

## Assessment Utility

ผศ. นพ. ชัยพฤกษ์ กุสุมาพรรณโณ  
คณะแพทยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

ผู้ปฏิบัติงานประกันคุณภาพการศึกษาด้านแพทยศาสตร์ควรรู้จักคำว่า assessment utility หรือ วัตถุประสงค์ประโยชน์ของเครื่องมือประเมินผล คำนี้ปรากฏในเกณฑ์มาตรฐานของสมาพันธ์แพทยศาสตรศึกษาโลก (World Federation for Medical Education: WFME) ทั้งในระดับก่อนปริญญาและหลังปริญญา ในองค์ประกอบที่ 3.1 ในส่วนอรรถาธิบาย ได้กำหนดนิยามของ assessment utility ไว้ว่า รูปแบบและเครื่องมือประเมินผลนั้น ควรประกอบด้วยคุณลักษณะด้านความตรง (validity) ความเที่ยง (reliability) ผลกระทบโดยตรงต่อการศึกษา (educational impact) ความยอมรับ (acceptability) และประสิทธิภาพ (efficiency) ตามรูปที่ 1

Van der Vleuten (1996) ได้ระบุองค์ประกอบเพิ่มเติมอีก 2 ข้อ คือ feasibility หรือ ความเป็นไปได้ที่จะนำเครื่องมืออื่น ๆ มาใช้ และค่าใช้จ่าย (cost) ของเครื่องมือ



รูปที่ 1 องค์ประกอบของ assessment utility

ดังนั้น ผู้บริหารการศึกษาคควรมองในภาพรวมว่า หลักสูตรมีการวางแผนการประเมินผลอย่างไร มีระบบรูปแบบ และการเลือกเครื่องมืออย่างไรที่เหมาะสม เพื่อให้บรรลุวัตถุประสงค์และผลสัมฤทธิ์ของหลักสูตรตามที่ได้กำหนดไว้ในหลักสูตร (มคอ.2) ผู้นิพนธ์ขอแสดงรายละเอียดในแต่ละองค์ประกอบ

การเลือกใช้รูปแบบและเครื่องมือการประเมินที่ **valid** เป็นเหมือนหญ้าปากคอก (เป็นเรื่องง่าย ไม่ซับซ้อน) อาจารย์เป็นผู้สอนและออกข้อสอบ ข้อสอบจึง valid แต่แนวคิดเรื่องความตรงหรือ validity มีหลายมิติ ไม่ง่ายเสียทีเดียว อาจเป็นเรื่องของเครื่องมือที่วัดนั้นตรงตามวัตถุประสงค์ที่นำมาใช้หรือไม่ ใช้ประเมิน competency ที่ต้องการประเมินจริงหรือไม่ (มี construct ถูกต้องเหมาะสมหรือไม่) หรือว่าเครื่องมือนั้นสุ่มจำนวนตัวอย่างของเนื้อหาที่เพียงพอและบอกได้ว่าเป็นการวัดความรู้ในเรื่องนั้น ๆ อย่างครอบคลุม (มีความเป็นตัวแทนหรือ representativeness หรือไม่) นอกจากนี้เครื่องมือมีความเสมือนจริง (authenticity) หรือไม่

โดยเฉพาะเครื่องมือที่ใช้ประเมินทักษะทางคลินิก สิ่งเหล่านี้ดูเหมือนไม่ยาก แต่ต้องระวังหลุมพราง ตัวอย่างเช่น เมื่อพิจารณาข้อสอบเรื่อง ยาชา อาจเป็นข้อสอบของรายวิชาเภสัชวิทยา หรือรายวิชาวิสัญญีวิทยาก็ได้ ขึ้นอยู่กับความซับซ้อนและวัตถุประสงค์ของโจทย์ หรือข้อสอบของรายวิชาทางศัลยศาสตร์ที่เกี่ยวกับผู้ป่วยที่มีภาวะ upper GI hemorrhage แต่โจทย์อาจไปพ้องกับการ approach ทางสาขาอายุรศาสตร์ ผู้ที่จะบอกได้ว่า โจทย์ข้อนี้ valid หรือไม่คือ อาจารย์ผู้สอน อาจารย์ผู้ออกข้อสอบ และ/หรือผู้เชี่ยวชาญด้านการศึกษา ถ้าไม่แน่ใจควรตั้งคณะกรรมการเฉพาะกิจเพื่อถ่วงถ่วงข้อสอบในเรื่อง validity โดยดูจากค่า content validity index ทั้งรายข้อ และทั้งชุด

ความเที่ยงหรือ **reliability** ของเครื่องมือประเมินเช่นกัน ขึ้นอยู่กับ test production เป็นหลัก ถ้าผู้สอนสร้างเครื่องมือที่มี objectivity ที่เพียงพอ คำตอบมีความจำเพาะและแน่นอน ข้อสอบมีจำนวนมากพอ ผู้เข้าสอบมีจำนวนมากพอ มีระยะเวลาในการทดสอบนานพอ (กรณีสอบเป็นรายทักษะ) อาจารย์ผู้คุมสอบมีความเป็นมืออาชีพ ไม่อคติ ไม่เอนเอียง จะทำให้ความเที่ยงมีค่าสูงขึ้นได้ ค่า reliability มีค่าตั้งแต่ 0 (ไม่มีความเที่ยง) ถึง 1 (ความเที่ยงสมบูรณ์) ข้อสอบที่ดีโดยเฉพาะข้อสอบที่มีความสำคัญหรือ high stake เช่น การสอบเพื่อรับใบอนุญาตประกอบวิชาชีพเวชกรรม ควรมีค่าความเที่ยงมากกว่า 0.8 นอกจากนี้การมีความเที่ยงที่ดีจะทำให้มี validity ที่ดีตามไปด้วย

การประเมินผลมีส่วนสำคัญในการกำหนดการเรียนการสอนด้วย เหมือนคำกล่าวที่ว่า “Assessment drives learning” นั่นคือ การประเมินผลมีผลกระทบต่อกระบวนการจัดการเรียนการสอนซึ่ง **educational impact** แสดงผลกระทบต่อ 4 ด้านคือ 1) เนื้อหาที่ประเมิน เช่น ถ้าประเมินเฉพาะระดับ “ความจำ” ผู้เรียนจะพยายามท่องจำอย่างเดียว ไม่พยายามทำความเข้าใจเนื้อหาหรือนำเนื้อหามาประยุกต์เพื่ออธิบายปรากฏการณ์ แต่ถ้าประเมินทักษะที่สูงขึ้น เช่น การประยุกต์ใช้ การวิเคราะห์ การประเมินค่า (ตาม Bloom’s taxonomy) ผู้เรียนจะเปลี่ยนวิธีการเรียนโดยศึกษาให้ถ่องแท้ ใช้ปัญหาเป็นฐาน ใช้ทฤษฎีมาอธิบายปรากฏการณ์ 2) รูปแบบการประเมิน เช่น การสอบรวบยอด comprehensive ปลายปี ทำให้ผู้เรียนต้องศึกษาทบทวนความรู้ทั้งหมดที่ศึกษาตลอดปี ไม่จำเพาะเจาะจงไปรายวิชาใดวิชาหนึ่ง 3) การนำผลการประเมินมาใช้ การนำผลการประเมินมาใช้ให้เกิดประโยชน์ เกิดจากการแปลผลการสอบให้มีความหมาย (meaningful interpretation of test results) และนำผลการแปลผลนั้นมาป้อนกลับให้ผู้เรียน อาจารย์ผู้สอน และผู้บริหารสถานศึกษาเพื่อใช้ในการปรับปรุงการเรียนการสอนต่อไป โดยการประเมิน formative assessment จะมีบทบาทมาก และ 4) เจ็อนไขของการประเมิน เช่น จำนวนครั้งของการสอบ สัดส่วนของคะแนนในการสอบแต่ละครั้ง ผู้สอบอาจไม่ให้น้ำหนักของความทุ่มเทแตกต่างกันตามสัดส่วนคะแนน การเปิดโอกาสในการสอบแก้ตัว เกณฑ์การเลื่อนชั้นปีจากผลการสอบ เกณฑ์การตัดสินผล สิ่งเหล่านี้จะเป็นปัจจัยในการกำหนดพฤติกรรมเรียนและการสอบ นอกจากนี้ negative effect ที่เกิดจากการสอบ เช่น จำนวนครั้งของการสอบที่ถึงจนไม่มีเวลาให้ผู้เรียนได้ศึกษาเนื้อหาเพื่อความเข้าใจ แต่เป็นการศึกษาเพื่อตอบคำตอบให้ได้คะแนนมากที่สุด ซึ่งจะนำไปสู่การท่องจำข้อสอบเก่า หรือการติวโดยรุ่นพี่หรือสถาบันกวดวิชา (ซึ่งหลายแห่งกวด “ข้อสอบเก่า”)

**ประสิทธิภาพ (efficiency)** หมายถึง การได้ผลลัพธ์ตามที่ตั้งไว้ในงบประมาณและเวลาที่เหมาะสม ดังนั้นการลงทุนใด ๆ ในการประเมินผลต้องคำนึงถึงความเป็นไปได้ (**feasibility**) ในด้านการเงินและการสิ้นเปลืองเวลา (และอาจรวมถึงบุคลากรด้วย) ผู้บริหารสถานศึกษาควรให้ความสำคัญ อย่างไรก็ตาม การจัดสอบ

บางอย่าง เช่น การสอบ OSCE จะใช้ทรัพยากรมาก ถ้าผู้เรียนมีจำนวนไม่มากการนำมาใช้ต้องคำนึงถึงความคุ้มค่าในการจัดสอบเองด้วย

เรื่องสุดท้ายคือ การใช้เครื่องมือประเมินนั้นเป็นที่ยอมรับในวงกว้าง (**acceptability**) หรือไม่ ในวงกว้างหมายถึง ผู้เรียน ผู้สอน และผู้บริหารสถานศึกษา ในประเด็นนี้มักถูกละเลยโดยมักคิดว่า ถ้าผู้สอนหรือผู้บริหารกำหนดการเลือกใช้เครื่องมือประเมินแล้วผู้เรียนน่าจะยอมรับได้ เพราะเป็นวิธีการที่ใช้มาโดยตลอด สิ่งเหล่านี้เกิดจาก “ความเชื่อ” และ “ประสบการณ์ส่วนตัว” ของทั้งผู้เรียนและผู้สอน โดยที่อาจไม่ได้ศึกษางานวิจัยหรือหลักฐานเชิงประจักษ์ทางวิทยาศาสตร์เกี่ยวกับการประเมินผลนั้น ๆ เช่น สมัยก่อนมักนิยมใช้คำถามในรูปแบบถูก-ผิด หรือ True/False format ซึ่งง่ายต่อการออกข้อสอบ ผู้เรียนสะดวกในการใช้วิธีการท่องจำ ต่อมาเมื่อมีงานวิจัยที่พบว่า ข้อสอบแบบนี้วัดผู้เรียนเก่ง-อ่อนไม่ได้ เพราะคาดเดาได้ง่าย บางครั้งข้อสอบมีความคลุมเครือ คำตอบอาจมีได้หลายแบบ ทำให้ข้อสอบชนิดนี้หมดความนิยมไป ซึ่งกว่าจะทำให้ “ความเชื่อ” นี้สลายไปต้องใช้เวลาานาน จนปัจจุบันแทบไม่มีการใช้ข้อสอบแบบนี้ในหลักสูตรแพทยศาสตรบัณฑิตแล้ว

การทำให้เครื่องมือหรือรูปแบบการประเมินเป็นที่ยอมรับต้องใช้ข้อมูลเชิงประจักษ์ในการทำความเข้าใจและเปลี่ยน “ความเชื่อ” ของผู้เรียน ผู้สอนและผู้บริหาร การมีเวทีหรือการประชุมที่รับฟังความคิดเห็นของผู้เรียนและผู้สอน โดยเฉพาะเมื่อมีการเปลี่ยนแปลงระบบ รูปแบบ หรือเครื่องมือการประเมิน ตลอดจนการอธิบายและทำความเข้าใจถึงประโยชน์ที่ได้จากการเปลี่ยนแปลงนั้นจะช่วยให้การประเมินผลเป็นที่ยอมรับมากขึ้น ลดแรงต้าน และสามารถนำไปใช้ได้อย่างราบรื่น ผู้นิพนธ์ขอแนะนำว่า คุณลักษณะของระบบ หรือเครื่องมือประเมินตามรูปที่ 1 เหล่านี้ควรบันทึกหลักฐานไว้ให้ครบในแต่ละประเด็นเพราะสามารถนำมาใช้อธิบาย *assessment utility* ตามเกณฑ์ของ WFME ได้

### แหล่งอ้างอิง

1. World Federation for Medical Education. Basic Medical Education WFME Global Standards for Quality Improvement. The 2015 Revision.
2. Van der Vleuten CPM, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39(3):309-17.
3. Van der Vleuten CP. The assessment of professional competence: developments, research, and practical implications. *Adv Health Sci Educ Theory Pract* 1996;1(1):41-67.
4. Yusof MSB. ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal* 2019; 11(2): 49-54.