

๑. เนื้อหา ความรู้ ทักษะ ที่ได้เรียนรู้จากการฝึกอบรม

จากการเข้าร่วมการอบรมเชิงปฏิบัติการ "Assessment Workshop for Clinical Teachers ๒๐๒๕" ระหว่างวันที่ ๒๔ – ๒๖ มีนาคม ๒๕๖๘ ข้าพเจ้าได้รับความรู้และทักษะในด้านการวัดผลและประเมินผลทางคลินิก โดยมีประเด็นสำคัญดังนี้:

๑.๑ หลักการพื้นฐานของการวัดผล (Basic Principles of Assessment)

หลักการวัดผลทางการศึกษาศาสนาคลินิกมีเป้าหมายเพื่อสะท้อนความรู้ ทักษะ และเจตคติของผู้เรียนอย่างถูกต้อง ยุติธรรม และเที่ยงตรง การประเมินผลที่ดีต้องมีความ Valid (วัดสิ่งที่ควรวัด), Reliable (ให้ผลสม่ำเสมอ), Feasible (ปฏิบัติได้จริง), และ Acceptable (ได้รับการยอมรับจากผู้เกี่ยวข้อง) การวัดผลสามารถใช้ได้ทั้งในเชิง Formative (เพื่อพัฒนา) และ Summative (เพื่อสรุปผล) โดยควรคำนึงถึงความต่อเนื่องและหลากหลาย เพื่อประเมินผลได้รอบด้านตามเป้าหมายการเรียนรู้

การวัดผลทางการศึกษาทางคลินิกมีเป้าหมายหลักเพื่อสะท้อนระดับความรู้ ทักษะ และเจตคติของผู้เรียนอย่างถูกต้องและรอบด้าน โดยหลักการสำคัญได้แก่:

Validity (ความตรง): วัดสิ่งที่ตั้งใจจะวัด ตัวอย่างเช่น หากต้องการวัดทักษะการซักประวัติ การใช้ข้อสอบ MCQ จะไม่เหมาะเท่าการประเมินจากการสังเกตในสถานการณ์จริง เช่น OSCE

Reliability (ความเชื่อถือได้): ผลการประเมินต้องคงที่แม้ผู้ประเมินหรือเวลาจะเปลี่ยนไป เช่น การใช้หลายผู้ประเมินและใช้ rubric ที่ชัดเจนจะช่วยเพิ่มความเชื่อถือได้

Feasibility (ความเป็นไปได้): เครื่องมือประเมินต้องสามารถใช้ได้จริงภายใต้ข้อจำกัด เช่น งบประมาณ เวลา หรือทรัพยากรบุคคล

Acceptability (การยอมรับ): ผู้เรียนและผู้สอนต้องยอมรับผลการประเมินและรู้สึกว่าเป็นธรรม

การใช้ Formative (เพื่อพัฒนา) และ Summative (เพื่อสรุปผล) ร่วมกัน จะช่วยให้นักศึกษาเติบโตอย่างต่อเนื่อง ทั้งด้านวิชาการและทักษะทางคลินิก

๑.๒ เทคนิคการกำหนดเกณฑ์คะแนน (Standard Setting)

การกำหนดเกณฑ์คะแนน (Standard Setting) คือกระบวนการที่ใช้กำหนดจุดแบ่งผ่านหรือระดับความสามารถที่ยอมรับได้ในแต่ละการประเมิน ซึ่งช่วยให้ผลการสอบสะท้อนความสามารถจริง ไม่ขึ้นกับความยากง่ายของข้อสอบ วิธีที่นิยม เช่น Angoff Method, Borderline Group Method และ Bookmark Method เหมาะกับบริบทที่ต้องการความโปร่งใสและเป็นธรรม โดยเฉพาะในการสอบแบบ OSCE หรือ MCQ การใช้เทคนิคนี้ช่วยให้การประเมินมีความยุติธรรมและสมเหตุสมผลมากขึ้น

การตั้งคะแนนผ่านโดยอ้างอิง “๕๐%” หรือ “๖๐%” อาจไม่เหมาะสมกับข้อสอบเชิงสมรรถนะในยุคปัจจุบัน เทคนิคการตั้งเกณฑ์แบบเป็นระบบจะช่วยให้การวัดผลมีความยุติธรรมและเที่ยงตรงมากขึ้น เช่น:

Modified Angoff Method: ให้ผู้เชี่ยวชาญประเมินความน่าจะเป็นที่ผู้เรียนระดับ "สมควรผ่าน" จะตอบคำถามแต่ละข้อถูก

Borderline Group Method: ใช้คะแนนเฉลี่ยของผู้เข้าสอบที่อยู่ในระดับ "หวุดหวิดผ่าน" มาเป็นเกณฑ์

Bookmark Method: ใช้กับข้อสอบเรียงตามระดับความยาก โดยกำหนดจุด "bookmark" ที่เหมาะสมที่สุด

การใช้ Standard Setting ช่วยให้ได้ Cut Score ที่มีความเหมาะสมกับระดับความยากของข้อสอบและสมรรถนะที่ต้องการวัด ไม่ตายตัวกับคะแนนดิบ ทำให้การสอบสะท้อนคุณภาพของผู้เรียนได้ดีกว่าเดิม

๑.๓ การวิเคราะห์ข้อสอบ (Item Analysis) ทั้งแบบ MCQ, MEQ และ OSCE

การวิเคราะห์ข้อสอบช่วยตรวจสอบคุณภาพของข้อสอบที่ใช้ว่าเหมาะสมหรือไม่ เช่น ค่าความยาก (Difficulty Index), ค่าจำแนก (Discrimination Index), และค่า Reliability โดยเฉพาะในข้อสอบแบบ MCQ (Multiple-choice questions) การวิเคราะห์จะช่วยปรับปรุงคำถามให้มีประสิทธิภาพมากขึ้น ขณะที่ MEQ (Modified Essay Questions) และ OSCE (Objective Structured Clinical Examination) จะเน้นวิเคราะห์ความชัดเจนของโจทย์ การกำหนดเกณฑ์การให้คะแนน และความสม่ำเสมอของผู้ประเมิน เพื่อให้การประเมินสะท้อนสมรรถนะผู้เรียนได้อย่างแท้จริง

การวิเคราะห์ข้อสอบเป็นกระบวนการที่จำเป็นเพื่อปรับปรุงคุณภาพของเครื่องมือประเมิน ตัวอย่างตัวชี้วัด ได้แก่:

Difficulty Index: ค่าระหว่าง ๐ – ๑ (ยิ่งใกล้ ๑ ยิ่งง่าย) ค่าที่ดีควรอยู่ระหว่าง ๐.๓ – ๐.๘

Discrimination Index: ค่าระหว่าง -๑ ถึง +๑ (ยิ่งใกล้ +๑ ยิ่งแยกผู้เรียนเก่งกับไม่เก่งได้ดี) ค่าเหมาะสมควร > ๐.๓

Distractor Efficiency: สัดส่วนของ distractor ที่ถูกเลือกโดยผู้เรียน หาก distractor ไม่ถูกเลือกเลย ควรพิจารณาปรับปรุง

สำหรับ OSCE การวิเคราะห์จะดูความสม่ำเสมอของผู้ประเมิน (inter-rater reliability), ความชัดเจนของ checklist และการครอบคลุมวัตถุประสงค์การเรียนรู้

การวิเคราะห์ข้อสอบจึงมีบทบาทสำคัญในการตรวจสอบและพัฒนาเครื่องมือประเมินให้มีคุณภาพสูงขึ้นอย่างต่อเนื่อง

๑.๔ แนว Programmatic Assessment คือการประเมินอย่างต่อเนื่องหลากหลายรูปแบบที่เน้นการเรียนรู้ (assessment for learning) โดยสะสมข้อมูลจากหลายแหล่งตลอดการศึกษาเพื่อพัฒนาผู้เรียน ในขณะที่ Authentic Assessment คือการประเมินที่สะท้อนการปฏิบัติงานจริง เช่น การดูแลผู้ป่วย การสื่อสาร หรือการวิเคราะห์เคส การประเมินทั้งสองแนวคิดนี้ช่วยให้การเรียนการสอนเป็นไปตามบริบทของโลกแห่งความเป็นจริง และสนับสนุนการเรียนรู้เชิงลึกของนักศึกษา

Programmatic Assessment คือแนวคิดที่มองการประเมินเป็นกระบวนการต่อเนื่อง โดยใช้ข้อมูลจากแหล่งต่าง ๆ อย่างมีโครงสร้าง เช่น Mini-CEX, DOPS, Logbook, Portfolio รวมถึงการสังเกตพฤติกรรมประจำวัน แล้วนำข้อมูลเหล่านี้มาวิเคราะห์ร่วมกันเพื่อตัดสินผลการเรียนในภาพรวม โดยมุ่งให้ feedback ที่มีคุณภาพและส่งเสริมการเรียนรู้

Authentic Assessment คือการประเมินในบริบทที่สะท้อนการปฏิบัติงานจริง เช่น การดูแลผู้ป่วย การสื่อสารกับทีม การวางแผนการรักษา เป็นการประเมินที่ไม่เน้นแค่ “ถูก-ผิด” แต่ประเมินจากการคิดวิเคราะห์ การลงมือปฏิบัติ และการแสดงออกที่เหมาะสม

ทั้งสองแนวคิดนี้ช่วยให้การประเมินเน้นที่สมรรถนะจริงของผู้เรียน และมีความยืดหยุ่น สอดคล้องกับบริบทวิชาชีพมากยิ่งขึ้น

๑.๕ การพัฒนาและตรวจสอบข้อสอบทั้งแบบ Multiple-choice, Constructed Response, OSCE

การออกข้อสอบที่ดีต้องยึดตาม Learning Outcomes อย่างชัดเจน MCQ ควรเขียนคำถามให้ตรงประเด็น หลีกเลี่ยงข้อความคลุมเครือหรือหลอกลวง ขณะที่ Constructed Response (ตอบคำถามปลายเปิด) ควรออกแบบให้สามารถประเมินกระบวนการคิด วิเคราะห์ และสังเคราะห์ได้ ส่วน OSCE ต้อง

ออกแบบสถานีให้สะท้อนสถานการณ์จริง พร้อมเกณฑ์การให้คะแนนที่ชัดเจน การตรวจสอบ (Item Review) เป็นขั้นตอนสำคัญเพื่อให้ข้อสอบมีคุณภาพ ตรงวัตถุประสงค์ และลดอคติ

การออกแบบข้อสอบที่ดีต้องสะท้อนเป้าหมายการเรียนรู้ (Learning Outcomes) และสมรรถนะที่ต้องการวัด โดยกระบวนการพัฒนาและตรวจสอบข้อสอบแบ่งออกเป็นหลายรูปแบบตามลักษณะของเครื่องมือ:

Multiple-choice Questions (MCQ) ข้อสอบควรมี “Stem” (คำถาม) ที่ชัดเจน และตัวเลือก (Options) ที่มี Distractor มีคุณภาพ หลีกเลี่ยงการใช้ภาษาคลุมเครือหรือคำถามที่ตอบได้ด้วยการคาดเดา ตรวจสอบให้แน่ใจว่าคำตอบถูกเพียงข้อเดียว และสามารถแยกแยะผู้เรียนที่มีความรู้จริงได้

Constructed Response เป็นข้อสอบปลายเปิด เช่น Short Answer หรือ MEQ (Modified Essay Question) ใช้เพื่อวัดความสามารถในการวิเคราะห์ สังเคราะห์ และสื่อสารความคิด ควรมี Rubric หรือ Guideline ในการตรวจข้อสอบที่ชัดเจน เพื่อลดความแปรปรวนระหว่างผู้ตรวจ

OSCE (Objective Structured Clinical Examination) ใช้สถานีจำลองสถานการณ์จริงเพื่อประเมินทักษะ เช่น การตรวจร่างกาย การสื่อสาร หรือ Clinical Reasoning ต้องมี Checklist หรือ Rating Scale ที่กำหนดไว้อย่างชัดเจน และผู้ประเมินควรได้รับการฝึกอบรม การเขียน Script หรือ Scenario ให้สมจริงเป็นหัวใจสำคัญของการประเมินแบบ OSCE

ในทุกประเภทข้อสอบ การ “Review” หรือการตรวจสอบร่วมกันระหว่างคณะผู้สอนมีความสำคัญมาก ทั้งในด้านความถูกต้อง ความเหมาะสม และความสอดคล้องกับจุดประสงค์การเรียนรู้

๑.๖ การประเมินผลการปฏิบัติงานในสถานที่ทำงานจริง (Workplace-based Assessment)

การประเมินผลในสถานที่ทำงานจริงเป็นวิธีที่มีประสิทธิภาพในการประเมินสมรรถนะในบริบทจริง ตัวอย่างเครื่องมือเช่น Mini-CEX, Direct Observation of Procedural Skills (DOPS) และ Case-based Discussion (CbD) ซึ่งผู้ประเมินสามารถให้ Feedback ทันทีหลัง การประเมิน ช่วยให้ผู้เรียนเห็นจุดแข็งจุดอ่อนของตนเอง และปรับปรุงได้ทันเวลา นอกจากนี้ยังส่งเสริมความสัมพันธ์อาจารย์-นักศึกษา และวัฒนธรรมการเรียนรู้ที่เน้นพัฒนาอย่างต่อเนื่อง

การประเมินในสถานการณ์จริงมีบทบาทสำคัญในการวัดสมรรถนะเชิงปฏิบัติ เพราะสะท้อนพฤติกรรมและความสามารถที่แท้จริงของผู้เรียนในบริบทของการทำงาน ตัวอย่างเครื่องมือที่นิยมใช้ ได้แก่:

Mini-CEX (Mini Clinical Evaluation Exercise) ประเมินทักษะคลินิกผ่านการสังเกตผู้เรียนดูแลผู้ป่วยจริง ใช้เวลาสั้น (๑๐-๑๕ นาที) และเน้น feedback หลังประเมิน

DOPS (Direct Observation of Procedural Skills) ใช้ประเมินทักษะหัตถการ เช่น การเจาะเลือด เย็บแผล พร้อม checklist การให้คะแนน

CbD (Case-based Discussion) เป็นการพูดคุยเชิงลึกกับผู้เรียนเกี่ยวกับกรณีผู้ป่วยที่ดูแล เพื่อประเมิน clinical reasoning และ decision making

จุดแข็งของการประเมินแบบนี้คือให้ “Real-time Feedback” และสามารถปรับปรุงพฤติกรรมได้ทันที ช่วยส่งเสริมวัฒนธรรม “การเรียนรู้ตลอดเวลา” (Lifelong Learning) และความเป็นมืออาชีพ (Professionalism)

๑.๗ การใช้ Portfolio และ Clinical Performance Rating เพื่อวัดสมรรถนะ

Portfolio คือเครื่องมือรวบรวมผลงานหรือกิจกรรมการเรียนรู้ของผู้เรียน เช่น Reflection, Feedback, หรือ Logbook ซึ่งช่วยสะท้อนการเรียนรู้อย่างต่อเนื่องและการเติบโตในวิชาชีพ ขณะที่ Clinical

Performance Rating เป็นการให้คะแนนตามพฤติกรรมหรือทักษะที่แสดงออก เช่น การสื่อสาร การวินิจฉัย หรือความเป็นมืออาชีพ การประเมินทั้งสองรูปแบบนี้เหมาะกับการวัดผลสมรรถนะที่ซับซ้อนและช่วยกระตุ้นให้ผู้เรียนพัฒนาตนเองอย่างรอบด้าน

Portfolio คือเครื่องมือรวบรวมหลักฐานการเรียนรู้ของผู้เรียนอย่างเป็นระบบ เช่น Reflection, Logbook, Case Summary, Feedback มีบทบาทสำคัญใน การประเมินระยะยาว โดยสะท้อนพัฒนาการการเรียนรู้ และการตั้งเป้าหมายส่วนบุคคล ต้องมีโครงสร้างชัดเจนและระบบการประเมินที่แน่นอน เช่น การใช้ Rubric ในการให้คะแนนแต่ละส่วน

Clinical Performance Rating การประเมินเชิงพฤติกรรมในการทำงาน เช่น Communication, Teamwork, Professionalism, Clinical Judgement ใช้แบบฟอร์มที่มีเกณฑ์ระดับความสามารถ เช่น ๑-๕ หรือ ๑-๗ Likert scale เป็นการประเมินที่ใช้ในสถานการณ์จริง โดยหัวหน้างานหรืออาจารย์ประจำที่ใกล้ชิดกับผู้เรียน

การประเมินสองรูปแบบนี้เหมาะสมอย่างยิ่งสำหรับ Competency-based Education เพราะวัดพฤติกรรมในบริบทจริง และส่งเสริมให้ผู้เรียนมีความรับผิดชอบต่อการเรียนรู้ของตนเอง

๒. การนำมาใช้ประโยชน์ในงานของหน่วยงาน ข้อเสนอแนะเพื่อการพัฒนา

เนื้อหาและกิจกรรมจากการอบรมสามารถนำมาใช้ในการออกแบบและปรับปรุงการประเมินผลนักศึกษาแพทย์ให้เหมาะสมและทันสมัยมากขึ้น โดยเฉพาะในด้านการตั้งเกณฑ์คะแนนที่โปร่งใสและตรวจสอบได้ การเขียนข้อสอบแบบวัดผลตามสมรรถนะ (Competency-based Assessment) การประเมินผลนักศึกษาด้วยวิธีหลากหลายเพื่อสะท้อนความสามารถที่แท้จริง

ข้อเสนอแนะคือควรจัดให้มีการอบรมเชิงปฏิบัติในหัวข้อเฉพาะทางอย่างต่อเนื่อง รวมทั้งมีการจัดทำคลังข้อสอบร่วมและระบบประเมินผลร่วมกันระหว่างภาควิชา

๓. ความคิดเห็นต่อหลักสูตรการฝึกอบรม

หลักสูตรนี้มีความครบถ้วนทั้งในเชิงทฤษฎีและปฏิบัติ วิทยากรมีความเชี่ยวชาญ ถ่ายทอดเนื้อหาได้อย่างชัดเจน กิจกรรมกลุ่มช่วยให้เกิดการเรียนรู้ร่วมกัน เนื้อหาเป็นประโยชน์และสามารถนำไปใช้ได้จริง มีความคุ้มค่าต่อเวลาและทรัพยากรที่ใช้ในการเข้าร่วม