacher education and teacher assessment were struggling with the same problem (Bird, 1990). Here too, the key challenge was how to assess performance in real life settings. Shulman (1998) describes the Teacher Assessment Project that was set up with the purpose of exploring and developing new approaches to the evaluation of teaching in primary and secondary education. He recounts that it was considered undesirable to assess teacher competence solely on the basis of ratings in assessment centres, because experiments showed that the information provided by assessment centres alone was not enough to identify competent and excellent teachers. Information about whether teachers succeeded in making the most of their pupils' learning opportunities *within* their own complex working environment was needed as well. It was also

Good performance at the Does level (of Miller's Pyramid) implies competence.

recognised that there can be striking variations among teaching settings. For instance, it makes quite a difference whether one teaches at an urban school in a deprived area with its myriad of social problems or at a high school in a middle class suburban environment. As part of efforts to achieve fair judgement of teacher performance in a broad array of settings and situations, the *portfolio* concept was borrowed from the arts and architecture (Box 2).

BOX 2 Portfolio

Portfolios that are used in education contain evidence of how learners fulfil tasks and their competence is progressing. They may be digital or paper based and content may be prescribed or left to the learners' discretion. Despite variations in content and format, portfolios basically report on work done, feedback received, progress made, and plans for improving competence (Driessen et al., 2007b).

Since portfolios were introduced in medical education in the early 1990s (Royal College of General Practitioners, 1993), their use as an instrument for both assessment and encouraging professional growth has increased enormously (Snadden et al., 1999; Friedman Ben David et al., 2001). However, the evidence to date suggests that the introduction of portfolios for these purposes has met with mixed success (Driessen et al., 2007b; Tochel, et al., 2009; Buckley et al., 2009). Although potentially powerful instruments in education, the use of portfolios has proved to be vulnerable.

The aim of this AMEE Guide is to help medical teachers and educators to make full use of the possibilities that portfolios offer and prevent difficulties occurring. Based on an analysis of what portfolios help achieve, it is our purpose to provide practical clues about the design, implementation and use of portfolios in medical education.

Firstly, we will describe how portfolio content and structure relate to the various goals that they are designed to achieve. Next, we will focus on the use of portfolios as instruments that can encourage professional growth by stimulating learning from experience and subsequently, we will elaborate on the use of portfolios as instruments for assessment. Each of these goals requires specific content and organization of portfolios. Finally, we will focus on the factors that are important for the successful introduction of portfolios in (medical) education.

Portfolio goals, content, and organization

Portfolios as a multipurpose instrument

- **Portfolios for assessment:** When portfolios were originally introduced in education as instruments for authentic assessment, they closely resembled the portfolios of architects and artists that Lyons (1998) describes as a portable case for keeping, usually without folding, loose sheets of papers, drawings or photographs. Building on the principle of triangulation (Denzin, 1978; Denzin & Lincoln, 2000) all kinds of evidence can be brought

together in those portfolios that, in combination, give the possibility to draw valid conclusions about competence (Box 3).

BOX 3

Combining evidence to improve the quality of conclusions

In the literature, combining data from various sources with the aim to improve the quality of conclusions is often referred to as triangulation. The aim of triangulation is to avoid biases and problems, such as those related to the reliability and trustworthiness of data that are derived from one single source.

Procedures for multisource feedback or 360-degree feedback use a similar strategy by stimulating learners to gather feedback from different sources. Lockyer & Clyman (2008) describe a procedure involving a questionnaire survey among medical colleagues, nurses, and patients and their families to collect data about learners' specific competencies. The same questionnaire is completed by the learners themselves. By aggregating these data, reliability is improved.

However, in one of the first explorations of portfolios for teacher assessment, Bird (1990) wrote that the portfolio procedures for assessment might easily degenerate into exercises in amassing paper. He suggested that the evidence in a portfolio should be organised according to the competencies that the person compiling the portfolio wants to show. Both for the learner compiling the portfolio and for an assessor this would be helpful. Instructions starting with "Show how you..." might clarify for portfolio owners that they are asked to provide specific evidence about their performance. A portfolio organised by tasks or competencies might be helpful for assessors, because it indicates what the material in the portfolio is supposed to show. Based on initial experiments with portfolios, Collins (1991) suggested that captions should be attached to the evidence in the portfolio:

One essential component of the portfolio was the document caption. The caption is a little sheet attached to each document stating what the document is (...) and why it is valuable evidence. (...) Captions proved to be essential to the portfolio development process. Documents without captions were meaningless to the raters. (p. 153)

- **Portfolios for learning:** Soon after the introduction of portfolios in medical education, Snadden & Thomas introduced the term "portfolio learning" (Snadden & Thomas, 1998b):

Portfolio learning is a method of encouraging adult and reflective learning for professionals. Derived from the graphic arts it is based on developing a collection of evidence that learning has taken place (p. 192)

They emphasise the importance the importance of supervision and critical reflection for portfolio learning:

The system works well when it operates through the interaction of a learner and mentor using the material as a catalyst to guide further learning. It is essential that the portfolio does not become a mere collection of events seen or experienced, but contains critical reflections on these and the learning that has been made from them (p.192).

...portfolio procedures for assessment might easily degenerate into exercises in amassing paper.

Portfolio learning is a method of encouraging adult and reflective learning for professionals. Derived from the graphic arts it is based on developing a collection of evidence that learning has taken place.

A portfolio can also stimulate reflection, because collecting and selecting work samples, evaluations and other types of materials that are illustrative of the work done, compels learners to look back on what they have done and analyse what they have and have not yet accomplished.

In many cases, portfolios are assembled over a longer period of time. That is why they can also be used to support planning and monitoring in professional development. One way to do so is to include learning objectives in the portfolio as well as a document trail of related learning activities and accomplishments (Mathers et al. 1999; Oermann, 2002).

As a consequence, reflections and overviews of personal development have secured a prominent place in many portfolios. Portfolios that are primarily geared to assessment will remain organised around all kinds of materials that provide 'evidence' of competencies. In portfolios that are primarily used to monitor and plan learners' development, overviews will take centre stage. Portfolios whose primary objective is to foster learning by stimulating learners to reflect on and discuss their development will be organised around learners' reflections.

- **A multipurpose instrument¹:** Inevitably, these developments have widened the applicability of the label *portfolio* to a broad range of instruments. Some portfolios might equally and aptly be labelled *Personal Development Plan or Reflective Essay*. Because of the tremendous variety in portfolios, careful and critical appraisal of the strengths and weaknesses of different portfolios is advisable before deciding which one to implement in a particular setting.

The question to be answered is whether a certain portfolio is fit for its intended purpose. And just as someone else's shoes are unlikely to fit comfortably, portfolios tailored to one particular educational setting may not fit into the educational configuration(s) of other settings (Spandel, 1997). An ill-fitting portfolio will inevitably be discarded sooner or later. To assist in determining whether a portfolio is appropriate for its intended purpose the triangle in Figure 2 helps to define the nature of a portfolio. It does so by inviting positioning of a portfolio in the area of the triangle where it is most likely to achieve its intended principal objectives.

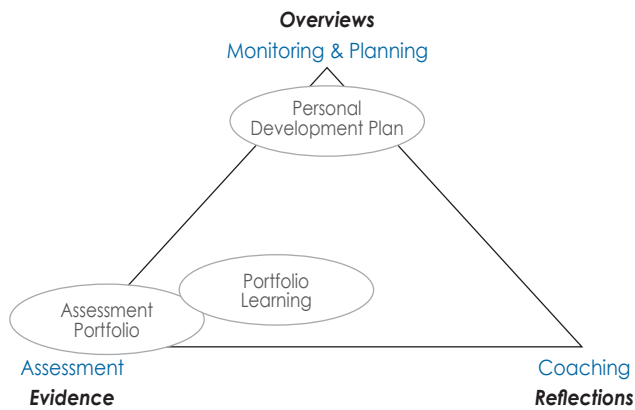
Obviously, a portfolio can be used to achieve more than one goal. When a portfolio is to serve a combination of goals, its position in the triangle will shift towards the centre because its strengths have to be distributed more evenly over evidence, overviews and reflections. In practice, the majority of portfolios are not situated in one of the corners of the triangle (Buckley et al., in press). A controversial issue in the literature on educational portfolios is whether it is acceptable to have one portfolio for both assessment and reflection (Snyder et al. 1998). An argument against this dual function is that assessment may jeopardise the quality of reflection thereby detracting from the portfolio's effectiveness for mentoring purposes. Learners may be reluctant to expose their less successful efforts at specific tasks and to reflect on strategies for addressing weaknesses if

A portfolio can also stimulate reflection...

¹ Parts of this section were published in the journal *Quality in Higher Education* (van Tartwijk, et al., 2007)

they believe they are at risk of having 'failures' turned against them in an assessment situation. Portfolios that are not assessed, on the other hand, do not "reward" learners for the time and energy they invest in them. As a result, learners are likely to take the portfolio and any associated learning activities less seriously. A recent BEME review showed that most portfolios were also assessed for summative purposes (Buckley et al., 2009).

FIGURE 2
Purposes and content of portfolios (van Tartwijk, et al., 2007)



An effective portfolio has a clear but flexible structure, giving individual learners opportunities to describe their own unique development (Pearson & Heywood, 2004; Driessen et al. 2005b; Grant et al. 2007). Clear instructions are important, but when the content of a portfolio is prescribed in detail, portfolios are often experienced as highly bureaucratic instruments (Davis et al., 2001; O'Sullivan et al. 2004; Pearson & Heywood, 2004; Kjaer et al. 2006). Portfolios meet with stronger appreciation when learners have a certain amount of freedom to determine the content of their own portfolios (Snadden & Thomas, 1998a; Driessen et al., 2005b).

An effective portfolio has a clear but flexible structure, giving individual learners opportunities to describe their own unique development.

Electronic portfolios

A growing number of medical schools use electronic portfolios (e-portfolios) instead of paper-based portfolios (Fung Kee Fung et al., 2000; Lawson et al., 2004; Woodward & Nanlohy, 2004; van Tartwijk et al., 2007; Driessen et al. 2007a). This preference is based on a number of considerations:

- In e-portfolios, hyperlinks can be inserted to make connections between evidence, overviews, and reflections. This can be useful, for instance, when learners want to illustrate reflections with evidence that is stored somewhere else in the portfolio, or want to illustrate a schematic overview of their development by making hyperlinks to materials and reflections. Hyperlinks can also be useful to make a table of contents of the portfolio. For instance by including a list of captions in the portfolio and making hyperlinks to related materials. Mentors or assessors can browse through this list of captions, obtain a quick overview of all the evidence in the portfolio, and just click on the evidence that is relevant to their specific purpose.

- A paper-based portfolio can be cumbersome because of its bulk. Imagine an assessor who needs to take 15 paper portfolios home! Furthermore, there is generally only one copy of a paper portfolio. Whenever learners hand their paper portfolios to their mentor or assessor, the portfolio is literally out of their hands. Not only do they run the risk of the portfolio getting lost, it is also more difficult for them to prepare to discuss the portfolio with their mentor or assessor. Another advantage of e-portfolios is that they are easier to keep up to date.

Of course there are disadvantages as well:

- Mentors who do not like to read a portfolio on screen will still have to print it. In most systems it is not possible to make notes on the portfolio itself (although making notes on the learner's paper portfolio might not be desirable as well).
- E-portfolios can only be used by learners and teachers who are sufficiently skilled in using the relevant software and hardware.
- An e-portfolio requires a stable and high quality information technology infrastructure that is not always available.

Nowadays, many dedicated portfolio systems are available, which are usually user-friendly (Dornan et al., 2002; www.eportfolioservice.nl). These systems can provide specific functionalities for specific portfolio goals: options to include work-based assessment instruments, such as multisource feedback or mini clinical evaluation exercises (mini-CEX) in portfolios for clinical training: to invite specific individuals to inspect the portfolio, either wholly or in part, while denying access to everyone else.

Apart from dedicated systems, learners can produce an e-portfolio using standard word-processors or HTML editors, preferably ones that they and their teachers are familiar with (Gibson & Barrett, 2003). The cost of dedicated portfolio software is not the only reason to support this choice: for many purposes the hyperlink functionality of generic software is all that learners need. Furthermore, generic software allows a learner to impart his or her own flavour to the portfolio. This can enhance the learners' motivation to work with the instrument. Another reason is that many portfolio systems are limited because they are built to accommodate no more than one or two portfolio types. Finally, portfolios built with dedicated software need to be accessible with generic software for later maintenance and presentation. This may well be the case after a learner has left the setting in which the portfolio was produced, or in the event that the vendor in question ceases to do business. In summary, standard software tools have disadvantages from the perspective of managing access to the portfolio using the internet or to include work-based assessment instruments, but they usually provide all the options learners need to produce a portfolio that works well and looks great.

In a study comparing web-based and paper-based portfolios (Driessen et al., 2007a), not only did learners add more personal touches to content and form and invested more time in their portfolios, but mentors were unanimous in their appreciation of the greater ease of use of web-based portfolios compared to the more familiar paper-based ones. Information was

...standard software tools have disadvantages from the perspective of managing access to the portfolio using the internet or to include work-based assessment instruments, but they usually provide all the options learners need to produce a portfolio that works well and looks great.

easy to locate without having to turn pages to find certain content and the portfolios could be accessed from different locations were two reasons cited for preferring web-based portfolios. Other authors have also reported on the user friendliness of electronic portfolios (Fung Kee Fung et al., 2000; Lawson et al., 2004). In these studies, tutors appreciated the easy electronic access and reduction in the amount of paper used. However, the same authors also reported certain situations that make web-based portfolios less user-friendly than paper-based portfolios. For instance, limited computer access in the clinical workplace cancels out the advantages of user-friendliness and may even have an opposite effect.

Portfolios and learning from experience

Research shows that the role of the mentor is crucial to the successful use of portfolios aimed at learning from experience (Finlay et al. 1998; Snadden & Thomas, 1998a Mathers et al., 1999; Pearson & Heywood, 2004; Driessen et al., 2005b; Grant et al., 2007). In this section, we focus on the strategies mentors can use to promote learning from experience with a portfolio.

Theoretical background

The contemporary view of learning, based on constructivism, is that people "construct" new knowledge and understanding based on what they already know and believe (Bransford et al. 2000). What people know and believe can be represented as cognitive structures that guide their perception of reality. Evidently, a perception of reality based on individual cognitive structures does not afford an objective view of reality, but, by definition, an individual, idiosyncratic view. It is this personal perception of reality that guides a person's actions.

Reflection is an important concept in this framework, which relates to changing cognitive structures. Research has shown that meta-cognitive skills, such as reflection, increase the degree to which learners transfer what they have learned to new settings and events (Bransford et al., 2000). Despite considerable confusion about the precise definition of the term reflection (Hatton & Smith, 1995; Mann et al. 2007) all authors writing about reflection share the constructivist view that human behaviour is guided by mental structures that are not static but flexible, evolving, and changing in response to experiences. Based on this consensus view, reflection can be defined as the mental process of organising or reorganising cognitive structures that represent existing knowledge and beliefs and guide perceptions of experiences, situations, and problems (Korthagen et al. 2001). To put it in simpler terms: reflection means exploring and elaborating one's *understanding* of an experience (Eva & Regehr, 2008). Building on Van Manen's work (1977), Hatton & Smith (1995) distinguish three types or levels of reflection. The first type is concerned with the *means* to achieve certain ends. The second type is not only about means, but also about *goals*, the *assumptions* upon which they are based, and the actual *outcomes*. The third type of reflection is referred to as *critical reflection*. Here, moral and ethical criteria are also taken into consideration. Judgements are made about whether professional activity is equitable, just, and respectful to persons or not. Hatton and Smith emphasise that these three types of reflection should

Research shows that the role of the mentor is crucial to the successful use of portfolios aimed at learning from experience.

...meta-cognitive skills, such as reflection, increase the degree to which learners transfer what they have learned to new settings and events.

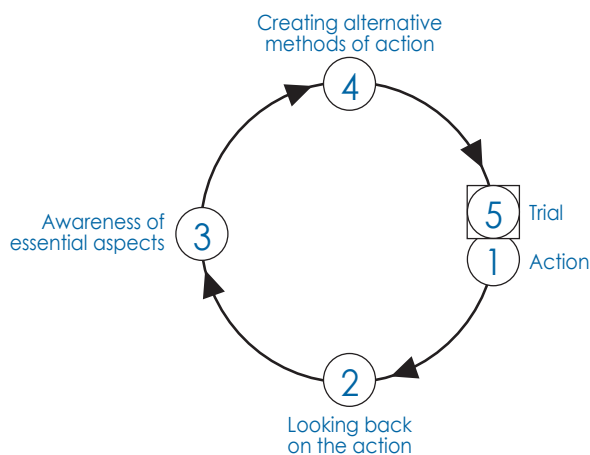
not be viewed as hierarchical. Different (educational) contexts and situations may lend themselves more to one kind of reflection than to another.

Reflection and professional development

For medical teachers who want to help learners learn from practice, the key question to answer is: "How can I stimulate my learners to reflect on their experiences and learn from them?" For this AMEE Guide the additional question is: "... and how can a portfolio help to improve the quality of reflection?".

Korthagen designed the **ALACT** model (**A**ction, **L**ooking back, **A**wareness, **C**reating alternative methods, **T**rial) (Figure 3) to describe the spiralling process that effective learners go through when faced with a situation for which no routine solution is available (Korthagen et al., 2001). This model resembles the three step model described by Snadden & Thomas (1998b) which focused on evaluation, reflection, and formulating a learning plan. We will describe the ALACT model, explain the potential contribution of working with a portfolio in each of the stages, and give suggestions for coaching strategies (Driessen et al., 2008).

FIGURE 3
ALACT model showing the phases of spiral professional development (Korthagen et al., 2001)



ALACT

A Action: The cycle starts with action undertaken for a specific purpose (e.g. for developing a specific competence). Learners can be helped to improve their existing routines and concurrently acquire new ones by pre-selecting experiences from which they can learn, for example a mixture of patients who are more or less easy to diagnose. Ericsson's research predicts that expertise will grow not just from the weight of experience but also from engaging in activities specifically designed or selected to improve performance (Ericsson, 2006).

Learners can be helped to improve their existing routines and concurrently acquire new ones by pre-selecting experiences from which they can learn.

L Looking back on action: self-directed assessment seeking: The ALACT cycle then moves to the stage where learners look back on a previous action, usually when that action was not successful or something unexpected happened. This looking back on action is assumed to be accompanied by an evaluation of whether the goals were realised and the learner's part in this. In many cases this can be regarded as a form of *self assessment*. Eva & Regehr (2008) write that most of the time self-assessment is conceptualised according to a "guess your grade" model of which the quality is generally poor (Davis et al., 2006). As an alternative they propose *self-directed assessment seeking*, which they describe as a process by which a learner takes personal responsibility for looking outward, explicitly seeking feedback and information from external sources of assessment data, to direct performance improvements that can help them to validate their self-assessment.

The role of the portfolio: Seeking and selecting evidence (documents, feedback, work-based assessments, etc.) for inclusion in a portfolio can be regarded as self-directed assessment seeking. To improve the quality of this process, it is important to use a variety of evidence from various sources. The validity of the results of self-directed assessment seeking will be maximised if the learner's self-reflections are consistent with all the information that is brought together in a portfolio.

Teaching strategies: Research has shown that a mentor can play a decisive role in determining whether the use of portfolios in education is successful or not (e.g. Driessen et al., 2007b). At the very least, learners may expect their mentors to pay serious attention to their portfolios, for after all they did spend a lot of time and energy to put their portfolio together. But even more importantly, careful scrutiny of their own performance may be confronting for learners. Effective mentors have an important role in this respect. In Box 4, we give suggestions for a number of strategies to be used by medical teachers in this phase, derived from the work by Korthagen and colleagues (Korthagen et al. 2002).

A Awareness of essential aspects: reflection: After conclusions have been drawn about the quality of performance and the characteristics of the situation, the next step in the ALACT model is to foster awareness of essential aspects. In this phase, learners try to develop a new and better understanding of what has happened, i.e. they reflect on their performance.

They can focus on the *means* they used to achieve a goal and try to understand why their strategy was successful or not. They can also consider whether they had selected a suitable goal for this particular situation. And finally they may consider what they want to achieve from a *moral or ethical* perspective.

Seeking and selecting evidence (documents, feedback, work-based assessments, etc.) for inclusion in a portfolio can be regarded as self-directed assessment seeking.

BOX 4**Strategies to stimulate self-directed assessment seeking**

- Provide a safe environment by distinguishing between learners as individuals and their performance.
- Focus on description.
- Stimulate learners to be concrete in their reports. When learners give general evaluations about a situation and their performance, ask questions:
 - What went well?
 - What went wrong?
 - How did you solve this?
 - What effect did this have?
- Stimulate learners to carefully scrutinise all the information in their portfolio. Learners could be asked to go through all the available evidence and answer questions:
 - Which information in your portfolio supports your answers/evaluation?
 - Which information in your portfolio contradicts your answers/evaluation?
- Stimulate learners to take the perspective of other stakeholders. Ask questions:
 - What did you want? What do you think the patient/your colleague/the nurse wanted?
 - What did you think? What did the others think?
 - What did you do? What did the others do?
 - What emotions did you experience? What emotions did the other people involved experience?

The role of the portfolio: Language is important in supporting thinking. Writing things down can help to stimulate reflection (Korthagen et al., 2001). Written reflections were not a part of the original portfolios, like the ones in which artists presented a selection from their work, but almost immediately after the introduction of portfolios in education, written reflections became a fixture of portfolios (Paulson et al. 1991). Embedding a written reflection in a portfolio has the advantage that it can be built on the self-assessment that was validated by the evidence in the portfolio. This is a form of facilitated reflection (Conlon, 2003). The learner can also use the evidence to illustrate a reflection with a concrete example.

Teaching strategies: To stimulate learners to reflect and learn from their experiences, mentors do not need to have all the right answers. The most important thing for them is to ask the right questions. In Box 5 we give a number of examples of questions that mentors can ask.

Language is important in supporting thinking. Writing things down can help to stimulate reflection.

To stimulate learners to reflect and learn from their experiences, mentors do not need to have all the right answers. The most important thing for them is to ask the right questions.

BOX 5**Questions to stimulate reflection****Means**

- Which strategies did you consider? Why did you select this strategy? Which are the advantages and disadvantages of the strategy you used?
- Which part of your strategy was effective and which part was not effective? Why was it effective / not effective?
- Would this strategy have been more /less effective in a different situation?

Goals, assumptions, outcomes

- What did you want to achieve? Were you successful? What do you consider successful?
- Why is this particular goal important?/Why did you pursue this goal?

Critical reflection

- Do you think patients / patients' families / medical colleagues / nurses / administrators are satisfied with these outcomes? What are their primary interests?

Confront with discrepancies

- I read in your portfolio that you are happy with the result, but when we talk about it, your face tells a different story.
- You write here that this is what you want to achieve, but you are pleased with your results even though they do not match your goals.
- You do not actually do what you say you want to do.

Generalize across experiences

- Which differences and similarities do you recognise between what is happening now and what happened in situations that you described in your portfolio?
- When do these things happen?
- Do you recognise a pattern?

C Creating or identifying alternative methods of action: change: Analysing previous actions may trigger a search for alternative strategies, or abandonment of original goals. It is important to explicate (new) goals and alternative strategies. A recent review showed that goal setting stimulates learning and that a mentor has an important role to play in this respect (Shute, 2008). Learners who work with a mentor set more specific goals and improve more than those who do not work with a mentor (Smither et al. 2003). Very often, agreement about what should be done differently and which goals should be achieved are written down in a document that is referred to as a Personal Development Plan (PDP).

The role of the portfolio: In many portfolios, the central goal is to keep track of the learner's development. In these portfolios, PDPs can have an important place. Snadden & Thomas for instance, (Snadden & Thomas 1998b) propose that when a portfolio is used for professional development and to track progress, it is important to attach to the portfolio some kind of learning plan.

Teaching Strategies: Both mentors and learners should commit to the agreements in the PDP and it should be on the agenda of their next progress meeting. The plans in the PDP are often too vague. It is important that mentors stimulate learners to be very concrete. It can be helpful to keep in mind that the learning goals in the plan should be formulated in a SMART way (Box 6).

Learners who work with a mentor set more specific goals and improve more than those who do not work with a mentor.

**BOX 6
SMART**

Specific	(Straightforward, not ambiguous)
Measurable	(It is clear under which conditions the goals are achieved)
Acceptable	(The goals should be acceptable to all stakeholders)
Realistic	(The learner should be able to achieve the goals)
Time-bound	(It should be clear when the goal is to be achieved)

T Trial: The last step in the ALACT cycle is trial. This is also the start of a new cycle in the spiral of professional development in this model.

Using portfolios as tools for assessment

In the introduction, we quoted Shulman (1998), who wrote that the reason for introducing portfolios in education as tools for assessment is that in a portfolio information can be brought together about how a person performs and how his or her competencies develop in his or her own complex working environment. From the perspective of assessment, the strength of the portfolio is also its weakness. The evidence held by a portfolio is often not standardised and its meaning often depends on the context from which it originates.

Assessing non-standardised portfolios requires a different perspective on assessment than the traditional quantitative perspective that is best suited for analysing quantitative test scores or results from standardised observations. Authors like Snadden (1999) and Webb (2003) all come to the conclusion that we should not try to fit non-standardised portfolios to standardised psychometric assessment criteria. They point out that portfolio assessment is primarily concerned with interpreting various forms of qualitative information and suggest that assessment procedures should be developed that are based on methods used in qualitative research.

In the next section, we will translate the insights of this literature into recommendations for portfolio assessment. We will structure this section according to five questions that, according to Harden (1979), should always be asked and answered by medical teachers in relation to assessment:

- What is assessed?
- Why is this assessed?
- How is this assessed?
- Who assesses?
- When is this assessed?

What? Although portfolios are also used in undergraduate medical education to assess reflective ability or communication skills (Driessen et al. 2003), portfolios are particularly suited to work-based assessment. In other words, they have added value at the does level of Miller's pyramid (Miller 1990).

The evidence held by a portfolio is often not standardised and its meaning often depends on the context from which it originates.

Many medical curricula are based on competency criteria developed by organisations such as the General Medical Council (GMC), the American Council of Graduate Medical Education (ACGME), and the Royal College of Physicians and Surgeons of Canada (RCPSC). More often than not, additional detail is required to fit the competency criteria to assessment procedures. In aligning competency descriptions with assessment procedures it is of the essence to strike the right balance between very concrete but also very detailed and long lists of "is able to" statements, on the one hand, and very global descriptions providing an overview but too little to support assessment, on the other hand. The extremes of this continuum have been referred to as an analytical versus a global approach. Both approaches have their pros and cons (Box 7).

BOX 7**Analytical versus global assessment**

In an analytical assessment, various aspects of a competency are assessed separately. A formula is used to combine the partial assessments into one final score.

Because the criteria are explicitly defined and each partial competence is explicitly assessed, the result is very transparent and usually more reliable and more informative for the learner. Criteria are usually defined in terms of: "The candidate is able to...".

Problems that may occur are:

- Learners may adapt their learning activities to 'ticking' specified criteria. This may result in unnatural activities in the workplace where competencies are acquired.
- Analytical assessment is very labour intensive. It may be experienced as bureaucratic.
- It can be difficult for assessors to give a truly distinct assessment of each individual criterion ('halo effect').
- Assessors have limited freedom to take account of specific competencies or extremely good (or poor) performance: if it is not in the criteria, it is not assessed. The assessor may feel curtailed in his/her freedom by the criteria.

In a global assessment, the assessors study the entire portfolio and give an assessment based on their overall impression. A global assessment is far less labour intensive than an analytical assessment. It also enables assessors to take account of learners' special qualities.

Disadvantages are:

- It is less clear to learners on which criteria the assessment is based. The assessment may also be less reliable. As a result the assessment will be less acceptable to learners.
- Some assessors will feel less certain about their judgement. As a result they will study the material over and over again, which will take even more time than an analytical assessment.
- This type of assessment is relatively vulnerable to assessor preferences and sequence effects (the contrast with the previous candidate may influence the assessment).

A way to combine the best of both approaches is to use scoring rubrics. A *scoring rubric* is a global performance descriptor that lists the criteria for a competency and articulates a limited number of gradations of quality for each criterion. Gradations can be unsatisfactory, sufficient, good, and excellent. Scoring rubrics can be presented as tables, with the criteria in the rows and the grades in the columns. In each cell of this table, performance at that particular level of competence is described. Box 8 provides an example.

BOX 8
Rubrics used for the assessment for final year medical students (source Maastricht University)

	BELOW EXPECTATION	AS EXPECTED	ABOVE EXPECTATION
Clinical performance	<p>Slow in taking a history and performing a physical examination. Considers Irrelevant aspects.</p> <p>Slow in making a diagnosis. Misses important conclusions.</p> <p>Frequently unable to formulate management plan and needs considerable guidance.</p>	<p>Adequate speed in taking a history and performing a physical examination. Relevant aspects are considered.</p> <p>Adequate speed in making a diagnosis. Diagnosis contains important conclusions.</p> <p>Formulates an adequate management plan for simple clinical presentations.</p> <p>Needs some guidance.</p> <p>Achieves these goals in the second half of the internship.</p>	<p>Conducts an adequate and efficient history and physical examination.</p> <p>Arrives at an accurate diagnosis within adequate time.</p> <p>Formulates an adequate management plan for simple clinical presentations.</p> <p>Needs little guidance.</p> <p>Has achieved these goals at the start of the internship.</p>
Professionalism (for instance as judged by 360 degree feedback)	<p>Does not keep commitments.</p> <p>Occasionally fails to ask for supervision when this is necessary. Reacts defensively to feedback.</p> <p>Is unable to cope with stress</p> <p>Does not pay attention to his/her personal appearance.</p> <p>Frequently shows awkward behaviour or behaves disrespectfully.</p>	<p>Keeps commitments.</p> <p>Asks for supervision when this is necessary.</p> <p>Needs help in reflecting and considering alternatives and responds adequately to feedback.</p> <p>Occasionally needs help in coping with stress.</p> <p>Appropriate personal appearance; behaves respectfully.</p>	<p>Keeps commitments.</p> <p>Asks for supervision when this is necessary.</p> <p>Is able to reflect critically; responds adequately to feedback and is prepared to acknowledge errors.</p> <p>Is able to cope with stress adequately.</p> <p>Looks well cared for and behaves respectfully.</p>
Has critically assessed his/her performance and formulated appropriate learning goals. This is evidenced by an adequate analysis of strengths and weaknesses and the development plan.	<p>Incomplete, limited or one-sided description of strengths and weaknesses in performance (e.g. only strengths or only weaknesses, limited to one competency).</p> <p>No explanations only lists of facts or situations.</p> <p>No learning goals, learning goals do not match the analysis or are not specific.</p>	<p>A fair number of strengths and weaknesses are not explained or explanations are limited to external attributions (for instance mini-CEX at the wrong moment)</p> <p>Some of the learning goals are not specified.</p>	<p>Above expectation (authentic, recognizable, and well explained). A good analysis of strengths and weaknesses. Also internal attributions and references to evidence in the portfolio.</p> <p>Logical, detailed (based on the analysis) and attainable learning goals.</p>

For learners and their mentors, scoring rubrics can be a roadmap for competence development. It can help them diagnose a learner's current level of competence and point the way to further development. Assessors should not use scoring rubrics as a checklist,

but as a list of arguments to underpin their assessment when they explain it to learners. Learners can also use scoring rubrics to organise their portfolio. They can organise the evidence in their portfolio in chapters corresponding to the different competencies to be assessed and use captions to explain what the evidence shows about a specific competency.

Why? Assessing competencies can be done for three reasons: selection, diagnosis, and certification.

Selection: Determining whether a person is suitable for a certain position. Assessments for selection purposes can take place before entering an educational programme, but also, for instance, before starting a new job.

Diagnosis: In the course of an education programme, the development of learners' competencies is assessed. The purpose of this type of assessment is to give feedback to learners and help them identify new learning goals. Sometimes, this assessment is also used to determine whether or not a learner is allowed to continue with a programme.

Certification: The goal of assessment at the end of an educational or training programme is to establish whether learners have attained the competencies required for graduation or certification. Obviously, the quality of any assessment is important. Poor quality of assessment for selection purposes, for instance, can harm the interests of prospective learners and waste talent. Similarly, poor quality of diagnostic assessment can cause frustration and delay in learners' development. Nevertheless, with graduation and certification decisions the quality of assessment is crucial. Learners who pass but should have failed will become (or continue to be) certified doctors and may become a risk to the community!

How? The quality of the assessment of competencies is crucially determined by the procedure that is used. In the introduction to this section about portfolio assessment, we wrote that the standard psychometric procedures that are used to determine the quality of tests and standardised observations are not very well suited to portfolios with their non-standardised content. In medical education, Webb and colleagues (2003) pointed out that portfolio assessment is primarily concerned with qualitative information and they introduced the idea to use routines developed for qualitative research. Guba & Lincoln's (1989) strategies to achieve *credibility* and *dependability* of assessment can be translated to portfolio assessment (Webb et al., 2003; Tigelaar et al. 2005). In Box 9, we discuss how these strategies can be used.

...the standard psychometric procedures that are used to determine the quality of tests and standardised observations are not very well suited to portfolios with their non-standardised content.

BOX 9**Strategies for portfolio assessment derived from the methodology of qualitative research**

- Incorporate feedback cycles into the mentoring process that accompanies the portfolio to ensure that the mentor's final recommendation does not come as a(n) - unpleasant - surprise to the learner; this approach relates to the credibility strategies of prolonged engagement and member checking.
- Maintain a careful balance between the roles of the mentor as coach and assessor. The aim is to ensure that the person who knows the learner best provides the most relevant information while minimizing any damaging effect on the mentor-learner relationship by using an assessment committee to assess the portfolio; this approach relates to the credibility strategy of prolonged engagement.
- Involve the learner in the decision process to ensure commitment on the part of the learner and allow the learner to communicate a different point of view to that of the mentor; this approach relates to the credibility strategy of member checking.
- Use a sequential judgement procedure in which conflicting information necessitates more information gathering. This ensures the efficient use of resources by limiting the use of additional resources to cases where this is necessary to achieve reliable judgement. This approach relates to the credibility strategy of triangulation.
- Document the different steps of the assessment process. For example a formal assessment plan approved by the Examination Board; portfolio and assessment guidelines; overviews of the results per phase, and written assessment forms per learner. This approach relates to the dependability strategy of audit trail.

The major problem with qualitative research methods as well as with portfolio assessment is the required substantial time investment. At Maastricht University, we developed a portfolio assessment procedure that uses many of these strategies while at the same time aiming for optimal efficiency (Driessen et al., 2005a). This procedure is described in Box 10.

Who? A problem that is much debated in the portfolio literature is the feasibility and acceptability of combining the roles of mentor and assessor into one person. Tigelaar et al. interviewed nine portfolio experts about their views on the use of portfolios in education (Tigelaar et al. 2004). While some of the experts agreed that the mentor is the most appropriate person to advise an assessment committee about a candidate, others argued that it is unethical for mentors to undertake the assessor role. The latter group argued that candidates must feel free to reflect on their professional development together with their mentors, knowing that the mentor will not pass any information on to others. For this reason, the majority of the experts were of the opinion that mentors should not be involved in summative assessment nor make recommendations to an assessment committee. However, there was a minority who agreed with Snyder and colleagues, who wrote that: *"The tension between assessment for support and assessment for high stakes decision making will never disappear. Still, that tension is constructively dealt with daily by teacher educators throughout the nation"* (Snyder, et al., 1998, p. 59).

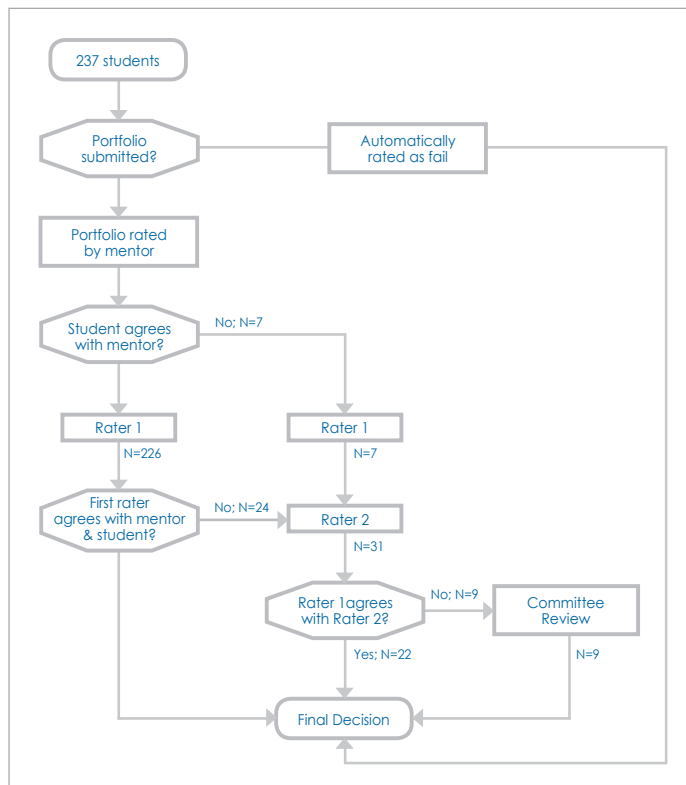
The tension between assessment for support and assessment for high stakes decision making will never disappear.

BOX 10
A procedure for portfolio assessment (Driessen et al., 2005a)

The student submits the portfolio to the mentor, who examines the portfolio and writes a recommendation regarding the grading of the portfolio to be submitted to the assessment committee.

In their final meeting of the academic year the student and the mentor discuss this recommendation. When student and mentor agree on the grade, the student signs the recommendation. If the student disagrees, he or she does not sign.

Subsequently, the portfolio is submitted to the assessment committee. This committee consists of all the mentors. The committee members do not grade the portfolios of the students they mentored themselves. Portfolios on which student and mentor agree are rated by one committee member, who does not study the portfolio in any great detail, but typically scans the work of the student and mentor and checks whether all procedures have been followed correctly. When rater and mentor agree on the grading, the recommendation becomes the final decision.



Striking the right balance between support and judgement is the challenge facing assessors/mentors with whom learners talk about their portfolios. A number of scenarios can be chosen in a procedure (Box 11). Which one is the most appropriate depends, amongst other things, on the educational context and the level of experience of the learners in question.

When? The answer to the question “when is this assessed?” depends on the answers to the other questions in this section.

Decisions about *selection* are made before the actual start of a programme or training period or after a first “trial” period, in which learners are observed and can prove themselves. The important question is whether a prospective learner matches the criteria for admission and whether this learner has the potential to finish an education or training programme.

Diagnostic assessment can be a frequent occurrence during an education or training programme. In fact, every time a mentor and a learner meet to discuss the learner's progress using information from the learner's portfolio, it can be qualified as diagnostic/ formative assessment. This implies that having easy access to a portfolio, for instance on-line, can be very helpful for mentors.

Decisions about *certification* are made when a learner's competencies match all the criteria or when the time available for a programme has run out. In an outcome based programme, this means that when the learner and his or her mentor conclude that the learner's competence meets all the criteria an assessment for certification purposes can take place. The logical consequence would be that if a person meets the competency criteria on entering an educational or training programme, he or she is exempt from training and awarded a certificate right away.

BOX 11**Portfolio assessors: scenarios**

Combining the role of the mentor and assessor is often considered problematic. On the hand, most people will agree that the mentor is probably the person who is best informed about the learner's competencies. As a consequence, ignoring the mentor's opinion in assessing the portfolio can be considered as missing the chance to improve the validity of the assessment. On the other hand, combining the roles of assessor and mentor can put a strain on the relationship between mentor and learner, because learners may be reluctant to discuss any difficulties they are facing for fear of repercussions in the assessment. Below we use the metaphors of the mentor as teacher, PhD supervisor, driving instructor, and coach to distinguish between four (non exclusive) scenarios. When mentors are in the role of a teacher, their role of assessor is most prominent. When they are in the role of a coach, they do not assess at all.

The teacher: This is the most common assessment scenario in education. Just like most teachers in primary, secondary, and higher education, mentors discuss their learners' performance and progress and assess their level of competence at the end of a course.

PhD supervisor: In some scenarios the role of the mentors in the assessment procedure of portfolios can be compared with the role of supervisors of PhD students. In many countries, the formal assessment of theses/portfolios is the responsibility of a committee. Supervisors invite their peers to sit on the committee but they themselves are not a member of the committee. A negative assessment of the thesis/portfolio would harm their reputation among their peers. For this reason they are highly unlikely to invite their peers to sit on the committee unless they are convinced the portfolio meets the criteria. As a consequence, mentors and students have the same interest: to produce a thesis or portfolio that merits a positive judgment.

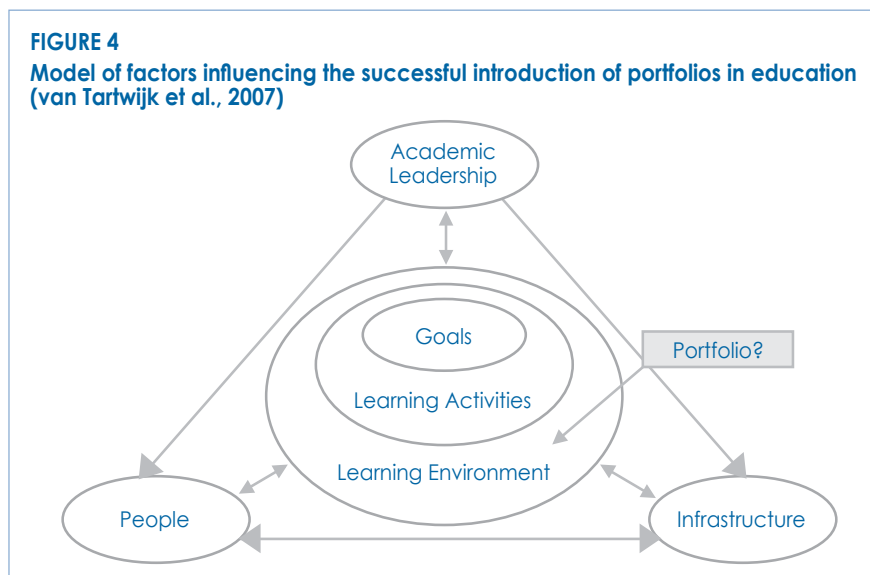
Driving license instructor: In this model the roles of the mentor and the assessor are strictly separated. The mentor/driving instructor mentors the learner in acquiring the required competencies, which are shown in the portfolio. If the mentor thinks the learner is competent, he invites an assessor from a professional body (i.e. the examiner from the Driver and Vehicle Licensing Agency) to assess the competence of the learner result. The learners can also approach the licensing agency themselves.

Coach: In this model, the learners themselves have the initiative. They can ask, for instance, a senior colleague to coach them until they have achieved the required level of competence. This scenario is likely, for instance, when a professional wants to acquire an additional qualification. The assessor would be someone from an external body.

Factors influencing the success of the introduction of a portfolio²

In the previous sections, we have argued that it is important to tailor portfolios to the intended purposes and to introduce portfolios only in situations in which they can serve a useful purpose. However, these conditions do not suffice to guarantee a successful introduction. In the literature on educational change, winning the hearts and minds of the people involved, both teachers and learners, as well as the quality of leadership are identified as key factors for lasting educational improvement (Martin et al. 2003; Hargreaves & Fink, 2004;).

Figure 4 presents a model in which portfolios are presented as part of the learning environment and in which three conditional factors are presented that influence whether an educational portfolio is introduced successfully or not: people (the teachers and learners), leadership, and infrastructure. The importance of these three conditional factors is discussed below.



People

Educational innovations involving the use of portfolios usually imply a transfer from teacher-directed education with a strong focus on conveying knowledge, to education in which the development of students' competencies in the workplace is emphasised. In most cases, teachers are expected to invest more time and effort in coaching and assessment than they were used to. Almost inevitably, this change in roles and routines will cause uncertainty and evoke resistance (Hammerness et al., 2005). Not only does it imply that teachers need to rethink key ideas, practices, and values, but for many teachers it also means that they need to invest in developing new competencies for coaching and assessment.

Educational innovations involving the use of portfolios usually imply a transfer from teacher-directed education with a strong focus on conveying knowledge, to education in which the development of students' competencies in the workplace is emphasised.

² Parts of this chapter were published before in *Quality in Higher Education* (van Tartwijk, et al., 2007)

In discussions about these innovations, the important questions are which educational problems need to be resolved and what is the most effective and efficient way to do that. Very often however, discussions concentrate on the portfolio, which becomes the visible “symbol” of the innovation. As a consequence, resistance to the innovation is likely to be projected onto the portfolio, while the important questions are not discussed.

Teachers are more likely to support and invest in educational changes if they acknowledge and subscribe to the educational value of the new learning approach, internalise and support the innovation, and are empowered to assume ownership of it. They are more likely to do so when it is clear to them how the innovation helps solve concrete problems that they have to cope with in their everyday teaching practice (Hargreaves et al. 1998). The risk that the important questions are not discussed can be reduced if teachers are involved in educational innovations at an early stage of decision-making. They are more likely to support and invest in working with a portfolio if the decision to work with this instrument was their own decision, based on their personal understanding and endorsement of the educational innovation and the role of the portfolio in it. From this perspective, the option should be kept of not using a portfolio when a better alternative is found. Teachers who have had a say in the decision to use a portfolio will feel a stronger commitment to it and will be more inclined to look for solutions and less likely to lay the instrument aside when faced with problems and inevitable design faults in the curriculum and the portfolio.

In the literature on educational change the importance of teachers as change agents is emphasised (Darling-Hammond et al., 2005) but the input of learners is crucial too. The successful introduction of a portfolio in education also depends on how much time and energy learners are willing to invest in their portfolios. In general, learners will only put effort into portfolios if this effort is rewarded in some way. The most obvious reward is that the portfolio is graded. In education, a very strong relationship exists between summative assessment and learning: assessment drives learning (Frederiksen, 1984; Driessen & van der Vleuten, 2000; van der Vleuten et al., 2000). Although assessment influences whether learners accept and put effort into a portfolio, assessment in itself is not enough. For learners, developing a portfolio implies putting a lot of effort into making their development visible. Thus, it is very frustrating for them if they discover that nobody takes a good look at the result of all their hard work. Mentors who take an interest in learners and their portfolios have been found to be a key factor in learners' appreciation of working with portfolios (Pearson & Heywood, 2004; Tigelaar et al. 2006).

A last condition for a successful introduction of portfolios related to learners and their mentors is their *understanding* of the portfolio and of what working with portfolios entails. Experience has shown that, although in theory portfolios can have a clear function in education, in practice the introduction of portfolios often leads to confusion and, consequently, frustration (Anderson & DeMeulle, 1998; Pearson & Heywood, 2004; Kjaer, et al., 2006; Davis et al. 2009). Most students who enrol in a medical curriculum are accustomed to teacher directed education. Self-assessment, asking for feedback, reflection and identifying personal learning needs, which are fundamental to portfolio learning (Snadden & Thomas, 1998b; Driessen et al. 2008), are perceived as

Although assessment influences whether learners accept and put effort into a portfolio, assessment in itself is not enough. For learners, developing a portfolio implies putting a lot of effort into making their development visible.

strange and sometimes even threatening by learners for whom education is synonymous with lectures and exams. Instructions are necessary that not only explain how to work with a portfolio, but also help learners and their mentors understand what a portfolio is and why it used in education. A study by Duque and colleagues (Duque et al., 2006) demonstrated that hands-on introduction with a proper briefing of learners by staff on the portfolio's purpose and procedures had a positive effect on portfolio scores and learner satisfaction with the portfolio. We have experimented with the use of the analogy between a portfolio and a CV to help learners better understand what a portfolio is and what working with a portfolio entails (van Tartwijk et al. 2008).

Academic leadership

Commitment by educational leaders is another vital condition for the successful introduction of portfolios. In a study on perceptions of leadership in academic contexts, Martin and her colleagues (2003) found that the quality of student learning is affected by the way leadership is constituted and experienced in academic contexts. A group of educational leaders was identified who were successful in stimulating teachers to adopt a student-focused approach to teaching. A characteristic of these educational leaders is that they discuss and negotiate these changes with the teachers. Similar findings are reported by Bland and her colleagues (2000), who reviewed the available literature with the aim to identify a set of characteristics that are associated with successful curricular change in medical education. They write that leadership comes up again and again as critical to the success of curricular change. The literature shows that successful and less successful leaders in medical education use organizational authority at about the same rate, but also that successful leaders more often seek input from others. When educational innovations ask teachers to change their roles and routines, these teachers must know that they can rely on educational leaders who support and value their commitment in every respect (Malden, 1994; van Veen et al. 2005). And finally, of course, commitment of the academic leaders is also reflected in the allocation of sufficient financial resources to ensure that the intended changes can actually be implemented.

Infrastructure

An increasing number of Faculties of Medicine are choosing to work with electronic rather than paper portfolios. In the section on e-portfolios, we described the reasons for this choice. We also wrote that research shows that adverse conditions like limited computer access in the workplace may cancel out the advantages of an e-portfolio. In general we conclude that e-portfolios are vulnerable to adverse conditions, because the demands of the technical infrastructure are large. If the electronic part of the portfolio system malfunctions, that is usually all the excuse that the adversaries of the use of portfolios need to drop the idea of a portfolio altogether, including the curriculum innovation for which the portfolio very often is a symbol.

Concluding remarks

In curricula with a strong focus on the development and assessment of competencies a portfolio can be a valuable instrument. They have the potential to make learning visible on the Does level of Miller's pyramid (Miller 1990), which describes independent performance in the workplace. However, portfolios are also vulnerable. Portfolio learning requires reflection by learners and investment in coaching by teachers. The quality of portfolio assessment depends on investing in the interpretation of and discussion about qualitative data. Not only does it require a new perspective on education from mentors and learners, many of whom are used to teacher-directed learning with a strong emphasis on the acquisition of knowledge, it also asks teachers and learners for a significant investment of time and energy. The literature shows that many conditions need to be fulfilled to enable successful introduction of a portfolio (Driessen et al., 2007b), and even then a portfolio is not a cure for all pains.

We conclude this Guide for using portfolios for assessment and learning by referring to Spandel once more (Spandel, 1997), who wrote:

"..... introducing portfolios is just like buying shoes: the best choice depends on purpose and comfort comes with wearing".

We would like to add that portfolios are like expensive shoes and even during the process of getting used to them, there will inevitably be times when one's toes are really hurting. However, for those owners who persist, the portfolio has the potential to become one of their best purchases.

Portfolio learning requires reflection by learners and investment in coaching by teachers. The quality of portfolio assessment depends on investing in the interpretation of and discussion about qualitative data.

"..... introducing portfolios is just like buying shoes: the best choice depends on purpose and comfort comes with wearing".

References

- ANDERSON RS & DEMEULLE L (1998). Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly*, 25: 23-32.
- BIRD T (1990). The schoolteacher's portfolio: an essay on possibilities. In: J Millman & L Darling-Hammond (Eds), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*, pp. 241-256 (Newbury Park, CA, Corwin Press, inc).
- BLAND CJ, STARNAMAN S, WERSAL L, MOORHEAD-ROSENBERG L, ZONIA S & HENRY R (2000). Curricula change in medical schools: How to succeed. *Academic Medicine*, 75: 575-594.
- BRANSFORD J, BROWN AL & COCKING RR (Eds) (2000). *How people learn: Brain, mind, experience, and school*. (Washington D.C., National Academy Press).
- BUCKLEY S, ASHCROFT T, DAVIS J, KHAN KS, MORLEY D, POLLARD D, POPOVIC C, SAYERS J, SUSARLA R, THOMAS H & ZAMORA J (in press). The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education systematic review, *Medical Teacher*.
- COLLINS A (1991). Portfolios for biology teacher assessment. *Journal of Personnel Evaluation in Education*, 5: 147-167.
- CONLON M (2003). Appraisal: The catalyst of personal development. *British Medical Journal*, 327: 389-391.
- DARLING-HAMMOND L, PACHECO A, MICHELLI N, LEPAGE P, HAMMERNESS K & YOUNG P (2005). Implementing curriculum renewal in teacher education: managing organizational and policy change. In: L Darling-Hammond, J Bransford, P LePage, K Hammerness & H Duffy (Eds), *Preparing teachers for a changing world: What teachers should learn and be able to do*, pp. 442-479 (San Francisco, Jossey-Bass).
- DAVIS DA, MAZMANIAN PE, FORDIS M, VAN HARRISON R, THORPE KE & PERRIER L (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*, 296: 1094-1102.
- DAVIS MH, FRIEDMAN BEN DAVID M, HARDEN RM, HOWIE P, KER J, MCGHEE C, et al. (2001). Portfolio assessment in medical students' final examinations. *Medical Teacher*, 23: 357-366.
- DAVIS MH, PONNAMPERUMA GG, & KER JS (2009). Student perceptions of a portfolio assessment process. *Medical Education*, 43: 89-98.
- DENZIN NK (1978). *Sociological Methods: A Sourcebook* (2nd ed.). New York: McGraw Hill.
- DENZIN NK & LINCOLN YS (2000). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- DORNAN T, CARROLL C & PARBOOSHING J (2002). An electronic learning portfolio for reflective continuing professional development. *Medical Education*, 36: 767-769.
- DREYFUS SE (2004). The five-stage model of adult skill acquisition. *Bulletin of Science Technology and Society*, 24: 117-181.
- DRIESSEN EW, MUIJTJENS AMM, VAN TARTWIJK J & VAN DER VLEUTEN CPM (2007a). Web- or paper-based portfolios: is there a difference? *Medical education*, 41: 1067-1073.
- DRIESSEN EW & VAN DER VLEUTEN CPM (2000). Matching student assessment to problem based learning: lessons from experience in a law faculty. *Studies in Continuing Education*, 22: 235-248.
- DRIESSEN EW, VAN DER VLEUTEN CPM, SCHUWIRTH L, VAN TARTWIJK J & VERMUNT JD (2005a). Credibility of portfolio assessment as an alternative for reliability evaluation: a case study. *Medical Education*, 39: 214-220.
- DRIESSEN EW, VAN TARTWIJK J & DORNAN T (2008). The self-critical doctor: Helping students become more reflective. *BMJ*, 336: 827-830.
- DRIESSEN EW, VAN TARTWIJK J, OVEREEM K, VERMUNT JD & VAN DER VLEUTEN CPM (2005b). Conditions for successful reflective use of portfolios in undergraduate medical education. *Medical Education*, 39: 1230-1235.
- DRIESSEN EW, VAN TARTWIJK J, VAN DER VLEUTEN CPM, & WASS V (2007b). Portfolios in medical education: Why do they meet with mixed success? A systematic review. *Medical Education*, 41: 1224-1233.

- DRIESSEN EW, VAN TARTWIJK J, VERMUNT JD & VAN DER VLEUTEN CPM (2003). Use of portfolio in early undergraduate medical training. *Medical Teacher*, 25: 18-23.
- DUQUE G, FINKELSTEIN A, ROBERT A, TABATABA D, GOLD SL & WINER LR (2006). Learning while evaluating: the use of an electronic evaluation portfolio in a geriatric medicine clerkship. *BMC Medical Education*, 6: 1-7.
- ERICSSON KA (2006). The influence of experience and deliberate practice on the development of expert performance. In: KA Ericsson, N Charness, PJ Feltovich & RR Hoffman (Eds), *The Cambridge handbook of expertise and expert performance* (pp. 683-704). New York: Cambridge University Press.
- EVA KW & REGEHR G (2008). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, 28: 14-19.
- FINLAY IG, MAUGHAN TS & WEBSTER DJ (1998). A randomized controlled study of portfolio learning in undergraduate cancer education. *Medical Education*, 32: 172-176.
- FREDERIKSEN N (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39: 193-202.
- FRIEDMAN BEN DAVID M, DAVIS MH, HARDEN RM, HOWIE PW, KER J & PIPPARD MJ (2001). *AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment* (Dundee, Association for Medical Education in Europe).
- FUNG KEE FUNG M, WALKER M, FUNG KEE FUNG K, TEMPLE L, LAJOIE F, BELLEMARE G, et al. (2000). An Internet-based learning portfolio in resident education: The KOALA-super (TM) multicentre programme. *Medical Education*, 34: 474-479.
- GENERAL MEDICAL COUNCIL (2000). *Revalidating doctors: Ensuring standards, securing the future*. London: GMC.
- GIBSON D & BARRETT H (2003). Directions in Electronic Portfolio Development. *Contemporary Issues in Technology and Teacher Education*, 2: 559-576.
- GRANT AJ, VERMUNT JD, KINNERSLEY P & HOUSTON H (2007). Exploring students' perceptions of the use of a significant event analysis as part of a portfolio assessment process in general practice, as a tool for learning how to use reflection in learning. *BMC Medical Education*: 7:5.
- GUBA EG & LINCOLN YS (1989). Judging the quality of fourth generation evaluation. In: EG Guba & YS Lincoln (Eds), *Fourth Generation Evaluation* (London, Sage).
- HAMMERNES K, DARLING-HAMMOND L, BRANSFORD J, BERLINER DC, COCHRAN-SMITH M, MCDONALD M, et al. (2005). How teachers learn and develop. In: L Darling-Hammond, J Bransford, P LePage, K Hammerness & H Duffy (Eds), *Preparing teachers for a changing world: What teachers should learn and be able to do*, pp. 358-389 (San Francisco, Jossey-Bass).
- HARDEN RM (1979). How to assess students: An overview. *Medical Teacher*, 1: 65-70.
- HARGREAVES A & FINK D (2004). The seven principles of sustainable leadership. *Educational Leadership*, April 2004: 8-13.
- HARGREAVES A, LIEBERMAN A, FULLAN M & HOPKINS D (Eds) (1998). *International handbook of educational change* (Dordrecht: Kluwer Academic Publishers).
- HATTON N & SMITH D (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11: 33-49.
- KJAER NK, MAAGARD R & WIES S (2006). Using an online portfolio in postgraduate training. *Medical Teacher*, 28: 708-712.
- KORTHAGEN FAJ, KESSELS J, KOSTER B, LAGERWERF B & WUBBELS T (2001). *Linking theory and practice: The pedagogy of realistic teacher education* (Mahwah, NY, Lawrence Erlbaum Associates).
- KORTHAGEN FAJ, KOSTER B, MELIEF K & TIGCHELAAR A (2002). *Teach teachers to reflect: Systematic reflection in the training and coaching of teachers* [In Dutch: Docenten leren reflecteren: Systematische reflectie in de opleiding en begeleiding van leraren] (Soest, Uitgeverij Nelissen).
- LAWSON M, NESTEL D & JOLLY B (2004). An e-portfolio in health professional education. *Medical Education*, 38: 569-570.
- LOCKYER JM & CLYMAN SG (2008). Multisource feedback (360-degree feedback). In: ES Holmboe & RE Hawkins (Eds), *Practical guide to the evaluation of clinical competence*, pp. 75-85 (Philadelphia, Pa, Mosby Elsevier).

- LYONS N (1998). Reflection in teaching: Can it be developmental? A portfolio perspective. *Teacher Education Quarterly*, Winter 1998: 115-127.
- MALDEN B (1994). The micropolitics of education: mapping the multiple dimensions of power relations in school policies. *Journal of Educational Policy*, 9: 147-167.
- MANN K, GORDON J & MACLEOD A (2007). Reflections and reflective practice in health profession education: A systematic review. *Advanced Health Science Education*, (First published online November 2007): 1-27.
- MARTIN E, TRIGWELL K, PROSSER M & RAMSDEN P (2003). Variations in the experience of leadership of teaching in higher education. *Studies in Higher Education*, 28: 247-259.
- MATHERS NJ, CHALLIS MC, HOWE AC & FIELD NJ (1999). Portfolios in continuing medical education – effective and efficient? *Medical Education*, 33: 521-530.
- MILLER GE (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65: S63-67.
- NORCINI JJ & BURCH VC (Eds) (2007). *Workplace-based assessment as an educational tool*, AMEE Guide 31 (Dundee, UK, AMEE).
- NORCINI JJ, HOLMBOE ES & HAWKINS RE (2008). Evaluation challenges in the era of outcome based education. In: ES Holmboe & RE Hawkins (Eds), *Practical guide to the evaluation of clinical competence*, pp. 1-9 (Philadelphia, PA, Mosby Elsevier).
- O'SULLIVAN PS, RECKASE MD, MCCLAIN T, SAVIDGE MA & CLARDY JA (2004). Demonstration of portfolios to assess competency of residents. *Advances in Health Sciences Education*, 9: 1-15.
- OERMANN MH (2002). Developing a professional portfolio in Nursing. *Orthopaedic Nursing*, 21: 73-78.
- PAULSON FL, PAULSON PR & MEYER CA (1991). What makes a portfolio a portfolio? Eight thoughtful guidelines will help educators encourage self directed learning. *Educational Leadership*, February 1991: 60-63.
- PEARSON DJ & HEYWOOD P (2004). Portfolio use in general practice vocational training: A survey of GP registrars. *Medical Education*, 38: 87-95.
- ROYAL COLLEGE OF GENERAL PRACTITIONERS (1993). *Portfolio-based learning in general practice: Report of a working group on higher professional education*, Occasional paper 63 (London, Royal College of General Practitioners).
- ROYAL COLLEGE OF PHYSICIANS AND SURGEONS OF CANADA (1996). *Canmeds 2000 Project: Skills for the New Millennium. Report on the societal needs working group* (Ottawa, The Royal College of Physicians and Surgeons of Canada).
- SHULMAN LS (1998). Teacher portfolios: a theoretical activity. In: N Lyons (Ed), *With portfolio in hand: validating the new teacher professionalism*, pp. 23-38 (New York, Teachers College Press).
- SHUTE VJ (2008). Focus on formative feedback. *Review of Educational Research*, 78: 153-189.
- SMITHER JW, LONDON M, FLAUTT R, VARGAS Y & KUCINE I (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personal Psychology*, 56: 23-44.
- SNADDEN D (1999). Portfolios – attempting to measure the unmeasurable? [Commentary]. *Medical Education*, 33(7): 478-479.
- SNADDEN D, CHALLIS M, & THOMAS ML (1999). *AMEE Medical Education Guide No. 11: Portfolio-based learning and assessment* (Dundee, Association for Medical Education in Europe).
- SNADDEN D & THOMAS ML (1998a). Portfolio learning in general practice vocational training - does it work? *Medical Education*, 32: 401-406.
- SNADDEN D & THOMAS ML (1998b). The use of portfolio learning in medical education. *Medical Teacher*, 20: 192-199.
- SNYDER J, LIPPINCOTT A & BOWER D (1998). The inherent tensions in the multiple uses of portfolios in teacher education. *Teacher Education Quarterly*, 25: 45-60.
- SPANDEL V (1997). Reflections on portfolios. In: GD Phye (Ed), *Handbook of academic learning: Construction of knowledge* (pp. 573-591). San Diego: Academic Press.

Series 2

- 30 Peer Assisted Learning: a planning and implementation framework**
Michael Ross & Helen Cameron (2007)
ISBN: 978-1-903934-38-8
Primarily designed to assist curriculum developers, course organisers and educational researchers develop and implement their own PAL initiatives.
- 31 Workplace-based Assessment as an Educational Tool**
John Norcini & Vanessa Burch (2008)
ISBN: 978-1-903934-39-5
Several methods for assessing work-based activities are described, with preliminary evidence of their application, practicability, reliability and validity.
- 32 e-Learning in Medical Education**
Rachel Ellaway & Ken Masters (2008)
ISBN: 978-1-903934-41-8
An increasingly important topic in medical education – a 'must read' introduction for the novice and a useful resource and update for the more experienced practitioner.
- 33 Faculty Development: Yesterday, Today and Tomorrow**
Michelle McLean, Francois Cilliers & Jacqueline M van Wyk (2010)
ISBN: 978-1-903934-42-5
Useful frameworks for designing, implementing and evaluating faculty development programmes.
- 34 Teaching in the clinical environment**
Subha Ramani & Sam Leinster (2008)
ISBN: 978-1-903934-43-2
An examination of the many challenges for teachers in the clinical environment, application of relevant educational theories to the clinical context and practical teaching tips for clinical teachers.
- 35 Continuing Medical Education**
Nancy Davis, David Davis & Ralph Bloch (2010)
ISBN: 978-1-903934-44-9
Designed to provide a foundation for developing effective continuing medical education (CME) for practicing physicians.
- 36 Problem-Based Learning: where are we now?**
David Taylor & Barbara Mifflin (2010)
ISBN: 978-1-903934-45-6
A look at the various interpretations and practices that claim the label PBL, and a critique of these against the original concept and practice.
- 37 Setting and maintaining standards in multiple choice examinations**
Raja C Bandaranayake (2010)
ISBN: 978-1-903934-51-7
An examination of the more commonly used methods of standard setting together with their advantages and disadvantages and illustrations of the procedures used in each, with the help of an example.
- 38 Learning in Interprofessional Terms**
Marilyn Hammick, Lorna Olickers & Charles Champion-Smith (2010)
ISBN: 978-1-903934-52-4
Clarification of what is meant by Interprofessional learning and an exploration of the concept of teams and team working.
- 39 Online eAssessment**
Reg Dennick, Simon Wilkinson & Nigel Purcell (2010)
ISBN: 978-1-903934-53-1
An outline of the advantages of on-line eAssessment and an examination of the intellectual, technical, learning and cost issues that arise from its use.
- 40 Creating effective poster presentations**
George Hess, Kathryn Tosney & Leon Liegel (2009)
ISBN: 978-1-903934-48-7
Practical tips on preparing a poster – an important, but often badly executed communication tool.
- 41 The Place of Anatomy in Medical Education**
Graham Louw, Norman Eizenberg & Stephen W Carmichael (2010)
ISBN: 978-1-903934-54-8
The teaching of anatomy in a traditional and in a problem-based curriculum from a practical and a theoretical perspective.
- 42 The use of simulated patients in medical education**
Jennifer A Cleland, Keiko Abe & Jan-Joost Rethans (2010)
ISBN: 978-1-903934-55-5
A detailed overview on how to recruit, train and use Standardized Patients from a teaching and assessment perspective.
- 43 Scholarship, Publication and Career Advancement in Health Professions Education**
William C McGaghie (2010)
ISBN: 978-1-903934-50-0
Advice for the teacher on the preparation and publication of manuscripts and twenty-one practical suggestions about how to advance a successful and satisfying career in the academic health professions.
- 44 The Use of Reflection in Medical Education**
John Sandars (2010)
ISBN: 978-1-903934-56-2
A variety of educational approaches in undergraduate, postgraduate and continuing medical education that can be used for reflection, from text based reflective journals and critical incident reports to the creative use of digital media and storytelling.
- 45 Portfolios for Assessment and Learning**
Jan van Tartwijk & Erik W Driessen (2010)
ISBN: 978-1-903934-57-9
An overview of the content and structure of various types of portfolios, including eportfolios, and the factors that influence their success.
- 46 Student Selected Components**
Simon C Riley (2010)
ISBN: 978-1-903934-58-6
An insight into the structure of an SSC programme and its various important component parts.
- 47 Using Rural and Remote Settings in the Undergraduate Medical Curriculum**
Maira Maley, Paul Worley & John Dent (2010)
ISBN: 978-1-903934-59-3
A description of an RRME programme in action with a discussion of the potential benefits and issues relating to implementation.
- 48 Effective Small Group Learning**
Sarah Edmunds & George Brown (2010)
ISBN: 978-1-903934-60-9
An overview of the use of small group methods in medicine and what makes them effective.

To see the full list of guides available, and to order, see the website www.amee.org.

About AMEE

What is AMEE?

AMEE is an association for all with an interest in medical and healthcare professions education, with members throughout the world. AMEE's interests span the continuum of education from undergraduate/basic training, through postgraduate/specialist training, to continuing professional development/continuing medical education.

- **Conferences:** Since 1973 AMEE has been organising an annual conference, held in a European city. The conference now attracts over 2300 participants from 80 countries.
- **Courses:** AMEE offers a series of courses at AMEE and other major medical education conferences relating to teaching, assessment, research and technology in medical education.
- **MedEdWorld:** AMEE's exciting new initiative has been established to help all concerned with medical education to keep up to date with developments in the field, to promote networking and sharing of ideas and resources between members and to promote collaborative learning between students and teachers internationally.
- **Medical Teacher:** AMEE produces a leading international journal, Medical Teacher, published 12 times a year, included in the membership fee for individual and student members.
- **Education Guides:** AMEE also produces a series of education guides on a range of topics, including Best Evidence Medical Education Guides reporting results of BEME Systematic Reviews in medical education.
- **Best Evidence Medical Education (BEME):** AMEE is a leading player in the BEME initiative which aims to create a culture of the use of best evidence in making decisions about teaching in medical and healthcare professions education.

Membership categories

- **Individual and student members (£85/£39 a year):** Receive Medical Teacher (12 issues a year, hard copy and online access), free membership of MedEdWorld, discount on conference attendance and discount on publications.
- **Institutional membership (£200 a year):** Receive free membership of MedEdWorld for the institution, discount on conference attendance for members of the institution and discount on publications.

See the website (www.amee.org) for more information.

If you would like more information about AMEE and its activities, please contact the AMEE Office:
Association for Medical Education in Europe (AMEE), Tay Park House, 484 Perth Road, Dundee DD2 1LR, UK
Tel: +44 (0)1382 381953; Fax: +44 (0)1382 381987; Email: amee@dundee.ac.uk

www.amee.org

Scottish Charity No. SC 031618

ผลลัพธ์การปฏิบัติงานของ



นายแพทย์ X

อาจารย์ที่ปรึกษา อาจารย์ A

ตามการประเมินด้วยแฟ้มสะสมผลงานการ (Portfolio)

ปีการศึกษา 2554-2556

Competency based portfolio assessment

Academic year 2011-2013

สาส์นจากหัวหน้าภาควิชา

ภาควิชาสูติศาสตร์-นรีเวชวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ขอแสดงความยินดีกับ **นายแพทย์ A** ที่สำเร็จการฝึกอบรมแพทย์ประจำบ้าน สาขาสูติศาสตร์-นรีเวชวิทยา ระหว่างปีการศึกษา 2553-2555

ตลอดระยะเวลาสามปีที่ผ่านมา ภาควิชาฯ ได้ดำเนินการประเมินคุณสมบัติด้านต่างๆ ของท่าน ได้แก่ ความรู้ ทักษะหัตถการ การวิจัย และพฤติกรรมการทำงาน ในรูปแบบ Portfolio ดังผลสรุปในเอกสารฉบับนี้

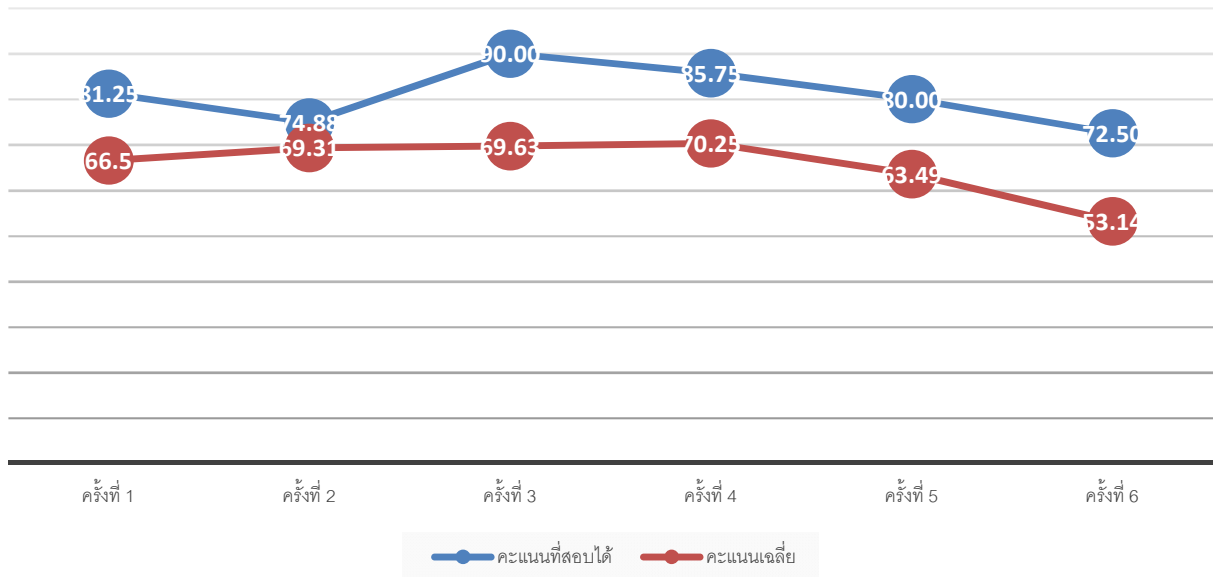
ภาควิชาฯ ขออัญวยพรให้ท่านประสบความสำเร็จในการดำเนินชีวิตครอบครัว และหน้าที่การงาน ตลอดไป

ศาสตราจารย์คลินิก นายแพทย์ชาญชัย วันทนาศิริ
หัวหน้าภาควิชาสูติศาสตร์-นรีเวชวิทยา
คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

การประเมินความรู้ทางสูติศาสตร์-นรีเวชวิทยา

(Knowledge assessment)

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 1

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	66.50	81.25	1
2	100	69.31	74.88	3
3	100	69.63	90.00	1
4	100	70.25	85.75	1
5	100	63.49	80.00	1
6	100	53.14	72.50	2

ผลการสอบตามหลักสูตรประกาศนียบัตรบัณฑิตชั้นสูงสาขาวิทยาศาสตร์การแพทย์คลินิก:

The Higher Graduate Diploma (Clinical Medical Sciences) คณะแพทยศาสตร์ศิริราชพยาบาล

ผ่าน ได้รับประกาศนียบัตรเมื่อ 25 พฤษภาคม 2555

ไม่ผ่าน

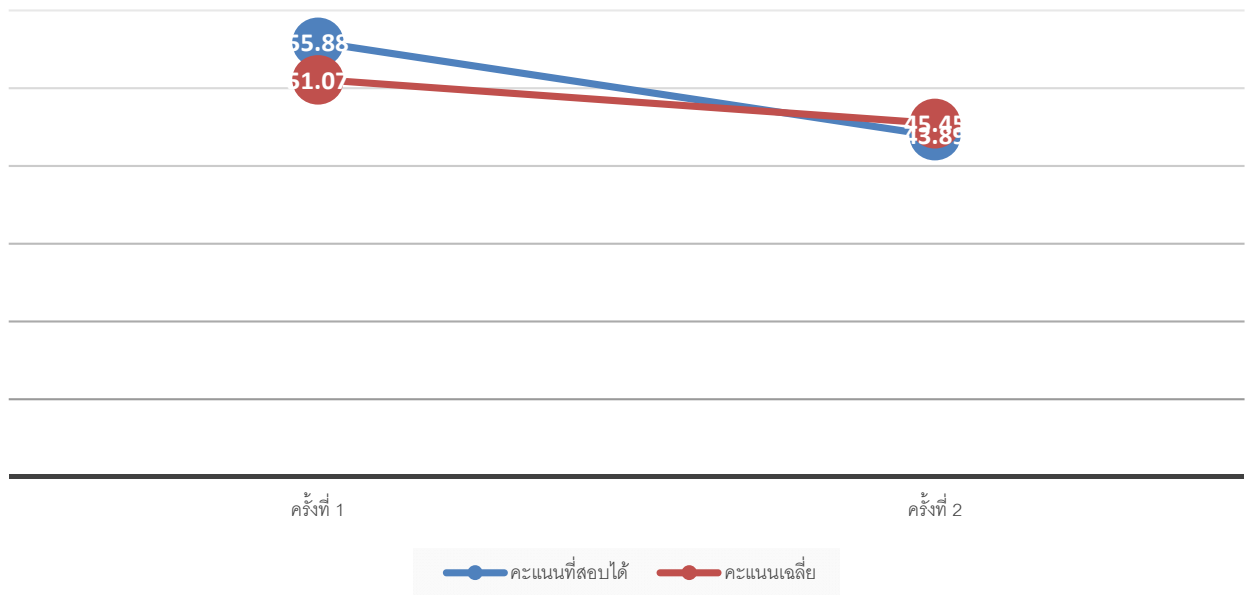
การสอบวิชาภาษาอังกฤษและกฎหมายทางการแพทย์ครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบวิชาภาษาอังกฤษและกฎหมายทางการแพทย์ครั้งที่ 2 (กรณีสอบไม่ผ่านครั้งแรก)

ผ่าน ไม่ผ่าน

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 2

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	51.07	55.88	5
2	100	45.45	43.89	10

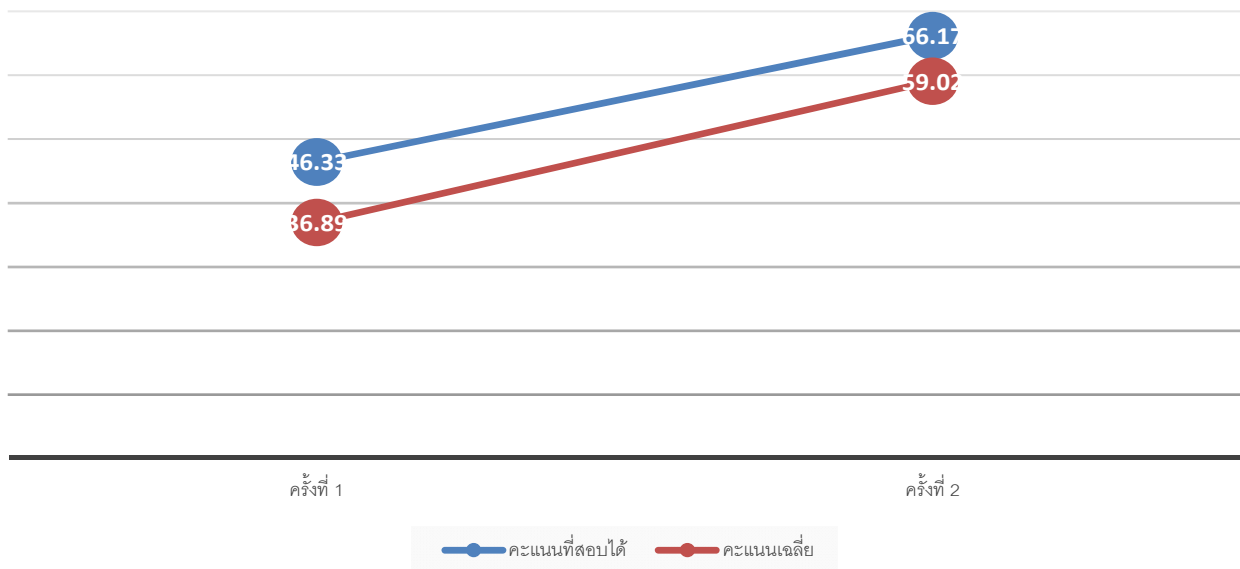
การสอบ OSLER ในสถาบัน ครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบ Basic science ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย

ผ่าน ไม่ผ่าน

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 3

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	36.89	46.33	2
2	100	59.02	66.17	1

การสอบ OSLER ในสถาบัน ครั้งที่ 2

ผ่าน ไม่ผ่าน

การสอบ OSLER ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทยครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบ OSLER ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทยครั้งที่ 2 (กรณีสอบครั้งแรกไม่ผ่าน)

ผ่าน ไม่ผ่าน

การสอบงานวิจัย ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย

ผ่าน ไม่ผ่าน

หัตถการสำคัญทางสูติศาสตร์-นรีเวชวิทยาที่ปฏิบัติ

ขณะเป็นแพทย์ประจำบ้านชั้นปีที่ 3

(Clinical skills assessment when being the 3rd year resident)

การผ่าตัดทางนรีเวช

การผ่าตัด	จำนวน
Total abdominal hysterectomy +/- bilateral salpingoophorectomy	19
Vaginal hysterectomy +/- AP repair	4
Adnexal surgery: Salpingectomy/Salpingotomy/Salpingostomy	21
Cervical conization	11

การผ่าตัดทางสูติศาสตร์

การผ่าตัด	จำนวน
Cesarean delivery	55
Tubal sterilization	3
Dilatation and curettage	16
Vacuum extraction/Forceps extraction	4
Breech assisting	
Manual removal of placenta	2

หมายเหตุ

จำนวนหัตถการเป็นจำนวนโดยประมาณ เนื่องจากอยู่ระหว่างกระบวนการพัฒนาและปรับปรุงระบบเก็บข้อมูลหัตถการแพทย์ประจำบ้าน ภาควิชาสูติศาสตร์-นรีเวชวิทยา

การทำงานวิจัยระดับแพทย์ประจำบ้าน
(Research competency)

เรื่อง **Prevalence and Associating Factors of Sexual Dysfunction in Women Who Use Intrapartum Device (IUD)**

อาจารย์ผู้ควบคุมผู้ช่วยศาสตราจารย์นายแพทย์ธันยารัตน์ วงศ์วนานุรักษ์

ข้อมูลสำคัญสำหรับงานวิจัย

1. ผ่าน SIRB เมื่อ 21 กุมภาพันธ์ 2555
เลขที่ 813/2554 (EC3)
2. ประกวดการนำเสนองานวิจัยในการประชุมราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย
วันที่ 26 พฤศจิกายน 2556
 เข้าร่วมนำเสนอ ไม่ได้รับรางวัล
 เข้าร่วมนำเสนอ ได้รับรางวัล ชมเชย
3. การตีพิมพ์ในวารสารวิชาการ
 ไม่ได้ตีพิมพ์
 ได้รับการตีพิมพ์ (ระบุรายละเอียดวารสาร) J Med Assoc Thai 2014
Full text. E-Journal: <http://Jmatonline.com>

ผลการประเมินเจตคติและพฤติกรรมการปฏิบัติงานของแพทย์ประจำบ้าน (Multisources feedback)

แพทย์ประจำบ้านจะได้รับการประเมินในประเด็นต่อไปนี้

1. ความรู้ความสามารถด้านวิชาการ

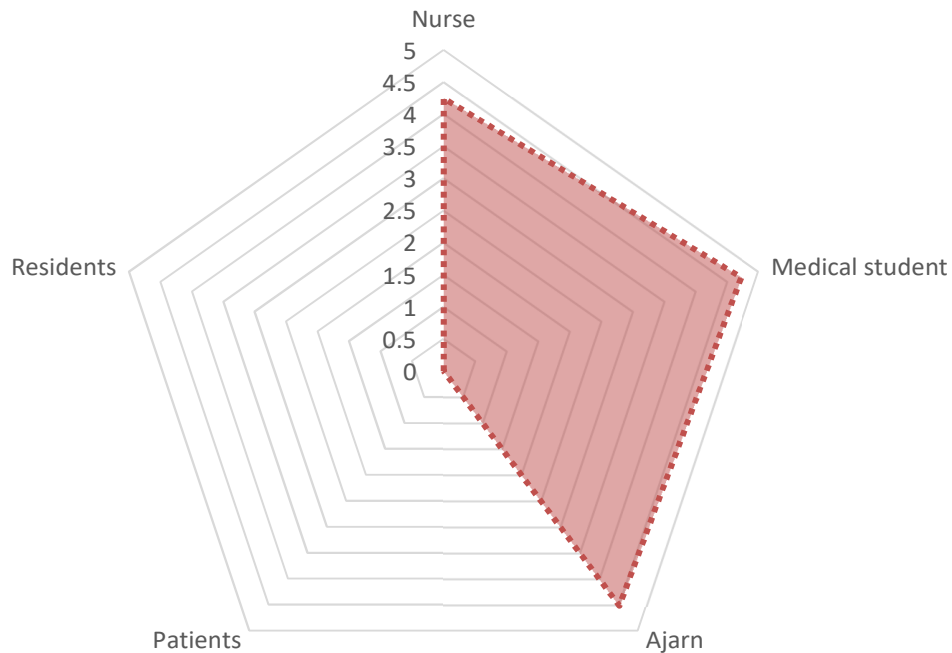
2. ทักษะพื้นฐานในการปฏิบัติงาน

ได้แก่ ทักษะการสื่อสารกับเพื่อนร่วมงานและผู้ป่วย/ญาติ การบันทึกรายงานผู้ป่วย การทำงานร่วมกับผู้อื่น และบุคลิกภาพขณะปฏิบัติงาน

3. คุณธรรมและจริยธรรม

ได้แก่ ความรับผิดชอบ ความเสียสละ ความตรงต่อเวลา ความซื่อสัตย์ การปฏิบัติตามระเบียบข้อบังคับ และอภัย/น้ำใจ/ความเอื้อเฟื้อต่อผู้อื่น

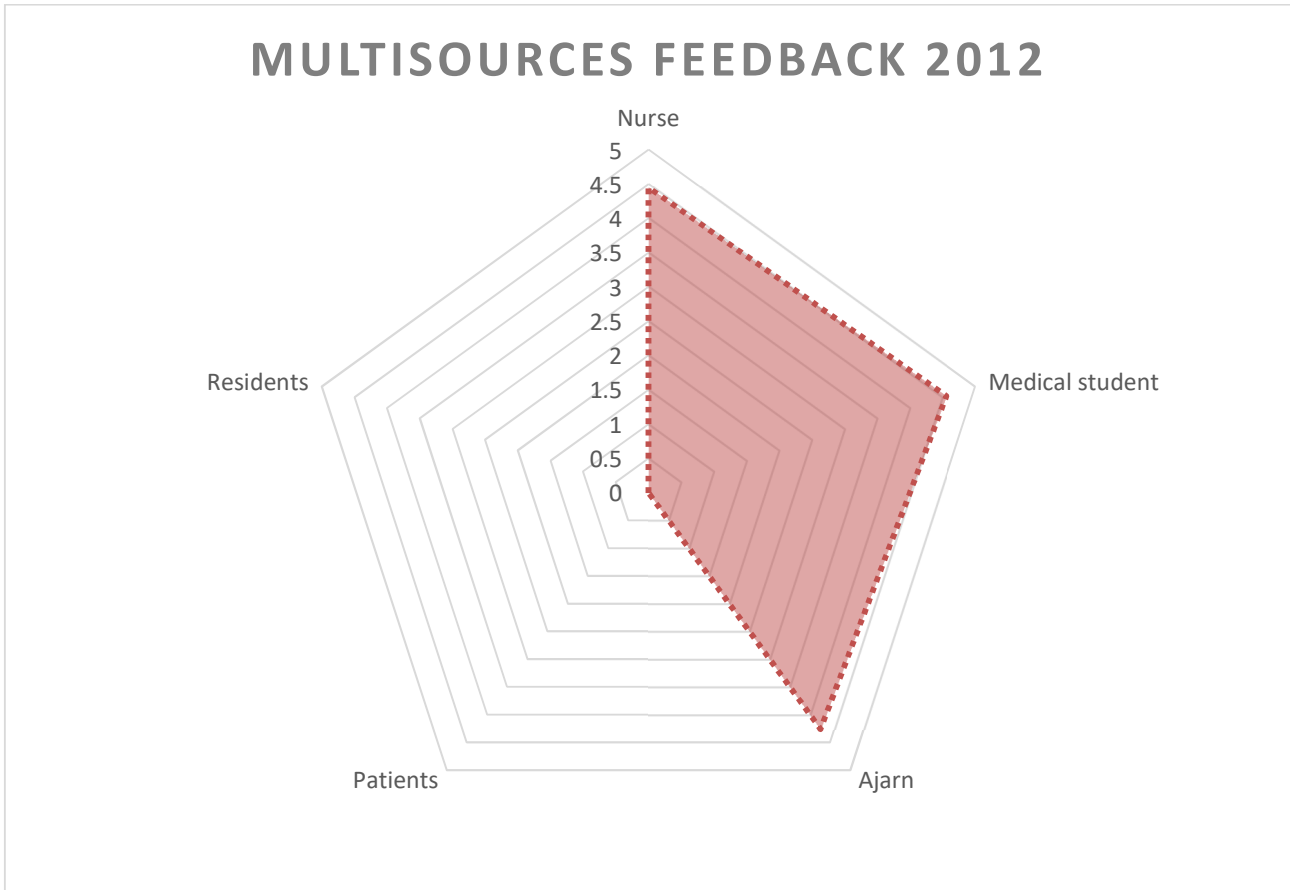
MULTISOURCES FEEDBACK 2011



ชั้นปีที่ 1 ปีการศึกษา 2554

Rotation	อาจารย์ (5 คะแนน)	นักศึกษาแพทย์ (5 คะแนน)	พยาบาล (5 คะแนน)	แพทย์ ประจำบ้าน	ผู้รับบริการ
พระศรีฯ 9/2			4.61		
LR เช้า		5.00	3.76		
LR พิเศษเช้า			4.61		
นรีเวช 1	4.90	5.00	4.00		
นรีเวช 1 (2)	4.50	4.90	4.00		
พระศรีฯ 10/2			4.46		
พระศรีฯ 9/1+ANC			5.00		
LR ดึก			4.00		
LR พิเศษบ่าย			4.30		
นรีเวช 2	4.20	4.50	4.56		
Onco	4.50	4.30	3.84		
พระศรีฯ 10/3		5.00	4.30		
พระศรีฯ 10/1		4.46	3.91		
คะแนนเฉลี่ย	4.52	4.73	4.25		

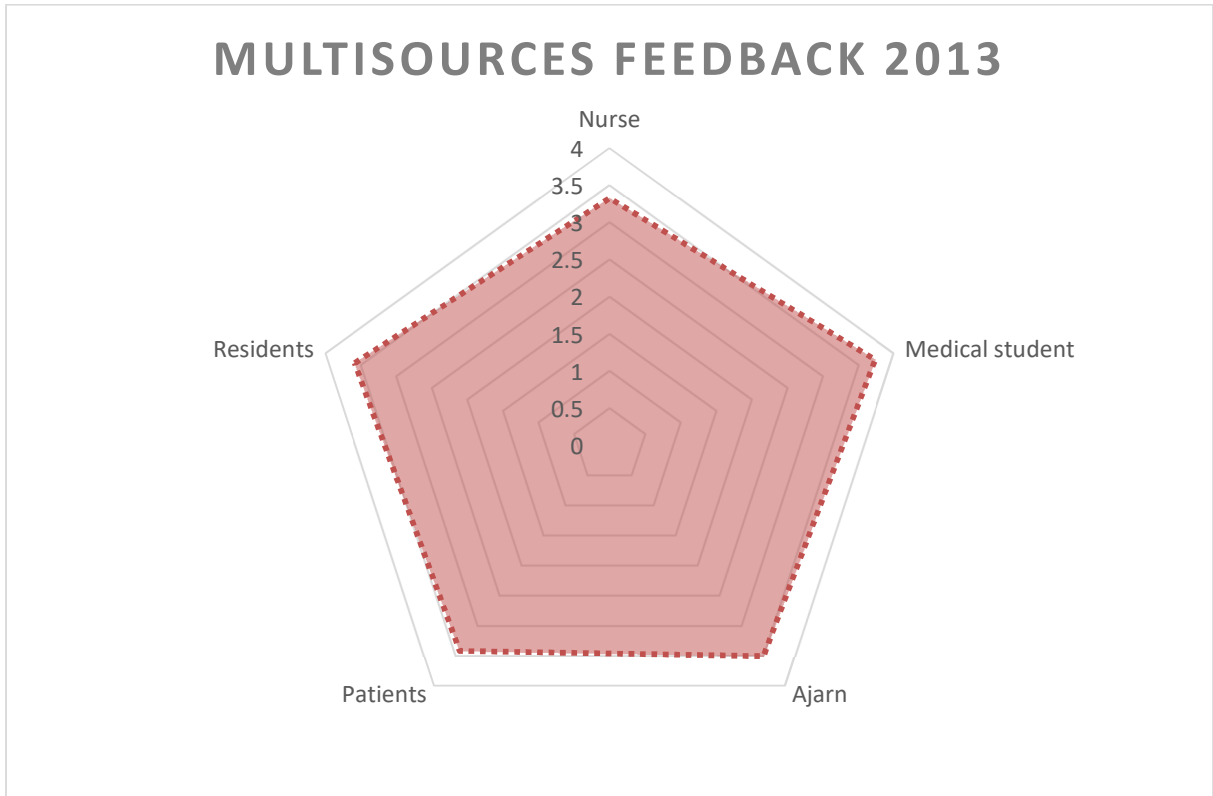
*ยังไม่มีผลการประเมินจากแพทย์ประจำบ้านและผู้รับบริการในปีการศึกษา 2554



ชั้นปีที่ 2 ปีการศึกษา 2555

Rotation	อาจารย์ (5 คะแนน)	นักศึกษาแพทย์ (5 คะแนน)	พยาบาล (5 คะแนน)	แพทย์ ประจำบ้าน	ผู้รับบริการ
นรีเวช 1	3.93	4.00	4.53		
เลิดสิน	4.67				
พระศรีฯ 13/1	4.35		4.61		
LR ดึก			4.00		
Onco	4.17	4.20	3.23		
พระศรีฯ 14/2			5.00		
นรีเวช 2	4.11	4.50	5.00		
สระบุรี	4.47				
พระศรีฯ 13/2	4.40		4.70		
พระศรีฯ 10/1		4.70	4.50		
พระศรีฯ 14/1	4.00		4.23		
LR เช้า		5.00	4.69		
พระศรีฯ 10/3		5.00	4.56		
คะแนนเฉลี่ย	4.26	4.56	4.45		

*ยังไม่มีผลการประเมินจากแพทย์ประจำบ้านและผู้รับบริการในปีการศึกษา 2555



ชั้นปีที่ 3 ปีการศึกษา 2556

Rotation	อาจารย์ (4 คะแนน)	แพทย์ประจำบ้าน (4 คะแนน)	พยาบาล (4 คะแนน)	นักศึกษาแพทย์ (4 คะแนน)	ผู้รับบริการ (4 คะแนน)
นรีเวช 1	3.50	3.80	3.40	3.90	3.03
STD	3.70		3.20		
พระศรีฯ 10/1		2.62	4.00	3.70	3.36
LR พิเศษ		3.90	3.08		
OPD GYN			2.90		3.40
Septic		3.75	3.10	4.00	3.26
วิสัญญี	3.75				
นรีเวช 2	3.90	4.00	3.85	3.87	3.74
Infertile	3.20				
นครปฐม	3.00				
OPD ANC			3.75		3.73
ONCO	3.60	3.81	3.02		
LR เข้า		3.25	3.08	3.50	
Surgery	3.47				
คะแนนเฉลี่ย	3.51	3.59	3.33	3.74	3.42

*เริ่มการประเมินจากนักศึกษาแพทย์และผู้รับบริการ ในปีการศึกษา 2556

ผลลัพธ์การปฏิบัติงานของ



แพทย์หญิง Y

อาจารย์ที่ปรึกษา อาจารย์ B

ตามการประเมินด้วยแฟ้มสะสมพัฒนาการ (Portfolio)

ปีการศึกษา 2554-2556

Competency based portfolio assessment

Academic year 2011-2013

สาส์นจากหัวหน้าภาควิชา

ภาควิชาสูติศาสตร์-นรีเวชวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ขอแสดงความยินดีกับ **แพทย์หญิง B** ที่สำเร็จการฝึกอบรมแพทย์ประจำบ้าน สาขาสูติศาสตร์-นรีเวชวิทยา ระหว่างปีการศึกษา 2553-2555

ตลอดระยะเวลาสามปีที่ผ่านมา ภาควิชาฯ ได้ดำเนินการประเมินคุณสมบัติด้านต่างๆ ของท่าน ได้แก่ ความรู้ ทักษะหัตถการ การวิจัย และพฤติกรรมการทำงาน ในรูปแบบ Portfolio ดังผลสรุปในเอกสารฉบับนี้

ภาควิชาฯ ขออัญวยพรให้ท่านประสบความสำเร็จในการดำเนินชีวิตครอบครัว และหน้าที่การงาน ตลอดไป

ศาสตราจารย์คลินิก นายแพทย์ชาญชัย วันทนาศิริ

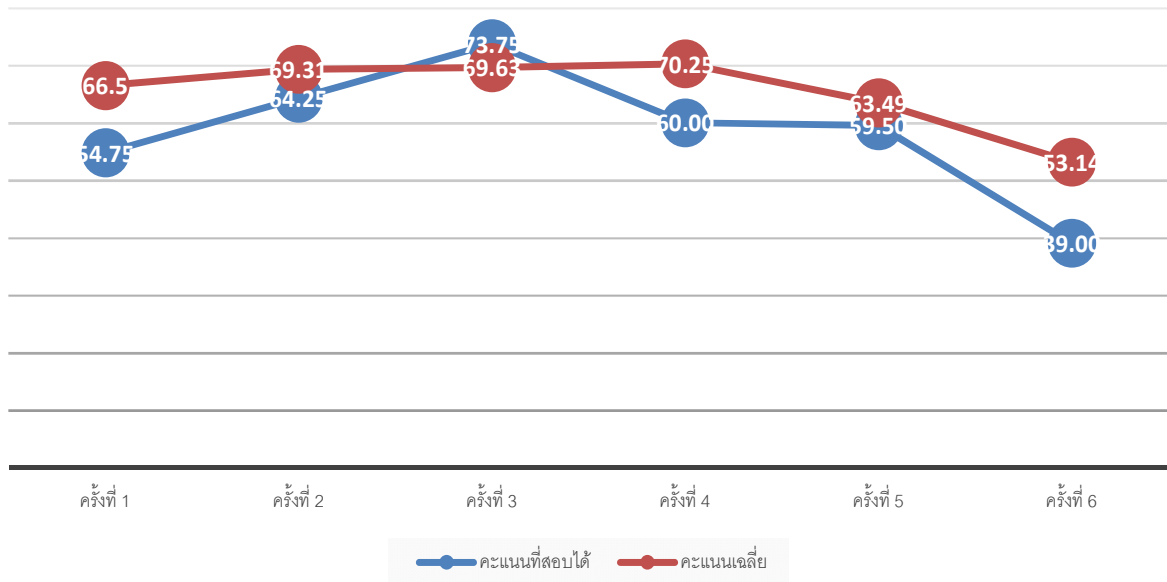
หัวหน้าภาควิชาสูติศาสตร์-นรีเวชวิทยา

คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

การประเมินความรู้ทางสูติศาสตร์-นรีเวชวิทยา

(Knowledge assessment)

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 1

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	66.50	54.75	13
2	100	69.31	64.25	11
3	100	69.63	73.75	4
4	100	70.25	60.00	13
5	100	63.49	59.50	11
6	100	53.14	39.00	13

ผลการสอบตามหลักสูตรประกาศนียบัตรบัณฑิตชั้นสูงสาขาวิทยาศาสตร์การแพทย์คลินิก:

The Higher Graduate Diploma (Clinical Medical Sciences) คณะแพทยศาสตร์ศิริราชพยาบาล

ผ่าน ได้รับประกาศนียบัตรเมื่อ 25 พฤษภาคม 2555

ไม่ผ่าน

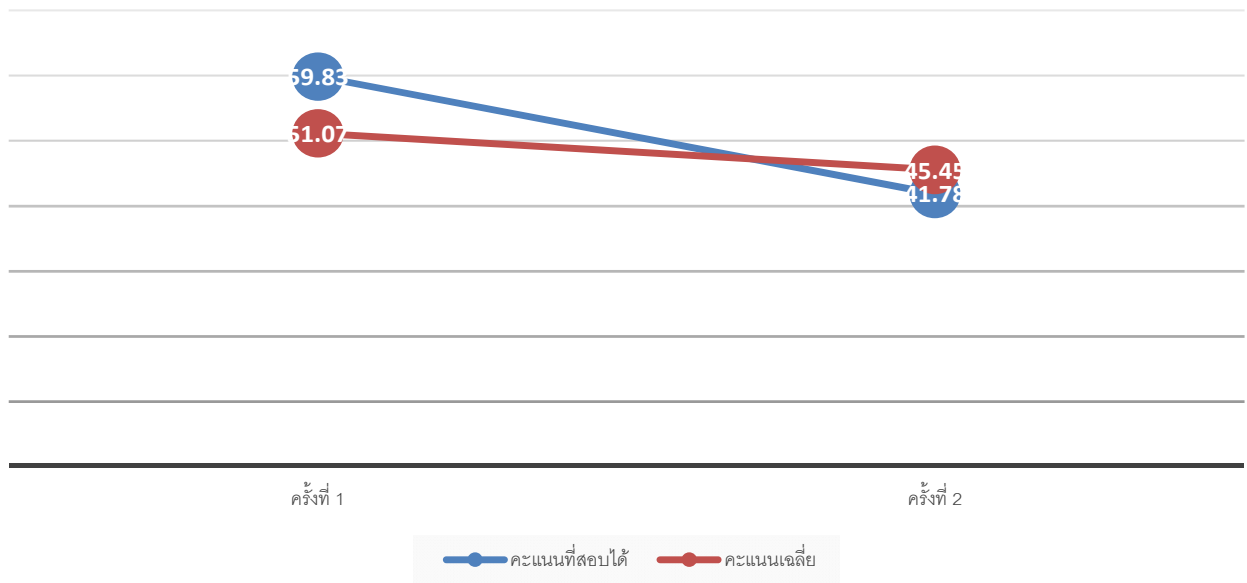
การสอบวิชาภาษาอังกฤษและกฎหมายทางการแพทย์ครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบวิชาภาษาอังกฤษและกฎหมายทางการแพทย์ครั้งที่ 2 (กรณีการสอบครั้งที่ 1 ไม่ผ่าน)

ผ่าน ไม่ผ่าน

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 2

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	51.07	59.83	4
2	100	45.45	41.78	12

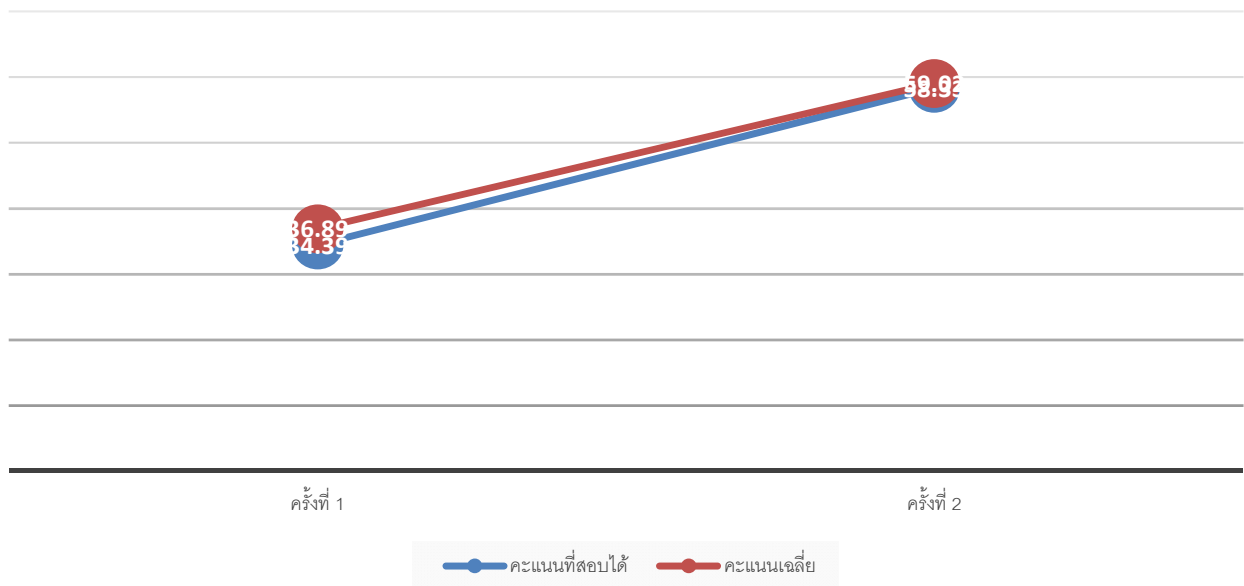
การสอบ OSLER ในสถาบัน ครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบ Basic science ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย

ผ่าน ไม่ผ่าน

ผลการสอบวัดระดับความรู้ทางวิชาการ



ชั้นปีที่ 3

การสอบครั้งที่	คะแนนรวม	คะแนนเฉลี่ย	คะแนนที่สอบได้	ลำดับที่
1	100	36.89	34.39	10
2	100	59.02	58.33	10

การสอบ OSLER ในสถาบัน ครั้งที่ 2

ผ่าน ไม่ผ่าน

การสอบ OSLER ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทยครั้งที่ 1

ผ่าน ไม่ผ่าน

การสอบ OSLER ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทยครั้งที่ 2 (กรณีสอบครั้งแรกไม่ผ่าน)

ผ่าน ไม่ผ่าน

การสอบงานวิจัย ราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย

ผ่าน ไม่ผ่าน

หัตถการสำคัญทางสูติศาสตร์-นรีเวชวิทยาที่ปฏิบัติ
ขณะเป็นแพทย์ประจำบ้านชั้นปีที่ 3
 (Clinical skills assessment when being the 3rd year resident)

การผ่าตัดทางนรีเวช

การผ่าตัด	จำนวน
Total abdominal hysterectomy +/- bilateral salpingoophorectomy	14
Vaginal hysterectomy +/- AP repair	7
Adnexal surgery: Salpingectomy/Salpingotomy/Salpingostomy	4
Cervical conization	2

การผ่าตัดทางสูติศาสตร์

การผ่าตัด	จำนวน
Cesarean delivery	43
Tubal sterilization	1
Dilatation and curettage	5
Vacuum extraction/Forceps extraction	5
Breech assisting	
Manual removal of placenta	6

หมายเหตุ

จำนวนหัตถการเป็นจำนวนโดยประมาณ เนื่องจากอยู่ระหว่างกระบวนการพัฒนาและปรับปรุงระบบเก็บข้อมูลหัตถการ
 แพทย์ประจำบ้าน ภาควิชาสูติศาสตร์-นรีเวชวิทยา

การทำงานวิจัยระดับแพทย์ประจำบ้าน
(Research competency)

เรื่อง Prevalence of Abnormal Menstrual Patterns among Copper Intrauterine Devices (IUDs) Users in Women Attending Family Planning Clinic, Siriraj Hospital

อาจารย์ผู้ควบคุม ผู้ช่วยศาสตราจารย์นายแพทย์สุรศักดิ์ อังสุวัฒนา

ข้อมูลสำคัญสำหรับงานวิจัย

- ผ่าน SIRB เมื่อ 28 สิงหาคม 2555
เลขที่ 415/2555(EC3)
- ประกวดการนำเสนองานวิจัยในการประชุมราชวิทยาลัยสูตินรีแพทย์แห่งประเทศไทย
วันที่ 26 พฤศจิกายน 2556
 เข้าร่วมนำเสนอ ไม่ได้รับรางวัล
 เข้าร่วมนำเสนอ ได้รับรางวัล
- การตีพิมพ์ในวารสารวิชาการ
 ไม่ได้ตีพิมพ์
 ได้รับการตีพิมพ์ (ระบุรายละเอียดวารสาร)

ผลการประเมินเจตคติและพฤติกรรมการปฏิบัติงานของแพทย์ประจำบ้าน (Multisources feedback)

แพทย์ประจำบ้านจะได้รับการประเมินในประเด็นต่อไปนี้

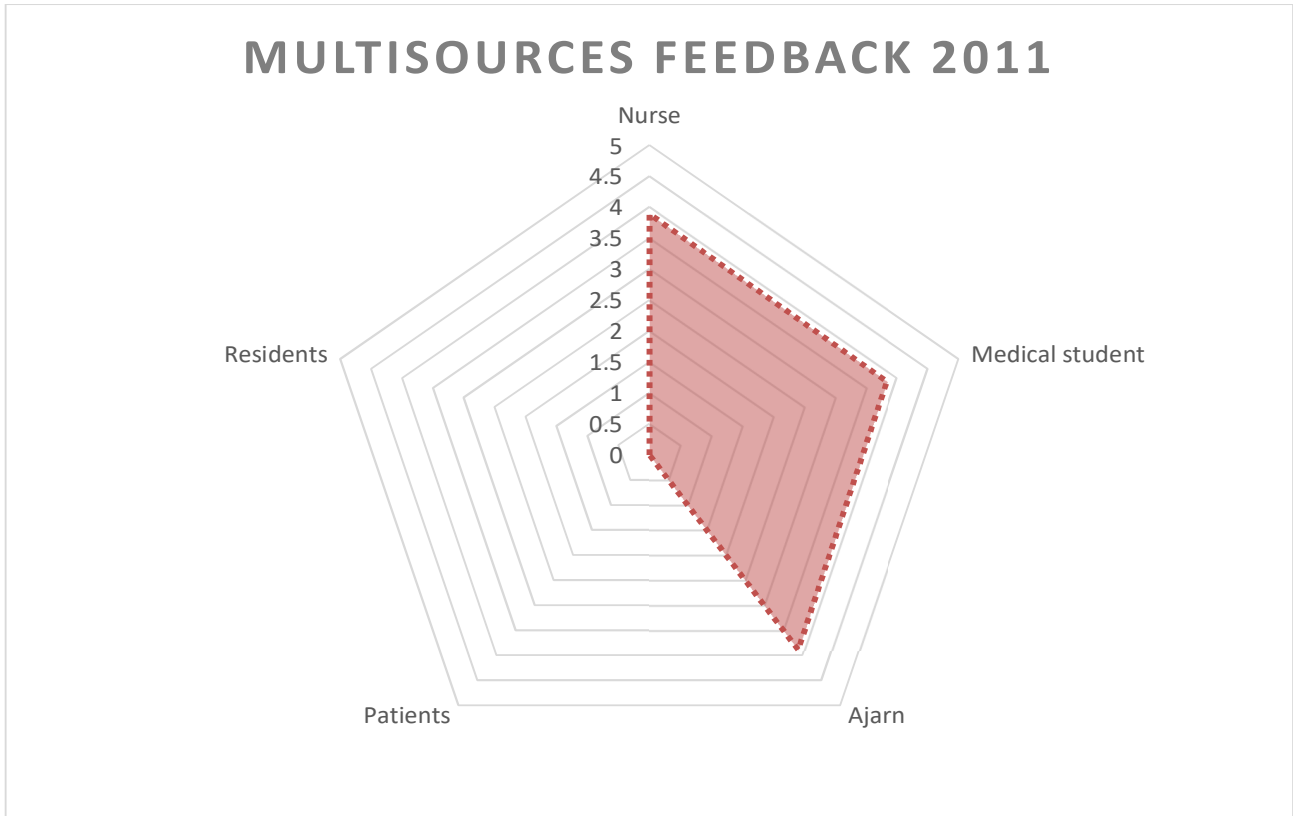
1. ความรู้ความสามารถด้านวิชาการ

2. ทักษะพื้นฐานในการปฏิบัติงาน

ได้แก่ ทักษะการสื่อสารกับเพื่อนร่วมงานและผู้ป่วย/ญาติ การบันทึกรายงานผู้ป่วย การทำงานร่วมกับผู้อื่น และบุคลิกภาพขณะปฏิบัติงาน

3. คุณธรรมและจริยธรรม

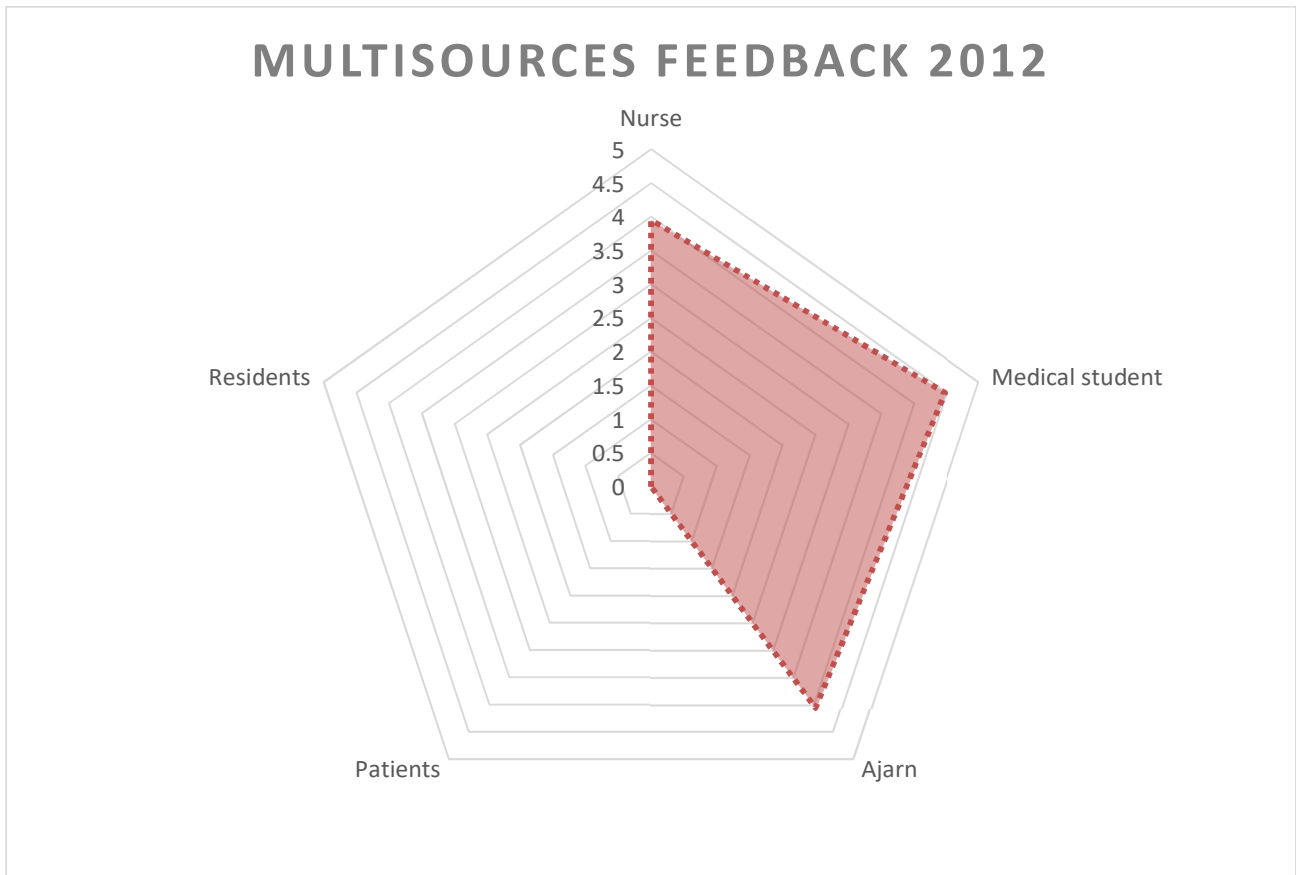
ได้แก่ ความรับผิดชอบ ความเสียสละ ความตรงต่อเวลา ความซื่อสัตย์ การปฏิบัติตามระเบียบข้อบังคับ และอภัย/น้ำใจ/ความเอื้อเฟื้อต่อผู้อื่น



ชั้นปีที่ 1 ปีการศึกษา 2554

Rotation	อาจารย์ (5 คะแนน)	นักศึกษาแพทย์ (5 คะแนน)	พยาบาล (5 คะแนน)	แพทย์ ประจำบ้าน	ผู้รับบริการ
พระศรีฯ 9/2			4.00		
LR เข้า		3.58	4.00		
LR พิเศษเข้า			3.90		
นรีเวช 1	3.50	3.40	4.10		
นรีเวช 1 (2)	4.00	3.41	3.92		
พระศรีฯ 10/2			3.92		
พระศรีฯ 9/1+ANC			4.03		
LR ดึก			3.76		
LR พิเศษบ่าย			3.23		
นรีเวช 2	4.20	3.17	5.00		
Onco	3.88	5.00	4.07		
พระศรีฯ 10/3		3.83	2.92		
พระศรีฯ 10/1		4.50	3.84		
คะแนนเฉลี่ย	3.89	3.84	3.89		

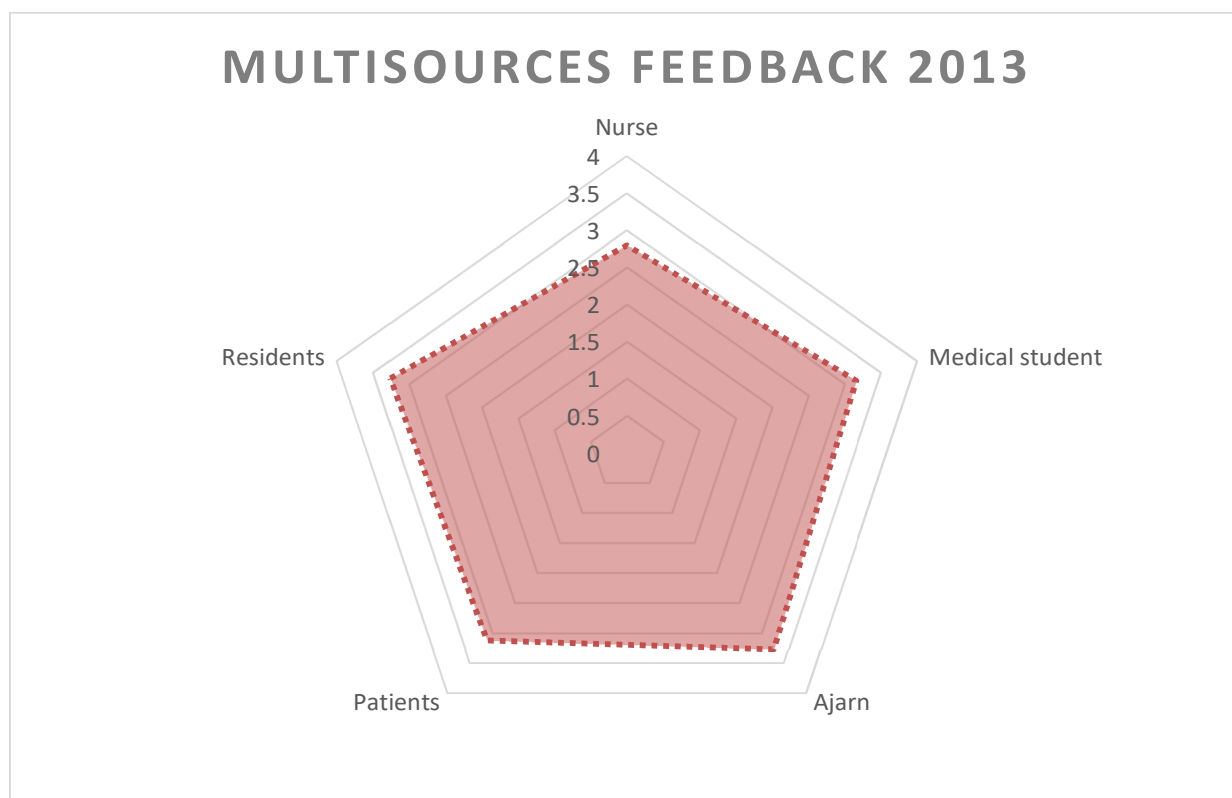
*ยังไม่มีผลการประเมินจากแพทย์ประจำบ้านและผู้รับบริการในปีการศึกษา 2554



ชั้นปีที่ 2 ปีการศึกษา 2555

Rotation	อาจารย์ (5 คะแนน)	นักศึกษาแพทย์ (5 คะแนน)	พยาบาล (5 คะแนน)	แพทย์ ประจำบ้าน	ผู้รับบริการ
นรีเวช 1	3.95	4.67	4.84		
เลิดสิน	4.00				
พระศรีฯ 13/1	4.00		3.69		
LR ดึก			4.07		
Onco	4.05	3.77	3.76		
พระศรีฯ 14/2			4.42		
นรีเวช 2	3.82	4.50	4.23		
พระศรีฯ 13/2	4.35		3.08		
พระศรีฯ 10/1		5.00	3.25		
พระศรีฯ 14/1	4.30		4.00		
LR เช้า		4.58	4.20		
พระศรีฯ 10/3		4.42	4.00		
คะแนนเฉลี่ย	4.06	4.49	3.95		

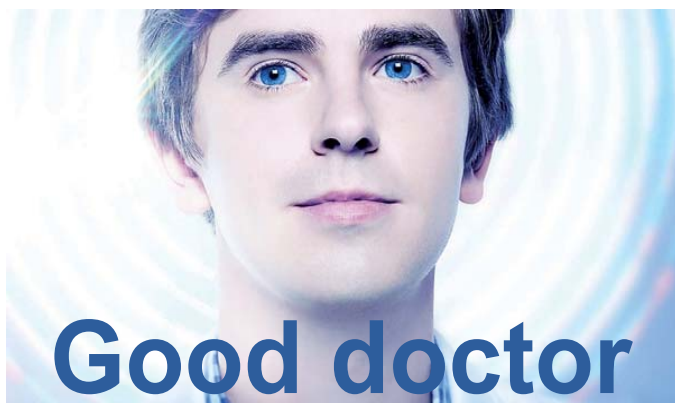
*ยังไม่มีผลการประเมินจากแพทย์ประจำบ้านและผู้รับบริการในปีการศึกษา 2555



ชั้นปีที่ 3 ปีการศึกษา 2556

Rotation	อาจารย์ (4 คะแนน)	แพทย์ประจำบ้าน (4 คะแนน)	พยาบาล (4 คะแนน)	นักศึกษาแพทย์ (4 คะแนน)	ผู้รับบริการ (4 คะแนน)
นรีเวช 1	3.30	3.40	2.29	3.12	3.28
STD	3.50		3.00		
พระศรีฯ 10/1		3.30	2.20	2.66	2.64
LR พิเศษ			3.06		
OPD GYN			3.40		3.07
Septic		3.50	3.00	3.40	3.33
วิสัญญี	2.70				
นรีเวช 2	3.40	3.48	3.13	3.44	3.40
Infertile	3.30				
นครปฐม	3.30				
OPD ANC			2.85		3.04
ONCO	3.05	2.80	2.21	3.75	
LR เข้า		3.10	2.95	2.63	
Surgery	3.62				
คะแนนเฉลี่ย	3.27	3.26	2.80	3.16	3.12

*เริ่มการประเมินจากนักศึกษาแพทย์และผู้รับบริการ ในปีการศึกษา 2556



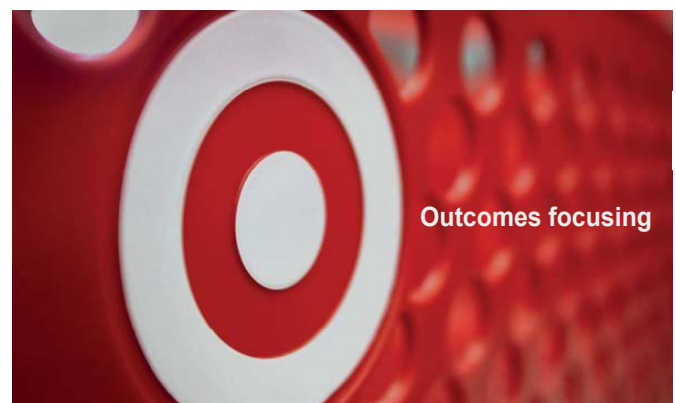
How could we assess all of their skills?



4 steps for portfolio development

Outcomes Evidences Reflection Assessment







- | | | |
|-----------------|----------------------|---------------------|
| Exam scores | Research skills | Presentation skills |
| Clinical skills | Multisource feedback | Social skills |
| Leadership | Communication | Medical record |
| Reflection | Teamwork | EPA & DOPS |



Learning without reflection is **waste**
Confucius



Formative **vs** Summative

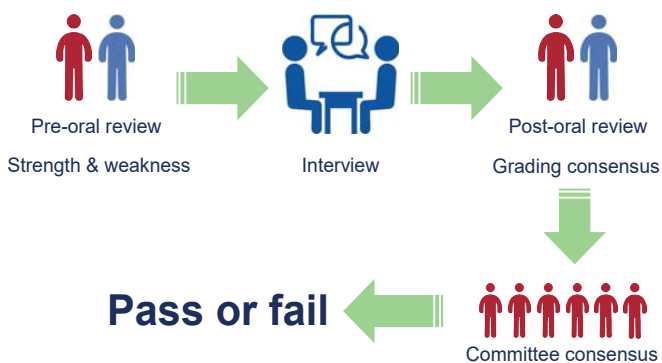
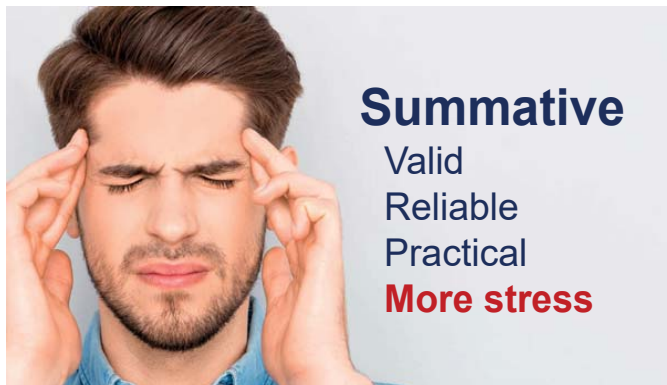
Formative

- Motivation
- Feedback
- Less stress
- Less corporation**



Summative

- Valid
- Reliable
- Practical
- More stress**



O utcomes
E vidences
R eflection
A sssessment

Teacher is a ...
GARDENER



Portfolio looks like
Shoes

► Question & Comments

ศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ (ศตว)
Siriraj Health science Education Excellence center (SHEE)

ฝ่ายการศึกษาก่อนปริญญา คณะแพทยศาสตร์ศิริราชพยาบาล
สำนักงาน: ตึกออดุลดะชวกรรม ชั้น 6 (ห้อง 656)
Tel. 02 419 9978, 02 419 96637 Fax. 02 412 3901



shee.si.mahidol.ac.th



shee.mahidol@gmail.com



mahidol.shee



SHEE FC



Siriraj Health science Education Excellence center