

เครื่องมือประเมินผลในการเรียนการสอน แบบใช้ปัญหาเป็นหลัก

บุญนาท ลายสนิทเสรีกุล*

Laisnitsarekul B. Measuring instruments for problem-based learning. Chula Med J 1988 Jun; 32(6): 553-566

This documentary research aimed to study measuring instruments for problem-based learning curriculum in foreign medical schools. The instruments were characterised by type, validity, reliability, feasibility, strength and weakness.

The measuring instruments used in problem-based curriculum were designed for integrated measurement. Many major characteristics and minor characteristics were measured including the content, process and product. The well-known instruments were Triple Jump Exercise, Objective Structured Clinical Examination (OSCE), Modified Essay Questions (MEQ), and Tutorial evaluation.

Reprint requests : Unit of Medical Education, Faculty of Medicine, Chulalongkorn University, Bangkok 10500, Thailand.

Received for publication. January 22, 1988.

หลักสูตรที่จัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก⁽¹⁾ เป็นหลักสูตรการศึกษาแบบบูรณาการ (Integration) ที่ผสมผสานเนื้อหาวิชา วิธีการเรียนการสอน และการประเมินผล โดยมุ่งหวังให้ผู้เรียนเกิดคุณลักษณะ ทั้งในด้านความรู้ ทักษะ เจตคติ ความสามารถในการแก้ปัญหา และความสามารถในการแสวงหาความรู้ด้วยตนเอง ควบคู่กันไป การจัดประสบการณ์การเรียนรู้ เน้นการเรียนรู้โดยใช้ปัญหาเป็นหลัก โดยสอดแทรกแนวคิดเชิงคุณธรรม และจริยธรรมของวิชาชีพ ไว้อย่างต่อเนื่อง เน้นการเรียนรู้ที่ผู้เรียนเป็นศูนย์กลาง (Student-centered)

เป็นที่ยอมรับกันว่า การประเมินผลเป็นเครื่องกำหนดพฤติกรรมการเรียนของผู้เรียน การที่จะใช้เครื่องมือประเมินผลในหลักสูตรแบบเก่า ที่แบ่งแยกเนื้อหาวิชาเป็นเรื่อง ๆ แบ่งแยกวัตถุประสงค์ออกเป็นด้านความรู้ ทักษะ และเจตคติ จึงไม่เหมาะสมและก่อให้เกิดปัญหาด้านการประเมินผลเป็นอย่างมาก ในการประเมินผลหลักสูตรที่จัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก โรงเรียนแพทย์หลายแห่งที่จัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก⁽²⁾ เช่น มหาวิทยาลัยแมคมาสเตอร์ ประเทศแคนาดา, มหาวิทยาลัยเซาเทิร์นอิลลินอยส์ ประเทศสหรัฐอเมริกา, มหาวิทยาลัยลิมเบอร์ก ประเทศเนเธอร์แลนด์, มหาวิทยาลัยนิวยอร์ก ประเทศสหรัฐอเมริกา, มหาวิทยาลัยนิวคาสเซิล ประเทศออสเตรเลีย เป็นต้น ได้พัฒนาเครื่องมือประเมินผลแบบบูรณาการ ใช้วัดคุณลักษณะของนิสิตแพทย์หลาย ๆ อย่างในเวลาเดียวกัน วัดทั้งในด้านเนื้อหา (Content) วิธีแสวงหาความรู้ (Process) และการนำเอาความรู้มาแก้ปัญหา (Product) เครื่องมือประเมินผลแบบบูรณาการที่มีใช้กันแพร่หลาย ได้แก่ Triple Jump Exercise⁽³⁾, Objective Structured Clinical Examination (OSCE)⁽⁴⁾, Modified Essay Question (MEQ)⁽⁵⁾, และ Tutorial evaluation⁽⁶⁾ ซึ่งชื่อเครื่องมือเหล่านี้เป็นชื่อที่ใหม่มากในวงการประเมินผลการศึกษา อย่างไรก็ตาม กิลเบิร์ต⁽⁷⁾ ได้ให้หลักในการประเมินผลการเรียนรู้ของผู้เรียนไว้ว่า การประเมินด้านความรู้ ทักษะและเจตคติ สามารถประเมินได้ทั้งทางตรงและทางอ้อม โดยด้านความรู้วิธีที่เหมาะสมที่สุดคือการประเมินทางอ้อมด้วยแบบทดสอบ และเสริมด้วยการประเมินจากการสังเกต ส่วนด้านทักษะและเจตคติ วิธีที่เหมาะสมที่สุดคือ การประเมินทางตรงด้วย

การสังเกต และเสริมด้วยการประเมินทางอ้อมด้วยแบบทดสอบ ดังแสดงรายละเอียดในแผนภูมิที่ 1

เครื่องมือประเมินผลแบบบูรณาการ ที่ใช้ในหลักสูตรที่จัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก เป็นเรื่องใหม่ในวงการประเมินผล จึงน่าสนใจที่จะศึกษาว่าเครื่องมือประเมินผลแต่ละชนิด ใช้วัดผู้เรียนด้านใดเป็นคุณลักษณะหลัก ด้านใดเป็นคุณลักษณะรอง ใช้วัดเนื้อหาความรู้ วิธีแสวงหาความรู้ ผลจากการนำความรู้มาแก้ปัญหา รวมกันไปหรือแยกเป็นส่วน ๆ คุณภาพของเครื่องมือในรูปความตรง ความเที่ยง ความสะดวกในการใช้ ตลอดจนจุดเด่น จุดด้อยของเครื่องมือ

วัตถุประสงค์การศึกษา

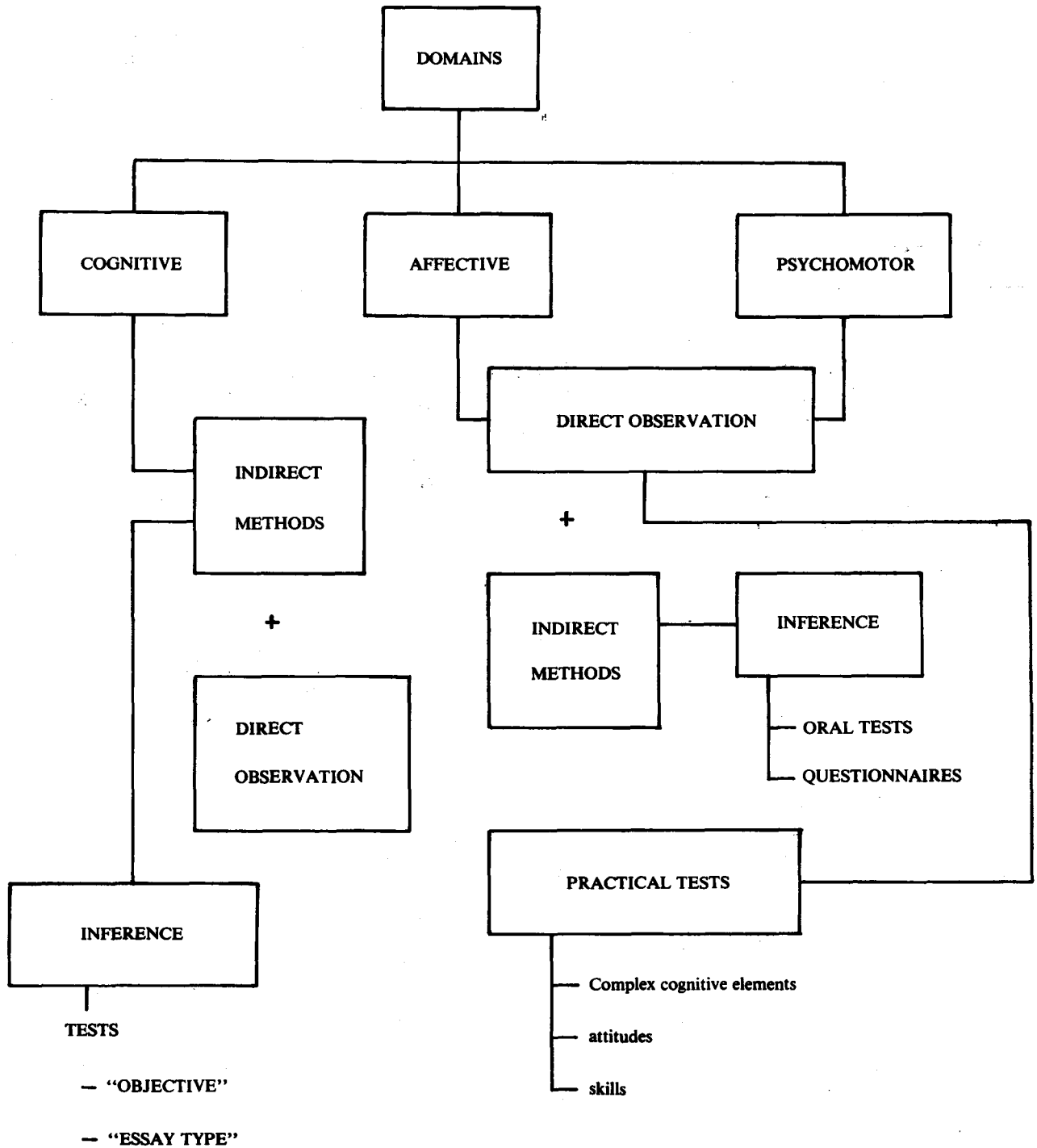
1. ศึกษาเครื่องมือประเมินผลนิสิต ที่มีใช้ในหลักสูตรที่มีการเรียนการสอนแบบใช้ปัญหาเป็นหลัก
2. ศึกษาจุดเด่น-จุดด้อย ของเครื่องมือแต่ละชนิด
3. ศึกษาคุณภาพของเครื่องมือประเมินผลนิสิตแต่ละชนิด ในรูปความตรง (Validity) ความเที่ยง (Reliability) ความสะดวกในการใช้ (Feasibility)

วัสดุและวิธีการ

การวิจัยนี้เป็นการวิจัยเอกสาร (Documentary research) ที่เกี่ยวกับเครื่องมือประเมินผลนิสิตแพทย์ ในโรงเรียนแพทย์ต่างประเทศ ที่มีหลักสูตรการเรียนการสอนแบบใช้ปัญหาเป็นหลัก โดยมีรายละเอียดดังนี้

1. วัสดุ ได้แก่
 - 1.1 หลักสูตรโรงเรียนแพทย์ต่างประเทศ ที่มีการจัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก
 - 1.2 หนังสือต่างประเทศ ที่เขียนเกี่ยวกับการเรียนการสอนแบบใช้ปัญหาเป็นหลัก
 - 1.3 วารสารวิจัยต่างประเทศ ที่เกี่ยวกับการประเมินผลนิสิตแพทย์ ที่เรียนแบบใช้ปัญหาเป็นหลัก
2. วิธีการ ดำเนินการวิเคราะห์ข้อมูล จากเอกสารต่าง ๆ ข้างต้น ตามหัวข้อต่อไปนี้
 - 2.1 ชนิดของเครื่องมือ
 - 2.2 วัตถุประสงค์ของเครื่องมือ
 - 2.3 คุณภาพของเครื่องมือ (ความตรง ความเที่ยง ความสะดวกในการใช้)
 - 2.4 จุดเด่น-จุดด้อย ของเครื่องมือ

Figure 1 Learning objectives and methods of evaluation



ผลการศึกษา

เครื่องมือประเมินผลนิสิตที่ใช้ในหลักสูตรที่มีการจัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก มีอยู่เป็นจำนวนมาก แต่ที่ใช้กันแพร่หลายในหลายโรงเรียนแพทย์ มีอยู่ 4 ชนิด ได้แก่

1. Triple Jump Exercise

เป็นเครื่องมือประเมินผลที่สร้างโดยมหาวิทยาลัยแมคมาสเตอร์ ประเทศแคนาดา มีลักษณะเป็นการสอบปากเปล่า

ที่มีโครงสร้างแบ่งออกเป็น 3 ชั้น⁽⁸⁾ โดยชั้นแรกจะประเมินความสามารถในการตั้งสมมติฐาน การรวบรวมข้อมูล การระบุปัญหาที่น่าจะเป็น การกำหนดหัวข้อความรู้ที่จะไปแสวงหาเพิ่มเติม ชั้นที่สองจะประเมินความสามารถในการแสวงหาความรู้ด้วยตนเอง และชั้นที่ 3 เป็นการประเมินความสามารถในการสังเคราะห์ข้อมูล การระบุปัญหาที่แท้จริง และการประเมินผลความสามารถของตนเอง ดังแสดงรายละเอียดในตารางที่ 1

Table 1 Triple Jump Exercise.

STEPS	ASSESSMENT ITEM
I. PROBLEM EXPLORATION	1. HYPOTHESIS GENERATION 2. DATA GATHERING AND INTERPRETATION 3. KNOWLEDGE (CURRENT) 4. INTERIM PROBLEM FORMULATION 5. ISSUE IDENTIFICATION
II. INFORMATION SEARCH	
III. SYNTHESIS	6. INFORMATION SEARCH 7. ISSUE SYNTHESIS (CUMULATIVE KNOWLEDGE) 8. PROBLEM FORMULATION (FINAL) 9. SELF ASSESSMENT ABILITY

1.1 วัตถุประสงค์ เครื่องมือนี้วัดคุณลักษณะหลัก ได้แก่ ความรู้ กระบวนการแก้ปัญหา การแสวงหาความรู้ด้วยตนเอง และการประเมินตนเอง ส่วนคุณลักษณะรอง ได้แก่ กระบวนการใช้เหตุผลทางคลินิก นอกจากนี้ Triple Jump Exercise ยังวัดทั้งในด้านเนื้อหา กระบวนการแสวงหาความรู้ และการนำความรู้ไปแก้ปัญหา

1.2 จุดเด่น Triple Jump Exercise ให้ประโยชน์หลายประการในการประเมินนิสิต ได้แก่ สามารถประเมินคุณลักษณะหลายอย่างในการประเมินเพียงครั้งเดียว ให้ข้อมูลป้อนกลับโดยตรงแก่นิสิตในขณะที่ประเมิน มีความสะดวกในการใช้เนื่องจากสามารถใช้ได้ในทุกสถานการณ์ ทุกระดับชั้นปีของนิสิต ปัญหาที่นำมาประเมินอาจใช้ปัญหาในกระดานชนวนใช้สมมติ คนใช้จริง เป็นต้น

1.3 จุดอ่อน เครื่องมือนี้จะต้องใช้ผู้ประเมินที่มีความสามารถสูง หรือเป็นผู้ที่ผ่านการอบรมมาเป็นอย่างดี การประเมินแต่ละครั้งต้องใช้ผู้ประเมินจำนวนมากโดยผู้

ประเมินหนึ่งคนต่อนิสิตหนึ่งคน และผู้ประเมินต้องใช้เวลาอย่างน้อย 1 ชั่วโมงต่อนิสิตหนึ่งคน

1.4 คุณภาพ จากการศึกษาของเพนวินในปี 1978⁽⁹⁾ และเพอร์เลส ในปี 1981⁽¹⁰⁾ ในด้านความตรง มีการใช้วิธีให้ผู้สังเกตประเมินเครื่องมือว่า เป็นที่พอใจหรือไม่ (Satisfactory inter-observer agreement) ซึ่งทางการศึกษา เรียกว่า ความตรงผิวเผิน⁽¹¹⁾ (Face validity) ส่วนค่าความเที่ยง พบว่ามีค่าต่ำเนื่องจากรูปแบบการเขียนเครื่องมือยังขาดมาตรฐาน

2. The Objective Structured Clinical Examination (OSCE)

เป็นวิธีการประเมินผลที่พัฒนาโดยมหาวิทยาลัยดันดี ประเทศสกอตแลนด์⁽¹²⁾ มีลักษณะเป็นการสอบพฤติกรรมคลินิกที่สถานีย่อยแต่ละสถานี ใช้เวลาสถานีละ 4-5 นาที แล้วจะมีสัญญาณให้เปลี่ยนไปสถานีถัดไป สถานีแบ่งได้เป็น 2 ชนิด ได้แก่ สถานีที่ประเมินหัตถการ เช่น

การตรวจร่างกาย การซักประวัติผู้ป่วย การอ่านฟิล์มเอ็กซเรย์ เป็นต้น และสถานีคำถาม ซึ่งจะให้นิสิตรอบคําถามจากเรื่องราวในสถานีหัตถการ ดังแสดงรายละเอียดในแผนภาพที่ 2 และ 3 ตามลำดับ จำนวนสถานีที่ใช้ประเมินประมาณ 20

สถานี เครื่องมือที่ใช้ประเมินสถานีหัตถการ คือแบบเลือกตรวจ (Check list) และแบบมาตราประมาณค่า (Rating scale) ส่วนสถานีคำถามใช้ข้อสอบปรนัย

Figure 2 Stations in the Objective Structured Clinical Examination are of two types, as shown.

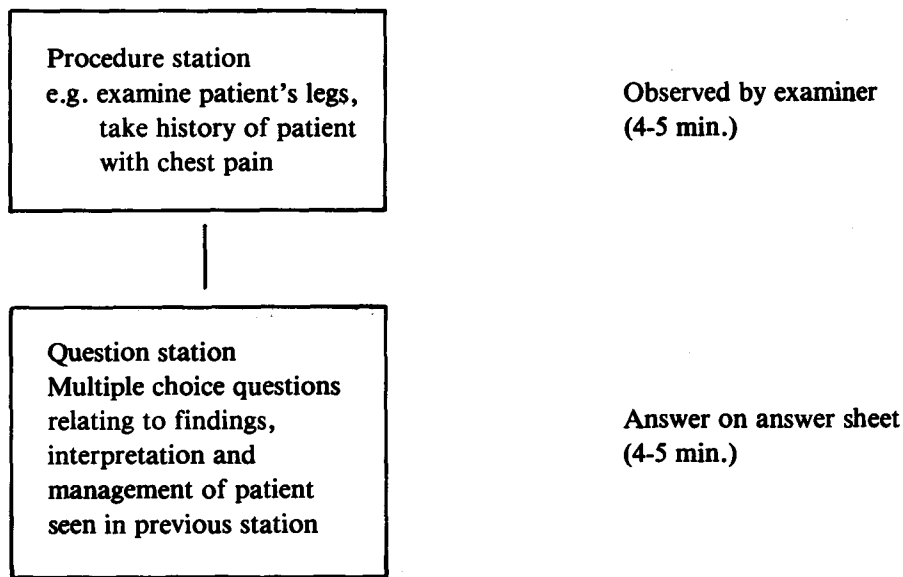
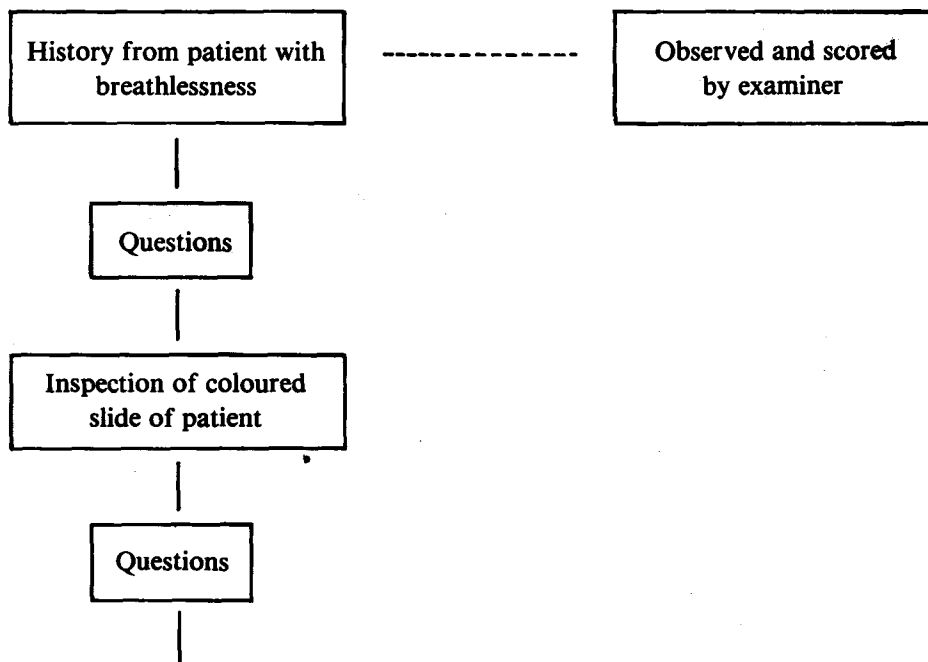
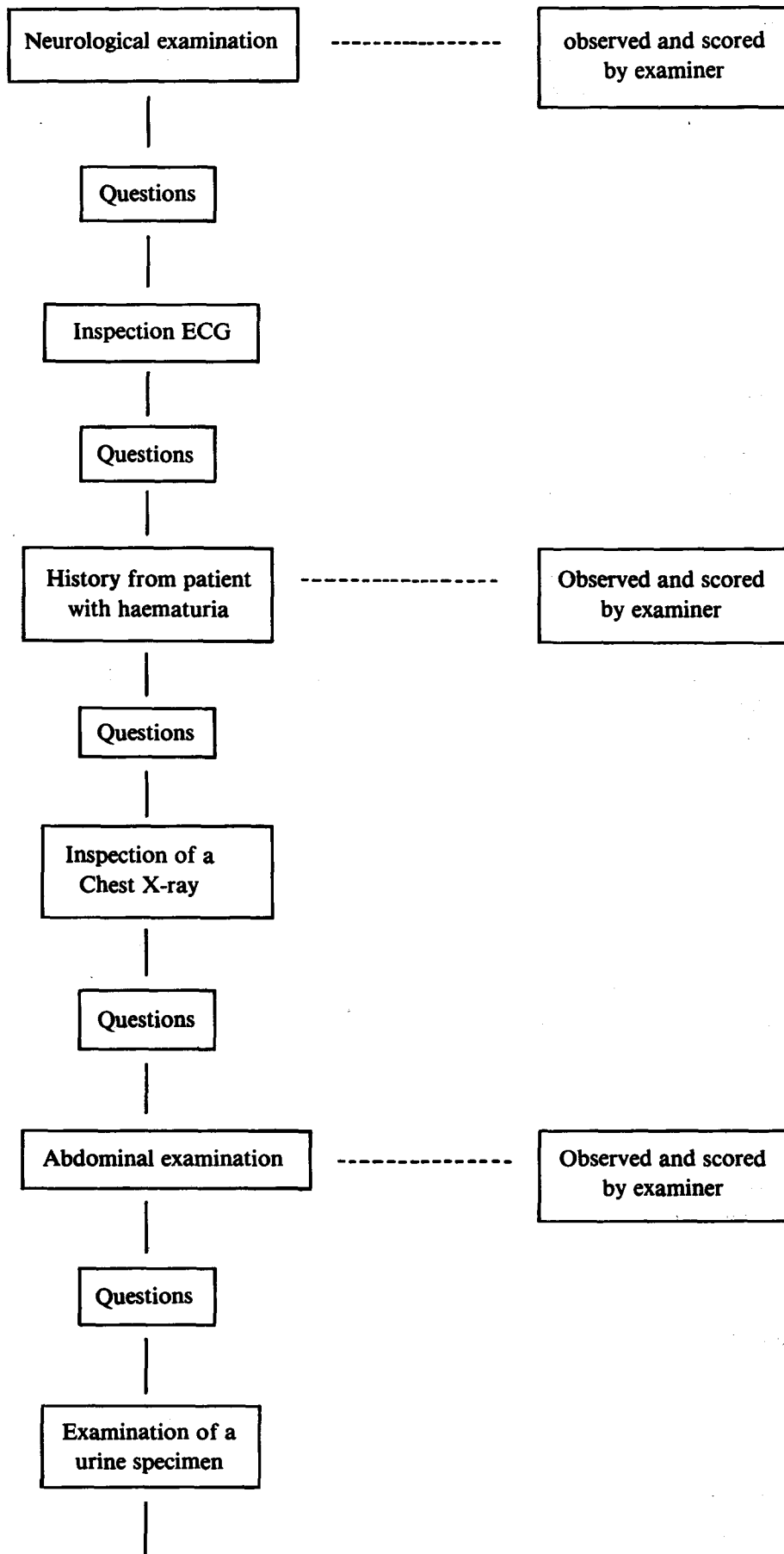
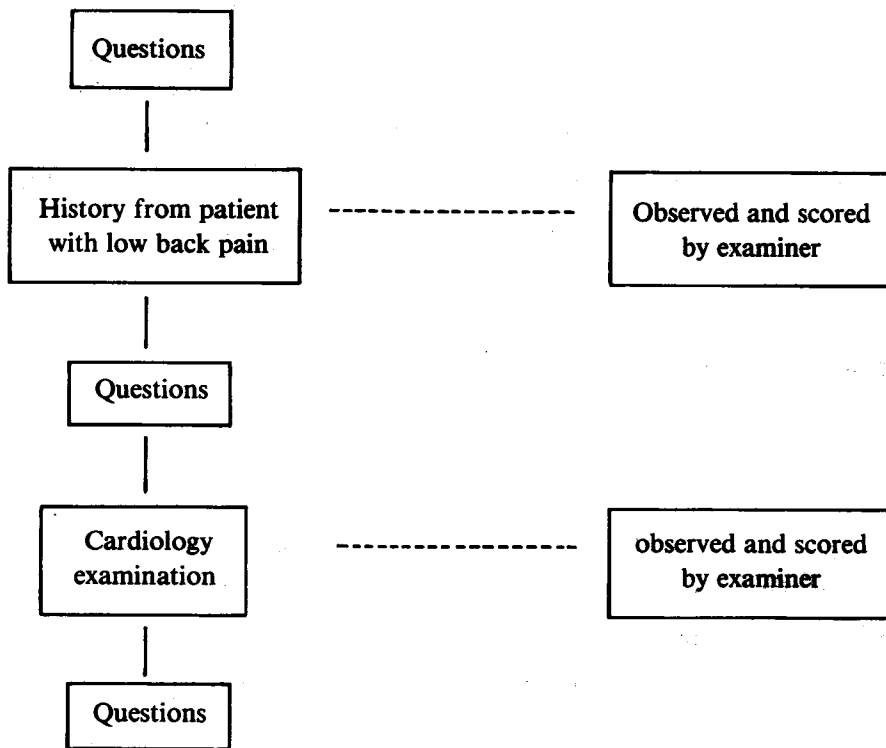


Figure 3 Complete structured clinical examination using twenty stations.







2.1 วัตถุประสงค์ วิธีการนี้วัดคุณลักษณะหลัก ได้แก่ พฤติกรรมคลินิก ซึ่งโอ๊คและคณะ⁽¹³⁾ หมายถึงการประเมินบุคลิกลักษณะ ท่าทาง วุฒิภาวะ การให้ความร่วมมือ ความสามารถทางวิชาการ วิริยะอดทนของนักเรียน ความสามารถทางวิชาชีพ ความสนใจให้บริการ เขียนและส่งรายงานตามเกณฑ์ที่กำหนด ความรับผิดชอบต่องานที่ได้รับมอบหมาย และมนุษยสัมพันธ์กับผู้อื่น พรินเทินและคณะ⁽¹⁴⁾ หมายถึงการประเมินการตัดสินใจเพื่อแก้ปัญหา การรวบรวมข้อมูล การเลือกตรวจทางห้องปฏิบัติการที่เหมาะสม ความรู้พื้นฐานเกี่ยวกับโรค มนุษยสัมพันธ์ต่อผู้ป่วย การวางแผนการรักษาผู้ป่วย ความเข้าใจเกี่ยวกับอุปสรรคในการดูแลผู้ป่วย มนุษยสัมพันธ์ต่อเพื่อนร่วมงาน ความพึงพอใจในวิชาชีพ มนุษยสัมพันธ์ต่อบุคลากรสาธารณสุขอื่น และทัศนคติที่มีต่อห้องผ่าตัด ฮอลส์ลอคและคณะ⁽¹⁵⁾ หมายถึงการประเมินความรู้พื้นฐาน ทักษะทางการแพทย์ การแก้ปัญหา มาตรฐานทางวิชาชีพและความเอื้ออารีต่อผู้ป่วย วิลสันบีและคณะ⁽¹⁶⁾ หมายถึงการประเมินทัศนคติต่อวิชาชีพ ความสัมพันธ์กับเพื่อนร่วมงาน ความเชื่อมั่นในตนเอง ความรู้ทางวิชาการแพทย์ ความคิดรวบยอด ทักษะ วุฒิภาวะ ความเอื้ออารีต่อผู้ป่วย ความมีไหวพริบ การปฏิบัติตามกฎข้อบังคับ และความซื่อสัตย์ อนุญาทและคณะ⁽¹⁷⁾ หมายถึงการประเมินความสามารถด้าน

การซักประวัติ การตรวจร่างกาย การวินิจฉัยและการรักษา ทัศนคติทางการแพทย์ ความรับผิดชอบ ความสัมพันธ์กับผู้ป่วย ความสัมพันธ์กับผู้ร่วมงาน คุณลักษณะสำหรับผู้จะเป็นแพทย์ และความสนใจในการศึกษาหาความรู้เพิ่มเติม ส่วนคุณลักษณะรอง ได้แก่ การประเมินความรู้ นอกจากนี้ OSCE ยังวัดทั้งในด้านเนื้อหาความรู้ ขั้นตอนการทำทักษะ ผลของการทำทักษะหรือการนำความรู้ไปแก้ปัญหา

2.2 จุดเด่น Objective Structured Clinical Examination มีประโยชน์หลายประการในการประเมินผล นิสิต ได้แก่ ผู้ประเมินสามารถควบคุมทั้งเนื้อหาและความยาก ให้เหมาะสมกับนิสิตแต่ละชั้นปี ประเมินทัศนคติทางการแพทย์ ได้หลายชนิดขึ้นอยู่กับจำนวนสถานี่ ใช้ได้กับนิสิตจำนวนมาก ๆ ในระยะเวลาสั้น ๆ ใช้เป็นการประเมินความก้าวหน้าเพื่อให้ข้อมูลป้อนกลับแก่นิสิต⁽¹⁸⁾ หรือการประเมินผลรวม ได้ทั้งสองแบบ

2.3 จุดด้อย วิธีการนี้มุ่งประเมินความรู้และทักษะของนิสิตเป็นส่วน ๆ ไม่ได้ประเมินความสามารถทั้งหมดของนิสิตในการดูแลรักษาผู้ป่วย การปฏิบัติงานบนหอพักผู้ป่วย นอกจากนี้ OSCE ใช้กำลังคนเป็นจำนวนมากทั้งอาจารย์ผู้ประเมินและผู้ป่วยที่นำมาให้นิสิตตรวจ อาจารย์ผู้ประเมินต้องสังเกตอย่างใกล้ชิด และประเมินนิสิตซ้ำ ๆ กัน

เท่ากับจำนวนนิสิตที่เข้าสอบ ผู้ป่วยต้องถูกตรวจซ้ำแล้วซ้ำอีก เท่ากับจำนวนที่นิสิตเข้าสอบ การเตรียมสถานที่ใช้เวลามากกว่า การสอบแบบเก่า การประเมินแบบนี้ต้องการสถานที่ที่ใหญ่พอสมควร ต้องการความช่วยเหลือจากคณาจารย์ภายในคณะ ในการเตรียมสถานที่ การเข้าร่วมเป็นอาจารย์ผู้ประเมิน ต้องมีการอบรมให้ความรู้แก่ผู้จะทำหน้าที่เป็นผู้ประเมิน และต้องมีการทดสอบเครื่องมือก่อนการใช้งานจริง

2.4 คุณภาพ จากการศึกษาของคณะแพทยศาสตร์ มหาวิทยาลัยแมคมาสเตอร์พบว่า ค่าความเที่ยงและความตรงของ OSCE ในการประเมินหัตถการทางการแพทย์ได้ค่าต่ำกว่าค่าความเที่ยงของสถานีคำถามเท่ากับ 0.63 ค่าความตรงตามเกณฑ์ซึ่งศึกษาโดยวิธีหาความสัมพันธ์ระหว่างคะแนนที่ได้จากคะแนนพฤติกรรมคลินิกเท่ากับ 0.63⁽¹⁹⁾ จากการศึกษาของคณะพยาบาลศาสตร์ มหาวิทยาลัยแมคมาสเตอร์พบว่า ค่าความเที่ยงจากผู้ประเมินในการประเมินนักศึกษาพยาบาลปีที่ 1 โดยเฉลี่ยเท่ากับ 0.72 แต่คะแนนจากการประเมินแบบ OSCE ไม่สัมพันธ์กับคะแนนที่ได้จากการปฏิบัติงานในห้องปฏิบัติการ และอันดับคะแนนทางคลินิก⁽²⁰⁾ อย่างไรก็ดีตามในปี 1985⁽²¹⁾ ค่าความเที่ยงจากผู้ประเมินที่ประเมินนักศึกษาพยาบาลปีที่ 3 มีค่าเท่ากับ 0.66 และ 0.74 และค่าความเที่ยงของสถานีคำถามเท่ากับ 0.62 และ 0.99 นอกจากนี้คะแนนที่ได้จากการประเมินแบบ OSCE ยังสัมพันธ์กับอันดับคะแนนทางคลินิกอย่างมีนัยสำคัญทางสถิติ แวน-เนี่ยคเกิลและคณะ⁽²²⁾ รายงานว่า OSCE มีความตรงเชิงทำนายสูงเมื่อศึกษาหาความสัมพันธ์กับอันดับคะแนนทางคลินิก อติเมย์-โคโร⁽²³⁾ รายงานว่า การประเมินผลแบบ OSCE มีความตรง ความเที่ยงและความสะดวกในการใช้ นายน่าและคณะ⁽²⁴⁾ รายงานว่า OSCE มีค่าความเที่ยงเท่ากับ 0.81 และ 0.78 และมีความตรงเชิงทำนายเท่ากับ 0.11 ถึง 0.60

3. Modified Essay Questions (MEQ)

เป็นเครื่องมือประเมินผลที่พัฒนาโดยราชวิทยาลัยเวชปฏิบัติทั่วไปในปี 1971⁽²⁵⁾ และน็อค⁽²⁶⁾ ในปี 1972 และคณะแพทยศาสตร์ มหาวิทยาลัยนิวคาสเซิล ประเทศออสเตรเลีย⁽²⁷⁾ ใช้เป็นเครื่องมือประเมินผลหลักชนิดหนึ่งในการประเมินผลนิสิต MEQ มีลักษณะเป็นข้อสอบแบบข้อเขียน นำเสนอด้วยปัญหาทางคลินิกแล้วตั้งคำถามให้ตอบสั้น ๆ คำถามจะเรียงลำดับขั้นตอน เริ่มตั้งแต่การตั้งสมมติฐาน การรวบรวมข้อมูล การทดสอบสมมติฐานและการแก้ปัญหา นิสิตจะต้องใช้ความรู้และข้อมูลที่กำหนดให้มาแก้ปัญหาตามคำถามและเวลาที่กำหนดให้ ดังตัวอย่างที่แสดงในแผนภาพที่ 4

3.1 วัตถุประสงค์ เครื่องมือนี้วัดคุณลักษณะหลัก ได้แก่ เหตุผลทางคลินิกและการแก้ปัญหา วัดคุณลักษณะรอง ได้แก่ ความรู้ นอกจากนี้ยังวัดในด้านเนื้อหาวิชา กระบวนการแก้ปัญหา และผลของการนำความรู้ไปแก้ปัญหา

3.2 จุดเด่น ข้อสอบแบบ MEQ สร้างง่ายคล้ายกับการสร้างข้อสอบแบบอัตนัย ใช้ประเมินนิสิตได้ทั้งระดับก่อนปริญญา หลังปริญญา และแพทย์ประจำบ้าน เนื้อหาของข้อสอบสามารถผสมผสานหลาย ๆ สาขาวิชาเข้าด้วยกัน ใช้เป็นเครื่องมือประเมินได้ทั้งแบบการประเมินความก้าวหน้าและประเมินผลรวม นอกจากนี้ยังใช้ประเมินนิสิตได้ครั้งละจำนวนมาก

3.3 จุดด้อย MEQ มีความยุ่งยากในการกำหนดเกณฑ์การให้คะแนน ครอบคลุมเนื้อหาได้ไม่กว้างขวางเท่าข้อสอบแบบปรนัย การสอบไม่อนุญาตให้นิสิตกลับมาแก้ไขคำตอบในข้ออื่น ๆ เมื่อได้ทำผ่านมาแล้ว ใช้เวลาตรวจข้อสอบค่อนข้างมาก⁽²⁸⁾ นิสิตไม่ทราบว่าต้องตอบลึกซึ้งเพียงใด⁽²⁹⁾

3.4 คุณภาพ พิเลทติ⁽³⁰⁾ ได้คำนวณหาค่าความเที่ยงด้วยวิธี Coefficient alpha ของครอนบาชพบว่า มีค่าระหว่าง 0.57 ถึง 0.91 และหาค่าความตรงตามโครงสร้างพบว่า มีค่าระหว่าง 0.34 ถึง 0.87 สแตรทฟอร์ดและคณะ⁽³¹⁾ รายงานว่าค่าความเที่ยงของ MEQ ที่มีจำนวนคำถาม 25 คำถามมีค่าเท่ากับ 0.63 และถ้า MEQ มีเกณฑ์การให้คะแนนที่ชัดเจน จะมีค่าความเที่ยงระหว่าง 0.73 ถึง 0.83 ทั้งนี้ค่าความเที่ยงจะมากหรือน้อย ขึ้นอยู่กับจำนวนคำถาม ถ้าคำถามมีมากข้อ ค่าความเที่ยงจะสูง เออร์วินและคณะ⁽³²⁾ รายงานว่า MEQ มีค่าความตรงเชิงทำนายในปี 1978 ระหว่าง 0.23 ถึง 0.46 และในปี 1980 ระหว่าง 0.21 ถึง 0.50

4. The Tutorial evaluation

เป็นวิธีการประเมินผลที่พัฒนาโดยมหาวิทยาลัยแมคมาสเตอร์ ประเทศแคนาดาและคณะแพทยศาสตร์ มหาวิทยาลัยนิวเม็กซิโก ประเทศสหรัฐอเมริกา⁽³³⁾ ใช้เป็นวิธีการประเมินผลหลักชนิดหนึ่งในการประเมินผลนิสิต มีลักษณะเป็นการประเมินโดยการสังเกต ใช้แบบมาตราประมาณค่า เป็นเครื่องมือประเมิน การประเมินแบบ Tutorial มีวัตถุประสงค์เพื่อให้ข้อมูลป้อนกลับ (Feedback) ความก้าวหน้าทางการเรียนของผู้เรียน เป็นข้อมูลเอกสารที่บันทึกเกี่ยวกับพฤติกรรมในการทำงานกลุ่ม ของผู้เรียนแต่ละคน (Self) กลุ่มผู้เรียน (Peer) และ Tutor เพื่อใช้ประกอบการประเมินผลรวม เมื่อสิ้นสุดหน่วยการเรียนรู้และการประเมินผลแบบนี้ จะให้ประสบการณ์แก่ผู้เรียน ในด้านการประเมินผลตนเองและ

เพื่อนร่วมงาน ลักษณะข้อคำถามในแบบมาตราประมาณค่า ที่ 2
เพื่อประเมินเพื่อนร่วมงาน⁽³⁴⁾ มีลักษณะดังแสดงในตาราง

Figure 4 A part of Modified Essay Questions.

You are a general practitioner. Mr. Henricks, aged 69, comes to see you accompanied by his daughter. You have not met him before. He complains of unsteadiness on his feet and difficulty passing urine. His daughter tells you that he has been depressed and off his food for some months.

Question 1. What hypotheses can you advance to explain what is going on?

Time : 5 minutes

Table 2 Items used for Peer-evaluation in problem-based learning groups

Part A: Peer-evaluation of effort

1. Did his/her best for the group
2. Sticked to the arrangements
3. Was a hard worker
4. worked scrupulously
5. Painstakingly prepared groups meetings

Part B: Peer-evaluation of knowledge:

6. Contributed valuable information
7. Showed factual knowledge of the topic under study
8. Asked good questions
9. Seemed to have a good grasp on complex matters
10. Made enlightening contributions to the discussion

4.1 วัตถุประสงค์ การประเมินผลนี้ใช้แบบมาตราประมาณค่าเป็นเครื่องมือในการประเมิน แบบมาตราประมาณค่าสามารถใช้ประเมินด้านต่าง ๆ ที่เครื่องมือชนิดอื่น ๆ ไม่เหมาะสมที่จะใช้ รวมถึงการประเมินความสามารถในการแก้ปัญหา ความคิดริเริ่มสร้างสรรค์ และการใฝ่หาความรู้ด้วยตนเอง คิวเลสและคณะ⁽³⁵⁾ ระบุว่าแบบมาตราประมาณค่าเหมาะที่จะใช้ประเมินคุณลักษณะหลัก ในด้านเจตคติ เช่น ความสามารถในการขอความร่วมมือจากผู้ป่วย ความรับผิดชอบ ความมั่นคงทางอารมณ์ ความซื่อสัตย์ จริยธรรม และความสามารถในการปรับปรุงตนเอง นอกจากนี้

ยังวัดในด้านเนื้อหาวิชา กระบวนการทำงาน และผลของการนำความรู้ไปแก้ปัญหา

4.2 จุดเด่น Tutorial evaluation สามารถให้ข้อมูลป้อนกลับทันทีในการประเมินผู้เรียน (Self) เพื่อนร่วมงาน (Peer) และ Tutor ใช้เป็นการประเมินความก้าวหน้าและการประเมินผลรวมได้ ในส่วนของแบบมาตราประมาณค่าซึ่งเป็นเครื่องมือ สามารถสร้างได้ง่าย ใช้เวลาน้อย สะดวกในการตอบ สามารถวัดคุณลักษณะต่าง ๆ ได้หลายชนิด⁽³⁶⁾ ดังแสดงในตารางที่ 3

Table 3 The strengths and weaknesses of the various evaluation methods.

	GLOBAL RATING	TRIPLE JUMP	MEQ	DIRECT OBSERV.	MCQ	OSCE
KNOWLEDGE	+	+	++		+++	+
PROBLEM-SOLVING	+	+++	++	+	+	+
INTERPERSONAL SKILLS	++			+++		+
TECHNICAL SKILLS	+			++		+++
ATTITUDES	++					

4.3 จุดด้อย การประเมินผลแบบนี้ต้องการ Tutor ที่มีคุณภาพสูงกล่าวคือ มีความรู้เกี่ยวกับวัตถุประสงค์ของหลักสูตร หน่วยการเรียนรู้ และตัวนิสิตเป็นอย่างดี ผ่านการอบรมการเป็น Tutor มาเป็นอย่างดี มีจิตใจที่จะเป็น Tutor มีความสามารถที่จะประเมินนิสิตว่า บรรลุตามวัตถุประสงค์หรือไม่ สามารถให้คำปรึกษา แนะนำแก่นิสิตได้ทั้งด้านวิชาการ สังคม อารมณ์ การปรับตัว ฯลฯ ในส่วนของแบบมาตราประมาณค่า การให้คะแนนขึ้นอยู่กับผู้ประเมิน อาจมีความลำเอียงในการให้คะแนน การใช้ภาษาหรือคำถามในแบบมาตราประมาณค่าและการแก้งคำตอบของผู้ตอบ จะทำให้เกิดความคลาดเคลื่อนได้ง่าย

4.4 คุณภาพ มีผู้รายงานในรูปคุณภาพของแบบมาตราประมาณค่าและการประเมินโดยเพื่อนร่วมงาน มหาวิทยาลัยแมคมาสเตอร์รายงานว่ แบบมาตราประมาณค่ามีความเที่ยงจากผู้ประเมิน (Interrater reliability) เท่ากับ

0.85 เคนและลอเลอร์⁽³⁷⁾ รายงานว่าการประเมินโดยเพื่อนร่วมงาน (Peer-rating) มีค่าความเที่ยงเท่ากับ 0.45 และค่าความตรงเท่ากับ 0.35 ไมสเลส⁽³⁸⁾ เบอห์นเน็ทและคาเวย์⁽³⁹⁾ ไลน์แมน⁽⁴⁰⁾ เทอร์นเนอร์และเรคคอร์ด⁽⁴¹⁾ อาร์โนลและคณะ⁽⁴²⁾ รายงานว่าการประเมินผลโดยเพื่อนร่วมงานมีค่าความเที่ยงสูง โดยอาร์โนลและคณะรายงานว่ามีความเที่ยงเท่ากับ 0.94 นอกจากนี้รายงานวิจัยหลายคณะยังพบว่า การประเมินโดยเพื่อนร่วมงานมีความสัมพันธ์สูงกับการประเมินโดยอาจารย์ เกรฟและโพลเดอร์⁽⁴³⁾ รายงานว่า การประเมินโดยเพื่อนร่วมงานมีค่าความเที่ยงเท่ากับ 0.99 และความตรงร่วมสมัย (Concurrent validity) มีค่าระหว่าง 0.20 ถึง 0.70 ซึ่งศึกษาโดยหาความสัมพันธ์ระหว่างคะแนนที่ได้จากการประเมินโดยเพื่อนร่วมงาน กับคะแนนที่ได้จากการสอบด้วยเครื่องมือประเมินผลชนิดอื่น ดังแสดงรายละเอียดในตารางที่ 4

Table 4 Pearson correlations between summative evaluation variables. Coefficients in underlines are significant at the 1% level.

Summative evaluation method	1	2	3	4	5	6
	Multi. choice	Essay	Peer. know.	Peer. effort	Cont.	Form.
1. Multiple choice test	1					
2. Essay-test	<u>.22</u>	1				
3. Peer-evaluation of knowledge	<u>.36</u>	<u>.30</u>	1			
4. Peer-evaluation of effort	.10	.00	<u>.61</u>	1		
5. Content of paper	.15	.08	<u>.02</u>	.10	1	
6. Form of paper	.10	.03	.05	.03	<u>.70</u>	1

วิจารณ์

เครื่องมือประเมินผลที่มีใช้ในโรงเรียนแพทย์ต่างประเทศ ที่จัดการเรียนการสอนแบบใช้ปัญหาเป็นหลักมีมากมายหลายชนิด เรียกชื่อแตกต่างกันไปตามสถาบันที่พัฒนาขึ้น นอกเหนือจาก Triple Jump Exercise, The Objective Structured Clinical Examination, Modified Essay Questions, The Tutorial evaluation ซึ่งเป็นเครื่องมือประเมินผลแบบบูรณาการที่มีผู้นิยมใช้กันอย่างแพร่หลายแล้วยังมี IPA⁽⁴⁴⁾ (Individual Process Assessment) ซึ่งใช้ในคณะแพทยศาสตร์ มหาวิทยาลัยนิวเม็กซิโก ประเทศสหรัฐอเมริกา, MRAT⁽⁴⁵⁾ และ CRT⁽⁴⁶⁾ (Medical Reasoning Aptitude Test and Clinical Reasoning Test) ซึ่งใช้ในคณะแพทยศาสตร์ มหาวิทยาลัยเซาเทิร์นอิลลินอยส์ ประเทศสหรัฐอเมริกา, Clinical Problem-solving Performance⁽⁴⁷⁾ ซึ่งใช้ในคณะพยาบาลศาสตร์ มหาวิทยาลัยแมคมาสเตอร์ ประเทศแคนาดา, Progress Test⁽⁴⁸⁾ และ SIMP (The Simulation of Initial Medical Problem-solving)⁽⁴⁹⁾ ซึ่งใช้ในคณะแพทยศาสตร์ มหาวิทยาลัยลิมเบอร์ก ประเทศเนเธอร์แลนด์, P4 (Portable Patient Problem Pack)⁽⁵⁰⁾, Self-Assessment Item Bank⁽⁵¹⁾, และ Contracts⁽⁵²⁾ ซึ่งใช้ในคณะแพทยศาสตร์ มหาวิทยาลัยแมคมาสเตอร์ ประเทศแคนาดา เครื่องมือประเมินผลแบบบูรณาการเหล่านี้ ยังไม่เป็นที่นิยมใช้มากนักในโรงเรียนแพทย์หรือสถาบันที่มีการจัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก อย่างไรก็ตามเป็นที่น่าสังเกตว่า เครื่องมือประเมินผลดังกล่าว ส่วนใหญ่มีการคำนวณหาค่าความตรง ความเที่ยง และความสะดวกในการนำไปใช้ ซึ่งก่อนหน้านั้นเครื่องมือประเมินผลจำนวนมากขาดการให้ความสนใจในด้านคุณภาพ ดังที่ เคทซ์และสโนว์⁽⁵³⁾

ได้วิเคราะห์เครื่องมือประเมินผลที่ใช้ในการประเมินผลการปฏิบัติงานของบุคลากรสาธารณสุข ซึ่งรวบรวมอยู่ในหนังสือ Assessing health workers' performance: a manual for training and supervision พบว่าเครื่องมือเหล่านี้ร้อยละ 66.67 ไม่ได้มีการหาค่าความตรง ความเที่ยง และความสะดวกในการใช้ ในด้านความเที่ยง ฮับบาร์ดและซูเมเกอร์⁽⁵⁴⁾ เสนอแนะว่า เครื่องมือประเมินผลที่มีคุณภาพจะต้องมีค่าความเที่ยงตั้งแต่ 0.70 ขึ้นไป

สรุป

การวิจัยนี้เป็นการศึกษาเอกสาร ผู้วิจัยได้ศึกษาเครื่องมือประเมินผลชนิดที่มีใช้ในหลักสูตรที่มีการจัดการเรียนการสอนแบบใช้ปัญหาเป็นหลัก ในโรงเรียนแพทย์ต่างประเทศ โดยศึกษาถึงชนิด วัตถุประสงค์ จุดเด่น-จุดด้อย คุณภาพ (ความตรง ความเที่ยง และความสะดวกในการใช้) ของเครื่องมือแต่ละชนิด เครื่องมือประเมินผลชนิดที่มีใช้ในหลักสูตรที่มีการเรียนการสอนแบบใช้ปัญหาเป็นหลัก เป็นเครื่องมือประเมินผลแบบบูรณาการ วัดคุณลักษณะหลักและคุณลักษณะรองหลายด้าน รวมทั้งวัดด้านเนื้อหาวิชา กระบวนการแสวงหาความรู้ และการนำความรู้ที่หาได้ไปแก้ปัญหา เครื่องมือประเมินผลแบบบูรณาการที่มีใช้กันแพร่หลาย ได้แก่ Triple Jump Exercise, The Objective Structured Clinical Examination, Modified Essay Questions และ The Tutorial evaluation

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ ศาสตราจารย์นายแพทย์ เจริญ วราวิทย์ หัวหน้าหน่วยแพทยศาสตรศึกษา คณะแพทยศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย ที่ให้การสนับสนุนและคำปรึกษา
ในการวิจัย

ขอขอบพระคุณ อาจารย์ยงศ์วรรณ วงศ์สุภา หัวหน้า

หน่วยบริการตอบคำถามและช่วยการค้นคว้าวิจัย หอสมุด
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ช่วยตรวจแก้ไข
รูปแบบของการเขียนเอกสารอ้างอิง

อ้างอิง

1. Waterman RE. Butler C. Curriculum : problems to stimulate learning. In: Kaufman A, ed. Implementing Problem-Based Medical Education: Lessons From Successful Innovations. New York: Springer, 1985. 16-17
2. Richards R. Fulop. T. Bannerman J. Greenholm G. Guilbert J-J. Wunderlich M. Innovative Schools for Health Personnel. Geneva: World Health Organization, 1987. 3-6
3. Neufeld VR. The design and use of assessment methods for problem based learning. In : Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning : New Directions in Training for the Health Professions*. Maastricht: Van Gorcum, Assen, 1984. 68-71
4. Harden RM. Gleesen FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979 Jan; 13(1): 41-54
5. Feletti GI. Smith EK. Modified essay questions: are they worth the effort?. *Med Educ* 1986 Mar; 20(2): 126-132
6. McMaster University. *Evaluation Methods: A Resource Manual*. Ontario: McMaster University, 1987. 10-12
7. Guilbert JJ. *Educational handbook for health personnel*. WHO offset publication No. 35. Geneva: World Health Organization 1977. 324
8. Schmidt HG. Volder ML. *Tutorials in Problem-Based Learning: New Directions in Training for the Health professions*. Maastricht: Van Gorcum, 1984. 69
9. Painvin C. Neufeld V. Norman G. Walker I. Whelan G. The Triple Jump Exercise: a structured measure of problem-solving and self-directed learning. *Proceedings of the Annual Conference on Research in Medical Education*. Washington: American Association of Medical Colleges, 1979.
10. Powles ACP. Wintrip N. Neufeld V. Wakefield JG. Coates G. Barrows J. The Triple Jump Exercise: Further Studies on an Evaluative Technique. *Proceeding of the Annual Conference on Research in Medical Education*. Washington: American Association of Medical Colleges, 1981.
11. บุญนาท ลายสนิทเสรีกุล. เจลิม วรวิทย์. การสร้างเครื่องมือทดสอบให้ได้มาตรฐาน. ในเจลิม วรวิทย์, เสรี ร่วมสุข บรรณาธิการ. *แพทยศาสตร์ศึกษา*. กรุงเทพมหานคร: คอมพิวเตอร์ไซน์แอนด์พริ้นท์, 2526. 414
12. Harden RM. Gleeson FA. Assessment of Clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979 Jan; 13(1): 41-54
13. Oaks WW. Scheinok PA. Husted Fl. Objective evaluation of assessing student performance in clinical clerkship. *J Med Educ* 1969 Mar; 44(3): 207-213
14. Printen KJ. Chappell W. Whitney DR. Clinical performance evaluation of junior medical students. *J Med Educ* 1973 Apr; 48(4): 343-348
15. Hallock JA. Christensen JA. Denker MW. Hochberg CJ. Trudeau WL. Williams JW. A comparison of the clinical performance of students in three-and four-year curricula. *J Med Educ* 1977 Aug; 32(8): 658-663
16. Willoughby TL. Gammon LC. Tonas HS. Correlates of clinical performance during medical school. *J Med Educ* 1979 Jun; 54(6): 453-460
17. บุญนาท ลายสนิทเสรีกุล. โทม รัตนวราภรณ์. เจลิม วรวิทย์. เครื่องมือการประเมินผลการปฏิบัติงานทางคลินิกของนิสิตแพทย์. *จุฬาลงกรณ์เวชสาร* 2525 กันยายน; 26(5): 397-404
18. Black NMI. Harden RM. Providing feedback to students on clinical skills by using the objective structured clinical examination. *Med Educ* 1986 Jan; 20(1): 48-52
19. McMaster University. *Evaluation Methods: A Resource Manual*. Ontario: McMaster University, 1987. 32-33
20. McKnight J. Rideout E. Brown B. Ciliska D. Patton P. Rankin J. Woodwarn CA. The objective structured clinical examination: an alternative approach in assessing student performance. *J Med Educ* (in press)
21. Woodward C. Patton D. Brown B. The Relationship Between Nursing Student's Evaluation of the OSCE and Performance. *Procee-*

- dings of the International Conference on Newer Developments in Assessing Clinical Competence, 1985.
22. Van Niekerk JGP. Lombard SA. The OSCE experiment at MEPUNSA. *Curations* 1982; 5: 44-48
 23. Adeyemi-Doro HO. Objective procedural clinical examination (OPCE) for assessing competence of medical and dental students in practical skills. *East Afr Med J* 1984 Jun; 61(6): 477-481
 24. Nayar U. Malik SL. Bylani RL. Objective structured practical examination: a new concept in assessment of laboratory exercises in preclinical sciences. *Med Educ* 1986 May; 20(3) : 204-209
 25. The Board of Censors of the Royal College of General Practitioners. The modified essay question. *Proc Roy Coll Gen Pract* 1971; 21: 373-385
 26. Knox JDE. How to use modified essay questions. *Med Teacher* 1980; 2(1): 20-24
 27. McMaster University. Evaluation of Student Learning. Ontario: McMaster University, 1985. 13-14
 28. Feletti GI. Engel CE. The modified essay questions for testing problem-solving skills. *Med J Aust* 1980 Jan 26; 1(2): 79-80
 29. Feletti GI. Smith EKM. Modified essay questions: are they worth the effort?. *Med Educ* 1986 Mar; 20(2): 126-132
 30. Feletti GI. Reliability and validity studies on Modified Essay Questions. *J Med Educ* 1980 Nov; 55(11): 933-941
 31. Stratford P. Pierce-Fenn H. Predictive validity of the modified essay questions. *Physiother Can* 1985; 37(6): 356-359
 32. Irwin WG. Bamber JH. The cognitive structure of the Modified Essay Questions. *Med Educ* 1982 Nov; 16(6): 326-331
 33. West DA. Umland BE. Lucero SM. Evaluating student performance. In: Kaufman A, ed. *Implementing Problem-Based Medical Education: Lessons from Successful Innovatives*. New York: Springer, 1985. 147-150
 34. Grave W. Volder M. Peer-evaluation and problem-based learning. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: New Directions in Training for the Health Professions*. Maastricht: Van Gorcum, Assen, 1984. 119
 35. Streiner DL. Global rating scales. In: Neufeld VR, Norman GR, eds. *Assessing Clinical competence*. New York: Springer, 1985. 119-141
 36. McMaster University. *Evaluation Methods: A Resource Manual*. Ontario: McMaster University, 1987. 5
 37. Kane JS. Lawler EE. Methods of peer assessment. *Psychol Bull* 1978; 85: 555-586
 38. Moses I. Assessment in Australian higher education 1974-1979. *Asses High Educ* 1980; 5: 294-315
 39. Burnett W. Cavaye G. Peer assessment by fifth year students of surgery. *Asses High Edu* 1980; 33: 273-278
 40. Fineman S. Reflections on peer teaching and peer assessment: an undergraduate experience. *Asses Evaluation High Educ* 1981; 6: 82-93
 41. Turner CJ. Record AL. Self-concept clarity: evaluation of self, Peer and Professional ratings. *Measurement of Evaluation in Guidance* 1981; 14 : 38-45
 42. Arnold L. Willoughby L. Calkins V. Gammon L. Eberhart G. Use of Peer-evaluation in the assessment of medical students. *J Med Educ* 1981 Jan; 56(1): 35-42
 43. Grave W. Volder M. Peer-evaluation and Problem-based learning. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: New Directions in Training for the Health Professions*. Maastricht: Van Gorcum, Assen, 1984. 119
 44. West DA. Umland BE. Lucero SM. Evaluating student performance. In : Kaufman A, ed. *Implementing Problem-Based Medical Education: Lessons from Successful Innovations*. New York: Springer, 1985. 150-158
 45. Vu NV. Barrows HS. Paiva REA. Dawson-Saunders BK. The Medical Reasoning Aptitude Test. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: A New Direction in Teaching the Health Professions*. Maastricht: Van Gorcum, Assen, 1984. 72-80
 46. Williams RG. Vu NV. Barrows HS. Verhulst S. Profile of the Clinical Reasoning Test. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: A New Direction in Teaching the Health Professions*. Maastricht: Van Gorcum, Assen, 1984. 81-90
 47. Tamblyn RM. Lewis KE. Increasing the objectivity of measuring clinical problem-solving performance in patient situations. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: A New Direction in Teaching*

- the Health Professions. Maastricht: Van Gorcum, Assen, 1984. 91-105
48. Imbos T. Drukker J. Mameren HV. Verwijnen M. The growth in knowledge of anatomy in a problem-based curriculum. In: Schmidt HG, Volder ML, eds. *Tutorials in Problem-Based Learning: A New Direction in teaching the health professions*. Maastricht ; Van Gorcum, Assen, 1984. 106-115
49. Leeuwen Y. Graaff E. Drop MJ. The construction of simulation initial medical problem-solving (SIMP). เอกสารประกอบการประชุมเชิงปฏิบัติการเรื่อง Student Assessment วันที่ 22,24-25 กันยายน 2530. ณ ดิโกอันโทมฮิดล ชั้น 2 และตึกเวชศาสตร์ป้องกันและสังคม คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
50. Barrows HS. Tamblin RM. The portable patient problem pack (P4): a problem-based learning unit. *J Med Educ* 1977 Dec; 52(12): 1002-1004
51. McMaster University. *Evaluation Methods: A Resource Manual*. Ontario: McMaster University, 1987. 13
52. McMaster University. *Evaluation Methods: A Resource Manual*. Ontario: McMaster University, 1987. 39-42
53. Katz FM. Snow. R. *Assessing Health Workers' Performance: A Manual for Training and Supervision*. Public Health Paper No. 72. Geneva: World Health Organization, 1980.
54. Hubbard JP. Schumacher CF. *Measuring Medical Education: the Tests and Test Procedures of the National Board of Medical Examiners*. Philadelphia: Lea & Febiger, 1971. 60