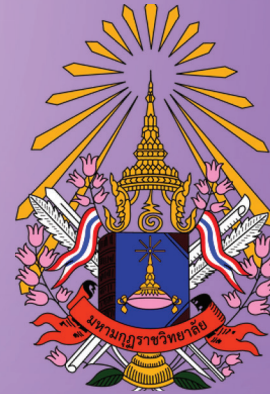




ED1006 การวัดและการประเมินผลการเรียนรู้ Learning Measurement and Evaluation



# การวัดและการประเมิน ผลการเรียนรู้

ED1006 Learning Measurement and Evaluation

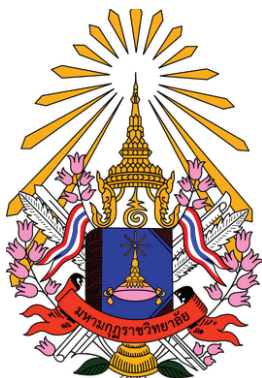


มหาวิทยาลัยมหามกุฏราชวิทยาลัย  
Mahamakut Buddhist University

ความเป็นเลิศทางวิชาการ  
ตามแนวทางพระพุทธศาสนา  
ACADEMIC EXCELLENCE  
BASED ON BUDDHISM



มหาวิทยาลัยมหามกุฏราชวิทยาลัย  
248 ม.1 ถ.ศาลายา -นครชัยศรี ต.ศาลายา  
อ.พุทธมณฑล จ.นครปฐม 73170  
โทร. 02-444-6000  
เว็บไซต์ : WWW.MBU.AC.TH



# **เอกสารประกอบการสอน รายวิชา ED1006**

## **การวัดและการประเมินผลการเรียนรู้ Learning Measurement and Evaluation**

**คณะศึกษาศาสตร์  
Faculty of Education  
มหาวิทยาลัยมหามกุฏราชวิทยาลัย  
Mahamakut Buddhist University  
ปีการศึกษา 2569**



## คำปรารภอธิการบดี

มหาวิทยาลัยมหามกุฏราชวิทยาลัย เป็นมหาวิทยาลัยทางพระพุทธศาสนา สังกัดอยู่ในกลุ่มที่ ๔ คือ กลุ่มพัฒนาปัญญาและคุณธรรมด้วยหลักศาสนา เป็นสถาบันหลักในการเผยแผ่พระพุทธศาสนาระดับสูงและการค้นคว้าหลักคำสอนเพื่อนำไปประยุกต์ใช้ในการดำเนินชีวิต หนึ่งในยุทธศาสตร์ที่สำคัญของการบริหารและพัฒนามหาวิทยาลัยนั้นคือ การพัฒนาสู่ความเป็นเลิศทางวิชาการทางพระพุทธศาสนาในระดับสากล ด้วยเหตุนี้มหาวิทยาลัย จึงตระหนักถึงความสำคัญของเอกสารประกอบการสอนรายวิชาชีพครู (ED) ในฐานะที่เป็นหลักสูตรผลิตครูผู้ช่วยของมหาวิทยาลัย โดยมีความสอดคล้องกับหลักสูตรที่เปิดการเรียนการสอนในวิทยาเขตและวิทยาลัยทั่วประเทศ

การผลิตเอกสารประกอบการสอนถือเป็นจุดเริ่มต้นของการพัฒนาคุณภาพการศึกษา เพราะจะทำให้มีทิศทางในการประเมินผลลัพธ์การเรียนรู้ของนักศึกษาให้เป็นไปตามวัตถุประสงค์ของหลักสูตร และสอดคล้องกับวิสัยทัศน์และพันธกิจของมหาวิทยาลัย อีกทั้งยังช่วยพัฒนาทักษะความเชี่ยวชาญของอาจารย์ผู้สอนให้สามารถถ่ายทอดองค์ความรู้ได้อย่างมีประสิทธิภาพและมีมาตรฐานเช่นเดียวกัน นอกจากนั้นเอกสารประกอบการสอนที่ทำขึ้นนี้ยังสามารถขยายผลไปสู่ระดับเอกสารวิชาการที่สูงขึ้นไป เช่น เอกสารคำสอน หนังสือ หรือตำรา เป็นต้น ซึ่งรายวิชาวิชาชีพครู (ED) สามารถเรียนร่วมกันและออกข้อสอบกลาง โดยตรวจสอบจากรายละเอียดของ PLO, CLO และ LLO ว่าเป็นไปตามมาตรฐานหลักสูตรแบบ OBE

เอกสารประกอบการสอนเล่มนี้ จะเป็นประโยชน์ต่อการจัดการศึกษาในหลักสูตรและเป็นประโยชน์ต่อการพัฒนาคุณภาพของนักศึกษาและคณาจารย์ให้มีความเข้าใจรายวิชา และยังสะดวกต่อการปรับปรุงเนื้อหาองค์ความรู้ให้ทันสมัย เหมาะสมกับบริบททางสังคมที่เปลี่ยนแปลงไปอย่างต่อเนื่อง ขออนุโมทนาชื่นชมคณะกรรมการบริหารจัดการตำรา เอกสารประกอบการสอนและสื่อการเรียนการสอนรายวิชาชีพครู (ED) และคณะทำงานคณะศึกษาศาสตร์ ผู้ทรงคุณวุฒิ ทุกรูป/คน ที่ร่วมกันทำเอกสารวิชาการจนเสร็จสมบูรณ์ในครั้งนี้



(พระพรหมวชิรเมธาภรณ์, รศ.ดร.)

อธิการบดีมหาวิทยาลัยมหามกุฏราชวิทยาลัย

## คำนำ

เอกสารประกอบการสอนรายวิชา ED1006 การวัดและการประเมินผลการเรียนรู้ เล่มนี้ เรียบเรียงขึ้นเพื่อใช้ประกอบการจัดการเรียนการสอนรายวิชาในกลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย โดยพัฒนาเนื้อหาให้สอดคล้องกับผลลัพธ์การเรียนรู้ของรายวิชา(Course Learning Outcomes) ตามแนวทางการจัดการศึกษาที่เน้นผลลัพธ์ (Outcome-Based Education) และหลักความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) เนื้อหาแบ่งออกเป็น 14 บท ครอบคลุมตั้งแต่ความรู้พื้นฐานของการวัดและการประเมินผล แนวคิดและประเภทของการประเมิน การกำหนดจุดประสงค์การเรียนรู้และการจำแนกพฤติกรรม การสร้างและการตรวจสอบคุณภาพเครื่องมือวัด การวัดและประเมินด้านพุทธิพิสัย ทักษะพิสัย และจิตพิสัย การประเมินตามสภาพจริงและการสร้างเกณฑ์การให้คะแนน สถิติพื้นฐานเพื่อการวัดและประเมินผล การตัดสินใจและการรายงานผลการเรียนรู้ ตลอดจนการใช้เทคโนโลยีและจริยธรรมในการวัดและประเมินผล

ผู้เรียบเรียงหวังเป็นอย่างยิ่งว่า เอกสารเล่มนี้จะเป็นประโยชน์ต่อนักศึกษาวิชาชีพครูและผู้สนใจ ในการนำความรู้ด้านการวัดและการประเมินผลไปประยุกต์ใช้เพื่อพัฒนาผู้เรียนได้อย่างมีคุณภาพและเป็นธรรม หากมีข้อเสนอแนะประการใด ผู้เรียบเรียงยินดีน้อมรับเพื่อปรับปรุงแก้ไขให้สมบูรณ์ยิ่งขึ้นต่อไป

อาจารย์ประจำรายวิชา

คณะผู้จัดทำ

## สารบัญ

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| คำนำ.....                                                                    | ก  |
| สารบัญ.....                                                                  | ข  |
| <b>บทที่ 1 ความรู้พื้นฐานการวัดและการประเมินผลการเรียนรู้.....</b>           |    |
| ความคิดรวบยอด.....                                                           | 2  |
| จุดประสงค์การเรียนรู้ประจำบท.....                                            | 2  |
| เนื้อหาวิชา.....                                                             | 2  |
| เอกสารอ้างอิงประจำบท.....                                                    | 15 |
| <b>บทที่ 2 แนวคิด หลักการ และประเภทของการวัดและประเมินผลการเรียนรู้.....</b> |    |
| ความคิดรวบยอด.....                                                           | 20 |
| จุดประสงค์การเรียนรู้ประจำบท.....                                            | 20 |
| คำสำคัญ.....                                                                 | 21 |
| เนื้อหาวิชา.....                                                             | 21 |
| เอกสารอ้างอิงประจำบท.....                                                    | 33 |
| <b>บทที่ 3 จุดประสงค์การเรียนรู้และการจำแนกพฤติกรรม.....</b>                 |    |
| ความคิดรวบยอด.....                                                           | 37 |
| เนื้อหาวิชา.....                                                             | 39 |
| เอกสารอ้างอิงประจำบท.....                                                    | 54 |
| <b>บทที่ 4 หลักการสร้างและเขียนข้อสอบวัดผลสัมฤทธิ์.....</b>                  |    |
| ความคิดรวบยอด.....                                                           | 56 |
| จุดประสงค์การเรียนรู้ประจำบท.....                                            | 57 |
| คำสำคัญ.....                                                                 | 59 |
| เนื้อหาวิชา.....                                                             | 60 |
| <b>บทที่ 5 การวัดและประเมินทักษะปฏิบัติ.....</b>                             |    |
| ความคิดรวบยอด.....                                                           | 78 |

## สารบัญ (ต่อ)

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| จุดประสงค์การเรียนรู้ประจำบท.....                                       | 80  |
| คำสำคัญ.....                                                            | 81  |
| สรุปบทเรียน.....                                                        | 95  |
| เอกสารอ้างอิงประจำบท.....                                               | 97  |
| <b>บทที่ 6 การวัดและประเมินด้านจิตพิสัยและเจตคติ.....</b>               |     |
| ความคิดรวบยอด.....                                                      | 102 |
| จุดประสงค์การเรียนรู้ประจำบท.....                                       | 102 |
| คำสำคัญ.....                                                            | 103 |
| เนื้อหาวิชา.....                                                        | 103 |
| <b>บทที่ 7 ความตรงของเครื่องมือวัดและประเมินผล.....</b>                 |     |
| ความคิดรวบยอด.....                                                      | 123 |
| เนื้อหาวิชา.....                                                        | 124 |
| สรุปบทเรียน.....                                                        | 134 |
| <b>บทที่ 8 ความเที่ยงของเครื่องมือวัดและประเมินผล.....</b>              |     |
| จุดประสงค์การเรียนรู้ประจำบท.....                                       | 139 |
| เนื้อหาวิชา.....                                                        | 140 |
| สรุปบทเรียน.....                                                        | 148 |
| <b>บทที่ 9 การวิเคราะห์ข้อสอบรายข้อ: ความยาก-อำนาจจำแนก-ตัวลวง.....</b> | 153 |
| เนื้อหาวิชา.....                                                        | 154 |
| สรุปบทเรียน.....                                                        | 162 |
| <b>บทที่ 10 การประเมินตามสภาพจริงและการสร้างรูบรีค.....</b>             |     |
| ความคิดรวบยอด.....                                                      | 167 |
| เนื้อหาวิชา.....                                                        | 168 |
| สรุปบทเรียน.....                                                        | 177 |

## สารบัญ (ต่อ)

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| <b>บทที่ 11 การใช้ผลการประเมินเพื่อพัฒนาผู้เรียน.....</b>             |     |
| ความคิดรวบยอด.....                                                    | 182 |
| เนื้อหาวิชา.....                                                      | 183 |
| สรุปบทเรียน.....                                                      | 190 |
| เอกสารอ้างอิงประจำบท.....                                             | 192 |
| <b>บทที่ 12 สถิติพื้นฐานเพื่อการวัดและประเมินผล.....</b>              |     |
| ความคิดรวบยอด.....                                                    | 195 |
| เนื้อหาวิชา.....                                                      | 196 |
| กิจกรรมการเรียนรู้การสอนและใบงาน.....                                 | 204 |
| เอกสารอ้างอิงประจำบท.....                                             | 206 |
| <b>บทที่ 13 การตัดสินผล การให้ระดับผลการเรียน และการรายงานผล.....</b> |     |
| ความคิดรวบยอด.....                                                    | 209 |
| คำสำคัญ.....                                                          | 210 |
| เนื้อหาวิชา.....                                                      | 210 |
| สรุปบทเรียน.....                                                      | 218 |
| เอกสารอ้างอิงประจำบท.....                                             | 220 |
| <b>บทที่ 14 การใช้เทคโนโลยีและจริยธรรมในการวัดและประเมินผล.....</b>   |     |
| ความคิดรวบยอด.....                                                    | 223 |
| จุดประสงค์การเรียนรู้ประจำบท.....                                     |     |
| คำสำคัญ.....                                                          | 224 |
| เนื้อหาวิชา.....                                                      | 224 |
| เอกสารอ้างอิงประจำบท.....                                             | 234 |



## บทที่ 1

### ความรู้พื้นฐานการวัดและการประเมินผลการเรียนรู้

#### *Fundamentals of Learning Measurement and Evaluation*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 1 จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

| รายการ                    | รายละเอียด                                                                  |
|---------------------------|-----------------------------------------------------------------------------|
| รหัส-รายวิชา              | ED1006 การวัดและการประเมินผลการเรียนรู้                                     |
| คณะ/หลักสูตร              | คณะศึกษาศาสตร์ กลุ่มวิชาชีพครู มหาวิทยาลัยมหาจุฬาราชวิทยาลัย                |
| จำนวนคาบ                  | 4 คาบ (บรรยาย 2 คาบ และปฏิบัติ/อภิปราย 2 คาบ)                               |
| ผลลัพธ์การเรียนรู้ประจำบท | CLO1 (อธิบายหลักการและความสำคัญ) และ CLO5 (คุณธรรม จริยธรรม ความเที่ยงธรรม) |
| ระดับพฤติกรรมหลัก         | จำ-เข้าใจ (พุทธิพิสัย) และ รับรู้-เห็นคุณค่า (จิตพิสัย)                     |

#### แผนบริหารการสอนประจำบทที่ 1

##### หัวข้อสาระการจัดการเรียนการสอน

- องค์ประกอบของการจัดการศึกษาและวงจรการวัดและประเมินผล (O-L-E)
- ความหมายของการทดสอบ การวัดผล การประเมิน และการตัดสินคุณค่า
- ธรรมชาติของการวัดผลการศึกษาและมาตรการวัด 4 ระดับ
- ประเภทและการเลือกใช้เครื่องมือวัดและประเมินผล
- คุณภาพของเครื่องมือวัดและประเมินผล: ความตรง ความเที่ยง ความเป็นปรนัย
- จริยธรรม จรรยาบรรณ และแนวโน้มการประเมินในศตวรรษที่ 21

## ความคิดรวบยอด

การวัดและการประเมินผลเป็นองค์ประกอบสำคัญของกระบวนการจัดการศึกษา ทำหน้าที่เป็นฟันเฟืองสุดท้ายในวงจร “จุดมุ่งหมาย-การจัดการเรียนรู้-การประเมินผล (O-L-E)” การวัดผล (Measurement) คือการกำหนดตัวเลขหรือสัญลักษณ์แทนคุณลักษณะของผู้เรียนตามกฎเกณฑ์ ส่วนการประเมินผล (Evaluation) คือการนำผลการวัดมาเทียบเกณฑ์เพื่อตัดสินคุณค่า ครูมืออาชีพจึงต้องเลือกใช้เครื่องมือให้ตรงกับพฤติกรรมที่ต้องการวัด คำนึงถึงคุณภาพของเครื่องมือทั้งความตรง ความเที่ยง และความเป็นปรนัยละเอียดยึดมั่นในคุณธรรมความยุติธรรมเพื่อให้ผลการประเมินน่าเชื่อถือและนำไปพัฒนาผู้เรียนได้จริง

## จุดประสงค์การเรียนรู้

1. อธิบายองค์ประกอบของการจัดการศึกษาและความสัมพันธ์ของวงจร O-L-E ได้ (CLO1)
2. จำแนกความหมายของการทดสอบ การวัดผล การประเมิน และการตัดสินคุณค่าได้ถูกต้อง (CLO1)
3. อธิบายธรรมชาติของการวัดผลการศึกษาและระดับมาตรฐานการวัดทั้ง 4 ระดับได้ (CLO1)
4. เลือกประเภทเครื่องมือวัดและประเมินให้เหมาะสมกับด้านพฤติกรรมที่ต้องการวัดได้ (CLO1)
5. อธิบายคุณภาพของเครื่องมือ (ความตรง ความเที่ยง ความเป็นปรนัย) พร้อมยกตัวอย่างข้อผิดพลาดที่พบบ่อยได้ (CLO1)
6. ตระหนักและแสดงออกถึงคุณธรรม จริยธรรม และความยุติธรรมในการประเมินผ่านการวิเคราะห์กรณีศึกษา (CLO5)

## เนื้อหาวิชา

เนื้อหาวิชาโดยละเอียดปรากฏในหัวข้อ 1.5 เนื้อหาวิชา ของเอกสารประกอบการสอนบทที่ 1 ซึ่งแบ่งเป็น 6 หัวข้อย่อย (1.5.1-1.5.6) ครอบคลุมตั้งแต่องค์ประกอบของการจัดการศึกษา ความหมายของศัพท์พื้นฐาน ธรรมชาติและมาตรฐานการวัด ประเภทเครื่องมือ คุณภาพเครื่องมือ จนถึงจริยธรรมและแนวโน้มร่วมสมัย

## กิจกรรมการเรียนรู้

1. ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน 10 ข้อ และอภิปรายประเด็น “ถ้าครูใช้ไม้บรรทัดวัดความฉลาด จะเกิดอะไรขึ้น”
2. ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ + กิจกรรมคิด-จับคู่-แลกเปลี่ยน (Think-Pair-Share) จำแนกการวัด/การประเมิน และเลือกเครื่องมือให้ตรงด้านพฤติกรรม

3. ชั้นระหว่างเรียน: วิเคราะห์กรณีศึกษาจริยธรรมการให้คะแนนเป็นกลุ่ม เชื่อมโยงหลักความเป็นปรนัยและความยุติธรรม
4. ชั้นหลังเรียน: เขียนบัตรออก (Exit Ticket) สรุป “ความยุติธรรมในการวัดผลของครูมืออาชีพ” และทำแบบทดสอบหลังเรียน

### สื่อการเรียนการสอน

- เอกสารประกอบการสอนบทที่ 1 และสไลด์นำเสนอประเด็นสำคัญ
- ตำราหลัก: พิชรินทร์ ชมภูวิเศษ (2559) และสื่อวีดิทัศน์ตัวอย่างการให้คะแนนด้วยรูบรีค
- ใบงานจำแนกการวัด-การประเมิน และใบกิจกรรมกรณีศึกษาจริยธรรม
- เว็บไซต์สถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) และฐานข้อมูลวารสาร ThaiJO

### การวัดและประเมินผล

| ผลลัพธ์การเรียนรู้                   | วิธีวัด                                        | เครื่องมือ                          |
|--------------------------------------|------------------------------------------------|-------------------------------------|
| CLO1 อธิบายหลักการและความสำคัญ       | แบบทดสอบก่อน-หลังเรียน และแบบฝึกหัดท้ายบท      | แบบทดสอบปรนัย/อัตนัย (เฉลยตามเกณฑ์) |
| CLO5 คุณธรรม จริยธรรม ความเที่ยงธรรม | วิเคราะห์กรณีศึกษา และสังเกตพฤติกรรมการอภิปราย | แบบสังเกตพฤติกรรม/รูบรีคจรรยาบรรณ   |

## 1.1 ความคิดรวบยอด

## 1.2 จุดประสงค์การเรียนรู้ประจำบท

การจัดการเรียนการสอนทุกครั้งประกอบด้วยองค์ประกอบสำคัญสามส่วนที่เชื่อมโยงกันเป็นวงจร ได้แก่ จุดมุ่งหมายการเรียนรู้ (Objective) การจัดประสบการณ์การเรียนรู้ (Learning Experience) และการวัดและประเมินผล (Evaluation) การวัดและประเมินผลจึงมิใช่เพียงปลายทางของการสอน แต่เป็นกลไกสะท้อนกลับที่บอกว่าผู้เรียนบรรลุจุดมุ่งหมายเพียงใด และช่วยให้ครูปรับปรุงการสอนได้ตรงจุด (พิชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

ในเชิงวิชาการต้องแยกแยะศัพท์สื่คำให้ชัดเจน การทดสอบ (Testing) คือการใช้สิ่งเร้ากระตุ้นให้ผู้เรียนแสดงพฤติกรรมออกมา การวัดผล (Measurement) คือการกำหนดตัวเลขหรือสัญลักษณ์แทนพฤติกรรมนั้นตามกฎเกณฑ์ การประเมิน (Assessment) คือการรวบรวมหลักฐานอย่างเป็นระบบเพื่อพัฒนาผู้เรียน และการ

ประเมินผล (Evaluation) คือการตัดสินคุณค่าโดยเทียบผลการวัดกับเกณฑ์ ครูที่เข้าใจความแตกต่างนี้จะเลือกเครื่องมือได้เหมาะสม คำนึงถึงคุณภาพของเครื่องมือทั้งความตรง (Validity) ความเที่ยง (Reliability) และความเป็นปรนัย (Objectivity) ตลอดจนยึดมั่นคุณธรรมความยุติธรรม เพื่อให้ผลการประเมินน่าเชื่อถือและนำไปใช้พัฒนาผู้เรียนได้อย่างแท้จริงตามเจตนารมณ์ของการจัดการศึกษาที่เน้นผลลัพธ์ (Outcome-Based Education) **เมื่อศึกษาบทที่ 1 จบแล้ว นักศึกษาสามารถ:**

1. อธิบายองค์ประกอบของการจัดการศึกษาและความสัมพันธ์ของวงจร O-L-E ได้
2. จำแนกความหมายของการทดสอบ การวัดผล การประเมิน และการตัดสินคุณค่าได้ถูกต้อง
3. อธิบายธรรมชาติของการวัดผลการศึกษาและระดับมาตรฐานการวัดทั้ง 4 ระดับได้
4. เลือกประเภทเครื่องมือวัดและประเมินให้เหมาะสมกับด้านพฤติกรรมที่ต้องการวัดได้
5. อธิบายคุณภาพของเครื่องมือ (ความตรง ความเที่ยง ความเป็นปรนัย) พร้อมยกตัวอย่างข้อผิดพลาดที่พบบ่อยได้
6. ตระหนักและแสดงออกถึงคุณธรรม จริยธรรม และความยุติธรรมในการประเมินผ่านการวิเคราะห์กรณีศึกษา

#### ตารางที่ 1.1 ความสอดคล้องระหว่างจุดประสงค์ประจำบท ผลลัพธ์การเรียนรู้ (CLO) และการวัดประเมิน

| จุดประสงค์ข้อที่ | CLO  | ระดับพฤติกรรม                | วิธีวัด / เครื่องมือ                            |
|------------------|------|------------------------------|-------------------------------------------------|
| 1-5              | CLO1 | จำ-เข้าใจ (พุทธิพิสัย)       | แบบทดสอบก่อน-หลังเรียน, แบบฝึกหัดท้ายบท         |
| 6                | CLO5 | รับรู้-เห็นคุณค่า (จิตพิสัย) | วิเคราะห์กรณีศึกษา + แบบสังเกต/ รุบริคจรรยาบรรณ |

ที่มา: ปรับจากผังความสอดคล้อง (Constructive Alignment) รายวิชา ED1006

### 1.3 คำสำคัญ

| คำสำคัญ (ไทย)        | Keyword     | คำสำคัญ (ไทย) | Keyword               |
|----------------------|-------------|---------------|-----------------------|
| การวัดผล             | Measurement | ความตรง       | Validity              |
| การประเมินผล         | Evaluation  | ความเที่ยง    | Reliability           |
| การประเมินเพื่อพัฒนา | Assessment  | ความเป็นปรนัย | Objectivity           |
| การทดสอบ             | Testing     | มาตรการวัด    | Scales of Measurement |

|                 |                               |                            |                         |
|-----------------|-------------------------------|----------------------------|-------------------------|
| การตัดสินคุณค่า | Value Judgment                | การประเมินเพื่อการเรียนรู้ | Assessment for Learning |
| วงจร O-L-E      | Objective-Learning-Evaluation | จรรยาบรรณการประเมิน        | Ethics in Assessment    |

## 1.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบจัดทำในระบบทดสอบออนไลน์แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

## 1.5 เนื้อหาวิชา

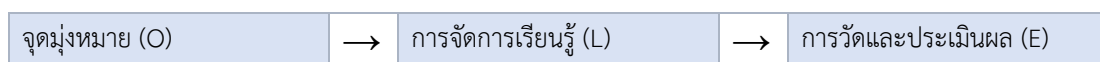
เนื้อหาในบทนี้แบ่งเป็น 6 หัวข้อย่อย โดยเดินเรื่องจากภาพรวมของกระบวนการจัดการศึกษาไปสู่ความหมายของศัพท์พื้นฐาน ธรรมชาติและมาตรการวัด ประสิทธิภาพและคุณภาพของเครื่องมือ และปิดท้ายด้วยจริยธรรมและแนวโน้มร่วมสมัย แต่ละหัวข้อนำเสนอตามลำดับ นิยาม → คำอธิบายเชิงลึก → ตัวอย่างในชั้นเรียนไทย → การประยุกต์ใช้และข้อควรระวัง เพื่อให้ นักศึกษาครูเชื่อมโยงทฤษฎีกับการฝึกประสบการณ์วิชาชีพได้

### 1.5.1 องค์ประกอบของการจัดการศึกษาและวงจรการวัดและประเมินผล (O-L-E)

การจัดการศึกษามีเป้าหมายสูงสุดเพื่อพัฒนาผู้เรียนให้เป็น “มนุษย์ที่สมบูรณ์” ทั้งร่างกาย จิตใจ สติปัญญา ความรู้ และคุณธรรม สามารถอยู่ร่วมกับผู้อื่นได้อย่างมีความสุข (พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542) เพื่อบรรลุเป้าหมายนี้ การจัดการเรียนการสอนทุกครั้งต้องดำเนินผ่านองค์ประกอบสำคัญ 3 ส่วนที่สัมพันธ์กันเป็นวงจร เรียกโดยย่อว่า วงจร O-L-E ได้แก่ จุดมุ่งหมาย (Objective) การจัดประสบการณ์การเรียนรู้ (Learning Experience) และการประเมินผล (Evaluation) (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

องค์ประกอบที่หนึ่ง จุดมุ่งหมาย (O) คือการกำหนดทิศทางว่าต้องการให้ผู้เรียนเปลี่ยนแปลงพฤติกรรมในด้านใด ครอบคลุมทั้งด้านพุทธิพิสัย (ความรู้–ความคิด) จิตพิสัย (เจตคติ–ค่านิยม) และทักษะพิสัย (การปฏิบัติ) องค์ประกอบที่สอง การจัดประสบการณ์การเรียนรู้ (L) คือกิจกรรมและวิธีสอนที่ครูออกแบบเพื่อให้ผู้เรียนบรรลุจุดมุ่งหมาย และองค์ประกอบที่สาม การประเมินผล (E) คือกระบวนการตรวจสอบว่าผู้เรียนบรรลุจุดมุ่งหมายมากน้อยเพียงใด แล้วส่งข้อมูลย้อนกลับไปปรับปรุงทั้งจุดมุ่งหมายและการสอน ครูจึงต้องวัดและประเมินผลทุกครั้งที่มีการสอน มิใช่ทำเฉพาะปลายภาค (ขวาล แพรัตกุล, 2520)

#### ภาพที่ 1.1 วงจรการจัดการศึกษา O–L–E และวงจรย้อนกลับเพื่อพัฒนา



ผลการประเมิน (E) ย้อนกลับเพื่อปรับปรุงจุดมุ่งหมาย (O) และการจัดการเรียนรู้ (L) อย่างต่อเนื่อง

ที่มา: ดัดแปลงจาก พิชรินทร์ ชมภูวิเศษ (2559, น. 8)

ตัวอย่างในชั้นเรียนไทย ครูคณิตศาสตร์ชั้น ป.5 ตั้งจุดมุ่งหมายว่า “นักเรียนบวกเศษส่วนที่มีตัวส่วนไม่เท่ากันได้” (O) จากนั้นจัดกิจกรรมใช้สื่อแท่งเศษส่วนและแบบฝึก (L) แล้วทดสอบย่อยท้ายคาบเพื่อตรวจว่านักเรียนทำได้หรือไม่ (E) หากพบว่าครึ่งห้องยังหาตัวส่วนร่วมไม่ได้ ครูย่อมนำผลนี้กลับไปปรับกิจกรรมในคาบถัดไป นี่คือหัวใจของวงจรย้อนกลับ (Feedback Loop) ที่ทำให้การสอนพัฒนาขึ้นเรื่อย ๆ ดังภาพที่ 1.1

ในการจัดการศึกษาที่เน้นผลลัพธ์ (Outcome-Based Education: OBE) ความสัมพันธ์ของทั้งสามองค์ประกอบถูกเรียกว่า “ความสอดคล้องเชิงสร้างสรรค์” (Constructive Alignment) กล่าวคือ วิธีการประเมินต้องสอดคล้องกับผลลัพธ์การเรียนรู้ที่ประกาศไว้และกับกิจกรรมการเรียนรู้ (Biggs & Tang, 2011) ครูที่เข้าใจหลักนี้จะออกแบบการประเมินให้ “วัดในสิ่งที่สอนและสอนในสิ่งที่จะวัด” การรู้เรื่องการประเมิน (Assessment Literacy) จึงเป็นสมรรถนะพื้นฐานที่ครูทุกคนต้องมีเพื่อขับเคลื่อนการศึกษาตามแนว OBE (พิเชษฐ์ พิณีจ, อนุศิษฐ์ อ้นมานะตระกูล, และเอกรัตน์ รวยรวย, 2566)

ข้อควรระวัง ความผิดพลาดที่พบบ่อยคือการประเมินที่ไม่สอดคล้องกับจุดมุ่งหมาย เช่น ตั้งจุดมุ่งหมายให้ผู้เรียน “คิดวิเคราะห์” แต่ข้อสอบกลับวัดเพียง “ความจำ” ทำให้ผลการประเมินไม่สะท้อนสิ่งที่ตั้งใจพัฒนา ครูจึงควรตรวจสอบความสอดคล้องระหว่าง O–L–E ทุกครั้งก่อนนำเครื่องมือไปใช้จริง

### 1.5.2 ความหมายของการทดสอบ การวัดผล การประเมิน และการตัดสินคุณค่า

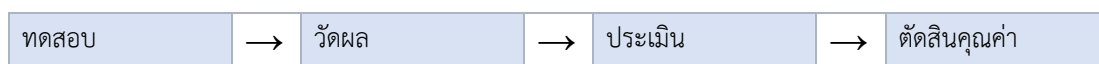
ศัพท์พื้นฐานสี่คำที่ต้องแยกให้ชัด ได้แก่ การทดสอบ (Testing) การวัดผล (Measurement) การประเมิน (Assessment) และการประเมินผลหรือการตัดสินคุณค่า (Evaluation) โดยมีความสัมพันธ์เป็นลำดับ กล่าวคือ การทดสอบใช้สิ่งเร้ากระตุ้นให้ผู้เรียนแสดงพฤติกรรม การวัดผลกำหนดตัวเลขแทนพฤติกรรมนั้นตามกฎเกณฑ์ การประเมินรวบรวมและเรียบเรียงสารสนเทศจากหลายแหล่ง และการประเมินผลตัดสินคุณค่าโดยเทียบกับเกณฑ์ ดังสรุปเปรียบเทียบในตารางที่ 1.2 และลำดับความสัมพันธ์ในภาพที่ 1.2 (พัชรินทร์ ชมภูวิเศษ, 2559; พิชิต ฤทธิ์จรูญ, 2564) ทั้งนี้ นิยามเชิงลึกของแต่ละแนวคิด หลักการสำคัญของการวัดและประเมินผลที่มีคุณภาพ และประเภทของการประเมินศึกษาโดยละเอียดในบทที่ 2 (หัวข้อ 2.5.1)

ตารางที่ 1.2 เปรียบเทียบการทดสอบ การวัดผล การประเมิน และการประเมินผล

| ประเด็น   | การทดสอบ (Testing)             | การวัดผล (Measurement)   | การประเมิน/ประเมินผล                   |
|-----------|--------------------------------|--------------------------|----------------------------------------|
| คำถามหลัก | ใช้สิ่งเร้าอะไรกระตุ้นพฤติกรรม | ได้ปริมาณ/ตัวเลขเท่าใด   | มีคุณค่า/อยู่ระดับใดเมื่อเทียบกับเกณฑ์ |
| ผลลัพธ์   | พฤติกรรมตอบสนอง                | ข้อมูลดิบเชิงตัวเลข      | ค่าตัดสินคุณค่า/การตัดสินใจ            |
| ตัวอย่าง  | ให้ทำข้อสอบ 50 ข้อ             | ตอบถูก 45 ข้อ = 45 คะแนน | $45/50 =$ ระดับ “ดีเยี่ยม” ผ่านเกณฑ์   |

ที่มา: ผู้เขียน สังเคราะห์จาก พัชรินทร์ ชมภูวิเศษ (2559, น. 10–12) และ พิชิต ฤทธิ์จรูญ (2564)

ภาพที่ 1.2 ลำดับความสัมพันธ์จากการทดสอบสู่การตัดสินคุณค่า



ที่มา: ผู้เขียน

ตัวอย่างในชั้นเรียนไทย ครูภาษาไทยให้นักเรียนทำข้อสอบอ่านจับใจความ 50 ข้อ (ทดสอบ) นักเรียนชื่อนาย ก ตอบถูก 45 ข้อ ครูนับทิกเป็น 45 คะแนน (วัดผล) จากนั้นนำ 45 คะแนนไปเทียบเกณฑ์ของโรงเรียนที่กำหนดว่า 80% ขึ้นไปถือว่า “ดีเยี่ยม” (ประเมิน) และสรุปว่านาย ก อยู่ในระดับดีเยี่ยมและผ่านรายวิชา (ตัดสินคุณค่า) จะเห็นว่าตัวเลข 45 เพียงอย่างเดียวยังไม่มีความหมาย จนกว่าจะถูกนำไปเทียบกับเกณฑ์ ดังภาพที่ 1.2

ข้อควรระวัง ครูจำนวนมากใช้คำว่า “วัดผล” และ “ประเมินผล” สลับกันจนเข้าใจคลาดเคลื่อน การจำไว้ว่า “วัด = ได้ตัวเลข, ประเมิน = ตัดสินค่าด้วยเกณฑ์” จะช่วยให้ออกแบบการสอบและการตัดสินผลได้รัดกุมขึ้น อีกทั้งควรตระหนักว่าการประเมินเพื่อพัฒนา (Assessment) กับการประเมินเพื่อตัดสิน (Evaluation) มีจุดเน้น

ต่างกัน หากใช้ผิดจุดประสงค์ เช่น เอาคะแนนสอบย่อยระหว่างเรียนไปตัดสินเกรดทันที ก็อาจบั่นทอนโอกาสพัฒนาของผู้เรียน

### 1.5.3 ธรรมชาติของการวัดผลการศึกษาและมาตรการวัด 4 ระดับ

การวัดผลการศึกษาแตกต่างจากการวัดทางกายภาพ เช่น การชั่งน้ำหนักหรือวัดส่วนสูง เพราะคุณลักษณะที่ครูต้องการวัด เช่น ความรู้ ความคิด เจตคติ ล้วนเป็นนามธรรมที่ซ่อนอยู่ภายในตัวผู้เรียน ครูจึงต้องเข้าใจธรรมชาติของการวัดผลการศึกษาเพื่อใช้และตีความผลอย่างระมัดระวัง ธรรมชาติสำคัญมีดังนี้ (บุญธรรมกิจปรีดาบริสุทธิ์, 2549; พชรินทร์ ชมภูวิเศษ, 2559)

- เป็นการวัดทางอ้อม (Indirect) วัดผ่านพฤติกรรมที่แสดงออก ไม่อาจวัดสติปัญญาหรือจิตใจได้โดยตรง
- เป็นการวัดเชิงสัมพัทธ์ (Relative) ผลการวัดมีความหมายต่อเมื่อนำไปเทียบกับกลุ่มหรือเกณฑ์
- เป็นการวัดที่ไม่สมบูรณ์ (Incomplete) วัดได้เพียงตัวอย่างของพฤติกรรม เพราะข้อจำกัดด้านเวลาและเนื้อหา
- ไม่มีศูนย์แท้ (No Absolute Zero) ได้ 0 คะแนน ไม่ได้แปลว่าผู้เรียนไม่มีความรู้ในเรื่องนั้นเลย
- มีความคลาดเคลื่อนเสมอ (Error) คะแนนที่ได้ = คะแนนจริง + ความคลาดเคลื่อนจากการวัด (ศิริชัย กาญจนวาสี, 2552)
- ต้องใช้วิธีวัดที่หลากหลาย เพราะพฤติกรรมมนุษย์ซับซ้อนและแตกต่างกันในแต่ละบุคคล/สถานการณ์

#### กล่องที่ 1.1 สมการคะแนนจากการวัด (Classical Test Theory)

$$X = T + E$$

X (Obtained Score) = คะแนนที่ผู้เรียนได้รับจริงจากการสอบ

T (True Score) = คะแนนที่เกิดจากความรู้ความสามารถที่แท้จริง

E (Error Score) = ความคลาดเคลื่อนจากเครื่องมือ/สภาพการสอบ

ครูที่ลดความคลาดเคลื่อน (E) ด้วยข้อสอบที่ดีและการคุมสอบที่รัดกุม  
จะได้คะแนนที่เข้าใกล้ความรู้จริงของผู้เรียนมากขึ้น

ที่มา: พชรินทร์ ชมภูวิเศษ (2559, น. 15)

นอกจากธรรมชาติข้างต้นแล้ว ผลการวัดที่อยู่ในรูปตัวเลขยังจัดอยู่ใน "มาตรการวัด" (Scales of Measurement) 4 ระดับ (Stevens, 1946) ได้แก่ นามบัญญัติ เรียงอันดับ อันตรภาค และอัตราส่วน ซึ่งเป็นตัวกำหนดขอบเขตของการแปลความหมายและสถิติที่ใช้ได้กับคะแนน รายละเอียด ตัวอย่างในชั้นเรียน และ

ตารางเปรียบเทียบทั้งสี่ระดับ ศึกษาในบทที่ 2 (หัวข้อ 2.5.2) ส่วนการเลือกใช้สถิติให้สอดคล้องกับระดับการวัด ศึกษาในบทที่ 12 (หัวข้อ 12.5.1)

#### 1.5.4 ประเภทและการเลือกใช้เครื่องมือวัดและประเมินผล

เป้าหมายของการศึกษาคือการพัฒนาผู้เรียนครบทั้งสามด้านของพฤติกรรม ได้แก่ ด้านพุทธิพิสัย (Cognitive Domain) ด้านจิตพิสัย (Affective Domain) และด้านทักษะพิสัย (Psychomotor Domain) (Bloom, 1956) เครื่องมือวัดและประเมินผลแต่ละชนิดถูกออกแบบมาเพื่อวัดพฤติกรรมต่างด้านกัน ครูจึงต้องรู้จักเครื่องมือที่หลากหลายและเลือกใช้ให้ตรงกับธรรมชาติของพฤติกรรมที่ต้องการวัด (สำเร็จ บุญเรืองรัตน์, 2534) ทั้งนี้ รายละเอียดของเครื่องมือแต่ละกลุ่ม ศึกษาเชิงลึกในบทที่ 4 ถึงบทที่ 6 และบทที่ 10 เครื่องมือวัดและประเมินผลทางการศึกษาที่ใช้กันแพร่หลายมีหลายชนิด เช่น แบบทดสอบ (Test) วัดความรู้ ความสามารถ แบบมาตราส่วนประมาณค่า (Rating Scale) ประเมินคุณภาพงานตามระดับ แบบสอบถาม (Questionnaire) สำหรับข้อเท็จจริงหรือความคิดเห็น แบบตรวจสอบรายการ (Checklist) ตรวจสอบการมีหรือไม่มีของพฤติกรรม แบบสังเกต (Observation) บันทึกพฤติกรรมที่แสดงออกโดยตรง แบบสัมภาษณ์ (Interview) เจาะลึกข้อมูล แบบบันทึกพฤติกรรม (Anecdotal Record) สังคมมิติ (Sociometry) วัดความสัมพันธ์ในกลุ่ม ตลอดจนการศึกษารายกรณี (Case Study) เพื่อหาสาเหตุของปัญหาเชิงลึก (สำเร็จ บุญเรืองรัตน์, 2534; พิชรินทร์ ชมภูวิเศษ, 2559)

#### ตารางที่ 1.3 การจับคู่ด้านพฤติกรรมกับเครื่องมือวัดและประเมินที่เหมาะสม

| ด้านพฤติกรรม              | ตัวอย่างสิ่งที่วัด                | เครื่องมือที่เหมาะสม                              |
|---------------------------|-----------------------------------|---------------------------------------------------|
| พุทธิพิสัย (ความรู้-คิด)  | ความจำ ความเข้าใจ การวิเคราะห์    | แบบทดสอบปรนัย/อัตนัย, ใบงาน, คำถามปากเปล่า        |
| จิตพิสัย (เจตคติ-ค่านิยม) | ความสนใจ เจตคติต่อวิชาชีพ คุณธรรม | แบบวัดเจตคติ/มาตราประมาณค่า, แบบสังเกต, แบบสอบถาม |
| ทักษะพิสัย (การปฏิบัติ)   | การทดลอง การนำเสนอ ทักษะกีฬา      | แบบสังเกตการปฏิบัติ + RUBRIC, การทดสอบภาคปฏิบัติ  |

ที่มา: ผู้เขียน ปรับจาก Bloom (1956) และ พิชรินทร์ ชมภูวิเศษ (2559)

ตัวอย่างในชั้นเรียนไทย ครูวิทยาศาสตร์ที่สอนหน่วย “การทดลองวงจรไฟฟ้า” ไม่ควรวัดผลด้วยข้อสอบปรนัยเพียงอย่างเดียว เพราะวัดได้แค่ความรู้ (พุทธิพิสัย) แต่ควรใช้แบบสังเกตการปฏิบัติพร้อมรูบริกขณะนักเรียนต่อวงจรจริง (ทักษะพิสัย) และอาจใช้แบบวัดเจตคติเพื่อดูความกระตือรือร้นต่อวิชา (จิตพิสัย) การใช้

เครื่องมือหลายชนิดประกอบกันทำให้เห็นภาพความสามารถของผู้เรียนรอบด้านยิ่งขึ้น (Nitko & Brookhart, 2019)

ข้อควรระวัง ความผิดพลาดที่พบบ่อยคือการใช้เครื่องมือไม่ตรงด้านพฤติกรรม เช่น ใช้ข้อสอบเขียนตอบวัดทักษะการพูด หรือใช้แบบทดสอบความรู้ตัดสินเจตคติ ซึ่งทำให้ผลการวัดไม่ตรงกับสิ่งที่ต้องการ (ขาดความตรง) ครูจึงต้องถามตนเองเสมอว่า “พฤติกรรมที่จะวัดอยู่ด้านใด และเครื่องมือนี้วัดด้านนั้นได้จริงหรือไม่”

#### 1.5.5 คุณภาพของเครื่องมือวัดและประเมินผล: ความตรง ความเที่ยง ความเป็นปรนัย

เครื่องมือที่ครูสร้างขึ้นจะเชื่อถือได้ต่อเมื่อมีคุณภาพตามมาตรฐาน คุณสมบัติพื้นฐานที่สำคัญที่สุดสามประการ ได้แก่ ความตรง ความเที่ยง และความเป็นปรนัย ซึ่งเป็นหลักสากลที่ปรากฏใน Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014)

ความตรง (Validity) หมายถึง เครื่องมือวัดได้ตรงตามสิ่งที่ต้องการวัดจริง เป็นคุณสมบัติที่สำคัญที่สุด เพราะหากวัดไม่ตรงเป้าหมาย ผลที่ได้ก็ไร้ความหมายแม้จะแม่นยำเพียงใด ความเที่ยง (Reliability) หรือความเชื่อมั่น หมายถึง ผลการวัดมีความคงเส้นคงวา (Consistency) เมื่อวัดซ้ำในสภาพเดิมหรือเปลี่ยนผู้ตรวจก็ได้ผลใกล้เคียงกัน และความเป็นปรนัย (Objectivity) หมายถึง เครื่องมือมีความชัดเจนไม่ขึ้นกับอคติของผู้ตรวจ ประกอบด้วย 3 ส่วน คือ คำถามชัดเจนไม่กำกวม เกณฑ์การให้คะแนนแน่นอน และการตีความหมายคะแนนตรงกัน (ไพศาล หวังพานิช, 2526) ทั้งนี้ สารเชิงลึกของความตรงและความเที่ยง ตลอดจนวิธีตรวจสอบเชิงประจักษ์ ศึกษาในบทที่ 7 และบทที่ 8 ส่วนการวิเคราะห์คุณภาพข้อสอบรายข้อ ศึกษาในบทที่ 9

#### ตารางที่ 1.4 คุณภาพของเครื่องมือ 3 ด้าน พร้อมตัวอย่างข้อผิดพลาดที่พบบ่อย

| คุณสมบัติ                   | ความหมายโดยย่อ          | ตัวอย่างข้อผิดพลาด                                                 |
|-----------------------------|-------------------------|--------------------------------------------------------------------|
| ความตรง (Validity)          | วัดตรงสิ่งที่ต้องการวัด | ต้องการวัดทักษะพูด แต่ใช้ข้อสอบปรนัย<br>ผลจึงไม่ใช่ทักษะที่แท้จริง |
| ความเที่ยง (Reliability)    | วัดซ้ำได้ผลคงเดิม       | ข้อสอบมีคำถามกำกวม นักเรียนต้องเดา<br>คะแนนจึงกว้างไม่น่าเชื่อถือ  |
| ความเป็นปรนัย (Objectivity) | ไม่ขึ้นกับอคติผู้ตรวจ   | ตรวจอัตรานัยโดยไม่มีรูบรีค<br>คะแนนขึ้นกับอารมณ์ครูหรือลายมือเด็ก  |

ที่มา: ผู้เขียน ปรับจาก ไพศาล หวังพานิช (2526) และ AERA, APA, & NCME (2014)

## กล่องที่ 1.2 ข้อพึงระวังในการสร้างเครื่องมือให้มีคุณภาพ

- ตรวจ “ความตรง” ก่อนเสมอ โดยเทียบข้อสอบกับจุดประสงค์/ผังข้อสอบ (Test Blueprint)
- เพิ่ม “ความเที่ยง” ด้วยการเขียนคำถามชัดเจน หลีกเลี่ยงข้อกำกวมและคำปฏิเสธซ้อน
- สร้าง “ความเป็นปรนัย” ในข้อสอบด้วยวิธีสุ่มและการตรวจแบบปิดข้อผู้สอบ
- คุณภาพทั้งสามเชื่อมโยงกัน เครื่องมือที่ไม่เที่ยงย่อมไม่ตรง แต่เครื่องมือที่เที่ยงอาจยังไม่ตรงก็ได้

การประยุกต์ใช้ ในรายวิชานี้ นักศึกษาจะได้ฝึกสร้างรูปรูปในหัวข้อเครื่องมือวัดและประเมินผลประจำบท

(1.10) ซึ่งเป็นกลไกสำคัญที่ช่วยยกระดับความเป็นปรนัยและความเที่ยงของการประเมินงานปฏิบัติและงานเขียน เครื่องมือที่มีคุณภาพครบสามด้านจะทำให้ผลการประเมินสะท้อนความสามารถจริงของผู้เรียนและนำไปใช้ตัดสินใจได้อย่างมั่นใจ (Nitko & Brookhart, 2019)

### 1.5.6 จริยธรรม จรรยาบรรณ และแนวโน้มนการประเมินในศตวรรษที่ 21

เครื่องมือที่ดีจะไร้ความหมายหากผู้ใช้ขาดคุณธรรม การวัดและประเมินผลเกี่ยวข้องโดยตรงกับอนาคตของผู้เรียน ครูผู้ทำหน้าที่ประเมินจึงต้องเป็นผู้มีคุณธรรม พัทรินทร์ ชมภูวิเศษ (2559, น. 19–20) ได้สรุปคุณธรรมของผู้วัดและประเมินผลไว้ 10 ประการ ซึ่งสอดคล้องกับผลลัพธ์การเรียนรู้ CLO5 ของรายวิชา

#### ตารางที่ 1.5 คุณธรรม 10 ประการของครูผู้วัดและประเมินผล

| คุณธรรม                   | ความหมายโดยย่อ                                               |
|---------------------------|--------------------------------------------------------------|
| 1. ความรับผิดชอบ          | ปฏิบัติหน้าที่วัดผลให้สำเร็จ ไม่ละทิ้งงาน วางแผนรอบคอบ       |
| 2. ความซื่อสัตย์สุจริต    | ไม่ใช้ความสามารถไปในทางทุจริตหรือเอื้อประโยชน์โดยมิชอบ       |
| 3. มีวินัยในตนเอง         | ยึดมั่นปฏิบัติในสิ่งถูกต้อง ไม่หวั่นไหวต่อสิ่งเร้า           |
| 4. ความขยันอดทน           | ทุ่มเทกับงานวัดผลที่มีรายละเอียดซับซ้อน                      |
| 5. ความยุติธรรม           | ตัดสิน/ให้คะแนนโดยไม่เห็นแก่ประโยชน์ส่วนตนหรือผู้สอบคนใด     |
| 6. เจตคติที่ดีต่อการวัดผล | มองการวัดผลเป็นเครื่องมือเชิงบวกเพื่อพัฒนา ไม่ใช่เพื่อลงโทษ  |
| 7. ความใจกว้าง            | ยอมรับความเห็นต่างเพื่อปรับปรุงงานและเผยแพร่ความรู้          |
| 8. มนุษยสัมพันธ์          | ประสานงานกับผู้เกี่ยวข้องในงานวัดผลได้ดี                     |
| 9. แสวงหาความรู้          | พัฒนาตนตามหลักวิชาการวัดผลที่ก้าวหน้าอยู่เสมอ                |
| 10. รักษาความลับ          | ปกปิดข้อมูลเครื่องมือและผลการประเมินเพื่อไม่ให้กระทบผู้เรียน |

ที่มา: พัทรินทร์ ชมภูวิเศษ (2559, น. 19–20)

## กล่องที่ 1.3 กรณีศึกษาจริยธรรม: ความยุติธรรมในการให้คะแนน

สถานการณ์: ครูสมชายตรวจข้อสอบพบว่า “เด็กหญิงปลา” ผู้ช่วยงานครูเสมอ ได้ 49 คะแนน (เกณฑ์ผ่าน 50) จึงปิดให้ผ่าน แต่ไม่ปิดให้ “เด็กชายเอก” ที่ได้ 49 เท่ากัน เพราะเขามีพฤติกรรมก้าวร้าว

ประเด็นวิเคราะห์: การกระทำนี้จะละเมิดหลักความยุติธรรมและความเป็นปรนัย ใช้ปัจจัยที่ไม่เกี่ยวกับผลสัมฤทธิ์ มาตัดสินคะแนนบนทอนความน่าเชื่อถือของระบบประเมิน แนวทางที่ถูกต้องคือใช้เกณฑ์เดียวกันกับผู้เรียนทุกคนอย่างโปร่งใส

ในด้านแนวโน้ม การประเมินในศตวรรษที่ 21 เคลื่อนจากการสอบเพื่อตัดสินเพียงอย่างเดียว ไปสู่การประเมินเพื่อการเรียนรู้ (Assessment for/as Learning) การประเมินตามสภาพจริง และการใช้เทคโนโลยี ดิจิทัลสนับสนุนการประเมิน (Black & Wiliam, 1998; สำนักวิชาการและมาตรฐานการศึกษา, 2553) ซึ่งสาระทั้งสามส่วนนี้ศึกษาโดยละเอียดในบทที่ 10 บทที่ 11 และบทที่ 14 ตามลำดับ

การจะใช้แนวทางเหล่านี้ได้จริง ครูต้องมี “การรู้เรื่องการประเมิน” (Assessment Literacy) ซึ่งงานวิจัยกับนักศึกษาครูพบว่าควรพัฒนาทั้งฐานความรู้ มโนทัศน์เกี่ยวกับการประเมิน และการเรียนรู้ของครูควบคู่กัน (Atjonen, Pöntinen, Kontkanen, & Ruotsalainen, 2022) ขณะเดียวกันเทคโนโลยีปัญญาประดิษฐ์ และการวิเคราะห์การเรียนรู้ (Learning Analytics) เริ่มเข้ามาช่วยครูออกแบบเกณฑ์และวิเคราะห์แนวโน้มผู้เรียนเป็นรายบุคคล แต่ครูต้องใช้อย่างมีจริยธรรม ระวังการรั่วไหลข้อมูลและความเป็นส่วนตัวของผู้เรียน (พิเชษฐ์ พินิจ, อนุศิษฐ์ อันมานะตระกูล, และเอกรัตน์ รวยรวย, 2566)

การประยุกต์ใช้และข้อควรระวัง นักศึกษาครูควรฝึกตั้งคำถามเชิงจริยธรรมก่อนตัดสินผลทุกครั้งว่า “เกณฑ์นี้ยุติธรรมกับผู้เรียนทุกคนหรือไม่” และ “ข้อมูลคะแนนถูกเก็บรักษาเป็นความลับเพียงพอหรือยัง” การยึดคุณธรรม 10 ประการควบคู่กับการใช้เทคโนโลยีอย่างรับผิดชอบ คือหลักประกันว่าการประเมินจะเป็นเครื่องมือพัฒนาผู้เรียนอย่างแท้จริง ไม่ใช่เครื่องมือสร้างความเหลื่อมล้ำ

## 1.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ขั้นก่อนเรียน (Pre-engagement)

1. ทำแบบทดสอบก่อนเรียน 10 ข้อ เพื่อสำรวจความรู้พื้นฐาน
2. ระดมสมองประเด็นกระตุ้นความคิด “ถ้าครูใช้ไม้บรรทัดวัดความฉลาด จะเกิดอะไรขึ้น”  
เพื่อโยงสู่ธรรมชาติการวัดทางอ้อม

### ขั้นระหว่างเรียน (Active Processing)

1. บรรยายเชิงปฏิสัมพันธ์เรื่องวงจร O-L-E และความหมายของศัพท์พื้นฐาน 4 คำ
2. ใบงานที่ 1: คิด-จับคู่-แลกเปลี่ยน (Think-Pair-Share) จำแนกสถานการณ์ว่าเป็น “การวัด” หรือ “การประเมิน” พร้อมระบุระดับมาตรการวัด

3. ใบงานที่ 2: เลือกเครื่องมือ 3 ชนิดให้เหมาะกับการวัดพฤติกรรมต่างด้าน (พุทธิ/จิต/ทักษะ)  
ในรายวิชาที่ตนจะสอน
4. กิจกรรมกลุ่ม: วิเคราะห์กรณีศึกษาจริยธรรมการให้คะแนน (กล่องที่ 1.3)  
ตามหลักความยุติธรรมและความเป็นปรณัย

### ชั้นหลังเรียน (Reflection)

1. เขียนบัตรออก (Exit Ticket) สรุป “ความยุติธรรมในการวัดผลของครูมีอาชีพ” คนละ 3-5 บรรทัด
2. ทำแบบทดสอบหลังเรียน 10 ข้อ และมอบหมายแบบฝึกหัดท้ายบท

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป

ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน

ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน

เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 1.7 สรุปบทเรียน

---

- การจัดการศึกษาดำเนินผ่านวงจร O-L-E คือ จุดมุ่งหมาย การจัดการเรียนรู้ และการประเมินผล  
ซึ่งสะท้อนกลับเพื่อพัฒนา
- การทดสอบคือการใช้สิ่งเร้ากระตุ้นพฤติกรรม การวัดผลคือการแทนพฤติกรรมด้วยตัวเลขตามกฎเกณฑ์
- การประเมิน (Assessment) เน้นรวบรวมหลักฐานเพื่อพัฒนา ส่วนการประเมินผล (Evaluation)  
คือการตัดสินคุณค่าด้วยเกณฑ์
- การวัดผลการศึกษาเป็นการวัดทางอ้อม มีความคลาดเคลื่อนเสมอ ( $X = T + E$ ) และไม่มีศูนย์แท้
- มาตรการวัดมี 4 ระดับ ได้แก่ นามบัญญัติ เรียงอันดับ อันตรภาค และอัตราส่วน ซึ่งกำหนดสถิติที่ใช้ได้
- ครูต้องเลือกเครื่องมือให้ตรงกับด้านพฤติกรรม (พุทธิพิสัย จิตพิสัย ทักษะพิสัย)
- เครื่องมือที่ดีต้องมีความตรง ความเที่ยง และความเป็นปรณัย โดยความตรงสำคัญที่สุด
- ครูผู้ประเมินต้องมีคุณธรรม 10 ประการ โดยเฉพาะความยุติธรรมและการรักษาความลับ
- แนวโน้มศตวรรษที่ 21 เน้นการประเมินเพื่อการเรียนรู้ (AfL) การให้ข้อมูลย้อนกลับ  
และการรู้เรื่องการประเมิน
- เทคโนโลยีและ AI เป็นเครื่องมือเสริม ต้องใช้อย่างมีจริยธรรมและคำนึงถึงความเป็นส่วนตัวของข้อมูลผู้เรียน

## 1.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 1.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

### ตอนที่ 1: คำถามทบทวน (ตอบสั้น)

1. อธิบายความสัมพันธ์ของวงจร O–L–E และระบุว่าหากขาดขั้นการประเมินผลจะส่งผลกระทบต่อคุณภาพการสอนอย่างไร
2. ครูประจำชั้นแจ้งผู้ปกครองว่า "เด็กชายเอได้คะแนนการอ่าน 12 จาก 20 คะแนน จึงยังไม่ผ่าน" จงวิเคราะห์ว่าข้อความนี้มีทั้งการวัดผลและการประเมินผลอยู่อย่างไร และครูควรสื่อสารข้อมูลใดเพิ่มเติมจึงจะเป็นธรรมและเป็นประโยชน์ต่อการพัฒนาผู้เรียน
3. ยกตัวอย่างคุณลักษณะที่วัดได้ในแต่ละมาตรการวัดทั้ง 4 ระดับ ระดับละ 1 ตัวอย่างจากชั้นเรียน
4. เหตุใด “ความตรง” จึงถือเป็นคุณสมบัติที่สำคัญที่สุดของเครื่องมือวัดผล
5. ยกตัวอย่างคุณธรรมของผู้ประเมิน 3 ข้อ พร้อมอธิบายความสำคัญต่อความน่าเชื่อถือของการประเมิน

### ตอนที่ 2: วิเคราะห์และประยุกต์

1. ครูสอบ “ทักษะการพูดภาษาอังกฤษ” ด้วยข้อสอบปรนัยเลือกตอบ จงวิเคราะห์ว่าเครื่องมือนี้บกพร่องด้านใด และควรปรับอย่างไร
2. ออกแบบสั้น ๆ ว่าจะใช้เครื่องมือใดบ้างเพื่อวัดผู้เรียนครบทั้งสามด้าน (พุทธิ/จิต/ทักษะ) ในหน่วยการเรียนรู้ที่ท่านสนใจ
3. จงเสนอแนวทางรักษา “ความเป็นปรนัย” ในการตรวจข้อสอบอัตนัยอย่างน้อย 3 วิธี

### ตอนที่ 3: กิจกรรมปฏิบัติ

ให้นักศึกษาจัดทำ “จรรยาบรรณผู้ประเมิน” 5 ข้อ ที่ครอบคลุมความยุติธรรม การรักษาความลับ และการใช้เทคโนโลยี/AI อย่างมีจริยธรรม พร้อมยกตัวอย่างสถานการณ์ประกอบแต่ละข้อ แล้วนำเสนอกลุ่มย่อย (ประเมินด้วยรูบริกในหัวข้อ 1.10)

#### 1.10 เครื่องมือวัดและประเมินผลประจำบท

การวัดและประเมินผลประจำบทนี้ผูกกับผลลัพธ์การเรียนรู้สองข้อ ได้แก่ CLO1 ที่วัดด้วยแบบทดสอบก่อน-หลังเรียนและแบบฝึกหัดท้ายบท ตรวจสอบให้คะแนนตามเฉลย โดยกำหนดเกณฑ์ผ่านที่ร้อยละ 70 ขึ้นไป และ CLO5 ที่วัดด้วยการวิเคราะห์กรณีศึกษาจริยธรรมและการสังเกตพฤติกรรมการอภิปราย ประเมินด้วยรูบริกแบบแยกองค์ประกอบ (Analytic Rubric) ดังตารางที่ 1.6

ตารางที่ 1.6 รูบริกประเมินการวิเคราะห์กรณีศึกษาจริยธรรม (CLO5)

| เกณฑ์การประเมิน                               | ดีมาก (4)                        | ดี (3)                  | พอใช้ (2)         | ปรับปรุง (1)        |
|-----------------------------------------------|----------------------------------|-------------------------|-------------------|---------------------|
| 1. ระบุประเด็นจริยธรรม                        | ระบุครบถ้วน ชัดเจน ตรงประเด็น    | ระบุประเด็นหลักได้      | ระบุได้บางส่วน    | ระบุไม่ตรงประเด็น   |
| 2. วิเคราะห์ตามหลักความยุติธรรม/ ความเป็นปณัย | วิเคราะห์ลึก อ้างหลักการถูกต้อง  | วิเคราะห์ถูกต้องตามหลัก | วิเคราะห์ผิวเผิน  | ขาดการวิเคราะห์     |
| 3. เสนอแนวทางแก้ไข                            | เสนอแนวทางเป็นรูปธรรม นำไปใช้ได้ | เสนอแนวทางเหมาะสม       | เสนอแนวทางกว้าง ๆ | ไม่เสนอ/ไม่สอดคล้อง |
| 4. การมีส่วนร่วมอภิปราย                       | ร่วมอภิปรายอย่างมีเหตุผลสม่ำเสมอ | ร่วมอภิปรายดี           | ร่วมเป็นบางครั้ง  | แทบไม่มีส่วนร่วม    |

ที่มา: ผู้เขียน (เกณฑ์สอดคล้องกับคุณธรรมความยุติธรรมและความเป็นปณัยในหัวข้อ 1.5.5–1.5.6)

#### เกณฑ์การผ่าน (Passing Criteria)

- CLO1: ทำแบบทดสอบ/แบบฝึกหัดได้คะแนนรวมตั้งแต่ร้อยละ 70 ขึ้นไป
- CLO5: ได้ระดับคุณภาพเฉลี่ยตั้งแต่ระดับ “ดี” (3) ขึ้นไปจากรูบริก (คะแนนรวม 12 จาก 16 ขึ้นไป)

#### 1.11 เอกสารอ้างอิงประจำบท

##### ภาษาไทย

ชวาล แพร์ตกุล. (2520). เทคนิคการเขียนข้อสอบ (พิมพ์ครั้งที่ 5). กรุงเทพฯ: โรงพิมพ์พิทักษ์อักษร.

บุญธรรม กิจปรีดาบริสุทธิ์. (2549). การวัดและประเมินผลการเรียนการสอน (พิมพ์ครั้งที่ 2).  
กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542. (2542, 19 สิงหาคม). ราชกิจจานุเบกษา. เล่ม 116  
ตอนที่ 74 ก, หน้า 1-23.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการศึกษา. อุตรธานี: คณะครุศาสตร์  
มหาวิทยาลัยราชภัฏอุตรธานี.

พิชิต ฤทธิ์จรรยา. (2564). หลักการวัดและการประเมินผลการศึกษา (พิมพ์ครั้งที่ 10). กรุงเทพฯ: แฮส  
ออฟ เคอร์มิสท์.

พิเชษฐ์ พินิจ, อนุศิษฐ์ อันมานะตระกูล, และเอกรัตน์ รวยรวย. (2566). การรู้การประเมิน:  
การประเมิน ความก้าวหน้าในการเรียนรู้  
และการกำหนดเกรดที่ถูกต้องและเป็นธรรมตามแนวคิดการจัดการศึกษาแบบเน้นผลลัพธ์.  
วารสารนวัตกรรมการเรียนรู้และเทคโนโลยี, 3(2), 6-20.

ไพศาล หวังพานิช. (2526). การวัดผลการศึกษา. กรุงเทพฯ: ไทยวัฒนาพานิช.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6). กรุงเทพฯ:  
สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำนักวิชาการและมาตรฐานการศึกษา. (2553).  
แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.  
กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

สำเริง บุญเรืองรัตน์ และคณะ. (2534). การวัดและประเมินผลการศึกษา. กรุงเทพฯ:  
สำนักพิมพ์พัฒนาศึกษา.

อนุวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่.  
สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

## ภาษาอังกฤษ

American Educational Research Association, American Psychological Association, &  
National Council on Measurement in Education. (2014). Standards for educational and  
psychological testing. Washington, DC: American Educational Research Association.

Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). Enhancing preservice teachers' assessment literacy: Focus on knowledge base, conceptions of assessment, and teacher learning. *Frontiers in Education*, 7, 891391. <https://doi.org/10.3389/feduc.2022.891391>

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Maidenhead, UK: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). New York: Pearson.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.

## บทที่ 2

### แนวคิด หลักการ และประเภทของการวัดและประเมินผลการเรียนรู้ *Concepts, Principles, and Types of Measurement and Evaluation*

#### รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้

(Learning Measurement and Evaluation)

##### ข้อมูลประจำบท

จำนวนคาบเรียน : 4 คาบ (บรรยายเพื่อสร้างฐานคิด 2 คาบ และฝึกปฏิบัติ/อภิปราย 2 คาบ)

ผลลัพธ์การเรียนรู้หลักประจำบท : CLO1 — อธิบายหลักการ แนวคิด

และความสำคัญของการวัดและประเมินผลการเรียนรู้

ระดับพฤติกรรมเป้าหมาย : จำ-เข้าใจ (Remember-Understand) ของพฤติกรรมด้านพุทธิพิสัย

เครื่องมือวัดเด่นประจำบท : แบบทดสอบก่อน-หลังเรียน และใบงานจำแนกประเภท (พร้อมรูบรีค)

#### แผนบริหารการสอนประจำบทที่ 2

##### 1. หัวข้อสาระการจัดการเรียนการสอน

- แนวคิดพื้นฐานของการวัดและประเมินผล: ความสัมพันธ์ระหว่างการทดสอบ การวัด และการประเมินผล และองค์ประกอบของการวัด
- มาตรการวัด 4 ระดับ: นามบัญญัติ เรียงอันดับ อันตรภาค และอัตราส่วน
- หลักการสำคัญ 5 ประการของการวัดและประเมินผลที่มีคุณภาพ
- ประเภทของการวัดผล: การวัดแบบอิงกลุ่ม (Norm-referenced) และอิงเกณฑ์ (Criterion-referenced) และตารางวิเคราะห์หลักสูตร (TOS)
- ประเภทของแบบทดสอบ: แบบทดสอบผลสัมฤทธิ์ (Achievement) และแบบทดสอบความถนัด (Aptitude)
- ประเภทของการประเมินตามช่วงเวลา: ก่อนเรียน ระหว่างเรียน วินิจฉัย และสรุปผล
- ประเภทของการประเมินตามวัตถุประสงค์: Assessment of / for / as Learning
- การประเมินตามสภาพจริง (Authentic Assessment) และแนวโน้มการประเมินในศตวรรษที่ 21

##### 2. ความคิดรวบยอด

การวัดและประเมินผลเป็นองค์ประกอบหนึ่งของระบบการจัดการเรียนรู้ที่ต้องสอดคล้องกับจุดมุ่งหมาย และกระบวนการเรียนรู้ (Objective-Learning-Evaluation)

การเลือกใช้เครื่องมือวัดต้องยึดโยงกับผลลัพธ์การเรียนรู้ และต้องมีคุณภาพตามหลักการสำคัญ

โดยเฉพาะความเที่ยงตรง (Validity) และความเชื่อมั่น (Reliability) จึงจะสร้างความยุติธรรมแก่ผู้เรียน และนำไปสู่การประเมิน เพื่อพัฒนาผู้เรียนมากกว่าการตัดสินเพียงอย่างเดียว

### 3. จุดประสงค์การเรียนรู้

1. อธิบายความหมายและความสัมพันธ์เชิงระบบของการทดสอบ การวัด และการประเมินผลได้ถูกต้อง (CLO1)
2. จำแนกองค์ประกอบของการวัดและมาตราการวัดทั้ง 4 ระดับ พร้อมยกตัวอย่างประกอบได้ (CLO1)
3. อธิบายหลักการสำคัญ 5 ประการของการวัดและประเมินผลที่มีคุณภาพได้ (CLO1)
4. เปรียบเทียบการวัดแบบอิงกลุ่มกับอิงเกณฑ์ และจำแนกประเภทของแบบทดสอบและการประเมินได้ (CLO1)
5. อธิบายแนวคิดการประเมินตามสภาพจริงและแนวโน้มการประเมินในศตวรรษที่ 21 ได้ (CLO1)

### 4. เนื้อหาวิชา

เนื้อหาโดยละเอียดปรากฏตามเอกสารประกอบการสอนบทที่ 2 หัวข้อ 2.5 (เนื้อหาวิชา) ประกอบด้วยหัวข้อย่อย 2.5.1 ถึง 2.5.8 ซึ่งจัดเรียงจากแนวคิดและนิยามพื้นฐาน สู่หลักการคุณภาพ ประเภทของการวัดและประเมิน และการประยุกต์ในชั้นเรียนยุคปัจจุบัน

### 5. กิจกรรมการเรียนการสอน

1. ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน และร่วมอภิปรายกรณี “การให้คะแนนโดยไม่มีเกณฑ์” เพื่อกระตุ้นความตระหนักเรื่องความยุติธรรมในการประเมิน
2. ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ประกอบตัวอย่าง ทำกิจกรรมจำแนกระดับมาตราการวัดของข้อมูลจริง และวิเคราะห์กรณีศึกษาการละเมิดหลักการวัดผล (Think-Pair-Share)
3. ขั้นหลังเรียน: ทำใบงานจำแนกประเภท สรุปเป็นผังมโนทัศน์ (Mind Map) และทำแบบทดสอบหลังเรียน

### 6. สื่อการเรียนการสอน

- เอกสารประกอบการสอนบทที่ 2 และสไลด์นำเสนอ (PowerPoint)
- ตำราหลัก: พัทธินทร์ ชมภูวิเศษ (2559); ขวาล แพรัตกุล (2520); พิชิต ฤทธิ์เจริญ (2564)
- ใบงานจำแนกประเภท ชุดข้อมูลตัวอย่าง และกรณีศึกษาสถานการณ์การประเมิน
- เครื่องมือดิจิทัล: Quizizz/Kahoot สำหรับแบบทดสอบ และโปรแกรมช่วยร่างรูปรีด

### 7. การวัดและประเมินผล

- แบบทดสอบหลังเรียน (ปรนัย 11 ข้อ) วัดความเข้าใจนิยาม หลักการ และประเภท — ผูกกับ CLO1 (น้ำหนัก 40%)
- ใบงานจำแนกประเภท (มาตราการวัด/ประเภทการประเมิน) ตรวจสอบด้วยรูปรีด — ผูกกับ CLO1 (น้ำหนัก 40%)
- การมีส่วนร่วมในการอภิปราย/กิจกรรมกลุ่ม — ผูกกับ CLO1 (น้ำหนัก 20%)

- เกณฑ์ผ่าน: ได้คะแนนรวมไม่น้อยกว่าร้อยละ 60 และใบงานอยู่ในระดับ “ดี” ขึ้นไป

## 2.1 ความคิดรวบยอด

การจัดการศึกษาที่มีคุณภาพต้องอาศัยความสอดคล้องกันขององค์ประกอบสามส่วน คือ การกำหนดจุดมุ่งหมาย (Objective) การจัดประสบการณ์การเรียนรู้ (Learning) และการประเมินผล (Evaluation) หรือกรอบ O-L-E การประเมินผลทำหน้าที่เป็น สารสนเทศย้อนกลับที่สะท้อนคุณภาพของกระบวนการทั้งหมด เพื่อระบุจุดเด่นจุดด้อยสำหรับการปรับปรุงและพัฒนา (พัชรินทร์ ชมภูวิเศษ, 2559)

การวัดและประเมินผลมิใช่เพียงการให้ตัวเลขหรือการตัดสินผลการเรียนเมื่อสิ้นภาคเรียน แต่เป็นเครื่องมือเชิงกลยุทธ์ในการพัฒนาศักยภาพผู้เรียน การเลือกใช้เครื่องมือวัดจึงต้องยึดโยงกับวัตถุประสงค์และผลลัพธ์การเรียนรู้ และต้องมีคุณภาพตามหลักการสำคัญ โดยเฉพาะความเที่ยงตรง (Validity) และความเชื่อมั่น (Reliability) มิฉะนั้นการประเมินย่อมไม่อาจสร้างความยุติธรรมแก่ผู้เรียนได้ (ขวาล แพรัตกุล, 2520; สุทธิพงศ์ บุญผดุง, 2568) แนวคิดปัจจุบันจึงเน้นการเปลี่ยนผ่านจากการประเมินเพื่อ “ตัดสิน” (Assessment of Learning) ไปสู่การประเมินเพื่อ “พัฒนา” และการประเมินที่เป็นส่วนหนึ่งของการเรียนรู้ (Assessment for/as Learning) (Earl, 2003)

## 2.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทนี้จบแล้ว นักศึกษาสามารถ

1. อธิบายความหมายและความสัมพันธ์เชิงระบบของการทดสอบ การวัด และการประเมินผลได้ถูกต้อง
2. จำแนกองค์ประกอบของการวัดและมาตราการวัดทั้ง 4 ระดับ พร้อมยกตัวอย่างประกอบได้
3. อธิบายหลักการสำคัญ 5 ประการของการวัดและประเมินผลที่มีคุณภาพได้
4. เปรียบเทียบการวัดแบบอิงกลุ่มกับอิงเกณฑ์ และจำแนกประเภทของแบบทดสอบและการประเมินได้
5. อธิบายแนวคิดการประเมินตามสภาพจริงและแนวโน้มการประเมินในศตวรรษที่ 21 ได้

ตารางความสอดคล้องระหว่างจุดประสงค์ ผลลัพธ์การเรียนรู้ (CLO) ระดับพฤติกรรม กิจกรรม และการวัดประเมินผล

| จุดประสงค์ | CLO  | ระดับพฤติกรรม      | กิจกรรมการเรียนรู้          | วิธีวัด / เครื่องมือ               |
|------------|------|--------------------|-----------------------------|------------------------------------|
| ข้อ 1      | CLO1 | เข้าใจ             | บรรยาย-อภิปราย, ผังมโนทัศน์ | แบบทดสอบ ข้อ 1-2                   |
| ข้อ 2      | CLO1 | เข้าใจ-จำแนก       | กิจกรรมจำแนกมาตราการวัด     | แบบทดสอบ ข้อ 4-5;<br>ใบงานตอน ก    |
| ข้อ 3      | CLO1 | เข้าใจ             | วิเคราะห์กรณีละเมิดหลักการ  | แบบทดสอบ ข้อ 6-7                   |
| ข้อ 4      | CLO1 | เข้าใจ-เปรียบเทียบ | Think-Pair-Share, ใบงาน     | แบบทดสอบ ข้อ 3,8-11;<br>ใบงานตอน ข |
| ข้อ 5      | CLO1 | เข้าใจ             | อภิปรายแนวโน้มและกรณีศึกษา  | คำถามทบทวน/อัตรัย                  |

หมายเหตุการจัดวางการประเมิน (Constructive Alignment)

บทนี้ประกาศและประเมินเฉพาะ CLO1 เท่านั้น ทุกจุดประสงค์จึงมีกิจกรรมและเครื่องมือวัดรองรับครบถ้วน ไม่มีจุดประสงค์ที่ไม่ถูกวัด และไม่มีการประเมิน CLO อื่นที่มีได้ประกาศไว้ในบทนี้

## 2.3 คำสำคัญ

| คำสำคัญ (ไทย)             | ภาษาอังกฤษ           | คำสำคัญ (ไทย)              | ภาษาอังกฤษ              |
|---------------------------|----------------------|----------------------------|-------------------------|
| การทดสอบ                  | Testing              | การวัดผล                   | Measurement             |
| การประเมินผล              | Evaluation           | มาตราการวัด                | Scales of Measurement   |
| ความเที่ยงตรง             | Validity             | ความเชื่อมั่น              | Reliability             |
| ความเป็นปรนัย             | Objectivity          | การวัดแบบอิงกลุ่ม          | Norm-referenced         |
| การวัดแบบอิงเกณฑ์         | Criterion-referenced | แบบทดสอบผลสัมฤทธิ์         | Achievement Test        |
| แบบทดสอบความถนัด          | Aptitude Test        | การประเมินเพื่อพัฒนา       | Formative Assessment    |
| การประเมินสรุปผล          | Summative Assessment | การประเมินเพื่อการเรียนรู้ | Assessment for Learning |
| การประเมินตามสภาพจริง     | Authentic Assessment | ตารางวิเคราะห์หลักสูตร     | Table of Specifications |
| ความฉลาดรู้ด้านการประเมิน | Assessment Literacy  |                            |                         |

## 2.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงลิงก์ที่ผู้สอน ประกาศในชั้นเรียน

## 2.5 เนื้อหาวิชา

### 2.5.1 แนวคิดพื้นฐาน: การทดสอบ การวัด และการประเมินผล

การเข้าใจนิยามของคำสามคำที่ผู้เรียนมักใช้สับสนกัน ได้แก่ การทดสอบ การวัดผล และการประเมินผล ถือเป็นจุดเริ่มต้นของความยุติธรรมทางการศึกษาเพราะการเลือกคำผิดมักนำไปสู่การตีความผล และการตัดสินใจที่ผิดพลาด พัทธินทร์ ชมภูวิเศษ (2559) และไพศาล หวังพานิช (2526) อธิบายตรงกันว่า ทั้งสามคำเป็นกระบวนการต่อเนื่องเชิงระบบ มิใช่คำที่ใช้แทนกันได้ และในยุคที่เน้นความรอบรู้ด้านการประเมิน (**assessment literacy**) ครูจำเป็นต้องแยกแยะสามคำนี้ให้ชัดเจนก่อนออกแบบการวัดผลใด ๆ (พิเชษฐ์ พิณิจ, อนุศิษฐ์ อันมานะตระกูล, และเอกรัตน์ รวยรวย, 2566)

การทดสอบ (Testing) คือกระบวนการใช้เครื่องมือหรือสถานการณ์ เช่น ข้อสอบ คำถาม หรือภาระงาน เป็นสิ่งเร้าเพื่อกระตุ้นให้ผู้เรียนแสดงพฤติกรรมที่ต้องการออกมาให้สังเกตและบันทึกได้ การทดสอบจึงเป็น “วิธีเก็บข้อมูล” ยังไม่ใช่ตัวเลขหรือการตัดสิน

การวัดผล (Measurement) คือการกำหนดตัวเลขหรือสัญลักษณ์ให้แก่พฤติกรรมหรือคุณลักษณะของผู้เรียนตามกฎเกณฑ์ เป็นการตอบคำถามว่า “มีมากน้อยเพียงใด” ผลลัพธ์ของการวัดคือข้อมูลเชิงปริมาณ เช่น ได้ 35 จาก 50 คะแนน (ชวาล แพรัตกุล, 2520)

การประเมินผล (Evaluation) คือการตัดสินคุณค่าของสิ่งที่วัดได้ โดยนำผลการวัดไปเทียบกับเกณฑ์ (criteria) แล้วลงข้อสรุปเพื่อการตัดสินใจ เช่น “ผ่าน/ไม่ผ่าน” หรือ “ควรสอนซ่อมเสริมเรื่องใด” การประเมินจึงเป็นการให้ความหมายแก่ตัวเลข มีใช้ตัวเลขเอง

เมื่อเรียงลำดับความสัมพันธ์จะได้กระบวนการที่ไหลต่อเนื่องดังภาพที่ 2.1 กล่าวคือ การทดสอบสร้างพฤติกรรมให้สังเกตได้ การวัดเปลี่ยนพฤติกรรมนั้นเป็นตัวเลข และการประเมินเปลี่ยนตัวเลขให้เป็นการตัดสินใจ ก่อนนำไปสู่การตัดสินผล (judgment) ในขั้นสุดท้าย

ภาพที่ 2.1 ความสัมพันธ์เชิงระบบของการทดสอบ การวัดผล และการประเมินผล

| 1) การทดสอบ<br>(Testing)                  | 2) การวัดผล<br>(Measurement)         | 3) การประเมินผล<br>(Evaluation)     | 4) การตัดสินผล<br>(Judgment)              |
|-------------------------------------------|--------------------------------------|-------------------------------------|-------------------------------------------|
| ใช้สิ่งเร้า/เครื่องมือ<br>กระตุ้นพฤติกรรม | กำหนดตัวเลขแก่<br>พฤติกรรมตามกฎเกณฑ์ | เทียบผลกับเกณฑ์<br>แล้วตัดสินคุณค่า | ตัดสินใจ/ดำเนินการ<br>เช่น ผ่าน-ซ่อมเสริม |

ที่มา: ดัดแปลงจาก พิชรินทร์ ชมภูวิเศษ (2559, น. 21) และ ไพศาล หวังพานิช (2526)

องค์ประกอบของการวัดที่ขาดไม่ได้มี 3 ส่วน (ไพศาล หวังพานิช, 2526) ได้แก่

- (1) สิ่งที่ต้องการวัด (Attribute) เช่น ความรู้ทางคณิตศาสตร์ เจตคติต่อวิชา หรือทักษะการทดลอง
- (2) เครื่องมือวัด (Tool) เช่น แบบทดสอบ แบบสังเกต หรือแบบประเมินการปฏิบัติ
- (3) ผลการวัด (Result) ตัวเลขหรือสัญลักษณ์ที่ได้ ซึ่งต้องนำไปแปลความหมายในขั้นประเมินต่อไป

ตารางที่ 2.1 เปรียบเทียบการทดสอบ การวัดผล และการประเมินผล

| ประเด็น       | การทดสอบ             | การวัดผล         | การประเมินผล                  |
|---------------|----------------------|------------------|-------------------------------|
| คำถามหลัก     | แสดงพฤติกรรมอะไร     | มีมากน้อยเพียงใด | มีคุณค่าผ่านเกณฑ์หรือไม่      |
| ลักษณะผลลัพธ์ | พฤติกรรมที่สังเกตได้ | ตัวเลข/สัญลักษณ์ | ข้อสรุปเชิงคุณค่า/การตัดสินใจ |
| ตัวอย่าง      | ให้ทำข้อสอบ 50 ข้อ   | ได้ 35 คะแนน     | ผ่านเกณฑ์ 60% จึง “ผ่าน”      |

ที่มา: สังเคราะห์จาก ชวาล แพรัตกุล (2520) และ พิชรินทร์ ชมภูวิเศษ (2559)

ตัวอย่างในชั้นเรียนไทย: ครูวิชาวิทยาศาสตร์ให้นักเรียนทำแบบทดสอบเรื่องระบบสุริยะ 20 ข้อ (การทดสอบ) เด็กชายก้องตอบถูก 14 ข้อ จึงได้ 14 คะแนน (การวัดผล) เมื่อครูเทียบกับเกณฑ์ผ่านที่ร้อยละ 60 (12 คะแนน) จึงสรุปว่า “ผ่าน” และวางแผนเสริมเฉพาะหัวข้อที่ก้องพลาด (การประเมินผลและการตัดสินใจ) จะเห็นว่าหากครูหยุดเพียงแค่ตัวเลข 14 คะแนนโดยไม่ตีความ ก็ยังไม่เกิดการประเมินที่นำไปใช้พัฒนาผู้เรียนได้

การประยุกต์และข้อควรระวัง: ความผิดพลาดที่พบบ่อยคือการใช้คำว่า “วัดผล” และ “ประเมินผล” สลับกัน ทำให้ครูรายงานเพียงคะแนนดิบโดยไม่ให้ความหมายเชิงพัฒนา หรือรับ “ตัดสิน” ผู้เรียนทั้งที่เครื่องมือยังด้อยคุณภาพ ครูจึงควรถามตนเองทุกครั้งว่ากำลังอยู่ในขั้นใดของกระบวนการ และผลการวัดถูกแปลความอย่างมีเกณฑ์รองรับหรือไม่ เพื่อให้การประเมินรับใช้การเรียนรู้ มิใช่เป็นเพียงพิธีกรรมของการให้คะแนน (พิเชษฐ พินิจ และคณะ, 2566) 2.5.2 มาตรการวัด 4 ระดับ (Scales of Measurement)

Stevens (1946) เสนอว่า ข้อมูลที่ได้จากการวัดจำแนกได้เป็น 4 ระดับตามคุณสมบัติทางคณิตศาสตร์ ที่กระทำกับตัวเลขนั้นได้ เรียงจากหยาบไปละเอียด ได้แก่ นามบัญญัติ เรียงอันดับ อันตรภาค และอัตราส่วน ซึ่งสอดคล้องกับที่ตำราไทยอธิบายไว้ (พัชรินทร์ ชมภูวิเศษ, 2559; ศิริชัย กาญจนวาสี, 2552) การรู้ระดับ มาตรการวัดสำคัญมาก เพราะเป็นตัวกำหนดว่าจะเลือกใช้สถิติชนิดใดได้อย่างถูกต้อง ดังสรุปในตารางที่ 2.2

ตารางที่ 2.2 คุณสมบัติ ตัวอย่างทางการศึกษา และสถิติที่เหมาะสมของมาตรการวัดทั้ง 4 ระดับ

| ระดับมาตรา                  | คุณสมบัติสำคัญ                                  | ตัวอย่างทางการศึกษา                 | สถิติที่ใช้ได้                                  |
|-----------------------------|-------------------------------------------------|-------------------------------------|-------------------------------------------------|
| 1. นามบัญญัติ<br>(Nominal)  | จำแนก/จัดกลุ่ม<br>ตัวเลขไม่มีความหมายเชิงปริมาณ | เพศ กลุ่มสาระ<br>หมายเลขประจำตัว    | ความถี่ ร้อยละ ฐานนิยม โค-สแควร์                |
| 2. เรียงอันดับ<br>(Ordinal) | บอกมาก-น้อย/ลำดับได้<br>แต่ช่วงห่างไม่เท่ากัน   | อันดับที่สอบได้<br>ระดับความพึงพอใจ | มัธยฐาน เปอร์เซ็นไทล์<br>สหสัมพันธ์อันดับ       |
| 3. อันตรภาค<br>(Interval)   | ช่วงเท่ากัน บวก-ลบได้<br>แต่ไม่มีศูนย์แท้       | คะแนนสอบ<br>อุณหภูมิเซลเซียส        | ค่าเฉลี่ย<br>ส่วนเบี่ยงเบนมาตรฐาน<br>สหสัมพันธ์ |
| 4. อัตราส่วน<br>(Ratio)     | มีศูนย์แท้ คูณ-หาร/เทียบเป็นเท่าได้             | น้ำหนัก ส่วนสูง<br>เวลาที่ใช้ทำงาน  | ใช้ได้กับสถิติทุกประเภท                         |

ที่มา: ดัดแปลงจาก พัทชรินทร์ ชมภูวิเศษ (2559, น. 240-242, ตารางที่ 7.2) และ Stevens (1946)

เหตุที่คะแนนสอบส่วนใหญ่จัดเป็นมาตราอันตรภาค ไม่ใช่อัตราส่วน เป็นประเด็นที่ครูเข้าใจคลาดเคลื่อนบ่อยที่สุด ดังอธิบายในกล่องที่ 2.1

กล่องที่ 2.1 ทำไมคะแนนสอบจึงเป็น “อันตรภาค” ไม่ใช่ “อัตราส่วน”

คะแนน 0 ในการสอบเป็น “ศูนย์สมมติ” (arbitrary zero) มิใช่ศูนย์แท้ การได้ 0 คะแนนไม่ได้แปลว่าผู้เรียนไม่มีความรู้เลย เพียงแต่ทำข้อสอบฉบับนั้นไม่ได้จึงเปรียบเทียบเป็น “จำนวนเท่า” ไม่ได้ ผู้ได้

80 คะแนนไม่ได้เก่งเป็นสองเท่าของผู้ได้ 40 คะแนน ทำได้เพียงบวก-ลบหาผลต่างและหาค่าเฉลี่ย/ ส่วนเบี่ยงเบนมาตรฐาน แต่คุณ-หารเทียบเป็นเท่าไม่ได้  
ที่มา: พชรินทร์ ชมภูวิเศษ (2559, น. 241)

ตัวอย่างในชั้นเรียนไทย: ในกิจกรรม “นักสืบมาตรฐานการวัด” ครูให้นักศึกษาจำแนกข้อมูล เช่น เลขประจำตัวนักเรียน (นามบัญญัติ) อันดับที่สอบได้ (เรียงอันดับ) คะแนนสอบกลางภาค (อันตรภาค) และเวลาที่ใช้วิ่ง 100 เมตร (อัตราส่วน) การฝึกเช่นนี้ช่วยให้นักศึกษาครุ่นคิดว่าตัวเลขที่หน้าตาเหมือนกันอาจอยู่คนละระดับมาตรฐาน และเลือกสถิติได้ถูกต้องเมื่อไปฝึกสอนจริง

**การประยุกต์และข้อควรระวัง:** ข้อผิดพลาดที่พบบ่อยคือการนำสถิติของมาตรฐานสูงไปใช้กับข้อมูลมาตรฐานต่ำ เช่น การหาค่าเฉลี่ยของ “เพศ” หรือ “กลุ่มสาระ” ซึ่งเป็นนามบัญญัติ หรือการเฉลี่ยลำดับความพึงพอใจแบบหยาบ ๆ โดยไม่ระวังว่าช่วงห่างไม่เท่ากัน ครูจึงควรระบุระดับมาตรฐานของข้อมูลก่อนเสมอ แล้วจึงเลือกสถิติให้สอดคล้อง เพื่อให้การแปลผลมีความถูกต้องตามหลักการวัด (ศิริชัย กาญจนวาสี, 2552)

### 2.5.3 หลักการสำคัญ 5 ประการของการวัดและประเมินผลที่มีคุณภาพ

คุณภาพของเครื่องมือคือรากฐานของความยุติธรรมทางการศึกษา ครูจึงต้องยึดหลักการสำคัญ 5 ประการในการสร้างและเลือกใช้เครื่องมือวัด (ขวาล แพร์ตกุล, 2520; พชรินทร์ ชมภูวิเศษ, 2559) ซึ่งสอดคล้องกับเกณฑ์คุณภาพการวัดในตารางต่างประเทศที่ถือว่าความตรงและความเที่ยงเป็นหัวใจ (Gronlund & Linn, 1990) หลักการทั้งห้าและคำถามที่ใช้ตรวจสอบสรุปไว้ในตารางที่ 2.3

ตารางที่ 2.3 หลักการคุณภาพ 5 ประการของเครื่องมือวัดและคำถามตรวจสอบ

| หลักการ                         | ความหมายโดยย่อ                                 | คำถามตรวจสอบ                                      |
|---------------------------------|------------------------------------------------|---------------------------------------------------|
| 1. ความเที่ยงตรง (Validity)     | วัดได้ตรงกับสิ่งที่ต้องการวัดจริง              | ข้อสอบครอบคลุมจุดประสงค์/เนื้อหาที่สอนหรือไม่     |
| 2. ความเชื่อมั่น (Reliability)  | วัดกี่ครั้งผลคงเส้นคงวาหากคุณลักษณะไม่เปลี่ยน  | สอบซ้ำหรือคนละชุดคู่ขนานแล้วผลใกล้เคียงกันหรือไม่ |
| 3. ความเป็นปรนัย (Objectivity)  | คำถามชัดเจน<br>เฉลย/เกณฑ์ตรวจตรงกันทุกผู้ตรวจ  | ครูคนละคนตรวจแล้วได้คะแนนตรงกันหรือไม่            |
| 4. ความยุติธรรม (Fairness)      | ไม่เปิดโอกาสให้ผู้เรียนได้เปรียบ-เสียเปรียบกัน | เครื่องมือมีอคติทางเพศ ภาษา หรือฐานะหรือไม่       |
| 5. ความเป็นไปได้ (Practicality) | ใช้เวลา ทรัพยากร<br>และการตรวจที่เหมาะสม       | นำไปใช้จริงในชั้นเรียนได้คุ้มค่าหรือไม่           |

ที่มา: สังเคราะห์จาก ขวาล แพร์ตกุล (2520), พชรินทร์ ชมภูวิเศษ (2559, น. 21) และ Gronlund & Linn (1990)

หลักการทั้งห้ามีความสัมพันธ์เชิงลำดับชั้น ดังกล่องที่ 2.2 กล่าวคือ ความเป็นปรนัยเป็นรากฐานที่ทำให้เกิดความเชื่อมั่น และความเชื่อมั่นเป็นเงื่อนไขจำเป็น (แต่ไม่เพียงพอ) ของความเที่ยงตรง เครื่องมือจึงอาจเชื่อมั่นสูงแต่ไม่เที่ยงตรงได้ หากวัดสิ่งที่คงเส้นคงวาแต่ผิดเป้าหมาย

กล่องที่ 2.2 ความสัมพันธ์: ปรนัย → เชื่อมั่น → เที่ยงตรง  
 ความเป็นปรนัย ช่วยให้การตรวจและแปลผลตรงกัน → ลดความคลาดเคลื่อนจากผู้ตรวจ  
 เมื่อความคลาดเคลื่อนน้อย ผลการวัดจึง คงเส้นคงวา (เชื่อมั่นสูง)  
 แต่เชื่อมั่นสูงยังไม่รับประกันความเที่ยงตรง เช่น ตาชั่งที่อ่านค่าคงที่แต่เพี้ยน 2 กก. ทุกครั้ง: เชื่อมั่นสูงแต่ไม่ตรง  
 ที่มา: ดัดแปลงจาก พิซิต ฤทธิ์จรรยา (2564) และ Gronlund & Linn (1990)

ตัวอย่างในชั้นเรียนไทย: ครูศิลปะตรวจผลงานวาดภาพโดยไม่มีเกณฑ์ชัดเจน คะแนนจึงขึ้นกับอารมณ์ขณะตรวจ กรณีนี้ขาดความเป็นปรนัยก่อน และเมื่อขาดความเป็นปรนัยก็ทำให้ความเชื่อมั่นต่ำตามไปด้วย ทางแก้คือจัดทำ rubric ที่ระบุระดับคุณภาพและคำบรรยายชัดเจน เพื่อให้ครูคนละคนให้คะแนนงานชิ้นเดียวกันได้ใกล้เคียงกัน

การประยุกต์และข้อควรระวัง: ครูมักเร่งสร้างข้อสอบให้ทันสอบโดยมองข้ามความเที่ยงตรงเชิงเนื้อหา ทำให้ออกข้อสอบนอกเรื่องที่สอน หรือเลือกเฉพาะข้อที่ออกง่าย ข้อควรระวังคืออย่าสับสนว่าข้อสอบที่ “ตรวจง่าย/ปรนัย” จะเที่ยงตรงเสมอไป ความเป็นปรนัยเป็นเพียงเงื่อนไขหนึ่ง ไม่ใช่หลักประกันว่าวัดตรงเป้าหมาย การวางตารางวิเคราะห์หลักสูตรก่อนออกข้อสอบ (ดูหัวข้อ 2.5.4) จึงช่วยคุมความเที่ยงตรงได้ตั้งแต่ต้น (พัชรินทร์ ชมภูวิเศษ, 2559)

## 2.5.4 ประเภทของการวัดผล: การวัดแบบอิงกลุ่มและอิงเกณฑ์

ยุทธศาสตร์การแปลความหมายคะแนนแบ่งได้เป็น 2 แนวทางหลักซึ่งเหมาะกับวัตถุประสงค์ต่างกัน คือ การวัดแบบอิงกลุ่ม (norm-referenced) และการวัดแบบอิงเกณฑ์ (criterion-referenced) (พิซิต ฤทธิ์จรรยา, 2564; ระพินทร์ โพธิ์ศรี, สุกัญญา รุจิเมธาภาส, และสุรเชษฐ์ บุญรักษ์, 2567) การเลือกแนวทางผิดต่อวัตถุประสงค์เป็นสาเหตุสำคัญที่ทำให้การตัดสินผลไม่เป็นธรรม จึงต้องเข้าใจความต่างดังตารางที่ 2.4 ทั้งนี้ การนำหลักอิงกลุ่ม-อิงเกณฑ์ไปใช้ตัดสินผลและตัดเกรด ศึกษาโดยละเอียดในบทที่ 13

ตารางที่ 2.4 เปรียบเทียบการวัดแบบอิงกลุ่มและแบบอิงเกณฑ์

| มิติเปรียบเทียบ | อิงกลุ่ม (Norm-referenced)        | อิงเกณฑ์ (Criterion-referenced)  |
|-----------------|-----------------------------------|----------------------------------|
| เป้าหมาย        | จำแนก/จัดลำดับผู้เรียนเทียบกับกัน | ตรวจสอบการบรรลุเกณฑ์/สมรรถนะ     |
| การแปลผล        | เทียบกับค่าเฉลี่ยหรือตำแหน่งใน    | เทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า |

|                | กลุ่ม                          |                                             |
|----------------|--------------------------------|---------------------------------------------|
| ตัวอย่างการใช้ | สอบแข่งขันคัดเลือก ชิงทุน      | สอบผ่านจุดประสงค์ การสอบใบขับขี่            |
| จุดเด่น        | คัดเลือกเมื่อนั่งจำกันได้      | สะท้อนว่าผู้เรียน “ทำอะไรได้” ชัดเจน        |
| ข้อจำกัด       | ไม่บอกว่าผู้เรียนรู้จริงแค่ไหน | ต้องตั้งเกณฑ์ที่เหมาะสม มิเช่นนั้นแปลผลคลาด |

ที่มา: สังเคราะห์จาก พิชิต ฤทธิ์เจริญ (2564), ระพินทร์ โพธิ์ศรี และคณะ (2567) และ Gronlund & Linn (1990)

เครื่องมือสำคัญที่ช่วยให้การวัดแบบอิงเกณฑ์มีความเที่ยงตรงเชิงเนื้อหาคือ ตารางวิเคราะห์หลักสูตร (Table of Specifications: TOS) หรือ “พิมพ์เขียวข้อสอบ” ซึ่งกำหนดสัดส่วนจำนวนข้อตามเนื้อหาและระดับพฤติกรรม ดังตัวอย่างในกล่องที่ 2.3

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>กล่องที่ 2.3 ตัวอย่างย่อตารางวิเคราะห์หลักสูตร (TOS) วิชาวิทยาศาสตร์ 20 ข้อ</b></p> <p>หน่วย “ระบบสุริยะ” : จำ 3 ข้อ / เข้าใจ 3 ข้อ / นำไปใช้ 2 ข้อ = 8 ข้อ</p> <p>หน่วย “แรงและการเคลื่อนที่” : จำ 2 ข้อ / เข้าใจ 4 ข้อ / นำไปใช้ 2 ข้อ = 8 ข้อ</p> <p>หน่วย “พลังงาน” : จำ 1 ข้อ / เข้าใจ 2 ข้อ / นำไปใช้ 1 ข้อ = 4 ข้อ</p> <p>ประโยชน์: คุณให้ข้อสอบกระจายตรงเนื้อหาที่สอนและระดับพฤติกรรมตามจุดประสงค์<br/>จึงเพิ่มความเที่ยงตรงเชิงเนื้อหา</p> <p><b>ที่มา:</b> ดัดแปลงจาก พัทรินทร์ ชมภูวิเศษ (2559) และ พิชิต ฤทธิ์เจริญ (2564)</p> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

ตัวอย่างในชั้นเรียนไทย: การสอบคัดเลือกเข้าศึกษาต่อชั้น ม.1 ของโรงเรียนที่มีผู้สมัครมากกว่าที่นั่งเป็นการวัดแบบอิงกลุ่ม เพราะต้องจัดลำดับเพื่อคัดผู้ที่ได้คะแนนสูงสุดตามจำนวนที่รับได้ ในทางกลับกัน การสอบเก็บคะแนนหลังเรียนหน่วย “การอ่านจับใจความ” ที่กำหนดว่าต้องได้ร้อยละ 70 จึงผ่าน เป็นการวัดแบบอิงเกณฑ์ เพราะสนใจว่าผู้เรียนบรรลุสมรรถนะหรือไม่ ไม่สนใจว่าใครได้อันดับเท่าใด

การประยุกต์และข้อควรระวัง: ในยุคการศึกษาฐานสมรรถนะ การประเมินในชั้นเรียนควรเน้นอิงเกณฑ์เป็นหลัก เพื่อบอกได้ว่าผู้เรียนแต่ละคน “ทำอะไรได้” แล้ววางแผนพัฒนาเป็นรายบุคคล ข้อควรระวังคืออย่านำคะแนนอิงกลุ่มมาตัดสินการบรรลุจุดประสงค์ เช่น ให้เกรดตามการจัดอันดับในห้องที่ทุกคนอาจบรรลุเกณฑ์แล้ว ระพินทร์ โพธิ์ศรี และคณะ (2567) เปรียบว่าหากเนื้อหาเป็นกรอบเล็กและเป็นเนื้อเดียวกันควรแปลผลแบบอิงเกณฑ์ แต่หากกว้างและหลากหลายมากจึงพิจารณาอิงกลุ่มประกอบ

## 2.5.5 ประเภทของแบบทดสอบ: แบบทดสอบผลสัมฤทธิ์และแบบทดสอบความถนัด

นอกจากจำแนกตามวิธีแปลความหมายคะแนน (อิงกลุ่ม/อิงเกณฑ์) แล้ว ยังจำแนกแบบทดสอบตามเป้าหมายของสิ่งที่มุ่งวัดได้อีก 2 ประเภท ซึ่งครูมักสับสน ได้แก่ แบบทดสอบผลสัมฤทธิ์และแบบทดสอบความถนัด (ซวาล แพร์ตกุล, 2520; ศิริชัย กาญจนวาสี, 2552)

แบบทดสอบผลสัมฤทธิ์ (Achievement Test) วัดสิ่งที่ผู้เรียนได้เรียนรู้มาแล้วในอดีต (past) เพื่อตอบว่า “รู้เท่าใด” เช่น การสอบปลายภาค การสอบเก็บคะแนนหลังเรียนจบหน่วย

แบบทดสอบความถนัด (Aptitude Test) วัดศักยภาพหรือแนวโน้มความสามารถในอนาคต (future) เพื่อพยากรณ์ว่า “จะไปได้ไกลเพียงใด” เช่น แบบวัดความถนัดทางการเรียน (scholastic aptitude) หรือความถนัดทางช่าง

## ตารางที่ 2.5 เปรียบเทียบแบบทดสอบผลสัมฤทธิ์กับแบบทดสอบความถนัด

| <p>กล่องที่ 2.4 อุปมาเชิงวิชาการ: คนกับรถยนต์</p> <p>“ความถนัด” เปรียบได้กับกำลังของเครื่องยนต์ (สมรรถภาพที่ซ่อนอยู่) ว่ารถคันนี้มีศักยภาพวิ่งได้เร็วเพียงใด</p> <p>“ผลสัมฤทธิ์” เปรียบได้กับระยะทางที่รถวิ่งมาแล้วจริง ว่าได้เดินทางมาไกลเท่าใดจากการเรียนที่ผ่านมา</p> <p>บทเรียนสำหรับครู: เด็กที่ผลสัมฤทธิ์ยังต่ำอาจมีความถนัดสูง เพียงแต่ยังไม่ได้มีโอกาส/การสอนที่เหมาะสม</p> <p>ที่มา: ดัดแปลงจาก ขวาล แพร์ตกุล (2520)</p> |                              |                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------------------------|
| มิติเปรียบเทียบ                                                                                                                                                                                                                                                                                                                                                                                                                   | แบบทดสอบผลสัมฤทธิ์           | แบบทดสอบความถนัด                  |
| มุ่งวัด                                                                                                                                                                                                                                                                                                                                                                                                                           | สิ่งที่เรียนรู้มาแล้ว (อดีต) | ศักยภาพ/แนวโน้มในอนาคต            |
| คำถามหลัก                                                                                                                                                                                                                                                                                                                                                                                                                         | รู้เท่าใด                    | จะไปได้ไกลเพียงใด                 |
| จุดมุ่งหมายการใช้                                                                                                                                                                                                                                                                                                                                                                                                                 | ตัดสินผลการเรียน/วินิจัย     | พยากรณ์ความสำเร็จ/แนะแนว          |
| ตัวอย่าง                                                                                                                                                                                                                                                                                                                                                                                                                          | สอบปลายภาควิชาคณิตศาสตร์     | แบบวัดความถนัดทางการเรียน/ทางช่าง |

ที่มา: สังเคราะห์จาก ขวาล แพร์ตกุล (2520), ศิริชัย กาญจนวาสี (2552) และ Nitko & Brookhart (2019)

ตัวอย่างในชั้นเรียนไทย: เมื่อครูแนะแนวต้องการช่วยนักเรียน ม.3 เลือกสายการเรียน ครูจะใช้ผลสัมฤทธิ์ (เกรดที่ผ่านมา) ร่วมกับแบบวัดความถนัด เพื่อดูทั้ง “สิ่งที่ทำได้แล้ว” และ “แนวโน้มที่จะไปได้ดี” การใช้เพียงผลสัมฤทธิ์อย่างเดียวอาจปิดโอกาสเด็กที่เริ่มช้าแต่มีศักยภาพสูง

การประยุกต์และข้อควรระวัง: ข้อควรระวังคืออย่านำแบบทดสอบผลสัมฤทธิ์ไปใช้ “พยากรณ์” อนาคตโดยตรง เพราะคนละจุดมุ่งหมายกัน และอย่าตีตราเด็กจากคะแนนความถนัดเพียงครั้งเดียว เนื่องจากความถนัดยังพัฒนาได้ด้วยการสอนและสภาพแวดล้อม ครูควรใช้ข้อมูลทั้งสองประเภทประกอบกันเพื่อการแนะแนวและออกแบบการสอนที่เป็นธรรม (Nitko & Brookhart, 2019)

## 2.5.6 ประเภทของการประเมินตามช่วงเวลา (ก่อน-ระหว่าง-หลังเรียน)

เมื่อจำแนกตามช่วงเวลาและจุดมุ่งหมายในการนำไปใช้ การประเมินแบ่งออกได้หลายประเภท โดยแนวคิดการแยกการประเมินเพื่อพัฒนา (formative) ออกจากการประเมินเพื่อตัดสิน (summative) มีรากฐานจาก Scriven (1967, อ้างถึงใน พัทธินันท์ ชมภูวิเศษ, 2559) และได้รับการตอกย้ำด้วยหลักฐานเชิง

ประจักษ์ว่า การประเมินเพื่อพัฒนาที่มีข้อมูลย้อนกลับช่วยยกระดับการเรียนรู้ได้อย่างมีนัยสำคัญ (Black & Wiliam, 1998) ความสัมพันธ์ตามช่วงเวลาแสดงดังภาพที่ 2.2

ภาพที่ 2.2 ไทม์ไลน์ประเภทการประเมินตามช่วงเวลาของการเรียนการสอน

| ก่อนเรียน                                                                                                                       | ระหว่างเรียน                                                                                                   | หลังเรียน                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• ประเมินเพื่อจัดวาง (Placement)</li> <li>• ประเมินเพื่อวินิจฉัย (Diagnostic)</li> </ul> | <ul style="list-style-type: none"> <li>• ประเมินเพื่อพัฒนา (Formative) ให้ข้อมูลย้อนกลับ ปรับการสอน</li> </ul> | <ul style="list-style-type: none"> <li>• ประเมินเพื่อตัดสิน (Summative) ตัดสินผล/รายงานผล</li> </ul> |
| ดูพื้นฐานความพร้อม และจุดบกพร่องก่อนสอน                                                                                         | ติดตามความก้าวหน้า ระหว่างกระบวนการเรียน                                                                       | สรุปผลสัมฤทธิ์ เมื่อสิ้นสุดการเรียนรู้                                                               |

ที่มา: คัดแปลงจาก พัทธวิเศษ (2559) อ้างอิงแนวคิดของ Scriven (1967) และ Black & Wiliam (1998)

การประเมินเพื่อจัดวาง (Placement) ประเมินพื้นฐานและความพร้อมก่อนเริ่มเรียน เพื่อจัดกลุ่มหรือออกแบบการสอนให้เหมาะสมกับผู้เรียน

การประเมินเพื่อวินิจฉัย (Diagnostic) ค้นหาจุดบกพร่องและสาเหตุของปัญหาการเรียนรู้ เพื่อวางแผนสอนซ่อมเสริมอย่างตรงจุดการประเมินเพื่อพัฒนา (Formative) ประเมินระหว่างกระบวนการเพื่อให้ข้อมูลย้อนกลับและปรับปรุงการสอน เช่น การใช้คำถาม Exit ticket หรือแบบฝึกย่อย

การประเมินเพื่อตัดสิน (Summative) ประเมินเมื่อสิ้นสุดการเรียนรู้เพื่อตัดสินผลและรายงานผล เช่น การสอบปลายภาคตัวอย่างในชั้นเรียนไทย: ครูภาษาไทยเริ่มหน่วย “การเขียนเรียงความ” ด้วยการให้เขียนย่อหน้าสั้น ๆ เพื่อดูพื้นฐาน (placement/diagnostic) ระหว่างหน่วยใช้การให้ข้อมูลย้อนกลับต่อรายงานเขียนแต่ละครั้งโดยยังไม่ให้คะแนนตัดสิน (formative) และเก็บคะแนนเรียงความฉบับสมบูรณ์เมื่อจบหน่วย (summative) การจัดวางเช่นนี้ทำให้คะแนนปลายทางสะท้อนพัฒนาการที่เกิดขึ้นจริง

**การประยุกต์และข้อควรระวัง:** ข้อผิดพลาดที่พบบ่อยคือการนำการประเมินระหว่างเรียนทุกครั้งไปคิดเป็นคะแนนตัดสิน ทำให้ผู้เรียนกลัวความผิดพลาดและไม่กล้าลองผิดลองถูก ซึ่งบั่นทอนประโยชน์ของการประเมินเพื่อพัฒนา Carney และคณะ (2022) ชี้ว่าครูควรเพิ่มน้ำหนักการประเมินเพื่อพัฒนาและคุณภาพของข้อมูลย้อนกลับให้มากขึ้น โดยแยกบทบาทของการให้ข้อมูลย้อนกลับออกจากการให้เกรดอย่างชัดเจน

## 2.5.7 ประเภทของการประเมินตามวัตถุประสงค์ (Assessment of / for / as Learning)

ภายใต้แนวคิดการศึกษาอิงผลลัพธ์ (Outcome-Based Education: OBE) บทบาทของการประเมินขยายจากการ “ตัดสิน” ไปสู่การ “พัฒนา” และการ “กำกับการเรียนรู้ด้วยตนเอง” จึงจำแนกได้เป็นสามมุมมอง

ที่เสริมกัน (Earl, 2003; สุทธิพงศ์ บุญผดุง, 2568) ดังตารางที่ 2.6 ทฤษฎีและกลยุทธ์ของการประเมินเพื่อพัฒนา และการให้ข้อมูลย้อนกลับ ศึกษาเชิงลึกในบทที่ 11

ตารางที่ 2.6 เปรียบเทียบการประเมิน Of / For / As Learning

| ประเภท                           | จุดมุ่งหมายหลัก                          | ผู้มีบทบาทหลัก           | ตัวอย่างกิจกรรม                             |
|----------------------------------|------------------------------------------|--------------------------|---------------------------------------------|
| Of Learning<br>(เพื่อตัดสิน)     | สรุป/ตัดสินผลสัมฤทธิ์                    | ครู                      | สอบปลายภาค ตัดเกรด                          |
| For Learning<br>(เพื่อพัฒนา)     | ให้ข้อมูลย้อนกลับเพื่อปรับการเรียนการสอน | ครู<br>(ร่วมกับผู้เรียน) | ตรวจรายงาน<br>ให้คำแนะนำระหว่างเรียน        |
| As Learning<br>(เพื่อกำกับตนเอง) | ผู้เรียนประเมินและกำกับการเรียนรู้ของตน  | ผู้เรียน                 | ประเมินตนเองด้วยรูบริก<br>ตั้งเป้าหมายถัดไป |

ที่มา: สังเคราะห์จาก Earl (2003), Black & Wiliam (1998) และ สุทธิพงศ์ บุญผดุง (2568)

ตัวอย่างในชั้นเรียนไทย: ในวิชาโครงงานวิทยาศาสตร์ ครูให้ผู้เรียนใช้รูบริกเดียวกับที่ครูใช้ตรวจ มาประเมินร่างโครงงานของตนเองก่อนส่ง (as learning) ครูให้ข้อมูลย้อนกลับเป็นข้อเสนอแนะระหว่างพัฒนางาน (for learning) และให้คะแนนชิ้นงานสมบูรณ์ตอนจบ (of learning) การวางสามบทบาทร่วมกันช่วยให้ผู้เรียนเป็นเจ้าของการเรียนรู้มากขึ้น

การประยุกต์และข้อควรระวัง: โรงเรียนไทยจำนวนมากยังหนักไปทาง of learning จนละเลย for/as learning ทำให้การประเมินกลายเป็นเพียงการเก็บคะแนน ข้อเสนอแนะคือเพิ่มโอกาสประเมินเพื่อพัฒนาและฝึกให้ผู้เรียนประเมินตนเองอย่างมีเกณฑ์ Carney และคณะ (2022) ย้ำว่าการประเมินเพื่อพัฒนาให้ผลคุ้มค่าต่อการเรียนรู้ แต่ต้องลงทุนกับการออกแบบข้อมูลย้อนกลับที่ผู้เรียนนำไปใช้ได้จริง

## 2.5.8 การประเมินตามสภาพจริงและแนวโน้มการประเมินในศตวรรษที่ 21

การประเมินตามสภาพจริง (Authentic Assessment) เน้นวัดสมรรถนะของผู้เรียนผ่านภาระงานที่ใกล้เคียงสถานการณ์จริง และครอบคลุมทั้งกระบวนการและผลงาน มิใช่วัดเพียงความจำด้วยข้อสอบปรนัย (Wiggins, 1998; พัทรินทร์ ชมภูวิเศษ, 2559) ลักษณะเด่นเมื่อเทียบกับการสอบแบบดั้งเดิมแสดงในตารางที่ 2.7 หลักการออกแบบการประเมินตามสภาพจริงและการสร้างรูบริกฉบับเต็ม ศึกษาในบทที่ 10

ตารางที่ 2.7 เปรียบเทียบการสอบแบบดั้งเดิมกับการประเมินตามสภาพจริง

| ประเด็น    | การสอบแบบดั้งเดิม      | การประเมินตามสภาพจริง            |
|------------|------------------------|----------------------------------|
| สิ่งที่วัด | ความรู้-ความจำเป็นหลัก | การนำความรู้ไปใช้ในสถานการณ์จริง |
| ลักษณะงาน  | เลือกตอบ/ตอบสั้น       | ภาระงาน/ชิ้นงาน/การปฏิบัติ       |

|                    |                 |                         |
|--------------------|-----------------|-------------------------|
| บริบท              | แยกจากชีวิตจริง | เชื่อมโยงกับชีวิต/ชุมชน |
| เครื่องมือให้คะแนน | เฉลยถูก-ผิด     | rubric หลายมิติ         |

ที่มา: สังเคราะห์จาก Wiggins (1998), Nitko & Brookhart (2019) และ พชรินทร์ ชมภูวิเศษ (2559)

### กล่องที่ 2.5 กรณีศึกษา: วิชาวิทยาศาสตร์ เรื่อง “มลพิษในชุมชน”

แผนการสอบปรนัย 4 ตัวเลือก ครูให้ผู้เรียนสำรวจแหล่งน้ำในชุมชน เก็บข้อมูล วิเคราะห์ และเสนอแนวทางการแก้ไขเป็นโครงการ วัตถุประสงค์การเรียนรู้วิทยาศาสตร์ ทักษะกระบวนการ การคิดวิเคราะห์ และการสื่อสาร ด้วย rubric หลายมิติ

ข้อดี: สะท้อนสมรรถนะจริงและเชื่อมโยงชีวิต / ข้อจำกัด: ใช้เวลาและทรัพยากรสูง ต้องคุมความเป็นปรนัยด้วย rubric

ที่มา: ดัดแปลงจาก พชรินทร์ ชมภูวิเศษ (2559) และ Wiggins (1998)

แนวโน้มการประเมินในศตวรรษที่ 21 กำลังเคลื่อนสู่สามทิศทางหลักที่ครูยุคใหม่ควรเตรียมพร้อม ดังนี้

- การประเมินฐานสมรรถนะ (Competency-based Assessment): เน้นสิ่งที่ผู้เรียนทำได้จริงมากกว่ากระดาษคำตอบ
- การวิเคราะห์การเรียนรู้ (Learning Analytics): ใช้ข้อมูล/แดชบอร์ดติดตามพัฒนาการรายบุคคลเพื่อวินิจฉัยปัญหาได้ทันทั่วทั้ง
- ปัญญาประดิษฐ์ในการประเมิน (AI in Assessment): ช่วยตรวจคำตอบอัตโนมัติและวิเคราะห์จุดแข็ง-จุดอ่อนรายบุคคล โดยครูยังต้องกำกับความเป็นธรรม

**ตัวอย่างในชั้นเรียนไทย:** ครูใช้แบบทดสอบออนไลน์ (เช่น Quizizz/Kahoot) เก็บข้อมูลการตอบของนักเรียนทั้งห้องแบบทันที แล้วใช้แดชบอร์ดดูว่าข้อใดผิดมากที่สุด เพื่อกลับไปสอนซ้ำเฉพาะจุด เป็นการนำ learning analytics อย่างง่ายมาใช้ในชั้นเรียนไทยได้จริง

**การประยุกต์และข้อควรระวัง:** แม้เทคโนโลยีและ AI ช่วยเพิ่มประสิทธิภาพ แต่ครูต้องระวังเรื่องความเป็นธรรม ความเป็นส่วนตัวของข้อมูล และอคติของระบบ และพึงตระหนักว่าการประเมินตามสภาพจริงต้องอาศัย rubric ที่ชัดเจนเพื่อคงความเป็นปรนัย ความรอบรู้ด้านการประเมินของครูจึงเป็นปัจจัยชี้ขาดว่าจะใช้เครื่องมือเหล่านี้ให้เกิดประโยชน์ต่อผู้เรียนได้เพียงใด (พิเชษฐ์ พิณิจ และคณะ, 2566; Nitko & Brookhart, 2019)

## 2.6 กิจกรรมการเรียนการสอนและใบงาน

### กิจกรรมตามขั้นการสอน

**ก่อนเรียน** — ทำแบบทดสอบก่อนเรียน และกิจกรรม “ความยุติธรรมที่มองไม่เห็น”: ครูจำลองการให้คะแนนโดยไม่มีเกณฑ์ ให้นักศึกษาวิจารณ์ผลกระทบต่อความยุติธรรม

ระหว่างเรียน — กิจกรรม “นักสืบมาตราการวัด”: วิเคราะห์ข้อมูลดิบ 10 ชนิด (เช่น รายได้ เลข โทรศัพท์ เกรดเฉลี่ย อันดับที) แล้วระบุระดับมาตราการวัด และวิเคราะห์กรณีศึกษาการละเมิดหลักการ วัดผลแบบ Think-Pair-Share

หลังเรียน — ทำใบงานจำแนกประเภท สรุปเป็นผังมโนทัศน์ และทำแบบทดสอบหลังเรียน  
ใบงานที่ 2.1 จำแนกประเภท (เครื่องมือวัด CLO1)

ตอน ก — จำแนกมาตราการวัด: ระบุว่าข้อมูลต่อไปนี้เป็นมาตราระดับใด (นามบัญญัติ/เรียงอันดับ/ อนุภาค/อัตราส่วน) พร้อมให้เหตุผล: (1) หมายเลขเสื้อนักกีฬา (2) อันดับการประกวด (3) คะแนนสอบกลาง ภาค (4) น้ำหนักตัว (5) ระดับความพึงพอใจ 5 ระดับ

ตอน ข — จำแนกประเภทการประเมิน: ระบุว่าสถานการณ์ต่อไปนี้เป็น Assessment of/for/as Learning พร้อมเหตุผล: (1) สอบปลายภาคเพื่อตัดเกรด (2) ครูเดินสังเกตและจดจุดที่ผู้เรียนยังไม่เข้าใจเพื่อ อธิบายซ้ำ (3) ผู้เรียนใช้รูปบัตรตรวจงานตนเองก่อนส่ง (4) การสอบคัดเลือกเข้าศึกษาต่อ (5) ผู้เรียนบันทึก สะท้อนการเรียนรู้ลงสมุด

ตอน ค — วิเคราะห์การละเมิดหลักการ: ระบุว่าแต่ละสถานการณ์ละเมิดหลักการใดใน 5 ประการ พร้อมเสนอแนวทางแก้ไข (เช่น สอบว่ายน้ำด้วยการให้เขียนอธิบายว่ายว่ายน้ำ; ตรวจภาพวาดโดยไม่มีเกณฑ์)

## 2.7 สรุปบทเรียน

---

- การทดสอบคือการกระตุ้นพฤติกรรม การวัดผลคือการให้ตัวเลขตามกฎเกณฑ์ และการประเมินผลคือการตัดสินคุณค่าเทียบกับเกณฑ์
- องค์ประกอบของการวัดมี 3 ส่วน: สิ่งที่วัด เครื่องมือวัด และผลการวัด
- มาตราการวัดมี 4 ระดับ: นามบัญญัติ เรียงอันดับ อนุภาค และอัตราส่วน โดยคะแนนสอบส่วนใหญ่เป็นมาตราอนุภาค (ไม่มีศูนย์แท้)
- หลักการคุณภาพ 5 ประการ ได้แก่ ความเที่ยงตรง ความเชื่อมั่น ความเป็นปรนัย ความยุติธรรม และความเป็นไปได้ โดยความเที่ยงตรงสำคัญที่สุด
- การวัดแบบอิงกลุ่มเน้นจัดลำดับ/คัดเลือก ส่วนอิงเกณฑ์เน้นการบรรลุสมรรถนะ และ TOS ช่วยประกันความเที่ยงตรงเชิงเนื้อหา
- แบบทดสอบผลสัมฤทธิ์วัดสิ่งที่เรียนรู้มาแล้ว (อดีต) ส่วนแบบทดสอบความถนัดวัดศักยภาพเพื่อพยากรณ์อนาคต
- การประเมินจำแนกตามช่วงเวลา (ก่อนเรียน-วินิจัย-ระหว่างเรียน-สรุปผล) และตามวัตถุประสงค์ (of/for/as learning)
- การประเมินตามสภาพจริงสะท้อนสมรรถนะเชิงบูรณาการ และเทคโนโลยี/AI ช่วยเพิ่มประสิทธิภาพและความเป็นปรนัยในยุคปัจจุบัน

## 2.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

## 2.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

### ตอนที่ 1 คำถามอัตนัยเชิงวิเคราะห์

1. ในการสอบปลายภาควิชาสังคมศึกษา ครูสรุปว่า "เด็กหญิงบีได้ 18 จาก 30 คะแนน ไม่ผ่าน ต้องเรียนซ่อมเสริม" จงจำแนกไว้ในสถานการณ์นี้ขั้นใดคือการทดสอบ การวัดผล และการประเมินผล และการตัดสินว่า "ไม่ผ่าน" ต้องอาศัยสิ่งใดเพิ่มเติมจากตัวคะแนนดิบ
2. เหตุใดคะแนนสอบจึงจัดเป็นมาตรา “อันตรภาค” ไม่ใช่ “อัตราส่วน” จงอธิบายแนวคิดศูนย์สมมุติพร้อมตัวอย่าง
3. จงวิเคราะห์ความสัมพันธ์ระหว่างความเที่ยงตรงและความเชื่อมั่น หากเครื่องมือเชื่อมั่นสูงแต่ขาดความเที่ยงตรงจะส่งผลอย่างไร
4. เปรียบเทียบแบบทดสอบผลสัมฤทธิ์กับแบบทดสอบความถนัด พร้อมยกตัวอย่างการนำไปใช้ของแต่ละประเภท
5. ในบริบทการศึกษาระดับสมรรถนะ ท่านจะออกแบบการประเมินตามสภาพจริงเรื่องใดเรื่องหนึ่ง และรักษาความเป็นปรนัยอย่างไร

### ตอนที่ 2 กิจกรรมปฏิบัติ

ให้นักศึกษาทำใบงานที่ 2.1 (ตอน ก-ค) ให้ครบถ้วน แล้วนำเสนอผลการจำแนกพร้อมเหตุผลในชั้นเรียน เพื่อรับข้อมูลย้อนกลับจากเพื่อนและผู้สอน

## 2.10 เครื่องมือวัดและประเมินผลประจำบท

บทนี้ประเมิน CLO1 ด้วยเครื่องมือที่เหมาะสมกับระดับพฤติกรรมจำ-เข้าใจ ได้แก่ แบบทดสอบ (จำ-เข้าใจ) และใบงานจำแนกประเภท (เข้าใจ-จำแนก) โดยมีสัดส่วนคะแนนรวม 100% ดังนี้

| องค์ประกอบการประเมิน             | CLO  | เครื่องมือ             | น้ำหนัก (%) |
|----------------------------------|------|------------------------|-------------|
| แบบทดสอบหลังเรียน (ปรนัย 11 ข้อ) | CLO1 | แบบทดสอบออนไลน์        | 40          |
| ใบงานจำแนกประเภท (ตอน ก-ค)       | CLO1 | ใบงาน + RUBRIC         | 40          |
| การมีส่วนร่วมในกิจกรรม/อภิปราย   | CLO1 | แบบสังเกตการมีส่วนร่วม | 20          |
| รวม                              |      |                        | 100         |

## รูบริกการประเมินใบงานจำแนกประเภท (Scoring Rubric)

| ระดับคุณภาพ  | เกณฑ์การพิจารณา (ความถูกต้องของการจำแนก + เหตุผลเชิงทฤษฎี)                          |
|--------------|-------------------------------------------------------------------------------------|
| 4 (ดีมาก)    | จำแนกถูกต้องเกือบทั้งหมด ( $\geq 90\%$ ) ให้เหตุผลอ้างอิงหลักการได้ลึกซึ้งและชัดเจน |
| 3 (ดี)       | จำแนกถูกต้อง 75–89% ให้เหตุผลได้ถูกต้องชัดเจนเป็นส่วนใหญ่                           |
| 2 (พอใช้)    | จำแนกถูกต้อง 60–74% ให้เหตุผลได้บางส่วน ยังมีความสับสนในบางโน้ต                     |
| 1 (ปรับปรุง) | จำแนกถูกต้อง $< 60\%$ หรือให้เหตุผลไม่สอดคล้องกับหลักการ                            |

เกณฑ์ผ่าน: ผู้เรียนต้องได้ระดับคุณภาพ “3 (ดี)” ขึ้นไป จึงถือว่าผ่านการประเมินใบงานนี้

### แนวเฉลยใบงานที่ 2.1 (สำหรับผู้สอน)

ตอน ก: (1) นามบัญญัติ (2) เรียงอันดับ (3) อันตรภาค (4) อัตราส่วน (5) เรียงอันดับ (อนุโลมเป็นอันตรภาค)

ตอน ข: (1) of learning (2) for learning (3) as learning (4) of learning (5) as learning

ตอน ค (ตัวอย่าง): สบว้ายน้ำด้วยการเขียน = ขาดความเที่ยงตรง (วัดไม่ตรงทักษะปฏิบัติ)

แก้ไขให้ปฏิบัติจริงพร้อมรูบริก; ตรวจภาพวาดโดยไม่มีเกณฑ์ = ขาดความเป็นปรนัย แก้โดยใช้รูบริกที่ชัดเจน

## 2.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

ชวาล แพรัตกุล. (2520). *เทคนิคการวัดผล* (พิมพ์ครั้งที่ 5). วัฒนาพานิช.

พัชรินทร์ ชมภูวิเศษ. (2559). *การวัดและประเมินผลการศึกษา*. คณะครุศาสตร์

มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์เจริญ. (2564). *เทคนิคการวัดและประเมินผลการเรียนรู้*. ศูนย์หนังสือจุฬาลงกรณ์

มหาวิทยาลัย.

พิเชษฐ์ พิณใจ, อนุศิษฐ์ อันมานะตระกูล, และเอกรัตน์ รวยรวย. (2566).

การรู้การประเมินของครูกับการจัดการเรียนรู้ฐานสมรรถนะ. *วารสารนวัตกรรมการเรียนรู้และเทคโนโลยี*, 3(2), 6–20.

ไพศาล หวังพานิช. (2526). *การวัดผลการศึกษา*. ไทยวัฒนาพานิช.

ระพีพันธ์ โพธิ์ศรี, สุภิญญา รุจิเมธาส, และสุรเชษฐ์ บุญรักษ์. (2567).

การวัดและประเมินผลแบบอิงกลุ่มและแบบอิงเกณฑ์. *วารสารมหาวิทยาลัยราชภัฏมหาสารคาม*, 18(1), 13–22.

ศิริชัย กาญจนวาสี. (2552). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 6).

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุทธิพงศ์ บุญผดุง. (2568). *เอกสารประกอบการสอนรายวิชาการวัดและประเมินผลการเรียนรู้*.

คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

## ภาษาอังกฤษ

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

Carney, E. A., Zhang, X., Charsha, A., Taylor, J. N., & Hoshaw, J. P. (2022). Formative assessment helps students learn over time: Why aren't we paying more attention to it? *Intersection: A Journal at the Intersection of Assessment and Learning*, 4(1).

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). Macmillan.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.

## บทที่ 3

### จุดประสงค์การเรียนรู้และการจำแนกพฤติกรรมตามแนวคิดของบลูม

(Bloom's Taxonomy)

### Learning Objectives and the Classification of Learning Behaviors

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สัปดาห์ที่ 3 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

รหัสวิชา ED1006 • จำนวน 4 คาบ (ทฤษฎี 2 คาบ / ปฏิบัติ 2 คาบ)

ผลลัพธ์การเรียนรู้หลักของบท: CLO1 และ CLO2

### แผนบริหารการสอนประจำบทที่ 3

จุดประสงค์การเรียนรู้และการจำแนกพฤติกรรมตามแนวคิดของบลูม (Bloom's Taxonomy)

#### 1. หัวข้อสาระจัดการเรียนการสอน

- ความหมาย ความสำคัญ และระดับของจุดประสงค์การเรียนรู้ (จุดมุ่งหมายทั่วไปและจุดประสงค์เชิงพฤติกรรม)
- ความสัมพันธ์เชิงระบบของการศึกษาและหลักความสอดคล้องเชิงโครงสร้าง (Constructive Alignment)
- การจำแนกพฤติกรรมการเรียนรู้ 3 ด้าน: พุทธิพิสัย จิตพิสัย และทักษะพิสัย
- Bloom's Taxonomy ฉบับเดิม (1956) และฉบับปรับปรุง (Anderson & Krathwohl, 2001)
- หลักการเขียนจุดประสงค์เชิงพฤติกรรม: แนวคิดของเมเกอร์และโมเดล ABCD พร้อมคำกริยาเชิงพฤติกรรม (Action Verbs)
- การจับคู่ระดับพฤติกรรมกับเครื่องมือวัด และตารางวิเคราะห์ข้อสอบ (Table of Specifications: TOS)

## 2. ความคิดรวบยอด

จุดประสงค์การเรียนรู้คือ “เข้มทิศ” ของการจัดการเรียนการสอน เป็นข้อความที่ระบุพฤติกรรม ที่คาดหวังให้เกิดขึ้นกับผู้เรียนอย่าง “สังเกตได้และวัดได้” การจำแนกพฤติกรรมตามแนวคิดของบลูมเป็นเครื่องมือ กำหนดระดับความซับซ้อนของพฤติกรรม ซึ่งเป็นรากฐานของความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) ที่ว่า “สอนสิ่งใด-วัดสิ่งนั้น และวัดในระดับพฤติกรรมที่ตั้งไว้”

## 3. จุดประสงค์การเรียนรู้

1. อธิบายความหมาย ความสำคัญ และระดับของจุดประสงค์การเรียนรู้ได้ถูกต้อง (CLO1)
2. อธิบายพัฒนาการและระดับพฤติกรรมตาม Bloom's Taxonomy ทั้งฉบับเดิมและฉบับปรับปรุงได้ (CLO1)
3. จำแนกระดับพฤติกรรมด้านพุทธิพิสัย จิตพิสัย และทักษะพิสัย จากสถานการณ์ที่กำหนดให้ได้ (CLO2)
4. วิเคราะห์ความสอดคล้องระหว่างจุดประสงค์ ระดับพฤติกรรม และเครื่องมือวัดได้อย่างมีเหตุผล (CLO2)

## 4. เนื้อหาวิชา

เนื้อหาโดยละเอียดเป็นไปตามเอกสารประกอบการสอนบทที่ 3 หัวข้อ 3.5 ซึ่งครอบคลุม ความหมายและความสำคัญของจุดประสงค์ การจำแนกพฤติกรรม 3 ด้าน ทฤษฎีของบลูมฉบับเดิมและฉบับปรับปรุง หลักการเขียนจุดประสงค์เชิงพฤติกรรม (ABCD) และการเลือกเครื่องมือวัดให้สอดคล้องกับระดับพฤติกรรม

## 5. กิจกรรมการเรียนการสอน

1. ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน และกิจกรรม “Action Verb” ระดมคำกริยาที่วัดได้/วัดไม่ได้
2. ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ + อภิปรายกรณีศึกษา + Workshop  
ฝึกจำแนกระดับพฤติกรรมและวิเคราะห์จุดประสงค์
3. ขั้นระหว่างเรียน: กิจกรรม Gallery Walk/Peer Review วิพากษ์ความสอดคล้องของจุดประสงค์-ระดับพฤติกรรม-เครื่องมือวัด
4. ขั้นหลังเรียน: ทำใบงานวิเคราะห์ความสอดคล้อง ทำแบบทดสอบหลังเรียน และสะท้อนคิด (Reflection)

## 6. สื่อการเรียนการสอน

- เอกสารประกอบการสอนบทที่ 3 และสไลด์นำเสนอ (PowerPoint)
- ใบงานวิเคราะห์จุดประสงค์เชิงพฤติกรรมและตารางจำแนกพฤติกรรมตามบลูม
- ตัวอย่างมาตรฐาน/ตัวชี้วัดจากหลักสูตรแกนกลาง และคลังคำกริยาเชิงพฤติกรรม (Action Verbs)
- รูบรีค (Analytical Rubric) สำหรับประเมินการวิเคราะห์จุดประสงค์

## 7. การวัดและประเมินผล

- CLO1 (จำ-เข้าใจ): แบบทดสอบก่อน-หลังเรียน (ปรนัย 10 ข้อ) — สัดส่วน 40%
- CLO2 (วิเคราะห์): ใบงานจำแนกพฤติกรรมและวิเคราะห์ความสอดคล้อง ประเมินด้วย Analytical Rubric — สัดส่วน 60%
- การสังเกตการมีส่วนร่วมในกิจกรรม (Formative) เพื่อให้ข้อมูลย้อนกลับระหว่างเรียน

### 3.1 ความคิดรวบยอด

การจัดการศึกษาที่มีคุณภาพดำเนินไปด้วยกระบวนการที่สัมพันธ์กันเป็นระบบ ซึ่งพัชรินทร์ ชมภูวิเศษ (2559) เรียกว่า “สามเหลี่ยมการศึกษา” ประกอบด้วยสามกระบวนการที่ต้องสมดุลและเชื่อมโยงกัน ได้แก่ (1) กระบวนการกำหนดวัตถุประสงค์ (Objective) (2) กระบวนการจัดประสบการณ์การเรียนรู้ (Learning) และ (3) กระบวนการประเมินผลการเรียนรู้ (Evaluation) โดยมี “จุดประสงค์การเรียนรู้” เป็นจุดเริ่มต้นที่กำหนดทิศทางของอีกสองกระบวนการที่เหลือ

หัวใจของการประกันคุณภาพภายในชั้นเรียนคือ ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) ตามแนวคิดของ Biggs และ Tang (2011) ที่กำหนดว่า ครูต้องจัดกิจกรรมและวัดผลให้ตรงกับระดับพฤติกรรมที่ระบุไว้ใน จุดประสงค์ หากตั้งจุดประสงค์ระดับ “วิเคราะห์” แต่สอนเพียงการบรรยายและวัดด้วยข้อสอบ “ความจำ” จะเกิดภาวะ ไม่สอดคล้อง (Misalignment) ทำให้ผลการวัดขาดความเที่ยงตรง (Validity) และไม่สะท้อนคุณภาพผู้เรียนที่แท้จริง (พิชิต ฤทธิ์จรูญ, 2564) การจำแนกพฤติกรรมตามแนวคิดของบลูมจึงเป็นเครื่องมือสำคัญที่ช่วยให้ครูกำหนด ระบุ และวัด พฤติกรรมได้ตรงตามระดับที่คาดหวัง

#### แก่นความคิดของบท

- จุดประสงค์การเรียนรู้ = ข้อความระบุพฤติกรรมที่ “สังเกตได้และวัดได้” ของผู้เรียนหลังเรียน
- พฤติกรรมการเรียนรู้มี 3 ด้าน: พุทธิพิสัย (สมอง) จิตพิสัย (จิตใจ) ทักษะพิสัย (ร่างกาย)
- บลูมฉบับปรับปรุง (2001) เปลี่ยนคำนามเป็นคำกริยา และเรียงลำดับสูงสุดเป็น “สร้างสรรค์”
- เขียนจุดประสงค์ให้ครบโมเดล ABCD แล้วเลือกเครื่องมือวัดให้ตรงระดับพฤติกรรม = Constructive Alignment

### 3.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทนี้จบแล้ว นักศึกษาสามารถ

1. อธิบายความหมาย ความสำคัญ และระดับของจุดประสงค์การเรียนรู้ได้อย่างถูกต้อง

2. อธิบายพัฒนาการและระดับพฤติกรรมตาม Bloom's Taxonomy ทั้งฉบับเดิมและฉบับปรับปรุง (Revised) ได้
3. จำแนกระดับพฤติกรรมด้านพุทธิพิสัย จิตพิสัย และทักษะพิสัย จากสถานการณ์ที่กำหนดให้ได้ถูกต้อง
4. วิเคราะห์ความสอดคล้องระหว่างจุดประสงค์ ระดับพฤติกรรม และเครื่องมือวัดผล โดยใช้หลักความสอดคล้องเชิงโครงสร้างได้

**ตารางที่ 3.1 ความเชื่อมโยงจุดประสงค์ประจำบทกับผลลัพธ์การเรียนรู้ระดับรายวิชา (CLO) กิจกรรม วิธีวัด เครื่องมือ และสัดส่วนคะแนน**

| จุดประสงค์ประจำบท                                   | CLO หลัก | ระดับพฤติกรรม          | กิจกรรม                         | วิธีวัด/เครื่องมือ                  | สัดส่วน |
|-----------------------------------------------------|----------|------------------------|---------------------------------|-------------------------------------|---------|
| ข้อ 1-2<br>อธิบายหลักการจุดประสงค์ และทฤษฎีบลูม     | CLO1     | จำ-เข้าใจ (พุทธิพิสัย) | บรรยายเชิงปฏิสัมพันธ์ / อภิปราย | แบบทดสอบก่อน-หลังเรียน (ปรนัย)      | 40%     |
| ข้อ 3 จำแนกพฤติกรรม 3 ด้านจากสถานการณ์              | CLO2     | วิเคราะห์ (พุทธิพิสัย) | Workshop<br>จำแนกพฤติกรรม       | ใบงานจำแนก + Analytical Rubric      | 25%     |
| ข้อ 4<br>วิเคราะห์ความสอดคล้องจุดประสงค์-เครื่องมือ | CLO2     | วิเคราะห์ (พุทธิพิสัย) | Peer Review / Gallery Walk      | ใบงานวิเคราะห์ความสอดคล้อง + Rubric | 35%     |

หมายเหตุ: บทนี้ประเมินเฉพาะ CLO1 และ CLO2 ตามผังความสอดคล้องของรายวิชา รวมสัดส่วนคะแนนประจำบทเท่ากับ 100% (CLO1 = 40%, CLO2 = 60%) ไม่มีจุดประสงค์ใดที่ไม่ถูกประเมิน และไม่ประเมินผลลัพธ์ที่มีได้ประกาศไว้

### 3.3 คำสำคัญ

| คำสำคัญ (ไทย)             | ภาษาอังกฤษ               | ความหมายโดยสังเขป                                   |
|---------------------------|--------------------------|-----------------------------------------------------|
| จุดประสงค์เชิงพฤติกรรม    | Behavioral Objective     | ข้อความระบุพฤติกรรมผู้เรียนที่สังเกตและวัดได้       |
| ความสอดคล้องเชิงโครงสร้าง | Constructive Alignment   | การจัดกิจกรรมและการวัดให้ตรงกับจุดประสงค์           |
| พุทธิพิสัย                | Cognitive Domain         | พฤติกรรมด้านสติปัญญา ความรู้ ความคิด                |
| จิตพิสัย                  | Affective Domain         | พฤติกรรมด้านความรู้สึก เจตคติ ค่านิยม               |
| ทักษะพิสัย                | Psychomotor Domain       | พฤติกรรมด้านทักษะการปฏิบัติและการเคลื่อนไหว         |
| อนุกรมวิธานของบลูม        | Bloom's Taxonomy         | ระบบจำแนกระดับพฤติกรรมการเรียนรู้                   |
| บลูมฉบับปรับปรุง          | Revised Bloom's Taxonomy | ฉบับ Anderson & Krathwohl (2001) เน้นคำกริยา 2 มิติ |

|                      |                               |                                                             |
|----------------------|-------------------------------|-------------------------------------------------------------|
| โมเดล ABCD           | ABCD Model                    | องค์ประกอบจุดประสงค์: Audience, Behavior, Condition, Degree |
| คำกริยาเชิงพฤติกรรม  | Action Verb                   | คำกริยาที่แสดงพฤติกรรมที่สังเกตและวัดได้                    |
| ตารางวิเคราะห์ข้อสอบ | Table of Specifications (TOS) | ตารางกระจายน้ำหนักเนื้อหา×พฤติกรรมเพื่อความเที่ยงตรง        |
| ความเที่ยงตรง        | Validity                      | คุณสมบัติของเครื่องมือที่วัดได้ตรงตามสิ่งที่ต้องการวัด      |

### 3.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอน ประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 3.5 เนื้อหาวิชา

#### 3.5.1 ความหมายและความสำคัญของจุดประสงค์การเรียนรู้

**นิยาม.** จุดประสงค์การเรียนรู้ (Learning Objective) หมายถึง ข้อความที่ระบุการเปลี่ยนแปลง พฤติกรรมที่ “สังเกตได้และวัดได้” ซึ่งคาดหวังให้เกิดขึ้นกับผู้เรียนหลังสิ้นสุดกระบวนการเรียนรู้ (พัชรินทร์ ชมภู วิชา, 2559) พัทธินทรขยายความว่า จุดประสงค์เชิงพฤติกรรม (Behavioral Objective) คือข้อความที่ระบุ “เกณฑ์ขั้นต่ำที่เฉพาะเจาะจง” ว่าหลังเรียนจบเนื้อหาหนึ่ง ผู้เรียนจะแสดงพฤติกรรมที่ครูวัดและสังเกตได้ อย่างไร ภายใต้เงื่อนไขและเกณฑ์ที่กำหนด ศาสตราจารย์ ดร.ชวาล แพร่ตฤกุล (2518) ประมาจารย์การวัดผล การศึกษาไทย จำแนกเป้าหมายทางการศึกษาออกเป็น 2 ระดับ คือ จุดมุ่งหมายทั่วไป (General Objective) และจุดประสงค์เชิงพฤติกรรม ดังเปรียบเทียบในตารางที่ 3.2

ตารางที่ 3.2 เปรียบเทียบจุดมุ่งหมายทั่วไปกับจุดประสงค์เชิงพฤติกรรม

| ประเด็นเปรียบเทียบ | จุดมุ่งหมายทั่วไป (General Objective) | จุดประสงค์เชิงพฤติกรรม (Behavioral Objective) |
|--------------------|---------------------------------------|-----------------------------------------------|
| ลักษณะ             | ความคาดหวังภาพกว้าง เป็นนามธรรม       | ระบุพฤติกรรมเฉพาะที่สังเกตและวัดได้           |

| ชนิดพฤติกรรม    | พฤติกรรมภายใน (Covert)<br>มองไม่เห็น  | พฤติกรรมภายนอก (Overt) สังเกตได้                                                                 |
|-----------------|---------------------------------------|--------------------------------------------------------------------------------------------------|
| ตัวอย่างคำ      | เข้าใจ ซาบซึ้ง ตระหนัก เห็นคุณค่า     | อธิบาย จำแนก คำนวณ เขียน<br>วิเคราะห์                                                            |
| ตัวอย่างข้อความ | เพื่อให้นักเรียนเข้าใจเรื่องระบบนิเวศ | เมื่อกำหนดภาพสายใยอาหารให้<br>นักเรียนเขียนอธิบาย<br>ความสัมพันธ์ได้ถูกต้องอย่างน้อย 3<br>รูปแบบ |
| การนำไปวัด      | วัดได้ยาก ตีความต่างกัน               | วัดได้ชัดเจน ลดความคลาดเคลื่อน                                                                   |

ที่มา: ดัดแปลงจาก ขวาล แพร่ดกุล (2518) และ พิชรินทร์ ชมภูวิเศษ (2559, น. 41)

**คำอธิบายเชิงลึก.** แก่นของความแตกต่างอยู่ที่ “ความสามารถในการสังเกต” จุดมุ่งหมายทั่วไปอาศัยค่านามธรรม เช่น เข้าใจ ซาบซึ้ง ตระหนัก ซึ่งเป็นพฤติกรรมภายใน (Covert Behavior) ที่เกิดในจิตใจผู้เรียนและมองไม่เห็นโดยตรง ครูสองคนจึงอาจตีความระดับ “ความเข้าใจ” ของนักเรียนคนเดียวกันไม่เท่ากัน ส่วนจุดประสงค์เชิงพฤติกรรมใช้คำกริยาที่บ่งบอกการกระทำภายนอก (Overt Behavior) เช่น อธิบาย จำแนก คำนวณ ซึ่งสังเกตและวัดได้ จึงทำหน้าที่เป็น “สะพาน” ที่แปลงเป้าหมายนามธรรมของหลักสูตรให้กลายเป็นหลักฐานเชิงประจักษ์ที่ตรวจสอบได้ การเขียนจุดประสงค์จึงไม่ใช่เรื่องของถ้อยคำสวยงาม แต่เป็นการกำหนด “หลักฐานความสำเร็จ” ล่วงหน้า

**ความสำคัญ.** จุดประสงค์ที่ชัดเจนส่งผลโดยตรงต่อความเที่ยงตรง (Validity) ของการวัดผล หากจุดประสงค์คลุมเครือ เครื่องมือวัดที่สร้างขึ้นย่อมขาดความแม่นยำและไม่อาจสะท้อนความสามารถจริงของผู้เรียน (สำเร็จ บุญเรืองรัตน์, 2534) สุทธิพงศ์ บุญผดุง (2568) ชี้ว่าจุดประสงค์ที่ดีทำหน้าที่สามด้านพร้อมกัน คือ เป็นทิศทางให้ครูออกแบบการสอน เป็นเกณฑ์ให้ครูเลือกและสร้างเครื่องมือวัด และเป็น “แผนที่นำทาง” ให้ผู้เรียนกำกับการเรียนรู้ของตนเอง (Self-regulated Learning) เมื่อผู้เรียนทราบความคาดหวังที่ชัดเจน ย่อมจัดสรรเวลาและความพยายามได้ตรงเป้าหมายมากขึ้น

**ตัวอย่างในชั้นเรียนไทย.** ครูคณิตศาสตร์ชั้น ม.2 เขียนจุดประสงค์ว่า “เพื่อให้นักเรียนเข้าใจสมการเชิงเส้น” เมื่อถึงเวลาออกข้อสอบ ครูพบว่าไม่อาจตัดสินได้ว่าระดับใดจึงเรียกว่า “เข้าใจ” การปรับเป็นจุดประสงค์เชิงพฤติกรรมว่า “เมื่อกำหนดโจทย์สมการเชิงเส้นตัวแปรเดียว 10 ข้อ นักเรียนแสดงวิธีหาค่าตัวแปรได้ถูกต้องอย่างน้อย 8 ข้อ” ทำให้ทั้งการสอน การออกข้อสอบ และการตัดสินผลมีเกณฑ์เดียวกัน นักเรียนก็ทราบล่วงหน้าว่าต้องฝึกให้ทำโจทย์ได้ในระดับใด

การประยุกต์และข้อควรระวัง. ก่อนออกแบบกิจกรรมและข้อสอบทุกครั้ง ครูควรแปลงจุดมุ่งหมายทั่วไปของหลักสูตรให้เป็นจุดประสงค์เชิงพฤติกรรมที่ขึ้นต้นด้วยคำกริยาที่วัดได้ อย่างไรก็ตาม การเลือกคำกริยาต้องอาศัยวิจารณ์ญาณ มิใช่เพียงหยิบจาก “คลังคำกริยา” อย่างกลไก ดังข้อค้นพบในกล่องที่ 3.1

### กล่องที่ 3.1 ข้อควรระวัง: คำกริยาเชิงพฤติกรรมไม่ใช่กลไกอัตโนมัติ

Stanny (2016) วิเคราะห์คลังคำกริยา 30 ชุดที่เผยแพร่บนเว็บไซต์ พบว่า ไม่มีคำกริยาใดเลยที่ทุกคลังจัดให้อยู่ระดับเดียวกัน และคำว่า **choose, relate, select** ปรากฏครบทั้ง 6 ระดับ ขึ้นกับว่าใช้คลังใด

Newton, Da Silva และ Peters (2020) ยืนยันปัญหาเดียวกัน และเสนอว่า ระดับพฤติกรรมขึ้นกับ “บริบทของงาน” มากกว่าตัวคำกริยา เช่น “อธิบาย” อาจเป็นระดับเข้าใจหรือวิเคราะห์ก็ได้

**บทเรียนสำหรับครู:** ให้พิจารณาคำกริยากับเงื่อนไขและสิ่งที่ผู้เรียนต้องทำจริง อย่าตัดสินระดับจากคำกริยาลอย ๆ

ที่มา: สังเคราะห์จาก Stanny (2016) และ Newton, Da Silva, & Peters (2020)

### 3.5.2 ความสัมพันธ์เชิงระบบและความสอดคล้องเชิงโครงสร้าง

**นิยาม.** ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) คือหลักการออกแบบการเรียนรู้ที่กำหนดให้ ผลลัพธ์การเรียนรู้ที่คาดหวัง กิจกรรมการเรียนการสอน และการวัดประเมินผล สอดคล้องไปในทิศทางเดียวกัน (Biggs & Tang, 2011) คำว่า “Constructive” มาจากฐานคิดสร้างนิยม (Constructivism) ที่ถือว่าผู้เรียนเป็นผู้สร้างความรู้ด้วยตนเองผ่านกิจกรรม ส่วน “Alignment” คือการจัดองค์ประกอบทั้งหมดให้เล็งไปยังจุดประสงค์เดียวกัน คุณภาพการศึกษาจะเกิดขึ้นเมื่อ “ลูกโซ่ความสอดคล้อง” ไม่ขาดตอน ดังภาพที่ 3.1

ภาพที่ 3.1 ลูกโซ่ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment Chain)

|                             |                                              |
|-----------------------------|----------------------------------------------|
| หลักสูตร (Curriculum)       | กรอบเป้าหมายและมาตรฐานระดับชาติ/สถาบัน       |
| ผลลัพธ์ระดับหลักสูตร (PLOs) | สิ่งที่ผู้สำเร็จการศึกษาทำได้เมื่อจบหลักสูตร |
| ผลลัพธ์ระดับรายวิชา (CLOs)  | สิ่งที่ผู้เรียนทำได้เมื่อจบรายวิชา           |
| จุดประสงค์ระดับคาบสอน       | พฤติกรรมที่คาดหวังในแต่ละบท/คาบ              |
| กิจกรรมการเรียนรู้          | ประสบการณ์ที่ “สร้าง” พฤติกรรมตามจุดประสงค์  |
| การวัดและประเมินผล          | หลักฐานยืนยันว่าผู้เรียนบรรลุจุดประสงค์      |

ที่มา: ดัดแปลงจาก Biggs & Tang (2011)

**คำอธิบายเชิงลึก.** ข้อต่อแต่ละข้อในลูกโซ่ต้องส่งทอดพฤติกรรมระดับเดียวกัน หากจุดประสงค์ตั้งไว้ระดับ “วิเคราะห์” แต่กิจกรรมเป็นเพียงการบรรยายให้จด และข้อสอบวัดเพียงระดับ “ความจำ” ย่อมเกิด

ช่องว่าง (Gap) ในลูกโซ่ เรียกว่า ภาวะนี้ว่า ความไม่สอดคล้อง (Misalignment) ซึ่งทำให้ข้อมูลจากการวัดไม่สะท้อนสมรรถนะจริง เปรียบได้กับหลักการ “ขยะเข้า-ขยะออก” (Garbage in, Garbage out) คือเมื่อต้นทาง (จุดประสงค์-กิจกรรม) บกพร่อง ผลลัพธ์ปลายทาง (คะแนน) ย่อมเชื่อถือไม่ได้ และนำไปพัฒนาผู้เรียนต่อไม่ได้ (พิชิต ฤทธิ์จัญญ, 2564)

**ตัวอย่างในชั้นเรียนไทย.** รายวิชาวิทยาศาสตร์ที่ตั้ง CLO ว่า “ออกแบบการทดลองเพื่อตรวจสอบสมมติฐานได้” แต่จัดการสอนด้วยการบรรยายและให้ท่องขั้นตอน แล้วสอบด้วยข้อสอบปรนัยตามนิยาม ย่อมไม่สอดคล้องกับ CLO การแก้คือเพิ่มกิจกรรมให้นักเรียนลงมือออกแบบและทำการทดลองจริง แล้วประเมินด้วย rubric ที่ดูกระบวนการออกแบบ ทำให้กิจกรรมและการวัด “เล็ง” ไปที่ CLO เดียวกัน

**การประยุกต์และข้อควรระวัง.** ครูควรจัดทำ “ตารางจับคู่” ระหว่างจุดประสงค์-กิจกรรม-เครื่องมือวัด เพื่อตรวจสอบความสอดคล้องก่อนสอนจริง และใช้เป็นหลักฐานในการประกันคุณภาพรายวิชา อย่างไรก็ตาม การนำหลักการนี้ไปใช้จริงในบริบทไทยมีความท้าทาย ดังกล่องที่ 3.2

**กล่องที่ 3.2 บทเรียนจากงานวิจัยไทย:** ความท้าทายของการนำความสอดคล้องเชิงโครงสร้างไปใช้ Pathamathamakul (2022) ศึกษาอาจารย์มหาวิทยาลัยไทย 9 คน ที่นำการศึกษาที่มุ่งผลลัพธ์ (OBE) และความสอดคล้องเชิงโครงสร้างไปใช้ในชั้นเรียน พบว่าอาจารย์เผชิญ “ภาวะขัดแย้งทางความคิด” (Cognitive Dissonance) หลายประการ

- ผู้เรียนมีเจตคติแบบ “เรียนเพื่อสอบ” (study-to-the-test) สวนทางกับเป้าหมายสมรรถนะ
- ความยุ่งยากเชิงเทคนิคในการให้คะแนนแบบอิงเกณฑ์และการใช้ rubric
- แรงกดดันจากระบบเกรดที่ยังอิงเวลาเรียนแบบเดิม

**บทเรียนสำหรับครู:** ความสอดคล้องเชิงโครงสร้างไม่ใช่เพียงเทคนิคบนกระดาษ แต่ต้องอาศัยการปรับเจตคติของผู้เรียนและการสนับสนุนเชิงระบบควบคู่กัน

ที่มา: สรุปจาก Pathamathamakul (2022)

### 3.5.3 การจำแนกพฤติกรรมการเรียนรู้ 3 ด้าน (Domains of Learning)

นิยาม. เบนจามิน บลูม และคณะ (Bloom et al., 1956) เสนอว่า พฤติกรรมการเรียนรู้ของมนุษย์ครอบคลุม 3 ด้าน เพื่อพัฒนาผู้เรียนอย่างรอบด้าน (Holistic Development) ได้แก่ ด้านพุทธิพิสัย (Cognitive Domain) ด้านจิตพิสัย (Affective Domain) และด้านทักษะพิสัย (Psychomotor Domain) ซึ่งมักเรียกอย่างเข้าใจง่ายว่า “สมอง-จิตใจ-มือ” (Head-Heart-Hand) ดังเปรียบเทียบในตารางที่ 3.3 (พัชรินทร์ ชมภูวิเศษ, 2559)

**ตารางที่ 3.3 เปรียบเทียบพฤติกรรมการเรียนรู้ 3 ด้านและเครื่องมือวัดที่เหมาะสม**

| ด้าน (Domain)               | จุดเน้น                                | สื่อความหมาย   | เครื่องมือวัดที่เหมาะสม                     |
|-----------------------------|----------------------------------------|----------------|---------------------------------------------|
| พุทธิพิสัย<br>(Cognitive)   | ความรู้ สติปัญญา<br>การคิด การแก้ปัญหา | สมอง (Head)    | แบบทดสอบปรนัย/อัตนัย                        |
| จิตพิสัย<br>(Affective)     | เจตคติ ค่านิยม<br>ความรู้สึก คุณลักษณะ | จิตใจ (Heart)  | แบบสังเกต<br>มาตราประมาณค่า<br>แบบวัดเจตคติ |
| ทักษะพิสัย<br>(Psychomotor) | ทักษะการปฏิบัติ<br>การประสานกล้ามเนื้อ | ร่างกาย (Hand) | แบบทดสอบภาคปฏิบัติ<br>Checklist RUBRIC      |

ที่มา: ดัดแปลงจาก Bloom et al. (1956) และ พิชรินทร์ ชมภูวิเศษ (2559, น. 33–34)

**คำอธิบายเชิงลึก.** ทั้งสามด้านมิได้แยกขาดจากกัน แต่ทำงานประสานกันในพฤติกรรมจริง บลูมและคณะพัฒนาด้านพุทธิพิสัยไว้สมบูรณ์ที่สุด ส่วนด้านจิตพิสัยได้รับการขยายโดย Krathwohl, Bloom และ Masia (1964) เป็น 5 ระดับ จากการรับรู้ไปสู่การสร้างเป็นคุณลักษณะนิสัยถาวร ดังภาพที่ 3.2 สำหรับด้านทักษะพิสัย นักวิชาการรุ่นต่อมา เช่น Simpson และ Harrow ได้จัดลำดับจากการเลียนแบบไปสู่การปฏิบัติที่เป็นธรรมชาติอัตโนมัติ (สุทธิพงศ์ บุญผดุง, 2568)

ภาพที่ 3.2 ลำดับขั้นพฤติกรรมด้านจิตพิสัย 5 ระดับ (จากล่างขึ้นบน)

|                                            |                                                                     |
|--------------------------------------------|---------------------------------------------------------------------|
| 1. การรับรู้<br>(Receiving)                | เต็มใจรับฟัง/สนใจสิ่งเร้า เช่น<br>ตั้งใจฟังเรื่องการอนุรักษ์        |
| 2. การตอบสนอง<br>(Responding)              | มีส่วนร่วมและตอบสนอง เช่น<br>เข้าร่วมกิจกรรมเก็บขยะ                 |
| 3. การเห็นคุณค่า<br>(Valuing)              | ยอมรับว่าสิ่งนั้นมีคุณค่า เช่น<br>เชื่อว่าการแยกขยะสำคัญ            |
| 4. การจัดระบบ<br>(Organizing)              | จัดลำดับค่านิยมให้สอดคล้องกัน เช่น<br>วางแผนใช้ชีวิตแบบรักษาสภาพโลก |
| 5. การสร้างลักษณะนิสัย<br>(Characterizing) | ค่านิยมกลายเป็นนิสัยถาวร เช่น<br>เป็นผู้นำด้านสิ่งแวดล้อม           |

ที่มา: ดัดแปลงจาก Krathwohl, Bloom, & Masia (1964)

**ตัวอย่างในชั้นเรียนไทย.** หัวข้อ “การทดลองวิทยาศาสตร์” สามารถตั้งจุดประสงค์ครบ 3 ด้านได้ คือ อธิบายหลักการของปฏิกิริยา (พุทธิพิสัย) ปฏิบัติการทดลองตามขั้นตอนอย่างปลอดภัย (ทักษะพิสัย) และแสดงความรับผิดชอบต่ออุปกรณ์และเพื่อนร่วมกลุ่ม (จิตพิสัย) การออกแบบเช่นนี้ทำให้ผู้เรียนได้รับการพัฒนาอย่างสมดุล ไม่เน้นเพียงความรู้ในสมองอย่างเดียว

**การประยุกต์และข้อควรระวัง.** เมื่อกำหนดจุดประสงค์ ครูควรระบุให้ชัดว่าพฤติกรรมเป้าหมายอยู่ด้านใด เพราะแต่ละด้านใช้เครื่องมือวัดต่างกัน ข้อผิดพลาดที่พบบ่อยคือ ตั้งจุดประสงค์ด้านจิตพิสัย เช่น “เห็นคุณค่า” แต่กลับวัดด้วยข้อสอบปรนัยซึ่งวัดได้เพียงพุทธิพิสัย ทำให้ผลการวัดไม่ตรงกับสิ่งที่ตั้งใจวัด การจับคู่ด้านพฤติกรรมกับเครื่องมือให้ถูกต้องจึงเป็นเงื่อนไขสำคัญของความเที่ยงตรง

### 3.5.4 Bloom's Taxonomy ฉบับเดิมและฉบับปรับปรุง

**นิยาม.** อนุกรมวิธานด้านพุทธิพิสัยของบลูม (Bloom et al., 1956) จัดลำดับพฤติกรรมทางปัญญาจากง่ายไปยาก 6 ชั้น โดยใช้ “คำนาม” เป็นชื่อชั้น ต่อมา Anderson และ Krathwohl (2001) ซึ่งเป็นลูกศิษย์และผู้ร่วมงานของบลูม ได้ปรับปรุงใน 3 ประเด็นสำคัญ คือ (1) เปลี่ยน “คำนาม” เป็น “คำกริยา” เพื่อสะท้อนว่าการคิดเป็นกระบวนการของการกระทำ (2) สลับลำดับสองชั้นสูงสุด ทำให้ “สร้างสรรค์” กลายเป็นชั้นสูงสุด และ (3) เปลี่ยนจากโครงสร้างมิติเดียวเป็น “สองมิติ” คือ มิติด้านความรู้ (Knowledge Dimension) กับมิติด้านกระบวนการทางปัญญา (Cognitive Process Dimension) (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ดังเปรียบเทียบในตารางที่ 3.4

ตารางที่ 3.4 เปรียบเทียบระดับพุทธิพิสัยของบลูมฉบับเดิมและฉบับปรับปรุง

| ชั้น | บลูม (1956)<br>คำนาม       | ฉบับปรับปรุง (2001)<br>คำกริยา | ตัวอย่างคำกริยาเชิงพฤติกรรม     |
|------|----------------------------|--------------------------------|---------------------------------|
| 1    | ความรู้-ความจำ (Knowledge) | จำ (Remember)                  | ระบุ บอกชื่อ เลือก นิยาม        |
| 2    | ความเข้าใจ (Comprehension) | เข้าใจ (Understand)            | สรุปความ ยกตัวอย่าง<br>แปลความ  |
| 3    | การนำไปใช้ (Application)   | ประยุกต์ใช้ (Apply)            | คำนวณ แก้ปัญหา สาธิต            |
| 4    | การวิเคราะห์ (Analysis)    | วิเคราะห์ (Analyze)            | แยกแยะ เปรียบเทียบ<br>เชื่อมโยง |
| 5    | การสังเคราะห์ (Synthesis)  | ประเมินค่า (Evaluate)          | วิจารณ์ ตัดสิน ให้เหตุผล        |
| 6    | การประเมินค่า (Evaluation) | สร้างสรรค์ (Create)            | ออกแบบ วางแผน ผลิต<br>สร้างใหม่ |

ที่มา: ดัดแปลงจาก Anderson & Krathwohl (2001) และ พัชรินทร์ ชมภูวิเศษ (2559, น. 56)

**คำอธิบายเชิงลึก:** มิใช่แค่เปลี่ยนคำ. ชั้นที่ 5 และ 6 มิได้เพียงเปลี่ยนคำ แต่สลับตำแหน่งกันด้วย ฉบับเดิมจบที่ “การประเมินค่า” แต่ฉบับปรับปรุงถือว่า “การสร้างสรรค์ (Create)” คือการประกอบส่วนต่าง ๆ ขึ้นเป็นสิ่งใหม่ จึงเป็นชั้นที่สูงที่สุด (Anderson & Krathwohl, 2001) นอกจากนี้ พิศิษฐ ตัณฑวนิช (2558) อธิบายว่าฉบับปรับปรุงได้แยกการเรียนรู้ออกเป็น 2 ส่วนชัดเจน คือ ส่วนที่หนึ่งเป็น “กระบวนการคิด” 6 ชั้น และส่วน

ที่สองเป็น “ชนิดของความรู้” 4 ประเภท ทำให้ครูระบุได้ว่าต้องการให้ผู้เรียน “ทำกระบวนการคิดใด” กับ “ความรู้ชนิดใด” ดังภาพที่ 3.3 และภาพที่ 3.4

ภาพที่ 3.3 บันไดกระบวนการคิด 6 ขั้นของพุทธิพิสัยฉบับปรับปรุง (จากต่ำขึ้นสูง)

|                          |                                                 |
|--------------------------|-------------------------------------------------|
| 1. จำ (Remember)         | ดึงความรู้จากความจำ เช่น บอกลีขาน               |
| 2. เข้าใจ (Understand)   | สร้างความหมาย เช่น สรุป/อธิบาย/ยกตัวอย่าง       |
| 3. ประยุกต์ใช้ (Apply)   | ใช้ความรู้ในสถานการณ์ใหม่ เช่น คำนวณ/แก้ปัญหา   |
| 4. วิเคราะห์ (Analyze)   | แยกส่วนและหาความสัมพันธ์ เช่น เปรียบเทียบ/จำแนก |
| 5. ประเมินค่า (Evaluate) | ตัดสินด้วยเกณฑ์ เช่น วิเคราะห์/ให้เหตุผล        |
| 6. สร้างสรรค์ (Create)   | ประกอบเป็นสิ่งใหม่ เช่น ออกแบบ/วางแผน/ผลิต      |

ที่มา: ดัดแปลงจาก Anderson & Krathwohl (2001)

โครงสร้างสองมิติและ 4 ชนิดของความรู้. มิติด้านความรู้แบ่งเป็น 4 ชนิด คือ ความรู้ที่เป็นข้อเท็จจริง (Factual) ความรู้ที่เป็นมโนทัศน์ (Conceptual) ความรู้ที่เป็นกระบวนการขั้นตอน (Procedural) และความรู้ อภิปัญญาหรือการรู้คิดของตน (Metacognitive) (พิศิษฐ์ ตันทวนิช, 2558) เมื่อไขว้สองมิติเข้าด้วยกันจะได้ “ตารางสองมิติ” ที่ช่วยให้ครูวางจุดประสงค์ได้แม่นยำ ดังภาพที่ 3.4

ภาพที่ 3.4 ตารางสองมิติของบลูมฉบับปรับปรุง (มิติความรู้ × มิติกระบวนการคิด)

| ชนิดความรู้ \ กระบวนการคิด | จำ | เข้าใจ                      | ประยุกต์                 | วิเคราะห์ | ประเมินค่า                  | สร้างสรรค์ |
|----------------------------|----|-----------------------------|--------------------------|-----------|-----------------------------|------------|
| ข้อเท็จจริง (Factual)      | ✓  |                             |                          |           |                             |            |
| มโนทัศน์ (Conceptual)      |    | เช่น อธิบาย<br>ความสัมพันธ์ |                          |           |                             |            |
| กระบวนการ (Procedural)     |    |                             | เช่น คำนวณ<br>ตามขั้นตอน |           |                             |            |
| อภิปัญญา (Metacognitive)   |    |                             |                          |           | เช่น ประเมิน<br>กลวิธีของตน |            |

ที่มา: ดัดแปลงจาก Anderson & Krathwohl (2001) และ พิศิษฐ์ ตันทวนิช (2558, น. 13–25)

ตัวอย่างและการประยุกต์. จุดประสงค์ “นักเรียนคำนวณพื้นที่สามเหลี่ยมจากสูตรได้ถูกต้อง” อยู่ที่ช่อง “ความรู้กระบวนการ × ประยุกต์ใช้” ขณะที่ “นักเรียนประเมินว่ากลวิธีแก้โจทย์ของตนมีจุดบกพร่องใด” อยู่ที่ช่อง “ความรู้อภิปัญญา × ประเมินค่า” การระบุตำแหน่งในตารางช่วยให้ครูเลือกกิจกรรมและข้อสอบให้ตรง

ระดับ ป้องกันไม่ให้ข้อสอบกระจุกอยู่เพียงช่อง “ข้อเท็จจริง x จำ” ซึ่งเป็นข้อผิดพลาดที่พบบ่อยที่สุดในการออกข้อสอบ

### 3.5.5 หลักการเขียนจุดประสงค์เชิงพฤติกรรม: เมเกอร์และโมเดล ABCD

นิยาม. เมเกอร์ (Mager, 1962) ในหนังสือ *Preparing Instructional Objectives* เสนอว่า จุดประสงค์ที่มีประสิทธิภาพต้องมีองค์ประกอบสำคัญ 3 ส่วน ได้แก่ พฤติกรรม (Behavior/Performance) เงื่อนไข (Condition) และเกณฑ์ (Criterion/Degree) แนวคิดนี้สอดคล้องกับตำราไทย โดย พัทรินทร์ ชมภูวิเศษ (2559, น. 42–43) ซึ่งอ้างถึง ล้วน สายยศ และอังคณา สายยศ (2543) ระบุว่าจุดประสงค์เชิงพฤติกรรมที่สมบูรณ์ประกอบด้วย พฤติกรรมที่คาดหวัง เงื่อนไขหรือสถานการณ์ (อาจมีหรือไม่มีก็ได้) และเกณฑ์ ต่อมานักออกแบบการสอนได้ขยายเป็น “โมเดล ABCD” โดยเพิ่มองค์ประกอบ ผู้เรียน (Audience) เพื่อระบุกลุ่มเป้าหมายให้ชัดเจน ดังกล่องที่ 3.3

#### กล่องที่ 3.3 องค์ประกอบของจุดประสงค์เชิงพฤติกรรมตามโมเดล ABCD

**A – Audience (ผู้เรียน):**

ใครคือผู้แสดงพฤติกรรม เช่น “นักศึกษาชั้นปีที่ 1”

**B – Behavior (พฤติกรรม):**

พฤติกรรมที่สังเกตได้ ต้องใช้คำกริยาเชิงพฤติกรรม เช่น “คำนวณ” “อธิบาย”

**C – Condition (เงื่อนไข):**

สถานการณ์/สื่อที่กำหนดให้ เช่น “เมื่อกำหนดตารางข้อมูลให้”

**D – Degree (เกณฑ์):**

มาตรฐานความสำเร็จ เช่น “ได้ถูกต้องอย่างน้อยร้อยละ 80”

หมายเหตุเชิงวิชาการ:

โมเดล 3 ส่วน (B–C–D) เป็นของเมเกอร์ (1962) ส่วน Audience เป็นส่วนเพิ่มภายหลังในกรอบ ABCD การอ้างอิงจึงควรระบุที่มาให้ตรง

ที่มา: ดัดแปลงจาก Mager (1962) และ พัทรินทร์ ชมภูวิเศษ (2559, น. 42–43)

**คำอธิบายเชิงลึก:** หลักการเขียน. พัทรินทร์ ชมภูวิเศษ (2559, น. 44) สรุปหลักการเขียนจุดประสงค์เชิงพฤติกรรมไว้ 5 ข้อ คือ (1) ใช้ประโยคสั้น ๆ ใดใจความสมบูรณ์ (2) แต่ละข้อระบุพฤติกรรมที่คาดหวังเพียง 1 พฤติกรรม (3) เป็นพฤติกรรมของผู้เรียนที่แสดงออกได้ (4) เป็นรูปธรรมที่ครูวัดและสังเกตได้ และ (5) ใช้คำที่

ชัดเจน เฉพาะเจาะจง ไม่ต้องตีความ หลักการเหล่านี้ช่วยให้จุดประสงค์ทำหน้าที่เป็นเกณฑ์ตรวจสอบความสอดคล้องได้จริง ส่วนการวิเคราะห์ตัวอย่างการเขียนที่ถูกและที่ควรปรับปรุงแสดงในตารางที่ 3.5

ตารางที่ 3.5 วิเคราะห์ตัวอย่างการเขียนจุดประสงค์ตามโมเดล ABCD

| ตัวอย่างที่ควรปรับปรุง          | ตัวอย่างที่ถูกต้องตาม ABCD                                                                        | เหตุผลเชิงวิเคราะห์                                                     |
|---------------------------------|---------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| เพื่อให้นักเรียนเข้าใจเรื่องแรง | เมื่อจบคาบเรียน (C) นักเรียน (A) อธิบายกฎการเคลื่อนที่ของนิวตัน (B) ได้ถูกต้องอย่างน้อย 2 ข้อ (D) | “เข้าใจ” เป็นพฤติกรรมภายในที่วัดตรงไม่ได้ ต้องใช้ “อธิบาย” ที่สังเกตได้ |
| ครูสอนให้เด็กแก้สมการได้        | หลังเรียนจบ (C) นักเรียน (A) แก้โจทย์ปัญหาสมการ (B) ได้ถูกต้อง 8 ใน 10 ข้อ (D)                    | ต้องเน้น “ผลการเรียนรู้ของผู้เรียน” ไม่ใช่ “กิจกรรมของครู”              |

ที่มา: ผู้เขียน ปรับจากแนวทางใน พัทรินทร์ ชมภูวิเศษ (2559) และ Mager (1962)

**ตัวอย่างในชั้นเรียนไทยและการประยุกต์.** ครูภาษาไทยอาจเขียนว่า “เมื่อกำหนดหัวข้อให้ (C) นักเรียน (A) แต่งกลอนสุภาพ (B) ได้ถูกต้องตามฉันทลักษณ์อย่างน้อย 1 บท (D)” คลังคำกริยา (Action Verbs) ช่วยให้ครูเลือกคำให้ตรงระดับพฤติกรรม เช่น ระดับ “จำ” ใช้ ระบุด/บอกชื่อ ระดับ “วิเคราะห์” ใช้ จำแนก/เปรียบเทียบ ระดับ “สร้างสรรค์” ใช้ ออกแบบ/วางแผน อย่างไรก็ตาม Newton, Da Silva และ Peters (2020) เตือนว่าไม่ควรยึดคลังคำกริยาอย่างตายตัว เพราะคำเดียวกันอาจอยู่คนละระดับเมื่อบริบทต่างกัน คำกริยาที่คลุมเครืออย่าง “รู้” “ซาบซึ้ง” ควรหลีกเลี่ยง (สำเร็จ บุญเรืองรัตน์, 2534)

### 3.5.6 การจับคู่ระดับพฤติกรรมกับเครื่องมือวัดและตารางวิเคราะห์ข้อสอบ (TOS)

นิยาม. การเลือกเครื่องมือวัดต้องสอดคล้องกับ “สภาพจริง” (Authenticity) ของพฤติกรรมที่ต้องการวัด การจับคู่ที่ถูกต้องเป็นหัวใจของความเที่ยงตรง (พิชิต ฤทธิ์จรรยา, 2564) หลักง่าย ๆ คือ พฤติกรรมระดับต่ำ (จำ-เข้าใจ) วัดด้วยแบบทดสอบ ส่วนพฤติกรรมระดับสูง (ประยุกต์ขึ้นไป) และด้านจิตพิสัย/ทักษะพิสัย ต้องอาศัยการปฏิบัติ ชิ้นงาน หรือการสังเกต ดังตารางที่ 3.6

ตารางที่ 3.6 การจับคู่ระดับพฤติกรรมกับวิธีวัดและเครื่องมือ

| ระดับพฤติกรรม        | วิธีวัด               | เครื่องมือที่เหมาะสม                                 |
|----------------------|-----------------------|------------------------------------------------------|
| จำ / เข้าใจ          | การทดสอบเขียนตอบ      | ข้อสอบเลือกตอบ ถูก-ผิด จับคู่ เต็มคำ                 |
| ประยุกต์ / วิเคราะห์ | การแก้สถานการณ์/ใบงาน | ข้อสอบอัตนัยเชิงสถานการณ์ ใบงานวิเคราะห์ ระบุริคำแนก |

|                         |                         |                                                     |
|-------------------------|-------------------------|-----------------------------------------------------|
| ประเมินค่า / สร้างสรรค์ | การประเมินผลงาน/โครงการ | ชิ้นงาน แฟ้มสะสมงาน<br>(Portfolio) + Scoring Rubric |
| จิตพิสัย (ทุกระดับ)     | การสังเกตพฤติกรรม       | แบบสังเกต แบบวัดเจตคติ<br>มาตรฐานประเมินค่า         |
| ทักษะพิสัย (ทุกระดับ)   | การปฏิบัติงานจริง       | แบบประเมินทักษะ Checklist<br>รูบริก                 |

ที่มา: ดัดแปลงจาก พิชิต ฤทธิ์จรูญ (2564)

**ความหลากหลายของเครื่องมือวัด.** นอกจากการจับคู่ตามระดับพฤติกรรมแล้ว ผู้สอนควรรู้จักประเภทเครื่องมือวัดที่หลากหลาย เพื่อเลือกใช้ให้ครอบคลุมพฤติกรรมทั้งสามด้าน สำเร็จ บุญเรืองรัตน์ (2534) จำแนกเครื่องมือวัดและประเมินผลทางการศึกษาไว้ 12 ชนิด ดังต่อไปนี้ 3.4

#### กล่องที่ 3.4 เครื่องมือวัดและประเมินผลทางการศึกษา 12 ชนิด

- 1) แบบทดสอบ (Test) — วัดความรู้ด้านพุทธิพิสัย
- 2) มาตรฐานส่วนประมาณค่า (Rating Scale) — วัดคุณภาพงาน เจตคติ การปฏิบัติ
- 3) แบบสอบถาม (Questionnaire) — สำรวจความคิดเห็นและความรู้สึก
- 4) แบบตรวจสอบรายการ (Checklist) — ตรวจการมี/ไม่มี หรือขั้นตอน
- 5) แบบสำรวจ (Inventory) — สำรวจความสนใจและบุคลิภาพ
- 6) แบบสังเกต (Observation) — บันทึกพฤติกรรมตามสภาพจริง
- 7) แบบสัมภาษณ์ (Interview) — เจาะลึกความเชื่อ เหตุผล ความรู้สึก
- 8) แบบบันทึกเหตุการณ์ (Anecdotal Record) — บันทึกพฤติกรรมเด่น/ด้อยต่อเนื่อง
- 9) สังคมมิติ (Sociometry) — วัดโครงสร้างความสัมพันธ์ในกลุ่ม
- 10) แบบทดสอบสถานการณ์ (Situational Test) — จำลองสถานการณ์ดูการแก้ปัญหา
- 11) แบบทดสอบฉายภาพ (Projective Test) — วัดบุคลิภาพเชิงลึก
- 12) การศึกษาเป็นรายกรณี (Case Study) — วิเคราะห์ปัญหาเชิงลึกเฉพาะบุคคล

ที่มา: สรุปจาก สำเร็จ บุญเรืองรัตน์ (2534)

**ตารางวิเคราะห์ข้อสอบ (TOS).** ตารางวิเคราะห์ข้อสอบ (Table of Specifications: TOS) คือเครื่องมือประกันคุณภาพที่กระจายน้ำหนักข้อสอบตามสองแกน คือ แกนเนื้อหา (หัวข้อที่สอน) และแกนพฤติกรรม (ระดับของบลูม) ช่วยป้องกันไม่ให้ข้อสอบกระจุกอยู่ที่ระดับ “ความจำ” มากเกินไป และทำให้

คะแนนสะท้อนความสำคัญของเนื้อหาอย่างเที่ยงตรง (ชวาล แพรัตกุล, 2518) การใช้โครงสร้างสองมิติของบลูมฉบับปรับปรุงมาเป็นแกนพฤติกรรมจะช่วยให้ผังข้อสอบครอบคลุมระดับการคิดที่หลากหลายยิ่งขึ้น (Anderson & Krathwohl, 2001) ดังตัวอย่างในตารางที่ 3.7

ตารางที่ 3.7 ตัวอย่างตารางวิเคราะห์ข้อสอบอย่างย่อ (หน่วยเป็นจำนวนข้อ)

| เนื้อหา \ พฤติกรรม       | จำ | เข้าใจ | ประยุกต์ | วิเคราะห์ | รวม |
|--------------------------|----|--------|----------|-----------|-----|
| ความหมายของจุดประสงค์    | 2  | 2      | 1        | 1         | 6   |
| การจำแนกพฤติกรรม (Bloom) | 2  | 3      | 2        | 3         | 10  |
| การเขียน ABCD            | 1  | 1      | 3        | 4         | 9   |
| รวม                      | 5  | 6      | 6        | 8         | 25  |

ที่มา: ผู้เขียน

### 3.5.7 กรณีศึกษาการวิเคราะห์และปรับปรุงจุดประสงค์

กรณีศึกษาที่ 1 (ระดับโรงเรียน). อาจารย์ท่านหนึ่งตั้งจุดประสงค์ว่า “นักเรียนเห็นคุณค่าของศิลปะไทย” ซึ่งเป็นจิตพิสัยระดับสูงที่วัดโดยตรงในคาบเดียวได้ยาก จึงเกิดปัญหาในการเลือกเครื่องมือวัด แนวทางแก้ไขแสดงในกล่องที่ 3.5

#### กล่องที่ 3.5 แนวทางปรับปรุงจุดประสงค์ที่วัดยาก (Improvement Plan)

ปัญหา: “เห็นคุณค่า” เป็นนามธรรม วัดตรงไม่ได้ และตีความต่างกัน

แนวทาง: แยกเป็นพฤติกรรมบ่งชี้ที่สังเกตได้ เช่น “นักเรียน (A) เขียนเรียงความเชิงสร้างสรรค์ (B) เพื่อเชิญชวนให้ผู้อื่นอนุรักษ์ศิลปะไทย (C) ได้อย่างมีเหตุผลและน่าสนใจ (D)”

ผล: ได้พฤติกรรมที่วัดได้ และเลือกเครื่องมือ (รูปรีเรียงความ) ได้สอดคล้องกับจุดประสงค์

ที่มา: ผู้เขียน

กรณีศึกษาที่ 2 (บริบทอุดมศึกษา). อาจารย์มักตั้งจุดประสงค์ว่า “นักศึกษามีความรับผิดชอบ” ซึ่งเป็นคุณลักษณะด้านจิตพิสัยที่วัดโดยตรงได้ยาก แนวทางแก้ไขคือแตกออกเป็นพฤติกรรมบ่งชี้ที่สังเกตได้ เช่น “นักศึกษาส่งงานตรงตามเวลาที่กำหนดไม่น้อยกว่าร้อยละ 90 ของงานทั้งหมด” ทำให้เก็บข้อมูลและตัดสินผลได้อย่างเป็นรูปธรรม (สุทธิพงศ์ บุญผดุง, 2568)

**บทเรียนจากกรณีศึกษา.** เมื่อพบจุดประสงค์ที่วัดยาก ครูควรดำเนินการ 3 ขั้นตอน คือ (1) ระบุระดับพฤติกรรมที่แท้จริงที่ต้องการ (2) เลือกคำกริยาเชิงพฤติกรรมที่สังเกตได้ และ (3) กำหนดเงื่อนไขและเกณฑ์ให้

ครบ เพื่อรักษาความสอดคล้องเชิงโครงสร้างไว้ตลอดกระบวนการ ตั้งแต่จุดประสงค์ กิจกรรม ไปจนถึงการวัด และการนำผลไปพัฒนาผู้เรียน (พัชรินทร์ ชมภูวิเศษ, 2559)

### 3.6 กิจกรรมการเรียนการสอนและใบงาน

---

#### ก. ชั้นก่อนเรียน (Pre-class)

- ทำแบบทดสอบก่อนเรียน (หัวข้อ 3.4) เพื่อตรวจสอบความรู้พื้นฐาน
- กิจกรรม “Action Verb Brainstorming”: ผู้เรียนเขียนคำกริยาที่คิดว่า “วัดได้” และ “วัดไม่ได้” แล้วร่วมจัดกลุ่ม

#### ข. ชั้นระหว่างเรียน (In-class Workshop)

1. Conceptual Building:  
บรรยายเชิงปฏิสัมพันธ์และอภิปรายความแตกต่างของจุดมุ่งหมายทั่วไปกับจุดประสงค์เชิงพฤติกรรม
2. Practice & Coach: แบ่งกลุ่มจำแนกระดับพฤติกรรมจากตัวชี้วัดในหลักสูตรแกนกลาง โดยมีครูเป็นผู้อำนวยความสะดวก (Facilitator)
3. Peer Review (Gallery Walk): แลกเปลี่ยนและวิพากษ์ความสอดคล้องระหว่างจุดประสงค์–ระดับพฤติกรรม–เครื่องมือวัดของเพื่อน

#### ค. ชั้นหลังเรียน (Post-class)

- ทำใบงานวิเคราะห์ความสอดคล้อง (ด้านล่าง) และทำแบบทดสอบหลังเรียน (หัวข้อ 3.8)
- สะท้อนคิด (Reflection): สิ่งที่ได้เรียนรู้ – สิ่งที่ต้องปรับปรุง – สิ่งที่น่าไปใช้ได้จริง

#### ใบงานที่ 3.1 วิเคราะห์ความสอดคล้องของจุดประสงค์ (สำหรับ CLO2)

คำชี้แจง: เลือกรายวิชา/หัวข้อ 1 หัวข้อ แล้วดำเนินการดังนี้

- 1) คัดเลือกตัวชี้วัด/จุดประสงค์มา 5 ข้อ แล้วจำแนกว่าแต่ละข้ออยู่ระดับพฤติกรรมใดของบลูม
- 2) ระบุว่าจุดประสงค์แต่ละข้อครอบคลุม ABCD หรือไม่ ขาดส่วนใด
- 3) เสนอเครื่องมือวัดที่สอดคล้องกับระดับพฤติกรรมของแต่ละข้อ พร้อมเหตุผล
- 4) สรุปว่ามีจุดประสงค์ข้อใดเกิดภาวะ Misalignment หรือไม่ และจะปรับปรุงอย่างไร

เกณฑ์การประเมิน: ใช้ Analytical Rubric ในหัวข้อ 3.10

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

### 3.7 สรุปบทเรียน

---

- จุดประสงค์การเรียนรู้คือ “เต็มทิศ” ที่ระบุพฤติกรรมที่สังเกตและวัดได้ของผู้เรียนหลังเรียน
- สามเหลี่ยมการศึกษา (กำหนดวัตถุประสงค์-จัดประสบการณ์-ประเมินผล) ต้องทำงานสัมพันธ์กัน
- ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) คือการให้จุดประสงค์ กิจกรรม และการวัดตรงทิศทางเดียวกัน
- พฤติกรรมการเรียนรู้มี 3 ด้าน: พุทธิพิสัย (สมอง) จิตพิสัย (จิตใจ) ทักษะพิสัย (ร่างกาย)
- บลุ่มฉบับปรับปรุง (2001) เปลี่ยนคำนามเป็นคำกริยา เป็นโครงสร้าง 2 มิติ และให้ “สร้างสรรค์” เป็นขั้นสูงสุด
- โมเดล ABCD (Audience, Behavior, Condition, Degree) คือมาตรฐานการเขียนจุดประสงค์ที่วัดได้
- เมเกอร์ (1962) เสนอองค์ประกอบ 3 ส่วน (พฤติกรรม-เงื่อนไข-เกณฑ์) ซึ่งภายหลังขยายเป็น ABCD
- คำกริยาเชิงพฤติกรรม (Action Verbs) เป็นกุญแจสำคัญที่ทำให้จุดประสงค์วัดได้จริง
- การเลือกเครื่องมือวัดต้องสอดคล้องกับระดับพฤติกรรม และใช้ TOS กระจายน้ำหนักเพื่อความเที่ยงตรง
- ภาวะ Misalignment ทำให้ผลการวัดไม่สะท้อนสมรรถนะจริง และนำไปพัฒนาผู้เรียนต่อไม่ได้

### 3.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการ

เรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบ  
ก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับ  
ผู้เรียนในสัปดาห์ถัดไป

### 3.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

#### ตอนที่ 1 คำถามอัตนัย (5 ข้อ)

1. จงอธิบายความสัมพันธ์เชิงระบบของ “สามเหลี่ยมการศึกษา” มาพอสังเขป

แนวคำตอบ: จุดประสงค์กำหนดทิศทางของกิจกรรมและการประเมิน ทั้งสามต้องสอดคล้องกัน  
หากส่วนใดไม่ตรงกันจะทำให้คุณภาพการศึกษารั่วไหล

2. เหตุใดการเขียนจุดประสงค์จึงต้องใช้คำกริยาเชิงพฤติกรรม (Action Verbs)

แนวคำตอบ: เพราะเปลี่ยนนามธรรมให้เป็นรูปธรรมที่สังเกตและวัดได้  
ลดความคลาดเคลื่อนและอคติในการตีความของผู้ประเมิน

3. จงเปรียบเทียบความแตกต่างระหว่างพฤติกรรมด้านพุทธิพิสัยและจิตพิสัย พร้อมยกตัวอย่างเครื่องมือวัด

แนวคำตอบ: พุทธิพิสัยเน้นสติปัญญา/การคิด วัดด้วยแบบทดสอบ ส่วนจิตพิสัยเน้นความรู้สึก/เจตคติ  
วัดด้วยการสังเกตหรือแบบวัดเจตคติ

4. จงวิเคราะห์การเปลี่ยนแปลงสำคัญจาก Bloom (1956) สู่ฉบับปรับปรุง (2001)

แนวคำตอบ: เปลี่ยนคำนามเป็นคำกริยา สลับลำดับให้ “สร้างสรรค์” เป็นขั้นสูงสุด  
และเพิ่มมิติด้านความรู้เป็นโครงสร้างสองมิติ

5. หากท่านต้องวัด “เจตคติเชิงอนุรักษ์” ของนักศึกษา ท่านจะเลือกเครื่องมือใด เพราะเหตุใด

แนวคำตอบ: เลือกแบบวัดเจตคติหรือการสังเกตพฤติกรรมอย่างต่อเนื่อง  
เพราะเป็นความรู้สึภายในที่ต้องวัดผ่านการแสดงออกที่สม่ำเสมอ

#### ตอนที่ 2 กิจกรรมปฏิบัติ (สำหรับประเมิน CLO2)

ให้นักศึกษาเลือกหัวข้อที่สนใจ 1 หัวข้อ แล้วคัดเลือกจุดประสงค์/ตัวชี้วัดมา 5 ข้อ จากนั้น (1) จำแนกระดับ  
พฤติกรรมตามบลูม (2) ตรวจสอบองค์ประกอบ ABCD และ (3) เสนอเครื่องมือวัดที่สอดคล้อง พร้อมระบุข้อที่เกิด  
Misalignment และแนวทางปรับปรุง ประเมินด้วย Analytical Rubric ในหัวข้อ 3.10

### 3.10 เครื่องมือวัดและประเมินผลประจำบท

บทนี้ใช้เครื่องมือวัดให้สอดคล้องกับระดับพฤติกรรมของจุดประสงค์ คือ แบบทดสอบ (วัด CLO1 ระดับจำ-เข้าใจ) และใบงานวิเคราะห์/จำแนกที่ประเมินด้วยรูบรีคเชิงวิเคราะห์ (วัด CLO2 ระดับวิเคราะห์)

ตารางที่ 3.8 รูบรีคเชิงวิเคราะห์สำหรับประเมินใบงานวิเคราะห์จุดประสงค์ (Analytical Rubric, รวม 16 คะแนน)

| เกณฑ์การประเมิน                              | ดีมาก (4)                              | ดี (3)                 | พอใช้ (2)             | ปรับปรุง (1)                |
|----------------------------------------------|----------------------------------------|------------------------|-----------------------|-----------------------------|
| การจำแนกระดับพฤติกรรมตามบลูม                 | จำแนกถูกต้องแม่นยำทุกข้อ               | ถูกต้องเป็นส่วนในใหญ่  | ผิดในระดับสำคัญบางข้อ | จำแนกไม่ได้/ผิดเป็นส่วนใหญ่ |
| การตรวจสอบองค์ประกอบ ABCD                    | ระบุครบ/ขาดได้ถูกต้องทุกข้อ            | ถูกต้องเป็นส่วนในใหญ่  | ระบุได้บางส่วน        | ระบุไม่ถูกต้อง              |
| การจับคู่เครื่องมือกับพฤติกรรม               | เครื่องมือสอดคล้องทุกข้อ พร้อมเหตุผล   | สอดคล้องเป็นส่วนในใหญ่ | สอดคล้องบางข้อ        | ไม่สอดคล้อง                 |
| การวิเคราะห์ Misalignment และข้อเสนอปรับปรุง | วิเคราะห์คมชัด เสนอแนวทางได้สมเหตุสมผล | วิเคราะห์ได้ดี         | วิเคราะห์ผิวเผิน      | ไม่สามารถวิเคราะห์ได้       |

เกณฑ์การผ่าน: ได้คะแนนรวมตั้งแต่ 10 จาก 16 คะแนน (ร้อยละ 62.5) ขึ้นไป ถือว่าผ่านเกณฑ์ CLO2 ของบทนี้ (ระดับคุณภาพ: 14-16 = ดีมาก, 11-13 = ดี, 10 = ผ่านเกณฑ์ขั้นต่ำ, ต่ำกว่า 10 = ต้องปรับปรุง)

ตารางที่ 3.9 สรุปสัดส่วนคะแนนประจำบท (รวม 100%)

| องค์ประกอบการประเมิน             | CLO  | ระดับพฤติกรรม | เครื่องมือ                | สัดส่วน |
|----------------------------------|------|---------------|---------------------------|---------|
| แบบทดสอบหลังเรียน (ปรนัย 10 ข้อ) | CLO1 | จำ-เข้าใจ     | แบบทดสอบ                  | 40%     |
| ใบงานจำแนกพฤติกรรม               | CLO2 | วิเคราะห์     | ใบงาน + Analytical Rubric | 25%     |
| ใบงานวิเคราะห์ความสอดคล้อง       | CLO2 | วิเคราะห์     | ใบงาน + Analytical Rubric | 35%     |
| รวม                              |      |               |                           | 100%    |

การตรวจสอบความครบถ้วน: CLO1 และ CLO2 ต่างมีกิจกรรม วิธีวัด เครื่องมือ และสัดส่วนคะแนนรองรับครบถ้วน ไม่มีผลลัพธ์การเรียนรู้ที่ไม่ถูกประเมิน และไม่ประเมินผลลัพธ์ที่มีได้ประกาศไว้ในบท

### 3.11 เอกสารอ้างอิงประจำบท

---

#### ภาษาไทย

ชวาล แพร์ตกุล. (2518). การวัดผลการศึกษา. กรุงเทพฯ: วัฒนาพานิช.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการเรียนรู้. อุตรธานี: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุตรธานี.

พิชิต ฤทธิ์เจริญ. (2564). หลักการวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 11). กรุงเทพฯ: แฮสส์ ออฟ เคอร์มิสท์.

พิศิษฐ์ ตัณฑวณิช. (2558). แนวคิดการจำแนกพฤติกรรมการเรียนรู้ตามวัตถุประสงค์การจัดการศึกษาด้านพุทธิพิสัยตามแนวคิดของบลูมและคณะฉบับปรับปรุง. วารสารมหาวิทยาลัยราชภัฏลำปาง, 3(2), 13-25.

สำเร็จ บุญเรืองรัตน์. (2534). การวัดและประเมินผลการศึกษา. กรุงเทพฯ: พัฒนาศึกษา.

สุทธิพงศ์ บุญผดุง. (2568). การวัดและประเมินผลการเรียนรู้. กรุงเทพฯ: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

อนุวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่. สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

#### ภาษาอังกฤษ

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York, NY: Longman.

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). Maidenhead, UK: Open University Press.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives, Handbook I: The cognitive domain. New York, NY: David McKay.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives, Handbook II: The affective domain. New York, NY: David McKay.

Mager, R. F. (1962). Preparing instructional objectives. Palo Alto, CA: Fearon.

Newton, P. M., Da Silva, A., & Peters, L. G. (2020). A pragmatic master list of action verbs for Bloom's taxonomy. *Frontiers in Education*, 5, 107.

<https://doi.org/10.3389/feduc.2020.00107>

Pathamathamakul, C. (2022). Cognitive dissonance and teachers' adoption of outcome-based teaching. *Journal of Education and Innovative Learning*, 2(1), 15–34.

Stanny, C. J. (2016). Reevaluating Bloom's taxonomy: What measurable verbs can and cannot say about student learning. *Education Sciences*, 6(4), 37.

<https://doi.org/10.3390/educsci6040037>

## บทที่ 4

### หลักการสร้างและเขียนข้อสอบวัดผลสัมฤทธิ์

### *Principles of Constructing and Writing Achievement Tests*

รายวิชา ED1006 การวัดและการประเมินผลการเรียนรู้

*Measurement and Evaluation of Learning*

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

เวลาเรียน 4 คาบ (บรรยายและฝึกปฏิบัติ)

ผลลัพธ์การเรียนรู้หลักของบท: CLO2 และ CLO3

### แผนบริหารการสอนประจำบทที่ 4

---

#### 1. หัวข้อสาระจัดการเรียนการสอน

- ความหมาย บทบาท และประเภทของข้อสอบวัดผลสัมฤทธิ์ทางการเรียน
- คุณลักษณะของเครื่องมือวัดผลที่มีคุณภาพ 5 ประการ (ความตรง ความเที่ยง ความเป็นปรนัย ความยุติธรรม และความครอบคลุม)
- ความสัมพันธ์ระหว่างจุดประสงค์การเรียนรู้กับข้อสอบตามหลักความสอดคล้องเชิงโครงสร้าง (Constructive Alignment)
- การสร้างพิมพ์เขียวการสอบ/ตารางวิเคราะห์หลักสูตร (Test Blueprint / Table of Specifications)
- หลักการเขียนข้อสอบแบบเลือกตอบ (Multiple Choice) และข้อสอบอัตนัย (Essay)
- ข้อผิดพลาดที่พบบ่อยและการตรวจสอบคุณภาพข้อสอบเบื้องต้นด้วยดัชนีความสอดคล้อง (IOC)

#### 2. ความคิดรวบยอด

ข้อสอบวัดผลสัมฤทธิ์ทางการเรียนคือ “เครื่องมือทางวิทยาศาสตร์” ที่ออกแบบขึ้นอย่างเป็นระบบเพื่อกระตุ้นให้ผู้เรียนแสดง “พฤติกรรมประติษฐ์” อันสะท้อนความรู้ความสามารถที่แท้จริงตามจุดประสงค์การเรียนรู้ การสร้างข้อสอบที่ดีต้องยึดหลักความสอดคล้องเชิงโครงสร้าง เริ่มจากการวางตารางวิเคราะห์หลักสูตร

(Test Blueprint) เพื่อควบคุมน้ำหนักเนื้อหาและระดับพฤติกรรม แล้วจึงเขียนข้อคำถามตามหลักวิชา โดยรักษาความตรง ความเที่ยง และความเป็นปรนัยของเครื่องมือไว้เป็นสำคัญ

### 3. จุดประสงค์การเรียนรู้

เมื่อศึกษาบทเรียนและผ่านกิจกรรมเชิงปฏิบัติแล้ว นักศึกษาสามารถ

- อธิบายความหมาย บทบาท และประเภทของข้อสอบวัดผลสัมฤทธิ์ได้อย่างถูกต้อง (CLO2)
- วิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์การเรียนรู้ เครื่องมือวัด และระดับพฤติกรรมพุทธิพิสัยได้ (CLO2)
- ออกแบบและสร้างตารางวิเคราะห์หลักสูตร (Test Blueprint) ที่สะท้อนน้ำหนักเนื้อหาและพฤติกรรมได้ (CLO3)
- เขียนข้อสอบแบบเลือกตอบและข้อสอบอัตนัยที่มีคุณภาพตามหลักวิชาการ (CLO3)
- ตรวจสอบคุณภาพข้อสอบเบื้องต้นด้วยดัชนีความสอดคล้อง (IOC) ได้ (CLO2, CLO3)

### 4. เนื้อหาวิชา

ศึกษาเนื้อหาโดยละเอียดตามเอกสารประกอบการสอนบทที่ 4 หัวข้อ 4.5 ความหมายและบทบาทของข้อสอบ (4.5.1) คุณลักษณะของเครื่องมือวัดผลที่ดี (4.5.2) ความสอดคล้องเชิงโครงสร้างกับ Bloom (4.5.3) การสร้างตารางวิเคราะห์หลักสูตร (4.5.4) ประเภทของข้อสอบ (4.5.5) หลักการเขียนข้อสอบเลือกตอบ (4.5.6) หลักการเขียนข้อสอบอัตนัย (4.5.7) และข้อผิดพลาดที่พบบ่อยกับการตรวจสอบคุณภาพด้วย IOC (4.5.8)

### 5. กิจกรรมการเรียนการสอน

- ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน 10 ข้อ และเล่นเกม “จับผิดข้อสอบ (Spot the Error)” จากข้อสอบตัวอย่าง
- ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์และอภิปรายกรณีศึกษาคุณภาพข้อสอบ (CLO2)
- ขั้นระหว่างเรียน: ปฏิบัติการสร้างตารางวิเคราะห์หลักสูตร (TOS Workshop) และคลินิกการเขียนข้อสอบ (Item Writing Clinic) (CLO3)
- ขั้นระหว่างเรียน: ตรวจสอบชุดข้อสอบของเพื่อนด้วยแบบประเมิน IOC และให้ข้อมูลย้อนกลับ (Peer Review)
- ขั้นหลังเรียน: ทำแบบทดสอบหลังเรียนและส่งชิ้นงานชุดข้อสอบพร้อมตารางวิเคราะห์หลักสูตร

### 6. สื่อการเรียนการสอน

- เอกสารประกอบการสอนบทที่ 4 และสไลด์นำเสนอ (PowerPoint)

- ใบงานสร้างตารางวิเคราะห์หลักสูตร และใบงานวิเคราะห์คุณภาพข้อสอบ
- ตัวอย่างชุดข้อสอบ (ดี/บกพร่อง) และแบบฟอร์มประเมินดัชนีความสอดคล้อง (IOC)
- rubric ประเมินตารางวิเคราะห์หลักสูตร และ rubric ประเมินชุดข้อสอบ
- ตำราหลัก: พัชรินทร์ ชมภูวิเศษ (2559); ชวาล แพรัตกุล (2520)

## 7. การวัดและประเมินผล

- แบบทดสอบก่อนเรียน–หลังเรียน (10 ข้อคู่ขนาน): วัดความรู้ความเข้าใจหลักการสร้างข้อสอบ (CLO2)
- ใบงานวิเคราะห์คุณภาพข้อสอบจากกรณีศึกษา + rubric จำแนกข้อบกพร่อง: วัดการวิเคราะห์ (CLO2) สัดส่วน 40%
- ชิ้นงานตารางวิเคราะห์หลักสูตรและชุดข้อสอบ + rubric และเกณฑ์ IOC: วัดการออกแบบและสร้างเครื่องมือ (CLO3) สัดส่วน 60%
- เกณฑ์ผ่าน: แบบทดสอบหลังเรียนไม่น้อยกว่าร้อยละ 80 และคะแนน rubric เฉลี่ยตั้งแต่ระดับ “ดี” (3.00) ขึ้นไป

### 4.1 ความคิดรวบยอด

ข้อสอบในทางการวัดและประเมินผลการศึกษามีได้เป็นเพียงชุดคำถามที่ครูจัดทำขึ้นตามความพอใจ หากแต่เป็น “เครื่องมือทางวิทยาศาสตร์” ที่ออกแบบมาเพื่อกระตุ้นให้ผู้เรียนแสดง “พฤติกรรมประดิษฐ์” (Constructed Behavior) ที่สะท้อนความสามารถที่แท้จริงออกมาให้สังเกตและวัดปริมาณได้ (ชวาล แพรัตกุล, 2520) คุณค่าของข้อมูลที่ได้จากข้อสอบจึงขึ้นอยู่กับคุณภาพของเครื่องมือเป็นด้านหลัก

หัวใจของการสร้างข้อสอบที่ดีคือความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) ระหว่าง “จุดประสงค์การเรียนรู้–การจัดการเรียนการสอน–การวัดและประเมินผล” (Biggs & Tang, 2011) หากข้อสอบขาดความตรง (Validity) หรือไม่สอดคล้องกับจุดประสงค์เชิงพฤติกรรม ข้อมูลที่ได้ย่อมบิดเบือน ไม่อาจนำไปตัดสินคุณค่าของผู้เรียนได้อย่างยุติธรรม (ไพศาล หวังพานิช, 2526) ด้วยเหตุนี้ “ตารางวิเคราะห์หลักสูตร” หรือ พิมพ์เขียวการสอบ (Test Blueprint) จึงเป็นกลไกสำคัญในการควบคุมน้ำหนักเนื้อหาและระดับพฤติกรรมให้ครอบคลุมและสมดุล อันเป็นหลักประกันความเที่ยงและความเป็นปรนัยของเครื่องมือ (พัชรินทร์ ชมภูวิเศษ, 2559)

### แก่นความคิดของบท

- ข้อสอบ = เครื่องมือกระตุ้นพฤติกรรมประดิษฐ์ให้ปรากฏและวัดได้
- ความตรง (Validity) คือคุณสมบัติที่สำคัญที่สุดของเครื่องมือวัด
- Test Blueprint คือพิมพ์เขียวที่ทำให้ข้อสอบครอบคลุมและยุติธรรม
- การเขียนข้อสอบต้องสอดคล้องกับระดับพฤติกรรมตามจุดประสงค์ (CLO)

## 4.2 จุดประสงค์การเรียนรู้ประจำบท

บทที่ 4 มุ่งพัฒนาผลลัพธ์การเรียนรู้หลัก CLO2 (วิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์ เครื่องมือวัด และการแปลผล) และ CLO3 (ออกแบบและสร้างเครื่องมือวัดที่เหมาะสมกับลักษณะพฤติกรรม) โดยจำแนกเป็นจุดประสงค์เชิงพฤติกรรม พร้อมระบุระดับพฤติกรรม วิธีจัดการเรียนรู้ และวิธีวัดที่สอดคล้องกัน ดังตาราง 4.1

ตาราง 4.1 ความสอดคล้องระหว่างจุดประสงค์ของบท-CLO-ระดับพฤติกรรม-กิจกรรม-วิธีวัด/เครื่องมือ

| จุดประสงค์เชิงพฤติกรรม                                   | CLO        | ระดับพฤติกรรม       | กิจกรรมการเรียนรู้         | วิธีวัด/เครื่องมือ                 |
|----------------------------------------------------------|------------|---------------------|----------------------------|------------------------------------|
| อธิบายความหมาย บทบาท ประเภทของข้อสอบวัดผลสัมฤทธิ์        | CLO2       | จำ-เข้าใจ           | บรรยายเชิงปฏิสัมพันธ์      | แบบทดสอบก่อน-หลังเรียน             |
| วิเคราะห์ความสัมพันธ์จุดประสงค์-เครื่องมือ-ระดับพฤติกรรม | CLO2       | วิเคราะห์           | อภิปราย/วิเคราะห์กรณีศึกษา | ใบงานวิเคราะห์ข้อสอบ + รุบริคจำแนก |
| ออกแบบและสร้างตารางวิเคราะห์หลักสูตร (Blueprint)         | CLO3       | ประยุกต์-สร้างสรรค์ | TOS Workshop               | ชิ้นงาน Blueprint + รุบริค         |
| เขียนข้อสอบเลือกตอบและอัตนัยที่มีคุณภาพ                  | CLO3       | สร้างสรรค์          | Item Writing Clinic        | ชิ้นงานชุดข้อสอบ + รุบริค          |
| ตรวจสอบคุณภาพข้อสอบด้วยดัชนีความสอดคล้อง (IOC)           | CLO2, CLO3 | วิเคราะห์-ประยุกต์  | Peer Review (IOC)          | แบบประเมิน IOC + เกณฑ์             |

## 4.3 คำสำคัญ

ข้อสอบวัดผลสัมฤทธิ์ (Achievement Test) • พฤติกรรมประดิษฐ์ (Constructed Behavior) • ความตรง (Validity) • ความเที่ยง (Reliability) • ความเป็นปรนัย (Objectivity) • ความยุติธรรม (Fairness) • ครอบคลุม (Content Coverage) • ตารางวิเคราะห์หลักสูตร/พิมพ์เขียวการสอบ (Table of Specifications / Test Blueprint) • จุดประสงค์เชิงพฤติกรรม (Behavioral Objective) • ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) • ข้อสอบเลือกตอบ (Multiple Choice Item) • ส่วนนำ (Stem) • ตัวลง

## 4.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอน ประกาศในชั้นเรียนหมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนใน สัปดาห์ถัดไป

## 4.5 เนื้อหาวิชา

---

### 4.5.1 ความหมายและบทบาทของข้อสอบวัดผลสัมฤทธิ์

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน (Achievement Test) หมายถึง เครื่องมือที่ใช้วัดปริมาณ ความรู้ ทักษะ และสมรรถภาพทางสมองที่ผู้เรียนได้รับหลังผ่านประสบการณ์การเรียนรู้ตามหลักสูตรในระยะเวลา หนึ่ง (พัชรินทร์ ขมภูวิเศษ, 2559; สำเริง บุญเรืองรัตน์, 2534) ศาสตราจารย์ ดร.ชวาล แพร่ตกุล (2520) อธิบาย ว่าข้อสอบที่ดีเปรียบเสมือน “เครื่องมือทางวิทยาศาสตร์” ที่ออกแบบขึ้นเพื่อกระตุ้นให้ผู้เรียนแสดง “พฤติกรรม ประดิษฐ์ (Constructed Behavior)” คือพฤติกรรมตัวแทนที่สะท้อนความสามารถภายในซึ่งมองไม่เห็นโดยตรง ออกมาให้สังเกตและวัดปริมาณได้ ดังนั้นข้อสอบจึงมิใช่ชุดคำถามที่ครูกัดขึ้นตามใจชอบ แต่ต้องผ่านการวางแผน อย่างพิถีพิถันเพื่อให้พฤติกรรมที่ถูกเร้าออกมานั้นตรงกับสิ่งที่ต้องการวัดจริง

เมื่อจำแนกตามผู้สร้างและมาตรฐาน ข้อสอบวัดผลสัมฤทธิ์แบ่งได้เป็นสองกลุ่ม คือ ข้อสอบที่ครูสร้างขึ้น (Teacher-made Test) ซึ่งใช้ในชั้นเรียนเพื่อวัดผลการเรียนรู้ตามหลักสูตรที่ครูสอนเอง กับ ข้อสอบมาตรฐาน (Standardized Test) ซึ่งสร้างโดยผู้เชี่ยวชาญ ผ่านการทดลองใช้และวิเคราะห์คุณภาพอย่างเป็นระบบ มีเกณฑ์ ปกติ (Norm) สำหรับเปรียบเทียบในวงกว้าง (Miller, Linn, & Gronlund, 2013) เมื่อพิจารณาตามบทบาท หน้าที่ในชั้นเรียน ข้อสอบทำหน้าที่สำคัญสามประการ ได้แก่ การวินิจฉัย (Diagnostic) เพื่อค้นหาจุดแข็งจุดอ่อน ของผู้เรียนเป็นรายบุคคล การประเมินเพื่อพัฒนา (Formative) เพื่อให้ข้อมูลย้อนกลับระหว่างเรียนและปรับการ สอน และการประเมินเพื่อสรุปผล (Summative) เพื่อตัดสินผลการเรียนและรับรองมาตรฐานเมื่อจบหน่วยหรือ รายวิชา (Nitko & Brookhart, 2019)

#### กล่องที่ 4.1 แยกให้ชัด: วัด-ทดสอบ-ประเมิน และ “พฤติกรรมประดิษฐ์”

- การวัดผล (Measurement): ให้ “ตัวเลข” แก่คุณลักษณะ เช่น สอบได้ 32 จาก 40 คะแนน
- การทดสอบ (Testing): การวัดที่ใช้ “แบบทดสอบ” เป็นเครื่องมือเราพฤติกรรม
- การประเมินผล (Evaluation): ตัดสินคุณค่าจากตัวเลข เช่น  $32/40 =$  ผ่านเกณฑ์ระดับดี
- พฤติกรรมประดิษฐ์ (Constructed Behavior): พฤติกรรมตัวแทนที่ข้อสอบเราให้ปรากฏ เพื่อสะท้อนความรู้ภายในที่วัดโดยตรงไม่ได้ (ชวาล แพรัตกุล, 2520)

ตัวอย่างในชั้นเรียนไทย: หากครูวิทยาศาสตร์ตั้งเป้าหมายให้นักเรียน “แก้ปัญหาเชิงวิทยาศาสตร์ได้” การออกข้อสอบที่ถามเพียงนิยามหรือสูตร ย่อมเราได้แค่พฤติกรรมการจำ ไม่สะท้อนความสามารถในการแก้ปัญหา ครูจึงต้องสร้างสถานการณ์ปัญหาให้ผู้เรียนวิเคราะห์และประยุกต์ความรู้ เพื่อให้ “พฤติกรรมประดิษฐ์” ที่ปรากฏตรงกับสมรรถนะที่ต้องการวัดจริง การออกแบบข้อสอบจึงต้องดำเนินอย่างเป็นระบบตั้งแต่ต้นจนจบ พัทธินทร์ ชมภูวิเศษ (2559, น. 73–75) สรุปขั้นตอนการสร้างแบบทดสอบวัดผลสัมฤทธิ์ไว้ 8 ขั้นตอน ดังภาพที่ 4.1 ซึ่งจะเป็นกรอบเดินเรื่องของทั้งบทนี้

ภาพที่ 4.1 ขั้นตอนการสร้างแบบทดสอบวัดผลสัมฤทธิ์ 8 ขั้นตอน

| ขั้นที่ | ขั้นตอน                                       | สาระสำคัญของขั้นตอน                                                         |
|---------|-----------------------------------------------|-----------------------------------------------------------------------------|
| 1       | วิเคราะห์หลักสูตร/สร้างตารางวิเคราะห์หลักสูตร | กำหนดเนื้อหา×พฤติกรรมและน้ำหนักความสำคัญ (Test Blueprint) เป็นกรอบออกข้อสอบ |
| 2       | กำหนดจุดประสงค์การเรียนรู้                    | ระบุพฤติกรรมที่คาดหวังให้วัดได้ ล่วงหน้าและสังเกตได้                        |
| 3       | กำหนดชนิดข้อสอบและศึกษาวิธีสร้าง              | เลือกชนิดข้อสอบให้สอดคล้องจุดประสงค์และวัยผู้เรียน                          |
| 4       | เขียนข้อสอบ                                   | ลงมือเขียนตามจำนวน/พฤติกรรมที่ระบุในตารางวิเคราะห์หลักสูตร                  |
| 5       | ตรวจทานข้อสอบ                                 | ทบทวนความถูกต้องตามหลักวิชาและความครบถ้วนก่อนพิมพ์                          |
| 6       | จัดพิมพ์แบบทดลอง                              | จัดรูปแบบ ใส่คำชี้แจง (Direction) ให้พร้อมทดลองใช้                          |
| 7       | ทดลองสอบและวิเคราะห์ข้อสอบ                    | นำไปสอบกลุ่มใกล้เคียง วิเคราะห์ความยาก-อำนาจจำแนกเพื่อปรับปรุง              |
| 8       | จัดทำแบบทดสอบฉบับจริง                         | คัดข้อที่มีคุณภาพไว้ใช้กับกลุ่มเป้าหมาย                                     |

ที่มา: ดัดแปลงจาก พัทธินทร์ ชมภูวิเศษ (2559, น. 73–75); พิชิต ฤทธิ์จรูญ (2557, น. 97–98)

#### 4.5.2 คุณลักษณะของเครื่องมือวัดผลที่มีคุณภาพ 5 ประการ

คุณภาพของข้อมูลที่น่าไปตัดสินผู้เรียนขึ้นอยู่กับคุณภาพของเครื่องมือวัดโดยตรง หากเครื่องมือบกพร่อง การสรุปผลย่อมคลาดเคลื่อนและส่งผลกระทบต่ออนาคตของผู้เรียน นักวัดผลจึงกำหนดคุณลักษณะของเครื่องมือวัดที่ดีไว้ 5 ประการ ได้แก่ ความตรง ความเที่ยง ความเป็นปรนัย ความยุติธรรม และความครอบคลุม

(พัชรินทร์ ชมภูวิเศษ, 2559; Miller, Linn, & Gronlund, 2013) คุณลักษณะเหล่านี้ไม่ได้มีน้ำหนักเท่ากันทั้งหมด แต่มีลำดับความสำคัญและความสัมพันธ์เชื่อมโยงกัน สารเชิงลึกของความตรงและความเที่ยง ศึกษาในบทที่ 7 และบทที่ 8

ความตรง (Validity) คือคุณสมบัติที่บ่งบอกว่าเครื่องมือวัดได้ “ตรง” ในสิ่งที่ต้องการวัดจริงตามจุดประสงค์ ถือเป็นคุณสมบัติที่สำคัญที่สุด เพราะหากวัดผิดเป้าหมายตั้งแต่ต้น ตัวเลขที่ได้ก็ไร้ความหมายแม้จะแม่นยำเพียงใด สำหรับข้อสอบวัดผลสัมฤทธิ์ ความตรงที่สำคัญที่สุดคือความตรงเชิงเนื้อหา (Content Validity) ซึ่งตรวจสอบว่าข้อสอบเป็นตัวแทนของเนื้อหาและพฤติกรรมตามหลักสูตรครบถ้วนหรือไม่ โดยเทียบกับตารางวิเคราะห์หลักสูตรและอาศัยดุลยพินิจของผู้เชี่ยวชาญ (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

ความเที่ยงหรือความเชื่อมั่น (Reliability) คือความคงเส้นคงวาของผลการวัด หากนำข้อสอบไปวัดซ้ำกับกลุ่มเดิมในสภาพใกล้เคียงกัน ผลควรใกล้เคียงกัน ความเป็นปรนัย (Objectivity) คือความชัดเจนของคำถาม การตรวจให้คะแนน และการแปลผลที่ตรงกันไม่ว่าผู้ตรวจจะเป็นใคร ความยุติธรรม (Fairness) คือการไม่เปิดโอกาสให้ผู้เรียนกลุ่มใดได้เปรียบหรือเสียเปรียบจากปัจจัยที่ไม่เกี่ยวกับความสามารถ เช่น เพศ ฐานะ หรือถิ่นที่อยู่ และความครอบคลุม (Content Coverage) คือการที่ข้อสอบเป็นตัวแทนที่ดีของเนื้อหาทั้งหมด ไม่กระจุกอยู่เพียงบางหัวข้อที่ง่าย ดังสรุปในตาราง 4.2

ตาราง 4.2 คุณลักษณะของเครื่องมือวัดผลที่มีคุณภาพ 5 ประการ และผลกระทบหากขาดคุณลักษณะนั้น

| คุณลักษณะ                   | นิยามเชิงวิชาการ                                  | แนวทางตรวจสอบ                                            | ผลกระทบหากขาด                                  |
|-----------------------------|---------------------------------------------------|----------------------------------------------------------|------------------------------------------------|
| ความตรง (Validity)          | วัดได้ตรงตามจุดประสงค์ที่ต้องการวัด (สำคัญที่สุด) | หาค่า IOC โดยผู้เชี่ยวชาญ เทียบกับตารางวิเคราะห์หลักสูตร | สรุปผลในสิ่งที่ไม่ได้วัดจริง ข้อมูลไร้ความหมาย |
| ความเที่ยง (Reliability)    | ผลการวัดคงเส้นคงวาเมื่อวัดซ้ำ                     | หาค่าความเชื่อมั่น เช่น KR-20, สัมประสิทธิ์แอลฟา         | ขาดความน่าเชื่อถือ การตัดสินใจมีความเสี่ยงสูง  |
| ความเป็นปรนัย (Objectivity) | คำถามชัดเจน ตรวจตรงกัน แปลผลตรงกัน                | ตรวจสอบภาษาและเกณฑ์ให้คะแนน (เฉลย/รูปрик)                | คะแนนขึ้นกับดุลยพินิจผู้ตรวจ เกิดอคติ          |
| ความยุติธรรม (Fairness)     | ไม่เอื้อหรือกีดกันผู้เรียนกลุ่มใดอย่างไม่เป็นธรรม | ตรวจเนื้อหาไม่ให้อคติทางเพศ/ฐานะ/ถิ่น                    | ผู้เรียนเสียสิทธิ์ที่พึงได้จากความสามารถจริง   |
| ความครอบคลุม (Coverage)     | เป็นตัวแทนเนื้อหา-พฤติกรรมครบตามน้ำหนัก           | เทียบสัดส่วนข้อสอบกับตารางวิเคราะห์หลักสูตร              | ประเมินคลาดเคลื่อน เก่งเฉพาะจุดที่สุ่มถาม      |

ความสัมพันธ์ระหว่างคุณลักษณะทั้งห้ามีหลักการสำคัญที่ควรต้องเข้าใจ คือ ความตรงสำคัญกว่าความเที่ยงเสมอ และความเที่ยงเป็นเงื่อนไขจำเป็นแต่ไม่เพียงพอของความตรง กล่าวคือ เครื่องมือที่ตรงต้องเที่ยงด้วย แต่เครื่องมือที่เที่ยงอาจไม่ตรงก็ได้ ดังภาพเปรียบเทียบในกล่องที่ 4.2 ครูควรระวังความเข้าใจผิดที่พบบ่อยว่า “ข้อสอบที่ให้คะแนนครั้งที่ทุกครั้งคือข้อสอบที่ดี” เพราะความคงที่ (เที่ยง) ไม่ได้รับประกันว่าวัดตรงจุดประสงค์ (ตรง) การยึดความเที่ยงโดยละเลยความตรง จึงเป็นกับดักที่ทำให้ได้ข้อสอบที่ “แม่นยำแต่ผิดเป้า”

#### กล่องที่ 4.2 ข้อพึงระวัง: “เที่ยงสูงแต่ไม่ตรง” อันตรายอย่างไร

เปรียบเทียบวัดเหมือนการยิงปืนเข้าเป้า:

- ตรงและเที่ยง = กระสุนรวมกลุ่มกลางเป้า → เครื่องมือในอุดมคติ
- เที่ยงแต่ไม่ตรง = กระสุนรวมกลุ่มแน่น แต่เกาะอยู่มุมเป้า → คงเส้นคงวาแต่ผิดเป้าหมาย
- ไม่เที่ยงและไม่ตรง = กระสุนกระจัดกระจาย → ใช้ไม่ได้

ตัวอย่าง: ข้อสอบที่วัด “การท่องจำสูตร” ได้คะแนนครั้งที่ทุกครั้ง (เที่ยงสูง) แต่ครูตั้งใจวัด “การแก้ปัญหา” (จึงไม่ตรง) ผลคือคะแนนน่าเชื่อถือแต่สรุปผิดสมรรถนะ

#### 4.5.3 ความสัมพันธ์ระหว่างจุดประสงค์การเรียนรู้กับข้อสอบ (Constructive Alignment)

หัวใจของการสร้างข้อสอบที่ดีคือ ความสอดคล้องเชิงโครงสร้าง (Constructive Alignment) ซึ่งเป็นแนวคิดที่เสนอโดย Biggs และ Tang (2011) ว่าด้วยการจัดองค์ประกอบสามส่วน คือ ผลลัพธ์การเรียนรู้ (Learning Outcomes) กิจกรรมการเรียนการสอน และการวัดประเมินผล ให้สอดคล้องเป็นแนวเดียวกัน กล่าวคือ “สิ่งที่ประกาศว่าจะสอน” “สิ่งที่จัดกิจกรรมให้เรียน” และ “สิ่งที่นำมาออกข้อสอบ” ต้องเป็นเรื่องเดียวกัน ข้อสอบที่ขาดความสอดคล้องนี้ ย่อมให้ข้อมูลที่บิดเบือนและไม่ยุติธรรมต่อผู้เรียน เพราะวัดในสิ่งที่ไม่ได้สอนหรือไม่ตรงกับจุดประสงค์ที่ตั้งไว้

การสร้างข้อสอบจึงต้องเริ่มจากจุดประสงค์เชิงพฤติกรรม (Behavioral Objective) ที่ระบุพฤติกรรมที่สังเกตและวัดได้อย่างชัดเจน (สำเร็จ บุญเรืองรัตน์, 2534) แล้วเชื่อมโยงต่อไปยังระดับพฤติกรรมตามอนุกรมวิธานของบลูมฉบับปรับปรุง (Revised Bloom’s Taxonomy) ของ Anderson และ Krathwohl (2001) ซึ่งจำแนกพฤติกรรมเป็น 6 ระดับ ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ การกำหนดระดับพฤติกรรมให้ถูกต้องเป็นสิ่งสำคัญ เพราะเป็นตัวกำหนดว่าควรใช้ข้อสอบชนิดใดจึงจะวัดได้ตรง ความเชื่อมโยงนี้เห็นเป็นลูกโซ่ดังภาพที่ 4.2

ภาพที่ 4.2 ลูกโซ่ความสอดคล้องเชิงโครงสร้าง: จากจุดประสงค์สู่ข้อสอบ

| ลำดับ | องค์ประกอบในลูกโซ่         | ตัวอย่าง                                                    |
|-------|----------------------------|-------------------------------------------------------------|
| 1     | ผลลัพธ์การเรียนรู้ (CLO)   | ผู้เรียนออกแบบและสร้างข้อสอบที่มีคุณภาพได้ (CLO3)           |
| 2     | จุดประสงค์เชิงพฤติกรรม     | เขียนข้อสอบเลือกตอบที่มีตัวลวงทำงานได้อย่างน้อย 8 ใน 10 ข้อ |
| 3     | ระดับพฤติกรรม (Bloom)      | สร้างสรรค์ (Create)                                         |
| 4     | เนื้อหา/กิจกรรมการเรียนรู้ | คลินิกการเขียนข้อสอบ (Item Writing Clinic)                  |
| 5     | ชนิดข้อสอบ/เครื่องมือวัด   | ชิ้นงานชุดข้อสอบ +<br>รูปรีประเมินคุณภาพข้อสอบ              |

ที่มา: ดัดแปลงจาก Biggs และ Tang (2011); Anderson และ Krathwohl (2001)

การจับคู่ระดับพฤติกรรมกับชนิดข้อสอบจึงต้องทำอย่างพิถีพิถัน ระดับ “จำ-เข้าใจ” เหมาะกับข้อสอบเลือกตอบหรือถูก-ผิดที่ตรวจได้รวดเร็ว ขณะที่ระดับ “วิเคราะห์-ประเมินค่า-สร้างสรรค์” ต้องอาศัยข้อสอบอัตนัยหรือชิ้นงานที่เปิดให้ผู้เรียนแสดงกระบวนการคิด นอกจากนี้ระดับ “เข้าใจ” ยังจำแนกย่อยได้อีกสามลักษณะ คือ การแปลความ (เปลี่ยนรูปแบบของสารโดยคงความหมายเดิม) การตีความ (จับใจความสำคัญ) และการขยายความ (คาดคะเนสิ่งที่จะเกิดต่อไปจากข้อมูลเดิม) ดังตัวอย่างการจับคู่ในตาราง 4.3

ตาราง 4.3 ตัวอย่างการจับคู่ระดับพฤติกรรมพุทธิพิสัยกับชนิดข้อสอบที่เหมาะสม

| ระดับพฤติกรรม (Bloom) | กริยาพฤติกรรมที่ใช้บ่อย               | ตัวอย่างจุดประสงค์                         | ชนิดข้อสอบที่เหมาะสม             |
|-----------------------|---------------------------------------|--------------------------------------------|----------------------------------|
| จำ (Remember)         | ระบุ บอก จำแนกชื่อ                    | ระบุส่วนประกอบของข้อสอบเลือกตอบได้         | เลือกตอบ จับคู่ เต็มคำ           |
| เข้าใจ (Understand)   | อธิบาย สรุป ยกตัวอย่าง                | อธิบายความต่างของความตรงกับความเที่ยงได้   | เลือกตอบเชิงสถานการณ์ อัตนัยสั้น |
| ประยุกต์ใช้ (Apply)   | คำนวณ ใช้ แก้ปัญหา                    | คำนวณค่า IOC จากคะแนนผู้เชี่ยวชาญได้       | เลือกตอบเชิงคำนวณ ข้อสอบปฏิบัติ  |
| วิเคราะห์ (Analyze)   | จำแนก เปรียบเทียบ ตรวจสอบหาจุดบกพร่อง | ระบุข้อบกพร่องของตัวลวงในข้อสอบที่กำหนดให้ | ใบงานวิเคราะห์ อัตนัย            |

|                          |                             |                                              |                                |
|--------------------------|-----------------------------|----------------------------------------------|--------------------------------|
| ประเมินค่า<br>(Evaluate) | วิจารณ์ ตัดสิน<br>ให้เหตุผล | ตัดสินคุณภาพข้อสอบพร้อมเหตุผลเชิงวิชา<br>การ | อรรถนัยเชิงวิพากษ์ +<br>รูปรีด |
| สร้างสรรค์<br>(Create)   | ออกแบบ สร้าง<br>ประพันธ์    | สร้างชุดข้อสอบตามตารางวิเคราะห์หลักสูตร      | ชิ้นงาน/โครงการ +<br>รูปรีด    |

ที่มา: ผู้เขียน ดัดแปลงจาก Anderson และ Krathwohl (2001); พัทรินทร์ ชมภูวิเศษ (2559)

ข้อควรระวังที่พบบ่อยคือการประกาศจุดประสงค์ระดับสูง เช่น “วิเคราะห์” หรือ “ประเมินค่า” แต่ออกข้อสอบที่วัดเพียงการจำ ทำให้เกิดช่องว่างระหว่างสิ่งที่ประกาศกับสิ่งที่วัดจริง งานวิจัยในบริบทอุดมศึกษาไทยพบว่า การสอนแบบเน้นผลลัพธ์อาจเผชิญปัญหา “สอนเพื่อสอบ (Teaching to the Test)” หากครูยึดข้อสอบเดิมเป็นตัวตั้ง แทนที่จะเริ่มจากจุดประสงค์ (Pathamathamakul, 2022) การยึดลูกโซ่ความสอดคล้องตั้งแต่ต้นจึงเป็นเครื่องป้องกันไม่ให้เกิดความล้าสมัย

#### 4.5.4 พิมพ์เขียวการสอบ: ตารางวิเคราะห์หลักสูตร (Test Blueprint / Table of Specifications)

ตารางวิเคราะห์หลักสูตร (Table of Specifications) หรือพิมพ์เขียวการสอบ (Test Blueprint) คือตารางสองทางที่แสดงความสัมพันธ์ระหว่างเนื้อหากับระดับพฤติกรรมที่ต้องการวัด พร้อมระบุน้ำหนักความสำคัญ จำนวนข้อ และคะแนนในแต่ละช่อง (สมบุญ ภู่นวล อ้างถึงใน พัทรินทร์ ชมภูวิเศษ, 2559) ตารางนี้ทำหน้าที่เป็น “แบบแปลน” ที่ควบคุมให้ข้อสอบครอบคลุมเนื้อหาและพฤติกรรมอย่างสมดุล ป้องกันไม่ให้ครูออกข้อสอบตามความถนัดส่วนตัวหรือกระจุกอยู่เฉพาะหัวข้อที่ง่าย จึงเป็นหลักประกันสำคัญของทั้งความตรงเชิงเนื้อหาและความยุติธรรมการสร้างตารางวิเคราะห์หลักสูตรอย่างเป็นระบบมีขั้นตอนชัดเจน โดยเริ่มจากการแต่งตั้งคณะกรรมการผู้เชี่ยวชาญในศาสตร์เดียวกัน 5–10 คน ร่วมกันให้น้ำหนักความสำคัญของแต่ละช่องเนื้อหา×พฤติกรรม แล้วเฉลี่ยและปรับมาตราส่วนให้ได้ผลรวมที่สะดวกต่อการคำนวณจำนวนข้อ ดังสรุปขั้นตอนในกล่องที่ 4.3 (พัชรินทร์ ชมภูวิเศษ, 2559; สุนันท์ ศลโกสุม, 2525)

##### กล่องที่ 4.3 6 ขั้นตอนการสร้างตารางวิเคราะห์หลักสูตรอย่างเป็นระบบ

- 1) แต่งตั้งคณะกรรมการผู้เชี่ยวชาญในรายวิชา/ศาสตร์เดียวกัน จำนวน 5–10 คน
- 2) วิเคราะห์เนื้อหาออกเป็นหน่วยย่อยที่เรียงลำดับก่อน–หลัง
- 3) สร้างตารางสองทางระหว่างเนื้อหากับพฤติกรรม (พุทธิ-จิต-ทักษะพิสัย)
- 4) ให้กรรมการแต่ละคนให้น้ำหนักช่องละ 0–10 คะแนน (9–10 = สำคัญมาก ... 0–1 = สำคัญน้อย)
- 5) หาค่าเฉลี่ยน้ำหนักของกรรมการทุกคนในแต่ละช่อง
- 6) ปรับมาตราส่วนเป็นฐาน 1,000 (หรือ 100) เพื่อแปลงเป็นจำนวนข้อ:  $\text{นำค่าเฉลี่ยช่องนั้น} \times 1000 \div \text{ผลรวมค่าเฉลี่ย}$

ตัวอย่างตารางวิเคราะห์หลักสูตรในตาราง 4.4 แสดงการกระจายข้อสอบ 20 ข้อ 20 คะแนน ของหน่วย “หลักการสร้างและเขียนข้อสอบ” ครูจะอ่านค่าจากตารางโดยดูว่าแต่ละหัวข้อเนื้อหา ควรออกข้อสอบกี่ข้อในแต่ละระดับพฤติกรรม เช่น หัวข้อ “การเขียนข้อสอบเลือกตอบ/อัตนัย” มีน้ำหนัก 30% จึงได้ 6 ข้อ โดยเน้นระดับนำไปใช้-วิเคราะห์มากกว่าการจำ การกำหนดเช่นนี้ทำให้ข้อสอบสะท้อนเวลาและความสำคัญที่ใช้สอนจริง

**ตาราง 4.4 ตัวอย่างตารางวิเคราะห์หลักสูตร หน่วย “หลักการสร้างและเขียนข้อสอบ” (รวม 20 ข้อ / 20 คะแนน)**

| หัวข้อเนื้อหา                       | จำ-เข้าใจ | นำไปใช้-วิเคราะห์ | ประเมินค่า-สร้างสรรค์ | น้ำหนัก (%) | ข้อ | คะแนน |
|-------------------------------------|-----------|-------------------|-----------------------|-------------|-----|-------|
| 1. ความหมายและบทบาทของข้อสอบ        | 3         | 1                 | 0                     | 20          | 4   | 4     |
| 2. คุณภาพเครื่องมือ 5 ประการ        | 2         | 2                 | 1                     | 25          | 5   | 5     |
| 3. การเขียนข้อสอบเลือกตอบ/อัตนัย    | 2         | 3                 | 1                     | 30          | 6   | 6     |
| 4. การตรวจสอบคุณภาพ (Blueprint/IOC) | 2         | 2                 | 1                     | 25          | 5   | 5     |
| รวม                                 | 9         | 8                 | 3                     | 100         | 20  | 20    |

ที่มา: ผู้เขียน ออกแบบตามวิธีการของ พัชรินทร์ ชมภูวิเศษ (2559, น. 51–53)

ข้อควรระวังในการใช้ตารางวิเคราะห์หลักสูตร คือ ครูต้องไม่กำหนดน้ำหนักตามความรู้สึกหรือความถนัดส่วนตัว แต่ควรอิงเวลาเรียน ความสำคัญของจุดประสงค์ และความเห็นของผู้เชี่ยวชาญ อีกทั้งเมื่อเขียนข้อสอบเสร็จต้องย้อนกลับมาตรวจว่าจำนวนข้อจริงตรงกับที่ระบุไว้ในตารางหรือไม่ การข้ามขั้นตอนนี้คือสาเหตุที่พบบ่อยที่สุดของข้อสอบที่ “ครอบคลุมไม่สมดุล” เช่น ออกข้อสอบความจำมากเกินไปเพราะเขียนง่าย แล้วละเลยพฤติกรรมระดับสูงที่เป็นเป้าหมายสำคัญของหลักสูตร

#### 4.5.5 ประเภทของข้อสอบวัดผลสัมฤทธิ์

ข้อสอบวัดผลสัมฤทธิ์จำแนกตามลักษณะการตอบได้เป็นสองกลุ่มใหญ่ คือ ข้อสอบปรนัย (Objective Test) ซึ่งกำหนดคำตอบไว้ให้เลือกหรือเติมสั้น ๆ มีเฉลยแน่นอนจนจึงตรวจได้ตรงกันไม่ว่าผู้ตรวจจะเป็นใคร และข้อสอบอัตนัย (Subjective/Essay Test) ซึ่งเปิดให้ผู้เรียนเรียบเรียงคำตอบด้วยตนเอง (พัชรินทร์ ชมภูวิเศษ, 2559; พิชิต ฤทธิ์จรูญ, 2557) ข้อสอบปรนัยยังจำแนกย่อยได้อีกหลายชนิด เช่น เลือกตอบ (Multiple Choice)

ถูก-ผิด (True-False) จับคู่ (Matching) และเติมคำ/ตอบสั้น (Completion/Short Answer) ส่วนข้อสอบอัตนัยจำแนกเป็นแบบจำกัดคำตอบ (Restricted Response) และแบบขยายคำตอบ (Extended Response)

การเลือกชนิดข้อสอบต้องพิจารณาจากระดับพฤติกรรมที่ต้องการวัดเป็นหลัก มิใช่เลือกตามความสะดวกในการตรวจ ข้อสอบปรนัยมีจุดเด่นที่วัดได้กว้าง ครอบคลุมเนื้อหาครบถ้วน ตรวจสอบเร็วและเป็นปรนัยสูง แต่สร้างตัวลองที่ดียากและวัดความคิดสร้างสรรค์ได้จำกัด ขณะที่ข้อสอบอัตนัยวัดพฤติกรรมระดับสูงอย่างการสังเคราะห์และประเมินค่าได้ดี และป้องกันการเดา แต่ตรวจยาก ใช้เวลานาน และมีความเป็นปรนัยต่ำหากไม่มีเกณฑ์การให้คะแนน ดังเปรียบเทียบในตาราง 4.5 (ล้วน สายยศ และอังคณา สายยศ, 2543)

ตาราง 4.5 การเปรียบเทียบข้อสอบแต่ละชนิด: ข้อดี ข้อจำกัด และพฤติกรรมที่เหมาะสม

| ชนิดข้อสอบ                  | ข้อดี                                           | ข้อจำกัด                            | เหมาะกับพฤติกรรม                |
|-----------------------------|-------------------------------------------------|-------------------------------------|---------------------------------|
| เลือกตอบ (Multiple Choice)  | วัดกว้าง ครอบคลุม ตรวจสอบเร็ว เป็นปรนัยสูง      | สร้างตัวลองที่ดียาก ใช้เวลาสร้างนาน | จำ-เข้าใจ ถึง วิเคราะห์         |
| ถูก-ผิด (True-False)        | สร้างและตอบเร็ว วัดได้หลายข้อ                   | มีโอกาสเดาสูงถึงร้อยละ 50           | จำ-เข้าใจ                       |
| จับคู่ (Matching)           | วัดความสัมพันธ์/ข้อเท็จจริงได้มากในเนื้อที่น้อย | เหมาะกับเนื้อหาที่เป็นหมวดเดียวกัน  | จำ-เข้าใจ                       |
| เติมคำ/ตอบสั้น (Completion) | ลดการเดา วัดความจำได้แม่นยำ                     | การสะกด/หลายคำตอบอาจกระทบการตรวจ    | จำ-เข้าใจ-ประยุกต์              |
| อัตนัย (Essay)              | วัดการคิดระดับสูง ป้องกันการเดา                 | ตรวจยาก ใช้เวลานาน เป็นปรนัยต่ำ     | วิเคราะห์-ประเมินค่า-สร้างสรรค์ |

ที่มา: ดัดแปลงจาก พัชรินทร์ ชมภูวิเศษ (2559, น. 80, 102); ล้วน สายยศ และอังคณา สายยศ (2543)

#### 4.5.6 หลักการเขียนข้อสอบแบบเลือกตอบ (Multiple Choice Item Writing)

ข้อสอบเลือกตอบเป็นข้อสอบปรนัยที่นิยมใช้มากที่สุด เพราะวัดได้ครอบคลุมและตรวจเป็นปรนัย โครงสร้างของข้อสอบเลือกตอบประกอบด้วยสองส่วนหลัก คือ ส่วนที่เป็นข้อคำถามหรือส่วนนำ (Stem) และตัวเลือก (Options) ซึ่งแบ่งเป็นตัวถูก (Key/Correct Choice) เพียงหนึ่งคำตอบ กับตัวลอง (Distractors/Foils) ที่เหลือ โดยทั่วไปมี 3-5 ตัวเลือกขึ้นกับวัยและระดับชั้น (พัชรินทร์ ชมภูวิเศษ, 2559;

อนุวัติ คุณแก้ว, 2566) ในบางกรณีอาจเพิ่มสถานการณ์ (Situation) เพื่อวัดพฤติกรรมระดับสูง  
 ดังกายวิภาคในภาพที่ 4.3

ภาพที่ 4.3 กายวิภาคของข้อสอบเลือกตอบ

| ส่วนประกอบ                    | หน้าที่และตัวอย่าง                                                           |
|-------------------------------|------------------------------------------------------------------------------|
| สถานการณ์ (Situation) [ถ้ามี] | ข้อมูล รูปภาพ หรือบทอ่านที่ใช้เป็นฐานการคิด เช่น ตารางคะแนนผู้เชี่ยวชาญ      |
| ส่วนนำ (Stem)                 | ประโยคคำถามที่สมบูรณ์ในตัว เช่น “ตัวลงที่ดีควรมีลักษณะอย่างไร”               |
| ตัวเลือก (Options)            | คำตอบที่กำหนดให้เลือก โดยทั่วไป 3–5 ตัว: ก. ข. ค. ง.                         |
| ตัวถูก (Key)                  | ตัวเลือกที่ถูกต้องที่สุดเพียงหนึ่งเดียว                                      |
| ตัวลวง (Distractors)          | ตัวเลือกที่เหลือ<br>ออกแบบให้สมเหตุสมผลพอจะดึงดูดผู้ที่ไม่รู้จักจริงให้เลือก |

ที่มา: ดัดแปลงจาก พชรินทร์ ชมภูวิเศษ (2559, น. 99–100)

หลักการเขียนส่วนนำที่ดีคือ ต้องมีใจความสมบูรณ์ในตัว เป็นคำถามที่ชัดเจนและตรงประเด็น ไม่เริ่มถามประเด็นหลักไว้ที่ตัวเลือก ไม่ใช้ภาษาชี้นำคำตอบ และหลีกเลี่ยงการใช้คำปฏิเสธซ้อนปฏิเสธที่ทำให้สับสน หากจำเป็นต้องใช้คำปฏิเสธควรขีดเส้นใต้หรือทำตัวหนาเน้นไว้ (พิชิต ฤทธิ์จรรยา, 2557; ล้วน สายยศ และอังคณา สายยศ, 2543) ส่วนการเขียนตัวเลือกนั้น ตัวเลือกทุกตัวต้องเป็นขนาน (Parallel) ทั้งโครงสร้างไวยากรณ์และความยาว มีคำตอบถูกต้องเพียงข้อเดียว เรียงลำดับอย่างมีระบบ และกระจายตำแหน่งคำตอบถูกต้องอย่างสม่ำเสมอไม่ให้ผู้สอบคาดเดาได้

หัวใจที่ยากที่สุดของการเขียนข้อสอบเลือกตอบคือการสร้างตัวลวงที่ “ทำงาน” (Functional Distractor) กล่าวคือ ตัวลวงต้องสมเหตุสมผลพอจะดึงดูดผู้ที่เข้าใจผิดให้เลือก มิใช่เขียนให้ผิดอย่างเห็นได้ชัดจนไม่มีใครเลือก หากตัวลวงใดไม่มีผู้เลือกเลย ถือว่าล้มเหลวในการทำหน้าที่ Gierl, Bulut, Guo และ Zhang (2017) เสนอว่ากลยุทธ์ที่ได้ผลที่สุด คือการสร้างตัวลวงจากความเข้าใจผิดที่พบบ่อย (Common Misconceptions) ของผู้เรียนจริง นอกจากนี้งานวิจัยเชิงสังเคราะห์ของ Rodriguez (2005) ยังพบว่าข้อสอบที่มี 3 ตัวเลือกที่มีคุณภาพ ให้ผลทางจิตมิติไม่ด้อยกว่า 4–5 ตัวเลือก และยังออกข้อสอบได้มากข้อขึ้นในเวลาเท่ากัน จึงไม่จำเป็นต้องฝึกสร้างตัวลวงที่อ่อนเพียงเพื่อให้ครบจำนวน (Haladyna, Downing, & Rodriguez, 2002)

#### กล่องที่ 4.4 ตัวอย่างการปรับปรุงข้อสอบเลือกตอบ (ก่อน → หลัง)

ข้อบกพร่อง (ก่อน): “ข้อใดกล่าวไม่ถูกต้องเกี่ยวกับความเที่ยง?”

ก. ความคงเส้นคงวาของคะแนน ข. วัดซ้ำได้ผลเหมือนเดิม

ค. เป็นหัวใจของการวัด ง. ถูกทุกข้อ

จุดอ่อน: ใช้คำปฏิเสธ “ไม่” ตัวเลือกยาวไม่เท่ากัน และมี “ถูกทุกข้อ” ที่ช่วยให้เดา

ปรับปรุง (หลัง): “หากนักเรียนกลุ่มเดิมทำข้อสอบชุดเดิมสองครั้งแล้วได้คะแนนต่างกันมาก แสดงว่าข้อสอบขาดคุณสมบัติใด”

ก. ความตรงเชิงเนื้อหา ข. ความเที่ยง ค. ความเป็นปรนัย ง. ความยาก

จุดแข็ง: เป็นสถานการณ์ระดับประยุกต์ ตัวเลือกขนานกัน วัดประเด็นเดียวชัดเจน เฉลย ข.

ตัวอย่างในชั้นเรียนไทย: ครูสังคมศึกษาที่ต้องการวัดความเข้าใจเรื่องประชาธิปไตย ควรเปลี่ยนจากคำถาม “ประชาธิปไตยหมายถึงอะไร” (วัดการจำ) เป็นการให้สถานการณ์การเลือกตั้งในหมู่บ้าน แล้วถามว่าการกระทำใดสอดคล้องกับหลักประชาธิปไตยมากที่สุด พร้อมออกแบบตัวลวงจากความเข้าใจผิดที่นักเรียนมักมี เช่น สับสนระหว่างเสียงข้างมากกับการละเมิดสิทธิเสียงข้างน้อย วิธีนี้ทำให้ข้อสอบยกระดับจากการจำสู่การวิเคราะห์ และมีตัวลวงที่ทำงานได้จริง

#### 4.5.7 หลักการเขียนข้อสอบอัตนัย (Essay Item Writing)

ข้อสอบอัตนัย (Essay Test) คือข้อสอบที่กำหนดเพียงโจทย์หรือสถานการณ์ แล้วเปิดให้ผู้เรียนคิดและเรียบเรียงคำตอบด้วยตนเอง โดยการเขียนบรรยาย อธิบาย วิเคราะห์ หรือวิพากษ์วิจารณ์ (Marshall & Hales, 1972 อ้างถึงใน พächรินทร์ ชมภูวิเศษ, 2559) ข้อสอบประเภทนี้เหมาะกับการวัดพฤติกรรมระดับสูง เช่น การจัดระเบียบความคิด การสังเคราะห์ และการประเมินค่า ซึ่งข้อสอบเลือกตอบทำได้ยาก ข้อสอบอัตนัยจำแนกเป็นแบบจำกัดคำตอบ (Restricted Response) ที่กำหนดขอบเขตคำตอบไว้แคบ กับแบบขยายคำตอบ (Extended Response) ที่เปิดกว้างให้แสดงความคิดอย่างอิสระ

การเขียนข้อสอบอัตนัยที่ดีต้องกำหนดคำถามให้ผู้เรียนแสดงกระบวนการคิด ไม่ใช่เพียงเล่าความจำระบุขอบเขตและเงื่อนไขของคำตอบให้ชัดเจน เช่น “จงวิเคราะห์...พร้อมยกตัวอย่าง 3 ประการ” และควรเขียนแนวคำตอบ (Model Answer) ไว้ล่วงหน้าเพื่อตรวจสอบความชัดเจนของโจทย์ (Gronlund, 1990 อ้างถึงใน พächรินทร์ ชมภูวิเศษ, 2559) เนื่องจากข้อสอบอัตนัยมีความเป็นปรนัยต่ำโดยธรรมชาติ การแนบเกณฑ์การให้คะแนนแบบรูบริก (Scoring Rubric) จึงเป็นสิ่งจำเป็นเพื่อยกระดับความเป็นปรนัยและลดความลำเอียงในการตรวจ (Nitko & Brookhart, 2019) ดังตัวอย่างรูบริกเชิงวิเคราะห์ในตาราง 4.6

ตาราง 4.6 ตัวอย่างเกณฑ์การให้คะแนนข้อสอบอัตนัยแบบเชิงวิเคราะห์ (Analytic Rubric 4 ระดับ)

| ประเด็นประเมิน         | ดีมาก (4)                                     | ดี (3)                                  | พอใช้ (2)                          | ปรับปรุง (1)                      |
|------------------------|-----------------------------------------------|-----------------------------------------|------------------------------------|-----------------------------------|
| เนื้อหาและการวิเคราะห์ | วิเคราะห์ลึกซึ้ง ครอบคลุม ถูกต้องตามหลัก วิชา | วิเคราะห์ถูกต้องชัดเจน แต่ยังไม่ลุ่มลึก | วิเคราะห์ได้บางส่วน ขาดความลุ่มลึก | ระบุประเด็นไม่ได้หรือ คลาดเคลื่อน |
| การจัดระบบความคิด      | เรียบเรียงเป็นระบบ มีเหตุผลเชื่อมโยง ชัด      | เป็นระบบ มีจุดเชื่อมโยงเล็กน้อยที่ขาด   | เรียบเรียงได้บ้าง แต่ลำดับสับสน    | ไม่เป็นระบบ จับประเด็นไม่ได้      |
| การใช้ภาษาวิชาการ      | ภาษาถูกต้อง สละสลวย ใช้ศัพท์เทคนิคเหมาะสม     | ภาษาชัดเจน มีข้อผิดพลาดเล็กน้อย         | พอเข้าใจ แต่เรียบเรียงไม่เป็นระบบ  | ใช้ภาษาไม่เหมาะสม สื่อความไม่ได้  |

ที่มา: ผู้เขียน ดัดแปลงจาก Nitko และ Brookhart (2019); พิชิต ฤทธิ์จรูญ (2557)

**กล่องที่ 4.5 ข้อพึงระวัง: ความลำเอียงในการตรวจข้อสอบอัตนัยและวิธีป้องกัน**

ความลำเอียงที่พบบ่อย (บุญเชิด ภิญญอนันตพงษ์, 2544 อ้างถึงใน พชรินทร์, 2559, น. 79):

- ผลสืบเนื่อง (Carry-over/Halo): ข้อแรกตอบดีทำให้ครูให้คะแนนข้ออื่นสูงตามไปด้วย
- ความใจดี (Generosity): ให้คะแนนสูงแม้คำตอบไม่ครบถ้วน
- องค์ประกอบแทรกซ้อน: ลายมือ ตัวสะกด มีอิทธิพลต่อคะแนนทั้งที่ไม่ใช่เกณฑ์

วิธีป้องกัน: ตรวจทีละข้อจนครบทุกคน, ปิดชื่อผู้สอบ, ยึดเฉลย/รูบริกเป็นประเด็น ๆ, แจ้งเกณฑ์ที่ใช้ล่วงหน้า

ตัวอย่างในชั้นเรียนไทย: ครูภาษาไทยที่ออกข้อสอบ “จงวิเคราะห์คุณค่าของวรรณคดีเรื่องหนึ่งที่อ่าน” ควรกำหนดขอบเขตให้ชัด เช่น ระบุให้วิเคราะห์ทั้งด้านวรรณศิลป์และด้านสังคม พร้อมแนบริบรูคสามประเด็น ดังตาราง 4.6 และตรวจโดยปิดชื่อผู้เขียน วิธีนี้ช่วยให้ผู้เรียนทราบทิศทางการตอบ และทำให้คะแนนสะท้อนคุณภาพคำตอบจริงมากกว่าความประทับใจส่วนตัวของครู

#### 4.5.8 ข้อผิดพลาดที่พบบ่อยและการตรวจสอบคุณภาพเบื้องต้นด้วยดัชนีความสอดคล้อง (IOC)

แม้ครูจะเขียนข้อสอบตามหลักการแล้ว ข้อบกพร่องบางอย่างก็ยังมีเล็ดลอดได้ ข้อผิดพลาดที่พบบ่อยในการเขียนข้อสอบ ได้แก่ คำถามสองนัยหรือกำกวม การใช้คำชี้้นำคำตอบ (Clues) ตัวลงที่ไม่สมเหตุผล ผลการใช้ “ถูกทุกข้อ/ผิดทุกข้อ” พร่ำเพรื่อ การใช้คำปฏิเสธซ้อนปฏิเสธ และการออกข้อสอบนอกเนื้อหาที่สอนซึ่งขัดต่อความยุติธรรม

(พัชรินทร์ ชมภูวิเศษ, 2559; Gierl, Bulut, Guo, & Zhang, 2017) แนวทางแก้ไขข้อบกพร่องเหล่านี้สรุปไว้ในตาราง

#### 4.7 ขั้นตอนการคำนวณและการแปลผลค่า IOC อย่างละเอียด ศึกษาในบทที่ 7 (หัวข้อ 7.5.7–7.5.8)

ตาราง 4.7 ข้อผิดพลาดที่พบบ่อยในการเขียนข้อสอบและแนวทางแก้ไข

| ข้อผิดพลาดที่พบบ่อย                      | แนวทางแก้ไขเชิงปฏิบัติ                                  |
|------------------------------------------|---------------------------------------------------------|
| คำถามสองนัย/กำกวม ดีความได้หลายทาง       | แยกให้เหลือประเด็นเดียวต่อข้อ<br>ระบุนิยาม/ขอบเขตให้ชัด |
| ใช้คำชี้้นำคำตอบ (Clues)                 | ตรวจไม่ให้คำใน Stem ไปปรากฏซ้ำในตัวถูก                  |
| ตัวลวงไม่สมเหตุสมผล ไม่มีคนเลือก         | สร้างตัวลวงจากความเข้าใจผิดจริงของผู้เรียน              |
| ใช้ “ถูกทุกข้อ/ผิดทุกข้อ” พร่ำเพรื่อ     | แทนด้วยตัวลวงที่วิเคราะห์จากเนื้อหา                     |
| คำปฏิเสธซ้อนปฏิเสธ ทำให้สับสน            | เปลี่ยนเป็นประโยคบอกเล่า หรือเน้นคำปฏิเสธให้เด่น        |
| ตัวเลือกยาวไม่เท่ากัน ตัวถูกเด่น         | ปรับทุกตัวเลือกให้ขนานกันทั้งความยาวและไวยากรณ์         |
| ออกข้อสอบนอกเนื้อหาที่สอน (ขัด Fairness) | ยึดตารางวิเคราะห์หลักสูตรเป็นกรอบออกข้อสอบ              |
| ความยากกระจุกที่ความจำ ไม่สมดุล          | กระจายข้อง่าย–ยากและระดับพฤติกรรมตาม<br>Blueprint       |

ที่มา: ดัดแปลงจาก พัชรินทร์ ชมภูวิเศษ (2559); Gierl และคณะ (2017)

เครื่องมือพื้นฐานที่ครูใช้ตรวจสอบคุณภาพข้อสอบก่อนนำไปใช้จริงคือ ดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ (Index of Item-Objective Congruence: IOC) ซึ่งพัฒนาโดย Rovinelli และ Hambleton (1977) วิธีการคือให้ผู้เชี่ยวชาญอย่างน้อย 3 ท่าน พิจารณาว่าข้อสอบแต่ละข้อวัดตรงตามจุดประสงค์หรือไม่ โดยให้คะแนน +1 (แน่ใจว่าวัดตรง) 0 (ไม่แน่ใจ) หรือ -1 (แน่ใจว่าวัดไม่ตรง) แล้วนำมาคำนวณตามสูตร ข้อที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไปจึงคัดไว้ใช้ ส่วนข้อที่ต่ำกว่า 0.50 ต้องปรับปรุงหรือตัดทิ้ง (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

ดังตัวอย่างการคำนวณในกล่องที่ 4.6

#### กล่องที่ 4.6 สูตรและตัวอย่างการคำนวณค่า IOC

สูตร:  $IOC = \sum R / N$

เมื่อ  $\sum R$  = ผลรวมคะแนนความเห็นของผู้เชี่ยวชาญ,  $N$  = จำนวนผู้เชี่ยวชาญ

ตัวอย่าง: ข้อสอบข้อหนึ่ง ผู้เชี่ยวชาญ 3 ท่านให้คะแนน +1, +1, 0

$IOC = (1+1+0) / 3 = 0.67 \rightarrow \geq 0.50$  จึง “คัดไว้ใช้”

อีกข้อหนึ่งได้ +1, 0, -1  $\rightarrow IOC = 0/3 = 0.00 \rightarrow < 0.50$  จึง “ปรับปรุงหรือตัดทิ้ง”

ข้อพึงระวัง: เดิม IOC ของ Rovinelli และ Hambleton ออกแบบสำหรับข้อสอบอิงเกณฑ์ที่ข้อหนึ่งวัดเพียงจุดประสงค์เดียว การนำไปใช้แบบหาค่าเฉลี่ยทั่วไปจึงต้องระวังเงื่อนไข (พิศิษฐ์ ตันทวนิช และ พนา จินดาศรี, 2561)

ตัวอย่างในชั้นเรียนไทย: ก่อนใช้ข้อสอบกลางภาค ครูคณิตศาสตร์ส่งร่างข้อสอบให้ครูในกลุ่มสาระ 3 ท่านประเมินความสอดคล้องกับตัวชี้วัด เมื่อพบว่าบางข้อมีค่า IOC ต่ำกว่า 0.50 เพราะถามนอกขอบเขตหรือกำกวม ครูจึงปรับแก้ก่อนนำไปใช้ การทำเช่นนี้คือการประกันคุณภาพเชิงเนื้อหาขั้นต้นที่ทำได้จริงในโรงเรียน และเป็นจุดเริ่มต้นก่อนการวิเคราะห์ค่าความยาก-อำนาจจำแนกหลังสอบ ซึ่งจะศึกษาต่อไปบทที่ว่าด้วยการวิเคราะห์ข้อสอบรายข้อ

#### กรณีศึกษา: การพัฒนาแบบสอบวินิจัยแบบเลือกตอบสามระดับ

กมลวรรณ รอดพันธ์ และณัฐภรณ์ หลาวทอง (2565) พัฒนาแบบสอบคู่ขนานวินิจัยโมดูลทัศน์ที่คลาดเคลื่อน วิชาวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น จำนวน 2 ฉบับ ฉบับละ 20 ข้อ โดยแต่ละข้อมีสามระดับ ได้แก่ ระดับคำตอบ ระดับเหตุผล และระดับความมั่นใจ

จุดเด่นของงานนี้คือผู้วิจัยเริ่มจากการสร้างผังข้อสอบและตารางลักษณะเฉพาะของข้อสอบก่อนลงมือเขียนข้อสอบ แล้วจึงตรวจสอบความตรงเชิงเนื้อหา ค่าความยาก อำนาจจำแนก และความเที่ยง กับนักเรียนตัวอย่าง 600 คน

บทเรียนสำหรับนักศึกษาคู: ข้อสอบที่ดีไม่ได้เริ่มจากการนั่งเขียนตามความเคยชิน แต่เริ่มจากผังข้อสอบ และตัวลงที่ออกแบบจากโมทัศน์ที่คลาดเคลื่อนจริงของผู้เรียน ซึ่งทำให้ข้อสอบเลือกตอบทำหน้าที่ "วินิจัย" ได้ ไม่ใช่เพียงให้คะแนน

คำถามชวนคิด: หากท่านจะพัฒนาแบบสอบวินิจัยในสาขาวิชาเอกของท่าน

จะเก็บข้อมูลโมทัศน์ที่คลาดเคลื่อนของผู้เรียนจากแหล่งใด และตัวลงของท่านจะสะท้อนความเข้าใจผิดได้บ้าง

## 4.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ขั้นก่อนเรียน (Pre-engagement)

- ทำแบบทดสอบก่อนเรียน 10 ข้อ และร่วมกิจกรรม “จับผิดข้อสอบ (Spot the Error)” จากตัวอย่างข้อสอบที่บกพร่อง

## ขั้นระหว่างเรียน (Active Workshop)

- ใบงานที่ 4.1 วิเคราะห์คุณภาพข้อสอบ: พิจารณาข้อสอบตัวอย่าง ระบุระดับ Bloom จุดอ่อน และเสนอแนวทางปรับปรุง (CLO2)
- ใบงานที่ 4.2 สร้างตารางวิเคราะห์หลักสูตร (TOS Workshop): กำหนดน้ำหนักเนื้อหา-พฤติกรรมและจำนวนข้อ (CLO3)
- ใบงานที่ 4.3 คลินิกการเขียนข้อสอบ (Item Writing Clinic): เขียนข้อสอบเลือกตอบและอัตนัยตาม Blueprint (CLO3)
- Peer Review: แลกเปลี่ยนชุดข้อสอบกับเพื่อน ประเมินด้วยแบบฟอร์ม IOC แล้วให้ข้อมูลย้อนกลับ

## ขั้นหลังเรียน (Reflection)

- เขียนบันทึกสะท้อนคิด (Exit Ticket) สรุปความเสี่ยงของการออกข้อสอบที่ไม่มีพิมพ์เขียว และส่งชิ้นงานชุดข้อสอบฉบับสมบูรณ์

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป

ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน

ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน

เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 4.7 สรุปบทเรียน

ข้อสอบวัดผลสัมฤทธิ์เป็นเครื่องมือทางวิทยาศาสตร์ที่กระตุ้นพฤติกรรมประติสุขให้ปรากฏและวัดได้ คุณภาพของข้อสอบขึ้นอยู่กับคุณลักษณะ 5 ประการ โดยมีความตรงเป็นคุณสมบัติสำคัญที่สุด การสร้างข้อสอบที่ดีต้องยึดความสอดคล้องเชิงโครงสร้าง เริ่มจากตารางวิเคราะห์หลักสูตร (Test Blueprint) เพื่อควบคุม น้ำหนักเนื้อหาและระดับพฤติกรรม แล้วเขียนข้อสอบเลือกตอบและอัตนัยตามหลักวิชา โดยส่วนน่าต้องสมบูรณ์ ตัวเลขต้องทำงาน และข้อสอบอัตนัยต้องมีรูปrikกำกับ ก่อนใช้จริงต้องตรวจสอบความตรงเชิงเนื้อหาด้วยค่า IOC ที่ตั้งแต่ 0.50 ขึ้นไป ทั้งหมดนี้เป็นหลักประกันความยุติธรรมและความน่าเชื่อถือของการประเมินผลในระบบการศึกษาที่เน้นผลลัพธ์ (OBE)

## 4.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึก

คะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่  
ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเผยแพร่พร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับ  
ผู้เรียนในสัปดาห์ถัดไป

## 4.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

### ตอนที่ 1 คำถามทบทวนเชิงทฤษฎี

1. อธิบายความแตกต่างระหว่างการวัดผล (Measurement) และการประเมินผล (Evaluation)  
พร้อมยกตัวอย่าง

2. เพราะเหตุใด “ความตรง” จึงเป็นคุณสมบัติที่สำคัญที่สุดของเครื่องมือวัด  
หากเครื่องมือมีความเที่ยงสูงแต่ความตรงต่ำจะเกิดผลอย่างไร

3. ตารางวิเคราะห์หลักสูตร (Test Blueprint)  
ช่วยสร้างความยุติธรรมและความครอบคลุมของข้อสอบได้อย่างไร

### ตอนที่ 2 ข้อสอบอัตนัยเชิงวิเคราะห์

1. วิเคราะห์บทบาทของตัวลวง (Distractor) ในการคัดแยกผู้เรียนกลุ่มเก่งและกลุ่มอ่อน
2. ออกแบบแนวทางป้องกันความลำเอียง (Bias) ในการตรวจข้อสอบอัตนัยอย่างน้อย 3 ข้อ
3. อธิบายความสัมพันธ์ของหลักความสอดคล้องเชิงโครงสร้าง (Constructive Alignment)  
กับการจัดการศึกษาแบบ OBE

### ตอนที่ 3 กิจกรรมปฏิบัติ (ชิ้นงานหลักของบท)

ให้นักศึกษาเลือกหน่วยการเรียนรู้ในสาขาวิชาเอกของตน 1 หน่วย แล้ว (1) สร้างตารางวิเคราะห์หลักสูตร (Test Blueprint) ที่มีน้ำหนักรวม 100% (2) เขียนข้อสอบเลือกตอบ 10 ข้อ และข้อสอบอัตนัย 1 ข้อ พร้อมระบุระดับ  
พฤติกรรมพุทธิพิสัยกำกับทุกข้อ และ (3) แนบเกณฑ์การให้คะแนน (Rubric) ของข้อสอบอัตนัย ส่งเป็นชิ้นงาน  
เพื่อประเมินตามรูปรีคในหัวข้อ 4.10

## 4.10 เครื่องมือวัดและประเมินผลประจำบท

---

การประเมินผลประจำบทผูกกับ CLO หลักของบทอย่างครบถ้วน โดยจับคู่ระดับพฤติกรรมกับเครื่องมือ  
ที่เหมาะสม และกำหนดสัดส่วนคะแนนรวม 100% ดังตาราง 4.8

ตาราง 4.8 แผนการวัดและประเมินผลประจำบท (ผูกกับ CLO)

| CLO  | สิ่งที่วัด                   | กิจกรรม/หลักฐาน                 | เครื่องมือวัด                  | สัดส่วน |
|------|------------------------------|---------------------------------|--------------------------------|---------|
| CLO2 | ความเข้าใจหลักการสร้างข้อสอบ | แบบทดสอบก่อน-หลังเรียน          | แบบทดสอบปรนัย 10 ข้อ (คู่ขนาน) | 20%     |
| CLO2 | การวิเคราะห์คุณภาพข้อสอบ     | ใบงานที่ 4.1<br>วิเคราะห์ข้อสอบ | rubric จำแนกข้อบกพร่อง         | 20%     |
| CLO3 | การออกแบบ Test Blueprint     | ใบงานที่ 4.2 / ชิ้นงาน          | rubric ประเมิน Blueprint       | 25%     |
| CLO3 | การเขียนข้อสอบที่มีคุณภาพ    | ใบงานที่ 4.3 / ชุดข้อสอบ        | rubric ชุดข้อสอบ + เกณฑ์ IOC   | 35%     |
| รวม  |                              |                                 |                                | 100%    |

rubric ประเมินชิ้นงาน “ตารางวิเคราะห์หลักสูตรและชุดข้อสอบ” (CLO3)

ตาราง 4.9 เกณฑ์การให้คะแนนแบบ rubric (4 = ดีมาก, 3 = ดี, 2 = พอใช้, 1 = ปรับปรุง)

| ประเด็นประเมิน               | ดีมาก (4)                                 | ดี (3)                       | พอใช้ (2)                            | ปรับปรุง (1)        |
|------------------------------|-------------------------------------------|------------------------------|--------------------------------------|---------------------|
| ความครบถ้วนของ Blueprint     | ครบทุกช่อง<br>น้ำหนักรวม 100%<br>สมเหตุผล | ครบแต่บางช่อง<br>ยังไม่สมดุล | ขาดบางช่อง/<br>น้ำหนักไม่รวม<br>100% | ไม่ปรากฏพิมพ์เขียว  |
| ความสอดคล้องกับระดับพฤติกรรม | ข้อสอบตรงระดับ Bloom ทุกข้อ               | ตรงเป็นส่วนใหญ่              | ตรงบางส่วน                           | ไม่ระบุ/ไม่ตรง      |
| คุณภาพการเขียนข้อสอบ         | Stem สมบูรณ์<br>ตัวลวงทำงานทุกข้อ         | บกพร่องเล็กน้อย 1-2 ข้อ      | บกพร่องหลายข้อ                       | บกพร่องเป็นส่วนใหญ่ |
| การตรวจสอบคุณภาพ (IOC)       | ทุกข้อมี IOC $\geq 0.50$<br>พร้อมหลักฐาน  | ส่วนใหญ่ผ่านเกณฑ์            | ผ่านบางส่วน                          | ไม่มีการตรวจสอบ     |

เกณฑ์ผ่าน: คะแนน rubric เฉลี่ยตั้งแต่ระดับ “ดี” (3.00) ขึ้นไป และเขียนข้อสอบผ่านเกณฑ์อย่างน้อย 8 ใน 10

ข้อ; แบบทดสอบหลังเรียนผ่านไม่น้อยกว่าร้อยละ 80

## 4.11 เอกสารอ้างอิงประจำบท

---

### ภาษาไทย

- กมลวรรณ รอดพันธ์ และณัฐภรณ์ หลาวทอง. (2565).  
การพัฒนาแบบสอบคู่ขนานวินิจัยมโนทัศน์ที่คลาดเคลื่อนแบบเลือกตอบสามระดับรายวิชาวิทยาศาสตร์  
สำหรับนักเรียนมัธยมศึกษาตอนต้น. วารสารการวัดผลการศึกษา, 39(105), 1–12.
- ชวาล แพรัตกุล. (2520). เทคนิคการเขียนข้อทดสอบ. โรงพิมพ์พิทักษ์อักษร.
- พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการศึกษา. คณะครุศาสตร์  
มหาวิทยาลัยราชภัฏอุดรธานี.
- พิชิต ฤทธิ์จัญญ. (2557). หลักการวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 8). แฮส ออฟ เคอร์มิสท์.
- พิศิษฐ ตัณทวนิช และ พนา จินดาศรี. (2561). ความหมายที่แท้จริงของค่า IOC.  
วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 24(2), 3–12.
- ไพศาล หวังพานิช. (2526). การวัดผลการศึกษา. ไทยวัฒนาพานิช.
- ล้วน สายยศ และ อังคณา สายยศ. (2543). เทคนิคการวัดผลการเรียนรู้ (พิมพ์ครั้งที่ 2). สุวีริยาสาส์น.
- สำเร็จ บุญเรืองรัตน์. (2534). การวัดและประเมินผลการศึกษา. สำนักพิมพ์พัฒนาศึกษา.
- สุนันท์ ศลโกสม. (2525). การวัดผลการศึกษา. โรงพิมพ์มหามกุฏราชวิทยาลัย.
- อนุวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่.  
สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

### ภาษาอังกฤษ

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). Open University Press.
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. Review of Educational Research, 87(6), 1082–1116. <https://doi.org/10.3102/0034654317726529>

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333. [https://doi.org/10.1207/S15324818AME1503\\_5](https://doi.org/10.1207/S15324818AME1503_5)

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). Pearson.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Pathamathamakul, C. (2022). Cognitive dissonance and teachers' adoption of outcome-based teaching. *Journal of Education and Innovative Learning*, 2(1), 15–34.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research*, 2(2), 49–60.

## บทที่ 5

### การวัดและประเมินทักษะปฏิบัติ

### Performance Assessment

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้

สัปดาห์ที่ 5 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

แผนบริหารการสอนประจำบทที่ 5

---

#### การวัดและประเมินทักษะปฏิบัติ (Performance Assessment) (จำนวน 4 คาบ)

##### หัวข้อสาระจัดการเรียนการสอน

- ความหมาย ความสำคัญ และธรรมชาติของการประเมินทักษะปฏิบัติ
- พฤติกรรมด้านทักษะพิสัย (Psychomotor Domain) และระดับของทักษะตามแนวคิดของ Simpson  
ลักษณะสำคัญของทักษะปฏิบัติและประเภทของการประเมิน (Performance / Authentic / Competency-based)
- วิธีการวัดและเครื่องมือประเมินทักษะปฏิบัติ (การสังเกต การสาธิต การปฏิบัติ สถานการณ์จำลอง แฟ้มสะสมงาน)
- การสร้างและพัฒนาเกณฑ์การให้คะแนนแบบรูบริก (Analytic / Holistic Rubric) แบบตรวจสอบรายการ และมาตรา  
ประมาณค่า
- การให้คะแนน การแปลผล การให้ข้อมูลย้อนกลับ และความลำเอียงที่ต้องระวัง

##### ความคิดรวบยอด

ทักษะปฏิบัติเป็นพฤติกรรมที่ต้องลงมือทำและสังเกตเห็นได้จริง การวัดด้วยข้อสอบข้อเขียนเพียงอย่างเดียวจึงไม่เพียงพอ ต้องใช้การประเมินทักษะปฏิบัติที่พิจารณาทั้ง “กระบวนการ” (Process) และ “ผลผลิต” (Product) ผ่านภาระงานที่สะท้อนสภาพจริง โดยมีเกณฑ์การให้คะแนน (Rubric) ที่ชัดเจนเชิงพฤติกรรมเป็นหัวใจของความเที่ยงตรงและความยุติธรรม เพื่อเปลี่ยนการวัดผลจากการตัดสิน (Judgment) ให้เป็นการพัฒนา (Development) ผู้เรียนอย่างแท้จริง

##### จุดประสงค์การเรียนรู้

1. อธิบายความหมาย ความสำคัญ และธรรมชาติของการประเมินทักษะปฏิบัติได้ถูกต้อง
2. วิเคราะห์ระดับพฤติกรรมทักษะพิสัยและจับคู่กับวิธีการวัดและเครื่องมือที่เหมาะสมได้ (CLO3)
3. ออกแบบภาระงานปฏิบัติและสร้างเครื่องมือประเมิน (รูบริก/แบบสังเกต) ที่เที่ยงตรงและเชื่อถือได้ (CLO3)
4. ปฏิบัติการประเมินการสอนในสถานการณ์จำลองและให้ข้อมูลย้อนกลับอย่างเป็นธรรมและเป็นระบบ (CLO4)

## เนื้อหาวิชา

ตามเอกสารประกอบการสอนบทที่ 5 หัวข้อ 5.5.1–5.5.10 (ความหมายและความสำคัญ · ทักษะพิสัย และระดับพฤติกรรม · ลักษณะของทักษะปฏิบัติ · ประเภทการประเมิน · วิธีการวัด · เครื่องมือประเมิน · การสร้างรูบรีค · การให้คะแนนและข้อมูลย้อนกลับ · ความลำเอียงและการลดความคลาดเคลื่อน · การประยุกต์ในวิชาชีพครู)

### กิจกรรมการเรียนรู้

1. ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน และวิเคราะห์คลิป/ตัวอย่างการประเมินที่บกพร่องเพื่อหาจุดอ่อนของเกณฑ์
2. ขั้นระหว่างเรียน: บรรยายเชิงโต้ตอบ + Workshop ออกแบบภาระงานและสร้างรูบรีค (ใบงาน 5.1) แบบจับคู่ Peer-coaching
3. ขั้นระหว่างเรียน: ฝึกใช้เครื่องมือประเมินการสอนจำลองของเพื่อน (Peer Assessment) และฝึกให้ข้อมูลย้อนกลับ (ใบงาน 5.2)
4. ขั้นหลังเรียน: สะท้อนคิด (Exit Ticket/Reflection) เรื่องการลดความลำเอียง และทำแบบทดสอบหลังเรียน

### สื่อการเรียนรู้

- เอกสารประกอบการสอนบทที่ 5 และสไลด์ประกอบการบรรยาย
- ตัวอย่างรูบรีค แบบตรวจสอบรายการ (Checklist) และแบบสังเกตการปฏิบัติ
- คลิปวิดีโอการสอนจำลอง/Micro-teaching สำหรับฝึกประเมิน
- ใบงาน 5.1 (ออกแบบภาระงาน + สร้างรูบรีค) และใบงาน 5.2 (ประเมินการสอนของเพื่อน)

### การวัดและประเมินผล

- แบบทดสอบก่อนเรียน–หลังเรียน (วัดความเข้าใจหลักการ — เชื่อมโยง CLO3)
- ชิ้นงานเครื่องมือประเมิน (ภาระงาน + รูบรีค) ตรวจสอบด้วยรูบรีคสำหรับตรวจรูบรีค (Meta-rubric) — CLO3
- การปฏิบัติประเมินการสอนจำลองและคุณภาพข้อมูลย้อนกลับ ประเมินด้วยแบบสังเกตการปฏิบัติ — CLO4

### 5.1 ความคิดรวบยอด

การวัดผล (Measurement) เป็นกระบวนการกำหนดตัวเลขหรือสัญลักษณ์แทนคุณลักษณะของผู้เรียนอย่างมีกฎเกณฑ์ เพื่อตอบคำถามว่า “มากน้อยเพียงใด (How much?)” ส่วนการประเมินผล (Evaluation) เป็นการตัดสินคุณค่าโดยเทียบข้อมูลกับเกณฑ์ เพื่อตอบคำถามว่า “ดีเพียงใด (How good?)” (สุทธิพงศ์ บุญผดุง,

2568) สำหรับ “ทักษะปฏิบัติ” ซึ่งเป็นพฤติกรรมด้านทักษะพิสัย (Psychomotor Domain) การวัดต้องสังเกต การลงมือทำจริง พิจารณาทั้งกระบวนการและผลผลิต และอาศัยเครื่องมือที่แม่นยำ เพราะพฤติกรรมด้านนี้มีความแปรผันสูงกว่าความรู้ (ขวาล แพรัตกุล, 2520)

#### แก่นความคิดของบท

- ประเมินทั้ง “กระบวนการ (Process)” และ “ผลผลิต (Product)” ผ่านภาระงานที่สะท้อนสภาพจริง
- rubric (Rubric) ที่ชัดเจนเชิงพฤติกรรม คือหัวใจของความเที่ยงตรงและความยุติธรรม
- เป้าหมายสูงสุดคือการประเมินเพื่อพัฒนา (Formative) มิใช่เพียงการตัดสินได้-ตก

## 5.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 5 จบแล้ว ผู้เรียนสามารถ

1. อธิบายความหมาย ความสำคัญ และธรรมชาติของการประเมินทักษะปฏิบัติได้ถูกต้องตามหลักวิชา
2. วิเคราะห์ระดับพฤติกรรมทักษะพิสัยและเลือกวิธีการวัด/เครื่องมือให้สอดคล้องกับลักษณะพฤติกรรมและสถานการณ์ได้
3. ออกแบบภาระงานปฏิบัติและสร้างเครื่องมือประเมิน (rubric แบบตรวจสอบรายการ แบบสังเกต) ที่มีความเที่ยงตรงและเชื่อถือได้
4. ปฏิบัติการประเมินการสอนในสถานการณ์จำลอง ให้คะแนน แผลผล และให้ข้อมูลย้อนกลับอย่างเป็นธรรมและเป็นระบบ

### ตารางความสอดคล้องจุดประสงค์-CLO-การวัด (Constructive Alignment)

| CLO ของบท                                                                 | ระดับพฤติกรรม                    | เนื้อหาหลักที่รองรับ | กิจกรรม                                                      | วิธีวัด + เครื่องมือ                                              |
|---------------------------------------------------------------------------|----------------------------------|----------------------|--------------------------------------------------------------|-------------------------------------------------------------------|
| CLO3<br>ออกแบบและสร้างเครื่องมือวัด/แบบประเมินที่เหมาะสมกับลักษณะพฤติกรรม | ประยุกต์-สร้างสรรค์ (ทักษะพิสัย) | 5.5.2-5.5.7          | ใบงาน 5.1<br>ออกแบบภาระงาน + สร้างrubric                     | ชิ้นงานเครื่องมือประเมิน + Meta-rubric;<br>แบบทดสอบก่อน-หลังเรียน |
| CLO4<br>ปฏิบัติการวัดและประเมินผลในสถานการณ์จำลอง/รายวิชา                 | ประยุกต์ (ทักษะพิสัย)            | 5.5.5-5.5.10         | ใบงาน 5.2<br>ประเมินการสอนจำลองของเพื่อน + ให้ข้อมูลย้อนกลับ | แบบสังเกตการปฏิบัติการประเมิน +<br>คุณภาพข้อมูลย้อนกลับ           |

### 5.3 คำสำคัญ

| คำสำคัญ (ไทย)                 | ภาษาอังกฤษ             | คำสำคัญ (ไทย)        | ภาษาอังกฤษ           |
|-------------------------------|------------------------|----------------------|----------------------|
| การประเมินทักษะปฏิบัติ        | Performance Assessment | ทักษะพิสัย           | Psychomotor Domain   |
| การประเมินตามสภาพจริง         | Authentic Assessment   | ภาระงานปฏิบัติ       | Performance Task     |
| เกณฑ์การให้คะแนนแบบ<br>รูบรีค | Scoring Rubric         | แบบตรวจสอบรายการ     | Checklist            |
| มาตรฐานค่า                    | Rating Scale           | ข้อมูลย้อนกลับ       | Feedback             |
| ความเที่ยงตรง                 | Validity               | ความเชื่อมั่น        | Reliability          |
| ความลำเอียงของผู้ประเมิน      | Rater Bias             | การประเมินเพื่อพัฒนา | Formative Assessment |

### 5.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอน ประกาศในชั้นเรียน

### 5.5 เนื้อหาวิชา

เนื้อหาในบทนี้แบ่งออกเป็น 10 หัวข้อย่อย เริ่มจากความหมายและความสำคัญของการประเมินทักษะ ปฏิบัติ ไล่เรียงสู่ธรรมชาติของพฤติกรรมทักษะพิสัย ประเภทและวิธีการประเมิน เครื่องมือและการสร้างเกณฑ์ การให้คะแนนแบบรูบรีค จนถึงการให้คะแนน การให้ข้อมูลย้อนกลับ การลดความลำเอียง และการประยุกต์ใช้ ในวิชาชีพครู โดยทุกหัวข้อมุ่งรองรับ จุดประสงค์ CLO3 (ออกแบบและสร้างเครื่องมือประเมินให้สอดคล้องกับ ลักษณะพฤติกรรม) และ CLO4 (ปฏิบัติการประเมินในสถานการณ์จำลองและให้ข้อมูลย้อนกลับอย่างเป็นธรรม)

#### 5.5.1 ความหมายและความสำคัญของการประเมินทักษะปฏิบัติ

นิยาม. การประเมินทักษะปฏิบัติ (Performance Assessment) หมายถึง วิธีการประเมินที่ครู มอบหมายงานหรือกิจกรรมให้ผู้เรียนลงมือปฏิบัติจริง แล้วสังเกตและตัดสินคุณภาพของพฤติกรรมจากทั้ง

“กระบวนการ” (process) และ “ผลงาน” (product) เพื่อให้ได้สารสนเทศว่าผู้เรียนเกิดการเรียนรู้มากน้อยเพียงใด (พร้อมพรรณ อุดมสิน, 2546 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559; บุญเชิด ภิญโญนนตพงษ์, 2545 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559) ตำราการวัดผลหลายเล่ม (พชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ย้ำไว้สอดคล้องกันว่าการประเมินปฏิบัติต้องวางอยู่บนเงื่อนไขและสถานการณ์ที่สอดคล้องกับสภาพจริง และเกณฑ์การประเมินต้องสร้างขึ้นจากประเด็นคุณลักษณะสำคัญของผลงานนั้น

องค์ประกอบและธรรมชาติ. การประเมินทักษะปฏิบัติมีองค์ประกอบหลัก 2 ส่วนเสมอ คือ (1) ตัวงานหรือภาระงาน (task) ที่กำหนดให้ผู้เรียนแสดงออก และ (2) เกณฑ์การให้คะแนน (scoring criteria) ที่ครูใช้ตัดสินคุณภาพของการปฏิบัติหรือผลงาน (ขวลิต ชูกำแพง, 2553 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559) นอกจากนี้การวัดทางการศึกษายังจัดเป็น “การวัดทางอ้อม” (indirect measurement) เพราะความสามารถภายในของผู้เรียนวัดโดยตรงไม่ได้ ต้องอนุมานจากพฤติกรรมที่สังเกตเห็น คะแนนที่ได้จึงมีความคลาดเคลื่อนของการวัด (measurement error) ผ่องอยู่เสมอ และเลข 0 เป็น “ศูนย์สัมพัทธ์” (relative zero) ไม่ได้แปลว่าผู้เรียนไม่มีความสามารถเลย (สุทธิพงศ์ บุญผดุง, 2568)

ความสำคัญในการผลิตครู. ศาสตราจารย์ ขวาล แพร่ตกุล (2520) ชี้ว่าการทดสอบสัมฤทธิ์ผลแบบข้อเขียน (paper-and-pencil test) มุ่งวัด “ความสำเร็จในอดีต” ของผู้เรียนเป็นหลัก แต่วิชาชีพครูตัดสินกันที่ “การกระทำ” (doing) ในชั้นเรียนจริง การสอบข้อเขียนบอกได้เพียงว่าใคร “รู้” หลักการสอน แต่มีเพียงการประเมินปฏิบัติเท่านั้นที่ยืนยันได้ว่าใคร “สอนเป็น” เช่น นักศึกษาอาจทำข้อสอบเรื่องการตั้งคำถามกระตุ้นคิดได้คะแนนเต็ม แต่เมื่อสอนจริงกลับใช้คำถามปลายปิดที่ตอบเพียง ‘ใช่/ไม่ใช่’ ตลอดคาบ ช่องว่างระหว่าง ‘รู้’ กับ ‘ทำได้’ เช่นนี้ตรวจพบได้ด้วยการประเมินปฏิบัติเท่านั้น

การประยุกต์และข้อควรระวัง. ในการผลิตครู การประเมินปฏิบัติคือกลไกประกันคุณภาพว่าบัณฑิตมีสมรรถนะการสอนจริงตามมาตรฐานวิชาชีพครูของคุรุสภา การละเลยการประเมินส่วนนี้จึงเท่ากับปล่อยให้ผู้ขาดสมรรถนะเข้าสู่ระบบการศึกษา อย่างไรก็ตาม เพราะเป็นการวัดทางอ้อมที่อาศัยดุลยพินิจของผู้ประเมินค่อนข้างสูง ครูจึงต้องลดความคลาดเคลื่อนด้วยเกณฑ์ที่ชัดเจนและการอบรมผู้ประเมิน (ดูหัวข้อ 5.5.9) ความแตกต่างเชิงหลักการระหว่างการทดสอบข้อเขียนกับการประเมินปฏิบัติสรุปได้ดังตารางที่ 5.1

ตารางที่ 5.1 ความแตกต่างระหว่างการทดสอบข้อเขียนกับการประเมินทักษะปฏิบัติ

| ประเด็นเปรียบเทียบ | การทดสอบข้อเขียน (Paper-and-Pencil)       | การประเมินทักษะปฏิบัติ (Performance)       |
|--------------------|-------------------------------------------|--------------------------------------------|
| สิ่งที่มุ่งวัด     | ความรู้/ความเข้าใจที่ระลึกหรือเลือกตอบได้ | ความสามารถลงมือทำทั้งกระบวนการและผลงาน     |
| ลักษณะหลักฐาน      | คำตอบบนกระดาษ (“รู้”)                     | พฤติกรรมที่สังเกตเห็นขณะปฏิบัติ (“ทำเป็น”) |

|                 |                         |                                                  |
|-----------------|-------------------------|--------------------------------------------------|
| เครื่องมือหลัก  | ข้อสอบปรนัย/อัตนัย      | แบบสังเกต แบบตรวจสอบรายการ<br>มาตรฐานค่า รูบริค  |
| จุดเน้นด้านเวลา | มักวัดผลสำเร็จในอดีต    | วัดสมรรถนะที่ใช้ได้จริงในอนาคต/สภาพจริง          |
| ความเป็นปรนัย   | สูง ตรวจสอบให้คะแนนง่าย | ต่ำกว่า ต้องอาศัยเกณฑ์ชัดเจน +<br>อบรมผู้ประเมิน |

ที่มา: ดัดแปลงจาก ขวาล แพร่ตฤกล (2520) และ พัทรินทร์ ชมภูวิเศษ (2559, น. 155–157)

### 5.5.2 ทักษะพิสัยและระดับพฤติกรรม (Psychomotor Domain)

ทักษะพิสัย (Psychomotor Domain) คือ พฤติกรรมด้านกล้ามเนื้อ-ประสาท (neuromuscular) ที่บ่งถึงความสามารถในการปฏิบัติงานได้อย่างคล่องแคล่วชำนาญ โดยมี “เวลา” และ “คุณภาพของงาน” เป็นตัวชี้ระดับของทักษะ (Bloom et al., 1956 อ้างถึงใน พัทรินทร์ ชมภูวิเศษ, 2559; พิชิต ฤทธิ์จรูญ, 2557 อ้างถึงใน พัทรินทร์ ชมภูวิเศษ, 2559) พฤติกรรมด้านนี้ต้องอาศัยการประสานสัมพันธ์ระหว่างประสาทสัมผัสกับการเคลื่อนไหว เช่น การใช้กล้องจุลทรรศน์ การเขียนกระดาน หรือการสาธิตการทดลอง

การจำแนกระดับ แบบจำแนกระดับทักษะพิสัยที่ละเอียดและใช้อ้างอิงมากที่สุดคือของ Simpson (1972) ซึ่งแบ่งเป็น 7 ระดับเรียงจากง่ายไปยาก ตั้งแต่การรับรู้ (perception) ผ่านการเตรียมพร้อม การตอบสนองตามแนวที่กำหนด การปรับระบบกลไก การตอบสนองที่ซับซ้อน การดัดแปลง ไปจนถึงการริเริ่ม (origination) ดังตารางที่ 5.2 ครูสามารถใช้ระดับเหล่านี้เป็น “บันได” เพื่อกำหนดจุดประสงค์และเลือกวิธีวัดให้สอดคล้องกับระดับพฤติกรรมที่คาดหวัง

ตารางที่ 5.2 ระดับพฤติกรรมทักษะพิสัยตามแนวคิดของ Simpson (1972)

| ระดับ (Simpson, 1972)                            | ความหมายโดยย่อ                        | ตัวอย่างพฤติกรรมของนักศึกษาครู          |
|--------------------------------------------------|---------------------------------------|-----------------------------------------|
| 1. การรับรู้ (Perception)                        | ใช้ประสาทสัมผัสนำทางการเคลื่อนไหว     | สังเกตวิธีครูที่เลี้ยงใช้สื่อ           |
| 2. การเตรียมพร้อม (Set)                          | ความพร้อมกาย-ใจ-เจตคติก่อนลงมือ       | จัดอุปกรณ์และเตรียมแผนก่อนสอน           |
| 3. การตอบสนองตามแนวที่กำหนด (Guided Response)    | ทำตามแบบ/คำชี้แนะลองฝึกดลองถูก        | สอนตามขั้นที่ครูสาธิตโดยมีพี่เลี้ยงช่วย |
| 4. การปรับระบบกลไก (Mechanism)                   | ทำได้คล่อง เป็นนิสัยความมั่นใจปานกลาง | ดำเนินกิจกรรมตามแผนได้ลื่นไหล           |
| 5. การตอบสนองที่ซับซ้อน (Complex Overt Response) | ทำงานซับซ้อนได้แม่นยำ รวดเร็ว         | คุมชั้นเรียนพร้อมใช้คำถามกระตุ้นคิด     |

|                             |                           |                                    |
|-----------------------------|---------------------------|------------------------------------|
| 6. การดัดแปลง (Adaptation)  | ปรับทักษะรับสถานการณ์ใหม่ | แก้ปัญหาเฉพาะหน้าเมื่อสื่อขัดข้อง  |
| 7. การริเริ่ม (Origination) | สร้างรูปแบบ/วิธีการใหม่   | ออกแบบกิจกรรมการสอนรูปแบบใหม่ของตน |

ที่มา: ดัดแปลงจาก Simpson (1972); เทียบเคียง พัทธินทร์ ชมภูวิเศษ (2559, น. 155–156)

ตัวอย่างในชั้นเรียนไทย. ในการฝึกประสบการณ์วิชาชีพครู สัปดาห์แรกนักศึกษา มักอยู่ชั้น ‘การรับรู้-เตรียมพร้อม’ คือสังเกตครูที่เลี้ยงและจัดเตรียมแผน เมื่อสอนตามขั้นที่ครูสาธิตได้คือขั้น ‘ตอบสนองตามแนวที่กำหนด’ และเมื่อดำเนินกิจกรรมได้สัปดาห์จนเป็นอัตโนมัติคือขั้น ‘ปรับระบบกลไก’ ครั้นวันที่โปรเจกเตอร์ดับกลางคาบแล้วยังปรับไปใช้กระดานและกิจกรรมกลุ่มแทนได้ทันที นั่นคือขั้น ‘การดัดแปลง’ ซึ่งเป็นทักษะระดับสูงที่บ่งบอกความเป็นมืออาชีพ

**ข้อควรระวัง** ตำราไทยบางเล่มสรุปทักษะพิสัยเป็น 5 ขั้นอย่างย่อ ครูจึงควรเลือกใช้แบบจำแนกให้เหมาะสมกับความซับซ้อนของงาน และไม่จำเป็นต้องประเมินครบทั้ง 7 ระดับในทุกภาระงาน ดังเปรียบเทียบในกล่องที่ 5.1

#### กล่องที่ 5.1 เทียบแบบจำแนกทักษะพิสัย: Simpson (7 ระดับ) กับแบบย่อ 5 ขั้น

Simpson (1972) 7 ระดับ: รับรู้ → เตรียมพร้อม → ตอบสนองตามแนวที่กำหนด → ปรับระบบกลไก → ตอบสนองซับซ้อน → ดัดแปลง → ริเริ่ม

แบบย่อ 5 ขั้น (อ้างใน พัทธินทร์, 2559): การรับรู้/เลียนแบบ → กระทำตามแบบ → หาความถูกต้อง → กระทำต่อเนื่องจนชำนาญ → กระทำเป็นธรรมชาติอัตโนมัติ

ข้อสังเกต: ทั้งสองแบบมีตรรกะเดียวกัน คือไล่จาก ‘ทำตามแบบ’ สู่ ‘ทำเองอย่างอัตโนมัติและสร้างสรรค์’ เลือกความละเอียดให้เหมาะกับงาน

ที่มา: สังเคราะห์จาก Simpson (1972) และ พัทธินทร์ ชมภูวิเศษ (2559, น. 40, 155–156)

### 5.5.3 ลักษณะสำคัญของทักษะปฏิบัติและการสังเคราะห์กับระดับพฤติกรรม

คุณภาพของการปฏิบัติงานพิจารณาได้จากองค์ประกอบสำคัญ 5 ประการ ซึ่งสัมพันธ์กับระดับทักษะพิสัยของ Simpson (1972) และช่วยให้ครูออกแบบเกณฑ์ประเมินได้ครอบคลุม ได้แก่

- **ขั้นตอนการทำงาน (Sequence):** ปฏิบัติตามลำดับขั้นที่ถูกต้องเหมาะสม ไม่ข้ามหรือสลับขั้นจนเสียคุณภาพงาน
- **การใช้เครื่องมือ (Tool Manipulating):** เลือกและใช้อุปกรณ์ได้คล่องแคล่ว แม่นยำ และเหมาะกับงาน
- **การแก้ปัญหา (Problem Solving):** ตัดสินใจแก้ไขได้เมื่อเผชิญสถานการณ์ที่ไม่คาดคิดขณะปฏิบัติ
- **ความชำนาญ (Proficiency):** ความรวดเร็วที่มาพร้อมความถูกต้อง สะท้อนการฝึกฝนจนเป็นอัตโนมัติ
- **ความปลอดภัย (Safety):** ตระหนักถึงอันตรายและระเบียบปฏิบัติ โดยเฉพาะความปลอดภัยของผู้เรียน

**คำอธิบายเชิงลึก.** ลักษณะทั้ง 5 ไม่ได้อยู่อย่างแยกส่วน แต่สอดคล้องกับระดับพัฒนาการของทักษะ กล่าวคือ การใช้เครื่องมือเริ่มปรากฏตั้งแต่ขั้นต้น ส่วนการแก้ปัญหาเป็นสัญญาณของทักษะระดับสูง (การดัดแปลง) การมอง

ลักษณะเหล่านี้คู่กับระดับ Simpson ช่วยให้ครูกำหนด ‘ตัวบ่งชี้ที่สังเกตได้’ (observable indicators) ในเกณฑ์ประเมินได้ตรงจุด ดังตารางที่ 5.3

ตารางที่ 5.3 การสังเคราะห์ลักษณะสำคัญ 5 ประการเข้ากับระดับทักษะพิสัยของ Simpson

| ลักษณะสำคัญของการปฏิบัติ | ระดับทักษะพิสัยที่สอดคล้อง              | สิ่งที่ครูสังเกต (ตัวบ่งชี้)                     |
|--------------------------|-----------------------------------------|--------------------------------------------------|
| การใช้เครื่องมือ         | การรับรู้-เตรียมพร้อม (Perception-Set)  | เลือกและจับอุปกรณ์ถูกวิธี ถูกท่าทาง              |
| ขั้นตอนการทำงาน          | ตอบสนองตามแนวที่กำหนด (Guided Response) | ทำตามลำดับขั้นที่สาธิตได้ถูกต้องครบถ้วน          |
| ความชำนาญ                | ปรับระบบกลไก (Mechanism)                | ทำได้คล่อง ต่อเนื่อง<br>เกือบเป็นอัตโนมัติ       |
| การแก้ปัญหา              | การดัดแปลง (Adaptation)                 | ปรับวิธีรับมือสถานการณ์ที่ไม่คาดคิดได้           |
| ความปลอดภัย              | แทรกอยู่ในทุกระดับ                      | ตระหนักอันตราย ปฏิบัติตามระเบียบ<br>ดูแลผู้เรียน |

ที่มา: ผู้เขียน สังเคราะห์จาก Simpson (1972) และ พัทรินทร์ ชมภูวิเศษ (2559, น. 155-157)

**ตัวอย่างในชั้นเรียนไทย.** ในการประเมินทักษะ ‘การทดลองวิทยาศาสตร์’ ของนักเรียน ครูสามารถแยกสังเกตได้ว่านักเรียนใช้เครื่องมือถูกวิธีหรือไม่ (การใช้เครื่องมือ) ทำตามขั้นตอนการทดลองครบหรือไม่ (ขั้นตอน) และเมื่ออุปกรณ์ขัดข้องสามารถปรับแก้ได้หรือไม่ (การแก้ปัญหา) การแยกประเด็นเช่นนี้ทำให้ข้อมูลย้อนกลับตรงจุดและพัฒนาได้จริง

**การประยุกต์และข้อควรระวัง.** ในวิชาชีพครู ‘ความปลอดภัย’ ของนักเรียนคือเกณฑ์ตัดสินความเป็นมืออาชีพที่ห้ามละเลยในทุกกิจกรรม ครูควรกำหนดน้ำหนักคะแนนให้ลักษณะที่เป็นหัวใจของงานนั้นมากกว่าด้านอื่น ไม่ใช่ใช้ทุกด้านเท่ากันเสมอไป เช่น งานทดลองเคมีควรให้น้ำหนักความปลอดภัยและขั้นตอนสูงกว่าความสวยงามของรายงาน

#### 5.5.4 ประเภทของการประเมินทักษะปฏิบัติ

นิยาม. คำที่มักใช้คู่กันสามคำในการประเมินทักษะปฏิบัติ ได้แก่ การประเมินการปฏิบัติ (Performance Assessment) ซึ่งเน้นกระบวนการและผลผลิตของภาระงานที่ครูกำหนด การประเมินตามสภาพจริง

(Authentic Assessment) ซึ่งเน้นภาระงานที่มีความหมายและสอดคล้องกับชีวิตจริง และการประเมินอิงสมรรถนะ (Competency-Based Assessment) ซึ่งมุ่งยืนยันว่าผู้เรียนถึงเกณฑ์ขั้นต่ำตามมาตรฐานวิชาชีพหรือไม่ (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

ความสัมพันธ์และความต่าง. ตำราการวัดผลหลายเล่ม (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ชี้ไว้สอดคล้องกันว่าการประเมินการปฏิบัติงานเป็นการประเมินพฤติกรรมที่ครูเป็นผู้ออกแบบหรือกำหนด ส่วนการประเมินตามสภาพจริงคำนึงถึงบริบทที่เป็นไปตามชีวิตประจำวันและให้ผู้เรียนมีส่วนตัดสินใจพัฒนาผลงานของตนตั้งแต่ต้น ดังนั้น “การประเมินการปฏิบัติไม่ใช่ทุกอย่างที่จะเป็นการประเมินตามสภาพจริง” การเปรียบเทียบทั้งสามรูปแบบสรุปดังตารางที่ 5.4

ตารางที่ 5.4 เปรียบเทียบการประเมินการปฏิบัติ การประเมินตามสภาพจริง และการประเมินอิงสมรรถนะ

| รูปแบบ                 | จุดเน้น/เป้าหมาย                                 | เครื่องมือหลัก                         |
|------------------------|--------------------------------------------------|----------------------------------------|
| Performance Assessment | กระบวนการ + ผลผลิตของภาระงานที่กำหนด             | Checklist, Rating Scale, Rubric        |
| Authentic Assessment   | ภาระงานที่มีความหมายและสอดคล้องกับชีวิตจริง      | Portfolio, Simulation, Real-world Task |
| Competency-Based       | ยืนยันว่าถึงเกณฑ์ขั้นต่ำตามมาตรฐานวิชาชีพหรือไม่ | Criteria-referenced + Rubric           |

ที่มา: ดัดแปลงจาก พัชรินทร์ ชมภูวิเศษ (2559, น. 164–167)

ระดับความเป็นจริงของการวัด. วิธีวัดการปฏิบัติเรียงตัวเป็น “เส้นความต่อเนื่อง” (continuum) จากความเป็นจริงต่ำไปสูง ดังภาพที่ 5.1 ยิ่งใกล้สภาพจริงมากเท่าใด ข้อมูลก็ยิ่งสะท้อนสมรรถนะที่แท้จริง แต่ก็ควบคุมยากและสิ้นเปลืองมากขึ้น ครูจึงควรไล่ระดับการฝึก (scaffolding) จากสถานการณ์จำลองที่ปลอดภัยไปสู่สภาพจริง

ภาพที่ 5.1 เส้นความต่อเนื่องของการวัดการปฏิบัติ (จากความเป็นจริงต่ำไปสูง)

| ระดับความเป็นจริง | รูปแบบการวัดการปฏิบัติ                     | ตัวอย่าง                                       |
|-------------------|--------------------------------------------|------------------------------------------------|
| ต่ำ               | 1) การทดสอบด้วยแบบทดสอบ (Paper-and-Pencil) | ตอบคำถาม/วางแผนการปฏิบัติบนกระดาษ              |
|                   | 2) การทดสอบเฉพาะส่วน (Identification)      | ระบุ/เลือกเครื่องมือที่ถูกต้องสำหรับงาน        |
| ปานกลาง           | 3) สถานการณ์จำลอง (Simulated)              | สอนจำลอง (micro-teaching), ฝึกปฐมพยาบาลกับหุ่น |
|                   | 4) ตัวอย่างงาน (Work Sample)               | ฝึกขับรถโดยมีครูฝึกนั่งประกบ                   |
| สูง               | 5) สถานการณ์จริง (Authentic/Real)          | สอนจริงในชั้นเรียนช่วงฝึกประสบการณ์            |

|  |  |         |
|--|--|---------|
|  |  | วิชาชีพ |
|--|--|---------|

ที่มา: ดัดแปลงจาก ส.วสสนา ประมวลพฤกษ์ (2544, น. 83) อ้างถึงใน พชรินทร์ ชมภูวิเศษ (2559, น. 158–159)

ตัวอย่างและข้อควรระวัง. ในการฝึกหัดครู นักศึกษาควรได้ทดลองเขียนแผนและวิจารณ์คลิปการสอน (ระดับต่ำ) ฝึกสอนจำลองหน้าชั้น (ระดับปานกลาง) ก่อนออกฝึกประสบการณ์ในโรงเรียนจริง (ระดับสูง) การข้ามขั้นไปสู่สภาพจริงทันทีโดยไม่ฝึกจำลองก่อน เสี่ยงต่อทั้งคุณภาพการสอนและความเชื่อมั่นของผลประเมิน เพราะสถานการณ์จริงเปลี่ยนแปลงทุกครั้งที่ทำให้ผลที่วัดอาจไม่ใช่ความสามารถที่แท้จริงของผู้เรียน

### 5.5.5 วิธีการวัดและประเมินทักษะปฏิบัติ

นิยามและภาพรวม. การเลือกวิธีวัดต้องสอดคล้องกับประเภทของทักษะและบริบท วิธีหลักที่ใช้กัน ได้แก่ การสังเกต (observation) การสาธิต (demonstration) การปฏิบัติงานจริง (performance) สถานการณ์จำลอง (simulation) และแฟ้มสะสมงาน (portfolio) (สมนึก ภัททิยธนี, 2558 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559; Nitko & Brookhart, 2019) แต่ละวิธีมีข้อดี-ข้อจำกัดต่างกัน ดังตารางที่ 5.5

ตารางที่ 5.5 วิธีการวัดทักษะปฏิบัติ: ข้อดีและข้อจำกัด

| วิธีการ                            | ข้อดี                                        | ข้อจำกัด                                   |
|------------------------------------|----------------------------------------------|--------------------------------------------|
| การสังเกต<br>(Observation)         | เห็นพฤติกรรมจริงเป็นข้อมูลปฐมภูมิ            | เสี่ยงต่ออคติของผู้สังเกต/ผู้เรียนสร้างภาพ |
| การสาธิต<br>(Demonstration)        | วัดทักษะเฉพาะด้านได้ตรงจุด                   | ผู้เรียนอาจตื่นเต้น<br>ขาดความเป็นธรรมชาติ |
| การปฏิบัติงานจริง<br>(Performance) | เห็นทั้งกระบวนการและผลผลิต                   | ใช้เวลาและงบประมาณสูง                      |
| สถานการณ์จำลอง<br>(Simulation)     | ปลอดภัยสำหรับทักษะที่เสี่ยง เช่น การสอนจำลอง | ขาดบรรยากาศจริงที่ซับซ้อนทางอารมณ์         |
| แฟ้มสะสมงาน<br>(Portfolio)         | เห็นพัฒนาการความงอกงามต่อเนื่อง              | ใช้เวลาประเมินมากและจัดการข้อมูลยาก        |

ที่มา: ดัดแปลงจาก สมนึก ภัททิยธนี (2558) อ้างถึงใน พชรินทร์ ชมภูวิเศษ (2559, น. 162–166)

**คำอธิบายเชิงลึก.** สมนึก ภัททิยธนี (2558 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559) จำแนกการทดสอบภาคปฏิบัติได้หลายมิติ ทั้งตามปัจจัยที่ประเมิน (กระบวนการ/ผลผลิต) ตามลักษณะสถานการณ์ (จำลอง/จริง) และตามการเกิดสิ่งเร้า (ธรรมชาติ/จัดขึ้น) การจัดสิ่งเร้าขึ้นช่วยลดเวลาการรอคอย เพราะไม่ต้องรอให้พฤติกรรมเกิดเองตามธรรมชาติ ทั้งนี้การประเมินแต่ละครั้งอาจวัดเฉพาะกระบวนการ เฉพาะผลผลิต หรือทั้งสองพร้อมกันก็ได้ ขึ้นกับวัตถุประสงค์เป็นหลัก

**ตัวอย่างในชั้นเรียนไทย.** การวัดทักษะ ‘การควบคุมชั้นเรียน’ ของนักศึกษาฝึกสอนควรใช้การสังเกต การสอนจริง/จำลองเป็นหลัก ไม่ใช่แบบสอบถามหรือการสัมภาษณ์ ส่วนทักษะ ‘การใช้กล้องจุลทรรศน์’ กับการสังเกตขณะปฏิบัติพร้อมแบบตรวจสอบรายการขั้นตอน

**การประยุกต์และข้อควรระวัง.** ในการวัดภาคปฏิบัติใด ๆ มักต้องใช้เครื่องมือมากกว่าหนึ่งอย่างเพื่อความครอบคลุมในประเด็นที่ต้องการวัด (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) และควรใช้การสังเกตควบคู่กับการบันทึกผลลงแบบประเมินที่สร้างไว้ทันที เพื่อป้องกันการลืมและลดอคติจากการประเมินย้อนหลัง

### 5.5.6 เครื่องมือประเมินทักษะปฏิบัติ

**นิยาม.** เครื่องมือบันทึกผลการสังเกตการปฏิบัติมี 3 กลุ่มหลัก เลือกใช้ตามความละเอียดของข้อมูลที่ต้องการ (สำเร็จ บุญเรืองรัตน์, 2534; สุทธิพงศ์ บุญผดุง, 2568) ได้แก่ แบบตรวจสอบรายการ (checklist) มาตรฐานค่า (rating scale) และรูบริก (rubric) ซึ่งให้สารสนเทศจากหายไปละเอียดตามลำดับ ดังตารางที่ 5.6

ตารางที่ 5.6 เปรียบเทียบเครื่องมือประเมินทักษะปฏิบัติ 3 กลุ่ม

| เครื่องมือ                   | บันทึกอะไร                               | เหมาะกับ                                   | ข้อจำกัด                        |
|------------------------------|------------------------------------------|--------------------------------------------|---------------------------------|
| แบบตรวจสอบรายการ (Checklist) | มี/ไม่มี หรือ ทำ/ไม่ทำ ในแต่ละขั้น       | ทักษะที่มีขั้นตอนชัดเจน ตัดสินผ่าน/ไม่ผ่าน | ไม่บอกระดับคุณภาพของการทำ       |
| มาตรฐานค่า (Rating Scale)    | ระดับคุณภาพ เช่น 1–5                     | งานที่ต้องการความละเอียดของระดับ           | อาจลำเอียงหากไม่มีคำบรรยายระดับ |
| รูบริก (Rubric)              | ระดับคุณภาพ + คำบรรยายพฤติกรรมแต่ละระดับ | งานซับซ้อน ต้องการ feedback เพื่อพัฒนา     | สร้างยาก ใช้เวลาพัฒนามาก        |

ที่มา: ดัดแปลงจาก สำเร็จ บุญเรืองรัตน์ (2534); สุทธิพงศ์ บุญผดุง (2568); พัทธรินทร์ ชมภูวิเศษ (2559)

**คำอธิบายเชิงลึก.** เครื่องมือทั้งสามมักใช้คู่กับ “การสังเกต” ซึ่งสุทธิพงศ์ บุญผดุง (2568) แบ่งเป็นการสังเกตแบบมีส่วนร่วมและไม่มีส่วนร่วม และแบบมีโครงสร้างกับไม่มีโครงสร้าง สำหรับห้องเรียนควรใช้การสังเกตแบบมีโครงสร้าง คือกำหนดรายการพฤติกรรมเป้าหมายไว้ล่วงหน้า เพื่อให้การจดบันทึกตรงประเด็นและได้ข้อมูลปฐมภูมิที่น่าเชื่อถือ ตัวอย่างเครื่องมือจริงสองชนิดแสดงในกล่องที่ 5.2 และ 5-3

#### กล่องที่ 5.2 ตัวอย่างแบบตรวจสอบรายการ (Checklist) — ทักษะการจัดการเรียนรู้

- ( ) 1. เตรียมสื่อและอุปกรณ์ครบตามแผนการสอน
- ( ) 2. ดำเนินกิจกรรมตามลำดับขั้นในแผน
- ( ) 3. ใช้คำถามกระตุ้นการคิดของผู้เรียน

( ) 4. สรุปบทเรียนและเชื่อมโยงสู่การประยุกต์ใช้

**วิธีอ่าน:** ทำเครื่องหมายเมื่อพบพฤติกรรม เหมาะกับการตัดสิน ‘ทำ/ไม่ทำ’ แต่ไม่บอกว่าทำได้ ‘ดีเพียงใด’

ที่มา: ดัดแปลงจาก สุทธิพงศ์ บุญผดุง (2568) และ พชรินทร์ ชมภูวิเศษ (2559)

### กล่องที่ 5.3 ตัวอย่างมาตรฐานค่า (Rating Scale) 5 ระดับ

เกณฑ์: 5 = ดีเยี่ยม 4 = ดี 3 = ปานกลาง 2 = พอใช้ 1 = ปรับปรุง

- การใช้น้ำเสียงและการสื่อสารกับผู้เรียน 5 4 3 2 1
- การควบคุมลำดับกิจกรรมตามแผน 5 4 3 2 1
- บุคลิกภาพและความเหมาะสมกับวิชาชีพครู 5 4 3 2 1

**วิธีอ่าน:** วงระดับที่ตรงกับคุณภาพที่สังเกตเห็น ละเอียดยกกว่า checklist

แต่ควรมีคำบรรยายระดับกำกับเพื่อกำหนดความลำเอียง

ที่มา: ดัดแปลงจาก สำเริง บุญเรืองรัตน์ (2534); สุทธิพงศ์ บุญผดุง (2568)

การประยุกต์และข้อควรระวัง. ครูควรเลือกเครื่องมือให้ตรงลักษณะพฤติกรรม เช่น ใช้ checklist กับ ‘ความปลอดภัย/ขั้นตอน’ ที่ตัดสินทำ/ไม่ทำ และใช้ rating scale หรือ rubric กับ ‘คุณภาพ’ ที่มีหลายระดับ การใช้มาตรฐานค่าโดยไม่มีคำบรรยายระดับเป็นสาเหตุสำคัญของความลำเอียง จึงควรยกระดับเป็นรูปรีดเมื่อต้องการความเที่ยงและการให้ข้อมูลย้อนกลับที่ชัดเจน (ดูหัวข้อ 5.5.7)

### 5.5.7 การสร้างและพัฒนาเกณฑ์การให้คะแนนแบบรูปรีด (Rubric)

**นิยาม.** คำว่า “รูปรีด” (Rubric) มีรากศัพท์หมายถึง “กฎ” หรือ “กติกา” (rule) หมายถึง เกณฑ์การให้คะแนน (scoring rubric) ที่เกิดจากการรวมกันระหว่างเกณฑ์การให้คะแนน (scoring criteria) กับมาตรฐานค่าที่เป็นระดับคะแนน (rating scale) โดยมีคำบรรยายระดับคุณภาพหรือ “คำบรรยายพฤติกรรม” (descriptor) กำกับแต่ละระดับอย่างชัดเจน (McMillan, 2001 อ้างถึงใน พชรินทร์ ชมภูวิเศษ, 2559; Brookhart, 2013) พิชิต ฤทธิ์จรูญ (2562) อธิบายว่ารูปรีดคือเครื่องมือที่ช่วยให้ครูประเมินผลลัพธ์การเรียนรู้ได้อย่างเที่ยงตรงและยุติธรรม เพราะผู้ประเมินต่างคนใช้เกณฑ์เดียวกันแล้วได้ผลสอดคล้องกัน หลักการสร้างและพัฒนารูปรีดฉบับเต็ม ศึกษาในบทที่ 10 (หัวข้อ 10.5.4–10.5.5)

**ประเภทของรูปรีด.** คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย (2567) และพิชิต ฤทธิ์จรูญ (2562) จำแนกรูปรีดออกเป็นหลายประเภท ที่ใช้บ่อยสรุปได้ดังตารางที่ 5.7 โดยรูปรีดแบบแยกองค์ประกอบ (analytic) ให้ข้อมูลละเอียด เหมาะกับการพัฒนา ส่วนแบบองค์รวม (holistic) ตัดสินเร็วเหมาะกับการสรุปผล

ตารางที่ 5.7 ประเภทของเกณฑ์การให้คะแนนแบบรูปรีด

| ประเภท                   | ลักษณะ                         | เหมาะสำหรับ                          |
|--------------------------|--------------------------------|--------------------------------------|
| รูปรีดองค์รวม (Holistic) | ให้คะแนนภาพรวมของงานครั้งเดียว | สรุปตัดสินผลอย่างรวดเร็ว (Summative) |

|                                   |                                                |                                            |
|-----------------------------------|------------------------------------------------|--------------------------------------------|
| รูปรีคแยกองค์ประกอบ<br>(Analytic) | ให้คะแนนแยกรายเกณฑ์<br>พร้อมคำบรรยายแต่ละระดับ | ให้ข้อมูลย้อนกลับเพื่อพัฒนา<br>(Formative) |
| รูปรีคแบบผสม<br>(Annotated)       | ประเมินองค์รวมแล้วระบุจุดเด่น-<br>จุดด้อยกำกับ | แก้จุดอ่อนของสองแบบแรก                     |
| รูปรีคสองชั้น (Double-<br>layer)  | มีเกณฑ์ย่อยซ้อนในเกณฑ์หลัก                     | งานซับซ้อนที่ต้องการความละเอียดสูง         |

ที่มา: ดัดแปลงจาก คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย (2567); พิชิต ฤทธิ์จรูญ (2562); เทียบเคียง พชรินทร์ ชมภูวิเศษ (2559, น. 168)

**ขั้นตอนการสร้าง.** การสร้างรูปรีคที่ใช้ได้จริงมีขั้นตอนเป็นระบบ ดังสรุปในกล่องที่ 5.4 หัวใจอยู่ที่การ ‘ตรึงระดับมาตรฐาน’ ไว้ที่ระดับ 3 (ปฏิบัติได้เองตามเกณฑ์) ก่อน แล้วจึงไล่ขึ้น-ลงเป็นระดับอื่น

#### กล่องที่ 5.4 ขั้นตอนการสร้างเกณฑ์การให้คะแนนแบบรูปรีค

- 1) กำหนดสิ่งที่ประเมิน โดยเขียนนิยามให้ชัดเจนว่าจะเน้นกระบวนการ ผลงาน หรือทั้งสอง
- 2) กำหนดจำนวนระดับ (นิยม 4 ระดับ; อาจมี 0 กรณีไม่ส่งงาน) ให้ระดับ 3–4 = ผ่าน, 1–2 = ไม่ผ่าน
- 3) กำหนดระดับความสามารถ โดยพิจารณาระดับ 3 เป็นมาตรฐานก่อน (ทำได้เองโดยไม่ต้องพึ่งผู้อื่น)
- 4) เขียนคำบรรยายแต่ละระดับด้วยถ้อยคำที่สังเกตได้ หลีกเลี่ยงคำขยายที่เป็นนามธรรม
- 5) ทดลองใช้กับผู้ประเมิน 2 คน (หรือคนเดียวประเมินสองครั้ง) ปรับปรุง แล้วจัดทำฉบับจริงพร้อมคู่มือ

ที่มา: ดัดแปลงจาก ส.วาสนา ประवालพฤกษ์ (2544) อ้างถึงใน พชรินทร์ ชมภูวิเศษ (2559, น. 169)

**ตัวอย่างเครื่องมือจริง.** ตารางที่ 5.8 เป็นตัวอย่างรูปรีคแบบแยกองค์ประกอบ (analytic rubric) 4 ระดับ สำหรับประเมินทักษะ ‘การนำเข้าสู่บทเรียน’ ของนักศึกษาฝึกสอน สังเกตว่าคำบรรยายแต่ละช่องระบุพฤติกรรมที่สังเกตเห็นได้จริง ทำให้ผู้ประเมินต่างคนตัดสินตรงกันมากขึ้น

ตารางที่ 5.8 ตัวอย่างรูปรีคแบบแยกองค์ประกอบ: ทักษะการนำเข้าสู่บทเรียน

| เกณฑ์                           | 4 (ดีเยี่ยม)                               | 3 (ดี)                               | 2 (พอใช้)                        | 1 (ปรับปรุง)                   |
|---------------------------------|--------------------------------------------|--------------------------------------|----------------------------------|--------------------------------|
| การกระตุ้น<br>ความสนใจ          | ใช้สื่อแปลกใหม่และเชื่อมโยงเนื้อหาได้ทันที | ใช้สื่อสอดคล้องเนื้อหาแต่ไม่แปลกใหม่ | ใช้เพียงคำพูดทักทายกระตุ้น       | ไม่มีการกระตุ้นความสนใจ        |
| ความต่อเนื่อง<br>เข้าสู่เนื้อหา | เชื่อมโยงสู่เนื้อหาใหม่อย่างไร้รอยต่อ      | เชื่อมโยงชัดเจนแต่กระโดดบ้าง         | บอกหัวข้อแต่ไม่เห็นความเชื่อมโยง | เริ่มเนื้อหาใหม่โดยไม่เกริ่นนำ |
| การใช้คำถาม                     | ใช้คำถามปลายเปิดกระตุ้นการคิดวิเคราะห์     | ใช้คำถามกระตุ้นความจำเนื้อหา         | ใช้คำถามตอบเพียงใช่/ไม่ใช่       | ไม่มีการใช้คำถามนำ             |

ที่มา: ผู้เขียน ดัดแปลงแนวทางจาก Brookhart (2013) และ พชรินทร์ ชมภูวิเศษ (2559)

**การประยุกต์และข้อควรระวัง.** ส.วาสนา ประवालพฤกษ์ (2544 อ้างถึงใน พัชรินทร์ ชมภูวิเศษ, 2559) เสนอว่าระหว่างจัดกิจกรรมควรใช้ *analytic rubric* เพื่อช่วยให้ผู้เรียนรู้ว่าตนบกพร่องเรื่องใด (*formative*) ส่วนการตัดสินขั้นสุดท้ายจึงใช้ *holistic rubric (summative)* ข้อควรระวังสำคัญคือ การเขียน *descriptor* ต้องเป็นพฤติกรรมที่สังเกตได้ หลีกเลี่ยงคำคลุมเครือ เช่น ‘ดี’ ‘เหมาะสม’ ลอย ๆ โดยไม่บอกว่าดีอย่างไร เพราะคำคลุมเครือคือบ่อเกิดของความลำเอียง (ดูหัวข้อ 5.5.9)

#### 5.5.8 การให้คะแนน การแปลผล และการให้ข้อมูลย้อนกลับ (Feedback)

**นิยาม.** การให้คะแนน (scoring) คือการแปลงพฤติกรรมที่สังเกตเป็นตัวเลขตามเกณฑ์ ส่วนการแปลผล (interpretation) คือการตัดสินคุณค่าโดยให้ความหมายแก่คะแนนนั้นเทียบกับเกณฑ์ (พิสนุ พองศรี, 2551) เพราะ ‘คะแนนดิบ’ (raw score) ไม่มีความหมายในตัวเอง และการให้ข้อมูลย้อนกลับ (feedback) คือการสื่อสารจุดเด่น-จุดที่ต้องพัฒนาเพื่อให้ผู้เรียนปรับปรุงได้ การประเมินปฏิบัติที่ดีต้องทำหน้าที่เป็นการประเมินเพื่อพัฒนา (Formative Assessment) มิใช่เพียงตัดสินได้-ตก (Brookhart, 2013; Nitko & Brookhart, 2019) โมเดลการให้ข้อมูลย้อนกลับและกลยุทธ์การนำผลไปพัฒนาผู้เรียน ศึกษาในบทที่ 11

**คำอธิบายเชิงลึก.** ก่อนประเมิน ครูควรกำหนด ‘สัดส่วนคะแนน’ ระหว่างกระบวนการกับผลงานให้ชัดเจนและเหมาะสมกับธรรมชาติของงาน (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) จากนั้นรายงานผลได้ 3 แนวทาง คือ เทียบกับสมาชิกในกลุ่ม เทียบกับเกณฑ์มาตรฐาน หรือเทียบพัฒนาการของผู้เรียนเองก่อน-หลัง ส่วนการให้ feedback ที่ดี สุทธิพงศ์ บุญผดุง (2568) เสนอ 3 ลักษณะ ดังกล่องที่ 5.5

#### กล่องที่ 5.5 หลักการให้คะแนนและการให้ข้อมูลย้อนกลับที่ส่งเสริมการเรียนรู้

**กำหนดสัดส่วนล่วงหน้า:** ตกลงน้ำหนักคะแนน ‘กระบวนการ : ผลงาน’ ก่อนประเมิน เพื่อความโปร่งใสและเป็นธรรม

**เชื่อมโยง:** ให้คำแนะนำที่เชื่อมโยง ‘ความรู้/ทักษะเดิม’ เข้ากับ ‘สิ่งที่ต้องพัฒนาใหม่’

**แก้ไขทันท่วงที:** ชี้ความเข้าใจที่คลาดเคลื่อน (misconception) ให้ผู้เรียนแก้ได้ทันที ไม่รอผลปลายภาค

**ตั้งเป้าหมาย:** ส่งเสริมให้ผู้เรียนตั้งเป้าพัฒนาตนเอง (self-regulation) จากข้อมูลย้อนกลับ

ที่มา: ดัดแปลงจาก สุทธิพงศ์ บุญผดุง (2568) และ Brookhart (2013)

**ตัวอย่างในชั้นเรียนไทย.** หลังการสอนจำลอง (micro-teaching) ครูนิเทศก์ไม่ควรบอกเพียงคะแนนรวม แต่ควรชี้ทันทีว่า ‘การนำเข้าสู่บทเรียนน่าสนใจ (จุดเด่น) แต่คำถามส่วนใหญ่เป็นปลายปิด (จุดพัฒนา) คราวหน้าลองตั้งคำถาม ‘เพราะอะไร/อย่างไร’ อย่างน้อย 3 คำถาม (เป้าหมาย)’ ข้อมูลย้อนกลับเช่นนี้เปลี่ยนการประเมินจากการตัดสินเป็นการพัฒนา

**การประยุกต์และข้อควรระวัง.** ข้อควรระวังคือการให้แต่ ‘คะแนนดิบ’ หรือการเปรียบเทียบผู้เรียนกับเพื่อนอย่างเดียว ซึ่งบั่นทอนกำลังใจและไม่ช่วยให้รู้แนวทางพัฒนา ครูควรให้ข้อมูลย้อนกลับที่เจาะจงพฤติกรรม และชี้แนวทางปรับปรุงควบคู่กับคะแนน เพื่อให้การประเมินปฏิบัติรับใช้การเรียนรู้อย่างแท้จริง

### 5.5.9 ความคลาดเคลื่อนและความลำเอียงที่ต้องระวัง

นิยาม. ความคลาดเคลื่อนในการวัด (Measurement Error) คือส่วนต่างระหว่างคะแนนที่วัดได้กับความสามารถที่แท้จริงของผู้เรียน ในการประเมินปฏิบัติซึ่งอาศัยดุลยพินิจสูง ความคลาดเคลื่อนส่วนใหญ่มาจาก ‘ความลำเอียงของผู้ประเมิน’ (rater bias) หากเกณฑ์ไม่ชัดเจนหรือผู้ตรวจลำเอียง ผลการประเมินจะขาดความเชื่อถือได้ (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ตามทฤษฎีการทดสอบแบบดั้งเดิม คะแนนที่สังเกตได้เท่ากับคะแนนจริงบวกกับความคลาดเคลื่อน ( $X = T + E$ ) ครูจึงควรออกแบบการประเมินให้มีความคลาดเคลื่อน (E) เหลือน้อยที่สุด (ศิริชัย กาญจนวาสี, 2552)

**ชนิดของความลำเอียงที่พบบ่อย.** ความลำเอียงที่พบบ่อยในการประเมินปฏิบัติมีหลายชนิด แต่ละชนิดมีผลเสียและวิธีลดต่างกัน ดังตารางที่ 5.9

ตารางที่ 5.9 ชนิดของความลำเอียงของผู้ประเมินและแนวทางลด

| ชนิดความลำเอียง  | ลักษณะ                                               | ผลเสีย                                  | แนวทางลด                                     |
|------------------|------------------------------------------------------|-----------------------------------------|----------------------------------------------|
| Halo Effect      | ให้คะแนนสูงทุกด้านเพราะประทับใจบุคลิก/ผลงานด้านเดียว | คะแนนไม่สะท้อนความจริงรายด้าน           | แยกประเมินที่ละเกณฑ์ด้วย analytic rubric     |
| Generosity Error | ใจดีให้คะแนนสูงเกินจริง                              | ผู้เรียนไม่เห็นความจำเป็นต้องพัฒนา      | ยึด descriptor + เทียบตัวอย่างงาน (exemplar) |
| Central Tendency | ให้คะแนนกลาง ๆ เหมือนกันทุกคน                        | แยกผู้เก่ง-อ่อนไม่ได้ ผู้เก่งขาดกำลังใจ | ใช้ระดับที่มีคำบรรยายชัด + อบรมผู้ประเมิน    |
| Logical Error    | ให้คะแนนคล้ายกันในเกณฑ์ที่คิดว่าสัมพันธ์กัน          | การตัดสินปนเปื้อนข้ามเกณฑ์              | นิยามแต่ละเกณฑ์ให้เป็นอิสระจากกัน            |

ที่มา: ดัดแปลงจาก พัชรินทร์ ชมภูวิเศษ (2559) และ Nitko & Brookhart (2019)

**ตัวอย่างในชั้นเรียนไทย.** ครูให้คะแนนการทำงานกลุ่มของนักเรียนที่เรียนเก่งคณิตศาสตร์สูงกว่าปกติ ทั้งที่นักเรียนคนนั้นไม่ได้ช่วยงานกลุ่มจริง นั่นคือ Halo Effect ในทางกลับกัน ครูที่ให้ทุกคนได้ระดับ ‘พอใช้’ เท่ากันหมดเพราะเกรงใจ คือ Central Tendency ซึ่งทำให้ผู้เรียนที่ตั้งใจรู้สึกว่าความพยายามไม่ถูกมองเห็น

**การประยุกต์: หลักฐานเชิงประจักษ์.** Jonsson และ Svingby (2007) ทบทวนงานวิจัย 75 เรื่อง เกี่ยวกับการใช้รูบรีค พบว่าการให้คะแนนงานปฏิบัติจะ ‘เชื่อถือได้’ (reliable) สูงขึ้นเมื่อรูบรีคเป็นแบบแยก องค์ประกอบ เจาะจงหัวข้อ มีตัวอย่างงานประกอบ และผู้ประเมินได้รับการอบรม อย่างไรก็ตามรูบรีคเพียงอย่างเดียวไม่ได้รับประกัน ‘ความตรง’ (validity) โดยอัตโนมัติ ครูยังต้องออกแบบภาระงานให้ตรงกับจุดประสงค์ แนวทางลดความคลาดเคลื่อนเชิงปฏิบัติสรุปไว้ในกล่องที่ 5.6

#### กล่องที่ 5.6 แนวทางลดความคลาดเคลื่อนในการประเมินปฏิบัติ

- กำหนดพฤติกรรมบ่งชี้ (observable indicators) ที่ชัดเจนในรูบรีค หลีกเลี่ยงคำคลุมเครือ
- ใช้ผู้ประเมินหลายคน (multiple raters) และทำ moderation รับความเข้าใจเกณฑ์ร่วมกันก่อนประเมิน
- อบรมผู้ประเมิน (rater training) และใช้ตัวอย่างงาน (exemplars) เทียบมาตรฐาน
- หลักฐานวิจัย: รูบรีคแบบ analytic + เจาะจงหัวข้อ + exemplar + rater training เพิ่มความเชื่อมั่นของการให้คะแนน (Jonsson & Svingby, 2007)

ที่มา: สังเคราะห์จาก พัทรินทร์ ชมภูวิเศษ (2559) และ Jonsson & Svingby (2007)

#### 5.5.10 การประยุกต์ใช้ในวิชาชีพครูและกรณีศึกษา

บริบทการใช้งานจริง. การประเมินปฏิบัติปรากฏในงานครูหลายบริบท ทั้งการสอนจำลอง (micro-teaching) การควบคุมและประเมินทักษะในห้องปฏิบัติการ การฝึกประสบการณ์วิชาชีพในสถานศึกษาจริง การแข่งขันทักษะ ตลอดจนการประเมินบุคลิกภาพความเป็นครู ซึ่งต้องอาศัยการสังเกตอย่างต่อเนื่องมากกว่าการสอบครั้งเดียว ตารางที่ 5.10 เชื่อมโยงบริบทเหล่านี้กับเครื่องมือและจุดประสงค์ของบท

ตารางที่ 5.10 การประยุกต์การประเมินปฏิบัติในบริบทวิชาชีพครู

| บริบทงานครู                    | สิ่งที่ประเมิน                 | เครื่องมือ/วิธี              | เชื่อมโยง CLO |
|--------------------------------|--------------------------------|------------------------------|---------------|
| การสอนจำลอง (Micro-teaching)   | ทักษะการจัดการเรียนรู้         | รูบรีค + แบบสังเกตการปฏิบัติ | CLO3, CLO4    |
| ปฏิบัติการในห้องเรียน/ห้องแล็บ | การใช้เครื่องมือและความปลอดภัย | แบบตรวจสอบรายการ + การสังเกต | CLO4          |
| การฝึกประสบการณ์วิชาชีพ        | สมรรถนะการสอนในสภาพจริง        | รูบรีค + แฟ้มสะสมงาน         | CLO4          |
| การแข่งขันทักษะ/โครงการ        | กระบวนการและผลงาน              | รูบรีคแบบแยกองค์ประกอบ       | CLO3          |

ที่มา: ผู้เขียน สังเคราะห์จาก พัทรินทร์ ชมภูวิเศษ (2559) และ Nitko & Brookhart (2019)

**กรณีศึกษาเชิงวิเคราะห์.** กล่องที่ 5.7 เป็นกรณีศึกษาที่ช่วยให้นักศึกษาฝึกเชื่อมโยงระดับทักษะพิสัย เข้ากับการเลือกเครื่องมือและการให้ข้อมูลย้อนกลับ

**กล่องที่ 5.7 กรณีศึกษาเชิงวิเคราะห์: การประเมินทักษะการจัดกิจกรรมกลุ่ม**

**สถานการณ์:** ครูที่เลี้ยงประเมินทักษะ ‘การจัดกิจกรรมกลุ่ม’ ของนักศึกษาฝึกสอน พบว่าทำตามขั้นตอนได้ดี แต่เมื่อนักเรียนสองคนขัดแย้งกลับแก้ปัญหาไม่ได้

**วิเคราะห์:** นักศึกษาอยู่ระดับ ‘การปรับระบบกลไก (Mechanism)’ แต่ยังไม่ถึง ‘การดัดแปลง (Adaptation)’ ตามแนวคิดของ Simpson (1972)

**เครื่องมือที่ควรใช้:** rubric แบบแยกองค์ประกอบ เพื่อแยกคะแนน ‘ขั้นตอนการจัดกิจกรรม’ ออกจาก ‘การแก้ปัญหาเฉพาะหน้า’ ทำให้ข้อมูลย้อนกลับตรงจุดและพัฒนาได้ (Brookhart, 2013)

ที่มา: ผู้เขียน ประยุกต์จาก Simpson (1972) และ Brookhart (2013)

**การประยุกต์และข้อควรระวัง.** เนื้อหาทั้งหมดในหัวข้อ 5.5 รับใช้จุดประสงค์ของบทโดยตรง ดังสรุปในกล่องท้ายหัวข้อ ครูจึงควรออกแบบการประเมินปฏิบัติให้เชื่อมโยงจุดประสงค์–เครื่องมือ–การให้คะแนนอย่างเป็นระบบ และตระหนักว่าความเที่ยงธรรม (fairness) คือคุณธรรมสำคัญของผู้ประเมิน

**การเชื่อมโยงเนื้อหา 5.5 กับจุดประสงค์ของบท (Constructive Alignment)**

**CLO3 (ออกแบบ/สร้างเครื่องมือ):** รองรับด้วยหัวข้อ 5.5.3–5.5.7 (ลักษณะทักษะ ประเภท วิธีวัด เครื่องมือ และการสร้าง rubric) ประเมินผ่านชิ้นงานในใบงาน 5.1

**CLO4 (ปฏิบัติการประเมิน + ให้ข้อมูลย้อนกลับ):** รองรับด้วยหัวข้อ 5.5.8–5.5.10 (การให้คะแนน feedback ลดความลำเอียง และการประยุกต์) ประเมินผ่านการปฏิบัติในใบงาน 5.2

**กรณีศึกษา (สาขาการศึกษาปฐมวัย):** การประเมินอย่างใคร่ครวญผ่านการสังเกตการปฏิบัติจริง  
บุษบง ตันติวงศ์ และศศิลักษณ์ ขยันกิจ (2561) เสนอแนวทาง "การประเมินอย่างใคร่ครวญ" สำหรับเด็กปฐมวัย ซึ่งครูเฝ้าสังเกตเด็กขณะทำกิจกรรมจริง บันทึกพฤติกรรมเชิงบรรยายอย่างเป็นรูปธรรมโดยไม่ด่วนตัดสิน แล้วจึงนำหลักฐานหลายชิ้นมาตีความร่วมกันระหว่างครูและผู้ปกครอง เพื่อสนับสนุนการเรียนรู้ของเด็กเป็นรายบุคคล แนวทางนี้สะท้อนหลักการประเมินทักษะปฏิบัติของบทนี้โดยตรง คือ เน้นกระบวนการควบคุมผลงาน ใช้การสังเกตอย่างมีโครงสร้าง และแยก "ข้อเท็จจริงที่เห็น" ออกจาก "การตีความของผู้ประเมิน" เพื่อลดความลำเอียง

**คำถามชวนคิด:** การวัดพัฒนาการเด็กปฐมวัยด้วยข้อสอบกระดามีข้อจำกัดใด และหากท่านเป็นครูปฐมวัย แบบบันทึกการสังเกตของท่านควรออกแบบช่องรายการอย่างไรให้บันทึกข้อเท็จจริงก่อนการตีความ

## 5.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ใบงานที่ 5.1 ออกแบบภาระงานปฏิบัติและสร้างรูบรีค (รายคู่) — เชื่อมโยง CLO3

1. เลือกทักษะปฏิบัติในวิชาเอก 1 ทักษะ แล้วเขียนคำอธิบายภาระงาน (Task Description) ที่สะท้อนสภาพจริง
2. สร้าง Analytic Rubric 4 ระดับคุณภาพ กำหนดเกณฑ์อย่างน้อย 3 ด้าน (เช่น กระบวนการ ความถูกต้อง ความปลอดภัย)
3. ตรวจสอบ Descriptor ว่าระบุพฤติกรรมที่สังเกตเห็นได้จริง แล้วทดลองใช้ประเมินผลงานของคู่

### ใบงานที่ 5.2 ปฏิบัติการประเมินการสอนจำลองของเพื่อน (Peer Assessment) — เชื่อมโยง CLO4

1. ใช้แบบสังเกต/รูบรีคที่สร้างขึ้น ประเมินการสอนจำลอง (Micro-teaching) ของเพื่อน 1 คน
  2. ให้ข้อมูลย้อนกลับเชิงสร้างสรรค์ตามหลัก 3 ข้อ (เชื่อมโยง-แก้ไข-ตั้งเป้าหมาย)
  3. เขียนบันทึกสะท้อนคิด (Reflection) เกี่ยวกับความลำเอียงที่เกิดขึ้นและวิธีลดความคลาดเคลื่อน
- การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 5.7 สรุปทบทวน

การประเมินทักษะปฏิบัติเน้นพฤติกรรมที่สังเกตได้ขณะลงมือทำ พิจารณาทั้งกระบวนการและผลผลิต เป็นการวัด “การกระทำ (Doing)” ที่ข้อสอบข้อเขียนทำไม่ได้ (ชาวล แพร์ตกุล, 2520) พฤติกรรมทักษะพิสัยได้ระดับตั้งแต่การรับรู้จนถึงการริเริ่มตามแนวคิดของ Simpson (1972) ลักษณะสำคัญของการปฏิบัติประกอบด้วยขั้นตอน การใช้เครื่องมือ การแก้ปัญหา ความชำนาญ และความปลอดภัย ครูเลือกวิธีวัด (สังเกต สานิต ปฏิบัติ สถานการณ์จำลอง แฟ้มสะสมงาน) และเครื่องมือ (Checklist, Rating Scale, Rubric) ให้สอดคล้องกับทักษะ โดยมีรูบรีคที่มี Descriptor ชัดเจนเป็นหัวใจของความเที่ยงตรงและความยุติธรรม (Brookhart, 2013) การให้คะแนนต้องมาพร้อมข้อมูลย้อนกลับเพื่อพัฒนา และต้องระวังความลำเอียง (Halo,

Generosity, Central Tendency) ด้วยการใช้ผู้ประเมินหลายคน การทำ Moderation และการอบรมผู้ประเมิน (สุทธิพงศ์ บุญผดุง, 2568)

## 5.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 5.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

- อธิบายความแตกต่างระหว่าง “การวัดผล” กับ “การประเมินผล” ในบริบทของทักษะปฏิบัติ พร้อมยกตัวอย่าง
- เหตุใดการวัดทักษะปฏิบัติจึงต้องพิจารณาทั้งกระบวนการและผลผลิต และเหตุใดจึงเรียกว่าเป็นการวัดทางอ้อม
- ในการสอบปฏิบัติ “การสอนหน้าชั้น 10 นาที” ของนักศึกษาครู จงออกแบบว่าจะประเมิน “กระบวนการ” และ “ผลผลิต” ด้วยเครื่องมือใด และหากมีกรรมการ 3 คน จะควบคุมความคลาดเคลื่อนระหว่างผู้ประเมินอย่างไร
- เปรียบเทียบ Analytic Rubric กับ Holistic Rubric  
ว่าเหมาะกับการประเมินเพื่อพัฒนาและการตัดสินผลอย่างไร
- หากเครื่องมือประเมินของท่านมีความเชื่อมั่น (Reliability) ต่ำ ท่านจะมีวิธีแก้ไขเชิงเทคนิคอย่างไรบ้าง
- ยกตัวอย่างความลำเอียงของผู้ประเมิน 3 แบบ  
พร้อมเสนอแนวทางลดความคลาดเคลื่อนในการประเมินการสอนจำลอง
- เพราะเหตุใดความเที่ยงธรรมและความยุติธรรมจึงเป็นคุณธรรมสำคัญของครูในการประเมินทักษะปฏิบัติ

## 5.10 เครื่องมือวัดและประเมินผลประจำบท

ก) รุบริคตรวจชิ้นงานใบงาน 5.1 (Meta-rubric: รุบริคสำหรับตรวจรุบริค) — เชื่อมโยง CLO3

| เกณฑ์                     | 4 (ดีเยี่ยม)                              | 3 (ดี)               | 2 (พอใช้)       | 1 (ปรับปรุง)             |
|---------------------------|-------------------------------------------|----------------------|-----------------|--------------------------|
| ความสอดคล้องกับจุดประสงค์ | ทุกเกณฑ์สะท้อนภาระงานและจุดประสงค์ครบถ้วน | สอดคล้องเป็นส่วนใหญ่ | สอดคล้องบางส่วน | ไม่สอดคล้องกับจุดประสงค์ |

|                          |                                     |                    |                   |                          |
|--------------------------|-------------------------------------|--------------------|-------------------|--------------------------|
| ความชัดเจนของ Descriptor | บรรยายพฤติกรรมสังเกตได้จริงทุกระดับ | ชัดเจนเป็นส่วนใหญ่ | คลุมเครือหลายช่อง | ไม่มีคำบรรยายระดับคุณภาพ |
| การจำแนกระดับคุณภาพ      | ระดับต่างกันชัดเจนวัดได้            | ต่างกันพอสังเกตได้ | ระดับเหมือนกัน    | แยกระดับไม่ได้           |

เกณฑ์ผ่าน: ได้คะแนนรวมไม่ต่ำกว่าร้อยละ 70 (ระดับ “ดี” ขึ้นไปทุกเกณฑ์)

#### ข) แบบสังเกตการปฏิบัติการประเมิน (ใบงาน 5.2) — เชื่อมโยง CLO4

| พฤติกรรมที่สังเกต                                                    | ปฏิบัติได้<br>(3) | ปฏิบัติได้บางส่วน<br>(2) | ยังไม่ปฏิบัติ<br>(1) |
|----------------------------------------------------------------------|-------------------|--------------------------|----------------------|
| ใช้เครื่องมือ/รูบรีคประเมินได้ถูกต้องตามเกณฑ์                        |                   |                          |                      |
| ให้คะแนนสอดคล้องกับพฤติกรรมที่สังเกต<br>(ลดความลำเอียง)              |                   |                          |                      |
| ให้ข้อมูลย้อนกลับครบ 3 องค์ประกอบ (เชื่อมโยง-<br>แก้ไข-ตั้งเป้าหมาย) |                   |                          |                      |
| แสดงความเที่ยงธรรมและมารยาทในการประเมิน                              |                   |                          |                      |

เกณฑ์ผ่าน: รวม  $\geq 9$  จาก 12 คะแนน และไม่มีรายการใดอยู่ระดับ 1

### 5.11 เอกสารอ้างอิงประจำบท

#### ภาษาไทย

- คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย. (2567). *เกณฑ์การให้คะแนนแบบรูบรีค (Scoring Rubric)*. เอกสารประกอบการประชุมคณาจารย์คณะจิตวิทยา ครั้งที่ 10/2567. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ชวาล แพรัตกุล. (2520). *เทคนิคการเขียนข้อทดสอบ*. กรุงเทพฯ: โรงพิมพ์พิทักษ์อักษร.
- บุษบง ต้นดวงศ์ และศศิลักษณ์ ชัยนิกิจ. (2561). การประเมินอย่างใคร่ครวญต่อเด็กปฐมวัย: แนวคิดและการปฏิบัติเพื่อสนับสนุนการเรียนรู้. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- พัชรินทร์ ชมภูวิเศษ. (2559). *การวัดและประเมินผลการเรียนรู้*. อุดรธานี: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.
- พิชิต ฤทธิ์เจริญ. (2562). *เกณฑ์การให้คะแนน: เครื่องมือสำหรับครูในการประเมินผลการเรียนรู้*. วารสารศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช, 12(1), 1–13.
- พิสนุ ฟองศรี. (2551). *การประเมินทางการศึกษา: แนวคิดสู่การปฏิบัติ* (พิมพ์ครั้งที่ 5). กรุงเทพฯ: ด้านสหวิชาการพิมพ์.

ศิริชัย กาญจนวาสี. (2552). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 6). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำเริง บุญเรืองรัตน์. (2534). *การวัดและประเมินผลการศึกษา*. กรุงเทพฯ: สำนักพิมพ์พัฒนาศึกษา.

สุทธิพงศ์ บุญผดุง. (2568). *เอกสารประกอบการสอน รายวิชาการวัดและประเมินผลการเรียนรู้ (ED12101)*. กรุงเทพฯ: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

อนุวัติ คุณแก้ว. (2566). *การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่*. สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

## ภาษาอังกฤษ

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Simpson, E. J. (1972). *The classification of educational objectives in the psychomotor domain*. Washington, DC: Gryphon House.

## บทที่ 6

### การวัดและประเมินด้านจิตพิสัยและเจตคติ

#### Measurement and Assessment of the Affective Domain and Attitudes

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สัปดาห์ที่ 6 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

บทนี้ว่าด้วยการวัดและประเมิน “พฤติกรรมด้านจิตพิสัย” (Affective Domain) ซึ่งครอบคลุมความรู้สึก เจตคติ ค่านิยม ความสนใจ และคุณธรรม อันเป็นคุณลักษณะนามธรรมที่วัดทางตรงไม่ได้ ต้องอนุมานจากพฤติกรรมที่แสดงออก เนื้อหาเริ่มจากความหมายและธรรมชาติของจิตพิสัย ระดับพฤติกรรม 5 ขั้น ตามแครธวอลและคณะ ความสำคัญต่อการจัดการศึกษาอิงผลลัพธ์ (OBE) และวิชาชีพครู เครื่องมือวัด (แบบสอบถาม มาตรฐานค่าแบบลิเคิร์ต แบบนัยจำแนก การสังเกต การสัมภาษณ์ การประเมินตนเอง/ โดยเพื่อน) การสร้างแบบวัดเจตคติและการให้คะแนน การประเมินคุณลักษณะอันพึงประสงค์ ตลอดจนความคลาดเคลื่อน อคติ และจรรยาบรรณในการประเมิน เพื่อให้นักศึกษาออกแบบและใช้เครื่องมือวัดจิตพิสัยได้อย่างมีคุณภาพและเที่ยงธรรม (เชื่อมโยง CLO3 และ CLO5)

#### แผนบริหารการสอนประจำบทที่ 6

##### หัวข้อสาระจัดการเรียนการสอน

- ความหมาย ธรรมชาติ และความสำคัญของพฤติกรรมจิตพิสัยและเจตคติ
- ระดับพฤติกรรมจิตพิสัย 5 ขั้นตามแนวคิดของแครธวอลและคณะ (Krathwohl Taxonomy)
- ความสำคัญของจิตพิสัยต่อการจัดการศึกษาอิงผลลัพธ์ (OBE) และคุณลักษณะอันพึงประสงค์
- เครื่องมือวัดด้านจิตพิสัยและเจตคติ และการเลือกใช้ให้เหมาะสมกับคุณลักษณะที่วัด
- มาตรฐานค่าและการสร้างแบบวัดเจตคติแบบลิเคิร์ต (การเขียนข้อความ การให้คะแนน การแปลผล)

- การสังเกตพฤติกรรมและการประเมินคุณลักษณะอันพึงประสงค์
- ความคลาดเคลื่อน อคติ และแนวทางลดอคติในการประเมินจิตพิสัย
- คุณธรรม จริยธรรม และความเที่ยงธรรมในการวัดและประเมินจิตพิสัย

### ความคิดรวบยอด

พฤติกรรมจิตพิสัยเป็นคุณลักษณะภายในที่เป็นนามธรรม (ความรู้สึก เจตคติ ค่านิยม คุณธรรม) วัดทางตรงไม่ได้ จึงต้องวัดทางอ้อมโดยอนุมานจากพฤติกรรมที่แสดงออก ด้วยเครื่องมือหลายชนิดและวัดอย่างต่อเนื่อง การประเมินที่ดีต้องอาศัยทั้งความเชี่ยวชาญด้านเครื่องมือและจรรยาบรรณวิชาชีพ เพื่อให้ผลการวัดเที่ยงตรง เป็นธรรม และนำไปพัฒนาผู้เรียนได้จริง มิใช่เพียงการตัดสินหรือตีตรา

### จุดประสงค์การเรียนรู้

1. อธิบายความหมาย ธรรมชาติ และความสำคัญของพฤติกรรมจิตพิสัยและเจตคติได้ (CLO3)
2. จำแนกระดับพฤติกรรมจิตพิสัย 5 ชั้นของแคธวอลพร้อมยกตัวอย่างพฤติกรรมแต่ละชั้นได้ (CLO3)
3. เลือกและเปรียบเทียบเครื่องมือวัดจิตพิสัยให้เหมาะสมกับคุณลักษณะที่ต้องการวัดได้ (CLO3)
4. ออกแบบและสร้างแบบวัดเจตคติแบบลิเคิร์ทและแบบสังเกตคุณลักษณะที่มีคุณภาพได้ (CLO3)
5. วิเคราะห์สาเหตุของความคลาดเคลื่อนและอคติ พร้อมเสนอแนวทางลดอคติในการประเมินได้ (CLO5)
6. ตระหนักและประพฤติตามจรรยาบรรณและความเที่ยงธรรมในการประเมินจิตพิสัยได้ (CLO5)

### เนื้อหาวิชา

ดำเนินการสอนตามเอกสารประกอบการสอนบทที่ 6 หัวข้อ 6.5 เนื้อหาวิชา (6.5.1–6.5.8) ซึ่งครอบคลุมความหมายและสอนตามเอกสารประกอบการสอนบทที่ 6 หัวข้อ 6.5 เนื้อหาวิชา (6.5.1–6.5.8) ซึ่งครอบคลุมความหมายและธรรมชาติของจิตพิสัย ระดับพฤติกรรมตามแคธวอล ความสำคัญต่อ OBE และ

วิชาชีพครู เครื่องมือวัด มาตรฐานค่าและการสร้างแบบวัดเจตคติ การสังเกตและการประเมินคุณลักษณะ ความคลาดเคลื่อน/อคติ และจรรยาบรรณการประเมิน

### กิจกรรมการเรียนรู้การสอน

1. ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน และกิจกรรมสำรวจค่านิยม/เจตคติของตนเอง (Value Clarification) เพื่อสร้างความสนใจ
2. ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ประกอบสไลด์และกรณีตัวอย่าง อภิปรายกลุ่มจำแนกระดับจิตพิสัย และเวิร์กช็อปวิเคราะห์ข้อความในแบบวัดเจตคติ
3. ขั้นระหว่างเรียน (ปฏิบัติ): ใบงาน 6.1 สร้างแบบวัดเจตคติแบบลิเคิร์ท และใบงาน 6.2 ออกแบบแบบสังเกตคุณลักษณะอันพึงประสงค์ พร้อมวิพากษ์โดยเพื่อน (Peer Review)
4. ขั้นหลังเรียน: ทำแบบทดสอบหลังเรียนและแบบฝึกหัดท้ายบท วิเคราะห์กรณีศึกษาอคติในการประเมิน และสะท้อนคิดด้านจรรยาบรรณการวัดผล

### สื่อการเรียนรู้การสอน

เอกสารประกอบการสอนบทที่ 6 และสไลด์นำเสนอ

ตัวอย่างเครื่องมือจริง: แบบวัดเจตคติแบบลิเคิร์ท แบบนัยจำแนก แบบสังเกต และแบบบันทึกพฤติกรรม (Anecdotal Record) ใบงาน 6.1–6.2 และรูปรีดประเมินชิ้นงาน

กรณีศึกษาอคติในการประเมิน (Halo Effect / Central Tendency) และตำราในรายการอ้างอิงประจำบท

### การวัดและประเมินผล

- แบบทดสอบก่อน-หลังเรียน วัดความเข้าใจโน้ตค้นจิตพิสัย เครื่องมือ และหลักการให้คะแนน (CLO3)
- ชิ้นงานใบงาน 6.1–6.2 ประเมินด้วยรูปรีดแบบแยกองค์ประกอบ (Analytic Rubric) — เกณฑ์ผ่านร้อยละ 70 (CLO3)
- รายงานวิเคราะห์กรณีศึกษาอคติและแนวทางแก้ไข (CLO5/CLO3)
- แบบสังเกตพฤติกรรมจรรยาบรรณและความเที่ยงธรรมในการประเมิน (CLO5)

## 6.1 ความคิดรวบยอด

พฤติกรรมด้านจิตพิสัย (Affective Domain) เป็นพฤติกรรมที่เกี่ยวกับความรู้สึกนึกคิด อารมณ์ เจตคติ ค่านิยม ความสนใจ และคุณธรรม ซึ่งเป็นคุณลักษณะ “นามธรรม” ที่อยู่ภายในตัวบุคคล ไม่สามารถวัดได้โดยตรงเหมือนส่วนสูงหรือน้ำหนัก แต่ต้องวัดทางอ้อม (Indirect Measurement) โดยอนุมานจากพฤติกรรม หรือการแสดงออกที่สังเกตได้ (สุทธิพงศ์ บุญผดุง, 2568, น. 27–30) แครธวอล บลูม และมาเซีย (Krathwohl, Bloom, & Masia, 1964) จำแนกพฤติกรรมจิตพิสัยเป็น 5 ระดับจากต่ำไปสูง ได้แก่ การรับรู้ การตอบสนอง การสร้างค่านิยม การจัดระบบค่านิยม และการสร้างลักษณะนิสัย โดยเป้าหมายสูงสุดคือการหล่อหลอมค่านิยม จนกลายเป็นบุคลิกภาพถาวรที่คงเส้นคงวา

เนื่องจากจิตพิสัยเป็นนามธรรมและมีความแปรปรวนสูง การวัดจึงมีข้อจำกัดสำคัญคือ “พฤติกรรมประดิษฐ์” (Constructed Behavior) และการตอบตามความพึงปรารถนาทางสังคม (Social Desirability) ที่ผู้ถูกวัดสร้างขึ้นเพื่อให้ดูดี ครูในฐานะนักวัดผลจึงต้องใช้เครื่องมือหลายชนิดร่วมกัน วัดอย่างต่อเนื่อง และตรวจสอบข้อมูลแบบสามเส้า (Triangulation) เพื่อยืนยันพฤติกรรมที่แท้จริง อีกทั้งต้องยึดมั่นในจรรยาบรรณ และความเที่ยงธรรม เพราะการประเมินจิตพิสัยที่ไม่เป็นธรรมอาจนำไปสู่การตีตรา (Labeling) ที่ส่งผลเสียต่อผู้เรียน (พิชิต ฤทธิ์จรรยา, 2564; Nitko & Brookhart, 2019)

## 6.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 6 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                         | ระดับพฤติกรรม           | เชื่อมโยง CLO | วัดด้วย                |
|---------------------------------------------------------------------|-------------------------|---------------|------------------------|
| 1) อธิบายความหมาย ธรรมชาติ และความสำคัญของจิตพิสัยและเจตคติได้      | เข้าใจ (พุทธิพิสัย)     | CLO3          | แบบทดสอบ ก่อน-หลัง     |
| 2) จำแนกระดับพฤติกรรมจิตพิสัย 5 ขั้นพร้อมตัวอย่างได้                | เข้าใจ-วิเคราะห์        | CLO3          | แบบทดสอบ / ใบงาน       |
| 3) เลือกและเปรียบเทียบเครื่องมือวัดจิตพิสัยให้เหมาะกับคุณลักษณะได้  | วิเคราะห์-ประยุกต์      | CLO3          | ใบงาน / อภิปราย        |
| 4) ออกแบบและสร้างแบบวัดเจตคติแบบลิเคิร์ต และแบบสังเกตที่มีคุณภาพได้ | สร้างสรรค์ (ทักษะพิสัย) | CLO3          | ใบงาน 6.1–6.2 + RUBRIC |
| 5) วิเคราะห์สาเหตุความคลาดเคลื่อน/อคติ และเสนอแนวทางลดอคติได้       | วิเคราะห์-ประเมินค่า    | CLO5/CLO3     | กรณีศึกษา + RUBRIC     |

|                                                                  |                              |      |                    |
|------------------------------------------------------------------|------------------------------|------|--------------------|
| 6) ตระหนักและประพจน์ตามจรรยาบรรณและความเที่ยงธรรมในการประเมินได้ | รับรู้-เห็นคุณค่า (จิตพิสัย) | CLO5 | แบบสังเกตจรรยาบรรณ |
|------------------------------------------------------------------|------------------------------|------|--------------------|

หมายเหตุ: ทุกจุดประสงค์มีทั้งกิจกรรมการเรียนรู้ (6.6) วิธีวัด และเครื่องมือ (6.4, 6.8, 6.10) รองรับครบถ้วน โดย CLO3 (ออกแบบและสร้างเครื่องมือวัด) และ CLO5 (คุณธรรม จริยธรรม ความเที่ยงธรรม) เป็นผลลัพธ์หลักของบนี้

### 6.3 คำสำคัญ

จิตพิสัย (Affective Domain) • เจตคติ (Attitude) • ค่านิยม (Values) • การจำแนกพฤติกรรมจิตพิสัยของแครทวอล (Krathwohl Taxonomy) • การสร้างลักษณะนิสัย (Characterization) • การวัดทางอ้อม (Indirect Measurement) • มาตรฐานค่า (Rating Scale) • มาตรฐานค่าแบบลิเคิร์ต (Likert Scale) • แบบนัยจำแนก (Semantic Differential) • การให้คะแนนแบบกลับทาง (Reverse Scoring) • นิยามเชิงปฏิบัติการ (Operational Definition) • พฤติกรรมประดิษฐ์ (Constructed Behavior) • การตอบตามความพึงปรารถนาทางสังคม (Social Desirability) • การบันทึกพฤติกรรม (Anecdotal Record) • การตรวจสอบสามเส้า (Triangulation) • อคติในการประเมิน (Assessment Bias) • ความเที่ยงธรรม (Fairness)

### 6.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 6.5 เนื้อหาวิชา

#### 6.5.1 ความหมายและธรรมชาติของพฤติกรรมจิตพิสัยและเจตคติ

พฤติกรรมด้านจิตพิสัย (Affective Domain) หมายถึง พฤติกรรมที่เกี่ยวกับความรู้สึกนึกคิดทางจิตใจ อารมณ์ และการจัดระเบียบของจิตใจ ครอบคลุมเจตคติ ความสนใจ ความซาบซึ้ง ค่านิยม คุณธรรม และลักษณะนิสัย ซึ่งเกิดจากการปลูกฝังอบรมอย่างต่อเนื่อง มิใช่เกิดขึ้นได้ทันทีเหมือนการจำเนื้อหา (สุทธิพงศ์ บุญผดุง, 2568, น. 27; พชรินทร์ ชมภูวิเศษ, 2559) เมื่อกล่าวถึงพฤติกรรมการเรียนรู้ทั้งสามด้าน จิตพิสัยเป็นด้านที่วัดได้ยากที่สุด แต่กลับเป็นด้านที่กำหนดทิศทางการเรียนรู้ของผู้เรียนในระยะยาว

ส่วน เจตคติ (Attitude) คือ แนวโน้มของความรู้สึกหรือการตอบสนองของบุคคลต่อสิ่งเร้าหนึ่ง ๆ ในทิศทางบวกหรือลบ เจตคติจึงเป็นองค์ประกอบสำคัญของจิตพิสัยและเป็นตัวขับเคลื่อน (Driver) พฤติกรรมการเรียนรู้ด้านอื่น กล่าวคือ ผู้เรียนที่มีเจตคติที่ดีต่อวิชาใดวิชาหนึ่งย่อมมีแนวโน้มตั้งใจเรียนและฝึกฝนมากกว่าปกติ ขณะที่ ค่านิยม (Values) คือ สิ่งที่บุคคลเห็นว่ามีคุณค่าและยึดถือเป็นแนวปฏิบัติจนตกผลึกเป็นคุณลักษณะประจำตัว

นักจิตวิทยาอธิบายว่าเจตคติประกอบด้วยองค์ประกอบสามด้านที่สัมพันธ์กัน (Rosenberg & Hovland, 1960) ซึ่งช่วยให้ครูเข้าใจว่าการวัดเจตคติต้องครอบคลุมมากกว่าการถามความรู้สึก ดังกล่อง 6-1

### กล่องที่ 6.1 องค์ประกอบ 3 ด้านของเจตคติ (ABC Model of Attitude)

**เจตคติประกอบด้วย 3 องค์ประกอบที่สอดคล้องกัน:**

- ด้านความคิดความเชื่อ (Cognitive): สิ่งที่คุณรู้/เชื่อเกี่ยวกับสิ่งเร้า เช่น “คณิตศาสตร์มีประโยชน์ต่ออาชีพ”
  - ด้านความรู้สึก (Affective): อารมณ์ที่มีต่อสิ่งเร้า เช่น ชอบ/ไม่ชอบ สนุก/กังวลเมื่อเรียนคณิตศาสตร์
  - ด้านแนวโน้มพฤติกรรม (Behavioral): ความพร้อมจะกระทำ เช่น เลือกทำโจทย์เพิ่มหรือหลีกเลี่ยงวิชานั้น
- เครื่องมือวัดเจตคติที่ดีจึงควรมีข้อความครอบคลุมทั้งสามด้าน ไม่ใช่ถามเฉพาะความชอบเพียงด้านเดียว

ที่มา: ดัดแปลงจาก Rosenberg & Hovland (1960) และ พชรินทร์ ชมภูวิเศษ (2559, น. 37)

เพื่อให้เห็นความต่างของจิตพิสัยจากพฤติกรรมอีกสองด้านอย่างชัดเจน จึงสรุปเปรียบเทียบไว้ดังตารางที่ 6.1

### ตารางที่ 6.1 เปรียบเทียบพฤติกรรมการเรียนรู้ 3 ด้าน

| ด้านพฤติกรรม                | สาระสำคัญ/สิ่งที่มุ่งวัด                         | ตัวอย่างและเครื่องมือที่เหมาะสม                            |
|-----------------------------|--------------------------------------------------|------------------------------------------------------------|
| พุทธิพิสัย<br>(Cognitive)   | ความรู้ ความจำ ความเข้าใจ การคิดวิเคราะห์        | จำสูตร อธิบายหลักการ →<br>แบบทดสอบ/ข้อสอบ                  |
| ทักษะพิสัย<br>(Psychomotor) | การปฏิบัติ การเคลื่อนไหว ความชำนาญ               | ทดลองวิทยาศาสตร์ เล่นดนตรี →<br>แบบสังเกตการปฏิบัติ/รูปรिक |
| จิตพิสัย<br>(Affective)     | ความรู้สึก เจตคติ ค่านิยม คุณธรรม<br>ลักษณะนิสัย | ความซื่อสัตย์ เจตคติต่อวิชา →<br>แบบวัดเจตคติ/แบบสังเกต    |

ที่มา: ดัดแปลงจาก พชรินทร์ ชมภูวิเศษ (2559, น. 33–37) และ สุทธิพงศ์ บุญผดุง (2568)

ธรรมชาติสำคัญของจิตพิสัยคือเป็น การวัดทางอ้อม (Indirect Measurement) เพราะไม่อาจวัดความรู้สึกภายในได้โดยตรง ต้องนิยามคุณลักษณะที่ต้องการวัดให้เป็น “พฤติกรรมบ่งชี้” (Behavioral Indicator) ที่สังเกตได้เสียก่อน นอกจากนี้ผลการวัดยังมีลักษณะเป็น “ศูนย์สัมพัทธ์” (Relative Zero) กล่าวคือ การที่ผู้เรียนได้คะแนนจิตพิสัยต่ำหรือเป็นศูนย์ มิได้แปลว่าไร้คุณลักษณะนั้นโดยสิ้นเชิง (Absolute Zero) แต่

อาจเกิดจากเครื่องมือยังวัดไม่ถึงหรือผู้เรียนยังไม่พร้อม ครูจึงต้องระมัดระวังอย่างยิ่งในการแปลผลและตัดสินคุณค่าของผู้เรียน (สุทธิพงศ์ บุญผดุง, 2568, น. 2-3)

อีกข้อจำกัดที่ครูต้องตระหนักคือ “พฤติกรรมประดิษฐ์” (Constructed/Faked Behavior) คือผู้ถูกวัดอาจแสดงออกให้ดูดีกว่าความเป็นจริงเพื่อให้ได้คะแนนหรือเพื่อให้สังคมยอมรับ ด้วยเหตุนี้การวัดจิตพิสัยจึงควรทำอย่างต่อเนื่อง (Continuous) ใช้เครื่องมือหลายชนิด และเก็บข้อมูลแบบสามเส้า (Triangulation) จากครู เพื่อน และตัวผู้เรียน เพื่อยืนยันความคงเส้นคงวาของพฤติกรรมก่อนตัดสินตัวอย่างในชั้นเรียนไทย: งานวิจัยกับนักเรียนชั้นประถมศึกษาปีที่ 6 พบว่าหลังจัดการเรียนรู้แบบร่วมมือด้วยเทคนิค TGT ร่วมกับสื่อมัลติมีเดีย ผู้เรียนมีเจตคติต่อการเรียนวิชาคณิตศาสตร์ “อยู่ในระดับมาก” เมื่อวัดด้วยแบบวัดเจตคติแบบมาตรประมาณค่า 5 ระดับ จำนวน 24 ข้อ (กานต์พิชชา ไตรพัด และวไลพร เมฆไตรรัตน์, 2568) สะท้อนว่าเจตคติเป็นผลการเรียนรู้ที่วัดได้และเปลี่ยนแปลงได้จากการออกแบบการสอน

**ข้อควรระวัง:** ก่อนสร้างเครื่องมือทุกครั้ง ครูต้องเขียนนิยามเชิงปฏิบัติการของคุณลักษณะให้ชัดเจน และไม่ด่วนตัดสินผู้เรียนจากการสังเกตเพียงครั้งเดียว เพราะจิตพิสัยมีความแปรปรวนสูงตามสถานการณ์และอารมณ์

## 6.5.2 ระดับพฤติกรรมจิตพิสัยตามแนวคิดของแครธวอลและคณะ (Krathwohl Taxonomy)

แครธวอล บลูม และมาเซีย (Krathwohl, Bloom, & Masia, 1964) เสนอลำดับชั้นพฤติกรรมจิตพิสัยเป็น 5 ระดับ เรียงจากต่ำไปสูงตามกระบวนการ “กล่อมเกลายภายใน” (Internalization) คือการที่สิ่งเร้าหรือค่านิยมภายนอกค่อย ๆ ซึมเข้าเป็นส่วนหนึ่งของตัวบุคคล จนกลายเป็นบุคลิกภาพถาวรในที่สุด (พัชรินทร์ ชมภูวิเศษ, 2559; สุทธิพงศ์ บุญผดุง, 2568, น. 27-28) แต่ละระดับยังจำแนกเป็นชั้นย่อยที่ละเอียดขึ้น ดังภาพที่ 6.1 และตารางที่ 6.2

ภาพที่ 6.1 บันไดระดับพฤติกรรมจิตพิสัย 5 ชั้น (อ่านจากชั้นสูงสุดลงล่าง)

| ชั้น                                   | แก่นของพฤติกรรม (ยิ่งสูงยิ่งเป็นเนื้อแท้ของบุคคล)    |
|----------------------------------------|------------------------------------------------------|
| 5. สร้างลักษณะนิสัย (Characterization) | ค่านิยมหล่อหลอมเป็นบุคลิกภาพถาวร คาดการณ์พฤติกรรมได้ |
| 4. จัดระบบคุณค่า (Organization)        | จัดลำดับ/เชื่อมโยงค่านิยมหลายอย่างให้เป็นระบบของตน   |
| 3. เห็นคุณค่า (Valuing)                | ยอมรับและเชื่อมั่นในคุณค่า แสดงออกอย่างสม่ำเสมอ      |
| 2. ตอบสนอง (Responding)                | มีปฏิกิริยาโต้ตอบและพึงพอใจที่จะทำตาม                |
| 1. รับรู้ (Receiving)                  | ตระหนัก เต็มใจรับฟัง และเลือกสนใจสิ่งเร้า            |

ที่มา: ผู้เขียน เรียบเรียงจาก Krathwohl, Bloom, & Masia (1964) และ พัชรินทร์ ชมภูวิเศษ (2559, น. 37-39)

เมื่อพิจารณารายชั้นพร้อมชั้นย่อยและตัวอย่างที่สังเกตได้จริงในชั้นเรียน จะช่วยให้ครูเขียนจุดประสงค์และออกแบบการประเมินได้ตรงระดับ ดังตารางที่ 6.2

ตารางที่ 6.2 ระดับพฤติกรรมจิตพิสัย 5 ชั้น ชั้นย่อย และตัวอย่างในชั้นเรียนไทย

| ระดับ (ชั้นย่อย)                                                                   | ความหมายสำคัญ                               | ตัวอย่างพฤติกรรมที่สังเกตได้                                    |
|------------------------------------------------------------------------------------|---------------------------------------------|-----------------------------------------------------------------|
| 1. การรับรู้ (Receiving)<br>- ตระหนัก<br>- เต็มใจรับ<br>- เลือกใส่ใจ               | เริ่มมีความตระหนัก สนใจ และเลือกรับสิ่งเร้า | นักเรียนหยุดคุยและตั้งใจฟังเมื่อครูเริ่มสาธิตการทดลอง           |
| 2. การตอบสนอง (Responding)<br>- ยินยอม<br>- เต็มใจ<br>- ฟังพอใจ                    | มีปฏิกิริยาโต้ตอบและพอใจที่จะทำตาม          | อาสาตอบคำถาม ช่วยเก็บอุปกรณ์ด้วยความเต็มใจ                      |
| 3. การเห็นคุณค่า (Valuing)<br>- ยอมรับ<br>- ชื่นชอบ<br>- เชื่อมั่น                 | ยอมรับคุณค่าและยึดมั่นแสดงออกสม่ำเสมอ       | ทักท้วงเมื่อเพื่อนทิ้งขยะ เพราะเชื่อในความสะอาด                 |
| 4. การจัดระบบคุณค่า (Organization)<br>- สร้างมโนทัศน์คุณค่า<br>- จัดระบบคุณค่า     | จัดลำดับและเชื่อมโยงค่านิยมให้เป็นระบบ      | เลือกสละเวลาเล่นเกมมาทำการบ้านเพื่อรักษาเป้าหมายการเรียนรู้     |
| 5. การสร้างลักษณะนิสัย (Characterization)<br>- สรุปนัยทั่วไป<br>- เป็นบุคลิกลักษณะ | ค่านิยมกลายเป็นนิสัยถาวรคาดการณ์ได้         | ซื่อสัตย์เป็นปกติแม้ไม่มีใครสังเกต ปฏิบัติตามจรรยาบรรณเป็นนิสัย |

ที่มา: ดัดแปลงจาก Krathwohl, Bloom, & Masia (1964) และ พัทธินันท์ ชมภูวิเศษ (2559, น. 37-39)

ระดับที่ 5 ถือเป็นเป้าหมายสูงสุด เพราะพฤติกรรมได้กลายเป็นลักษณะนิสัยถาวรที่คาดการณ์ได้ ซึ่งสอดคล้องกับเป้าหมายของการจัดการศึกษาอิงผลลัพธ์ (OBE) ที่มุ่งสร้างคุณลักษณะอันยั่งยืนในตัวผู้เรียน (Anderson & Krathwohl, 2001) ทั้งนี้ระดับชั้นจิตพิสัยมีความสัมพันธ์คู่ขนานกับระดับพุทธิพิสัย กล่าวคือ

การจะ “เห็นคุณค่า” ของสิ่งใดมักต้องมีความรู้ความเข้าใจสิ่งนั้นเป็นฐานก่อน (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

**การประยุกต์ใช้:** ครูสามารถใช้ลำดับขั้นนี้เป็นกรอบเขียนจุดประสงค์เชิงพฤติกรรมด้านจิตพิสัยไล่จากระดับต่ำไปสูง และเลือกกริยาแสดงพฤติกรรมให้ตรงระดับ เช่น “เต็มใจเข้าร่วม” (ตอบสนอง) ต่างจาก “ยึดถือเป็นแนวปฏิบัติประจำ” (สร้างลักษณะนิสัย) ข้อควรระวัง: ไม่ควรคาดหวังให้ผู้เรียนกระโดดถึงระดับสูงสุดในคาบเดียว เพราะการกล่อมเกลายภายในต้องอาศัยเวลาและการปลูกฝังอย่างต่อเนื่อง

### 6.5.3 ความสำคัญของจิตพิสัยต่อการศึกษอิงผลลัพธ์ (OBE) และวิชาชีพครู

ในกระบวนการจัดการศึกษาอิงผลลัพธ์ (Outcome-Based Education: OBE) การประเมินไม่ได้สิ้นสุดเพียงความรู้หรือทักษะปฏิบัติ แต่ให้ความสำคัญกับจิตพิสัยในฐานะ รากฐานของความยั่งยืนในตัวผู้เรียน เพราะคุณลักษณะ เจตคติ และค่านิยมที่ดีจะคงอยู่และส่งผลต่อการเรียนรู้และการดำรงชีวิตในระยะยาวมากกว่าความรู้เฉพาะเรื่อง (สุทธิพงศ์ บุญผดุง, 2568, น. 27)

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จึงกำหนด “คุณลักษณะอันพึงประสงค์” 8 ประการ ที่สถานศึกษาทุกแห่งต้องปลูกฝังและประเมินควบคู่กับผลสัมฤทธิ์ทางวิชาการ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553) คุณลักษณะเหล่านี้คือเป้าหมายจิตพิสัยที่เป็นรูปธรรมในบริบทไทย ดังตารางที่ 6.3 ซึ่งเชื่อมโยงแต่ละคุณลักษณะกับระดับพฤติกรรมจิตพิสัยที่มุ่งหวัง

ตารางที่ 6.3 คุณลักษณะอันพึงประสงค์ 8 ประการ เชื่อมโยงกับระดับจิตพิสัยและพฤติกรรมบ่งชี้

| คุณลักษณะอันพึงประสงค์    | ระดับจิตพิสัยที่มุ่งหวัง    | ตัวอย่างพฤติกรรมบ่งชี้ที่สังเกตได้            |
|---------------------------|-----------------------------|-----------------------------------------------|
| 1. รักชาติ ศาสน์ กษัตริย์ | เห็นคุณค่า-สร้างลักษณะนิสัย | เข้าร่วมกิจกรรมของชาติด้วยความเต็มใจ สม่ำเสมอ |
| 2. ซื่อสัตย์สุจริต        | สร้างลักษณะนิสัย            | ไม่ทุจริตในการสอบแม้ไม่มีผู้ควบคุม            |
| 3. มีวินัย                | จัดระบบคุณค่า               | ส่งงานตรงเวลาและปฏิบัติตามกติกาห้องเรียน      |
| 4. ใฝ่เรียนรู้            | ตอบสนอง-เห็นคุณค่า          | ค้นคว้าเพิ่มเติมและซักถามด้วยความกระตือรือร้น |
| 5. อยู่อย่างพอเพียง       | จัดระบบคุณค่า               | วางแผนใช้ทรัพยากร/เวลาอย่างคุ้มค่า            |
| 6. มุ่งมั่นในการทำงาน     | เห็นคุณค่า-จัดระบบ          | ทำงานที่ได้รับมอบหมายจนสำเร็จแม้มีอุปสรรค     |

|                   |                          |                                      |
|-------------------|--------------------------|--------------------------------------|
| 7. รักความเป็นไทย | เห็นคุณค่า               | ใช้ภาษาไทยถูกต้องและภูมิใจในวัฒนธรรม |
| 8. มีจิตสาธารณะ   | ตอบสนอง-สร้างลักษณะนิสัย | อาสาช่วยเหลือส่วนรวมโดยไม่ต้องร้องขอ |

ที่มา: ดัดแปลงจาก สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (2553) และจัดระดับตามแนวคิดของ Krathwohl, Bloom, & Masia (1964)

**ตัวอย่างในชั้นเรียนไทย:** โรงเรียนทั่วไปจะประเมินคุณลักษณะอันพึงประสงค์ตามสภาพจริงตลอดภาคเรียน โดยครูประจำวิชาและครูที่ปรึกษาร่วมกันสังเกตและบันทึก แล้วสรุปผลเป็นระดับ “ดีเยี่ยม-ดี-ผ่าน-ไม่ผ่าน” ผู้เรียนต้องผ่านเกณฑ์ทุกข้อจึงจะจบช่วงชั้น สะท้อนว่าจิตพิสัยเป็นผลการเรียนรู้ที่มีน้ำหนักจริงในระบบการศึกษาไทย

**ความสำคัญต่อวิชาชีพครูและ CLO5:** มาตรฐานวิชาชีพครูกำหนดให้ครูต้องมีเจตคติที่ดีต่อวิชาชีพและประเมินผู้เรียนด้วยความเที่ยงธรรม การที่นักศึกษาครูเข้าใจและสร้างเครื่องมือวัดจิตพิสัยได้จึงเป็นสมรรถนะที่รับใช้ทั้งการพัฒนาผู้เรียนและการเป็นแบบอย่างด้านคุณธรรม ข้อควรระวัง: ต้องประเมินคุณลักษณะด้วยหลักฐานพฤติกรรมที่สังเกตได้ มิใช่ความประทับใจส่วนตัวของครู มิฉะนั้นผลการประเมินจะขาดความเที่ยงธรรม

#### 6.5.4 เครื่องมือวัดด้านจิตพิสัยและเจตคติ

เนื่องจากจิตพิสัยเป็นนามธรรมและวัดทางอ้อม จึงควร ใช้เครื่องมือหลายชนิดร่วมกันและวัดอย่างต่อเนื่อง เพื่อเพิ่มความน่าเชื่อถือและลดอคติ (สำเร็จ บุญเรืองรัตน์ และคณะ, 2534; พิชิต ฤทธิ์จรูญ, 2564; ล้วน สายยศ และอังคณา สายยศ, 2543) พชรินทร์ ชมภูวิเศษ (2559, น. 114–128) จำแนกเครื่องมือวัดจิตพิสัยที่ใช้บ่อยไว้หลายชนิด แต่ละชนิดเหมาะกับพฤติกรรมที่ต่างกัน สรุปจุดเด่น-ข้อจำกัดได้ดังตารางที่ 6.4

ตารางที่ 6.4 เครื่องมือวัดจิตพิสัย: จุดเด่น ข้อจำกัด และความเหมาะสม

| เครื่องมือ                   | จุดเด่น                                    | ข้อจำกัด/ข้อควรระวัง                             | เหมาะกับการวัด                       |
|------------------------------|--------------------------------------------|--------------------------------------------------|--------------------------------------|
| แบบสอบถาม (Questionnaire)    | เก็บข้อมูลจำนวนมาก<br>ประมวลผลทางสถิติง่าย | อาจได้คำตอบที่ “ควรจะเป็น” (Social Desirability) | ความคิดเห็น ความสนใจ เจตคติกลุ่มใหญ่ |
| แบบตรวจสอบรายการ (Checklist) | ใช้ง่าย ระบุการมี/ไม่มี<br>พฤติกรรมชัดเจน  | ไม่บอกระดับหรือคุณภาพ<br>ของพฤติกรรม             | พฤติกรรมพื้นฐานที่ตอบ มี/ไม่มี       |
| มาตรประมาณค่า (Rating Scale) | ประเมินระดับความเข้ม/<br>ความถี่ได้ละเอียด | เกิดอคติผู้ประเมิน (Halo/รวมศูนย์) ได้           | เจตคติ คุณลักษณะที่มีระดับ           |

|                                      |                                           |                                       |                                     |
|--------------------------------------|-------------------------------------------|---------------------------------------|-------------------------------------|
| แบบวัดเชิงสถานการณ์<br>(Situational) | ลดการเสแสร้ง<br>วัดการตัดสินใจในบริบทจริง | สร้างยาก<br>ต้องตรวจคุณภาพรอบคอบ      | ค่านิยม/จริยธรรมเชิง<br>การตัดสินใจ |
| การสังเกต<br>(Observation)           | ได้พฤติกรรมจริงตามสภาพ<br>ธรรมชาติ        | ใช้เวลานาน<br>ผู้สังเกตอาจมีอคติ      | พฤติกรรมที่แสดง<br>ออกในชั้นเรียน   |
| การสัมภาษณ์<br>(Interview)           | เจาะลึกเหตุผล/ความรู้สึก<br>เบื้องลึก     | ใช้เวลามาก<br>ขึ้นกับทักษะผู้สัมภาษณ์ | แรงจูงใจ<br>ความเชื่อเชิงลึก        |
| ประเมินตนเอง/โดย<br>เพื่อน           | ส่งเสริมการสะท้อนคิด<br>ได้มุมมองใกล้ชิด  | ลำเอียงเข้าข้างตน/เกรงใจ<br>กัน       | การกำกับตนเอง<br>พฤติกรรมกลุ่ม      |

ที่มา: ดัดแปลงจาก พัทรินทร์ ชมภูวพิเศษ (2559, น. 114–128), สมนึก ภัททิยธนี (2558) และ สุทธิพงศ์ บุญผดุง (2568, น. 30)

ในบรรดาเครื่องมือเหล่านี้ แบบวัดเชิงสถานการณ์ (Situational Test) เป็นทางเลือกที่ช่วยลดพฤติกรรมประดิษฐ์ได้ดี เพราะกำหนดสถานการณ์ให้ผู้ตอบเลือกตอบสนอง แล้วอนุมานค่านิยม/คุณลักษณะจากการตัดสินใจ ตัวอย่างเช่น งานวิจัยพัฒนา “แบบวัดความสามารถทางวิชาชีพครู” ชนิดสถานการณ์ 3 ตัวเลือก จำนวน 30 ข้อ ซึ่งมีค่าความเชื่อมั่นทั้งฉบับเท่ากับ .731 และจำแนกผู้ที่เรียนวิชาชีพครูออกจากผู้ที่ไม่ได้เรียนได้อย่างมีนัยสำคัญ (อุไร จักร์ตรึงมล, 2559) แสดงให้เห็นว่าคุณลักษณะเชิงจิตพิสัยสามารถวัดได้อย่างเป็นระบบเมื่อออกแบบเครื่องมืออย่างถูกต้อง

**การประยุกต์ใช้และข้อควรระวัง:** ครูควรเลือกเครื่องมือให้ตรงกับธรรมชาติของพฤติกรรมและวัยผู้เรียน แล้วใช้หลายเครื่องมือประกอบกันแบบสามเส้า โดยเฉพาะเมื่อใช้แบบรายงานตนเอง (Self-report) ต้องระวัง ความลำเอียงจากการตอบให้ดูดี (Social Desirability Bias) ซึ่งเป็นแนวโน้มที่ผู้ตอบปรับคำตอบเพื่อสร้างภาพลักษณ์ที่ดี การไม่ระบุชื่อผู้ตอบและการใช้ข้อความเชิงพฤติกรรมที่เป็นรูปธรรมจะช่วยลดอคตินี้ได้

### 6.5.5 มาตรฐานค่าและการสร้างแบบวัดเจตคติแบบลิเคิร์ต

มาตรฐานค่าแบบลิเคิร์ต (Likert Scale) หรือวิธีประมาณค่ารวม (Method of Summated Rating) เป็นเครื่องมือวัดเจตคติที่นิยมที่สุด พัฒนาโดยเรนชิส ลิเคิร์ต (Likert, 1932) ประกอบด้วยชุดข้อความแสดงความคิดเห็นทั้งทางบวกและทางลบ ให้ผู้ตอบเลือกระดับความเห็นของตน แล้วนำคะแนนทุกข้อมารวมกันเป็นคะแนนเจตคติ นิยมใช้ 5 ระดับ คือ เห็นด้วยมากที่สุด-มาก-ปานกลาง-น้อย-น้อยที่สุด (Likert, 1932; พัทรินทร์ ชมภูวพิเศษ, 2559) เพื่อให้เห็นภาพเครื่องมือจริง ตารางที่ 6.5 แสดงตัวอย่างแบบวัดเจตคติต่อวิชาชีพครูแบบลิเคิร์ต 5 ระดับ ที่มีทั้งข้อความทางบวกและทางลบ

## ตารางที่ 6.5 ตัวอย่างแบบวัดเจตคติต่อวิชาชีพครู (มาตรลิเคิร์ท 5 ระดับ)

| ข้อความ                                     | ทิศทาง | มากที่สุด | มาก | ปานกลาง | น้อย | น้อยที่สุด |
|---------------------------------------------|--------|-----------|-----|---------|------|------------|
| 1. ฉันภูมิใจที่ได้เลือกเรียนวิชาชีพครู      | บวก    | 5         | 4   | 3       | 2    | 1          |
| 2. การเป็นครูทำให้ฉันมีความสุข              | บวก    | 5         | 4   | 3       | 2    | 1          |
| 3. ฉันรู้สึกเบื่อหน่ายเมื่อต้องเตรียมการสอน | ลบ     | 1         | 2   | 3       | 4    | 5          |
| 4. ฉันคิดว่างานครูเป็นภาระที่หนักเกินไป     | ลบ     | 1         | 2   | 3       | 4    | 5          |
| 5. ฉันตั้งใจพัฒนาตนเองสู่การเป็นครูที่ดี    | บวก    | 5         | 4   | 3       | 2    | 1          |

ที่มา: ผู้เขียน (ออกแบบตามหลักการของ Likert, 1932 และ พิชรินทร์ ชมภูวิเศษ, 2559, น. 115–116)

จากตาราง สังเกตว่าข้อความทางลบ (ข้อ 3–4) ต้องให้คะแนน “กลับทาง” (Reverse/Reciprocal Scoring) กล่าวคือ ผู้ที่ตอบ “เห็นด้วยมากที่สุด” กับข้อความทางลบจะได้เพียง 1 คะแนน เพื่อให้ทิศทางคะแนนของทุกข้อชี้ไปทาง “เจตคติที่ดี” เหมือนกันก่อนนำมารวม การคำนวณและการแปลผลคะแนนเฉลี่ยสรุปไว้ในกล่องที่ 6.2

### กล่องที่ 6.2 การให้คะแนน การกลับคะแนน และเกณฑ์แปลผลค่าเฉลี่ย

#### ขั้นการให้คะแนนและแปลผลแบบลิเคิร์ท 5 ระดับ

- 1) ให้คะแนนข้อความบวก  $5 \rightarrow 1$  และข้อความลบกลับทางเป็น  $1 \rightarrow 5$
  - 2) รวมคะแนนทุกข้อ แล้วหาค่าเฉลี่ย (Mean) ต่อคน หรือเฉลี่ยรายข้อ
  - 3) กำหนดความกว้างอันตรภาคชั้น =  $(\text{คะแนนสูงสุด} - \text{ต่ำสุด}) \div \text{จำนวนระดับ} = (5 - 1) \div 5 = 0.80$
- เกณฑ์แปลผล: 4.21–5.00 = เจตคติดีมาก | 3.41–4.20 = ดี | 2.61–3.40 = ปานกลาง | 1.81–2.60 = ควรปรับปรุง | 1.00–1.80 = ต้องปรับปรุงเร่งด่วน

ที่มา: ผู้เขียน คำนวณตามหลักการมาตรฐานค่า (พิชรินทร์ ชมภูวิเศษ, 2559, น. 115–116)

จำนวนระดับควรมีเท่าใด? งานทบทวนวรรณกรรม 60 เรื่องของ Kusmaryono, Wijayanti, และ Maharani (2022) พบว่า รวบรวม 90 ของงานวิจัยใช้จำนวนระดับเป็นเลขคี่ (5, 7, 9) และสรุปว่ามาตรเลขคี่ที่มากกว่า 5 ระดับ (โดยเฉพาะ 7 ระดับ) ให้ค่าความเที่ยงและความตรงสูงที่สุด ส่วนมาตรเลขคู่ (เช่น 6 ระดับ)

เหมาะเมื่อต้องการ “บังคับ” ให้ผู้ตอบเลือกทิศทางบวกหรือลบโดยไม่มีตัวเลือกกลาง สำหรับเด็กเล็กอาจใช้เพียง 3 ระดับเพื่อความเข้าใจง่าย (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

นอกจากลิเคิร์ทแล้ว แบบนัยจำแนก (Semantic Differential) ของออสกู๊ดและคูนะ (Osgood, Suci, & Tannenbaum, 1957) ใช้คู่คำคุณศัพท์ที่มีความหมายตรงข้ามกัน เช่น ดี-เลว สนุก-น่าเบื่อ โดยมี 5-7 ช่องให้ขีดเลือก จำแนกเป็น 3 มิติ คือ ด้านประเมินค่า (Evaluation: ดี-เลว) ด้านศักยภาพ (Potency: ใหญ่-เล็ก) และด้านกิจกรรม (Activity: เร็ว-ช้า) เหมาะกับการวัดความหมายเชิงความรู้สึกที่ผู้เรียนมีต่อสิ่งหนึ่ง (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

ขั้นตอนการสร้างแบบวัดเจตคติที่มีคุณภาพมี 5 ขั้นตอน ดังนี้

- 1) กำหนดคุณลักษณะ/เป้าหมายที่จะวัดให้ชัดเจน และเขียนนิยามเชิงปฏิบัติการของสิ่งที่วัด
- 2) เขียนข้อความให้ครอบคลุมทุกองค์ประกอบ มีทั้งทางบวกและทางลบในสัดส่วนใกล้เคียงกัน
- 3) ทดลองใช้กับกลุ่มผู้เรียน แล้ววิเคราะห์คุณภาพรายข้อด้วยอำนาจจำแนก (t-test กลุ่มสูง-ต่ำ หรือ สหสัมพันธ์ข้อกับคะแนนรวม)
- 4) ตรวจสอบความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญ (ค่า IOC) และหาความเที่ยงด้วยสัมประสิทธิ์แอลฟาของครอนบาค
- 5) จัดทำแบบวัดฉบับสมบูรณ์ พร้อมเกณฑ์การให้คะแนน การกลับคะแนน และคู่มือการใช้หลักการเขียนข้อความที่ดี (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566): เกี่ยวข้องกับเรื่องที่วัด มีความหมายชัดเจนไม่กำกวม มีเพียงประเด็นความคิดเห็นเดียว เป็นปัจจุบัน และหลีกเลี่ยงคำรอบจักรวาล เช่น “เสมอ” “ทั้งหมด” “เท่านั้น” ที่ผู้ตอบตีความต่างกัน

**ตัวอย่างในชั้นเรียนไทย:** งานวิจัยของกานต์พิชชา ไตรพัต และวไลพร เมฆไตรรัตน์ (2568) ใช้แบบวัดเจตคติต่อการเรียนวิชาคณิตศาสตร์แบบมาตรประมาณค่า 5 ระดับ จำนวน 24 ข้อ กับนักเรียนชั้นประถมศึกษาปีที่ 6 และรายงานผลเป็นค่าเฉลี่ยที่แปลความได้ว่าเจตคติ “อยู่ในระดับมาก” ซึ่งเป็นแนวปฏิบัติมาตรฐานในการวัดเจตคติของไทย ข้อควรระวัง: อย่าลืมกลับคะแนนข้อความทางลบก่อนรวม และตรวจสอบว่าทุกข้อวัดเจตคติเรื่องเดียวกันจริง มิฉะนั้นคะแนนรวมจะตีความไม่ได้

#### 6.5.6 การสังเกตและการประเมินคุณลักษณะอันพึงประสงค์

การประเมินคุณลักษณะอันพึงประสงค์ใช้ การสังเกต (Observation) ตามสภาพจริงเป็นหลัก จำแนกได้เป็นการสังเกตแบบมีส่วนร่วม/ไม่มีส่วนร่วม และแบบมีโครงสร้าง/ไม่มีโครงสร้าง การสังเกตที่ดีต้องอาศัย

ความตั้งใจที่ปราศจากอคติ ผู้สังเกตมีสภาพพร้อม และมีเครื่องมือบันทึกที่ชัดเจน (สุทธิพงศ์ บุญผดุง, 2568, น. 30; พชรินทร์ ชมภูวิเศษ, 2559) ครูควรเลือกเครื่องมือบันทึกการสังเกตให้เหมาะสมกับธรรมชาติของพฤติกรรม โดยเปรียบเทียบได้ดังตารางที่ 6.6

ตารางที่ 6.6 เปรียบเทียบเครื่องมือบันทึกการสังเกตพฤติกรรม

| เครื่องมือ                         | สิ่งที่บันทึก                    | เหมาะกับ                                               |
|------------------------------------|----------------------------------|--------------------------------------------------------|
| แบบตรวจสอบรายการ (Checklist)       | มี/ไม่มีพฤติกรรม (ทำเครื่องหมาย) | พฤติกรรมพื้นฐานที่ชัดเจน เช่น ส่งงาน แต่งกายถูกระเบียบ |
| มาตราประมาณค่า (Rating Scale)      | ระดับความถี่/คุณภาพ (มาก-น้อย)   | คุณลักษณะที่มีระดับ เช่น ความรับผิดชอบ การมีส่วนร่วม   |
| ระเบียนพฤติกรรม (Anecdotal Record) | บรรยายเหตุการณ์ตามข้อเท็จจริง    | พฤติกรรมเด่น/เหตุการณ์สำคัญที่ไม่คาดคิด                |

ที่มา: ดัดแปลงจาก พชรินทร์ ชมภูวิเศษ (2559, น. 117–126) และ สมนึก ภัททิยธนี (2558)

### กล่องที่ 6.3 หลักการเขียนระเบียนพฤติกรรม (Anecdotal Record)

#### บันทึก “ข้อเท็จจริง” แยกจาก “การตีความ”

ตัวอย่างที่ดี (ข้อเท็จจริง): “วันที่ 5 มิ.ย. เวลาพักกลางวัน ด.ช.ก. เก็บขยะบริเวณสนามใส่ถังโดยไม่มีผู้ใดร้องขอ”

ตัวอย่างที่ควรเลี่ยง (ตีความล่วงหน้า): “ด.ช.ก. เป็นเด็กดีมีจิตสาธารณะ” — เป็นการสรุปก่อนมีหลักฐานพอ

หลักปฏิบัติ: บันทึกทันทีหลังเหตุการณ์ ระบุนวัน-เวลา-บริบท เก็บหลายเหตุการณ์ก่อนสรุป เพื่อให้การตัดสินมีหลักฐานเชิงประจักษ์

ที่มา: ผู้เขียน เรียบเรียงจาก สมนึก ภัททิยธนี (2558) และ สุทธิพงศ์ บุญผดุง (2568)

**ตัวอย่างในชั้นเรียนไทย:** การประเมิน “จิตสาธารณะ” ครูอาจกำหนดพฤติกรรมบ่งชี้ 4–6 ข้อ (เช่น อาสาช่วยงานส่วนรวม แบ่งปันอุปกรณ์ ดูแลความสะอาด) แล้วใช้มาตราประมาณค่า 3 ระดับบันทึกตลอดภาคเรียน ประกอบกับระเบียนพฤติกรรมเมื่อพบเหตุการณ์เด่น ข้อควรระวัง: การสังเกตต้องใช้เวลาและหลายโอกาสจึงจะได้ข้อมูลที่เที่ยงตรง ครูควรหลีกเลี่ยงการสรุปจากเหตุการณ์เดียวและการปล่อยให้ความประทับใจส่วนตัวครอบงำการบันทึก

### 6.5.7 ความคลาดเคลื่อน อคติ และแนวทางลดอคติในการประเมินจิตพิสัย

การวัดจิตพิสัยมีโอกาสคลาดเคลื่อน (Error) สูงจากทั้งสองฝั่ง คือ ฝั่งผู้ถูกวัด (เช่น ตอบให้ดูดีตามที่สังคมคาดหวัง) และ ฝั่งผู้ประเมิน (เช่น ลำเอียงจากความประทับใจส่วนตัว) ครูจึงต้องรู้จักประเภทของอคติและวิธีป้องกันอย่างเป็นระบบ (Nitko & Brookhart, 2019; พิชิต ฤทธิ์จรูญ, 2564) อคติที่พบบ่อยและแนวทางลดสรุปไว้ในตารางที่ 6.7

ตารางที่ 6.7 ประเภทของอคติในการประเมินจิตพิสัยและแนวทางลด

| ประเภทอคติ/ความคลาดเคลื่อน                    | ลักษณะปัญหา                                                                | แนวทางลดอคติ                                               |
|-----------------------------------------------|----------------------------------------------------------------------------|------------------------------------------------------------|
| ความลำเอียงจากความประทับใจ (Halo Effect)      | ตัดสินทุกด้านตามภาพรวมที่ประทับใจ เช่น เด็กเรียนเก่งได้คะแนนจิตพิสัยสูงตาม | ใช้รูปрик/BARS อิงพฤติกรรมเฉพาะ ประเมินทีละคุณลักษณะ       |
| ความลำเอียงเข้าสู่ส่วนกลาง (Central Tendency) | ให้คะแนนกลาง ๆ ทุกคนเพื่อความปลอดภัย                                       | ใช้มาตรเลขคู่ (4/6 ระดับ) หรือ BARS เพื่อบังคับเลือกทิศทาง |
| ความใจดี/เข้มงวดเกินไป (Leniency/Severity)    | ให้คะแนนสูงหรือต่ำกว่าจริงอย่างเินระบบ                                     | ฝึกผู้ประเมินด้วยตัวอย่างมาตรฐาน เทียบเกณฑ์ร่วมกัน         |
| ความลำเอียงเชิงตรรกะ (Logical Error)          | เหมาเอาว่าคุณลักษณะที่ดูคล้ายกันต้องไปด้วยกัน                              | นิยามแต่ละคุณลักษณะแยกชัด มีพฤติกรรมบ่งชี้เฉพาะ            |
| ตอบให้ดูดี (Social Desirability)              | ผู้ตอบปรับคำตอบให้สังคมยอมรับ                                              | ไม่ระบุชื่อผู้ตอบ ใช้การสังเกตทางอ้อม/หลายแหล่ง            |
| แนวโน้มตอบซ้ำทาง (Acquiescence)               | มักเลือก “เห็นด้วย” โดยไม่อ่านละเอียด                                      | สลับข้อความบวก-ลบ มีข้อกลับทางคอยตรวจสอบ                   |

ที่มา: ดัดแปลงจาก Nitko & Brookhart (2019) และ พชรินทร์ ชมภูวิเศษ (2559)

เครื่องมือสำคัญที่ช่วยลดอคติของผู้ประเมินได้ดีคือ มาตรประมาณค่าที่ผูกกับพฤติกรรม (Behaviorally Anchored Rating Scale: BARS) ซึ่งเสนอโดยสมิธและเคนดอลล์ (Smith & Kendall, 1963) โดยกำหนดคำบรรยายพฤติกรรมเฉพาะกำกับแต่ละระดับคะแนน ทำให้ผู้ประเมินหลายคนตีความตรงกัน ดังตัวอย่างในกล่องที่ 6.4

#### กล่องที่ 6.4 ตัวอย่างมาตรประมาณค่าแบบ BARS: คุณลักษณะ “ความรับผิดชอบต่องานกลุ่ม”

แต่ละระดับผูกกับพฤติกรรมที่สังเกตได้ชัดเจน

ระดับ 4 (ดีเยี่ยม): ทำส่วนของตนครบถ้วนตรงเวลา และช่วยเพื่อนที่ติดขัดจนงานกลุ่มสำเร็จ

ระดับ 3 (ดี): ทำส่วนของตนครบถ้วนตรงเวลาตามที่ได้รับมอบหมาย

ระดับ 2 (พอใช้): ทำส่วนของตนแต่ล่าช้าหรือไม่ครบ ต้องมีผู้เตือน

ระดับ 1 (ควรปรับปรุง): ไม่ทำส่วนของตน ปล่อยให้เป็นการของเพื่อน

ที่มา: ผู้เขียน (ออกแบบตามแนวคิด BARS ของ Smith & Kendall, 1963)

ตัวอย่างในชั้นเรียนไทย: เมื่อครูพบว่าตนเองให้คะแนนการมีส่วนร่วม “ระดับ 3” กับนักเรียนเกือบทุกคน (อาการ Central Tendency) การเปลี่ยนมาใช้ BARS ข้างต้นจะบังคับให้ครูเทียบพฤติกรรมจริงกับคำบรรยายแต่ละระดับ ทำให้คะแนนสะท้อนความแตกต่างของผู้เรียนได้ตรงขึ้น ข้อควรระวัง: รูปรีค/ BARS จะลดอคติได้จริงก็ต่อเมื่อครูฝึกใช้และเทียบเกณฑ์ให้เข้าใจตรงกันก่อนประเมินจริง

#### 6.5.8 คุณธรรม จริยธรรม และความเที่ยงธรรมในการประเมินจิตพิสัย (CLO5)

ครูในฐานะ ผู้ตัดสินคุณค่าทางการศึกษา ต้องตระหนักว่าผลการประเมินจิตพิสัยกระทบต่อจิตใจ ภาวะภูมิใจ และอนาคตของผู้เรียนโดยตรง การประเมินที่ไม่เที่ยงธรรมอาจนำไปสู่การ “ติดตรา” (Labeling) ที่บั่นทอนผู้เรียนอย่างรุนแรง พัทธินทร์ ชมภูวิเศษ (2559, น. 19–21) จึงระบุคุณธรรมของผู้ทำหน้าที่วัดและประเมินผลไว้ 10 ประการ ดังตารางที่ 6.8

#### ตารางที่ 6.8 คุณธรรม 10 ประการของผู้วัดและประเมินผล

| คุณธรรม                   | สาระสำคัญ                                          |
|---------------------------|----------------------------------------------------|
| 1. ความรับผิดชอบ          | ดำเนินการวัดผลอย่างรอบคอบและรับผิดชอบต่อผล         |
| 2. ความซื่อสัตย์สุจริต    | ไม่บิดเบือนหรือปลอมแปลงผลการประเมิน                |
| 3. มีวินัยในตนเอง         | ปฏิบัติตามระเบียบและกำหนดเวลาอย่างเคร่งครัด        |
| 4. ความขยันอดทน           | ทุ่มเทเก็บข้อมูลอย่างต่อเนื่องแม้ใช้เวลานาน        |
| 5. ความยุติธรรม           | ตัดสินด้วยเกณฑ์เดียวกัน ไม่เห็นแก่ผลประโยชน์ส่วนตน |
| 6. เจตคติที่ดีต่อการวัดผล | มองการวัดผลเป็นเครื่องมือพัฒนา ไม่ใช่การลงโทษ      |
| 7. ความใจกว้าง            | ยอมรับความเห็นต่างเพื่อปรับปรุงงาน                 |
| 8. มีมนุษยสัมพันธ์        | ประสานความร่วมมือกับผู้เกี่ยวข้องด้วยดี            |
| 9. แสวงหาความรู้          | พัฒนาวิชาชีพและใช้หลักวิชาที่ทันสมัย               |
| 10. รักษาความลับ          | ปกปิดข้อมูลและผลการประเมินเพื่อคุ้มครองผู้เรียน    |

ที่มา: ดัดแปลงจาก พัทธินทร์ ชมภูวิเศษ (2559, น. 19–21)

หลักจรรยาบรรณสำคัญในการประเมินจิตพิสัย ได้แก่ ความโปร่งใส (ผู้เรียนทราบเกณฑ์ล่วงหน้า) ความเป็นธรรมและไม่เลือกปฏิบัติ (ปราศจากอคติทางเพศ ฐานะ หรือภูมิหลัง) การรักษาความลับของข้อมูล และ การใช้ผลเพื่อ “พัฒนา” มิใช่เพื่อ “ตีตรา” ผู้เรียน (พิชิต ฤทธิ์จรรยา, 2564; Nitko & Brookhart, 2019)

ศาสตราจารย์ ดร.ชวาล แพรัตกุล ผู้วางรากฐานการวัดผลการศึกษาของไทย เสนอหลักการสำคัญว่า การวัดผลเป็นส่วนหนึ่งของการเรียนการสอน ผู้วัดผลควรเป็นผู้สอนเอง และไม่ควรประเมินเมื่อไม่ประสงค์จะนำผลไปใช้ให้เกิดประโยชน์ หลักการเหล่านี้ตอกย้ำว่าการประเมินจิตพิสัยต้องกระทำเพื่อพัฒนาผู้เรียนด้วยความเที่ยงธรรม มิใช่เพื่อจับผิดหรือตีตรา (ชวาล แพรัตกุล, 2518)

## กล่องที่ 6.5 กรณีศึกษา: เมื่ออคติทำให้การประเมินจิตพิสัยคลาดเคลื่อน

### วิเคราะห์ปัญหาและทางออกเชิงระบบ

กรณีที่ 1 (Halo Effect): ครูให้คะแนน “ความรับผิดชอบ” ของนักเรียนที่เรียนเก่งสูงกว่าความจริง — แก้ไขด้วยรูปรีด/BARS ที่อิงพฤติกรรมเฉพาะ และประเมินที่ละคุณลักษณะแยกจากผลการเรียน

กรณีที่ 2 (Central Tendency): ครูให้ “ระดับ 3” กับทุกคนเพื่อเลี่ยงปัญหา — แก้ไขด้วยมาตรเลขคู่หรือ BARS เพื่อบังคับให้ระบุทิศทางคุณภาพ

บทเรียน: เครื่องมือที่มีเกณฑ์ชัดเจน + การเก็บข้อมูลหลายแหล่ง (Triangulation) + จรรยาบรรณของผู้ประเมิน คือสามเสาหลักของความเที่ยงธรรม

ที่มา: ผู้เขียน เรียบเรียงจาก ชวาล แพรัตกุล (2518) และ Nitko & Brookhart (2019)

### กรณีศึกษา (สาขาสังคมศึกษา/พุทธศาสตร์): การประเมินคุณธรรมจริยธรรมตามหลักอิทธิบาท 4

ชาญณรงค์ ไชยดิษฐ์ และคณะ (2567) พัฒนารูปแบบการจัดการเรียนรู้เพื่อพัฒนาคุณธรรมและจริยธรรมตามหลักอิทธิบาท 4 (ฉันทะ วิริยะ จิตตะ วิมังสา) ในกลุ่มสาระสังคมศึกษาของโรงเรียนพระปริยัติธรรมแห่งหนึ่ง โดยเก็บข้อมูลจากนักเรียน 123 คน ด้วยแบบสอบถามมาตรฐานค่า แบบวัดความพึงพอใจ และการสัมภาษณ์ผู้ให้ข้อมูลสำคัญ แล้วนำผลมาปรับกิจกรรมการเรียนรู้ให้เน้นการฝึกปฏิบัติจริง งานนี้ชี้ว่าการวัดด้านจิตพิสัยในบริบทพุทธศาสตร์ทำได้จริง เมื่อครุญนิยมหลักธรรมให้เป็นพฤติกรรมบ่งชี้ที่สังเกตได้ และใช้เครื่องมือหลายชนิดร่วมกัน (แบบวัดเจตคติ การสังเกต การสัมภาษณ์) แทนการใช้ความรู้สึกของครูเพียงอย่างเดียว

**คำถามชวนคิด:** หากต้องวัด “วิริยะ” ของผู้เรียนในรายวิชาของท่าน จงนิยามเชิงพฤติกรรม 2 ข้อ พร้อมระบุเครื่องมือและช่วงเวลาเก็บข้อมูลที่เหมาะสม

## 6.6 กิจกรรมการเรียนรู้การสอนและใบงาน

---

ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน (6.4) และกิจกรรม Value Clarification ให้ผู้เรียนเลือก ค่านิยมที่ยึดมั่น 3 ประการ

ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ • อภิปรายกลุ่มจำแนกระดับจิตพิสัยจากสถานการณ์ • เวิร์กช็อปวิเคราะห์ความตรงของข้อความในแบบวัดเจตคติ • ฝึกเขียนแบบบันทึกพฤติกรรมที่ปราศจากอคติ

### ใบงานที่ 6.1 สร้างแบบวัดเจตคติแบบลิเคิร์ท (รายกลุ่ม) — CLO3

เลือกเจตคติที่ต้องการวัด 1 เรื่อง (เช่น เจตคติต่อวิชาชีพครู) ให้กลุ่มเขียนนิยามเชิงปฏิบัติการ แล้วสร้าง ข้อความอย่างน้อย 10 ข้อ (บวก 5 / ลบ 5) กำหนดระดับการตอบ การให้คะแนน (รวมการกลับทางข้อความ ลบ) และเกณฑ์การแปลผล พร้อมส่งให้เพื่อนกลุ่มอื่นวิพากษ์ความตรง

### ใบงานที่ 6.2 ออกแบบแบบสังเกตคุณลักษณะอันพึงประสงค์ (รายกลุ่ม) — CLO3

เลือกคุณลักษณะอันพึงประสงค์ 1 ข้อ (เช่น มีจิตสาธารณะ) เขียนนิยามเชิงปฏิบัติการ กำหนด พฤติกรรมบ่งชี้ที่สังเกตได้ 4–6 รายการ และออกแบบมาตราประมาณค่าหรือแบบตรวจสอบรายการพร้อมเกณฑ์ การตัดสิน

ขั้นหลังเรียน: ทำแบบทดสอบหลังเรียน (6.8) แบบฝึกหัดท้ายบท (6.9) และเขียนสะท้อนคิดด้าน จรรยาบรรณการประเมิน

**การสะท้อนคิดและการให้ข้อมูลย้อนกลับ:** ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผล ใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน–หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 6.7 สรุปบทเรียน

พฤติกรรมจิตพิสัยเป็นคุณลักษณะภายในที่เป็นนามธรรม (ความรู้สึก เจตคติ ค่านิยม คุณธรรม) วัด ทางตรงไม่ได้ ต้องอนุมานจากพฤติกรรมที่แสดงออก เครือข่ายและคณะจำแนกเป็น 5 ระดับจากการรับรู้ถึงการ สร้างลักษณะนิสัย โดยมีเป้าหมายสูงสุดคือการหล่อหลอมค่านิยมจนเป็นบุคลิกภาพถาวร จิตพิสัยมีความสำคัญ ต่อ OBE และการพัฒนาคุณลักษณะอันพึงประสงค์ 8 ประการตามหลักสูตรแกนกลางฯ เครื่องมือวัดมีหลาย

ชนิด (การสังเกต แบบสอบถาม/มาตราประมาณค่า การสัมภาษณ์ การประเมินตนเอง/โดยเพื่อน) โดยมาตรลิเคิร์ตเป็นที่ยอมรับที่สุด การให้คะแนนต้องกลับทางข้อความลบ และตรวจสอบคุณภาพด้วยความตรง (IOC) และความเที่ยง (แอลฟา) การวัดจิตพิสัยมีความคลาดเคลื่อนสูงจากพฤติกรรมประดิษฐ์และอคติของผู้ประเมิน (Halo Effect, Central Tendency) จึงต้องวัดต่อเนื่อง ใช้หลายเครื่องมือ ตรวจสอบสามเ้า และยึดมั่นในจรรยาบรรณและความเที่ยงธรรม เพื่อให้การประเมินนำไปสู่การพัฒนาผู้เรียนอย่างแท้จริง

## 6.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 6.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

1. ปลายภาคเรียน ครูต้องรายงานคุณลักษณะ "ความมีวินัย" ของนักเรียนเป็นรายบุคคล จงออกแบบการเก็บหลักฐานเชิงจิตพิสัยที่น่าเชื่อถือกว่าการประมาณจากความรู้สึกครั้งเดียว โดยระบุเครื่องมืออย่างน้อย 2 ชนิด ช่วงเวลาเก็บข้อมูล และวิธีสรุปผล

2. จำแนกระดับพฤติกรรมจิตพิสัย 5 ชั้นของแครธวอล พร้อมยกตัวอย่างพฤติกรรมที่สังเกตได้ในชั้นเรียนแต่ละชั้น

3. เปรียบเทียบเครื่องมือวัดจิตพิสัยอย่างน้อย 3 ชนิด ว่าเหมาะกับการวัดสิ่งใด พร้อมจุดเด่นและข้อจำกัด

4. อธิบายหลักการสร้างและการให้คะแนนแบบวัดเจตคติแบบลิเคิร์ต รวมถึงการกลับทางข้อความลบ และการแปลผลค่าเฉลี่ย

5. ออกแบบแบบวัดเจตคติแบบลิเคิร์ต 5 ข้อ (บวก 3 / ลบ 2) ในเรื่องที่สนใจ พร้อมเกณฑ์การให้คะแนนและการแปลผล

## 6. วิเคราะห์อคติแบบ Halo Effect และ Central Tendency พร้อมเสนอแนวทางลดอคติด้วย

รูปรีค/BARS และการตรวจสอบสามเส้า

## 7. อภิปรายเหตุผลที่จรรยาบรรณและความเที่ยงธรรมจึงสำคัญเป็นพิเศษในการวัดด้านจิตพิสัย

### 6.10 เครื่องมือวัดและประเมินผลประจำบท

#### ก) รูปรีคประเมินแบบวัดเจตคติ (ใบงาน 6.1) แบบแยกองค์ประกอบ (Analytic) — เชื่อมโยง CLO3

| ประเด็นประเมิน                | ดีมาก (4)                  | ดี (3)                  | พอใช้ (2)          | ปรับปรุง (1) |
|-------------------------------|----------------------------|-------------------------|--------------------|--------------|
| ความตรงของข้อความ<br>กับนิยาม | ทุกข้อตรง<br>ประเด็น       | ตรงส่วนใหญ่             | ตรงบางส่วน         | ไม่ตรง       |
| สมดุลข้อความบวก-ลบ            | สมดุลและเหมาะสม            | ค่อนข้างสมดุล           | ไม่ค่อยสมดุล       | ทางเดียวล้วน |
| เทคนิคการเขียนข้อความ         | ชัดเจน<br>ความคิดเห็นเดียว | ชัดเจนเป็นส่วน<br>ใหญ่  | กำกวมบางข้อ        | กำกวมมาก     |
| การให้คะแนนและการแปล<br>ผล    | ถูกต้องครบถ้วน             | ถูกต้องเป็นส่วน<br>ใหญ่ | ผิดพลาดบาง<br>ส่วน | ไม่ถูกต้อง   |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 12$  จาก 16 (ร้อยละ 70 ขึ้นไป)

#### ข) แบบสังเกตพฤติกรรมจรรยาบรรณและความเที่ยงธรรมในการประเมิน — เชื่อมโยง CLO5

| พฤติกรรมบ่งชี้                                  | ปฏิบัติสม่ำเสมอ<br>(3) | บางครั้ง<br>(2) | ควรปรับปรุง<br>(1) |
|-------------------------------------------------|------------------------|-----------------|--------------------|
| ประเมินด้วยความเป็นธรรม ไม่ลำเอียง              |                        |                 |                    |
| ตระหนักว่าการวัดจิตใจต้องระมัดระวัง<br>ไม่ตีตรา |                        |                 |                    |
| รักษาความลับของข้อมูลผู้เรียน                   |                        |                 |                    |
| ใช้เกณฑ์โปร่งใสและแจ้งผู้เรียนล่วงหน้า          |                        |                 |                    |

เกณฑ์: ดี = 10–12, ปานกลาง = 7–9, ควรปรับปรุง = ต่ำกว่า 7 (คะแนนเต็ม 12)

#### สัดส่วนคะแนนประจำบท (รวม 100%)

- ใบงาน 6.1 สร้างแบบวัดเจตคติแบบลิเคิร์ท (CLO3) — 35%

- ใบงาน 6.2 ออกแบบแบบสังเกตคุณลักษณะอันพึงประสงค์ (CLO3) — 25%
- รายงานวิเคราะห์กรณีศึกษาอคติและแนวทางแก้ไข (CLO5/CLO3) — 20%
- แบบสังเกตพฤติกรรมจรรยาบรรณและการมีส่วนร่วม (CLO5) — 20%

## 6.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

กานต์พิชชา ไตรพัทธ์, และวไลพร เมฆไตรรัตน์. (2568).

ผลการจัดการเรียนรู้แบบร่วมมือด้วยเทคนิคที่จิตที่ร่วมกับสื่อมัลติมีเดียที่มีต่อผลสัมฤทธิ์ทางการเรียนและเจตคติต่อการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6.

วารสารการบริหารการศึกษาและนวัตกรรมการศึกษา, 5(3), 95–106.

ชวาล แพร์ตกุล. (2518). เทคนิคการวัดผล (พิมพ์ครั้งที่ 6). วัฒนาพานิช.

ชาญณรงค์ ไชยดิษฐ์, วิทยา ทองดี และสมควร นามสีฐาน. (2567).

รูปแบบการจัดการเรียนรู้เพื่อพัฒนาคุณธรรมและจริยธรรมตามหลักอิทธิบาท 4

ของนักเรียนชั้นมัธยมศึกษาตอนต้น โรงเรียนเจ้าคุณปริยัติมุนีวิทยา อำเภอบางบาล จังหวัดนครราชสีมา.

วารสาร มจร อุบลปริทรรศน์, 9(2), 1683–1696.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จรูญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้.

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ถ้วน สายยศ, และอังคณา สายยศ. (2543). เทคนิคการวัดผลการเรียนรู้. สุวีริยาสาส์น.

สมนึก ภัททิยธนี. (2558). การวัดผลการศึกษา (พิมพ์ครั้งที่ 10). ประสานการพิมพ์.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2553).

แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.

ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

สำเริง บุญเรืองรัตน์, และคณะ. (2534). การวัดและประเมินผลการศึกษา.

โครงการสารานุกรมศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

สุทธิพงศ์ บุญผดุง. (2568). เอกสารประกอบการสอนรายวิชาการวัดและประเมินผลการเรียนรู้.

คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

อนุวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่.  
สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

อุไร จักร์ตรีมงคล. (2559). การพัฒนาแบบวัดความสามารถทางวิชาชีพครู. วารสารวิจัยและพัฒนา  
มหาวิทยาลัยราชภัฏสวนสุนันทา, 8(1), 1–14.

## ภาษาอังกฤษ

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives, Handbook II: Affective domain. David McKay.

Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of response options, reliability, validity, and potential bias in the use of the Likert scale education and social science research: A literature review. International Journal of Educational Methodology, 8(4), 625–637.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 5–55.

Nitko, A. J., & Brookhart, S. M. (2019). Educational assessment of students (8th ed.). Pearson.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. University of Illinois Press.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), Attitude organization and change: An analysis of consistency among attitude components (pp. 1–14). Yale University Press.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47(2), 149–155.

## บทที่ 7

### ความตรงของเครื่องมือวัดและประเมินผล

#### *Validity of Measurement and Evaluation Instruments*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สัปดาห์ที่ 7 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

บทนี้ว่าด้วยการตรวจสอบ “ความตรง” (Validity) ซึ่งเป็นคุณภาพที่สำคัญที่สุดของเครื่องมือวัดและประเมินผล เริ่มจากความหมายและความสำคัญของความตรง แนวคิดความตรงในมุมมองสมัยใหม่ที่ถือเป็นความคิดรวบยอดเชิงเดี่ยว (Unitary Concept) เกี่ยวกับความเหมาะสมของการตีความและการใช้ผลคะแนนประเภทของความตรง 3 ประการ (เชิงเนื้อหา เชิงโครงสร้าง และเชิงเกณฑ์สัมพันธ์) การประกันความตรงเชิงเนื้อหาด้วยตารางวิเคราะห์ข้อสอบ (Test Blueprint) และการตรวจสอบด้วยดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ (IOC) ตลอดจนการวิเคราะห์ แปลผล และใช้ข้อมูลเพื่อปรับปรุงคุณภาพเครื่องมือเพื่อให้นักศึกษาวิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์ เครื่องมือวัด และการแปลผลได้ และออกแบบ/ตรวจสอบคุณภาพเครื่องมือได้อย่างมีหลักวิชา (เชื่อมโยง CLO2 และ CLO3)

#### แผนบริหารการสอนประจำบทที่ 7

##### หัวข้อสาระจัดการเรียนการสอน

- ความหมาย ธรรมชาติ และความสำคัญของความตรง (Validity) ในฐานะคุณภาพหลักของเครื่องมือวัด
- แนวคิดความตรงในการวัดผลสมัยใหม่: ความคิดรวบยอดเชิงเดี่ยวและหลักฐานสนับสนุนการตีความคะแนน
- ประเภทของความตรง: เชิงเนื้อหา เชิงโครงสร้าง และเชิงเกณฑ์สัมพันธ์
- ความตรงเชิงเนื้อหาและการใช้ตารางวิเคราะห์ข้อสอบ (Test Blueprint)
- ความตรงเชิงโครงสร้าง ตัวแปรแฝง และเทคนิคการตรวจสอบ (Known Group/สหสัมพันธ์)
- ความตรงเชิงเกณฑ์สัมพันธ์: แบบตามสภาพและแบบพยากรณ์
- การตรวจสอบความตรงเชิงเนื้อหาด้วยดัชนีความสอดคล้อง (IOC): สูตร เกณฑ์ และการคำนวณ
- การวิเคราะห์ แปลผล และการใช้ข้อมูลเพื่อปรับปรุงคุณภาพเครื่องมือวัด

## ความคิดรวบยอด

การวัดทางการศึกษาเป็นการวัดทางอ้อมที่มีความคลาดเคลื่อนแฝงอยู่เสมอ “ความตรง” (Validity) จึงเป็นคุณภาพอันดับแรกและสำคัญที่สุดที่ทำหน้าที่เชื่อมโยงระหว่างสิ่งที่ผู้สอนตั้งใจจะวัดกับผลที่วัดได้จริง ความตรงในมุมมองสมัยใหม่มิใช่คุณสมบัติของตัวข้อสอบโดยตรง แต่เป็นความเหมาะสมของการตีความและการใช้ผลคะแนนตามวัตถุประสงค์ ครูจึงต้องตรวจสอบความตรงอย่างเป็นระบบ ทั้งเชิงเนื้อหาด้วยตารางวิเคราะห์ข้อสอบและค่า IOC เชิงโครงสร้างด้วยการนิยามตัวแปรแฝง และเชิงเกณฑ์สัมพันธ์ด้วยการหาความสัมพันธ์กับเกณฑ์ภายนอก เพื่อประกันความยุติธรรมและความน่าเชื่อถือของการตัดสินผลการเรียนรู้

## จุดประสงค์การเรียนรู้

1. อธิบายความหมาย ความสำคัญ และแนวคิดสมัยใหม่ของความตรงในฐานะคุณภาพหลักของเครื่องมือวัดได้ (CLO2)
2. วิเคราะห์และจำแนกความตรงทั้ง 3 ประเภทพร้อมยกตัวอย่างและวิธีตรวจสอบที่เหมาะสมได้ (CLO2)
3. วิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์ เครื่องมือวัด และการแปลผลผ่านกรณีศึกษาได้ (CLO2)
4. ออกแบบและใช้ตารางวิเคราะห์ข้อสอบ (Test Blueprint) เพื่อประกันความตรงเชิงเนื้อหาได้ (CLO3)
5. คำนวณค่าดัชนีความสอดคล้อง (IOC) จากผลการประเมินของผู้เชี่ยวชาญได้อย่างถูกต้อง (CLO3)
6. ประเมินคุณภาพข้อสอบจากค่า IOC และเสนอแนวทางปรับปรุงเพื่อเพิ่มความตรงของเครื่องมือได้ (CLO3)

## เนื้อหาวิชา

ดำเนินการสอนตามเอกสารประกอบการสอนบทที่ 7 หัวข้อ 7.5 เนื้อหาวิชา (7.5.1–7.5.10) ซึ่งครอบคลุมความหมายและความสำคัญของความตรง แนวคิดสมัยใหม่ ประเภทของความตรง ความตรงเชิงเนื้อหากับ Test Blueprint ความตรงเชิงโครงสร้าง ความตรงเชิงเกณฑ์สัมพันธ์ การตรวจสอบด้วย IOC การวิเคราะห์และแปลผล ปัญหาและข้อควรระวัง และกรณีศึกษา

## กิจกรรมการเรียนการสอน

### สื่อการเรียนการสอน

- เอกสารประกอบการสอนบทที่ 7 และสไลด์นำเสนอ
- ตัวอย่างจริง: ตารางวิเคราะห์ข้อสอบ (Test Blueprint) แบบฟอร์มประเมิน IOC และชุดข้อสอบสำหรับวิเคราะห์
- ใบงาน 7.1–7.3 พร้อมรูปรีดประเมินชิ้นงาน และเครื่องคิดเลข/สเปรดชีตสำหรับคำนวณ IOC กรณีศึกษาข้อสอบที่ขาดความตรง และตำราในรายการอ้างอิงประจำบท

## การวัดและประเมินผล

- แบบทดสอบก่อน-หลังเรียน วัดความเข้าใจโน้ตศน์ความตรงและการแปลผล IOC (CLO2)
- ใบงานวิเคราะห์/จำแนกประเภทความตรงจากกรณีศึกษา ประเมินด้วยรูบรีคจำแนก (CLO2) — เกณฑ์ผ่านร้อยละ 70
- ชิ้นงานใบงาน 7.2 (Test Blueprint) และใบงาน 7.3 (คำนวณ IOC + แนวทางปรับปรุง) ประเมินด้วยรูบรีค (CLO3)
- การมีส่วนร่วมในเวิร์กช็อปและการอภิปราย (CLO2)

### 7.1 ความคิดรวบยอด

การวัดทางการศึกษาเป็นการวัด “ทางอ้อม” (Indirect Measurement) เพราะคุณลักษณะที่ต้องการวัด เช่น ความรู้ ความสามารถ หรือเจตคติ เป็นสิ่งที่เป็นนามธรรม ไม่สามารถวัดได้โดยตรงเหมือนส่วนสูงหรือน้ำหนัก ผลการวัดจึงมีความคลาดเคลื่อน (Error) แฝงอยู่เสมอ ภารกิจของผู้สอนคือทำให้ความคลาดเคลื่อนนี้เหลือน้อยที่สุด “ความตรง” (Validity) คือคุณภาพที่ทำหน้าที่เชื่อมโยงระหว่างสิ่งที่ผู้สอนตั้งใจจะวัดกับผลที่วัดได้จริง และถือเป็นคุณสมบัติที่สำคัญที่สุดของเครื่องมือวัด ทุกชนิด เพราะหากเครื่องมือไม่สามารถวัดได้ตรงตามวัตถุประสงค์ แม้จะให้ผลคงที่เพียงใด คะแนนนั้นก็ไม่มีคามหมายในการสะท้อนสมรรถภาพที่แท้จริงของผู้เรียน (สุทธิพงศ์ บุญผดุง, 2568; พิชรินทร์ ชมภูวิเศษ, 2559)

ในมุมมองสมัยใหม่ ความตรงมิได้เป็นเพียง “คุณสมบัติ” ของตัวข้อสอบที่สร้างเสร็จแล้ว แต่เป็น “ความคิดรวบยอดเชิงเดียว” (Unitary Concept) ที่อาศัยหลักฐานหลายแหล่งมาสนับสนุนความเหมาะสมของการตีความและการใช้ผลคะแนนตามวัตถุประสงค์ (Messick, 1989; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) การตรวจสอบความตรงจึงเป็นกระบวนการสะสมหลักฐาน ทั้งเชิงเนื้อหา เชิงโครงสร้าง และเชิงเกณฑ์สัมพันธ์ เพื่อยืนยันว่าการสรุปผลการเรียนรู้ของผู้เรียนเป็นไปอย่างถูกต้องและยุติธรรม อันเป็นหัวใจของการประกันคุณภาพการวัดและประเมินผลตามแนวคิดการจัดการศึกษาอิงผลลัพธ์ (OBE)

### 7.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 7 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                | ระดับพฤติกรรม       | เชื่อมโยง CLO | วัดด้วย              |
|------------------------------------------------------------|---------------------|---------------|----------------------|
| 1) อธิบายความหมาย ความสำคัญ และแนวคิดสมัยใหม่ของความตรงได้ | เข้าใจ (พุทธิพิสัย) | CLO2          | แบบทดสอบ ก่อน-หลัง   |
| 2) วิเคราะห์และจำแนกความตรง 3 ประเภทพร้อมวิธีตรวจสอบได้    | วิเคราะห์           | CLO2          | แบบทดสอบ / ใบงาน 7.1 |

|                                                                     |                                  |      |                    |
|---------------------------------------------------------------------|----------------------------------|------|--------------------|
| 3) วิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์ เครื่องมือ และการแปลผลได้ | วิเคราะห์                        | CLO2 | ใบงาน 7.1 + RUBRIC |
| 4) ออกแบบและใช้ตารางวิเคราะห์ข้อสอบ (Test Blueprint) ได้            | ประยุกต์-สร้างสรรค์ (ทักษะพิสัย) | CLO3 | ใบงาน 7.2 + RUBRIC |
| 5) คำนวณค่าดัชนีความสอดคล้อง (IOC) ได้อย่างถูกต้อง                  | ประยุกต์ (ทักษะพิสัย)            | CLO3 | ใบงาน 7.3 + เฉลย   |
| 6) ประเมินคุณภาพข้อสอบจากค่า IOC และเสนอแนวทางปรับปรุงได้           | ประเมินค่า-สร้างสรรค์            | CLO3 | ใบงาน 7.3 + RUBRIC |

หมายเหตุ: ทุกจุดประสงค์มีทั้งกิจกรรมการเรียนรู้ (7.6) วิชัวิต และเครื่องมือ (7.4, 7.8, 7.10) รองรับครบถ้วน โดย CLO2 (วิเคราะห์ความสัมพันธ์ระหว่างจุดประสงค์ เครื่องมือวัด และการแปลผล) และ CLO3 (ออกแบบและสร้างเครื่องมือวัด/แบบประเมินที่เหมาะสมกับลักษณะพฤติกรรม) เป็นผลลัพธ์หลักของบทนี้

## 7.3 คำสำคัญ

ความตรง (Validity) • การวัดทางอ้อม (Indirect Measurement) • ความคลาดเคลื่อน (Error) • ศูนย์สัมพัทธ์ (Relative Zero) • ความคิดรวบยอดเชิงเดี่ยว (Unitary Concept) • หลักฐานความตรง (Validity Evidence) • ความตรงเชิงเนื้อหา (Content Validity) • ความตรงเชิงโครงสร้าง (Construct Validity) • ความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-related Validity) • ความตรงตามสภาพ (Concurrent Validity) • ความตรงเชิงพยากรณ์ (Predictive Validity) • ตารางวิเคราะห์ข้อสอบ (Test Blueprint) • ดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ (IOC) • ตัวแปรแฝง (Latent Variable) • เทคนิคกลุ่มรู้จัก (Known Group) • นิยามเชิงปฏิบัติการ (Operational Definition) • ดุลยพินิจผู้เชี่ยวชาญ (Expert Judgment)

## 7.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอน ประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

## 7.5 เนื้อหาวิชา

### 7.5.1 ความหมายและความสำคัญของความตรง

ความตรง (Validity) หมายถึง ระดับที่หลักฐานเชิงประจักษ์และทฤษฎีสนับสนุนการตีความคะแนนจากเครื่องมือวัดได้อย่างเหมาะสมตามวัตถุประสงค์ของการใช้ หรือกล่าวอย่างง่ายว่า เครื่องมือนั้น “วัดได้ตรงในสิ่งที่ต้องการจะวัด” (American Educational Research Association [AERA], American Psychological

Association [APA], & National Council on Measurement in Education [NCME], 2014) นักวัดผลของไทยนิยมใช้คำว่า “ความเที่ยงตรง” โดย ล้วน สายยศ และอังคณา สายยศ (2543, น. 246) อธิบายว่าเป็นคุณสมบัติของแบบทดสอบที่วัดได้ตรงตามลักษณะหรือจุดประสงค์ที่ต้องการวัด ส่วน พชรินทร์ ชมภูวิเศษ (2559, น. 185) และ สุทธิพงศ์ บุญผดุง (2568) ยืนยันตรงกันว่าความตรงเป็น “คุณสมบัติที่สำคัญที่สุด” ในบรรดาคุณภาพของเครื่องมือวัดทุกชนิด

เหตุที่ความตรงถูกจัดเป็นคุณภาพอันดับแรก เพราะการวัดทางการศึกษาเป็นการวัดทางอ้อม (Indirect Measurement) กล่าวคือ เราไม่สามารถวัดความรู้ เจตคติ หรือทักษะการคิดได้โดยตรงเหมือนการวัดส่วนสูงหรือน้ำหนัก แต่ต้องอนุมาน (Inference) จากพฤติกรรมที่ผู้เรียนแสดงออกผ่านการตอบข้อสอบ การวัดทางอ้อมจึงมีความคลาดเคลื่อน (Error) ผ่างอยู่เสมอ หากเครื่องมือวัด “ผิดพลาด” ตั้งแต่ต้น คะแนนที่ได้ย่อมไร้ความหมายและที่สำคัญคือ แม้เครื่องมือจะมีความเที่ยง (Reliability) สูงเพียงใด ก็ให้คะแนนคงที่ทุกครั้งที่วัด ก็เป็นเพียง “การวัดที่แม่นยำในสิ่งที่ผิด” ความตรงจึงเป็นเงื่อนไขจำเป็น (Necessary Condition) ที่ต้องมาก่อนคุณภาพด้านอื่น (Messick, 1989)

ตัวอย่างในชั้นเรียนไทย หากครูคณิตศาสตร์ต้องการวัด “ทักษะการคำนวณ” แต่ออกโจทย์ปัญหาด้วยภาษาที่ซับซ้อนเกินระดับชั้น เด็กที่คำนวณเก่งแต่ยังอ่านจับใจความไม่คล่องจะทำได้ ข้อสอบนั้นกำลังวัด “ทักษะทางภาษา” แทน “ทักษะการคำนวณ” จึงขาดความตรง คะแนนศูนย์ของเด็กคนนี้เป็นเพียง “ศูนย์สัมพัทธ์” (Relative Zero) มิได้หมายความว่าเขาไม่มีความรู้คณิตศาสตร์เลย แต่เกิดจากเครื่องมือที่บกพร่องส่งผลให้การตัดสินผลการเรียนคลาดเคลื่อนไปจากความสามารถจริง

**การประยุกต์ใช้และข้อควรระวัง:** ก่อนนำเครื่องมือไปใช้จริง ครูควรตรวจสอบความตรงเสมอ เพื่อประกันความยุติธรรมในการตัดสินผลการเรียน โดยตรวจว่าข้อสอบครอบคลุมจุดประสงค์และเนื้อหาที่สอนจริงหรือไม่ ใช้ภาษาที่เหมาะสมกับผู้เรียนหรือไม่ และเกณฑ์การให้คะแนนสอดคล้องกับสิ่งที่มุ่งวัดหรือไม่ ข้อพึงระวังที่พบบ่อยคือการเขียนจุดประสงค์ไว้ระดับหนึ่งแต่ออกข้อสอบอีกระดับหนึ่ง เช่น ตั้งจุดประสงค์ให้ “วิเคราะห์” แต่ข้อสอบวัดเพียง “ความจำ” ทำให้ผลการวัดไม่สะท้อนความสามารถที่มุ่งหมาย (เชื้อมโยง CLO2)

#### ตารางที่ 7.1 นิยาม “ความตรง/ความเที่ยงตรง” ตามทัศนะของนักวิชาการ

| นักวิชาการ (ปี)                           | สาระสำคัญของนิยาม                                                                 |
|-------------------------------------------|-----------------------------------------------------------------------------------|
| ไอเซนค์ (Eysenck, 1972, น. 97)            | แบบทดสอบที่สามารถวัดได้ตรงตามสิ่งที่ต้องการวัด                                    |
| ล้วน สายยศ และอังคณา สายยศ (2543, น. 246) | แบบทดสอบที่วัดได้ตรงตามลักษณะหรือจุดประสงค์ที่ต้องการวัด                          |
| ไพศาล หวังพานิช (2526)                    | ความถูกต้องแม่นยำที่เครื่องมือวัดได้ตามจุดประสงค์ที่วางไว้                        |
| AERA, APA และ NCME (2014)                 | ระดับที่หลักฐานและทฤษฎีสนับสนุนการตีความคะแนนตามการใช้ที่มุ่งหมาย                 |
| สุทธิพงศ์ บุญผดุง (2568)                  | คุณภาพของเครื่องมือที่วัดได้สอดคล้องกับวัตถุประสงค์และเป็นคุณสมบัติที่สำคัญที่สุด |

ที่มา: ประมวลจากนิยามที่รวบรวมไว้ใน พächรินทร์ ชมภูวิเศษ (2559, น. 187–188); ล้วน สายยศ และอังคณา สายยศ (2543); AERA, APA และ NCME (2014); และ สุทธิพงศ์ บุญผดุง (2568)

## 7.5.2 แนวคิดเกี่ยวกับความตรงในการวัดผลสมัยใหม่

เดิมนักวัดผลมองความตรงเป็น “ประเภท” ที่แยกขาดจากกัน (Trinitarian View) คือ ความตรงเชิงเนื้อหา เชิงโครงสร้าง และเชิงเกณฑ์สัมพันธ์ เป็นคนละเรื่องกัน แต่ปัจจุบันถือว่าความตรงเป็น “ความคิดรวบยอดเชิงเดี่ยว” (Unitary Concept) ที่อาศัยหลักฐานหลายแหล่งมาสนับสนุนความเหมาะสมของ “การตีความผลคะแนน” (Inference from Scores) ว่าคะแนนนั้นสะท้อนความสามารถที่มุ่งวัดจริงหรือไม่ (Messick, 1989) จุดเปลี่ยนสำคัญของมุมมองนี้คือ ความตรงไม่ใช่คุณสมบัติของ “ตัวข้อสอบ” แต่เป็นคุณสมบัติของ “การตีความและการใช้คะแนน” ในบริบทหนึ่ง ๆ ข้อสอบชุดเดียวกันจึงอาจตรงสำหรับการใช้หนึ่งแต่ไม่ตรงสำหรับอีกการใช้หนึ่ง

มาตรฐานการทดสอบทางการศึกษาและจิตวิทยาของ AERA, APA และ NCME (2014) ได้เสนอกรอบ “แหล่งหลักฐานความตรง” ไว้ 5 แหล่ง ซึ่งช่วยให้ครูสะสมหลักฐานได้อย่างเป็นระบบ แทนการมองความตรงเป็นเพียงป้ายชื่อประเภท ดังภาพที่ 7.1 และตารางที่ 7.2 ที่นำเสนอความสัมพันธ์ของแนวคิดนี้

ภาพที่ 7.1 มโนทัศน์ความตรงเชิงเดี่ยว: หลักฐาน 3 เสาค้ำยัน “ความถูกต้องของการตีความคะแนน”

| เสาหลักเชิงเนื้อหา<br>(Content)                                    | เสาหลักเชิงโครงสร้าง<br>(Construct) | เสาหลักเชิงเกณฑ์สัมพันธ์<br>(Criterion) |
|--------------------------------------------------------------------|-------------------------------------|-----------------------------------------|
| ข้อคำถามครอบคลุมและเป็นตัวแทนของสิ่งที่สอน                         | วัดตัวแปรแฝงได้ตรงตามทฤษฎี          | ผลคะแนนสัมพันธ์กับเกณฑ์ในโลกจริง        |
| ความถูกต้องของการตีความและการใช้คะแนน (Valid Interpretation & Use) |                                     |                                         |

ที่มา: ดัดแปลงจาก Messick (1989) และ AERA, APA และ NCME (2014)

หากเสาต้นใดบกพร่อง ความตรงของการสรุปผลการเรียนรู้ย่อมสั่นคลอน สำหรับครูผู้สอน หลักฐานเชิงเนื้อหา ซึ่งตรวจสอบผ่านตารางวิเคราะห์ข้อสอบ (Test Blueprint) และดัชนีความสอดคล้อง (IOC) เป็นแหล่งที่ตรวจสอบได้ง่ายและจำเป็นที่สุดในการสร้างข้อสอบประจำวิชา

ตารางที่ 7.2 แหล่งหลักฐานความตรง 5 ประการ ตามมาตรฐาน AERA, APA และ NCME (2014)

| แหล่งหลักฐาน                                                    | พิจารณาอะไร                                         | ตัวอย่างหลักฐานในชั้นเรียน               |
|-----------------------------------------------------------------|-----------------------------------------------------|------------------------------------------|
| 1) ด้านเนื้อหา (Content)                                        | ข้อสอบเป็นตัวแทนขอบเขตเนื้อหา/พฤติกรรมที่สอน        | Test Blueprint และผล IOC จากผู้เชี่ยวชาญ |
| 2) ด้านกระบวนการตอบสนอง (Response Process)                      | ผู้เรียนใช้กระบวนการคิดตรงกับที่มุ่งวัดหรือไม่      | การสัมภาษณ์ให้เด็กคิดออกเสียงขณะทำข้อสอบ |
| 3) ด้านโครงสร้างภายใน (Internal Structure)                      | ความสัมพันธ์ระหว่างข้อสอบสอดคล้องกับโครงสร้างที่วัด | การวิเคราะห์รายข้อ/วิเคราะห์องค์ประกอบ   |
| 4) ด้านความสัมพันธ์กับตัวแปรอื่น (Relations to Other Variables) | คะแนนสัมพันธ์กับเกณฑ์ภายนอกตามทฤษฎี                 | สหสัมพันธ์กับเกรดเฉลี่ยหรือแบบวัดมาตรฐาน |

|                                     |                                                       |                                               |
|-------------------------------------|-------------------------------------------------------|-----------------------------------------------|
| 5) ด้านผลที่ตามมา<br>(Consequences) | ผลของการใช้คะแนนก่อนผลที่พึง<br>ประสงค์/ไม่พึงประสงค์ | ตรวจสอบว่าการตัดสินผลไม่ลำเอียง<br>ต่อกลุ่มใด |
|-------------------------------------|-------------------------------------------------------|-----------------------------------------------|

ที่มา: ดัดแปลงจาก AERA, APA และ NCME (2014)

### 7.5.3 ประเภทของความตรง

แม้แนวคิดสมัยใหม่จะมองความตรงเป็นความคิดรวบยอดเชิงเดี่ยว แต่เพื่อความเข้าใจที่เป็นระบบ ตำราการวัดผลของไทยและสากลยังคงจำแนกหลักฐานความตรงออกเป็น 3 ประเภทหลักตามวัตถุประสงค์ของการใช้เครื่องมือ ได้แก่ ความตรงเชิงเนื้อหา (Content) ความตรงเชิงโครงสร้าง (Construct) และความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-related) โดยประเภทสุดท้ายยังแบ่งย่อยตามช่วงเวลาของเกณฑ์เป็นแบบตามสภาพและแบบพยากรณ์ (Nitko & Brookhart, 2019; พัทรินทร์ ชมภูวิเศษ, 2559) ทั้งนี้ สุทธิพงศ์ บุญผดุง (2568) จำแนกออกเป็น 4 ประเภทโดยนับความตรงตามสภาพและเชิงพยากรณ์แยกกัน ดังแสดงความสัมพันธ์ในภาพที่ 7.2

ภาพที่ 7.2 แผนภาพชนิดของความตรง (Types of Validity)

| ความตรง (Validity)    |                           |                                                     |
|-----------------------|---------------------------|-----------------------------------------------------|
| เชิงเนื้อหา (Content) | เชิงโครงสร้าง (Construct) | เชิงเกณฑ์สัมพันธ์ (Criterion)                       |
| —                     | —                         | ตามสภาพ (Concurrent)   เชิง<br>พยากรณ์ (Predictive) |

ที่มา: ผู้เขียน ประมวลจาก Nitko & Brookhart (2019) และ สุทธิพงศ์ บุญผดุง (2568)

การจำแนกนี้ช่วยให้ครูเลือกวิธีหาหลักฐานความตรงให้เหมาะกับจุดมุ่งหมายของเครื่องมือ เช่น ข้อสอบวัดผลสัมฤทธิ์ประจำวิชานับความตรงเชิงเนื้อหา แบบวัดเจตคติหรือความถนัดเน้นความตรงเชิงโครงสร้าง ส่วนแบบทดสอบคัดเลือกเข้าศึกษาต่อเน้นความตรงเชิงพยากรณ์ รายละเอียดเปรียบเทียบทั้งสามประเภทแสดงในตารางที่ 7.3

ตารางที่ 7.3 เปรียบเทียบความตรง 3 ประเภทหลัก

| ประเภทความตรง                 | จุดมุ่งหมาย                                             | วิธีตรวจสอบที่นิยม                                               | ตัวอย่างการใช้                        |
|-------------------------------|---------------------------------------------------------|------------------------------------------------------------------|---------------------------------------|
| เชิงเนื้อหา (Content)         | ตรวจสอบความเป็นตัวแทนของเนื้อหาและจุดประสงค์การเรียนรู้ | Test Blueprint และการหาค่า IOC โดยผู้เชี่ยวชาญ                   | ข้อสอบกลางภาค/ปลายภาคประจำวิชา        |
| เชิงโครงสร้าง (Construct)     | ตรวจว่าวัดคุณลักษณะแฝง (เจตคติ เชาวน์ปัญญา) ตรงตามทฤษฎี | เทคนิคกลุ่มรู้ชัด การวิเคราะห์องค์ประกอบ สหสัมพันธ์กับแบบมาตรฐาน | แบบวัดเจตคติต่อวิชาวิทยาศาสตร์        |
| เชิงเกณฑ์สัมพันธ์ (Criterion) | ตรวจสอบความสัมพันธ์ระหว่างคะแนนกับเกณฑ์ภายนอก           | หาค่าสัมประสิทธิ์สหสัมพันธ์กับเกณฑ์                              | ข้อสอบคัดเลือกทำนายผล การเรียนในอนาคต |

ที่มา: ประมวลจาก สุทธิพงศ์ บุญผดุง (2568); พัทรินทร์ ชมภูวิเศษ (2559, น. 189); และ Nitko & Brookhart (2019)

#### 7.5.4 ความตรงเชิงเนื้อหา (Content Validity) และ Test Blueprint

ความตรงเชิงเนื้อหา (Content Validity) หมายถึง การที่ข้อสอบเป็นตัวแทนที่ดีของขอบเขตเนื้อหาและระดับพฤติกรรมที่กำหนดไว้ในหลักสูตร เป็นคุณภาพที่จำเป็นอย่างยิ่งสำหรับข้อสอบวัดผลสัมฤทธิ์ (สุทธิพงศ์ บุญผดุง, 2568; พชรินทร์ ชมภูวิเศษ, 2559) พชรินทร์ ชมภูวิเศษ (2559, น. 189) ระบุวิธีตรวจสอบความตรงเชิงเนื้อหาไว้ 2 วิธี คือ (1) ตรวจสอบสอดคล้องของข้อสอบกับเนื้อหาในตารางวิเคราะห์หลักสูตร รวมทั้งสัดส่วนจำนวนข้อในแต่ละเนื้อหา และ (2) อาศัยดุลยพินิจของผู้เชี่ยวชาญอย่างน้อย 3 ท่าน หากเห็นตรงกันว่าใช้ได้ จึงถือว่ามีความตรงเชิงเนื้อหา

เครื่องมือสำคัญที่ใช้ประกันความตรงเชิงเนื้อหาตั้งแต่ขั้นวางแผนคือ “ตารางวิเคราะห์ข้อสอบ” (Test Blueprint หรือ Table of Specifications) ซึ่งช่วยกระจายจำนวนข้อสอบให้สอดคล้องกับน้ำหนักเนื้อหาและระดับพฤติกรรมที่สอนจริง หากครูใช้เวลาสอนเรื่อง “การวิเคราะห์” มากแต่ข้อสอบกลับมีแต่ระดับ “ความจำ” เครื่องมือนั้นย่อมล้มเหลวด้านความตรงเชิงเนื้อหา ตัวอย่างตารางวิเคราะห์ข้อสอบบทที่ 7 (รวม 20 ข้อ) แสดงดังตารางที่ 7.4 ซึ่งครูสามารถอ่านค่าได้โดยดูการกระจายจำนวนข้อในแต่ละหน่วยและแต่ละระดับพฤติกรรม แล้วตรวจว่าผลรวมตรงตามน้ำหนักที่ตั้งใจหรือไม่

ตารางที่ 7.4 ตัวอย่างตารางวิเคราะห์ข้อสอบ (Test Blueprint) ประจำปีที่ 7

| หน่วยการเรียนรู้                        | จำ-เข้าใจ | ประยุกต์ | วิเคราะห์ขึ้นไป | รวม (ข้อ) | ร้อยละ |
|-----------------------------------------|-----------|----------|-----------------|-----------|--------|
| 1) ความหมายและประเภทของความตรง          | 3         | 1        | 1               | 5         | 25%    |
| 2) ความตรงเชิงเนื้อหาและ Test Blueprint | 2         | 2        | 1               | 5         | 25%    |
| 3) IOC: สูตรและการคำนวณ                 | 2         | 3        | 1               | 6         | 30%    |
| 4) การแปลผลและการปรับปรุงเครื่องมือ     | 1         | 1        | 2               | 4         | 20%    |
| รวม                                     | 8         | 7        | 5               | 20        | 100%   |

ที่มา: ผู้เขียน ปรับใช้แนวทางตารางวิเคราะห์หลักสูตรของ พชรินทร์ ชมภูวิเศษ (2559, น. 189)

นอกจากการหาค่า IOC แล้ว ในงานวิจัยยังมีดัชนีเชิงปริมาณอื่นที่ใช้ยืนยันความตรงเชิงเนื้อหา ซึ่งครูควรรู้จักไว้เพื่อเลือกใช้ให้เหมาะกับจำนวนผู้เชี่ยวชาญและลักษณะข้อมูล ดังกล่องที่ 7.1

### กล่องที่ 7.1 ดัชนีเชิงปริมาณของความตรงเชิงเนื้อหา

- IOC (Index of Item-Objective Congruence): วัดความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ ใช้ผู้เชี่ยวชาญ  $\geq 3$  ท่าน ให้  $+1/0/-1$  เกณฑ์ผ่าน  $\geq 0.50$  (Rovinelli & Hambleton, 1977)
  - CVR (Content Validity Ratio): สัดส่วนผู้เชี่ยวชาญที่เห็นว่าข้อคำถาม “จำเป็น” มีค่าวิกฤตเปลี่ยนตามจำนวนผู้เชี่ยวชาญ (Lawshe, 1975)
  - CVI (Content Validity Index): สัดส่วนผู้เชี่ยวชาญที่ให้คะแนนความสอดคล้องในระดับสูง นิยมใช้ในงานวิจัยทางพยาบาลและสังคมศาสตร์
- การเลือกใช้ขึ้นกับจำนวนผู้ประเมินและประเภทมาตรวัด สุพัฒน์ สุขมลสันต์ (2567) รวบรวมไว้ถึง 15 วิธี ส่วน Almanasreh, Moles และ Chen (2019) ทบทวนจุดเด่น-จุดด้อยของแต่ละวิธีอย่างเป็นระบบ

**ข้อควรระวัง:** ความตรงเชิงเนื้อหาอาศัย “ดุลยพินิจ” ของผู้เชี่ยวชาญเป็นหลัก จึงต้องคัดเลือกผู้เชี่ยวชาญที่มีความรู้ตรงสาขา และเขียนจุดประสงค์ให้ชัดเจนก่อนส่งให้ประเมิน มิฉะนั้นผลที่ได้อาจสะท้อนความเห็นส่วนตัวมากกว่าคุณภาพที่แท้จริงของข้อสอบ

### 7.5.5 ความตรงเชิงโครงสร้าง (Construct Validity)

ความตรงเชิงโครงสร้าง (Construct Validity) มุ่งตรวจสอบว่าเครื่องมือวัด “ตัวแปรแฝง” (Latent Variable) ที่เป็นนามธรรมได้ตรงตามทฤษฎีหรือไม่ เช่น ความคิดสร้างสรรค์ ความใฝ่เรียนรู้ แรงจูงใจใฝ่สัมฤทธิ์ หรือเขาวนปัญญา แนวคิดนี้ได้รับการวางรากฐานโดย Cronbach และ Meehl (1955) ซึ่งเสนอว่าการตรวจสอบความตรงเชิงโครงสร้างเป็นการสะสมหลักฐานเพื่อยืนยันว่าคะแนนสะท้อน “มโนทัศน์” (Construct) ตามทฤษฎีที่อ้างอิงจริง

ขั้นตอนเริ่มจากการนิยามคุณลักษณะให้เป็น “นิยามเชิงปฏิบัติการ” (Operational Definition) คือ แปลงสิ่งนามธรรมให้เป็นพฤติกรรมบ่งชี้ที่สังเกตและวัดได้ แล้วจึงหาหลักฐานยืนยันด้วยวิธีต่าง ๆ (ไพศาล หวังพานิช, 2526; พัทรินทร์ ชมภูวิเศษ, 2559) ตัวอย่างในชั้นเรียนไทย หากครูต้องการวัด “ความรับผิดชอบ” อาจนิยามเป็นพฤติกรรมบ่งชี้ เช่น การส่งงานตรงเวลา การเตรียมอุปกรณ์มาเรียน และการรับผิดชอบเมื่อทำงานกลุ่มไม่เสร็จ จากนั้นใช้เทคนิคกลุ่มรู้ชัด (Known-Group Technique) นำแบบวัดไปใช้กับนักเรียนสองกลุ่มที่ครูทราบชัดเจนว่าต่างกัน หากแบบวัดจำแนกสองกลุ่มได้ชัดเจน แสดงว่ามีความตรงเชิงโครงสร้างสูง วิธีตรวจสอบที่นิยมสรุปไว้ในตารางที่ 7.5

ตารางที่ 7.5 เทคนิคการตรวจสอบความตรงเชิงโครงสร้าง

| เทคนิค                    | หลักการ                                                           | ตัวอย่างการใช้                                              |
|---------------------------|-------------------------------------------------------------------|-------------------------------------------------------------|
| กลุ่มรู้ชัด (Known-Group) | นำไปวัดกลุ่มที่ทราบว่ามีคุณลักษณะต่างกัน แล้วดูว่าจำแนกได้หรือไม่ | แบบวัดความวิตกในการสอบ เทียบเด็กที่ครูทราบว่าเครียดมาก-น้อย |

|                                       |                                                      |                                                   |
|---------------------------------------|------------------------------------------------------|---------------------------------------------------|
| วิเคราะห์องค์ประกอบ (Factor Analysis) | ตรวจสอบว่าข้อคำถามรวมกลุ่มตามองค์ประกอบที่ทฤษฎีกำหนด | แบบวัดเจตคติ 3 ด้านจัดกลุ่มได้ตรง 3 องค์ประกอบ    |
| หลักฐานสอดคล้อง (Convergent)          | สัมพันธ์สูงกับแบบวัดอื่นที่วัดสิ่งเดียวกัน           | แบบวัดเขาวนชุดใหม่สัมพันธ์สูงกับแบบมาตรฐานเดิม    |
| หลักฐานจำแนก (Discriminant)           | สัมพันธ์ต่ำกับแบบวัดที่วัดคนละสิ่ง                   | แบบวัดความถนัดเลขสัมพันธ์ต่ำกับแบบวัดความถนัดภาษา |

ที่มา: ประมวลจาก Cronbach & Meehl (1955); พิชรินทร์ ชมภูวิเศษ (2559, น. 189–190, 194)

**ข้อควรระวัง:** ภัยคุกคามความตรงเชิงโครงสร้างมี 2 ลักษณะ คือ การวัดไม่ครอบคลุมมโนทัศน์ (Construct Underrepresentation) เช่น วัดความคิดสร้างสรรค์เฉพาะด้านภาษาแต่ละเลยด้านภาพ และการปนเปื้อนของตัวแปรแปลกปลอม (Construct-Irrelevant Variance) เช่น ข้อสอบยากด้านภาษาจนไปวัดทักษะการอ่านแทนมโนทัศน์เป้าหมาย (Messick, 1989) 7.5.6 ความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-related Validity)

ความตรงเชิงเกณฑ์สัมพันธ์ (Criterion-related Validity) เป็นการพิสูจน์ความแม่นยำของเครื่องมือโดยเทียบคะแนนกับ “เกณฑ์ภายนอก” (External Criterion) ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งเรียกว่า “สัมประสิทธิ์ความตรง” (Validity Coefficient) ยิ่งสัมพันธ์สูงยิ่งแสดงว่าเครื่องมือตรงตามเกณฑ์ (สุทธิพงศ์ บุญผดุง, 2568; พิชรินทร์ ชมภูวิเศษ, 2559) จำแนกได้ 2 ลักษณะตามช่วงเวลาของเกณฑ์ที่นำมาเทียบดังตารางที่ 7.6

#### ตารางที่ 7.6 เปรียบเทียบความตรงตามสภาพและความตรงเชิงพยากรณ์

| ประเด็น          | ตามสภาพ (Concurrent)                                   | เชิงพยากรณ์ (Predictive)                      |
|------------------|--------------------------------------------------------|-----------------------------------------------|
| ช่วงเวลาของเกณฑ์ | เกณฑ์ที่มีอยู่ในปัจจุบัน                               | เกณฑ์ที่จะเกิดขึ้นในอนาคต                     |
| คำถามหลัก        | วัดสภาพปัจจุบันได้ตรงหรือไม่                           | ทำนายความสำเร็จในอนาคตได้หรือไม่              |
| สถิติที่ใช้      | สหสัมพันธ์กับแบบมาตรฐานเดิม หรือเทียบ 2 กลุ่ม (t-test) | สหสัมพันธ์ระหว่างคะแนนกับผลสำเร็จภายหลัง      |
| ตัวอย่าง         | แบบวัดเขาวนชุดใหม่สัมพันธ์สูงกับแบบมาตรฐานที่ใช้อยู่   | คะแนนสอบคัดเลือก ม.4 ทำนายเกรดเฉลี่ยตอนจบ ม.6 |

ที่มา: ประมวลจาก สุทธิพงศ์ บุญผดุง (2568); พิชรินทร์ ชมภูวิเศษ (2559, น. 194); และ ขวาล แพร่ตกุล (2518)

ความตรงเชิงพยากรณ์มีความสำคัญมากในงานแนะแนวและการคัดเลือกบุคคล เพราะช่วยให้สถานศึกษาตัดสินใจรับผู้เรียนได้อย่างมีหลักฐาน อย่างไรก็ตาม ค่าสัมประสิทธิ์ความตรงขึ้นอยู่กับคุณภาพของ “เกณฑ์” ที่นำมาเทียบโดยตรง หากเกณฑ์เองคลาดเคลื่อน (เช่น เกรดเฉลี่ยที่มาจากการให้คะแนนไม่เป็นมาตรฐาน) ค่าที่ได้ก็เชื่อถือไม่ได้ ปัญหานี้เรียกว่า “ปัญหาของเกณฑ์” (Criterion Problem) ซึ่งครูควรพิจารณาก่อนสรุปผล

#### 7.5.7 การตรวจสอบความตรงเชิงเนื้อหาด้วยค่า IOC

ดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ (Index of Item-Objective Congruence: IOC) เป็นวิธีหาหลักฐานความตรงเชิงเนื้อหาในเชิงปริมาณที่นิยมที่สุดในประเทศไทย พัฒนาขึ้นโดย Rovinelli และ

Hambleton (1977) โดยให้ผู้เชี่ยวชาญ (Expert Judgment) อย่างน้อย 3 ท่าน พิจารณาความสอดคล้องระหว่างข้อคำถามแต่ละข้อกับจุดประสงค์ที่ต้องการวัด แล้วให้คะแนนตามเกณฑ์ +1, 0 และ -1 จากนั้นนำมาคำนวณด้วยสูตรในกล่องที่ 7.2 (พัชรินทร์ ชมภูวิเศษ, 2559; สุทธิพงศ์ บุญผดุง, 2568)

#### กล่องที่ 7.2 สูตรและเกณฑ์การให้คะแนน IOC

สูตร:  $IOC = \Sigma R / N$

เมื่อ  $IOC$  = ดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์

$\Sigma R$  = ผลรวมคะแนนการพิจารณาจากผู้เชี่ยวชาญทุกคน

$N$  = จำนวนผู้เชี่ยวชาญทั้งหมด

เกณฑ์การให้คะแนนของผู้เชี่ยวชาญแต่ละคน:

+1 = แน่ใจว่าข้อคำถามสอดคล้อง (วัดตรง) กับจุดประสงค์

0 = ไม่แน่ใจว่าข้อคำถามสอดคล้องกับจุดประสงค์หรือไม่

-1 = แน่ใจว่าข้อคำถามไม่สอดคล้องกับจุดประสงค์

ค่า IOC มีพิสัยตั้งแต่ -1.00 ถึง +1.00

อย่างไรก็ตาม พิเศษฐ ตัณทวนิช และ พนา จินดาศรี (2561) ชี้ว่า “ความหมายที่แท้จริง” ของ IOC ตามต้นฉบับของ Rovinelli และ Hambleton ออกแบบไว้สำหรับข้อสอบอิงเกณฑ์ที่ข้อหนึ่ง ๆ มุ่งวัดจุดประสงค์เดียว การให้ +1 จึงควรหมายถึง “แน่ใจว่าข้อนี้วัดจุดประสงค์เป้าหมาย และไม่ไปวัดจุดประสงค์อื่น” การใช้ IOC เป็นเพียงค่าเฉลี่ยอย่างหลวม ๆ จึงอาจคลาดเคลื่อนจากเจตนาเดิม ในกรณีข้อสอบที่วัดหลายมิติ Turner และ Carlson (2003) ได้ขยายสูตรให้รองรับได้ บทเรียนสำคัญสำหรับครูคือ ต้องเขียนจุดประสงค์ให้คมชัดและให้ผู้เชี่ยวชาญพิจารณาบนฐานจุดประสงค์นั้นจริง ๆ

**ข้อควรระวัง:** ควรเลือกผู้เชี่ยวชาญที่มีความรู้ทั้งด้านเนื้อหาและด้านการวัดผล อย่างน้อย 3–5 ท่าน และแนบจุดประสงค์เชิงพฤติกรรมพร้อมเนื้อหาที่มุ่งวัดไปกับแบบประเมิน เพื่อให้ผู้เชี่ยวชาญตัดสินได้บนฐานเดียวกัน (สุพัฒน์ สุกมลสันต์, 2567)

#### 7.5.8 การวิเคราะห์และแปลผลค่า IOC

เกณฑ์มาตรฐานในการตัดสินคุณภาพข้อสอบจากค่า IOC คือ ค่าตั้งแต่ 0.50–1.00 ถือว่าข้อสอบสอดคล้องกับจุดประสงค์ (ผ่านเกณฑ์ คัดเลือกไว้ใช้ได้) ส่วนค่าต่ำกว่า 0.50 ถือว่าไม่สอดคล้อง ควรปรับปรุงแก้ไขหรือตัดทิ้ง (พัชรินทร์ ชมภูวิเศษ, 2559; ศิริชัย กาญจนวาสี, 2552; สุทธิพงศ์ บุญผดุง, 2568) ตัวอย่างการ

คำนวณ สมมติข้อสอบ 5 ข้อ ประเมินโดยผู้เชี่ยวชาญ 3 ท่าน (N = 3) ผลการคำนวณและการตัดสินใจแสดงดังตารางที่ 7.7

ตารางที่ 7.7 ตัวอย่างการคำนวณและแปลผลค่า IOC (ผู้เชี่ยวชาญ 3 ท่าน)

| ข้อที่ | ผู้เชี่ยวชาญ<br>1 | ผู้เชี่ยวชาญ<br>2 | ผู้เชี่ยวชาญ<br>3 | $\Sigma R$ | $IOC = \Sigma R/N$ | ผลการ<br>พิจารณา                  |
|--------|-------------------|-------------------|-------------------|------------|--------------------|-----------------------------------|
| 1      | +1                | +1                | +1                | 3          | 1.00               | ผ่าน —<br>คัดเลือกไว้             |
| 2      | +1                | +1                | 0                 | 2          | 0.67               | ผ่าน —<br>คัดเลือกไว้             |
| 3      | +1                | 0                 | 0                 | 1          | 0.33               | ไม่ผ่าน —<br>ปรับปรุง/<br>ตัดทิ้ง |
| 4      | +1                | 0                 | -1                | 0          | 0.00               | ไม่ผ่าน —<br>ปรับปรุง/<br>ตัดทิ้ง |
| 5      | +1                | +1                | -1                | 1          | 0.33               | ไม่ผ่าน —<br>ปรับปรุง/<br>ตัดทิ้ง |

ที่มา: ผู้เขียน ปรับใช้แนวทางการคำนวณของ พัชรินทร์ ชมภูวิเศษ (2559, น. 191, ตัวอย่างที่ 6.2)

จากตารางที่ 7.7 ข้อ 1 และ 2 ผ่านเกณฑ์ ( $IOC \geq 0.50$ ) ส่วนข้อ 3–5 ไม่ผ่าน ครูควรกลับไปทบทวนนิยามจุดประสงค์ ปรับภาษาให้ชัดเจน หรือปรึกษาผู้เชี่ยวชาญเพิ่มเติมก่อนนำไปใช้จริง ทั้งนี้ การตีความค่า IOC ควรพิจารณาควบคู่กับข้อเสนอแนะเชิงคุณภาพของผู้เชี่ยวชาญด้วย เพราะตัวเลขเพียงอย่างเดียวไม่ได้บอกว่าการแก้ไขข้อสอบอย่างไร

### 7.5.9 ปัญหา ข้อควรระวัง และการพัฒนาคุณภาพเครื่องมือ

การตรวจสอบความตรงอาจคลาดเคลื่อนได้หากไม่ระวังปัจจัยหลายประการ ทั้งจากตัวผู้เชี่ยวชาญ จากการเขียนจุดประสงค์ และจากตัวข้อสอบเอง สรุปปัญหาที่พบบ่อยพร้อมแนวทางแก้ไขไว้ในตารางที่ 7.8

ตารางที่ 7.8 ปัญหาที่พบบ่อยในการตรวจสอบความตรงและแนวทางแก้ไข

| ปัญหาที่พบบ่อย                                                      | แนวทางแก้ไข                                                                                          |
|---------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| อคติหรือความเห็นไม่ตรงกันของผู้เชี่ยวชาญ                            | คัดเลือกผู้เชี่ยวชาญ 3–5 ท่านที่เป็นกลางและมีประสบการณ์ตรง ชี้แจงนิยามจุดประสงค์ให้ชัดเจนก่อนประเมิน |
| จุดประสงค์การเรียนรู้กำกวม                                          | เขียนจุดประสงค์เชิงพฤติกรรมที่ระบุการกระทำชัดเจน (Action Verbs) ตามหลัก Bloom                        |
| ภาษาในข้อสอบซับซ้อน วัดทักษะอื่นแทน (Construct-Irrelevant Variance) | ปรับคำถามให้ใช้ภาษาง่าย กระชับ สื่อความหมายเดียว เหมาะกับระดับผู้เรียน                               |

|                                                       |                                                                  |
|-------------------------------------------------------|------------------------------------------------------------------|
| ข้อสอบไม่ครอบคลุมโน้ต (Construct Underrepresentation) | ใช้ Test Blueprint กำกับให้ครอบคลุมทุกหน่วยและทุกระดับพฤติกรรม   |
| เกณฑ์ภายนอกคลาดเคลื่อน (Criterion Problem)            | เลือกเกณฑ์ที่มีคุณภาพและเป็นมาตรฐานก่อนนำมาหาสัมประสิทธิ์ความตรง |

ที่มา: ประมวลจาก Messick (1989); AERA, APA และ NCME (2014); พิชิต ฤทธิ์จรูญ (2564); และ สมนึก ภัททิยธนี (2558)

### 7.5.10 กรณีศึกษา (Case Study)

เพื่อให้เห็นความเชื่อมโยงระหว่างจุดประสงค์ เครื่องมือ และการแปลผล ขอเสนอกรณีศึกษา 2 กรณีในกล่องที่ 7.3 ซึ่งสะท้อนปัญหาความตรงคนละลักษณะ

#### กล่องที่ 7.3 กรณีศึกษาความตรงของเครื่องมือวัด

**กรณีที่ 1 (ความตรงเชิงเนื้อหา/IOC ต่ำ):** ครูประณีตสร้างแบบประเมิน “ทักษะความร่วมมือ” 3 ข้อ ตั้งจุดประสงค์ว่า “ผู้เรียนทำงานร่วมกับผู้อื่นได้” เมื่อผู้เชี่ยวชาญ 3 ท่านประเมินข้อ 2 เรื่อง “การรับฟังความเห็นผู้อื่น” ได้ +1, 0, -1 ( $\Sigma R = 0$ ,  $IOC = 0.00$ )

วิเคราะห์: ค่า  $IOC = 0.00$  สะท้อนว่าผู้เชี่ยวชาญเห็นไม่ตรงกัน เพราะคำว่า “การรับฟัง” กว้างและเป็นนามธรรมเกินไปจนไม่เห็นพฤติกรรมที่วัดได้

ปรับปรุง: เปลี่ยนนิยามเป็นพฤติกรรมที่สังเกตได้ เช่น “การจดบันทึกประเด็นที่เพื่อนนำเสนอในกลุ่ม” ผู้เชี่ยวชาญจึงเห็นตรงกันมากขึ้นและค่า  $IOC$  สูงพอผ่านเกณฑ์

**กรณีที่ 2 (ความตรงเชิงโครงสร้าง/ปนเปื้อน):** จุดประสงค์ว่า “วิเคราะห์สาเหตุของภาวะโลกร้อนได้” แต่ข้อสอบถามว่า “ก๊าซใดเป็นก๊าซเรือนกระจก” (วัดความจำ)

วิเคราะห์: แม้ผู้เชี่ยวชาญอาจให้  $IOC$  สูงหากดูแต่หัวข้อเนื้อหา แต่เกิด Construct-Irrelevant Variance เพราะข้อสอบวัด “ความจำ” ไม่ใช่ “การวิเคราะห์” ตามจุดประสงค์ หากไม่ปรับ จะตัดสินเด็กที่วิเคราะห์เก่งแต่จำชื่อก๊าซไม่ได้อย่างไม่เป็นธรรม (Messick, 1989)

ทั้งสองกรณีแสดงให้เห็นความสัมพันธ์โดยตรงระหว่าง “จุดประสงค์ที่ชัดเจน” “เครื่องมือที่ตรง” และ “การแปลผลที่เชื่อถือได้” การตรวจสอบความตรงจึงไม่ใช่เพียงพิธีกรรมทางสถิติ แต่เป็นกลไกประกันความยุติธรรมในการประเมินผู้เรียน (เชื่อมโยง CLO2 และ CLO3)

## 7.6 กิจกรรมการเรียนรู้การสอนและใบงาน

ขั้นก่อนเรียน: ทำแบบทดสอบก่อนเรียน (7.4) และกิจกรรม “ข้อสอบเจ้าปัญหา” ให้ผู้เรียนแชร์ประสบการณ์ข้อสอบที่รู้สึกว่า “ไม่ตรง” กับสิ่งที่เรียน เพื่อเชื่อมโยงสู่แนวคิดความตรง

ขั้นระหว่างเรียน: บรรยายเชิงปฏิสัมพันธ์ • ใช้ Think-Pair-Share

วิเคราะห์จำแนกประเภทความตรงจากตัวอย่างข้อสอบ • เวิร์กช็อปออกแบบ Test Blueprint และฝึกคำนวณค่า  $IOC$  จากสถานการณ์จำลอง

ใบงานที่ 7.1 วิเคราะห์และจำแนกประเภทความตรงจากกรณีศึกษา (รายกลุ่ม) — CLO2

กำหนดสถานการณ์การวัด 4–5 กรณี ให้กลุ่มจำแนกว่าแต่ละกรณีเกี่ยวข้องกับความจริงประเภทใด พร้อมให้เหตุผลเชื่อมโยงระหว่างจุดประสงค์ เครื่องมือวัด และการแปลผล แล้วเสนอวิธีตรวจสอบที่เหมาะสม

### ใบงานที่ 7.2 ออกแบบตารางวิเคราะห์ข้อสอบ (Test Blueprint) (รายกลุ่ม) — CLO3

หน่วยการเรียนรู้ 1 หน่วย ให้กลุ่มออกแบบ Test Blueprint กระจายจำนวนข้อสอบตามน้ำหนักเนื้อหาและระดับพฤติกรรม (Bloom) จำนวนรวม 20 ข้อ พร้อมอธิบายเหตุผลของการกระจายน้ำหนักเพื่อประกันความตรงเชิงเนื้อหา

### ใบงานที่ 7.3 คำนวณค่า IOC และเสนอแนวทางปรับปรุง (รายบุคคล/รายกลุ่ม) — CLO3

ออกแบบข้อสอบ 5 ข้อพร้อมระบุจุดประสงค์ นำคะแนนสมมติจากผู้เชี่ยวชาญ 3 ท่านมาคำนวณค่า IOC ตัดสินคุณภาพตามเกณฑ์ และเขียนสรุปผลพร้อมแนวทางปรับปรุงข้อที่ไม่ผ่าน (ประเมินด้วยรูบริค 4 ระดับในหัวข้อ 7.10)

ขั้นหลังเรียน: กิจกรรม Peer Review แลกเปลี่ยนเครื่องมือที่สร้างเพื่อวิพากษ์ความตรง ทำแบบทดสอบหลังเรียน (7.8) และแบบฝึกหัดท้ายบท (7.9)

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน–หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 7.7 สรุปบทเรียน

ความตรง (Validity) เป็นคุณภาพที่สำคัญที่สุดของเครื่องมือวัด หมายถึงการวัดได้ตรงตามสิ่งที่ต้องการวัดและตามวัตถุประสงค์ การวัดทางการศึกษาเป็นการวัดทางอ้อมที่มีความคลาดเคลื่อนเสมอ คะแนนต่ำจึงเป็น “ศูนย์สัมพัทธ์” ที่ไม่ได้แปลว่าผู้เรียนไร้ความรู้ ในมุมมองสมัยใหม่ความตรงเป็นความคิดรวบยอดเชิงเดี่ยวที่อาศัยหลักฐานหลายแหล่งสนับสนุนการตีความคะแนน (Messick, 1989; AERA, APA, & NCME, 2014) ความตรงจำแนกเป็น 3 ประเภท คือ เชิงเนื้อหา (ใช้ Test Blueprint และ IOC) เชิงโครงสร้าง (วัดตัวแปรแฝงด้วยเทคนิคกลุ่มรู้จัด/วิเคราะห์องค์ประกอบ) และเชิงเกณฑ์สัมพันธ์ (แบบตามสภาพและแบบพยากรณ์) ดัชนี IOC ที่พัฒนาโดย Rovinelli และ Hambleton (1977) ใช้ดุลยพินิจผู้เชี่ยวชาญอย่างน้อย 3 ท่าน ให้คะแนน +1/0/–1 แล้วคำนวณจากสูตร  $IOC = \sum R/N$  โดยข้อสอบที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไปจึงผ่านเกณฑ์ การตรวจสอบความตรงอย่างเป็นระบบจึงเป็นกระบวนการสร้างความยุติธรรมและความน่าเชื่อถือในการตัดสินผลการเรียนรู้ของผู้เรียน

## 7.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึก

คะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่  
ผู้สอนประกาศในชั้นเรียน

## 7.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

1. ครูวิทยาศาสตร์สร้างข้อสอบที่มีโจทย์เป็นบทอ่านยาวและภาษาซับซ้อน  
จนนักเรียนที่อ่อนภาษาไทยทำไม่ได้ทั้งที่เข้าใจเนื้อหาวิทยาศาสตร์  
จงวิเคราะห์ว่าสถานการณ์นี้กระทบความตรงอย่างไร  
และเสนอวิธีตรวจสอบและแก้ไขโดยใช้ตารางวิเคราะห์ข้อสอบและค่า IOC
2. เปรียบเทียบมุมมอง “ความตรงแบบดั้งเดิม” กับ “ความตรงเชิงเดียวสมัยใหม่” ตามแนวคิดของ Messick  
และมาตรฐาน AERA/APA/NCME (2014)
3. จำแนกความตรง 3 ประเภท พร้อมระบุจุดมุ่งหมายและวิธีตรวจสอบที่เหมาะสมของแต่ละประเภท
4. อธิบายหลักการและประโยชน์ของตารางวิเคราะห์ข้อสอบ (Test Blueprint)  
ต่อการประกันความตรงเชิงเนื้อหา
5. ผู้เชี่ยวชาญ 3 ท่านให้คะแนนข้อสอบข้อหนึ่งเป็น +1, 0, +1 จงคำนวณค่า IOC  
และตัดสินว่าผ่านเกณฑ์หรือไม่ พร้อมแสดงวิธีทำ
6. วิเคราะห์ความแตกต่างระหว่างความตรงตามสภาพ (Concurrent) และความตรงเชิงพยากรณ์  
(Predictive) พร้อมยกตัวอย่างสถานการณ์จริง
7. หากท่านพบว่าข้อสอบที่สร้างมีค่า IOC ต่ำ ท่านจะมีขั้นตอนปรับปรุงข้อสอบนั้นอย่างไรบ้าง
8. อภิปรายข้อความ “ความตรงสำคัญกว่าความเที่ยง” ท่านเห็นด้วยหรือไม่ เพราะเหตุใด

## 7.10 เครื่องมือวัดและประเมินผลประจำบท

ก) รุบริคประเมินใบงาน 7.3 (คำนวณ IOC และเสนอแนวทางปรับปรุง) แบบแยกองค์ประกอบ (Analytic) —  
เชื่อมโยง CLO3

| ประเด็นประเมิน                   | ดีมาก (4)                        | ดี (3)                  | พอใช้ (2)        | ปรับปรุง (1) |
|----------------------------------|----------------------------------|-------------------------|------------------|--------------|
| ความถูกต้องของการ<br>คำนวณ IOC   | ถูกต้องทุกข้อ                    | ถูกต้องเป็นส่วน<br>ใหญ่ | ผิดพลาดบางข้อ    | ผิดพลาดมาก   |
| การตัดสินคุณภาพตาม<br>เกณฑ์ 0.50 | ตัดสินถูกต้องครบ<br>ถ้วน         | ถูกต้องเป็นส่วน<br>ใหญ่ | ถูกบางส่วน       | ไม่ถูกต้อง   |
| คุณภาพข้อเสนอการ<br>ปรับปรุง     | เป็นเหตุเป็นผล<br>นำไปใช้ได้จริง | เหมาะสมเป็น<br>ส่วนใหญ่ | ยังกว้าง/ผิวเผิน | ไม่สอดคล้อง  |

|                                    |                           |                        |                      |              |
|------------------------------------|---------------------------|------------------------|----------------------|--------------|
| ความเชื่อมโยงจุดประสงค์-<br>ข้อสอบ | เชื่อมโยงชัดเจน<br>ทุกข้อ | ชัดเจนเป็นส่วน<br>ใหญ่ | เชื่อมโยงบาง<br>ส่วน | ไม่เชื่อมโยง |
|------------------------------------|---------------------------|------------------------|----------------------|--------------|

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 12$  จาก 16 (ร้อยละ 70 ขึ้นไป)

#### ข) RUBRIC จําแนกประเมินใบงาน 7.1 (วิเคราะห์/จําแนกประเภทความตรง) — เชื่อมโยง CLO2

| ประเด็นประเมิน                                          | ดี (3)                   | พอใช้ (2)               | ควรปรับปรุง (1) |
|---------------------------------------------------------|--------------------------|-------------------------|-----------------|
| ความถูกต้องในการจําแนกประเภท<br>ความตรง                 | จําแนกถูกต้องทุก<br>กรณี | ถูกต้องเป็นส่วน<br>ใหญ่ | ผิดพลาดหลายกรณี |
| การให้เหตุผลเชื่อมโยงจุดประสงค์-<br>เครื่องมือ-การแปลผล | เหตุผลชัดเจน<br>ครบถ้วน  | เหตุผลพอใช้             | เหตุผลไม่ชัดเจน |
| การเสนอวิธีตรวจสอบที่เหมาะสม                            | เหมาะสมและเป็นไป<br>ได้  | เหมาะสมบางส่วน          | ไม่เหมาะสม      |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 7$  จาก 9 (ร้อยละ 70 ขึ้นไป)

#### ค) สัดส่วนคะแนนประจําบท (รวม 100%)

| รายการประเมิน                                                   | เชื่อมโยง CLO | สัดส่วน |
|-----------------------------------------------------------------|---------------|---------|
| ใบงาน 7.1 วิเคราะห์/จําแนกประเภทความตรงจากกรณีศึกษา             | CLO2          | 20%     |
| ใบงาน 7.2 ออกแบบ Test Blueprint<br>และตรวจสอบความตรงเชิงเนื้อหา | CLO3          | 25%     |
| ใบงาน 7.3 คํานวณ IOC และเสนอแนวทางปรับปรุง                      | CLO3          | 30%     |
| แบบทดสอบหลังเรียน (วัดมโนทัศน์ความตรง)                          | CLO2          | 15%     |
| การมีส่วนร่วมในเวิร์กช็อปและการอภิปราย                          | CLO2          | 10%     |
| รวม                                                             | —             | 100%    |

## 7.11 เอกสารอ้างอิงประจําบท

### ภาษาไทย

ชวาล แพทย์กุล. (2518). เทคนิคการวัดผล (พิมพ์ครั้งที่ 6). วัฒนาพานิช.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏ  
เลย.

พิชิต ฤทธิจรูญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้. ศูนย์หนังสือจุฬาลงกรณ์  
มหาวิทยาลัย.

พิศิษฐ์ ตัฒทวนิช และพนา จินดาศรี. (2561). ความหมายที่แท้จริงของค่า IOC. วารสารการวัดผล  
การศึกษา มหาวิทยาลัยมหาสารคาม, 24(2), 3–12.

ไพศาล หวังพานิช. (2526). การวัดผลการศึกษา. ไทยวัฒนาพานิช.

ล้วน สายยศ และอังคณา สายยศ. (2543). เทคนิคการวัดผลการเรียนรู้. สุวีริยาสาส์น.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6). โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมนึก ภัททิยธนี. (2558). การวัดผลการศึกษา (พิมพ์ครั้งที่ 10). ประสานการพิมพ์.

สุทธิพงษ์ บุญผดุง. (2568). เอกสารประกอบการสอนรายวิชาการวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

สุพัฒน์ สุขมลสันต์. (2567). นานาวิธีในการหาความตรงเชิงเนื้อหาของเครื่องมือวิจัยเชิงปริมาณและการเลือกใช้วิธีที่เหมาะสม. วารสารภาษาปริทัศน์, 39, 75–102.

## ภาษาอังกฤษ

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214–221. <https://doi.org/10.1016/j.sapharm.2018.03.066>

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education & Macmillan.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research*, 2(2), 49–60.

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing*, 3(2), 163–171. [https://doi.org/10.1207/S15327574IJT0302\\_5](https://doi.org/10.1207/S15327574IJT0302_5)

## บทที่ 8

### ความเที่ยงของเครื่องมือวัดและประเมินผล

#### *Reliability of Measurement and Evaluation Instruments*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สัปดาห์ที่ 9 (หลังสอบกลางภาคสัปดาห์ที่ 8) • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

ต่อเนื่องจากความจริง (บทที่ 7) บทที่ 8 มุ่งที่คุณภาพอีกด้านที่สำคัญ คือ “ความเที่ยง” (Reliability) — ความคงเส้น คงวาของผลการวัด ผู้เรียนจะเข้าใจความหมายและความสัมพันธ์ระหว่างความเที่ยงกับความเที่ยง รู้จักวิธีหาความเที่ยง ทั้งวิธีสอบซ้ำ แบบทดสอบคู่ขนาน แบ่งครึ่งฉบับ (Spearman-Brown) และความสอดคล้องภายใน (KR-20, KR-21, สัมประสิทธิ์แอลฟา) ตลอดจนความเชื่อมั่นของการสังเกต/ผู้ประเมิน และฝึกคำนวณค่าความเที่ยงพร้อมแปลผล

#### แผนบริหารการสอนประจำบทที่ 8

| รายการ                | รายละเอียด                                                                                                                                                                                                                                                  |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) ความหมายและความสำคัญของความเที่ยง 2) ความสัมพันธ์ระหว่างความเที่ยงกับความเที่ยง 3) วิธีหาความเที่ยง (สอบซ้ำ/คู่ขนาน/แบ่งครึ่ง/ความสอดคล้องภายใน) 4) การคำนวณ KR-20, KR-21 และแอลฟา 5) ความเชื่อมั่นของการสังเกต/ผู้ประเมิน และปัจจัยที่มีผลต่อความเที่ยง |
| ความคิดรวบยอด         | ความเที่ยงคือความคงเส้นคงวาของผลการวัด เครื่องมือที่ดีต้องทั้งตรงและเที่ยง ความตรงสำคัญกว่าแต่ความเที่ยง เป็นเงื่อนไขจำเป็น มีหลายวิธีหาค่าความเที่ยงให้เลือกตามลักษณะเครื่องมือและการเก็บข้อมูล                                                            |
| จุดประสงค์การเรียนรู้ | วิเคราะห์วิธีหาความเที่ยงและความสัมพันธ์กับความเที่ยงได้ (CLO2) และคำนวณค่าความเที่ยง (เช่น KR-20) พร้อมแปลผลได้ (CLO3)                                                                                                                                     |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 8.1–8.5 ของบท)                                                                                                                                                                                                   |

|                       |                                                                                                                                                                                                                                                        |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| กิจกรรมการเรียนการสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + สาธิตการคำนวณ 3) ใบงาน 8.1 คำนวณความเที่ยงด้วย KR-20 จากข้อมูลจริง 4) ใบงาน 8.2 เลือกวิธีหาความเที่ยงให้เหมาะกับเครื่องมือ 5) อภิปรายความสัมพันธ์ตรง-เที่ยง 6) สรุป + แบบทดสอบหลังเรียน + แบบฝึกหัด |
| สื่อการเรียนการสอน    | เอกสารประกอบการสอนบทที่ 8 • สไลด์นำเสนอ • ตัวอย่างการคำนวณ KR-20/Spearman-Brown • เครื่องคิดเลข/สเปรดชีต • ใบงาน 8.1-8.2 • ตำราที่อ้างอิง                                                                                                              |
| การวัดและประเมินผล    | แบบทดสอบก่อน-หลังเรียน • ใบงาน 8.1 (เฉลยการคำนวณ KR-20) • ใบงาน 8.2 (rubric เลือกวิธี) • การอภิปราย • แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70)                                                                                                            |

## 8.1 ความคิดรวบยอด

ความเที่ยงหรือความเชื่อมั่น (Reliability) หมายถึง ความคงเส้นคงวาของผลการวัด กล่าวคือ หากนำเครื่องมือไปวัดซ้ำ ในเงื่อนไขเดียวกันย่อมได้ผลใกล้เคียงเดิม (ศิริชัย กาญจนวาสี, 2552) ความเที่ยงเกี่ยวข้องกับ ความคลาดเคลื่อนของการวัด ( $X = T + E$ ) ยิ่งความคลาดเคลื่อนน้อย ความเที่ยงยิ่งสูง วิธีหาความเที่ยงมีหลายแบบ ได้แก่ วิธีสอบซ้ำ (Test-retest) วิธีแบบทดสอบคู่ขนาน (Parallel forms) วิธีแบ่งครึ่งฉบับ (Split-half) ที่ปรับด้วยสูตรสเปียร์แมน-บราวน์ และวิธีความ สอดคล้องภายใน (Internal consistency) เช่น KR-20, KR-21 และสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach, 1951) เครื่องมือที่ดีต้องมีทั้งความตรงและความเที่ยง โดยความตรงสำคัญกว่า แต่ความเที่ยงเป็นเงื่อนไขจำเป็น เพราะเครื่องมือ ที่ไม่เที่ยงย่อมไม่อาจตรงได้ (Gronlund & Linn, 1990; AERA, APA, & NCME, 2014)

## 8.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 8 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                  | เชื่อมโยง CLO           |
|--------------------------------------------------------------|-------------------------|
| 1) อธิบายความหมายและความสำคัญของความเที่ยงได้                | CLO2 (เข้าใจ)           |
| 2) อธิบายความสัมพันธ์ระหว่างความตรงกับความเที่ยงได้          | CLO2 (เข้าใจ-วิเคราะห์) |
| 3) จำแนกวิธีหาความเที่ยงและเลือกให้เหมาะกับเครื่องมือได้     | CLO2 (วิเคราะห์)        |
| 4) คำนวณค่าความเที่ยง (เช่น KR-20) และแปลผลได้               | CLO3 (ประยุกต์)         |
| 5) ระบุปัจจัยที่มีผลต่อความเที่ยงและแนวทางเพิ่มความเที่ยงได้ | CLO2 (เข้าใจ)           |

### 8.3 คำสำคัญ

ความเที่ยง/ความเชื่อมั่น (Reliability) • วิธีสอบซ้ำ (Test-retest) • แบบทดสอบคู่ขนาน (Parallel forms) • วิธีแบ่งครึ่งฉบับ (Split-half) • สูตรสเปียร์แมน-บราวน์ (Spearman-Brown) • ความสอดคล้องภายใน (Internal consistency) • KR-20 / KR-21 • สัมประสิทธิ์แอลฟา (Cronbach's Alpha) • ความเชื่อมั่นของผู้ประเมิน (Inter-rater)

### 8.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 8.5 เนื้อหาวิชา

---

#### 8.5.1 ความหมายและความสำคัญของความเที่ยง

ความเที่ยงหรือความเชื่อมั่น (Reliability) หมายถึง ความคงเส้นคงวาหรือความสม่ำเสมอของผลการวัด กล่าวคือ เมื่อใช้เครื่องมือเดิมวัดสิ่งเดิมซ้ำภายใต้เงื่อนไขเดียวกัน ย่อมได้ผลใกล้เคียงกันทุกครั้ง (ศิริชัย กาญจนวาสี, 2552; พิชิต ฤทธิ์จรูญ, 2564) ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) คะแนนที่สังเกตได้ (Observed score:  $X$ ) ประกอบด้วยคะแนนจริง (True score:  $T$ ) และคะแนนความคลาดเคลื่อน (Error score:  $E$ ) ตามสมการ  $X = T + E$  ความเที่ยงจึงนิยามเป็นสัดส่วนของความแปรปรวนของคะแนนจริงต่อความแปรปรวนของคะแนนที่สังเกตได้ ( $\rho = \sigma^2 T / \sigma^2 X$ ) มีพิสัย 0.00–1.00 ยิ่งความคลาดเคลื่อนน้อย ความเที่ยงยิ่งเข้าใกล้ 1.00

ความคลาดเคลื่อนในการวัดแบ่งได้เป็น 2 ชนิด ความเที่ยงเกี่ยวข้องโดยตรงกับความคลาดเคลื่อนเชิงสุ่ม (random error) ซึ่งเกิดขึ้นอย่างไม่เป็นระบบและผลักให้คะแนนสูงหรือต่ำกว่าความจริงอย่างไม่แน่นอน ส่วนความคลาดเคลื่อนเชิงระบบ (systematic error) จะลำเอียงไปทางเดียวอย่างคงที่และเป็นเรื่องของ "ความตรง"

มากกว่า (Nitko & Brookhart, 2019; AERA, APA, & NCME, 2014) ดังนั้นเครื่องมือที่เที่ยงคือเครื่องมือที่ลดความคลาดเคลื่อนเชิงสุ่มให้เหลือน้อยที่สุด ดังแสดงการเปรียบเทียบในตารางที่ 8.1

ตารางที่ 8.1 เปรียบเทียบความคลาดเคลื่อนเชิงสุ่มกับเชิงระบบ

| ประเด็น        | ความคลาดเคลื่อนเชิงสุ่ม (Random)            | ความคลาดเคลื่อนเชิงระบบ (Systematic)        |
|----------------|---------------------------------------------|---------------------------------------------|
| ทิศทาง         | ไม่แน่นอน สูง/ต่ำสลับกันไป                  | ลำเอียงไปทางเดียวอย่างคงที่                 |
| ตัวอย่างสาเหตุ | การเดา ความเหนื่อยล้า<br>เสียงรบกวนชั่วคราว | ข้อสอบออกนอกเนื้อหา เฉลยผิด<br>เกณฑ์ลำเอียง |
| ผลกระทบหลักต่อ | ความเที่ยง (Reliability) ลดลง               | ความตรง (Validity) เสียไป                   |
| แนวทางลด       | เพิ่มจำนวนข้อ จัดสภาพสอบมาตรฐาน             | ทบทวนเนื้อหา/เฉลย<br>ใช้ผังข้อสอบและรูบรีค  |

ที่มา: ดัดแปลงจาก Nitko & Brookhart, 2019

เมื่อพิจารณาที่มาของความคลาดเคลื่อน (E) ในทางปฏิบัติ ครูสามารถจำแนกแหล่งที่ควบคุมได้ออกเป็น 4 ด้าน การรู้ว่าคุณสมบัติความคลาดเคลื่อนมาจากแหล่งใดช่วยให้วางแผนลดความคลาดเคลื่อนและยกระดับความเที่ยงได้ตรงจุด (ศิริชัย กาญจนวาสี, 2552; Nitko & Brookhart, 2019) ดังกล่องด้านล่าง

#### กล่อง: แหล่งของความคลาดเคลื่อนในการวัด 4 ด้าน

- 1) ตัวผู้สอบ (examinee): ความเหนื่อยล้า ความวิตกกังวล การเดา สุขภาพ/ความพร้อมในวันสอบ
- 2) ตัวผู้ตรวจ (rater): อคติ ความไม่คงเส้นคงวาในการให้คะแนน การไม่มีเกณฑ์/รูบรีคที่ชัดเจน
- 3) ตัวเครื่องมือ (instrument): ข้อคำถามกำกวม ข้อสอบสั้นเกินไป ตัวเลือก/เฉลยบกพร่อง
- 4) สภาพแวดล้อม (environment): เสียงรบกวน แสง อุณหภูมิ เวลาไม่เพียงพอ คำชี้แจงไม่ชัด

ตัวอย่างในชั้นเรียนไทย: ครูให้นักเรียนคนเดิมซึ่งมีความรู้เท่าเดิมทำข้อสอบเก็บคะแนนชุดเดิม 40 ข้อ ในสองสัปดาห์ติดกัน ครั้งแรกได้ 27 คะแนน แต่ครั้งที่สองได้เพียง 15 คะแนน ทั้งที่ไม่ได้เรียนเพิ่มหรือลืม ความผันผวนที่มากเช่นนี้สะท้อนว่าเครื่องมือมีความคลาดเคลื่อนเชิงสุ่มสูงและมีความเที่ยงต่ำ คะแนนที่ได้จึงเชื่อถือไม่ได้ หากนำไปตัดสินได้-ตก ย่อมไม่ยุติธรรมต่อผู้เรียน

ความสำคัญของความเที่ยงจึงอยู่ที่การเป็นหลักประกันความยุติธรรมและความน่าเชื่อถือของคะแนน ก่อนนำผลไปตัดสินใจทางการศึกษา เช่น การตัดเกรด การคัดเลือก หรือการวินิจฉัยจุดอ่อนของผู้เรียน อย่างไรก็ตาม

ตามพิ้งระวังว่าความเที่ยงไม่ใช่คุณสมบัติถาวรของข้อสอบ แต่ขึ้นกับกลุ่มผู้สอบและสภาพการสอบด้วย ค่าความเที่ยงที่รายงานไว้จึงใช้ได้กับบริบทที่ใกล้เคียงกับตอนทดลองใช้เท่านั้น (พัชรินทร์ ชมภูวิเศษ, 2559)

## 8.5.2 ความสัมพันธ์ระหว่างความตรงกับ ความเที่ยง

ความตรง (Validity) คือการวัดได้ตรงในสิ่งที่มุ่งวัด ส่วนความเที่ยง (Reliability) คือการวัดได้คงเส้นคงวา ทั้งสองสัมพันธ์กันแต่ไม่ใช่สิ่งเดียวกัน วิธีเข้าใจที่นิยมใช้คือการเปรียบกับการยิงเป้า ความเที่ยงคือการยิงให้ลูกธนูเกาะกลุ่มกัน ส่วนความตรงคือการยิงให้เข้าจุดศูนย์กลางของเป้า (Gronlund & Linn, 1990) ดังจำแนกสี่สถานการณ์ในตารางที่ 8.2

ตารางที่ 8.2 ความสัมพันธ์ความตรง-ความเที่ยงผ่านการเปรียบกับการยิงเป้า

| สถานการณ์ยิงเป้า       | ตรง / เที่ยง              | ความหมายเชิงการวัด                              |
|------------------------|---------------------------|-------------------------------------------------|
| ลูกเกาะกลุ่มกลางเป้า   | ตรง + เที่ยง              | เครื่องมือที่ดีที่สุด วัดถูกสิ่งและคงที่        |
| ลูกเกาะกลุ่มแต่นอกกลาง | เที่ยง แต่ไม่ตรง          | คงที่แต่วัดผิดสิ่ง (อันตราย เพราะดูน่าเชื่อถือ) |
| ลูกกระจายรอบกลาง       | ตรงโดยเฉลี่ย แต่ไม่เที่ยง | เกิดได้ยากและไม่เสถียร                          |
| ลูกกระจายและนอกกลาง    | ไม่ตรง + ไม่เที่ยง        | เครื่องมือใช้ไม่ได้                             |

ที่มา: ดัดแปลงจาก Gronlund & Linn, 1990; ศิริชัย กาญจนวาสี, 2552

จากตารางสรุปเป็นหลักการสำคัญสามข้อ ดังนี้

- เครื่องมือที่ตรง ต้องเที่ยงเสมอ ความเที่ยงจึงเป็นเงื่อนไขจำเป็น (necessary condition) ของความตรง
- เครื่องมือที่เที่ยง อาจไม่ตรงก็ได้ คือวัดคงเส้นคงวาแต่วัดผิดสิ่ง (ความเที่ยงไม่เพียงพอต่อความตรง)
- ความตรงสำคัญกว่าความเที่ยง แต่เครื่องมือที่ดีต้องมีทั้งสองคุณสมบัติคู่กัน

**ตัวอย่างในชั้นเรียนไทย:** หากครูคณิตศาสตร์ออกข้อสอบที่มีโจทย์ปัญหาเป็นภาษายาวและซับซ้อนเกินระดับชั้น ผลที่วัดได้อาจคงที่ทุกครั้ง (เที่ยงสูง) แต่แท้จริงกำลังวัด "ทักษะการอ่านจับใจความ" มากกว่า "ความสามารถทางคณิตศาสตร์" คะแนนจึงเที่ยงแต่ไม่ตรงตามจุดประสงค์ การตีความว่านักเรียนอ่อนคณิตศาสตร์จึงผิดพลาด

การประยุกต์และข้อควรระวัง: ในทางปฏิบัติควรตรวจสอบความตรงเชิงเนื้อหาด้วยผังข้อสอบและผู้เชี่ยวชาญ (บทที่ 7) ควบคู่กับการหาความเที่ยง อย่าด่วนสรุปว่าเครื่องมือดีเพียงเพราะค่าความเที่ยงสูง เพราะค่าความเที่ยงหรือค่าสัมประสิทธิ์แอลฟาที่สูงไม่ได้รับประกันว่าเครื่องมือวัดตรงตามที่ตั้งใจ (Taber, 2018) ความเที่ยงและความตรงจึงต้องพิจารณาประกอบกันเสมอ

### 8.5.3 วิธีหาความเที่ยง

วิธีหาความเที่ยงจัดกลุ่มได้ 4 แนวทางตามแหล่งความคลาดเคลื่อนที่ต้องการควบคุม ได้แก่ (1) ความคงที่ตามเวลา ด้วยวิธีสอบซ้ำ (2) ความเท่าเทียมของแบบทดสอบ ด้วยวิธีคู่ขนาน (3) ความสอดคล้องภายในฉบับ ด้วยวิธีแบ่งครึ่งฉบับ KR-20/KR-21 และสัมประสิทธิ์แอลฟา และ (4) ความสอดคล้องของผู้ประเมิน สำหรับเครื่องมือที่ใช้ดุลยพินิจ (ล้วน สายยศ และอังคณา สายยศ, 2543; พวงรัตน์ ทวีรัตน์, 2540) สรุปได้ดังตารางที่ 8.3

ตารางที่ 8.3 วิธีหาความเที่ยงและสถิติที่ใช้

| วิธี                         | ลักษณะการเก็บข้อมูล                             | สถิติ/สูตรที่ใช้              |
|------------------------------|-------------------------------------------------|-------------------------------|
| สอบซ้ำ (Test-retest)         | ใช้ข้อสอบชุดเดิมสอบ 2 ครั้ง วัดความคงที่ตามเวลา | สหสัมพันธ์เพียร์สัน (r)       |
| คู่ขนาน (Parallel forms)     | ใช้ข้อสอบ 2 ฉบับที่เท่าเทียมกัน                 | สหสัมพันธ์เพียร์สัน (r)       |
| แบ่งครึ่งฉบับ (Split-half)   | แบ่งข้อสอบเป็น 2 ครึ่ง หาความสัมพันธ์           | เพียร์สัน + สเปียร์แมน-บราวน์ |
| ความสอดคล้องภายใน (KR-20/21) | สอบครั้งเดียว ข้อสอบตรวจ 0-1                    | KR-20, KR-21                  |
| ความสอดคล้องภายใน (แอลฟา)    | สอบครั้งเดียว เครื่องมือหลายระดับ (มาตรฐานค่า)  | สัมประสิทธิ์แอลฟาของครอนบาค   |

ที่มา: ดัดแปลงจาก พัทรินทร์ ชมภูวิเศษ, 2559, น.201-203

#### กล่องสูตรสำคัญ

สเปียร์แมน-บราวน์ (ปรับค่าแบ่งครึ่งให้เต็มฉบับ):  $r(tt) = 2 \cdot r(hh) / (1 + r(hh))$

KR-20:  $r(tt) = [k/(k-1)] \cdot [1 - (\sum pq / S^2t)]$  เมื่อ k = จำนวนข้อ, p = สัดส่วนตอบถูก, q = 1-p

KR-21:  $r(tt) = [k/(k-1)] \cdot [1 - (\bar{X}(k-\bar{X}) / (k \cdot S^2t))]$  ใช้เมื่อข้อสอบยากใกล้เคียงกัน

สัมประสิทธิ์แอลฟา:  $\alpha = [k/(k-1)] \cdot [1 - (\sum S^2i / S^2t)]$  เมื่อ  $S^2i$  = ความแปรปรวนรายข้อ

(ที่มา: พัทรินทร์ ชมภูวิเศษ, 2559, น.202, 208)

#### ระดับของมาตรวัดกับการเลือกสถิติหาความเที่ยง

การเลือกสถิติหาความเที่ยงยังขึ้นอยู่กับ "ระดับของมาตรวัด" (level of measurement) ของคะแนนด้วย สตีเวนส์ (Stevens, 1946) จำแนกมาตรวัดเป็น 4 ระดับจากหยาบไปละเอียด ได้แก่ นามบัญญัติ จัดอันดับ อัตรภาค และอัตราส่วน (สมนึก ภัททิยธนี, 2551) คะแนนจากแบบทดสอบและแบบวัดทางการศึกษาส่วน

ใหญ่อยู่ในระดับอันตรภาคหรืออัตราส่วน จึงใช้สถิติความเที่ยงเชิงปริมาณได้ และต้องเลือกสูตรให้ตรงกับระดับข้อมูล ดังสรุปในตารางที่ 8.4

ตารางที่ 8.4 ระดับของมาตรวัดกับสถิติความเที่ยงที่เหมาะสม

| ระดับมาตรวัด         | ลักษณะสำคัญ               | ตัวอย่าง                | สถิติความเที่ยงที่เหมาะสม                  |
|----------------------|---------------------------|-------------------------|--------------------------------------------|
| นามบัญญัติ (Nominal) | จัดกลุ่ม ไม่มีปริมาณ      | เพศ หมู่เลือด           | ความสอดคล้องผู้ประเมิน (K, Scott's $\pi$ ) |
| จัดอันดับ (Ordinal)  | บอกลำดับ ช่วงไม่เท่ากัน   | ดี/ปานกลาง/อ่อน         | สหสัมพันธ์อันดับ, Kendall's W              |
| อันตรภาค (Interval)  | ช่วงเท่ากัน ไม่มีศูนย์แท้ | คะแนนสอบ<br>อุณหภูมิ °C | KR-20 (0–1), แอลฟา (มาตรประมาณค่า)         |
| อัตราส่วน (Ratio)    | มีศูนย์แท้ คูณหารได้      | จำนวนข้อถูก เวลา        | เพียร์สัน r, แอลฟา                         |

ที่มา: ดัดแปลงจาก Stevens, 1946; สมนึก ภัททิยธนี, 2551

ข้อพึงระวังคือ มาตรอันตรภาคของคะแนนสอบมี "ศูนย์สมมติ" (relative zero) ไม่ใช่ "ศูนย์แท้" (absolute zero) กล่าวคือ การได้ 0 คะแนนไม่ได้แปลว่าผู้เรียนไม่มีความรู้เลย แต่อาจเพราะข้อสอบยากเกินไปหรือวัดไม่ตรงจุด การวัดทางการศึกษาเป็นการวัดทางอ้อมที่มีความคลาดเคลื่อนแฝงเสมอ การตรวจสอบความเที่ยงจึงเป็นขั้นตอนที่ขาดไม่ได้เพื่อยืนยันว่าคะแนนที่ได้ไม่ใช่เรื่องบังเอิญ (สมนึก ภัททิยธนี, 2551)

แต่ละวิธีมีจุดเด่นและข้อจำกัดต่างกัน วิธีสอบซ้ำสะท้อนความคงที่ตามเวลาแต่ต้องเว้นช่วงให้พอเหมาะ หากเว้นสั้นไปผู้สอบจะจำคำตอบเดิม (carry-over) หากเว้นยาวไปความรู้จะเปลี่ยน วิธีคู่ขนานเสี่ยงปัญหาการจำได้แต่สร้างข้อสอบสองฉบับที่เท่าเทียมกันจริงได้ยาก ส่วนวิธีแบ่งครึ่งฉบับและความสอดคล้องภายในสะดวกเพราะสอบเพียงครั้งเดียว โดย KR-20 และ KR-21 พัฒนาขึ้นสำหรับข้อสอบที่ตรวจเป็น 0–1 (Kuder & Richardson, 1937) ขณะที่สัมประสิทธิ์แอลฟาเป็นการขยาย KR-20 ให้ใช้กับเครื่องมือที่ให้คะแนนหลายระดับได้ (Cronbach, 1951)

ตัวอย่างการเลือกวิธีให้เหมาะกับเครื่องมือในชั้นเรียนไทย: ข้อสอบปรนัยเลือกตอบที่ตรวจถูก–ผิด ควรใช้ KR-20 แบบสอบถามเจตคติชนิดมาตรประมาณค่า 5 ระดับ ควรใช้สัมประสิทธิ์แอลฟา ส่วนการให้คะแนนชิ้นงานหรือการสังเกตพฤติกรรมด้วยรูบริก ควรตรวจความเชื่อมั่นระหว่างผู้ประเมิน (inter-rater) วิธีแบ่งครึ่งฉบับเหมาะเมื่อต้องการประมาณความเที่ยงอย่างรวดเร็วจากการสอบครั้งเดียวโดยปรับค่าด้วยสูตรสเปียร์แมน–บราวน์ เช่น ถ้าความสัมพันธ์ของคะแนนสองครั้งเท่ากับ 0.60 ค่าความเที่ยงทั้งฉบับจะเท่ากับ  $2(0.60)/(1+0.60) = 0.75$

**ข้อควรระวัง:** ต้องเลือกสูตรให้ตรงกับชนิดของคะแนน หากนำ KR-20 ไปใช้กับมาตราประมาณค่าหรือข้อสอบอัตนัยที่ให้คะแนนหลายระดับจะได้ค่าที่ผิด และการสอบซ้ำในเวลากระชั้นชิดจะทำให้ค่าความเที่ยงสูงเกินจริง เพราะผู้สอบจำคำตอบเดิมได้

#### 8.5.4 ตัวอย่างการคำนวณและการแปลผล

ตัวอย่างที่ 1 (KR-20): ในการสอบวิชาคณิตศาสตร์ด้วยข้อสอบปรนัย 5 ข้อ กับผู้สอบ 10 คน คำนวณได้ผลรวมของ pq รายข้อ  $\sum pq = 1.14$  และความแปรปรวนของคะแนนรวม  $S^2t = 2.71$  แทนค่าในสูตร KR-20 ดังกล่อง จะได้ค่าความเที่ยงประมาณ 0.72 อยู่ในเกณฑ์ "ดี" (ดัดแปลงจาก พัทรินทร์ ชมภูวิเศษ, 2559)

##### การแทนค่า KR-20

$$r(tt) = [k/(k-1)] \cdot [1 - (\sum pq / S^2t)] = [5/4] \cdot [1 - (1.14 / 2.71)]$$

$$= 1.25 \cdot [1 - 0.4207] = 1.25 \cdot 0.5793 \approx 0.72$$

แปลผล: แบบทดสอบฉบับนี้มีความเที่ยงระดับดี (ดูเกณฑ์การแปลผลในตารางที่ 8.6)

ตัวอย่างที่ 2 (สัมประสิทธิ์แอลฟา): แบบวัดเจตคติต่อวิชาที่เรียนชนิดมาตราประมาณค่า 5 ระดับ จำนวน 5 ข้อ นำไปทดลองใช้กับนักเรียน 6 คน ได้คะแนนดังตารางที่ 8.5 เนื่องจากเป็นข้อมูลหลายระดับจึงใช้สัมประสิทธิ์แอลฟา

ตารางที่ 8.5 คะแนนแบบวัดเจตคติ 5 ข้อ ของผู้ตอบ 6 คน (มาตราประมาณค่า 1-5)

| ผู้ตอบ | ข้อ 1 | ข้อ 2 | ข้อ 3 | ข้อ 4 | ข้อ 5 | รวม |
|--------|-------|-------|-------|-------|-------|-----|
| คน A   | 5     | 3     | 4     | 5     | 4     | 21  |
| คน B   | 4     | 4     | 5     | 3     | 4     | 20  |
| คน C   | 3     | 2     | 2     | 3     | 3     | 13  |
| คน D   | 2     | 3     | 3     | 2     | 2     | 12  |
| คน E   | 4     | 5     | 4     | 5     | 3     | 21  |
| คน F   | 3     | 2     | 3     | 2     | 4     | 14  |

ที่มา: ผู้เขียน

##### การคำนวณสัมประสิทธิ์แอลฟา

ความแปรปรวนรายข้อ ( $S^2i$ ): ข้อ 1 = 1.10, ข้อ 2 = 1.37, ข้อ 3 = 1.10, ข้อ 4 = 1.87, ข้อ 5 = 0.67 →

$$\sum S^2i = 6.10$$

ความแปรปรวนของคะแนนรวม:  $S^2t = 18.17$

$$\alpha = [k/(k-1)] \cdot [1 - (\sum S^2i / S^2t)] = [5/4] \cdot [1 - (6.10 / 18.17)] = 1.25 \cdot 0.664 \approx 0.83$$

แปลผล: แบบวัดเจตคติฉบับนี้มีความเที่ยงระดับดี (สูงกว่าเกณฑ์ 0.70)

นอกจากค่าความเที่ยงแล้ว ควรพิจารณาความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement: SEM) ควบคู่กัน เพราะ SEM บอก "ช่วงความผันผวน" ของคะแนนรายบุคคลได้เป็นรูปธรรม ดังกล่องด้านล่าง

#### หมายเหตุ: ความคลาดเคลื่อนมาตรฐานของการวัด (SEM)

สูตร:  $SEM = S \cdot \sqrt{1 - r}$  เมื่อ  $S$  = ส่วนเบี่ยงเบนมาตรฐานของคะแนนรวม,  $r$  = ค่าความเที่ยง

ตัวอย่าง: ถ้า  $S = 8$  และ  $r = 0.84 \rightarrow SEM = 8 \cdot \sqrt{1 - 0.84} = 8 \cdot 0.4 = 3.2$  คะแนน

การแปลความ: นักเรียนที่ได้ 30 คะแนน มีคะแนนจริงน่าจะอยู่ราว  $30 \pm 3.2$  (ประมาณ 26.8–33.2)

ยิ่งความเที่ยงสูง SEM ยิ่งเล็ก คะแนนยิ่งแม่นยำ

(ที่มา: ดัดแปลงจาก Nitko & Brookhart, 2019)

ตารางที่ 8.6 เกณฑ์การแปลผลค่าความเที่ยง (โดยทั่วไป)

| ค่าความเที่ยง | การแปลผล                                       |
|---------------|------------------------------------------------|
| 0.80–1.00     | สูง/ดีมาก (แบบทดสอบผลสัมฤทธิ์ควร $\geq 0.70$ ) |
| 0.61–0.79     | ดี/ปานกลางค่อนข้างสูง                          |
| 0.40–0.60     | ปานกลาง/พอใช้                                  |
| ต่ำกว่า 0.40  | ต่ำ ควรปรับปรุงเครื่องมือ                      |

(ที่มา: ดัดแปลงจาก พิชิต ฤทธิ์จรูญ, 2564; ล้วน สายยศ และอังคณา สายยศ, 2543)

ตัวอย่างการแปลผลในงานจริง: งานวิจัยไทยที่พัฒนาแบบทดสอบวัดผลสัมฤทธิ์ 30 ข้อ รายงานค่าความเที่ยง KR-20 = 0.70 และ KR-21 = 0.68 ซึ่งผ่านเกณฑ์ขั้นต่ำ 0.70 พอดี จึงสรุปได้ว่าแบบทดสอบมีความเที่ยงเพียงพอที่จะนำไปใช้จริง (ไชยวัฒน์ อารีโรจน์, 2568) หากครูดำนวนแล้วพบค่าต่ำกว่าเกณฑ์ ควรปรับปรุงข้อสอบที่กำกวม ตัดข้อที่ง่ายหรือยากเกินไป และเพิ่มจำนวนข้อที่มีคุณภาพ

**ข้อควรระวัง:** ค่าความเที่ยงหรือค่าแอลฟาที่สูงผิดปกติ (เช่น เกิน 0.95 ในข้อสอบทั่วไป) อาจไม่ใช่เรื่องดีเสมอ เพราะมักบ่งชี้ว่าข้อคำถามหลายข้อซ้ำซ้อนกัน (redundancy) วัดสิ่งเดียวกันมากเกินไปจนเสียความครอบคลุมของเนื้อหา การรายงานค่าแอลฟาจึงควรตีความอย่างระมัดระวังประกอบกับหลักฐานความตรง ไม่ใช่ยึดเพียงตัวเลขเดียว (Taber, 2018)

### 8.5.5 ความเชื่อมั่นของผู้ประเมินและปัจจัยที่มีผลต่อความเที่ยง

สำหรับเครื่องมือที่ตรวจให้คะแนนด้วยดุลยพินิจ เช่น การเรียงความ ชี้นงาน หรือแบบสังเกตพฤติกรรม การให้คะแนนของผู้ประเมินเป็นแหล่งความคลาดเคลื่อนสำคัญ จึงต้องตรวจสอบความเชื่อมั่นของผู้ประเมิน 2 ลักษณะ คือ ความสอดคล้องในตัวผู้ประเมินคนเดียวเมื่อตรวจซ้ำ (Intra-rater) และความสอดคล้องระหว่างผู้ประเมินตั้งแต่ 2 คนขึ้นไป (Inter-rater) (บุญธรรม กิจปรีดาบริสุทธิ์, 2549; พัทรินทร์ ชมภูวิเศษ, 2559) ดัชนีที่ใช้สรุปไว้ในตารางที่ 8.7

ตารางที่ 8.7 ดัชนีตรวจสอบความเชื่อมั่นของผู้ประเมิน

| ดัชนี/สถิติ                                             | ใช้เมื่อ                           | ลักษณะข้อมูล               |
|---------------------------------------------------------|------------------------------------|----------------------------|
| สหสัมพันธ์เพียร์สัน (r)                                 | ผู้ประเมิน 2 คน คะแนนต่อเนื่อง     | คะแนนเป็นช่วง/อัตราส่วน    |
| ดัชนีของสก็อตต์ (Scott's $\pi$ ) /<br>โคเฮน (Cohen's K) | ผู้ประเมิน 2 คน จัดประเภท/จัดระดับ | ข้อมูลนามบัญญัติ/จัดอันดับ |
| สัมประสิทธิ์ความสอดคล้องเคนดอลล์<br>(W)                 | ผู้ประเมินตั้งแต่ 3 คน จัดอันดับ   | ข้อมูลอันดับที่            |
| สหสัมพันธ์ภายในชั้น (ICC)                               | ผู้ประเมินหลายคน RUBRICหลายระดับ   | คะแนนเป็นช่วง              |

ที่มา: ดัดแปลงจาก บุญธรรม กิจปรีดาบริสุทธิ์, 2549; พัทรินทร์ ชมภูวิเศษ, 2559, น.216

ตัวอย่างในชั้นเรียนไทย: ครูภาษาไทยสองคนช่วยกันตรวจเรียงความของนักเรียน 30 คน โดยใช้ RUBRICชุดเดียวกัน แล้วนำคะแนนของครูทั้งสองมาหาสหสัมพันธ์ ถ้าได้ค่าสูง (เช่น 0.85) แสดงว่า RUBRICชัดเจนและผู้ประเมินเข้าใจเกณฑ์ตรงกัน ผลการให้คะแนนจึงเชื่อถือได้ ไม่ขึ้นกับว่าใครเป็นผู้ตรวจ

ปัจจัยที่มีผลต่อความเที่ยงและแนวทางเพิ่มความเที่ยงสรุปไว้ในตารางที่ 8.8 โดยหลักการคือการลดความคลาดเคลื่อนเชิงสุ่มทุกทาง ทั้งจากตัวข้อสอบ ผู้สอบ ผู้ตรวจ และสภาพการสอบ (ศิริชัย กาญจนวาสี, 2552; สมนึก ภัททิยธนี, 2551)

ตารางที่ 8.8 ปัจจัยที่มีผลต่อความเที่ยงและแนวทางเพิ่มความเที่ยง

| ปัจจัย                       | ผลต่อความเที่ยง                                       | แนวทางเพิ่มความเที่ยง                         |
|------------------------------|-------------------------------------------------------|-----------------------------------------------|
| จำนวนข้อ                     | ข้อมากพอช่วยเพิ่มความเที่ยง<br>(ตามสเปียร์แมน-บราวน์) | เพิ่มจำนวนข้อที่มีคุณภาพอย่างเหมาะสม          |
| ความชัดเจน/ความเป็นปร<br>นัย | ข้อกำกวมและเกณฑ์คลุมเครือลดความเที่ยง                 | เขียนข้อให้ชัด<br>ใช้ RUBRIC/เฉลยที่เป็นปรนัย |

|                           |                                                |                                                           |
|---------------------------|------------------------------------------------|-----------------------------------------------------------|
| ความยากของข้อ             | ข้อง่ายหรือยากเกินไปทำให้คะแนนเกาะกลุ่ม        | คุมความยากให้ $p \approx 0.20-0.80$ (เฉลี่ย $\sim 0.50$ ) |
| ความเป็นเอกพันธ์          | เนื้อหา/พฤติกรรมที่วัดกระจายลดความเที่ยง       | ให้ข้อในฉบับวัดโครงสร้างเดียวกัน                          |
| การกระจายของกลุ่ม/สภาพสอบ | กลุ่มคล้ายกันมากหรือสภาพสอบไม่นิ่งลดความเที่ยง | จัดสภาพสอบให้เป็นมาตรฐานเดียวกัน                          |

ที่มา: ดัดแปลงจาก ศิริชัย กาญจนวาสี, 2552; สมนึก ภัททิยธนี, 2551

การประยุกต์และข้อควรระวัง: แม้การเพิ่มจำนวนข้อจะช่วยเพิ่มความเที่ยง แต่มีจุดเหมาะสม (optimum point) หากข้อสอบยาวจนผู้สอบเหนื่อยล้าและเริ่มเดา ความคลาดเคลื่อนใหม่จะเกิดขึ้นและความเที่ยงกลับลดลง ครูจึงควรสมดุลระหว่างความยาวของแบบทดสอบกับเวลาและสมาธิของผู้เรียน และควรหาความเที่ยงทุกครั้งที่น่าเครื่องมือไปใช้กับกลุ่มใหม่ เพราะค่าความเที่ยงอาจเปลี่ยนไปตามบริบท

## 8.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ใบงานที่ 8.1 คำนวณความเที่ยงด้วย KR-20 (รายบุคคล)

คำชี้แจง: กำหนดตารางผลการตอบข้อสอบ (0/1) ของผู้สอบ 10 คน 5 ข้อ ให้ผู้เรียนคำนวณ  $p$ ,  $q$ ,  $\sum pq$ ,  $S^2$  และค่า KR-20 พร้อมแปลผลตามเกณฑ์ (มีเฉลยขั้นตอนในหัวข้อ 8.10) — เชื่อมโยง CLO3

### ใบงานที่ 8.2 เลือกวิธีหาความเที่ยงให้เหมาะกับเครื่องมือ (รายกลุ่ม)

คำชี้แจง: กำหนดสถานการณ์เครื่องมือหลายชนิด (ข้อสอบปรนัย/มาตราประมาณค่า/รูบริกการสังเกต) ให้กลุ่มเลือกวิธีหา ความเที่ยงที่เหมาะสมพร้อมเหตุผล และระบุเกณฑ์การแปลผล (ประเมินด้วยรูบริกในหัวข้อ 8.10) — เชื่อมโยง CLO2

**การสะท้อนคิดและการให้ข้อมูลย้อนกลับ:** ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงาน และกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 8.7 สรุปบทเรียน

ความเที่ยงหรือความเชื่อมั่น (Reliability) คือความคงเส้นคงวาของผลการวัด มีพิสัย 0.00–1.00 ค่ายิ่งสูงยิ่งดี เครื่องมือ ที่ดีต้องมีทั้งความตรงและความเที่ยง โดยความตรงสำคัญกว่าแต่ความเที่ยงเป็นเงื่อนไขจำเป็น (เครื่องมือที่ไม่เที่ยงย่อม ไม่ตรง) วิธีหาความเที่ยงมีหลายแบบ ได้แก่ วิธีสอบซ้ำ วิธีคู่นาน (ใช้สหสัมพันธ์เพียร์

สัน) วิธีแบ่งครึ่งฉบับ (ปรับด้วย สเปียร์แมน-บราวน์) และความสอดคล้องภายใน (KR-20, KR-21 สำหรับข้อสอบ 0-1 และสัมประสิทธิ์แอลฟาสำหรับเครื่องมือ หลายระดับ) สำหรับเครื่องมือที่ใช้ดุลยพินิจต้องตรวจสอบความเชื่อมั่นของผู้ประเมิน (Intra/Inter-rater) ปัจจัยที่เพิ่มความเที่ยง ได้แก่ จำนวนข้อที่มากพอ ความชัดเจนของข้อคำถามและเกณฑ์ และสภาพการสอบที่เป็นมาตรฐาน เมื่อเครื่องมือมีทั้งความตรง และความเที่ยงแล้ว บทถัดไปจะวิเคราะห์คุณภาพข้อสอบรายข้อ

## 8.8 แบบทดสอบหลังเรียน

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 8.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

1. ครูให้นักเรียนกลุ่มเดิมทำแบบทดสอบชุดเดียวกันซ้ำหลังผ่านไปหนึ่งสัปดาห์ พบว่าลำดับที่ของคะแนนสลับกันอย่างมาก จงวิเคราะห์ว่าปรากฏการณ์นี้สะท้อนปัญหาคุณภาพด้านใด มีแหล่งความคลาดเคลื่อนใดเป็นไปได้บ้าง (อธิบายผ่านสมการ  $X = T + E$ ) และเสนอแนวทางแก้ไขอย่างน้อย 2 วิธี
2. อธิบายความสัมพันธ์ระหว่างความตรงกับความเที่ยง โดยใช้การเปรียบเทียบการยิงเป้า
3. เปรียบเทียบวิธีหาความเที่ยง 4 วิธี พร้อมระบุว่าวิธีใดเหมาะกับเครื่องมือแบบใด
4. กำหนดข้อมูลข้อสอบ 5 ข้อ ผู้สอบ 10 คน คำนวณค่า KR-20 และแปลผล
5. อธิบายความแตกต่างระหว่าง Intra-rater กับ Inter-rater reliability
6. ระบุปัจจัยที่มีผลต่อความเที่ยงอย่างน้อย 3 ข้อ และแนวทางเพิ่มความเที่ยง

## 8.10 เครื่องมือวัดและประเมินผลประจำบท

### ก) เฉลยใบงาน 8.1 (ขั้นตอนการคำนวณ KR-20)

1) หา  $p$ ,  $q$  ของแต่ละข้อ  $\rightarrow$  2)  $\Sigma pq \rightarrow$  3) หา  $S^2$  ของคะแนนรวม  $\rightarrow$  4) แทนในสูตร  $KR-20 = [k/(k-1)] \cdot [1 - \Sigma pq / S^2] \rightarrow$  5) แปลผลตามเกณฑ์ (เช่น  $k=5$ ,  $\Sigma pq=1.14$ ,  $S^2=2.71 \rightarrow KR-20 \approx 0.72 =$  ดี)

### ข) rubric ประเมินการเลือกวิธีหาความเที่ยง ใบงาน 8.2 (Analytic) — เชื่อมโยง CLO2

| ประเด็น                         | ดีมาก (4)                  | ดี (3)          | พอใช้ (2)      | ปรับปรุง (1) |
|---------------------------------|----------------------------|-----------------|----------------|--------------|
| เลือกวิธีเหมาะกับชนิดเครื่องมือ | เหมาะสมทุกกรณี พร้อมเหตุผล | เหมาะสมส่วนใหญ่ | เหมาะสมบางส่วน | ไม่เหมาะสม   |
| ความถูกต้องของเหตุผล/หลักการ    | ถูกต้องครบถ้วน             | ถูกส่วนใหญ่     | มีข้อผิดพลาด   | ไม่ถูกต้อง   |
| การระบุเกณฑ์แปลผล               | ระบุชัดและถูกต้อง          | ระบุได้         | ไม่ชัดเจน      | ไม่ระบุ      |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 8$  จาก 12 (ร้อยละ 70 ขึ้นไป)

## 8.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

บุญธรรม กิจปรีดาบริสุทธิ์. (2549). การวัดและประเมินผลการเรียนการสอน. สามเจริญพานิช.

พวงรัตน์ ทวีรัตน์. (2540). วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์.

มหาวิทยาลัยศรีนครินทรวิโรฒ.

พัชรินทร์ ชมภูวิเศษ. (2559). เอกสารประกอบการสอนการวัดและประเมินผลการศึกษา.

คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จัญญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้.

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ. (2543). เทคนิคการวัดผลการเรียนรู้. สุวีริยาสาส์น.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6).

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมนึก ภัททิยธนี. (2551). การวัดผลการศึกษา. ประสานการพิมพ์.

ไชยวัฒน์ อารีโรจน์. (2568).

ผลของการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนและการศึกษาความคงทนในการเรียนรู้ เรื่อง

ลักษณะภาษา ด้วยแนวทางภาษาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5. วารสารการวัดผลการศึกษา  
สำนักทดสอบทางการศึกษาและจิตวิทยา, 42(111), 13–40.

## ภาษาอังกฤษ

American Educational Research Association, American Psychological Association,  
& National Council on Measurement in Education. (2014). Standards for educational  
and psychological testing. AERA.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.  
Psychometrika, 16(3), 297–334.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching  
(6th ed.). Macmillan.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test  
reliability. Psychometrika, 2(3), 151–160.

Nitko, A. J., & Brookhart, S. M. (2019). Educational assessment of students (8th  
ed.). Pearson.

Stevens, S. S. (1946). On the theory of scales of measurement. Science,  
103(2684), 677–680.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and  
reporting research instruments in science education. Research in Science Education,  
48(6), 1273–1296.

## บทที่ 9

### การวิเคราะห์ข้อสอบรายข้อ: ความยาก-อำนาจจำแนก-ตัวลวง

#### *Item Analysis: Difficulty, Discrimination and Distractors*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 10 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

หลังตรวจสอบความตรง (บทที่ 7) และความเที่ยง (บทที่ 8) ของแบบทดสอบทั้งฉบับแล้ว บทที่ 9 มุ่งวิเคราะห์คุณภาพ ข้อสอบ “รายข้อ” เพื่อคัดเลือก ปรับปรุง หรือตัดข้อที่ด้อยคุณภาพ ผู้เรียนจะเข้าใจและคำนวณค่าความยาก ( $p$ ) ค่าอำนาจจำแนก ( $r/D$ ) และวิเคราะห์ประสิทธิภาพของตัวลวง (Distractor) ตามเกณฑ์มาตรฐาน ซึ่งเป็นทักษะสำคัญ ในการพัฒนาคลังข้อสอบ (Item Bank) ที่มีคุณภาพ

#### แผนบริหารการสอนประจำบทที่ 9

| รายการ                | รายละเอียด                                                                                                                                                                                                 |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) ความหมายและความสำคัญของการวิเคราะห์ข้อสอบรายข้อ 2) ความยากของข้อสอบ ( $p$ ) 3) อำนาจจำแนกของข้อสอบ ( $r/D$ ) 4) การวิเคราะห์ตัวลวง (Distractor) 5) ตัวอย่างการวิเคราะห์และการคัดเลือก/ปรับปรุงข้อสอบ    |
| ความคิดรวบยอด         | ข้อสอบที่มีคุณภาพต้องมีความยากพอเหมาะ (ไม่ง่าย/ยากเกินไป) มีอำนาจจำแนกกลุ่มเก่ง-อ่อนได้ และมีตัวลวงที่ทำงาน (ดึงกลุ่มอ่อนมากกว่ากลุ่มเก่ง) การวิเคราะห์รายข้อช่วยคัดเลือก ปรับปรุง หรือตัดข้อที่ด้อยคุณภาพ |
| จุดประสงค์การเรียนรู้ | คำนวณและแปลผลค่าความยาก อำนาจจำแนก และวิเคราะห์ตัวลวงของข้อสอบจากข้อมูลจริงได้ (CLO3)                                                                                                                      |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 9.1–9.5 ของบท)                                                                                                                                                  |
| กิจกรรมการเรียนการสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + สาธิตการคำนวณ 3) ใบงาน 9.1 วิเคราะห์ความยาก-อำนาจจำแนกจากข้อมูลจริง 4) ใบงาน 9.2                                                                        |

|                    |                                                                                                                                     |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------|
|                    | วิเคราะห์ตัวलगและตัดสินคุณภาพข้อสอบ 5) อภิปรายการปรับปรุงข้อสอบ 6) สรุป +<br>แบบทดสอบหลังเรียน + แบบฝึกหัด                          |
| สื่อการเรียนการสอน | เอกสารประกอบการสอนบทที่ 9 • สไลด์นำเสนอ • ตัวอย่างตารางวิเคราะห์ข้อสอบ •<br>เครื่องคิดเลข/สเปรดชีต • ใบงาน 9.1–9.2 • ตำราที่อ้างอิง |
| การวัดและประเมินผล | แบบทดสอบก่อน–หลังเรียน • ใบงาน 9.1–9.2 (เฉลยการคำนวณ p, r, ตัวलग) • การอภิปราย<br>• แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70)           |

## 9.1 ความคิดรวบยอด

การวิเคราะห์ข้อสอบรายข้อ (Item Analysis) เป็นการตรวจสอบคุณภาพของข้อสอบแต่ละข้อหลังนำไปทดลองสอบ เพื่อคัดเลือก ข้อที่ดี ปรับปรุงข้อที่พอใช้ และตัดข้อที่ด้อยคุณภาพ ดัชนีหลักที่ใช้มี 2 ค่า คือ ค่าความยาก (Difficulty index: p) ซึ่งเป็นสัดส่วนผู้ตอบถูก ค่ายิ่งสูงข้อยิ่งง่าย ข้อที่ดีควรมีความยากพอเหมาะ (p ระหว่าง 0.20–0.80 อุดมคติประมาณ 0.50) และค่าอำนาจจำแนก (Discrimination index: r หรือ D) ซึ่งบอกความสามารถในการจำแนกกลุ่มเก่งกับกลุ่มอ่อน ข้อที่ดี ควรมี r ตั้งแต่ 0.20 ขึ้นไป นอกจากนี้ในข้อสอบเลือกตอบยังต้องวิเคราะห์ตัวलग (Distractor) ว่าทำงานหรือไม่ โดยตัวलग ที่ดีต้องมีผู้เลือกบ้างและกลุ่มอ่อนเลือกมากกว่ากลุ่มเก่ง (พิชิต ฤทธิจรูญ, 2564; ศิริชัย กาญจนวาสี, 2552; Ebel & Frisbie, 1991)

## 9.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 9 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                 | เชื่อมโยง CLO               |
|-------------------------------------------------------------|-----------------------------|
| 1) อธิบายความหมายและความสำคัญของการวิเคราะห์ข้อสอบรายข้อได้ | CLO3 (เข้าใจ)               |
| 2) คำนวณและแปลผลค่าความยาก (p) ของข้อสอบได้                 | CLO3 (ประยุกต์)             |
| 3) คำนวณและแปลผลค่าอำนาจจำแนก (r/D) ของข้อสอบได้            | CLO3 (ประยุกต์)             |
| 4) วิเคราะห์ประสิทธิภาพของตัวलगในข้อสอบเลือกตอบได้          | CLO3 (วิเคราะห์)            |
| 5) ตัดสินคัดเลือก/ปรับปรุง/ตัดข้อสอบตามเกณฑ์คุณภาพได้       | CLO3 (วิเคราะห์–ประเมินค่า) |

### 9.3 คำสำคัญ

การวิเคราะห์ข้อสอบรายข้อ (Item Analysis) • ค่าความยาก (Difficulty index: p) • ค่าอำนาจจำแนก (Discrimination index: r/D) • ตัวลวง (Distractor) • กลุ่มสูง-กลุ่มต่ำ (High-Low groups) • สัมประสิทธิ์สหสัมพันธ์รายข้อกับคะแนนรวม (Item-total correlation) • คลังข้อสอบ (Item Bank)

### 9.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 9.5 เนื้อหาวิชา

---

#### 9.5.1 ความหมายและความสำคัญของการวิเคราะห์ข้อสอบรายข้อ

การวิเคราะห์ข้อสอบรายข้อ (item analysis) คือกระบวนการตรวจสอบคุณภาพของข้อสอบ “ทีละข้อ” โดยนำกระดาษคำตอบจากการนำข้อสอบไปทดลองใช้ (try-out) มาคำนวณค่าสถิติประจำข้อ ได้แก่ ค่าความยาก (difficulty index: p) ค่าอำนาจจำแนก (discrimination index: r หรือ D) และการกระจายการเลือกตัวลวง (distractor) เพื่อบอกว่าข้อสอบแต่ละข้อ “ทำงาน” ตามที่ตั้งใจไว้หรือไม่ (สุทธิพงศ์ บุญผดุง, 2568) ขณะที่การตรวจสอบความตรง (บทที่ 7) และความเที่ยง (บทที่ 8) มองคุณภาพของแบบทดสอบ “ทั้งฉบับ” การวิเคราะห์รายข้อจะเจาะลึกลงไป “อีฐแต่ละก้อน” ที่ประกอบกันเป็นแบบทดสอบ ทำให้ครูทราบว่าจะควรเก็บ ปรับปรุง หรือตัดข้อใดทิ้ง

การวิเคราะห์รายข้อมีรากฐานจากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ซึ่งอธิบายว่าคะแนนที่สอบได้จริง (observed score: X) ประกอบด้วยคะแนนแท้ (true score: T) กับความคลาดเคลื่อน (error: E) ตามสมการ  $X = T + E$  (ศิริชัย กาญจนวาสี, 2552) ข้อสอบที่ด้อยคุณภาพ เช่น ใช้ภาษากำกวมหรือมีตัวเลือกที่ถูกได้มากกว่าหนึ่งตัวเลือก จะเพิ่มค่าความคลาดเคลื่อน (E) และบิดเบือนคะแนน ทำให้คะแนนไม่สะท้อนความสามารถที่แท้จริงของผู้เรียน หน้าที่หลักของการวิเคราะห์รายข้อจึงเป็นการ “ค้นหาและ

ลด” ข้อที่สร้างความคลาดเคลื่อนสูง เพื่อให้คะแนนรวมเที่ยงตรงและยุติธรรมที่สุด แนวคิดนี้สอดคล้องกับเป้าหมายของการวัดผลสมัยใหม่ที่เปลี่ยนจากการถามว่าผู้เรียนได้คะแนนเท่าใด (how much) ไปสู่การถามว่าเครื่องมือวัดมีคุณภาพดีเพียงใด (how good) (Gronlund & Linn, 1990)

ในทางปฏิบัติแบ่งการวิเคราะห์รายข้อได้ 2 แบบตามจุดมุ่งหมายของการสอบ ได้แก่ การวิเคราะห์แบบอิงกลุ่ม (norm-referenced) ที่พิจารณาทั้งค่าความยากและค่าอำนาจจำแนกเพื่อจำแนกผู้เรียนกลุ่มเก่ง-อ่อน และการวิเคราะห์แบบอิงเกณฑ์ (criterion-referenced) ที่เน้นค่าอำนาจจำแนกระหว่างการสอบ “ก่อนเรียน-หลังเรียน” เป็นหลัก โดยไม่ถือว่าข้อที่ง่ายเป็นข้อเสีย เพราะมุ่งดูว่าผู้เรียนบรรลุจุดประสงค์หรือไม่ (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ขั้นตอนมาตรฐานของการวิเคราะห์แบบอิงกลุ่มสรุปได้ดังตารางที่ 9.1 ตารางที่ 9.1 ขั้นตอนการวิเคราะห์ข้อสอบรายข้อแบบอิงกลุ่ม

| ขั้นที่ | รายละเอียดของขั้นตอน                                                                           |
|---------|------------------------------------------------------------------------------------------------|
| 1       | นำข้อสอบที่สร้างเสร็จไปทดลองสอบ (try-out) กับผู้เรียนกลุ่มที่มีลักษณะใกล้เคียงกับกลุ่มเป้าหมาย |
| 2       | ตรวจให้คะแนนรายข้อ (ถูก = 1, ผิด = 0) และรวมคะแนนของผู้สอบแต่ละคน                              |
| 3       | เรียงกระดาษคำตอบจากคะแนนสูงสุดไปต่ำสุด                                                         |
| 4       | แบ่งกลุ่มสูงและกลุ่มต่ำ (นิยมใช้ 27% บน-ล่างเมื่อกลุ่มใหญ่ หรือ 50% เมื่อกลุ่มเล็ก)            |
| 5       | นับความถี่การเลือกแต่ละตัวเลือก (ก ข ค ง) ของกลุ่มสูงและกลุ่มต่ำลงในแบบฟอร์ม                   |
| 6       | คำนวณค่าความยาก (p) และอำนาจจำแนก (r) แล้วแปลผลและตัดสินคุณภาพรายข้อ                           |

ที่มา: ดัดแปลงจาก พัทธรินทร์ ชมภูวิเศษ (2559, น. 195)

ตัวอย่างในชั้นเรียน: ครูคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 2 ออกข้อสอบปรนัย 30 ข้อ นำไปทดลองสอบกับนักเรียน 40 คน เมื่อตรวจและเรียงคะแนนแล้ว ครูเลือกกลุ่มสูง 11 คน (ราว 27% บน) และกลุ่มต่ำ 11 คน (ราว 27% ล่าง) เพื่อนำเฉพาะสองกลุ่มนี้มาวิเคราะห์รายข้อ ครูพบว่าข้อสอบ 6 ข้อที่กลุ่มอ่อนตอบถูกพอ ๆ กับกลุ่มเก่ง จึงนำ 6 ข้อนั้นไปทบทวนภาษาและตัวलगก่อนนำกลับมาใช้ใหม่ ส่วนข้อที่มีค่าสถิติดีจะถูกเก็บเข้าคลังข้อสอบของห้องเรียน

การประยุกต์ใช้และข้อควรระวัง: การวิเคราะห์รายข้อไม่ได้มีไว้เพียง “ตัดสิน” ข้อสอบ แต่เป็นเครื่องมือพัฒนาการสอนด้วย หากผู้เรียนส่วนใหญ่ตอบผิดในจุดประสงค์ที่สำคัญ ครูควรย้อนกลับไปทบทวนวิธีสอนและสื่อ ไม่ใช่สรุปว่าเป็นความบกพร่องของผู้เรียนเพียงอย่างเดียว (พิชิต ฤทธิ์จรูญ, 2564) ข้อควรระวังสำคัญคือ

ค่าสถิติรายข้อขึ้นอยู่กับกลุ่มผู้สอบและจำนวนผู้สอบ ถ้าทดลองกับคนน้อยเกินไป (ต่ำกว่าประมาณ 30 คน) ค่าที่ได้จะไม่เสถียร ครูจึงควรตีความอย่างระมัดระวังและใช้ร่วมกับการพิจารณาความเหมาะสมเชิงเนื้อหาเสมอ

### 9.5.2 ความยากของข้อสอบ (Difficulty Index: p)

ค่าความยาก (difficulty index) แทนด้วยสัญลักษณ์  $p$  คือสัดส่วนของผู้ที่ตอบข้อนั้นถูก ต่อจำนวนผู้เข้าสอบทั้งหมด มีค่าตั้งแต่ 0.00 ถึง 1.00 ค่า  $p$  ยิ่งมากยิ่งแสดงว่าข้อนั้นง่าย (เพราะมีคนตอบถูกมาก) ค่า  $p$  ยิ่งน้อยยิ่งแสดงว่าข้อนั้นยาก (สุทธิพงศ์ บุญผดุง, 2568) คำว่า “ความยาก” จึงสวนทางกับค่าตัวเลขเล็กน้อย นักศึกษาควรควรเข้าใจว่า  $p$  แท้จริงแล้วเป็น “ดัชนีความง่าย” จะตีความได้ถูกต้องกว่า

#### สูตรคำนวณค่าความยาก ( $p$ )

กรณีใช้ผู้สอบทั้งหมด:  $p = R / N$

$R$  = จำนวนผู้ตอบถูก,  $N$  = จำนวนผู้เข้าสอบทั้งหมด

กรณีแบ่งกลุ่มสูง-ต่ำ:  $p = (RU + RL) / (NU + NL)$

$RU, RL$  = ผู้ตอบถูกในกลุ่มสูง/กลุ่มต่ำ;  $NU, NL$  = จำนวนคนในแต่ละกลุ่ม

เกณฑ์ที่ใช้ได้: 0.20–0.80 • เหมาะสมที่สุดราว 0.50

ที่มา: สุทธิพงศ์ บุญผดุง (2568); พชรินทร์ ชมภูวิเศษ (2559, น. 196)

ข้อสอบที่ดีตามทฤษฎีอิงกลุ่มควรมีค่าความยากอยู่ระหว่าง 0.20–0.80 และเหมาะสมที่สุดเมื่ออยู่ราว 0.50 เพราะเป็นช่วงที่ข้อสอบให้สารสนเทศ (information) เกี่ยวกับความสามารถของผู้เรียนได้มากที่สุด และเปิดโอกาสให้คำอำนาจจำแนกสูงได้ (ศิริชัย กาญจนวาสี, 2552; บุญชม ศรีสะอาด, 2560) การแปลความหมายของค่า  $p$  อย่างละเอียดเป็นช่วง ๆ แสดงไว้ในตารางที่ 9.2

ตารางที่ 9.2 เกณฑ์การแปลความหมายค่าความยาก ( $p$ )

| ช่วงค่า $p$ | ความหมาย       | แนวทางการตัดสินใจ                            |
|-------------|----------------|----------------------------------------------|
| 0.81–1.00   | ง่ายมาก        | ปรับปรุงให้ยากขึ้น หรือใช้เป็นข้ออุ่นเครื่อง |
| 0.61–0.80   | ค่อนข้างง่าย   | ใช้ได้ (เหมาะกับการวัดพื้นฐาน/วินิจฉัย)      |
| 0.40–0.60   | ยากง่ายพอเหมาะ | ดีมาก ให้สารสนเทศสูงสุด ควรเก็บไว้           |
| 0.20–0.39   | ค่อนข้างยาก    | ใช้ได้ (เหมาะกับการสอบคัดเลือก)              |
| 0.00–0.19   | ยากมาก         | ตรวจสอบ/ปรับปรุง หรือตัดทิ้ง                 |

ที่มา: ดัดแปลงจาก สุทธิพงศ์ บุญผดุง (2568) และ พชรินทร์ ชมภูวิเศษ (2559, น. 196)

ตัวอย่างในชั้นเรียน: ข้อสอบวิทยาศาสตร์ชั้นประถมศึกษาปีที่ 6 ข้อหนึ่ง มีนักเรียนเข้าสอบ 40 คน ตอบถูก 30 คน คำนวณได้  $p = 30/40 = 0.75$  แปลว่าเป็นข้อค่อนข้างง่ายซึ่งยังอยู่ในเกณฑ์ใช้ได้ แต่ถ้าครูต้องการข้อที่ทำหายและจำแนกกลุ่มเก่งได้ดีขึ้น อาจปรับถ้อยคำหรือตัวลงให้ค่า  $p$  เข้าใกล้ 0.50–0.60 ในทางกลับกัน หากมีอีกข้อหนึ่งที่นักเรียนตอบถูกเพียง 6 คน จะได้  $p = 6/40 = 0.15$  ซึ่งยากมาก ครูควรตรวจสอบว่าเป็นเพราะเนื้อหายังไม่ได้สอน ภาษากำกวม หรือเฉลยผิด ก่อนจะสรุปว่าเป็นข้อที่ยากเกินไป

การประยุกต์ใช้และข้อควรระวัง: (1) ค่า  $p$  เป็นค่าที่ขึ้นกับกลุ่มผู้สอบ (sample-dependent) ข้อเดียวกันอาจง่ายสำหรับห้องเก่งแต่ยากสำหรับห้องอ่อน จึงนำค่ามาเทียบข้ามกลุ่มกันตรง ๆ ไม่ได้ (2) ค่า  $p$  บอกเพียง “จำนวนคนที่ตอบถูก” แต่ไม่บอกว่า “ใครตอบถูก” ข้อที่  $p$  พอเหมาะอาจยังจำแนกผู้เรียนได้ไม่ดีหากการตอบถูกเกิดจากการเดา จึงต้องพิจารณาค่าอำนาจจำแนกประกอบเสมอ (3) สำหรับข้อสอบอิงเกณฑ์ที่มุ่งให้ผู้เรียนผ่านจุดประสงค์ ค่า  $p$  ที่สูงหลังเรียนถือเป็นเรื่องดี ไม่ใช่ข้อบกพร่อง (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566)

### 9.5.3 อำนาจจำแนกของข้อสอบ (Discrimination Index: $r/D$ )

ค่าอำนาจจำแนก (discrimination index) แทนด้วยสัญลักษณ์  $r$  หรือ  $D$  คือดัชนีที่บอกความสามารถของข้อสอบในการ “แยก” ผู้สอบกลุ่มเก่ง (กลุ่มสูง) ออกจากกลุ่มอ่อน (กลุ่มต่ำ) ข้อที่ดีนั้นกลุ่มเก่งควรตอบถูกมากกว่ากลุ่มอ่อน ค่า  $r$  มีพิสัยตั้งแต่  $-1.00$  ถึง  $+1.00$  ค่าบวกสูงแปลว่าจำแนกได้ดี ค่าใกล้ศูนย์แปลว่าจำแนกไม่ได้ และค่าติดลบแปลว่าผิดปกติ คือกลุ่มอ่อนกลับตอบถูกมากกว่ากลุ่มเก่ง (สุทธิพงศ์ บุญผดุง, 2568)

#### สูตรคำนวณค่าอำนาจจำแนก ( $r/D$ )

วิธีกลุ่มสูง-ต่ำ:  $r = (RU - RL) / n$

$RU, RL$  = ผู้ตอบถูกในกลุ่มสูง/กลุ่มต่ำ,  $n$  = จำนวนคนในแต่ละกลุ่ม

หรือเขียนเป็น:  $D = pU - pL$  (ผลต่างค่าความยากของสองกลุ่ม)

อีกวิธี: สหสัมพันธ์รายข้อกับคะแนนรวม (point-biserial correlation)

ที่มา: ล้วน สายยศ และอังคณา สายยศ (2543); สุทธิพงศ์ บุญผดุง (2568)

เกณฑ์การแปลผลที่นิยมใช้คือเกณฑ์ของเอเบลและฟรีสบี (Ebel & Frisbie, 1991) ดังตารางที่ 9.3 โดยถือว่าค่า  $r$  ตั้งแต่ 0.20 ขึ้นไปจึงจะใช้ได้ ค่าที่ต่ำกว่านี้ควรปรับปรุงหรือตัด และค่าติดลบต้องตรวจสอบแก้ไขเสมอ

ตารางที่ 9.3 เกณฑ์การแปลความหมายค่าอำนาจจำแนก ( $r$ )

| ช่วงค่า r    | ความหมาย      | แนวทางการตัดสินใจ              |
|--------------|---------------|--------------------------------|
| 0.40 ขึ้นไป  | จำแนกได้ดีมาก | คัดเก็บเข้าคลังข้อสอบมาตรฐาน   |
| 0.30–0.39    | จำแนกได้ดี    | คงไว้ใช้งานได้                 |
| 0.20–0.29    | จำแนกพอใช้    | ควรปรับปรุงตัวलगง/ถ้อยคำ       |
| ต่ำกว่า 0.20 | จำแนกไม่ได้   | ปรับปรุงอย่างมากหรือตัดทิ้ง    |
| ติดลบ        | ผิดปกติ       | ตรวจเฉลย/ภาษา แล้วแก้ไขก่อนใช้ |

ที่มา: ดัดแปลงจาก Ebel & Frisbie (1991) และ สุทธิพงศ์ บุญผดุง (2568)

การแบ่งกลุ่มสูง-ต่ำนิยมใช้สัดส่วน 27% บนและล่างเมื่อกลุ่มผู้สอบมีจำนวนมาก ซึ่งเคลลี (Kelley, 1939) พิสูจน์ทางสถิติว่าเป็นสัดส่วนที่ให้ความสมดุลระหว่าง “ความต่าง” ของสองกลุ่มกับ “ความคงที่” ของค่าที่คำนวณได้ดีที่สุด แต่ถ้ากลุ่มเล็ก เช่น ทั้งห้องมีเพียง 40–50 คน มักใช้ 50% บน-ล่างเพื่อให้มีจำนวนคนพอวิเคราะห์ (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ข้อสังเกตสำคัญคือความสัมพันธ์ระหว่างค่า p กับ r กล่าวคือ ข้อที่ง่ายมาก (p เข้าใกล้ 1) หรือยากมาก (p เข้าใกล้ 0) แทบทุกคนจะตอบเหมือนกัน ทำให้ค่า r ต่ำลงโดยปริยาย ค่าอำนาจจำแนกจะมีโอกาสขึ้นสูงสุดเมื่อค่าความยากอยู่ราว 0.50

ตัวอย่างในชั้นเรียน: กำหนดกลุ่มสูง 10 คน และกลุ่มต่ำ 10 คน ในข้อหนึ่ง หากกลุ่มสูงตอบถูก 8 คน และกลุ่มต่ำตอบถูก 3 คน จะได้  $r = (8 - 3)/10 = 0.50$  ถือว่าจำแนกได้ดีมาก แต่ถ้ากลุ่มสูงถูก 5 คน กลุ่มต่ำถูก 5 คน จะได้  $r = 0$  คือจำแนกไม่ได้เลย และหากกลุ่มสูงถูก 3 คน แต่กลุ่มต่ำถูก 7 คน จะได้  $r = (3 - 7)/10 = -0.40$  ซึ่งเป็นสัญญาณว่าข้อสอบผิดปกติอย่างรุนแรง ควรตรวจสอบทันที

การประยุกต์ใช้และข้อควรระวัง: ค่า r ติดลบเป็นสัญญาณอันตรายที่พบบ่อย มักเกิดจากการเฉลยผิด (mis-keyed) ภาษาที่กำกวมจนทำให้กลุ่มเก่งคิดมากและตอบพลาด หรือมีตัวเลือกที่ถูกได้มากกว่าหนึ่งตัว ครูต้องตรวจสอบและแก้ไขก่อนนำข้อสอบไปใช้อีกครั้ง (สมนึก ภัททิยธนี, 2558) นอกจากนี้ค่า r ที่คำนวณจากกลุ่มเล็กจะกว้างได้มาก จึงควรใช้ผลการวิเคราะห์ประกอบกับการพิจารณาความสำคัญเชิงเนื้อหา ไม่ตัดสินจากตัวเลขเพียงอย่างเดียว

#### 9.5.4 การวิเคราะห์ตัวलगง (Distractor Analysis)

ตัวलगง (distractor หรือ foil) คือตัวเลือกที่ไม่ใช่คำตอบที่ถูกต้องในข้อสอบแบบเลือกตอบ ตัวलगงที่ดีต้อง “ดูน่าเชื่อ” มากพอจนดึงดูดผู้ที่ยังไม่รู้จริงให้เลือกตอบ การวิเคราะห์ตัวलगง (distractor analysis) คือการพิจารณาความถี่ที่กลุ่มสูงและกลุ่มต่ำเลือกแต่ละตัวलगง เพื่อประเมินว่าตัวलगงนั้น “ทำงาน” หรือไม่ หากตัวलगงทำงานดีจะช่วยให้อ่านข้อสอบมีความยากพอเหมาะและจำแนกผู้เรียนได้ดีขึ้น (Rezigalla et al., 2024)

หลักการวิเคราะห์ตัวลวงมีสองข้อสำคัญ ได้แก่ กฎ 5% (5% rule) คือตัวลวงที่มีประสิทธิภาพควรมีผู้เลือกอย่างน้อยประมาณร้อยละ 5 ของผู้เข้าสอบ และหลักทิศทางการกลับ (inverse direction) คือกลุ่มต่ำควรเลือกตัวลวงมากกว่ากลุ่มสูง (ค่าอำนาจจำแนกของตัวลวงคำนวณจาก  $pL - pH$  ควรเป็นบวก) (สุทธิพงศ์ บุญผดุง, 2568) ตัวอย่างตารางวิเคราะห์ตัวลวงรายตัวเลือกแสดงไว้ในตารางที่ 9.4

ตารางที่ 9.4 ตัวอย่างตารางวิเคราะห์ตัวลวงรายตัวเลือก (ตัวถูก = ข)

| ตัวเลือก  | กลุ่มสูง (10) | กลุ่มต่ำ (10) | การแปลผล                |
|-----------|---------------|---------------|-------------------------|
| ก (ลวง)   | 1             | 3             | ตัวลวงทำงาน (ต่ำ > สูง) |
| ข (ถูก) * | 8             | 3             | ตัวถูกดี (สูง > ต่ำ)    |
| ค (ลวง)   | 1             | 4             | ตัวลวงทำงานดี           |
| ง (ลวง)   | 0             | 0             | ตัวลวงตาย ควรปรับปรุง   |

ที่มา: ดัดแปลงจาก สุทธิพงศ์ บุญผดุง (2568) และ Rezigalla et al. (2024)

ตัวลวงที่ไม่มีใครเลือกเลย เรียกว่า “ตัวลวงตาย” (dead หรือ non-functioning distractor) เช่น ตัวเลือก ง ในตารางที่ 9.4 การมีตัวลวงตายทำให้ข้อสอบง่ายขึ้นโดยปริยาย เพราะเหลือตัวเลือกที่แข่งขันกันจริงน้อยลง และมีผลให้ค่าอำนาจจำแนกลดลงตามไปด้วย งานวิจัยของเรซิกัลลาและคณะ (Rezigalla et al., 2024) ซึ่งวิเคราะห์ข้อสอบเลือกตอบจำนวนมาก พบว่าข้อที่มีตัวลวงทำงานครบทุกตัวมีแนวโน้มจำแนกได้ดีกว่าและมีความยากพอเหมาะมากกว่าข้อที่มีตัวลวงตาย จึงยืนยันได้ว่าคุณภาพของตัวลวงส่งผลต่อทั้งค่า  $p$  และค่า  $r$  ของข้อสอบ

ตัวอย่างเชิงวินิจฉัยในชั้นเรียน: ในข้อสอบวิชาภาษาไทยชั้นประถมศึกษาปีที่ 5 หากพบว่าตัวลวง ค ถูกเลือกโดยกลุ่มเก่งจำนวนมากผิดปกติ แสดงว่าตัวลวงนั้นอาจ “ถูกบางส่วน” หรือคำถามกำกวมจนทำให้ผู้เรียนเก่งลังเล ครูสามารถใช้ข้อมูลการเลือกตัวลวงนี้เพื่อวินิจฉัย “มโนทัศน์ที่คลาดเคลื่อน” (misconception) ของผู้เรียน แล้วนำไปออกแบบการสอนซ่อมเสริมได้ตรงจุด ซึ่งเป็นการใช้ผลการวิเคราะห์เพื่อพัฒนาการเรียนรู้ ไม่ใช่เพียงเพื่อให้คะแนน (Nitko & Brookhart, 2019)

การประยุกต์ใช้และข้อควรระวัง: ไม่ควรตัดสินคุณภาพตัวลวงจากกลุ่มที่เล็กเกินไปเพราะค่าจะไม่เสถียร เมื่อพบตัวลวงตายควร “ปรับให้สมเหตุสมผลขึ้น” เช่น เปลี่ยนเป็นข้อผิดที่ผู้เรียนพลาดบ่อยจริง ไม่ใช่เพิ่มจำนวน

ตัวเลือกอย่างพร่ำเพรื่อ และควรเขียนตัวลงจากความเข้าใจผิดที่พบได้จริงของผู้เรียน เพื่อให้ตัวลงทำงานและให้ข้อมูลเชิงวินิจัยที่มีประโยชน์ (สมนึก ภัททิยธนี, 2558)

### 9.5.5 ตัวอย่างการวิเคราะห์และการคัดเลือก/ปรับปรุงข้อสอบ

เมื่อเข้าใจค่าทั้งสามแล้ว ขั้นตอนไปคือการนำค่าความยาก ค่าอำนาจจำแนก และผลการวิเคราะห์ตัวลงมาพิจารณาพร้อมกันเพื่อ “ตัดสิน” คุณภาพของข้อสอบแต่ละข้อ ดังตัวอย่างต่อไปนี้ กำหนดข้อสอบเลือกตอบ 4 ตัวเลือก (ตัวถูก = ข) ทดลองกับกลุ่มสูง 10 คนและกลุ่มต่ำ 10 คน ได้ผลการเลือกดังตารางที่ 9.5

| ตัวเลือก  | กลุ่มสูง (10) | กลุ่มต่ำ (10) | การแปลผล               |
|-----------|---------------|---------------|------------------------|
| ก (ลง)    | 1             | 3             | ตัวลงทำงาน (ต่ำ > สูง) |
| ข (ถูก) * | 8             | 3             | ตัวถูก: สูง > ต่ำ ดี   |
| ค (ลง)    | 1             | 2             | ตัวลงทำงาน             |
| ง (ลง)    | 0             | 2             | ตัวลงพอใช้ (ควรติดตาม) |

ตารางที่ 9.5 ตัวอย่างการวิเคราะห์ข้อสอบหนึ่งข้อแบบครบวงจร

#### การคำนวณข้อตัวอย่าง (ตัวถูก = ข)

ค่าความยาก  $p = (RU + RL)/(NU + NL) = (8 + 3)/(10 + 10) = 11/20 = 0.55 \rightarrow$  เหมาะสม

ค่าอำนาจจำแนก  $r = (RU - RL)/n = (8 - 3)/10 = 5/10 = 0.50 \rightarrow$  ดีมาก

สรุป: ข้อนี้  $p = 0.55, r = 0.50$  ตัวลงทำงานเกือบครบ  $\rightarrow$  คัดเลือกไว้ใช้ได้

ที่มา: ผู้เขียน

หลักการคัดเลือกคือ เก็บข้อที่ค่าความยากอยู่ในช่วง 0.20–0.80 และค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป ปรับปรุงข้อที่ค่ากำกึ่งหรือมีตัวลงตาย และตัดข้อที่อำนาจจำแนกต่ำกว่า 0.20 หรือติดลบ (Ebel & Frisbie, 1991) เกณฑ์การตัดสินใจแบบบูรณาการสรุปไว้ในตารางที่ 9.6

ตารางที่ 9.6 เกณฑ์การตัดสินใจคัดเลือก/ปรับปรุง/ตัดข้อสอบ

| สถานะข้อสอบ | ค่า p          | ค่า r                   | การดำเนินการ                 |
|-------------|----------------|-------------------------|------------------------------|
| คุณภาพดี    | 0.20–0.80      | 0.20 ขึ้นไป             | เก็บเข้าคลังข้อสอบ           |
| ควรปรับปรุง | นอกช่วง/กำกึ่ง | 0.20 ขึ้นไป             | ปรับภาษา/ตัวลง แล้วทดลองใหม่ |
| ตัดทิ้ง     | ทุกค่า         | ต่ำกว่า 0.20 หรือ ติดลบ | ตัดออก หรือเขียนข้อใหม่      |

ที่มา: ดัดแปลงจาก Ebel & Frisbie (1991) และ พิชิต ฤทธิ์จรูญ (2564)

ในทางปฏิบัติ ครูจะสรุปผลการวิเคราะห์ของทั้งฉบับลงในตารางเดียวเพื่อมองภาพรวมและจัดการคลังข้อสอบ ดังตัวอย่างผลการวิเคราะห์ข้อสอบ 8 ข้อในตารางที่ 9.7 ซึ่งช่วยให้เห็นชัดว่าควรเก็บ ปรับปรุง หรือตัดข้อใด

ตารางที่ 9.7 ตัวอย่างสรุปผลการวิเคราะห์ข้อสอบ 8 ข้อและการตัดสินใจ

| ข้อ | p    | r     | ข้อสังเกตตัวलग       | การตัดสินใจ          |
|-----|------|-------|----------------------|----------------------|
| 1   | 0.55 | 0.50  | ตัวलगทำงานครบ        | เก็บเข้าคลัง         |
| 2   | 0.12 | 0.05  | ยากมาก จำแนกไม่ได้   | ตัดทิ้ง              |
| 3   | 0.92 | 0.08  | ง่ายมาก              | ตัด/ปรับปรุงใหญ่     |
| 4   | 0.48 | 0.45  | ตัวलग ค ไม่มีคนเลือก | ปรับตัวलग ค แล้วเก็บ |
| 5   | 0.60 | 0.33  | ดีตามเกณฑ์           | เก็บไว้              |
| 6   | 0.30 | 0.42  | ยากแต่จำแนกดี        | เก็บ (ข้อคัดเลือก)   |
| 7   | 0.50 | -0.20 | กลุ่มอ่อนถูกมากกว่า  | ตัดทิ้ง/ตรวจเฉลย     |
| 8   | 0.70 | 0.28  | ใช้ได้               | เก็บไว้              |

ที่มา: ผู้เขียน

ผลการวิเคราะห์ที่สะสมอย่างต่อเนื่องจะถูกนำไปสร้าง “คลังข้อสอบ” (item bank) ซึ่งเก็บเฉพาะข้อที่มีคุณภาพพร้อมค่าสถิติกำกับไว้ ช่วยให้ครูประกอบข้อสอบฉบับใหม่ได้รวดเร็วและรักษามาตรฐานการวัดผลในระยะยาว (ชนาพร ไชยสุวรรณ, เบญจมาภรณ์ เสนารัตน์, และสมประสงค์ เสนารัตน์, 2568) ทักษะการคำนวณค่า p และ r การวิเคราะห์ตัวलग และการตัดสินใจคัดเลือก/ปรับปรุงข้อสอบจากข้อมูลจริงนี้ คือหัวใจของผลลัพธ์การเรียนรู้ CLO3 ที่ผู้เรียนต้องฝึกปฏิบัติในใบงานที่ 9.1–9.2

การประยุกต์ใช้และข้อควรระวังในการตัดสินใจ: ไม่ควรตัดข้อสอบโดยพิจารณาค่าสถิติเพียงอย่างเดียว ต้องพิจารณาความสำคัญเชิงเนื้อหา (ความตรง) ประกอบด้วย บางข้อสำคัญต่อจุดประสงค์การเรียนรู้แม้ค่า p หรือ r จะยังไม่สวยงาม ก็ควร “ปรับปรุง” มากกว่า “ตัดทิ้ง” และพึงระลึกว่าค่าที่ได้จากกลุ่มเล็กเป็นเพียงแนวโน้ม ไม่ใช่ข้อสรุปเด็ดขาด ครูที่ดีจึงใช้ทั้งหลักฐานเชิงสถิติและวิจารณญาณเชิงวิชาชีพควบคู่กัน (พิชิต ฤทธิ์จรูญ, 2564)

### กรณีศึกษา: การวิเคราะห์คุณภาพข้อสอบระดับชาติ (O-NET)

สังวรณ จิตกระโทก และคณะ (2561) วิเคราะห์คุณภาพข้อสอบโอเน็ต 5 วิชา รวมทั้งภาษาไทยและภาษาอังกฤษ ระหว่างปีการศึกษา 2554–2558 ด้วยทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ พบว่า ข้อสอบส่วนใหญ่มีค่าความยากและอำนาจจำแนกผ่านเกณฑ์ ค่าความเที่ยงส่วนใหญ่อยู่ระหว่าง 0.6–0.8 แต่บางฉบับที่มีจำนวนข้อน้อยมีความเที่ยงต่ำถึง .37 คณะผู้วิจัยจึงเสนอให้เพิ่มจำนวนข้อ จัดทำแบบทดสอบคู่ขนาน ระหว่างปี และตรวจสอบความสอดคล้องของข้อสอบกับมาตรฐานหลักสูตรก่อนใช้จริง

กรณีนี้แสดงว่าค่า  $p$  ค่า  $r$  และความเที่ยงที่เรียนในบทนี้ถูกใช้จริงในการประกันคุณภาพข้อสอบระดับประเทศ และย้ำความสัมพันธ์ระหว่างจำนวนข้อสอบกับความเที่ยงที่กล่าวถึงในบทที่ 8

**คำถามชวนคิด:** หากโรงเรียนของท่านจะจัดทำคลังข้อสอบกลางวิชาภาษาอังกฤษ ท่านจะวางกระบวนการ วิเคราะห์ข้อสอบรายข้อและเกณฑ์คัดเข้าคลังอย่างไร ตั้งแต่การสอบครั้งแรกจนถึงการปรับปรุงข้อสอบรอบถัดไป

## 9.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ใบงานที่ 9.1 วิเคราะห์ความยากและอำนาจจำแนก (รายบุคคล)

**คำชี้แจง:** กำหนดตารางผลตอบข้อสอบของกลุ่มสูงและกลุ่มต่ำ (กลุ่มละ 10 คน) จำนวน 5 ข้อ ให้ผู้เรียนคำนวณ ค่า  $p$  และ  $r$  ของแต่ละข้อ พร้อมแปลผลและตัดสินใจข้อใดใช้ได้/ปรับปรุง/ตัด (มีเฉลยในหัวข้อ 9.10) — เชื่อมโยง CLO3

### ใบงานที่ 9.2 วิเคราะห์ตัวलगและสรุปคุณภาพข้อสอบ (รายกลุ่ม)

**คำชี้แจง:** กำหนดตารางการเลือกตัวเลือก ก–ง ของกลุ่มสูง–ต่ำ ให้กลุ่มวิเคราะห์ว่าตัวलगใดทำงาน/ไม่ทำงาน เสนอแนวทาง ปรับปรุง พร้อมสรุปคุณภาพข้อสอบโดยรวม (ประเมินด้วยรูบริกในหัวข้อ 9.10) — เชื่อมโยง CLO3

**การสะท้อนคิดและการให้ข้อมูลย้อนกลับ:** ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและ กิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน–หลังเรียน เพื่อให้ ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 9.7 สรุปบทเรียน

การวิเคราะห์ข้อสอบรายข้อเป็นขั้นตอนสำคัญในการพัฒนาแบบทดสอบที่มีคุณภาพ โดยนำผลการทดลองสอบ มาคำนวณ ค่าความยาก ( $p$  = สัดส่วนผู้ตอบถูก เกณฑ์ 0.20–0.80 อุดมคติ ~0.50) และค่าอำนาจจำแนก ( $r$  =

ผลต่างสัดส่วนผู้ตอบถูก ของกลุ่มสูงกับกลุ่มต่ำ เกณฑ์  $\geq 0.20$ ) สำหรับข้อสอบเลือกตอบยังต้องวิเคราะห์ตัวलग โดยตัวलगที่ดีต้องมีผู้เลือก และกลุ่มอ่อนเลือกมากกว่ากลุ่มเก่ง ส่วนตัวถูกกลุ่มเก่งต้องเลือกมากกว่ากลุ่มอ่อน ผลการวิเคราะห์ใช้คัดเลือกข้อที่ดี ปรับปรุงข้อกำกวม และตัดข้อที่มีอำนาจจำแนกต่ำหรือติดลบ เพื่อจัดทำคลังข้อสอบที่มีคุณภาพ การวิเคราะห์รายข้อนี้ ทำงานร่วมกับการตรวจสอบความตรงและความเที่ยงทั้งฉบับ เป็นการประกันคุณภาพเครื่องมือก่อนนำไปใช้ตัดสินผลจริง

## 9.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลี่ยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 9.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

1. หลังการสอบท้ายหน่วยแบบอิงเกณฑ์ ครูพบว่าผู้เรียนเกือบทุกคนตอบข้อ 1–5 ถูก จะสรุปทันทีได้หรือไม่ว่าข้อสอบ "ง่ายเกินไปและต้องตัดทิ้ง" จงวิเคราะห์โดยใช้แนวคิดการแปลค่าความยากตามจุดมุ่งหมายอิงกลุ่ม–อิงเกณฑ์ พร้อมเสนอขั้นตอนตรวจสอบก่อนตัดสินใจ
2. อธิบายสูตรและเกณฑ์ของค่าความยาก ( $p$ ) พร้อมยกตัวอย่างการคำนวณ
3. อธิบายสูตรและเกณฑ์ของค่าอำนาจจำแนก ( $r$ ) และความหมายของค่าติดลบ
4. กำหนดข้อมูลกลุ่มสูง–ต่ำ จำนวน  $p$  และ  $r$  ของข้อสอบ 1 ข้อ พร้อมตัดสินคุณภาพ
5. อธิบายหลักการวิเคราะห์ตัวलग และยกตัวอย่างตัวलगที่ทำงาน/ไม่ทำงาน
6. อธิบายหลักการคัดเลือก ปรับปรุง และตัดข้อสอบเพื่อจัดทำคลังข้อสอบ

## 9.10 เครื่องมือวัดและประเมินผลประจำบท

### ก) เฉลยใบงาน 9.1 (ขั้นตอนการคำนวณ p และ r)

$p = (RU+RL)/(NU+NL) \cdot r = (RU-RL)/n$  เช่น  $RU=8, RL=3, n=10 \rightarrow p=(8+3)/20=0.55$  (เหมาะสม);  
 $r=(8-3)/10=0.50$  (ดีมาก)  $\rightarrow$  คัดเลือกไว้ใช้

### ข) รุบริคประเมินการวิเคราะห์ตัวลง ใบงาน 9.2 (Analytic) — เชื่อมโยง CLO3

| ประเด็น                     | ดีมาก (4)                   | ดี (3)          | พอใช้ (2)    | ปรับปรุง (1)     |
|-----------------------------|-----------------------------|-----------------|--------------|------------------|
| ความถูกต้องของการคำนวณ p, r | ถูกต้องทุกข้อ               | ถูกต้องส่วนใหญ่ | มีข้อผิดพลาด | ผิดมาก           |
| การวิเคราะห์ตัวลง           | วิเคราะห์ถูกต้องและมีเหตุผล | ถูกต้องส่วนใหญ่ | ไม่ครบ       | ไม่ถูกต้อง       |
| การตัดสินคุณภาพ/ปรับปรุง    | ตัดสินสมเหตุผลตามเกณฑ์      | ตัดสินได้       | ไม่ชัดเจน    | ไม่สอดคล้องเกณฑ์ |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 8$  จาก 12 (ร้อยละ 70 ขึ้นไป)

## 9.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

ชนาพร ไชยสุวรรณ, เบญจมาภรณ์ เสนารัตน์, และสมประสงค์ เสนารัตน์. (2568).

การพัฒนาลักษณะข้อสอบด้านคณิตศาสตร์ ตามแนวข้อสอบ National Test (NT)

สำหรับนักเรียนชั้นประถมศึกษาปีที่ 3 ในจังหวัดศรีสะเกษ. วารสารนิเชษฐาราม วัดจอมมณี, 8(5), 798–808.

บุญชม ศรีสะอาด. (2560). การวิจัยเบื้องต้น (พิมพ์ครั้งที่ 10). สุวีริยาสาส์น.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จัญญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้.

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ล้วน สายยศ และอังคณา สายยศ. (2543). เทคนิคการวัดผลการเรียนรู้. สุวีริยาสาส์น.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6).

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมนึก ภัททิยธนี. (2558). การวัดผลการศึกษา (พิมพ์ครั้งที่ 10). ประสานการพิมพ์.

สังวรณั์ จัดกระโทก, กาญจนา วัฒนสุนทร, นลินี ณ นคร, ศศิธร ชูตินันทกุล, ปารรณนา พลอภิชาติ, อนุสรณั์ เกิดศรี, ปิยนานา สิทธิฤทธิ, ณัฏฐร ชัยมมกค และนพรัตน์ ไบยา. (2561). คุณภาพข้อสอบระดับชาติ (โอเน็ต) และแนวทางการพัฒนาคุณภาพการทดสอบ. วารสารสุโขทัยธรรมาธิราช, 31(2), 110–123.

สุทธิพมกั บัญผดุง. (2568). การวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา.

อนวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่. สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

## ภาษาอังกฤษ

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Prentice-Hall.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6th ed.). Macmillan.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. Journal of Educational Psychology, 30(1), 17–24.

Nitko, A. J., & Brookhart, S. M. (2019). Educational assessment of students (8th ed.). Pearson.

Reziggalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. BMC Medical Education, 24, Article 445. <https://doi.org/10.1186/s12909-024-05433-y>

## บทที่ 10

### การประเมินตามสภาพจริงและการสร้างรูบริก

#### *Authentic Assessment and Rubric Construction*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 11 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

บทที่ 10 บูรณาการแนวคิดการวัดทักษะปฏิบัติ (บทที่ 5) และการประกันคุณภาพเครื่องมือ (บทที่ 7-9) สู่ “การประเมิน ตามสภาพจริง” (Authentic Assessment) ซึ่งเน้นการประเมินความสามารถที่แท้จริงผ่านภาระงานในบริบทจริง ผู้เรียนจะ เข้าใจแนวคิด สิ่งที่ควรประเมิน วิธีการและเครื่องมือ (ภาระงาน/ชิ้นงาน/แฟ้มสะสมงาน) และเจาะลึก “การสร้างรูบริก” (Scoring Rubric) ทั้งแบบภาพรวมและแบบแยกส่วน พร้อมขั้นตอนการพัฒนา เพื่อให้ออกแบบและใช้ประเมินตามสภาพจริง ได้อย่างเที่ยงตรงและยุติธรรม

#### แผนบริหารการสอนประจำบทที่ 10

| รายการ                | รายละเอียด                                                                                                                                                                                                           |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) ความหมายและแนวคิดการประเมินตามสภาพจริง 2) สิ่งที่ควรประเมินและลักษณะภาระงานตามสภาพจริง 3) วิธีการและเครื่องมือ (ภาระงาน/ชิ้นงาน/แฟ้มสะสมงาน) 4) รูบริก: นิยาม ประเภท และตัวอย่าง 5) ขั้นตอนการสร้างและพัฒนารูบริก |
| ความคิดรวบยอด         | การประเมินตามสภาพจริงประเมินความสามารถที่แท้จริงผ่านภาระงานในบริบทจริง<br>ครอบคลุมความรู้ การใช้เหตุผล ทักษะ เจตคติ และนิสัยการทำงาน<br>เครื่องมือสำคัญคือรูบริกที่บรรยายคุณภาพแต่ละระดับชัดเจน                      |
| จุดประสงค์การเรียนรู้ | ออกแบบและสร้างภาระงานตามสภาพจริงพร้อมรูบริกได้ (CLO3)<br>และปฏิบัติการประเมินตามสภาพจริงด้วยรูบริก ได้อย่างถูกต้องและยุติธรรม (CLO4)                                                                                 |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 10.1-10.5 ของบท)                                                                                                                                                          |

|                          |                                                                                                                                                                                                                                                                |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| กิจกรรมการเรียนรู้การสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + วิเคราะห์ตัวอย่างภาระงาน/รูบรีค 3) ใบงาน 10.1 ออกแบบภาระงานตามสภาพจริง + รูบรีค 4) ใบงาน 10.2 ทดลองใช้รูบรีคและประเมินตนเอง/เพื่อน 5) อภิปรายความเที่ยงตรง-ยุติธรรม 6) สรุป + แบบทดสอบหลังเรียน + แบบฝึกหัด |
| สื่อการเรียนรู้การสอน    | เอกสารประกอบการสอนบทที่ 10 • สไลด์นำเสนอ • ตัวอย่างภาระงานและรูบรีค (holistic/analytic) • แบบประเมินผลงาน • ใบงาน 10.1-10.2 • ตำราที่อ้างอิง                                                                                                                   |
| การวัดและประเมินผล       | แบบทดสอบก่อน-หลังเรียน • ใบงาน 10.1 ชิ้นงานภาระงาน+รูบรีค (Scoring Rubric) • ใบงาน 10.2 การประเมินจริง • การอภิปราย • แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70)                                                                                                    |

## 10.1 ความคิดรวบยอด

การประเมินตามสภาพจริง (Authentic Assessment) คือการประเมินความสามารถที่แท้จริงของผู้เรียนจากผลงานหรือการกระทำในสภาพที่เป็นจริง ซึ่งสะท้อนกระบวนการคิดที่ซับซ้อน การทำงาน การแก้ปัญหา และการตัดสินใจ (สถาบันส่งเสริม การสอนวิทยาศาสตร์และเทคโนโลยี, 2542) ต่างจากการประเมินแบบดั้งเดิมที่เน้นข้อสอบ โดยมุ่งให้ผู้เรียนประยุกต์ความรู้ และทักษะในบริบทที่มีความหมาย (Wiggins, 1989) สิ่ง que ประเมินครอบคลุมทั้งความรู้ การใช้เหตุผล ทักษะ เจตคติ และ นิสัยการทำงาน เครื่องมือสำคัญคือ “รูบรีค” (Scoring Rubric) ที่บรรยายคุณภาพของงานในแต่ละระดับอย่างชัดเจน ทั้ง แบบภาพรวม (Holistic) และแบบแยกส่วน (Analytic) ช่วยให้การประเมินภาระงานที่ใช้ดุลยพินิจมีความเป็นปรนัย เที่ยงตรง และยุติธรรม ตลอดจนเปิดโอกาสให้ผู้เรียนประเมินตนเองและเพื่อนได้ (Nitko & Brookhart, 2019)

## 10.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 10 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                     | เชื่อมโยง CLO           |
|-----------------------------------------------------------------|-------------------------|
| 1) อธิบายความหมายและแนวคิดของการประเมินตามสภาพจริงได้           | CLO3 (เข้าใจ)           |
| 2) ระบุสิ่งที่ควรประเมินและลักษณะของภาระงานตามสภาพจริงได้       | CLO3 (เข้าใจ-วิเคราะห์) |
| 3) เลือกวิธีการและเครื่องมือประเมินตามสภาพจริงให้เหมาะกับงานได้ | CLO3 (วิเคราะห์)        |
| 4) ออกแบบและสร้างภาระงานตามสภาพจริงพร้อมรูบรีคได้               | CLO3 (สร้างสรรค์)       |
| 5) ปฏิบัติการประเมินตามสภาพจริงด้วยรูบรีคได้อย่างยุติธรรม       | CLO4 (ประยุกต์)         |

### 10.3 คำสำคัญ

การประเมินตามสภาพจริง (Authentic Assessment) • ภาระงาน (Task) • ชิ้นงาน (Product) • แฟ้มสะสมงาน (Portfolio) • rubric (Scoring Rubric) • เกณฑ์ภาพรวม (Holistic) • เกณฑ์แยกส่วน (Analytic) • การประเมินตนเอง/เพื่อน (Self/Peer Assessment) • การประยุกต์ในบริบทจริง

### 10.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 10.5 เนื้อหาวิชา

---

#### 10.5.1 ความหมายและแนวคิดของการประเมินตามสภาพจริง

การประเมินตามสภาพจริง (Authentic Assessment) หมายถึง กระบวนการตัดสินความรู้ ความสามารถ และทักษะต่าง ๆ ของผู้เรียนในสภาพที่สอดคล้องกับชีวิตจริง โดยใช้เรื่องราว เหตุการณ์ หรือสถานการณ์จริงหรือคล้ายจริงที่ประสบในชีวิตประจำวันเป็นสิ่งที่เราให้ผู้เรียนตอบสนองด้วยการแสดงออก ลงมือกระทำ หรือผลิตผลงานจากกระบวนการทำงานตามที่คาดหวัง แล้วใช้ผลผลิตที่มีคุณภาพนั้นสะท้อนความรู้ความสามารถของผู้เรียนว่ามีมากน้อยเพียงใด (ศิริชัย กาญจนวาสี, 2556; พัทรินทร์ ชมภูวิเศษ, 2559) แนวคิดนี้มีรากฐานจากข้อเสนอของ Wiggins (1989) ที่เรียกร้องให้การสอบเป็น “ภารกิจที่แท้จริง” (true test) กล่าวคือ ให้ผู้เรียนได้ประยุกต์ความรู้ในบริบทที่มีความหมายและใกล้เคียงชีวิตจริง มากกว่าการตอบข้อสอบที่แยกขาดจากบริบท (decontextualized) ซึ่งวัดได้เพียงความจำระยะสั้น

ในเชิงทฤษฎี การประเมินตามสภาพจริงตั้งอยู่บนปรัชญาการสร้างความรู้ด้วยตนเอง (Constructivism) ที่เชื่อว่าผู้เรียนสร้างความหมายผ่านการลงมือปฏิบัติ จึงต้องประเมินจาก “สิ่งที่ผู้เรียนทำได้” ไม่ใช่เพียง “สิ่งที่ผู้เรียนจำได้” (สมศักดิ์ ภูวิภาดาวรรณ, 2545; สุวิมล ว่องวานิช, 2546) อนึ่ง พึงแยกความต่างระหว่างการประเมินการปฏิบัติงาน (Performance Assessment) กับการประเมินตามสภาพจริง ตำราการวัดผลหลายเล่ม

(พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ซึ่งสอดคล้องกันว่า การประเมินการปฏิบัติงานเป็นการประเมินพฤติกรรมที่ครูออกแบบหรือกำหนดขึ้น ส่วนการประเมินตามสภาพจริงเป็นการประเมินการแสดงออกที่คำนึงถึงบริบทตามสภาพจริงในชีวิตประจำวัน โดยผู้เรียนมีส่วนตัดสินใจพัฒนาผลงานของตนเองตั้งแต่เริ่มต้น ดังนั้นการประเมินการปฏิบัติงานจึงไม่ใช่การประเมินตามสภาพจริงเสมอไป

ความสำคัญของการประเมินรูปแบบนี้สอดคล้องกับการจัดการศึกษาที่มุ่งผลลัพธ์ (Outcome-Based Education) และการประเมินเพื่อพัฒนาการเรียนรู้ (Assessment for Learning) เพราะมุ่งให้ข้อมูลย้อนกลับ (feedback) เพื่อให้ผู้เรียนเห็นช่องว่างระหว่างความสามารถปัจจุบันกับเป้าหมาย แล้วกำกับการเรียนรู้ของตนเอง (self-regulation) มิใช่เพียงตัดสินผลปลายทาง การออกแบบประเมินจึงควรเริ่มจากผลลัพธ์การเรียนรู้ที่ต้องการก่อน แล้วจึงกำหนดหลักฐานและภาระงานย้อนกลับมา (backward design) (Wiggins & McTighe, 2005)

ตารางที่ 10.1 เปรียบเทียบการสอบแบบดั้งเดิมกับการประเมินตามสภาพจริง

| ประเด็นเปรียบเทียบ | การสอบแบบดั้งเดิม                | การประเมินตามสภาพจริง                     |
|--------------------|----------------------------------|-------------------------------------------|
| ลักษณะงาน          | งานที่ประดิษฐ์ขึ้น แยกจากบริบท   | งานในชีวิตจริง/สถานการณ์จำลองที่สมจริง    |
| เป้าหมายการวัด     | ความรู้ ความจำ การจำได้/จำแนกได้ | สมรรถนะและการบูรณาการความรู้สู่การปฏิบัติ |
| รูปแบบคำตอบ        | เลือกตอบ/ตอบสั้น มีคำตอบตายตัว   | สร้างสรรค์ผลงานหรือลงมือปฏิบัติจริง       |
| การเก็บข้อมูล      | สอบครั้งเดียว มีเวลาจำกัด        | สังเกต/เก็บหลักฐานต่อเนื่องหลายครั้ง      |
| การใช้ผล           | คะแนนเพื่อจัดลำดับ/ตัดสิน        | ข้อมูลย้อนกลับเพื่อพัฒนาผู้เรียน          |
| บทบาทผู้เรียน      | ผู้รับการทดสอบ (passive)         | ผู้ลงมือปฏิบัติและร่วมประเมิน (active)    |

ที่มา: ปรับจาก พัชรินทร์ ชมภูวิเศษ (2559) และ Wiggins (1989)

ดังตารางที่ 10.1 ความต่างที่เป็นหัวใจคือ “บริบท” และ “บทบาทของผู้เรียน” ตัวอย่างในชั้นเรียนไทย เช่น แทนที่จะสอบข้อเขียนเรื่องการเขียนจดหมายราชการในวิชาภาษาไทย ครูอาจมอบหมายให้ผู้เรียนเขียนจดหมายเชิญวิทยากรจริงให้แก่กิจกรรมของโรงเรียน แล้วประเมินจากจดหมายที่ใช้ได้จริง หรือในวิชาวิทยาศาสตร์ ให้ผู้เรียนออกแบบและทดลองแก้ปัญหาในสถานการณ์การตอบข้อสอบปรนัยเรื่องคุณภาพน้ำ

**ข้อควรระวัง:** “สภาพจริง” ไม่ได้หมายถึงเพียงการเปลี่ยนสถานที่สอบหรือเพิ่มกิจกรรมปฏิบัติ แต่ต้องมีเกณฑ์การประเมินที่ชัดเจนและแจ้งล่วงหน้า มิฉะนั้นผลการประเมินจะขาดความเที่ยงตรงและความยุติธรรม

นอกจากนี้ภาระงานตามสภาพจริงมักใช้เวลาและทรัพยากรมากกว่าแบบทดสอบทั่วไป ครูจึงควรเลือกใช้กับผลลัพธ์การเรียนรู้ที่เป็นทักษะ การปฏิบัติ หรือสมรรถนะที่แบบทดสอบกระดาษวัดไม่ได้ (พิชิต ฤทธิ์จัญญ, 2564)

### 10.5.2 สิ่งที่ควรประเมินและระดับของสภาพจริง

การประเมินตามสภาพจริงต้องประเมินผู้เรียนอย่างรอบด้าน ไม่จำกัดเฉพาะความรู้เชิงวิชาการ สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2542) เสนอว่าควรครอบคลุมสิ่งที่ผู้เรียนรู้และทำได้ใน 5 ด้านสำคัญ ดังกล่องที่ 10.1 ซึ่งสะท้อนการบูรณาการทั้งความรู้ (Knowledge) ทักษะ (Skill) และคุณลักษณะ (Attribute) เข้าด้วยกัน การกำหนดสิ่งที่ประเมินให้ครบทั้ง 5 ด้านช่วยป้องกันการประเมินที่เอนเอียงไปทางความจำเพียงอย่างเดียว

#### กล่องที่ 10.1 สิ่งที่ควรประเมินตามสภาพจริง 5 ด้าน

- 1) ผลการเรียนรู้ด้านวิชาการ — ความรู้ ความเข้าใจในสาระสำคัญของวิชา
- 2) การใช้เหตุผล — กระบวนการคิด การแก้ปัญหา และการสร้างความรู้ด้วยตนเอง
- 3) ทักษะและสมรรถนะ — การนำเสนอ การเขียน การทำงานเป็นทีม การวิจัย การใช้เทคโนโลยี
- 4) เจตคติ — เจตคติต่อการเรียน ความเป็นพลเมืองดี ความใฝ่รู้
- 5) นิสัยการทำงาน — ความรับผิดชอบ ความเพียร การทำงานสำเร็จตรงเวลา

ที่มา: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2542)

ภาระงาน (task) ที่ดีสำหรับการประเมินตามสภาพจริงควรเป็นสถานการณ์จริงหรือใกล้เคียงจริง มีความหมายต่อผู้เรียน เปิดให้ประยุกต์ความรู้ได้หลายด้าน และระบอบองค์ประกอบสำคัญ 3 ส่วนให้ชัดเจน ได้แก่ พฤติกรรมที่ต้องการให้แสดงออก (behavior) เงื่อนไขหรือสถานการณ์ที่กำหนด (condition) และเกณฑ์ความสำเร็จที่ยอมรับได้ (criterion) การเขียนภาระงานให้ครบสามส่วนนี้ช่วยให้ผู้เรียนเข้าใจตรงกันว่าจะถูกประเมินอย่างไร และทำให้ครูเก็บหลักฐานการเรียนรู้ (evidence of learning) ได้ตรงจุด (Marzano, Pickering, & McTighe, 1993; สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553)

อนึ่ง งานหรือกิจกรรมที่นำมาประเมินมี “ระดับของสภาพจริง” (levels of authenticity) แตกต่างกันไปเป็นช่วงต่อเนื่องจากสภาพจริงน้อยไปสู่สภาพจริงมาก บุญเชิด ภิญญอนันตพงษ์ (2545, อ้างถึงใน พิชรินทร์ ชมภูวิเศษ, 2559) เสนอให้พิจารณา 4 มิติ คือ สถานการณ์การทดสอบ รูปแบบคำตอบ วิธีเก็บรวบรวมข้อมูล และความถี่ในการทดสอบ ดังตารางที่ 10.2 ครูสามารถใช้ตารางนี้เป็น “บันได” ยกระดับเครื่องมือเดิมให้สมจริงยิ่งขึ้นทีละขั้น

## ตารางที่ 10.2 ระดับของสภาพจริง 4 มิติ (จากสภาพจริงน้อย → สภาพจริงมาก)

| มิติที่พิจารณา       | สภาพจริงน้อย                  | สภาพจริงมาก                               |
|----------------------|-------------------------------|-------------------------------------------|
| 1) สถานการณ์การทดสอบ | โจทย์สมมติบนกระดาษในห้องสอบ   | สถานการณ์จริง/บทบาทสมมติที่ใช้ของจริง     |
| 2) รูปแบบคำตอบ       | เลือกตอบหรือเติมคำตามที่กำหนด | สร้างสรรค์ผลงาน/ลงมือปฏิบัติอย่างอิสระ    |
| 3) วิธีเก็บข้อมูล    | เก็บทางอ้อมด้วยแบบทดสอบ       | เก็บทางตรงด้วยการสังเกต/สาธิต/ปฏิบัติจริง |
| 4) ความถี่ในการทดสอบ | ทดสอบครั้งเดียว มีเวลาจำกัด   | เก็บข้อมูลต่อเนื่องระหว่างเรียนหลายครั้ง  |

ที่มา: ปรับจาก บุญเชิด ภิญญอนันตพงษ์ (2545) อ้างถึงใน พชรินทร์ ชมภูวิเศษ (2559, น. 364–366)

ตัวอย่างในชั้นเรียนไทย: หัวข้อ “การทอนเงิน” ในวิชาคณิตศาสตร์ ระดับสภาพจริงน้อยคือให้นักเรียนแก้โจทย์บนกระดาษว่า “ซื้อของ 65 บาท จ่าย 100 บาท ได้เงินทอนเท่าไร” ส่วนระดับสภาพจริงมากคือจัดร้านค้าจำลองให้นักเรียนเป็นผู้ซื้อ-ผู้ขายด้วยธนบัตรจริง แล้วต้องทอนเงินให้ถูกต้องตามราคาสินค้า ดังตารางที่ 10.2 ครูจึงเลือกระดับให้เหมาะกับจุดประสงค์ ทรัพยากร และเวลา

**ข้อควรระวัง:** การยกระดับสภาพจริงต้องไม่ละเลยความเป็นปรนัยของการให้คะแนน ยิ่งงานเปิดกว้างเท่าใด ยิ่งต้องมีเกณฑ์ (รูบรีค) ที่ชัดเจนกำกับ เพื่อให้คะแนนสะท้อนความสามารถจริง ไม่ใช่ความประทับใจของผู้ประเมิน และควรเลือกประเมินเฉพาะด้านที่สอดคล้องกับจุดประสงค์ของบทเรียน ไม่ประเมินทุกด้านพร้อมกันจนเป็นภาระเกินจำเป็น (พิชิต ฤทธิ์จรูญ, 2564)

### 10.5.3 วิธีการและเครื่องมือประเมินตามสภาพจริง

เนื่องจากการประเมินตามสภาพจริงมุ่งวัดสมรรถนะที่หลากหลาย ครูจึงต้องมี “เครื่องมือที่หลากหลาย” และเลือกใช้ให้ตรงกับธรรมชาติของพฤติกรรมที่ต้องการวัด พชรินทร์ ชมภูวิเศษ (2559) และ พิชิต ฤทธิ์จรูญ (2564) สรุปรูปแบบที่ใช้บ่อยไว้ 5 รูปแบบ พร้อมข้อดีและข้อจำกัด ดังตารางที่ 10.3 หลักการเลือกคือจับคู่เครื่องมือกับระดับพฤติกรรม เช่น ทักษะปฏิบัติใช้การประเมินการปฏิบัติ ส่วนพัฒนาการระยะยาวใช้แฟ้มสะสมงาน

### ตารางที่ 10.3 รูปแบบการประเมินตามสภาพจริง: ข้อดีและข้อจำกัด

| รูปแบบการประเมิน         | ข้อดี                                   | ข้อจำกัด                                 |
|--------------------------|-----------------------------------------|------------------------------------------|
| การปฏิบัติ (Performance) | เห็นทักษะและกระบวนการทำงานจริง          | ใช้เวลา/ทรัพยากรสูง เสี่ยงอคติผู้ประเมิน |
| แฟ้มสะสมงาน (Portfolio)  | สะท้อนพัฒนาการระยะยาวและการสะท้อนคิด    | ภาระจัดเก็บ/ตรวจมาก ต้องมีเกณฑ์ชัดเจน    |
| โครงการ (Project)        | พัฒนาการคิดขั้นสูงและการแก้ปัญหา        | ให้คะแนนรายบุคคลในงานกลุ่มยาก            |
| การนำเสนอ (Presentation) | วัดการสื่อสารและการจัดระบบความคิด       | บุคลิกภาพ/ความตื่นตัวอาจกระทบคะแนน       |
| กรณีศึกษา (Case-based)   | ฝึกตัดสินใจในสถานการณ์วิชาชีพที่ซับซ้อน | ออกแบบโจทย์ให้สมจริงได้ยาก               |

ที่มา: ปรับจาก พชรินทร์ ชมภูวิเศษ (2559) และ พิชิต ฤทธิ์จรูญ (2564)

ในบรรดาเครื่องมือเหล่านี้ แฟ้มสะสมงาน (Portfolio) เป็นที่นิยมในการผลิตครู เพราะแสดงพัฒนาการต่อเนื่อง พัชรินทร์ ชมภูวิเศษ (2559) จำแนกแฟ้มสะสมงานเป็น 2 ประเภท คือ แฟ้มที่เก็บผลงานที่ดีที่สุด (Best-Work Portfolio) และแฟ้มที่แสดงความก้าวหน้าทางการเรียนรู้ (Growth Portfolio) อย่างไรก็ตาม หัวใจของแฟ้มสะสมงานไม่ได้อยู่ที่การรวบรวมชิ้นงาน แต่อยู่ที่การ “สะท้อนคิด” (reflection) ที่ให้ผู้เรียนวินิจัยจุดเด่น จุดด้อยและวางแผนพัฒนาตนเอง ทำให้แฟ้มสะสมงานกลายเป็นเครื่องมือเพื่อการเรียนรู้ ไม่ใช่เพียงที่เก็บเอกสาร (สมศักดิ์ ภูวิภาดาวรรณ, 2545)

เพื่อให้การจัดทำแฟ้มสะสมงานเป็นระบบ ชัยฤทธิ์ ศิลาเดช (2539, อ้างถึงใน พัชรินทร์ ชมภูวิเศษ, 2559) เสนอกระบวนการตามรูปแบบ COSDE ซึ่งย่อมาจาก 5 ขั้นตอน ดังกล่องที่ 10.2 ครูสามารถใช้กรอบนี้ ออกแบบให้ผู้เรียนรวบรวม คัดเลือก พัฒนา และประเมินผลงานของตนอย่างมีทิศทาง

#### กล่องที่ 10.2 รูปแบบ COSDE สำหรับการจัดทำแฟ้มสะสมงาน

C – Context (บริบท): กำหนดบริบท เป้าหมาย และเหตุผลของการจัดทำแฟ้ม

O – Objective (วัตถุประสงค์): ระบุวัตถุประสงค์และผลการเรียนรู้ที่ต้องการสะท้อน

S – Structure (โครงสร้าง): วางโครงสร้างแฟ้ม เนื้อหา ผู้เกี่ยวข้อง และหลักฐาน/ผลงาน

D – Development (การพัฒนา): สร้าง-สะสม-ประเมินผลงานตนเอง และรับข้อคิดเห็นจากผู้อื่น

E – Evaluation (การประเมินผล): คัดเลือกผลงานเด่น ใช้เกณฑ์ตัดสิน และสรุปผลการประเมิน

ที่มา: ชัยฤทธิ์ ศิลาเดช (2539) อ้างถึงใน พัชรินทร์ ชมภูวิเศษ (2559, น. 395)

ตัวอย่างในชั้นเรียนไทย: ในรายวิชาฝึกปฏิบัติวิชาชีพครู ให้นักศึกษาจัดทำแฟ้มสะสมงานอิเล็กทรอนิกส์ (e-Portfolio) รวบรวมแผนการจัดการเรียนรู้ คลิปวิดีโอการสอนจำลอง (micro-teaching) บันทึกหลังสอน และบันทึกสะท้อนคิด แล้วคัดเลือกชิ้นงานที่แสดงพัฒนาการด้านการตั้งคำถามและการจัดการชั้นเรียน พร้อมเขียนอธิบายว่าได้เรียนรู้อะไรและจะพัฒนาต่ออย่างไร

**ข้อควรระวัง:** ข้อจำกัดสำคัญของแฟ้มสะสมงานคือเกณฑ์ที่ไม่ชัดเจนจะทำให้ผลการประเมินไม่สอดคล้องกันและใช้เวลานาน (พัชรินทร์ ชมภูวิเศษ, 2559) ครูจึงต้องกำหนดรูบรีคและจำนวนชิ้นงานที่ต้องการล่วงหน้า ตลอดจนกำหนดช่วงเวลาให้ผู้เรียนสะท้อนคิดเป็นระยะ มิให้แฟ้มกลายเป็นเพียงการสะสมเอกสารจำนวนมากโดยขาดการเรียนรู้

#### 10.5.4 รูบรีค (Scoring Rubric): นิยาม ประเภท และตัวอย่าง

คำว่า “Rubric” มาจากความหมายเดิมว่า “กฎ” หรือ “กติกา” (rule) รูบรีคหรือเกณฑ์การให้คะแนน คือ เครื่องมือที่ระบุเกณฑ์การประเมินและคำอธิบายคุณภาพของชิ้นงานในแต่ละระดับ เพื่อใช้เป็นแนวทางให้

คะแนนการปฏิบัติงานหรือผลงานของผู้เรียนอย่างเป็นปรนัย (ส.วาสนา ประวาลพฤกษ์, 2544; สมศักดิ์ ภูวิภา ดาวรรณ, 2545) Suskie (2009) อธิบายว่ารูบรีคคือ “แนวทางการให้คะแนน” (scoring guide) ที่อธิบายรายการเกณฑ์เพื่อให้ผู้ประเมินใช้ตัดสินคุณภาพผลงานได้สอดคล้องกัน รูบรีคที่สมบูรณ์ประกอบด้วย 3 ส่วน ได้แก่ เกณฑ์/ประเด็นที่ประเมิน (criteria) ระดับคุณภาพ (performance levels) และคำอธิบายคุณภาพในแต่ละระดับ (descriptors) (พิชิต ฤทธิ์จัญญ, 2562)

รูบรีคใช้ประเมินได้ทั้งกระบวนการ (process) ผลผลิต (product) และการปฏิบัติ (performance) ตามหลัก 3P เช่น ประเมินกระบวนการทำงานกลุ่ม ประเมินรายงานหรือนวัตกรรมที่ผลิตขึ้น และประเมินการนำเสนอหรือการสาธิต (พิชิต ฤทธิ์จัญญ, 2562) การกำหนดรูบรีคจึงช่วยแปลงนิยามของ “คุณภาพ” ที่เป็นนามธรรมให้เป็นเกณฑ์ที่สังเกตและให้คะแนนได้ ลดความลำเอียงและเพิ่มความโปร่งใส

เมื่อจำแนกตามลักษณะการให้คะแนน รูบรีคแบ่งได้เป็น 4 ประเภทหลัก ตามแนวคิดของ Nitko (อ้างถึงใน พิชิต ฤทธิ์จัญญ, 2562) ได้แก่ แบบภาพรวม (holistic) แบบแยกส่วน (analytic) แบบทั่วไป (generic) และแบบเจาะจง (task-specific) ดังตารางที่ 10.4 นอกจากนี้ พัทธรินทร์ ชมภูวิเศษ (2559) ยังกล่าวถึงแบบผสม (annotated holistic) ที่ประเมินภาพรวมก่อนแล้วเสริมคำอธิบายจุดเด่นจุดด้อยรายประเด็น

ตารางที่ 10.4 เปรียบเทียบรูบรีค 4 ประเภท

| ประเภท                 | นิยามโดยย่อ                         | เหมาะกับ/ข้อสังเกต                        |
|------------------------|-------------------------------------|-------------------------------------------|
| ภาพรวม (Holistic)      | ให้คะแนนคุณภาพงานในภาพรวมครั้งเดียว | ตรวจเร็ว เหมาะกับการตัดสินสรุปรวม         |
| แยกส่วน (Analytic)     | ให้คะแนนแยกตามแต่ละเกณฑ์/องค์ประกอบ | ให้ข้อมูลย้อนกลับละเอียด เหมาะเพื่อพัฒนา  |
| ทั่วไป (Generic)       | บรรยายคุณภาพแบบกว้าง ใช้ได้หลายงาน  | นำกลับมาใช้ซ้ำได้ แต่ความเที่ยงอาจต่ำกว่า |
| เจาะจง (Task-specific) | บรรยายคุณภาพเฉพาะงานใดงานหนึ่ง      | ความเที่ยงสูง แต่ต้องสร้างใหม่ทุกงาน      |

ที่มา: ปรับจาก พิชิต ฤทธิ์จัญญ (2562) และ Nitko & Brookhart (2019)

การเลือกรูบรีคควรพิจารณาจุดมุ่งหมาย หากต้องการตัดสินภาพรวมอย่างรวดเร็วให้ใช้แบบภาพรวม แต่หากต้องการวินิจฉัยจุดเด่นจุดด้อยเพื่อพัฒนาผู้เรียนให้ใช้แบบแยกส่วน หลักฐานเชิงประจักษ์จากการสังเคราะห์งานวิจัย 75 เรื่องโดย Jonsson and Svingby (2007) ระบุว่า การให้คะแนนการปฏิบัติมีความเที่ยง (reliability) สูงขึ้นเมื่อใช้รูบรีคแบบแยกส่วนที่เจาะจงเนื้อหา ประกอบกับตัวอย่างชิ้นงานอ้างอิง (exemplars) และการฝึกอบรมผู้ประเมิน สอดคล้องกับ Brookhart (2018) ที่ย้ำว่ารูบรีคที่ดีต้องมีเกณฑ์ที่สอดคล้องกับผลลัพธ์การเรียนรู้ ใช้ภาษาที่วัดได้ และมีระดับคุณภาพราว 3–5 ระดับเพื่อให้ใช้งานได้จริง

ตารางที่ 10.5 ตัวอย่างรูบริกแบบแยกส่วน (Analytic) สำหรับประเมินการนำเสนอผลงาน

| ประเด็นการประเมิน | ดีมาก (4)                                  | ดี (3)                              | พอใช้ (2)                             | ปรับปรุง (1)                        |
|-------------------|--------------------------------------------|-------------------------------------|---------------------------------------|-------------------------------------|
| เนื้อหาสาระ       | ถูกต้อง ครบถ้วน<br>สอดคล้องหัวข้อทุกประการ | ถูกต้อง<br>ครบตามหัวข้อเป็นส่วนใหญ่ | ถูกต้องบางส่วน<br>ขาดสาระสำคัญบางจุด  | เนื้อหาคลาดเคลื่อน/<br>ไม่ตรงหัวข้อ |
| ทักษะการสื่อสาร   | ใช้ภาษา/น้ำเสียงชัดเจน<br>สบตาผู้ฟังดี     | สื่อสารเข้าใจได้<br>สบตาเป็นบางช่วง | สื่อสารได้บ้าง<br>น้ำเสียง/ภาษาติดขัด | สื่อสารไม่ชัด<br>อ่านตามสไลด์       |
| สื่อ/เทคนิคนำเสนอ | สื่อเหมาะสม น่าสนใจ<br>คุมเวลาได้ดี        | สื่อเหมาะสม<br>คุมเวลาได้           | สื่อพอใช้<br>คุมเวลาคลาดเคลื่อน       | สื่อไม่สอดคล้อง/เกินเวลา            |

ที่มา: ปรับจาก พิชิต ฤทธิ์จรูญ (2562, น. 13)

ดังตารางที่ 10.5 รูบริกแยกส่วนช่วยให้ผู้เรียนทราบว่าตนได้คะแนนใน “แต่ละด้าน” เพราะเหตุใด เช่น นักศึกษาที่นำเสนอแผนการสอนได้คะแนนเนื้อหาาระดับ 4 แต่ทักษะการสื่อสารระดับ 2 จะเห็นชัดว่าต้องฝึกการสบตาและการคุมน้ำเสียง ครูจึงให้ข้อมูลย้อนกลับได้ตรงจุดและผู้เรียนนำไปพัฒนาตนเองได้

**ข้อควรระวัง:** รูบริกแยกส่วนใช้เวลาสร้างและตรวจมากกว่าแบบภาพรวม หากเกณฑ์ย่อยมากเกินไปจะเป็นภาระและสับสน Brookhart (2018) จึงแนะนำให้จำกัดจำนวนเกณฑ์ที่จำเป็น และเขียนคำอธิบายแต่ละระดับให้จำแนกความแตกต่างได้จริง ไม่ใช่เพียงเติมคำว่า “มาก-น้อย” โดยไม่มีพฤติกรรมบ่งชี้

### 10.5.5 ขั้นตอนการสร้างและพัฒนารูบริก

การสร้างรูบริกที่มีคุณภาพต้องดำเนินการอย่างเป็นระบบ ไม่ใช่กำหนดคะแนนตามความรู้สึก พิชรินทร์ ชมภูวิเศษ (2559) เสนอขั้นตอนหลัก ส่วน พิชิต ฤทธิ์จรูญ (2562) ขยายให้ละเอียดขึ้น เมื่อประมวลร่วมกันได้แนวปฏิบัติ 8 ขั้น ดังนี้

- 1) กำหนดสิ่งที่ประเมินให้ชัดเจน โดยวิเคราะห์จากผลลัพธ์การเรียนรู้ว่าจะประเมินกระบวนการผลิต หรือการปฏิบัติ
- 2) กำหนดจุดมุ่งหมายของการประเมิน เพื่อเลือกจะใช้รูบริกแบบภาพรวมหรือแบบแยกส่วน
- 3) นิยามสิ่งที่ประเมิน และกำหนดประเด็น/องค์ประกอบให้ครอบคลุมและสอดคล้องกับจุดประสงค์
- 4) กำหนดจำนวนระดับคุณภาพ (นิยม 4 ระดับ และอาจมีระดับ 0 กรณีไม่ส่งงาน) พร้อมความหมายของแต่ละระดับ
- 5) เขียนคำอธิบายคุณภาพ (descriptors) ในแต่ละระดับให้สังเกตได้และจำแนกความแตกต่างชัดเจน

- 6) ตรวจสอบคุณภาพของเกณฑ์โดยผู้สอน เพื่อนครู และผู้เชี่ยวชาญด้านเนื้อหา/การวัดผล
- 7) ทดลองใช้เกณฑ์กับงานจริง และตรวจสอบความสอดคล้องระหว่างผู้ประเมิน (inter-rater)
- 8) ปรับปรุงเกณฑ์จากผลการใช้ และจัดทำคลังรูบรีคไว้ใช้/พัฒนาต่อในระดับสถานศึกษา

ขั้นที่ละเอียดอ่อนที่สุดคือการเขียนคำอธิบายคุณภาพ (ขั้นที่ 5) พิชิต ฤทธิ์จรูญ (2562) เสนอแนวทางเขียน 3 แบบ ดังกล่องที่ 10.3 ซึ่งช่วยให้ครูเขียนเกณฑ์ลดหลั่นกันอย่างมีระบบ โดยข้อความที่ใช้ต้องเป็นพฤติกรรมที่วัดหรือสังเกตได้

#### กล่องที่ 10.3 เทคนิคเขียนคำอธิบายระดับคุณภาพ (Descriptors) 3 แบบ

แบบที่ 1 ลดลง (เชิงลบ): เริ่มจากระดับสูงสุด (ดีมาก) แล้วลดคุณภาพลงทีละระดับ

แบบที่ 2 เพิ่มขึ้น (เชิงบวก): เริ่มจากระดับต่ำสุด (ปรับปรุง) แล้วเพิ่มคุณภาพขึ้นทีละระดับ

แบบที่ 3 ผสม: เริ่มจากระดับกลางที่ผ่านเกณฑ์ แล้วเพิ่ม/ลดคุณภาพไปทั้งสองทาง

ที่มา: พิชิต ฤทธิ์จรูญ (2562)

นอกจากแนวเขียนข้างต้น Nitko (อ้างถึงใน พิชิต ฤทธิ์จรูญ, 2562) ยังเสนอ 2 วิธีพัฒนาเกณฑ์ คือ วิธีจากบนสู่ล่าง (top-down) ที่เริ่มจากการออกแบบคิดของผลลัพธ์การเรียนรู้ และวิธีจากล่างสู่บน (bottom-up) ที่เริ่มจากการนำชิ้นงานจริงของผู้เรียน 10–12 ชิ้นมาจัดกลุ่มตามระดับคุณภาพแล้วสกัดเป็นเกณฑ์ การใช้ตัวอย่างชิ้นงานอ้างอิง (anchor papers) ประกอบกับการฝึกผู้ประเมินยังช่วยยกระดับความเที่ยงของการให้คะแนนอย่างมีนัยสำคัญ (Jonsson & Svingby, 2007)

ตัวอย่างในชั้นเรียนไทย: ก่อนมอบหมายงานเขียนแผนการจัดการเรียนรู้ ครูนำแผนตัวอย่างของรุ่นพี่ที่มีคุณภาพต่างกันมาให้นักศึกษาร่วมจัดระดับและอภิปรายว่าเหตุใดจึงจัดเช่นนั้น แล้วช่วยกันร่างรูบรีค จากนั้นให้นักศึกษาใช้รูบรีคเดียวกันประเมินงานของเพื่อน (peer review) เพื่อฝึกความสอดคล้องระหว่างผู้ประเมิน วิธีนี้ทำให้ผู้เรียนเข้าใจ “ภาพความสำเร็จ” ตั้งแต่ก่อนเริ่มงาน

ข้อควรระวังในการใช้รูบรีคให้ยุติธรรม สรุปได้ดังกล่องที่ 10.4 ครูควรตระหนักถึงความลำเอียงที่พบบ่อยและป้องกันด้วยการแจ้งเกณฑ์ล่วงหน้า ใช้ตัวอย่างอ้างอิง และใช้ผู้ประเมินมากกว่าหนึ่งคนเมื่อทำได้

#### กล่องที่ 10.4 ข้อพึงระวังในการใช้รูบรีค

- Halo effect: ตัดสินทุกด้านตามความประทับใจรวม — แก้โดยให้คะแนนทีละเกณฑ์
- ความโน้มเอียงสู่ค่ากลาง/ความเข้มงวด-ผ่อนปรน — แก้โดยใช้ตัวอย่างอ้างอิงและฝึกผู้ประเมิน
- ไม่แจ้งเกณฑ์ล่วงหน้า — ทำให้ผู้เรียนไม่เห็นเป้าหมายและเสียโอกาสพัฒนาตน
- ใช้ผู้ประเมินคนเดียวกับงานซับซ้อน — ควรเพิ่มผู้ประเมินและตรวจ inter-rater

ที่มา: ปรับจาก Brookhart (2018) และ Jonsson & Svingby (2007)

### เชื่อมโยงกับผลลัพธ์การเรียนรู้ของบท (CLO)

CLO3 (ออกแบบ/สร้างเครื่องมือ): เนื้อหา 10.5.2–10.5.5

รองรับการออกแบบภาระงานตามสภาพจริงและการสร้างรูบริค ประเมินด้วยชิ้นงาน “ภาระงาน+รูบริค” (ใบงาน 10.1) และรูบริคแยกส่วนในหัวข้อ 10.10

CLO4 (ปฏิบัติการวัดและประเมิน): เนื้อหา 10.5.3–10.5.5 รองรับการนำรูบริคไปใช้ประเมินจริง ประเมินด้วยการทดลองใช้รูบริค (ใบงาน 10.2) และการตรวจสอบความสอดคล้องระหว่างผู้ประเมิน

**กรณีศึกษา (สาขาการศึกษาปฐมวัย):** การประเมินตามสภาพจริงในโครงการบ้านนักวิทยาศาสตร์น้อย ศุภิสรา พวงทอง (2560) ศึกษาการประเมินผลการเรียนรู้ของผู้เรียนระดับปฐมวัยในโครงการ“บ้านนักวิทยาศาสตร์น้อย” โดยการใช้การประเมินพหุพื้นที่ ซึ่งเก็บหลักฐานการเรียนรู้ของเด็กจากการลงมือทำกิจกรรม การทดลองจริงในหลายบริบท แทนการใช้แบบทดสอบเพียงอย่างเดียว สะท้อนหลักการสำคัญของการประเมินตามสภาพจริงที่ว่า หลักฐานต้องมาจากภาระงานที่มีความหมายและใกล้เคียงชีวิตจริงของผู้เรียน สำหรับครูปฐมวัย งานประเภทนี้ยืนยันว่าการสังเกต ชิ้นงาน และการบันทึกพฤติกรรมระหว่างกิจกรรม เป็นแหล่งหลักฐานหลักของการประเมิน และ การใช้หลายพื้นที่/หลายสถานการณ์ช่วยเพิ่มความเชื่อมั่นของข้อสรุปเกี่ยวกับพัฒนาการของเด็ก

**คำถามชวนคิด:** จงออกแบบภาระงานตามสภาพจริง 1 งานสำหรับเด็กปฐมวัยหรือผู้เรียนในสาขาของท่าน พร้อมระบุหลักฐานการเรียนรู้ที่จะเก็บอย่างน้อย 3 ชนิด และวิธีสรุปผลจากหลักฐานหลายชิ้น

## 10.6 กิจกรรมการเรียนรู้การสอนและใบงาน

---

### ใบงานที่ 10.1 ออกแบบภาระงานตามสภาพจริงพร้อมรูบรีค (รายกลุ่ม)

คำชี้แจง: เลือกหน่วยการเรียนรู้ในกลุ่มสาระที่ตนถนัด ออกแบบภาระงานตามสภาพจริง 1 งาน (ระบุสถานการณ์จริง สิ่งที่ ประเมิน และผลงานที่คาดหวัง) พร้อมสร้างรูบรีคแบบแยกส่วน (Analytic) อย่างน้อย 3 ด้าน 4 ระดับ มีคำอธิบาย แต่ละระดับ (ประเมินด้วยรูบรีคในหัวข้อ 10.10) — เชื่อมโยง CLO3

### ใบงานที่ 10.2 ทดลองใช้รูบรีคประเมินตามสภาพจริง (รายบุคคล)

คำชี้แจง: ใช้รูบรีคที่กลุ่มสร้างประเมินผลงาน/การปฏิบัติของเพื่อน (หรือชิ้นงานตัวอย่าง) บันทึกคะแนนพร้อมเหตุผล และสะท้อนว่าการประเมินตามสภาพจริงด้วยรูบรีคช่วยให้ผู้เรียนอย่างไร — เชื่อมโยง CLO4

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและ กิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคุมกับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 10.7 สรุปบทเรียน

การประเมินตามสภาพจริงเป็นการประเมินความสามารถที่แท้จริงของผู้เรียนจากผลงานหรือการกระทำ ในบริบทจริง ครอบคลุมความรู้ การใช้เหตุผล ทักษะ เจตคติ และนิสัยการทำงาน ต่างจากการประเมินแบบ ดั้งเดิมที่เน้นข้อสอบ วิธีการและเครื่องมือที่ใช้ ได้แก่ การประเมินจากภาระงาน/ชิ้นงาน การแสดงออก/ กระบวนการ แฟ้มสะสมงาน และการ ประเมินตนเอง/เพื่อน เครื่องมือสำคัญคือรูบรีค (Scoring Rubric) ที่ บรรยายคุณภาพแต่ละระดับชัดเจน มีทั้งแบบภาพรวม (Holistic) และแบบแยกส่วน (Analytic) การสร้างรูบรีค ที่ดีเริ่มจากกำหนดสิ่งที่ประเมิน องค์ประกอบ จำนวนระดับ และ คำอธิบายแต่ละระดับ พร้อมทดลองใช้และ ตรวจสอบความสอดคล้องระหว่างผู้ประเมิน รูบรีคช่วยเพิ่มความเป็นปรนัย เที่ยงตรง และยุติธรรม ตลอดจนเปิด โอกาสให้ผู้เรียนรู้เกณฑ์ล่วงหน้าและพัฒนาตนเอง

## 10.8 แบบทดสอบหลังเรียน

---

หลังเรียนประจำบทเป็นแบบทดสอบ

สอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้ นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบ พัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนใน สัปดาห์ถัดไป

## 10.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

1. ครูสองคนต้องการวัด "ทักษะการทำโครงการ" ครูคนแรกใช้ข้อสอบเลือกตอบกลางภาค ส่วนครูคนที่สองใช้แฟ้มสะสมงานร่วมกับรูบริก จงวิเคราะห์จุดแข็งและจุดอ่อนของทั้งสองแนวทางในมิติความสอดคล้องกับสภาพจริง ความตรง และภาระงานของครู แล้วเสนอแนวทางผสมที่เหมาะสม
2. ระบุสิ่งที่ควรประเมินตามสภาพจริง 5 ด้าน พร้อมยกตัวอย่าง
3. อธิบายวิธีการและเครื่องมือประเมินตามสภาพจริง รวมถึงบทบาทของแฟ้มสะสมงาน
4. เปรียบเทียบรูบริกแบบภาพรวมกับแบบแยกส่วน พร้อมข้อดี-ข้อจำกัด
5. ออกแบบภาระงานตามสภาพจริง 1 งาน พร้อมรูบริกแยกส่วน 3 ด้าน 4 ระดับ
6. อธิบายขั้นตอนการสร้างและพัฒนารูบริก และเหตุผลที่ควรแจ้งเกณฑ์แก่ผู้เรียนล่วงหน้า

## 10.10 เครื่องมือวัดและประเมินผลประจำบท

ก) รูบริกประเมินภาระงาน+รูบริก ใบงาน 10.1 (Analytic) — เชื่อมโยง CLO3

| ประเด็น                             | ดีมาก (4)                  | ดี (3)        | พอใช้ (2)   | ปรับปรุง (1)    |
|-------------------------------------|----------------------------|---------------|-------------|-----------------|
| ความเป็นสภาพจริงของภาระงาน          | บริบทจริง<br>มีความหมายชัด | ค่อนข้างจริง  | จริงบางส่วน | ไม่เป็นสภาพจริง |
| ความครอบคลุมสิ่งที่ประเมิน          | ครอบคลุมหลายด้าน           | ครอบคลุมพอควร | จำกัด       | แคบมาก          |
| คุณภาพของรูบริก<br>(ระดับ/คำอธิบาย) | ชัดเจนจำแนกได้ทุกด้าน      | ชัดส่วนใหญ่   | ไม่ครบ      | ไม่ชัด          |
| ความเป็นปรนัย/ยุติธรรม              | เกณฑ์ลดอคติได้ดี           | ค่อนข้างดี    | พอใช้       | ยังลำเอียง      |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 11$  จาก 16 (ร้อยละ 70 ขึ้นไป)

## ข) แบบบันทึกการประเมินตามสภาพจริง ใบบาง 10.2 (โดยย่อ) — เชื่อมโยง CLO4

บันทึก: ใช้รูปรีคให้คะแนนแต่ละด้านอย่างมีเหตุผล • ให้ข้อมูลย้อนกลับเชิงพัฒนา • ประเมินสม่ำเสมอ ไม่ลำเอียง • เปิดให้ผู้เรียนประเมินตนเอง/เพื่อน

### 10.11 เอกสารอ้างอิงประเภท

---

#### ภาษาไทย

บุญเชิด ภิญโญนันตพงษ์. (2545). การประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ: แนวคิดและวิธีการ. อมรินทร์พริ้นติ้งแอนด์พับลิชชิ่ง.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการศึกษา. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิจรรณ. (2562). เกณฑ์การให้คะแนน: เครื่องมือสำหรับครูเพื่อประเมินผลลัพธ์การเรียนรู้ที่เที่ยงตรงและยุติธรรม. วารสารศึกษาศาสตร์ มสธ., 12(1), 1–13.

พิชิต ฤทธิจรรณ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้. ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 7). โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศุภิสรา พวงทอง. (2560). การประเมินผลการเรียนรู้ของผู้เรียนระดับปฐมวัยโครงการบ้านนักวิทยาศาสตร์น้อยโดยใช้การประเมินพหุพื้นที่. วารสารการวัดผลการศึกษา, 34(96), 55–67.

ส.วาสนา ประवालพฤษ. (2544). หลักการและเทคนิคการประเมินทางการศึกษา. เดอะมาสเตอร์กรุ๊ป แมเนจเม้นท์.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2542). การประเมินผลการเรียนรู้ตามสภาพจริง. สสวท.

สมศักดิ์ ภูวิภาดารรณ์. (2545). การยึดผู้เรียนเป็นศูนย์กลางและการประเมินตามสภาพจริง. เดอะโนว์เลจเซ็นเตอร์.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2553). แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

สุวิมล ว่องวานิช. (2546). การประเมินผลการเรียนรู้แนวใหม่. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

อนวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่.  
สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

## ภาษาอังกฤษ

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, Article 22. <https://doi.org/10.3389/feduc.2018.00022>

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.  
<https://doi.org/10.1016/j.edurev.2007.05.002>

Marzano, R. J., Pickering, D., & McTighe, J. (1993). Assessing student outcomes: Performance assessment using the dimensions of learning model. ASCD.

Nitko, A. J., & Brookhart, S. M. (2019). Educational assessment of students (8th ed.). Pearson.

Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.). Jossey-Bass.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703–713.

Wiggins, G., & McTighe, J. (2005). Understanding by design (2nd ed.). ASCD.

## บทที่ 11

### การใช้ผลการประเมินเพื่อพัฒนาผู้เรียน

#### Using Assessment Results for Learning Development

#### (Assessment for Learning & Feedback)

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สัปดาห์ที่ 12 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

หลังจากสร้างและตรวจสอบเครื่องมือ ตลอดจนเก็บผลการประเมินแล้ว คำถามสำคัญคือ “จะนำผลไปใช้พัฒนาผู้เรียนอย่างไร” บทที่ 11 มุ่งที่แนวคิด “การประเมินเพื่อการเรียนรู้” (Assessment for Learning: AfL) และหัวใจของมันคือ “การให้ข้อมูล ย้อนกลับ” (Feedback) ที่มีคุณภาพ ผู้เรียนจะเข้าใจความต่างของ AfL/AaL/AoL กลยุทธ์การประเมินเพื่อพัฒนา โมเดล การให้ข้อมูลย้อนกลับของแฮตตีและทิมเพอร์ลีย์ และการนำผลประเมินไปปรับการสอนและพัฒนาผู้เรียนอย่างสร้างสรรค์

#### แผนบริหารการสอนประจำบทที่ 11

| รายการ                   | รายละเอียด                                                                                                                                                                                                                                                                  |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ               | 1) แนวคิดการประเมินเพื่อการเรียนรู้ (AfL/AaL/AoL) 2) การประเมินเพื่อพัฒนา (Formative Assessment) และกลยุทธ์ 3) การให้ข้อมูลย้อนกลับ (Feedback) และโมเดลของแฮตตีและทิมเพอร์ลีย์ 4) การนำผลประเมินไปพัฒนาการสอนและผู้เรียน 5) บทบาทครู-ผู้เรียน และจรรยาบรรณในการใช้ผลประเมิน |
| ความคิดรวบยอด            | การประเมินมีคุณค่าสูงสุดเมื่อนำผลไปพัฒนาผู้เรียน ไม่ใช่เพียงตัดสิน การประเมินเพื่อการเรียนรู้เน้นการให้ข้อมูล ย้อนกลับที่ตอบ 3 คำถาม (ไปไหน-ไปถึงไหนแล้ว-ก้าวต่ออย่างไร) เพื่อให้ผู้เรียนพัฒนาได้จริง                                                                       |
| จุดประสงค์การเรียนรู้    | ใช้ผลการวัดและประเมินเพื่อวางแผนและพัฒนาการเรียนรู้ของผู้เรียนได้อย่างสร้างสรรค์ ผ่านการให้ข้อมูลย้อนกลับ ที่มีคุณภาพ (CLO6)                                                                                                                                                |
| เนื้อหาวิชา              | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 11.1–11.5 ของบท)                                                                                                                                                                                                                 |
| กิจกรรมการเรียนรู้การสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + วิเคราะห์ตัวอย่าง feedback ดี/ไม่ดี 3) ใบงาน 11.1 เขียน feedback เชิงพัฒนาจากกรณีศึกษา 4) ใบงาน 11.2                                                                                                                     |

|                    |                                                                                                                                                           |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | วางแผนใช้ผลประเมินพัฒนาผู้เรียน + นำเสนอ 5) อภิปรายจรรยาบรรณการให้ข้อมูลย้อนกลับ 6) สรุป + แบบทดสอบหลังเรียน + แบบฝึกหัด                                  |
| สื่อการเรียนการสอน | เอกสารประกอบการสอนบทที่ 11 • สไลด์นำเสนอ • กรณียศึกษา • ตัวอย่าง feedback • โมเดล Hattie & Timperley • ใบงาน 11.1–11.2 • ตำราที่อ้างอิง                   |
| การวัดและประเมินผล | แบบทดสอบก่อน–หลังเรียน • ใบงาน 11.1 (รูปบริการให้ข้อมูลย้อนกลับ) • ใบงาน 11.2 (การนำเสนอแนวทางพัฒนา) • การอภิปราย • แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70) |

## 11.1 ความคิดรวบยอด

การประเมินมีคุณค่าสูงสุดเมื่อนำผลไปใช้พัฒนาผู้เรียน ไม่ใช่เพียงตัดสินได้–ตก แนวคิด “การประเมินเพื่อการเรียนรู้” (Assessment for Learning: AfL) มุ่งใช้สารสนเทศจากการประเมินเพื่อปรับการสอนและส่งเสริมการเรียนรู้ขณะที่ยังเรียนอยู่ งานวิจัยคลาสสิกของ Black และ Wiliam (1998) ใน “Inside the Black Box” ยืนยันว่าการประเมินเพื่อพัฒนาที่มีคุณภาพ ช่วยยกระดับผลสัมฤทธิ์ของผู้เรียนได้อย่างมีนัยสำคัญ หัวใจของ AfL คือ “การให้ข้อมูลย้อนกลับ” (Feedback) ที่มีคุณภาพ ซึ่ง Hattie และ Timperley (2007) เสนอว่าต้องตอบ 3 คำถาม คือ ฉันกำลังจะไปไหน (Feed Up) ฉันไปถึงไหนแล้ว (Feed Back) และฉันจะก้าวต่อไปอย่างไร (Feed Forward) ครุจึงต้องนำผลประเมินมาวางแผนสอนซ่อมเสริม ปรับวิธีสอน และส่งเสริมให้ผู้เรียนตั้งเป้าหมายและกำกับการเรียนรู้ของตนเอง (พิชิต ฤทธิ์จรรยา, 2564; สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553)

## 11.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 11 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                               | เชื่อมโยง CLO              |
|---------------------------------------------------------------------------|----------------------------|
| 1) อธิบายแนวคิดและความต่างของการประเมินเพื่อการเรียนรู้ (AfL/AaL/AoL) ได้ | CLO6 (เข้าใจ)              |
| 2) อธิบายกลยุทธ์ของการประเมินเพื่อพัฒนา (Formative Assessment) ได้        | CLO6 (เข้าใจ)              |
| 3) ให้ข้อมูลย้อนกลับเชิงพัฒนาตามโมเดล 3 คำถามได้                          | CLO6 (ประยุกต์)            |
| 4) วางแผนนำผลการประเมินไปพัฒนาการสอนและผู้เรียนได้อย่างสร้างสรรค์         | CLO6 (ประยุกต์–ประเมินค่า) |
| 5) ตระหนักในจรรยาบรรณของการให้ข้อมูลย้อนกลับที่สร้างกำลังใจและเป็นธรรม    | CLO6 (จิตพิสัย)            |

### 11.3 คำสำคัญ

การประเมินเพื่อการเรียนรู้ (Assessment for Learning: AfL) • การประเมินในฐานะการเรียนรู้ (Assessment as Learning) • การประเมินผลการเรียนรู้ (Assessment of Learning) • การประเมินเพื่อพัฒนา (Formative Assessment) • การให้ข้อมูลย้อนกลับ (Feedback) • Feed Up/Back/Forward • การสอนซ่อมเสริม (Remedial Teaching) • การกำกับกับการเรียนรู้ด้วยตนเอง (Self-regulation)

### 11.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 11.5 เนื้อหาวิชา

---

#### 11.5.1 แนวคิดการประเมินเพื่อการเรียนรู้ (AfL / AaL / AoL)

เมื่อพิจารณา “บทบาทของการประเมินที่มีต่อการเรียนรู้” สามารถจำแนกการประเมินออกเป็น 3 ลักษณะที่เสริมกัน ได้แก่ การประเมินเพื่อตัดสินผลการเรียนรู้ (Assessment of Learning: AoL) ซึ่งสรุปและรับรองสมรรถนะเมื่อจบหน่วยหรือภาคเรียน การประเมินเพื่อพัฒนาการเรียนรู้ (Assessment for Learning: AfL) ซึ่งเก็บหลักฐานระหว่างเรียนเพื่อปรับการสอนและช่วยให้ผู้เรียนก้าวหน้า และการประเมินในฐานะการเรียนรู้ (Assessment as Learning: AaL) ซึ่งให้ผู้เรียนประเมินและกำกับกับการเรียนรู้ของตนเอง การจำแนกสามลักษณะนี้เป็นกรอบที่ได้รับการยอมรับอย่างกว้างขวางในวรรณกรรมด้านการวัดและประเมินผลการศึกษา (Nitko และ Brookhart, 2019)

จุดมุ่งหมายพื้นฐานของการวัดและประเมินผลในชั้นเรียนมี 2 ประการ คือ การประเมินเพื่อพัฒนาผู้เรียน และการประเมินเพื่อตัดสินผลการเรียน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553) โดย ดำรงการวัดผลหลายเล่ม (พัชรินทร์ ชมภูวิเศษ, 2559; อนุวัติ คุณแก้ว, 2566) ชี้ว่าสอดคล้องกันว่าการประเมินเพื่อพัฒนาคือการประเมินย่อย (Formative Assessment) ที่เกิดขึ้นในห้องเรียนทุกวัน เพื่อให้รู้จุดเด่นและจุดที่ต้อง

ปรับปรุง แล้วนำสารสนเทศนั้นไปใช้พัฒนาต่อ ส่วน สมพร มณีอ่อน (2559) ย้ำว่า การประเมินเพื่อการเรียนรู้ เป็นพลวัตในห้องเรียน การสอนและการประเมินต้องวางแผนให้ไปด้วยกัน และต้องเปลี่ยนมุมมองของครูจาก “การประเมินเพื่อจัดอันดับ” มาเป็น “การประเมินเพื่อพัฒนาการเรียนรู้”

ตัวอย่างในชั้นเรียนไทย: ครูภาษาไทยชั้น ม.2 มอบหมายให้เขียนเรียงความ หากครูเก็บงานไปให้ คะแนนแล้วคืนตอนปลายภาค ถือเป็น AoL แต่ถ้าครูเก็บ “ร่างแรก” มาอ่านระหว่างทาง แล้วเขียนคำแนะนำให้ ปรับก่อนส่งฉบับจริง ถือเป็น AfL และถ้าครูให้นักเรียนใช้รายการตรวจสอบ (checklist) ประเมินเรียงความของตนเองเทียบเกณฑ์ก่อนส่ง ก็คือ AaL ซึ่งฝึกให้ผู้เรียนเป็นเจ้าของการเรียนรู้

ข้อควรระวังที่พบบ่อย คือ ครูจำนวนมากให้น้ำหนัก AoL จนการประเมินย่อยกลายเป็นเพียง “การเก็บคะแนนย่อย” โดยไม่ได้นำผลไปปรับการสอน (สมพร มณีอ่อน, 2559) ครูจึงควรออกแบบให้ทั้งสามลักษณะ สมดุลกัน และต้องแน่ใจว่าทุกครั้งที่เก็บข้อมูล ต้องมี “การกระทำตามผล” ตามมาเสมอ ความแตกต่างของทั้ง สามลักษณะสรุปได้ดังตารางที่ 11.1

ตารางที่ 11.1 เปรียบเทียบการประเมิน 3 ลักษณะ: AoL / AfL / AaL

| ลักษณะ                   | จุดมุ่งหมาย           | ผู้ใช้ผลหลัก            | ช่วงเวลา              | คำถามนำ                     |
|--------------------------|-----------------------|-------------------------|-----------------------|-----------------------------|
| AoL<br>เพื่อตัดสิน       | ตัดสิน/รับรองสมรรถนะ  | ครู ผู้บริหาร ผู้ปกครอง | สิ้นสุดหน่วย/ภาคเรียน | “ผู้เรียนได้ระดับใด”        |
| AfL<br>เพื่อพัฒนา        | วินิจฉัยและปรับการสอน | ครูและผู้เรียน          | ระหว่างเรียน          | “จะช่วยให้ก้าวต่ออย่างไร”   |
| AaL<br>ในฐานะการเรียนรู้ | สร้างการกำกับตนเอง    | ผู้เรียน                | ตลอดกระบวนการ         | “ฉันรู้อะไรแล้ว/ยังขาดอะไร” |

ที่มา: ปรับจาก สมพร มณีอ่อน (2559) และ พัทรินทร์ ขมภูวิเศษ (2559, น. 13)

### 11.5.2 การประเมินเพื่อพัฒนา (Formative Assessment) และกลยุทธ์

การประเมินเพื่อพัฒนา (Formative Assessment) หมายถึง การเก็บและตีความหลักฐานการเรียนรู้ ระหว่างการสอน เพื่อนำมาใช้ปรับการสอนของครู และปรับการเรียนรู้ของผู้เรียนให้ก้าวหน้าขึ้น มิใช่เพื่อตัดสิน ได้-ตก พัทรินทร์ ขมภูวิเศษ (2559, น. 13) ระบุว่า “สิ่งสำคัญที่สุด” ของการประเมินเพื่อพัฒนา คือ การให้ ข้อมูลย้อนกลับแก่ผู้เรียนในลักษณะคำแนะนำที่เชื่อมโยงความรู้เดิมกับความรู้ใหม่ จนผู้เรียนสามารถ ตั้งเป้าหมายและพัฒนาตนเองได้หลักฐานเชิงประจักษ์ที่มีชื่อเสียงที่สุดมาจากงานสังเคราะห์ของ Black และ Wiliam (1998) เรื่อง Inside the Black Box ซึ่งทบทวนงานวิจัยหลายร้อยชิ้น และพบว่า การประเมินเพื่อ

พัฒนาที่มีคุณภาพช่วยยกลสัมฤทธิ์ได้อย่างมีนัยสำคัญ โดยมีขนาดอิทธิพล (effect size) ราว 0.4–0.7 ส่วน เบี่ยงเบนมาตรฐาน และช่วยกลุ่มผู้เรียนอ่อนได้มากเป็นพิเศษ ต่อมา Wiliam (2011) ได้ตกผลึกแนวปฏิบัติ ออกเป็นกลยุทธ์สำคัญ 5 ประการ ดังตารางที่ 11.2

ตารางที่ 11.2 กลยุทธ์สำคัญ 5 ประการของการประเมินเพื่อพัฒนา

| กลยุทธ์                                      | แนวปฏิบัติในชั้นเรียน                                              |
|----------------------------------------------|--------------------------------------------------------------------|
| 1) ทำเป้าหมาย/เกณฑ์ความสำเร็จให้ชัด          | แจ้งเกณฑ์และยกตัวอย่างงานที่ดีให้ผู้เรียนเห็นภาพร่วมกัน            |
| 2) ตั้งหลักฐานการเรียนรู้                    | ใช้คำถาม/กิจกรรมที่ทำให้เห็นความเข้าใจของทั้งห้อง เช่น กระดานคำตอบ |
| 3) ให้ข้อมูลย้อนกลับที่ทำให้ก้าวหน้า         | เขียนคำแนะนำเชิงพัฒนา แทนการให้เพียงคะแนน                          |
| 4) ใช้เพื่อนเป็นแหล่งเรียนรู้                | จัดให้ผู้เรียนตรวจและช่วยกันพัฒนางานตามเกณฑ์                       |
| 5) ส่งเสริมให้ผู้เรียนเป็นเจ้าของการเรียนรู้ | ฝึกตั้งเป้าหมายและประเมินตนเอง (self-regulation)                   |

ที่มา: ปรับจาก Wiliam (2011) และ Black และ Wiliam (1998)

ตัวอย่างในชั้นเรียนไทย: ครูคณิตศาสตร์ชั้น ป.6 ขณะสอนการบวกเศษส่วน ใช้เทคนิค “ไฟจราจร” ให้นักเรียนชูการ์ดสี เขียว–เหลือง–แดง เพื่อบอกระดับความเข้าใจของตน เมื่อครูเห็นว่าครึ่งห้องชูการ์ดเหลือง จึงหยุดเพื่อยกตัวอย่างเพิ่มและจับคู่ให้เด็กที่เข้าใจช่วยอธิบายให้เพื่อน ก่อนเดินหน้าสู่เนื้อหาถัดไป นี่คือการตั้งหลักฐาน (กลยุทธ์ที่ 2) และใช้เพื่อนเป็นแหล่งเรียนรู้ (กลยุทธ์ที่ 4) ในเวลาเดียวกัน

**ข้อควรระวัง:** การประเมินเพื่อพัฒนาจะได้ผลต่อเมื่อมี “การกระทำตามผล” คือครูหรือผู้เรียนเปลี่ยน พฤติกรรมจากสารสนเทศที่ได้ หากเก็บข้อมูลแล้วสอนต่อตามแผนเดิม ก็เป็นเพียงการประเมินสรุปย่อย ๆ ที่ไม่ เกิดการพัฒนา นอกจากนี้ Black และ Wiliam (1998) เตือนว่า การให้คะแนนกำกับทุกชิ้นงานมักเบนความ สนใจของผู้เรียนออกจากคำแนะนำเชิงพัฒนา ครูจึงควรลดการให้คะแนนย่อย และเพิ่มการให้ข้อมูลย้อนกลับเชิง คุณภาพแทน โดยสรุปหลักฐานสำคัญไว้ในกล่อง 11-1

#### กล่อง 11-1 หลักฐานวิจัยที่ครูควรรู้

Black และ Wiliam (1998): การประเมินเพื่อพัฒนาที่มีคุณภาพ ยกลสัมฤทธิ์ราว 0.4–0.7 SD และช่วยกลุ่ม อ่อนมากที่สุด นัยต่อการสอน: ลงทุนกับ ‘คำแนะนำ’ มากกว่า ‘คะแนน’ และเปิดโอกาสให้ผู้เรียนแก้ไขงาน หลังได้รับข้อมูลย้อนกลับ

### 11.5.3 การให้ข้อมูลย้อนกลับ (Feedback) — โมเดลของแฮตตีและทิมเพอร์ลีย์

ข้อมูลย้อนกลับ (Feedback) หมายถึง สารสนเทศที่ครู เพื่อน หรือตัวผู้เรียนเอง ให้เกี่ยวกับผลการ ปฏิบัติเทียบกับเป้าหมาย เพื่อลดช่องว่างระหว่าง “จุดที่ผู้เรียนอยู่” กับ “จุดที่ควรไปถึง” Hattie และ

Timperley (2007) เสนอว่าข้อมูลย้อนกลับที่มีพลังต้องตอบ 3 คำถามสำคัญ ได้แก่ Feed Up (ฉันกำลังจะไปไหน — เป้าหมาย), Feed Back (ฉันทำได้ถึงไหนแล้ว — สถานะปัจจุบัน) และ Feed Forward (ฉันจะก้าวต่อไปอย่างไร — ขั้นถัดไป) ดังแสดงตัวอย่างถ้อยคำของครูในตารางที่ 11.3

ตารางที่ 11.3 สามคำถามของข้อมูลย้อนกลับที่มีพลัง พร้อมตัวอย่างคำพูดของครู

| คำถาม                                 | มุ่งตอบเรื่อง              | ตัวอย่างคำพูดของครู                                              |
|---------------------------------------|----------------------------|------------------------------------------------------------------|
| Feed Up<br>ฉันกำลังจะไปไหน            | เป้าหมาย/เกณฑ์ความสำเร็จ   | “งานนี้ต้องการให้นักเรียนยกเหตุผลสนับสนุนข้อสรุปอย่างน้อย 2 ข้อ” |
| Feed Back<br>ฉันทำได้ถึงไหนแล้ว       | สถานะปัจจุบันเทียบเป้าหมาย | “ย่อหน้าสองมีเหตุผล 1 ข้อแล้ว ยังขาดอีก 1 ข้อจึงจะครบเกณฑ์”      |
| Feed Forward<br>ฉันจะก้าวต่อไปอย่างไร | ขั้นถัดไปที่ทำได้จริง      | “ลองเพิ่มตัวอย่างจากข่าวจริงหนึ่งเรื่องเพื่อสนับสนุนข้อสรุป”     |

ที่มา: ปรับจาก Hattie และ Timperley (2007)

นอกจาก 3 คำถามแล้ว Hattie และ Timperley (2007) ยังชี้ว่าข้อมูลย้อนกลับทำงานได้ใน 4 ระดับ ซึ่งมีพลังต่อการเรียนรู้ไม่เท่ากัน ครูควรเล็งไปที่ระดับกระบวนการและการกำกับตนเองเป็นหลัก และหลีกเลี่ยงคำชม-ตำหนิที่ตัวบุคคล ดังสรุปในกล่อง 11-2

**การประเมินเพื่อพัฒนา (Formative Assessment)** หมายถึง การเก็บและตีความหลักฐานการเรียนรู้ระหว่างการสอน เพื่อนำมาใช้ปรับการสอนของครู และปรับการเรียนรู้ของผู้เรียนให้ก้าวหน้าขึ้น มิใช่เพื่อตัดสินได้-ตก พัทธินทร ชมภูวิเศษ (2559, น. 13) ระบุว่า “สิ่งสำคัญที่สุด” ของการประเมินเพื่อพัฒนา คือ การให้ข้อมูลย้อนกลับแก่ผู้เรียนในลักษณะคำแนะนำที่เชื่อมโยงความรู้เดิมกับความรู้ใหม่ จนผู้เรียนสามารถตั้งเป้าหมายและพัฒนาตนเองได้

ตัวอย่างในชั้นเรียนไทย: ครูสังคมศึกษาชั้น ม.3 ตรวจผังความคิด (mind map) ของนักเรียน แทนที่จะเขียนเพียง “7/10 ดี” ครูเขียนตามสามคำถามว่า “เป้าหมายคือเชื่อมโยงสาเหตุกับผลให้ครบ (Feed Up) ตอนนี้หนูโยงได้ 3 คู่แล้ว แต่คู่ที่ 4 ยังขาดลูกศร (Feed Back) ลองเพิ่มเส้นเชื่อมระหว่างเศรษฐกิจกับการย้ายถิ่น (Feed Forward)” ข้อมูลย้อนกลับเช่นนี้ชี้ชี้ทางการพัฒนาได้ชัดเจนกว่าเพียงตัวเลข

อย่างไรก็ตาม ข้อมูลย้อนกลับมิได้มีพลังเสมอไป Wisniewski, Zierer และ Hattie (2020) สังเคราะห์งานวิจัย 435 เรื่อง พบว่าโดยเฉลี่ย feedback มีอิทธิพลขนาดกลาง ( $d = 0.48$ ) แต่ผลแปรปรวนสูงมากและขึ้นอยู่กับ “สาระของข้อมูล” มากกว่ารูปแบบ กล่าวคือ ข้อมูลย้อนกลับที่ให้สารสนเทศเชิงกระบวนการและชี้แนะ

ถัดไป ได้ผลกว่าการให้คะแนนหรือคำชมลอย ๆ ข้อพึงระวังของครู คือ ต้องให้ข้อมูลย้อนกลับให้ทันเวลาขณะที่ผู้เรียนยังแก้ไขงานได้ เจาะจงที่งานและกระบวนการ และต้องเปิดโอกาสให้ผู้เรียนนำคำแนะนำไปปรับปรุงจริง

#### 11.5. 4 การนำผลการประเมินไปพัฒนาการสอนและผู้เรียน

เมื่อได้ผลการประเมินแล้ว ขั้นที่สำคัญที่สุดคือการแปลคะแนนให้กลายเป็น “สารสนเทศ” เพื่อการตัดสินใจ พัทธินทร์ ชมภูวิเศษ (2559, น. 344) ระบุว่าครูสามารถนำผลการทดสอบไปใช้พัฒนาผู้เรียนได้

3 ลักษณะ คือ การวินิจฉัยผู้เรียน (Diagnosis) ก่อนเรียนเพื่อซ่อมเสริมความรู้พื้นฐาน การปรับปรุงการเรียนรู้ระหว่างเรียน (Formative) เมื่อพบจุดบกพร่อง และการตัดสินผลพร้อมส่งเสริม/รายงานผลหลังเรียน ดังตารางที่ 11.4

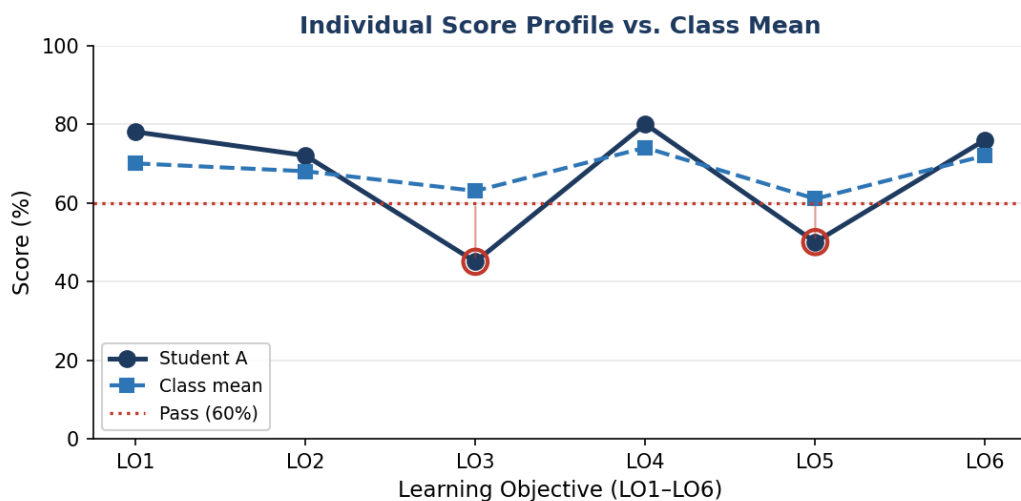
ตารางที่ 11.4 สามลักษณะของการนำผลการประเมินไปใช้พัฒนาผู้เรียน

| ลักษณะการใช้ผล           | ช่วงเวลา     | การกระทำของครู                                               |
|--------------------------|--------------|--------------------------------------------------------------|
| วินิจฉัย (Diagnosis)     | ก่อนเรียน    | ตรวจความพร้อม/ความรู้พื้นฐาน แล้วซ่อมเสริมก่อนเริ่มหน่วยใหม่ |
| ปรับปรุงการเรียนรู้      | ระหว่างเรียน | พบจุดบกพร่องแล้วปรับวิธีสอน/จังหวะ และจัดซ่อมเสริมทันที      |
| ตัดสิน + ส่งเสริม/รายงาน | หลังเรียน    | สรุปผล ส่งเสริมผู้เรียนเก่งให้ก้าวหน้า และรายงานผู้ปกครอง    |

ที่มา: ปรับจาก พัทธินทร์ ชมภูวิเศษ (2559, น. 344)

การวินิจฉัยที่แม่นยำต้องดูผลเป็น “รายจุดประสงค์” มิใช่ดูเพียงคะแนนรวม เครื่องมือที่ช่วยได้คือการทำ “เส้นภาพผลรายบุคคล” (profile) ที่นำคะแนนของผู้เรียนแต่ละคนในแต่ละจุดประสงค์มาวางเทียบกับค่าเฉลี่ยของชั้นและเกณฑ์ผ่าน ดังภาพที่ 11.1 จุดที่เส้นของผู้เรียนต่ำกว่าเส้นเกณฑ์ผ่าน คือจุดบกพร่องที่ควรซ่อมเสริมเป็นลำดับแรก การมองความแตกต่างระหว่างผู้เรียนยังอาศัยค่าสถิติพื้นฐาน เช่น ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ซึ่งบ่งบอกการกระจายของผลการเรียน (ศิริชัย กาญจนวาสี, 2552)

ภาพที่ 11.1 เส้นภาพแสดงผลการประเมินรายบุคคลเทียบค่าเฉลี่ยชั้นและเกณฑ์ผ่าน



ตัวอย่างในชั้นเรียนไทย: ครูวิทยาศาสตร์ชั้น ม.1 หลังสอบเก็บคะแนนหน่วย “สารและสมบัติของสาร” พบว่า นักเรียน ก ได้คะแนนจุดประสงค์ที่ 3 (ทักษะการทดลอง) และจุดประสงค์ที่ 5 ต่ำกว่าเกณฑ์ ขณะที่จุดประสงค์ด้านความรู้ความจำผ่านหมด (ดูจุดที่วงไว้ในภาพที่ 11.1) ครูจึงวางแผนซ่อมเสริมเฉพาะ “ทักษะปฏิบัติการทดลอง” ด้วยการให้ฝึกกับอุปกรณ์จริงเป็นกลุ่มย่อย แทนการสอนซ้ำทั้งหน่วย ซึ่งเปลืองเวลาและไม่ตรงจุด ส่วนนักเรียนที่ผ่านทุกจุดประสงค์ ครูมอบโจทย์ท้าทาย (Enrichment) ให้ออกแบบการทดลองของตนเอง

ข้อควรระวัง: ครูควรนำผลย้อนกลับมาพัฒนา “การสอนของตนเอง” ด้วย เช่น หากผู้เรียนส่วนใหญ่พลาดจุดประสงค์เดียวกัน แสดงว่ากิจกรรมหรือสื่อในจุดนั้นอาจยังไม่เหมาะ ต้องปรับแผนการสอนและเครื่องมือวัด (พิชิต ฤทธิ์จรูญ, 2564) นอกจากนี้การจัดกลุ่มผู้เรียนตามผลประเมินต้องเป็นการจัดแบบยืดหยุ่นชั่วคราวเพื่อช่วยเหลือ ไม่ใช่การตีตราถาวรที่บั่นทอนกำลังใจ

#### 11.5.5 บทบาทครู-ผู้เรียน และจรรยาบรรณในการใช้ผลประเมิน

ภายใต้แนวคิดการประเมินเพื่อการเรียนรู้ บทบาทของทั้งสองฝ่ายเปลี่ยนไป ครูเปลี่ยนจาก “ผู้ตัดสิน” มาเป็น “ผู้สนับสนุนการเรียนรู้” ส่วนผู้เรียนเปลี่ยนจาก “ผู้ถูกประเมิน” มาเป็น “ผู้ร่วมประเมินและกำกับการเรียนรู้ของตนเอง” ผ่านการประเมินตนเอง (self-assessment) และการประเมินโดยเพื่อน (peer assessment) ซึ่งเป็นกลไกหลักของการประเมินในฐานการเรียนรู้ (AaL)

งานวิชาการไทยสนับสนุนแนวทางนี้ พินดา วราสุนันท์ (2558) อธิบายว่า การให้ข้อมูลย้อนกลับโดยเพื่อน (peer feedback) ส่งเสริมการมีส่วนร่วมระหว่างผู้เรียนกับครู ช่วยเปลี่ยนพฤติกรรมการเรียนรู้ในระยะยาวได้มากกว่าการรับข้อมูลย้อนกลับจากครูเพียงฝ่ายเดียว และสร้างบรรยากาศการเรียนรู้แบบมิตร สอดคล้องกับหลักฐานเชิงปริมาณของ Double, McGrane และ Hopfenbeck (2020) ที่สังเคราะห์งานวิจัย 54 เรื่อง พบว่าการประเมินโดยเพื่อนช่วยยกผลสัมฤทธิ์ ( $g = 0.31$ ) สูงกว่าทั้งกลุ่มที่ไม่มีการประเมินและกลุ่มที่ประเมินโดยครูอย่างเดียว

ตัวอย่างในชั้นเรียนไทย: ก่อนส่งเรียงความให้ครู ครูภาษาไทยจัดให้นักเรียนจับคู่ตรวจงานของเพื่อนตามรูปรีด 3 เกณฑ์ (ใจความ-การลำดับความ-ภาษา) แล้วเขียนคำแนะนำเชิงพัฒนาให้กันคนละ 2 ข้อ กิจกรรมนี้ทำให้ผู้เรียนเข้าใจ “เกณฑ์คุณภาพ” ลึกซึ้งจากการสวมบทบาทผู้ประเมิน และได้ฝึกให้ข้อมูลย้อนกลับอย่างสุภาพ

เมื่อผลการประเมินถูกนำไปใช้ ครูต้องยึดจรรยาบรรณอย่างเคร่งครัด ได้แก่ การให้ข้อมูลย้อนกลับที่ “งานและกระบวนการ” ไม่ใช่ที่ “ตัวบุคคล” การรักษาความลับของคะแนนโดยไม่ประจานผู้เรียนต่อหน้าชั้น ความเที่ยงธรรมและการไม่ลำเอียง และการใช้ผลเพื่อพัฒนามิใช่เพื่อลงโทษ (สมชาย วรภิเษมสกุล, 2559) ตัวอย่างถ้อยคำที่ดีและที่ควรเลี่ยงแสดงไว้ในตารางที่ 11.5

ตารางที่ 11.5 ตัวอย่างถ้อยคำข้อมูลย้อนกลับ: ที่ควรเลี่ยง vs ที่ดี

| สถานการณ์      | ถ้อยคำที่ควรเลี่ยง              | ถ้อยคำที่ดี (เจาะจง + ชี้ทางพัฒนา)                                                        |
|----------------|---------------------------------|-------------------------------------------------------------------------------------------|
| ตรวจงานเขียน   | “ยังไม่ดี ไปแก้มาใหม่”          | “ย่อหน้าสรุปยังไม่เชื่อมกับวัตถุประสงค์<br>ลองเพิ่มประโยคโยงกลับไปข้อ 1”                  |
| ตอบผิดหน้าชั้น | “ผิด นั่งลงได้”                 | “ใกล้แล้ว<br>ลองทบทวนสูตรพื้นที่สามเหลี่ยมอีกครั้งแล้วตอบใหม่”                            |
| ชมผลงานดี      | “เก่งมาก ฉลาดจัง” (ที่ตัวบุคคล) | “วิธีที่หนูแยกโจทย์เป็นขั้น ๆ ช่วยให้คำตอบถูก<br>ลองใช้วิธีนี้กับข้อต่อไป” (ที่กระบวนการ) |

ที่มา: ปรับจาก Hattie และ Timperley (2007) และ พินดา วราสุนันท์ (2558)

**ข้อควรระวัง:** การประเมินโดยเพื่อนจะได้ผลก็ต่อเมื่อมีการฝึกฝนและมีเกณฑ์ที่ชัดเจน (พินดา วราสุนันท์, 2558) ครูต้องกำกับให้คำวิจารณ์สุภาพและสร้างสรรค์ ป้องกันการให้คะแนนตามความสัมพันธ์ส่วนตัว หรืออคติ และควรใช้ผลจากเพื่อนเป็นข้อมูลประกอบการพัฒนา ไม่ใช่ตัวตัดสินผลการเรียนเพียงอย่างเดียว

**กรณีศึกษา (สาขาภาษาไทย):** การให้ข้อมูลย้อนกลับเพื่อแก้ปัญหาการเขียนเชิงวิชาการ

อังศุมลีย์ ทวีเหลือ และคณะ (2562) ทำวิจัยปฏิบัติการในชั้นเรียนวิชาภาษาไทย ชั้นมัธยมศึกษาปีที่ 1 ซึ่งมีนักเรียน 4 คนจาก 38 คน ไม่ผ่านเกณฑ์การประเมินทักษะการเขียนเชิงวิชาการที่ร้อยละ 70 ผู้วิจัยใช้กระบวนการสืบสอบร่วมกับเทคนิคการให้ข้อมูลย้อนกลับ (นวัตกรรมผ่านการตรวจสอบจากผู้เชี่ยวชาญ ด้วยค่าความสอดคล้อง 0.86) ให้ผู้เรียนแก้ไขงานเขียนเป็นรอบ ผลปรากฏว่านักเรียนทั้ง 4 คนผ่านเกณฑ์ที่กำหนดทุกคน

กรณีนี้เป็นวงจรการประเมินเพื่อพัฒนา (AfL) ครบวงจรตามโมเดลของ Hattie และ Timperley กล่าวคือ ระบุเป้าหมาย ให้ข้อมูลย้อนกลับเฉพาะจุดที่งานยังไม่ถึงเกณฑ์ และเปิดโอกาสให้แก้ไขจนสำเร็จ มิใช่การคืนคะแนนโดยไม่มีคำอธิบาย

คำถามชวนคิด: จงเขียนข้อมูลย้อนกลับ (feed-back และ feed-forward) ที่ท่านจะให้แก่ผู้เรียนหนึ่งคนทำงานเขียนยังไม่ผ่านเกณฑ์ โดยให้ผู้เรียนแก้ไขได้เองโดยไม่ต้องพึ่งครูตลอดเวลา

## 11.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ใบงานที่ 11.1 เขียนข้อมูลย้อนกลับเชิงพัฒนาจากกรณีศึกษา (รายบุคคล)

คำชี้แจง: กำหนดชิ้นงานผู้เรียนพร้อมคะแนน/ข้อสังเกต ให้ผู้เรียนเขียนข้อมูลย้อนกลับตามโมเดล 3 คำถาม (Feed Up/Back/Forward) ที่เฉพาะเจาะจง อยู่ระดับงาน/กระบวนการ และสร้างกำลังใจ (ประเมินด้วยรูปรีดในหัวข้อ 11.10) — เชื่อมโยง CLO6

### ใบงานที่ 11.2 วางแผนใช้ผลการประเมินพัฒนาผู้เรียน (รายกลุ่ม)

คำชี้แจง: จากกรณีผลการประเมินของชั้นเรียนสมมุติ (มีทั้งกลุ่มอ่อน-เก่ง) ให้กลุ่มวิเคราะห์จุดบกพร่องวางแผนปรับการสอน การสอนซ่อมเสริม และการส่งเสริมผู้เรียนเก่ง พร้อมนำเสนอแนวทางพัฒนา — เชื่อมโยง CLO6 การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 11.7 สรุปบทเรียน

การประเมินมีคุณค่าสูงสุดเมื่อนำผลไปใช้พัฒนาผู้เรียน ไม่ใช่เพียงตัดสิน แนวคิดการประเมินเพื่อการเรียนรู้ (AfL) ร่วมกับการประเมินในฐานะการเรียนรู้ (AaL) และการประเมินผลการเรียนรู้ (AoL) ต้องสมดุลกัน

โดย AfL/AaL มีพลังต่อการ เรียนรู้มากที่สุด งานของ Black และ Wiliam (1998) ยืนยันว่าการประเมินเพื่อพัฒนาที่มีคุณภาพย่อมสัมฤทธิ์ได้ หัวใจคือ การให้ข้อมูลย้อนกลับที่มีคุณภาพ ซึ่ง Hattie และ Timperley (2007) เสนอให้ตอบ 3 คำถาม (Feed Up/Back/Forward) และให้ในระดับงาน/กระบวนการ/การกำกับตนเอง มากกว่าระดับตัวบุคคล ครูต้องนำผลประเมินไปวินิจฉัย ปรับการสอน จัดสอนซ่อมเสริมและส่งเสริมผู้เรียนเก่ง พร้อมเปลี่ยนบทบาทเป็นผู้สนับสนุนการเรียนรู้และส่งเสริมให้ผู้เรียนกำกับตนเอง ทั้งนี้ต้องยึดจรรยาบรรณ ให้ข้อมูลย้อนกลับเชิงสร้างสรรค์ที่ตัวงาน รักษาศักดิ์ศรีผู้เรียน และใช้ผลเพื่อพัฒนาอย่างเป็นธรรม

## 11.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 11.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

1. ในสถานการณ์การสอนภาษาอังกฤษแบบ Active Learning หนึ่งหน่วยการเรียนรู้ จงออกแบบว่าจะใช้การประเมินแบบ AfL, AaL และ AoL ในจังหวะใดบ้าง ใช้เครื่องมือใด และเหตุใดการจัดจังหวะเช่นนั้นจึงช่วยพัฒนาผู้เรียนได้มากกว่าการสอบปลายหน่วยเพียงครั้งเดียว
2. สรุปผลการวิจัยของ Black และ Wiliam (1998) เกี่ยวกับการประเมินเพื่อพัฒนา
3. อธิบายโมเดลการให้ข้อมูลย้อนกลับของ Hattie และ Timperley (3 คำถาม + ระดับ)
4. เขียนตัวอย่างข้อมูลย้อนกลับที่ดี 1 ตัวอย่าง และที่ควรเลี่ยง 1 ตัวอย่าง พร้อมเหตุผล
5. อธิบายแนวทางการนำผลประเมินไปสอนซ่อมเสริมและส่งเสริมผู้เรียนเก่ง
6. เหตุใดจรรยาบรรณในการให้ข้อมูลย้อนกลับจึงสำคัญต่อการพัฒนาผู้เรียน อภิปราย

## 11.10 เครื่องมือวัดและประเมินผลประจำบท

### ก) RUBRIC การให้ข้อมูลย้อนกลับ ใบงาน 11.1 (Analytic) — เชื่อมโยง CLO6

| ประเด็น                          | ดีมาก (4)                 | ดี (3)      | พอใช้ (2)       | ปรับปรุง (1)    |
|----------------------------------|---------------------------|-------------|-----------------|-----------------|
| ครบ 3 คำถาม<br>(Up/Back/Forward) | ครบและชัดทุกคำถาม         | ครบส่วนใหญ่ | ขาดบางคำถาม     | ไม่ครบ          |
| อยู่ระดับงาน/กระบวนการ           | ชี้ที่งาน/กระบวนการชัดเจน | ส่วนใหญ่    | ปนระดับตัวบุคคล | ต่ำเป็นตัวบุคคล |
| ความเฉพาะเจาะจง/นำไปใช้ได้       | เฉพาะเจาะจง ปฏิบัติได้    | ค่อนข้างชัด | กว้างเกินไป     | คลุมเครือ       |
| สร้างกำลังใจเป็นธรรมชาติ         | สร้างสรรค์และให้กำลังใจ   | ค่อนข้างดี  | เฉย ๆ           | บั่นทอน         |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 11$  จาก 16 (ร้อยละ 70 ขึ้นไป)

### ข) RUBRIC การนำเสนอแนวทางพัฒนา ใบงาน 11.2 (โดยย่อ) — เชื่อมโยง CLO6

พิจารณา: วิเคราะห์จุดบกพร่องถูกต้อง • แผนปรับการสอน/ซ่อมเสริมเหมาะสมและสร้างสรรค์ • ส่งเสริมผู้เรียนเก่ง • นำเสนอชัดเจน (เกณฑ์ผ่าน ร้อยละ 70)

## 11.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการศึกษา. อุดรธานี: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จรูญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้ (พิมพ์ครั้งที่ 11). กรุงเทพฯ:

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

พินดา วราสุนันท์. (2558). การพัฒนาผู้เรียนด้วยการให้ข้อมูลย้อนกลับโดยเพื่อน.

วารสารมหาวิทยาลัยราชภัฏมหาสารคาม, 9(1), 33–40.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6). กรุงเทพฯ:

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมชาย วรภิเกษมสกุล. (2559). การวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 7). อุดรธานี: โรงพิมพ์อักษรศิลป์.

สมาพร มณีอ่อน. (2559). การประเมินเพื่อการเรียนรู้. วารสารศึกษาศาสตร์ มหาวิทยาลัยศิลปากร, 14(1), 15–25.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2553).

แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.

กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

อนวัติ คุณแก้ว. (2566). การทดสอบ การวัดผลและประเมินผลการศึกษาแนวใหม่.

สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

อังศุมาลี ทวีเหลื่อ, ทศนีย์ เศรษฐพงษ์ และวิภาวรรณ เอกวรรณัง. (2562).

การแก้ปัญหาการเขียนเชิงวิชาการโดยใช้กระบวนการสืบสอบร่วมกับเทคนิคการให้ข้อมูลย้อนกลับของนักเรียนชั้นมัธยมศึกษาปีที่ 1/7 โรงเรียนโยธินบูรณะ. วารสารวิชาการ ครุศาสตร์สวนสุนันทา, 3(2), 1–10.

#### ภาษาอังกฤษ

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment.

Phi Delta Kappan, 80(2), 139–148.

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. Educational Psychology Review, 32(2), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>

Nitko, A. J., & Brookhart, S. M. (2019). Educational assessment of students (8th ed.). New York: Pearson.

Wiliam, D. (2011). Embedded formative assessment. Bloomington, IN: Solution Tree Press.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>

## บทที่ 12

### สถิติพื้นฐานเพื่อการวัดและประเมินผล

#### *Basic Statistics for Measurement and Evaluation*

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 13 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

ผลการวัดทางการศึกษาออกมาเป็นตัวเลข จึงต้องใช้สถิติช่วยสรุปและแปลความหมาย บทที่ 12 พื้นฐานสถิติเชิงบรรยาย ที่ครูทุกคนต้องใช้ ได้แก่ การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง (ค่าเฉลี่ย มัชยฐาน ฐานนิยม) การวัดการกระจาย (พิสัย ส่วนเบี่ยงเบนมาตรฐาน) และการวัดตำแหน่ง/คะแนนมาตรฐาน (เปอร์เซ็นต์ไทล์ คะแนน  $z$  คะแนน  $T$ ) พร้อมฝึกคำนวณจากชุดคะแนนจริง เพื่อเป็นพื้นฐานสู่การตัดสินผลและรายงานผลในบทถัดไป

#### แผนบริหารการสอนประจำบทที่ 12

| รายการ                | รายละเอียด                                                                                                                                                                                                                                                |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) ความหมายและบทบาทของสถิติในการวัดและประเมินผล 2) การแจกแจงความถี่ 3) การวัดแนวโน้มเข้าสู่ส่วนกลาง (ค่าเฉลี่ย/มัชยฐาน/ฐานนิยม) 4) การวัดการกระจาย (พิสัย/ส่วนเบี่ยงเบนมาตรฐาน) 5) การวัดตำแหน่งและคะแนนมาตรฐาน (เปอร์เซ็นต์ไทล์/ $z$ / $T$ ) และโค้งปกติ |
| ความคิดรวบยอด         | สถิติเชิงบรรยายช่วยสรุปและแปลความหมายคะแนน ครูต้องเลือกสถิติให้เหมาะกับระดับการวัดและจุดมุ่งหมาย ค่ากลางบอกตำแหน่งศูนย์กลาง การกระจายบอกความแตกต่าง และคะแนนมาตรฐานช่วยเปรียบเทียบข้ามวิชา                                                                |
| จุดประสงค์การเรียนรู้ | คำนวณและแปลความหมายค่าสถิติพื้นฐาน (ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนมาตรฐาน) เพื่อการวัดและ ประเมินผลได้ (CLO4)                                                                                                                                       |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 12.1–12.5 ของบท)                                                                                                                                                                                               |

|                       |                                                                                                                                                                                                                                                    |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| กิจกรรมการเรียนการสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + สาธิตการคำนวณ 3) ใบงาน 12.1 คำนวณค่ากลางและการกระจายจากชุดคะแนน 4) ใบงาน 12.2 แปลงคะแนนดิบเป็นคะแนนมาตรฐาน z/T 5) อภิปรายการเลือกสถิติให้เหมาะกับข้อมูล 6) สรุป + แบบทดสอบหลังเรียน + แบบฝึกหัด |
| สื่อการเรียนการสอน    | เอกสารประกอบการสอนบทที่ 12 • สไลด์นำเสนอ • ตัวอย่างการคำนวณ • เครื่องคิดเลข/สเปรดชีต • ใบงาน 12.1-12.2 • ตำราที่อ้างอิง                                                                                                                            |
| การวัดและประเมินผล    | แบบทดสอบก่อน-หลังเรียน • ใบงาน 12.1-12.2 (เฉลยการคำนวณ) • การอภิปราย • แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70)                                                                                                                                       |

## 12.1 ความคิดรวบยอด

เนื่องจากการวัดทางการศึกษาให้ผลเป็นตัวเลข ครูจึงต้องใช้สถิติเชิงบรรยาย (Descriptive Statistics) เพื่อสรุปและ แปลความหมายคะแนน (ชูศรี วงศ์รัตน์, 2552) สถิติพื้นฐานที่จำเป็นมี 3 กลุ่ม ได้แก่ การวัดแนวโน้มเข้าสู่ส่วนกลาง (ค่าเฉลี่ย มัธยฐาน ฐานนิยม) ที่บอกตำแหน่งศูนย์กลางของข้อมูล การวัดการกระจาย (พิสัย ส่วนเบี่ยงเบนมาตรฐาน) ที่บอกความแตกต่างของข้อมูล และการวัดตำแหน่ง/คะแนนมาตรฐาน (เปอร์เซ็นต์ไทล์ คะแนน z คะแนน T) ที่ช่วยระบุตำแหน่ง ของผู้เรียนและเปรียบเทียบข้ามวิชาได้ การเลือกสถิติต้อง สอดคล้องกับระดับการวัดของข้อมูลและจุดมุ่งหมาย (พิชิต ฤทธิ์จรรยา, 2564; ศิริชัย กาญจนวาสี, 2552)

## 12.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 12 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                  | เชื่อมโยง CLO   |
|--------------------------------------------------------------|-----------------|
| 1) อธิบายบทบาทของสถิติและเลือกสถิติให้เหมาะกับระดับการวัดได้ | CLO4 (เข้าใจ)   |
| 2) จัดทำการแจกแจงความถี่จากชุดคะแนนได้                       | CLO4 (ประยุกต์) |
| 3) คำนวณค่าเฉลี่ย มัธยฐาน และฐานนิยม พร้อมแปลผลได้           | CLO4 (ประยุกต์) |
| 4) คำนวณพิสัยและส่วนเบี่ยงเบนมาตรฐาน พร้อมแปลผลได้           | CLO4 (ประยุกต์) |
| 5) แปลงคะแนนดิบเป็นเปอร์เซ็นต์ไทล์และคะแนนมาตรฐาน (z, T) ได้ | CLO4 (ประยุกต์) |

## 12.3 คำสำคัญ

สถิติเชิงบรรยาย (Descriptive Statistics) • การแจกแจงความถี่ (Frequency Distribution) • ค่าเฉลี่ย (Mean) • มัธยฐาน (Median) • ฐานนิยม (Mode) • พิสัย (Range) • ส่วนเบี่ยงเบนมาตรฐาน (Standard

Deviation) • เปอร์เซ็นไทล์ (Percentile) • คะแนนมาตรฐาน  $z$  และ  $T$  (Standard Scores) • โค้งปกติ (Normal Curve)

## 12.4 แบบทดสอบก่อนเรียน

---

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัสนักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าลิงก์ที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

## 12.5 เนื้อหาวิชา

---

### 12.5.1 ความหมาย บทบาทของสถิติ และระดับของการวัด

คำว่า สถิติ (Statistics) มีความหมายสองนัย นัยแรกคือ “ข้อมูลสถิติ” (statistical data) ซึ่งหมายถึงตัวเลขที่แสดงข้อเท็จจริงของสิ่งต่าง ๆ เช่น จำนวนนักเรียนที่สอบผ่านในแต่ละห้อง นัยที่สองคือ “สถิติศาสตร์” อันเป็นศาสตร์ที่ว่าด้วยระเบียบวิธีเก็บรวบรวม จัดระบบ นำเสนอ วิเคราะห์ และแปลความหมายข้อมูล (ชูศรีวงศ์รัตน์, 2552; พชรินทร์ ชมภูวิเศษ, 2559) สำหรับครู สถิติทำหน้าที่เป็น “ภาษา” ที่เปลี่ยนคะแนนดิบ (raw score) ที่ยังกระจัดกระจายให้กลายเป็นสารสนเทศ (information) ที่อ่านความหมายได้ เช่น เปลี่ยนคะแนนสอบของนักเรียนทั้งห้องให้เป็นข้อสรุปว่าโดยรวมเก่งหรืออ่อนเพียงใด และมีความแตกต่างกันมากน้อยเพียงใด มโนทัศน์พื้นฐานของมาตรการวัดทั้งสี่ระดับ ทบทวนได้ในบทที่ 2 (หัวข้อ 2.5.2)

เมื่อจำแนกตามหน้าที่ สถิติแบ่งเป็นสองประเภทใหญ่ ได้แก่ สถิติเชิงบรรยายหรือเชิงพรรณนา (descriptive statistics) ที่มุ่งสรุปและบรรยายลักษณะของกลุ่มข้อมูลที่เก็บมาได้จริง เช่น การแจกแจงความถี่ ค่ากลาง และการกระจาย กับสถิติเชิงอนุมานหรือเชิงอ้างอิง (inferential statistics) ที่ใช้ข้อมูลจากกลุ่มตัวอย่างสรุปอ้างอิงไปยังประชากรโดยอาศัยทฤษฎีความน่าจะเป็น (พชรินทร์ ชมภูวิเศษ, 2559) ในการวัดและประเมินผลชั้นเรียนตามปกติ ครูใช้สถิติเชิงบรรยายเป็นหลัก เพราะต้องการสรุปผลของ “นักเรียนกลุ่มที่สอนจริง” มิได้สุ่มตัวอย่างเพื่ออ้างอิงไปยังประชากร เนื้อหาในบทนี้จึงเน้นสถิติเชิงบรรยายเป็นสำคัญ

บทบาทของสถิติต่อการวัดและประเมินผลมีอย่างน้อยสามประการ ประการแรก ช่วยสรุปภาพรวมของชั้นเรียนได้อย่างกระชับ ประการที่สอง ช่วยให้การตัดสินใจมีความโปร่งใสและเป็นธรรม เพราะอ้างอิงตัวเลขที่ตรวจสอบได้แทนความรู้สึกส่วนตัว จึงลดอคติ (bias) ของผู้สอน และประการที่สาม ช่วยวินิจฉัยจุดเด่นจุดด้อยของผู้เรียนเพื่อวางแผนพัฒนาต่อ ความสามารถในการอ่าน ทำความเข้าใจ และสื่อสารสารสนเทศเชิงสถิติเช่นนี้เรียกว่า “การรู้สถิติ” (statistical literacy) ซึ่งเป็นสมรรถนะพื้นฐานที่ครูทุกคนพึงมี เพราะหากครูแปลผลสถิติผิด การตัดสินใจเกี่ยวกับผู้เรียนก็จะผิดตามไปด้วย (Ziegler & Garfield, 2018)

การจะเลือกใช้สถิติให้ถูกต้องต้องเริ่มจากการรู้ว่าข้อมูลที่มีอยู่ใน “ระดับการวัด” (level of measurement) ไต Stevens (1946) จำแนกระดับการวัดออกเป็น 4 ระดับเรียงจากหยาบไปละเอียด ได้แก่ ระดับนามบัญญัติ ระดับเรียงอันดับ ระดับอันตรภาค และระดับอัตราส่วน ซึ่งแต่ละระดับมีคุณสมบัติและสถิติที่เหมาะสมต่างกัน ดังสรุปในตารางที่ 12.1 ตารางที่ 12.1 ระดับการวัด 4 ระดับ คุณสมบัติ ตัวอย่าง และสถิติที่เหมาะสม

| ระดับการวัด              | คุณสมบัติสำคัญ                          | ตัวอย่างทางการศึกษา            | สถิติที่เหมาะสม                |
|--------------------------|-----------------------------------------|--------------------------------|--------------------------------|
| นามบัญญัติ<br>(Nominal)  | จำแนก/จัดกลุ่ม<br>นำมาบวกลบคูณหารไม่ได้ | เพศ รหัสนักเรียน ศาสนา         | ความถี่ ร้อยละ ฐานนิยม         |
| เรียงอันดับ<br>(Ordinal) | บอกลำดับสูง-ต่ำ แต่ระยะห่างไม่เท่ากัน   | เกรด (A, B, C) อันดับที่สอบได้ | มัธยฐาน เปอร์เซ็นไทล์          |
| อันตรภาค<br>(Interval)   | ช่วงเท่ากัน แต่ไม่มีศูนย์แท้            | คะแนนสอบ คะแนนเจตคติ           | ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน |
| อัตราส่วน<br>(Ratio)     | ช่วงเท่ากัน และมีศูนย์แท้               | เวลาทำข้อสอบ น้ำหนัก ส่วนสูง   | ใช้ได้ทุกประเภทสถิติ           |

ที่มา: ดัดแปลงจาก พัชรินทร์ ขมภูวิเศษ (2559, น. 242); Stevens (1946)

ดังตารางที่ 12.1 คะแนนสอบในโรงเรียนส่วนใหญ่จัดเป็นข้อมูลระดับอันตรภาค เพราะมีหน่วยเท่ากัน แต่ไม่มี “ศูนย์แท้” (absolute zero) กล่าวคือ นักเรียนที่ได้ 0 คะแนนในวิชาคณิตศาสตร์ มิได้แปลว่าไม่มีความรู้คณิตศาสตร์เลย และผู้ที่ได้ 80 คะแนนก็มีได้เก่งเป็น “สองเท่า” ของผู้ได้ 40 คะแนน (พัชรินทร์ ขมภูวิเศษ, 2559) ความเข้าใจข้อนี้สำคัญมาก เพราะเป็นเหตุผลที่ทำให้เราต้องแปลงคะแนนดิบเป็นคะแนนมาตรฐานก่อนนำไปเปรียบเทียบข้ามวิชา ข้อผิดพลาดที่พบบ่อยในชั้นเรียนคือการนำข้อมูลระดับนามบัญญัติ เช่น เลขประจำห้องหรือรหัสเพศ มาหาค่าเฉลี่ย ซึ่งไม่มีความหมายใด ๆ ครูจึงต้องตรวจสอบระดับการวัดของข้อมูลก่อนเลือกสถิติทุกครั้ง

### กล่องที่ 12.1 ศัพท์พื้นฐานทางสถิติที่ควรรู้

ประชากร (population): สมาชิกทั้งหมดที่ต้องการศึกษา

กลุ่มตัวอย่าง (sample): ส่วนหนึ่งของประชากรที่เป็นตัวแทน

พารามิเตอร์ (parameter): ค่าที่บรรยายลักษณะประชากร ใช้สัญลักษณ์กรีก เช่น  $\mu$  (มิว),  $\sigma$  (ซิกมา)

ค่าสถิติ (statistic): ค่าที่คำนวณจากกลุ่มตัวอย่าง เช่น  $\bar{X}$  (เอกซ์บาร์), S.D.

ข้อมูลดิบ (raw data / raw score): ข้อมูลที่เก็บมาแต่ยังไม่จัดระเบียบหรือวิเคราะห์

### 12.5.2 การแจกแจงความถี่ (Frequency Distribution)

การแจกแจงความถี่ (frequency distribution) คือการนำข้อมูลดิบมาจัดเป็นหมวดหมู่ในรูปตารางที่แสดงค่าของข้อมูลคู่กับจำนวนครั้งที่ข้อมูลนั้นเกิดขึ้น (ความถี่) เพื่อให้ข้อมูลเป็นระเบียบ เห็นภาพรวมและรูปร่างการกระจายได้ง่ายขึ้น (สมชาย วรกิจเกษมสกุล, 2559; บุญชม ศรีสะอาด, 2560) การแจกแจงความถี่จึงเป็นขั้นแรกของการ “อ่าน” ข้อมูลก่อนคำนวณค่าสถิติอื่น ๆ

การแจกแจงความถี่แบ่งเป็นสองแบบ แบบแรกคือการแจกแจงความถี่ที่ไม่จัดหมวดหมู่ (ungrouped data) ซึ่งแสดงคะแนนทุกค่าพร้อมความถี่ เหมาะกับข้อมูลจำนวนน้อยและค่าคะแนนไม่กระจายมาก แบบที่สองคือการแจกแจงความถี่แบบจัดหมวดหมู่เป็นอันตรภาคชั้น (grouped data) ซึ่งยุบคะแนนเป็นช่วง ๆ เหมาะกับข้อมูลจำนวนมาก เช่น คะแนนของนักเรียนทั้งระดับชั้น ขั้นตอนการสร้างตารางแบบจัดหมวดหมู่สรุปไว้ในกล่องที่ 12.2 (พัชรินทร์ ชมภูวิเศษ, 2559)

### กล่องที่ 12.2 ขั้นตอนสร้างตารางแจกแจงความถี่แบบจัดหมวดหมู่

1) หาพิสัย (range) = คะแนนสูงสุด - คะแนนต่ำสุด

2) กำหนดจำนวนชั้น (นิยม 5–15 ชั้น ตามจำนวนข้อมูล)

3) หาความกว้างของอันตรภาคชั้น = พิสัย ÷ จำนวนชั้น (ปัดขึ้น)

4) กำหนดช่วงคะแนนของแต่ละชั้นให้ครอบคลุมข้อมูล ไม่คาบเกี่ยวกัน

5) ชีตรอยขีด (tally) แทนข้อมูลแต่ละตัว แล้วนับเป็นความถี่

ตัวอย่าง: ข้อมูล 32 จำนวน พิสัย =  $40 - 23 = 17$  ใช้ 6 ชั้น ความกว้างชั้น =  $17 \div 6 \approx 3$

ตัวอย่างเช่น เมื่อนำคะแนนสอบของนักศึกษา 30 คนมาแจกแจงความถี่แบบไม่จัดหมวดหมู่ จะได้ผลดังตารางที่ 12.2 ซึ่งช่วยให้เห็นทันทีว่าคะแนนกระจุกตัวอยู่ที่ค่าใด (คะแนน 15 และ 17 มีผู้ได้มากที่สุดคือ 6 คน) และมีผู้ได้คะแนนสูงสุดและต่ำสุดกี่คน

ตารางที่ 12.2 ตัวอย่างตารางแจกแจงความถี่คะแนนสอบของนักศึกษา 30 คน

| คะแนนที่ได้ | รอยขีด (tally) | ความถี่ (f) |
|-------------|----------------|-------------|
| 13          | //             | 2           |
| 14          | ///            | 3           |
| 15          | //// /         | 6           |
| 16          | ////           | 5           |
| 17          | //// /         | 6           |
| 18          | ////           | 5           |
| 19          | ///            | 3           |
| รวม         |                | 30          |

ที่มา: พิชรินทร์ ชมภูวิเศษ (2559, น. 244)

นอกจากตาราง ครูยังนำเสนอการแจกแจงความถี่ด้วยกราฟ เช่น ฮิสโตแกรม (histogram) รูปหลายเหลี่ยมความถี่ (frequency polygon) หรือเส้นโค้งความถี่ ซึ่งช่วยให้เห็น “รูปร่าง” ของการแจกแจงได้รวดเร็วว่าเป็นโค้งปกติ เบ้ซ้าย หรือเบ้ขวา ในทางปฏิบัติ หากครูพล็อตฮิสโตแกรมคะแนนแล้วพบว่าคะแนนเบ้ขวา (นักเรียนส่วนใหญ่ได้คะแนนต่ำ มีน้อยคนได้คะแนนสูง) ก็เป็นสัญญาณว่าข้อสอบอาจยากเกินไปหรือเนื้อหายังไม่ได้รับการเรียนรู้อย่างทั่วถึง ข้อควรระวังคือควรกำหนดจำนวนชั้นให้พอเหมาะ เพราะชั้นมากเกินไปทำให้ตารางกระจายจนมองไม่เห็นแนวโน้ม ส่วนชั้นน้อยเกินไปทำให้สูญเสียรายละเอียด และทุกชั้นต้องกว้างเท่ากันเพื่อไม่ให้เกิดความคลาดเคลื่อน

### 12.5.3 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

การวัดแนวโน้มเข้าสู่ส่วนกลาง (measure of central tendency) คือการหา “ค่ากลาง” ที่เป็นตัวแทนของข้อมูลทั้งชุด เพื่อสรุปภาพรวมความสามารถของกลุ่มด้วยตัวเลขเพียงค่าเดียว ค่ากลางที่ใช้บ่อยมี 3 ค่า ได้แก่ ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม (สมชาย วรกิจเกษมสกุล, 2559; บุญชม ศรีสะอาด, 2560) การเลือกใช้ค่ากลางให้ถูกต้องขึ้นอยู่กับระดับการวัดและรูปร่างการแจกแจงของข้อมูล

ค่าเฉลี่ยเลขคณิต (arithmetic mean, แทนด้วย  $\bar{X}$ ) คือผลรวมของคะแนนทุกตัวหารด้วยจำนวนข้อมูล ( $\bar{X} = \sum X \div N$ ) เป็นค่ากลางที่ใช้ข้อมูลทุกตัวจึงมีความละเอียดสูง แต่ก็ไวต่อค่าผิดปกติ (outlier) มัธยฐาน (median, Mdn) คือค่าที่อยู่ตำแหน่งกึ่งกลางเมื่อเรียงข้อมูลจากน้อยไปมาก หาตำแหน่งได้จาก  $(N+1) \div 2$  ส่วนฐานนิยม (mode, Mo) คือค่าที่มีความถี่สูงสุด คุณสมบัติและความเหมาะสมของค่ากลางทั้งสามสรุปไว้ในตารางที่ 12.3 ตารางที่ 12.3 เปรียบเทียบค่ากลาง 3 ชนิดและความเหมาะสมกับลักษณะข้อมูล

| ค่ากลาง             | วิธีหา / สูตร                                                       | เหมาะกับข้อมูล                              |
|---------------------|---------------------------------------------------------------------|---------------------------------------------|
| ค่าเฉลี่ย<br>(Mean) | นำคะแนนทุกตัวรวมกันแล้วหารด้วยจำนวน ( $\bar{X} = \Sigma X \div N$ ) | ระดับอันดับภาค/อัตราส่วนที่ไม่มีค่าผิดปกติ  |
| มัธยฐาน<br>(Median) | ค่าตำแหน่งกึ่งกลางเมื่อเรียงข้อมูล ตำแหน่ง $(N+1) \div 2$           | ข้อมูลเบ้ มีค่าผิดปกติ หรือระดับเรียงอันดับ |
| ฐานนิยม<br>(Mode)   | ค่าที่มีความถี่สูงสุดในชุดข้อมูล                                    | ข้อมูลระดับนามบัญญัติ หรือค่าที่พบบ่อย      |

ที่มา: ดัดแปลงจาก สมชาย วรภิเกษมสกุล (2559, น. 196); พัทรินทร์ ชมภูวิเศษ (2559)

### กล่องที่ 12.3 ตัวอย่างการคำนวณค่ากลาง (คะแนนสอบ 10 คน)

ชุดคะแนน (เรียงแล้ว): 15, 18, 20, 20, 22, 25, 28, 30, 32, 40

ค่าเฉลี่ย:  $\bar{X} = (15+18+20+20+22+25+28+30+32+40) \div 10 = 250 \div 10 = 25.0$

มัธยฐาน: ตำแหน่ง  $(10+1) \div 2 = 5.5 \rightarrow$  เฉลี่ยค่าตำแหน่งที่ 5 และ 6  $= (22+25) \div 2 = 23.5$

ฐานนิยม: ค่าที่ซ้ำมากที่สุด = 20

ข้อสังเกต: ค่าเฉลี่ย (25.0) สูงกว่ามัธยฐาน (23.5) เพราะคะแนน 40 ดึงค่าเฉลี่ยขึ้น

ในการเลือกใช้ค่ากลาง ครูควรพิจารณารูปร่างการแจกแจง หากข้อมูลมีการแจกแจงใกล้เคียงโค้งปกติ ค่าเฉลี่ย มัธยฐาน และฐานนิยมจะมีค่าใกล้เคียงกัน จึงนิยมใช้ค่าเฉลี่ยเพราะเสถียรและละเอียดที่สุด แต่ถ้าข้อมูลเบ้ (skewed) หรือมีค่าผิดปกติ เช่น ห้องหนึ่งมีนักเรียนคนหนึ่งขาดสอบจึงได้ 0 คะแนน ค่าเฉลี่ยจะถูกดึงให้ต่ำลงจนไม่เป็นตัวแทนที่ดี กรณีเช่นนี้ควรใช้มัธยฐานซึ่งทนทานต่อค่าผิดปกติมากกว่า (Nitko & Brookhart, 2019) ส่วนข้อมูลเชิงคุณภาพ เช่น อาชีพของผู้ปกครองหรือสีที่นักเรียนชอบ ใช้ได้เพียงฐานนิยมเท่านั้น ทั้งนี้เมื่อข้อมูลเบ้หาความสัมพันธ์จะเป็น ฐานนิยม < มัธยฐาน < ค่าเฉลี่ย และเบ้ซ้ายจะกลับกัน (พัชรินทร์ ชมภูวิเศษ, 2559)

#### 12.5.4 การวัดการกระจาย (Measures of Dispersion)

การวัดการกระจาย (measure of dispersion หรือ variability) เป็นการบอกว่าข้อมูลแต่ละตัวแตกต่างกันหรือกระจายออกจากค่ากลางมากน้อยเพียงใด การรู้ค่ากลางเพียงอย่างเดียวยังไม่พอที่จะตัดสินคุณภาพของชั้นเรียน เพราะสองห้องที่มีค่าเฉลี่ยเท่ากันอาจมีความแตกต่างภายในห้องไม่เท่ากันเลย การวัดการกระจายจึงช่วยให้ครูเข้าใจ “ความแตกต่างระหว่างบุคคล” (individual differences) ของผู้เรียน (สมชาย วรภิเกษมสกุล, 2559; Gronlund & Linn, 1990)

ค่าที่ใช้วัดการกระจายมีหลายตัว ที่พบบ่อยได้แก่ พิสัย (range) ซึ่งเท่ากับคะแนนสูงสุดลบคะแนนต่ำสุด คำนวณง่ายแต่หยابเพราะใช้ข้อมูลเพียงสองค่า ความแปรปรวน (variance, แทนด้วย  $S^2$ ) ซึ่งเป็นค่าเฉลี่ยของ

ผลต่างกำลังสองระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน (standard deviation, แทนด้วย S.D. หรือ S) ซึ่งเป็นรากที่สองของความแปรปรวน ส่วนเบี่ยงเบนมาตรฐานเป็นค่าที่นิยมใช้มากที่สุดเพราะมีหน่วยเดียวกับคะแนนเดิมและสะท้อนการกระจายของข้อมูลทุกตัว สูตรและตัวอย่างการคำนวณแสดงไว้ในกล่องที่ 12.4

#### กล่องที่ 12.4 สูตรและตัวอย่างการคำนวณส่วนเบี่ยงเบนมาตรฐาน

พิสัย (Range) = คะแนนสูงสุด - คะแนนต่ำสุด

ความแปรปรวน:  $S^2 = \sum(X - \bar{X})^2 \div N$

ส่วนเบี่ยงเบนมาตรฐาน:  $S.D. = \sqrt{\sum(X - \bar{X})^2 \div N}$

ตัวอย่าง: คะแนน 6 คน = 2, 4, 6, 7, 8, 9  $\rightarrow \bar{X} = 36 \div 6 = 6$

ผลต่างกำลังสอง:  $(-4)^2 + (-2)^2 + 0^2 + 1^2 + 2^2 + 3^2 = 16 + 4 + 0 + 1 + 4 + 9 = 34$

$S^2 = 34 \div 6 = 5.67$  และ  $S.D. = \sqrt{5.67} \approx 2.38$

หมายเหตุ: เมื่อใช้ข้อมูลเป็นกลุ่มตัวอย่างเพื่อประมาณค่าประชากร ใช้ตัวหาร N-1 แทน N

การแปลผลส่วนเบี่ยงเบนมาตรฐานทำได้โดยเทียบกับค่าเฉลี่ย ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าน้อยแสดงว่าคะแนนเกาะกลุ่มอยู่ใกล้ค่าเฉลี่ย กลุ่มมีความเป็นเอกพันธ์ (homogeneous) คือเก่ง-อ่อนใกล้เคียงกัน แต่ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่ามาก แสดงว่าคะแนนกระจาย กลุ่มมีความหลากหลาย (heterogeneous) ความแตกต่างระหว่างบุคคลสูง ดังตัวอย่างเปรียบเทียบสองห้องในตารางที่ 12.4 ที่มีค่าเฉลี่ยเท่ากันแต่ส่วนเบี่ยงเบนมาตรฐานต่างกันตารางที่ 12.4 เปรียบเทียบสองห้องที่มีค่าเฉลี่ยเท่ากันแต่การกระจายต่างกัน

| รายการ                      | ห้อง ก                               | ห้อง ข                           |
|-----------------------------|--------------------------------------|----------------------------------|
| ค่าเฉลี่ย ( $\bar{X}$ )     | 25                                   | 25                               |
| ส่วนเบี่ยงเบนมาตรฐาน (S.D.) | 3                                    | 9                                |
| การแปลผล                    | คะแนนเกาะกลุ่ม เก่ง-อ่อนใกล้เคียงกัน | คะแนนกระจาย ต่างกันมาก           |
| แนวทางการสอน                | สอนรวมทั้งห้องได้                    | ควรสอนซ่อมเสริม/จัดกลุ่มตามระดับ |

ที่มา: ผู้เขียน (สังเคราะห์จากแนวคิดของ สมชาย วรกิจเกษมสกุล, 2559)

ในทางปฏิบัติ ครูสามารถใช้ส่วนเบี่ยงเบนมาตรฐานเป็นข้อมูลตัดสินใจเรื่องการจัดการเรียนการสอนได้ทันที เช่น ถ้าห้องเรียนมีส่วนเบี่ยงเบนมาตรฐานสูงผิดปกติ ก็ควรออกแบบการสอนที่ตอบสนองความแตกต่างระหว่างบุคคล (*differentiated instruction*) ควบคู่กับการสอนซ่อมเสริม ข้อควรระวังคือ พิสัยเป็นค่าหายาที่เปลี่ยนแปลงง่ายเมื่อมีค่าผิดปกติ จึงควรใช้เพียงประกอบการสร้างตารางแจกแจงความถี่หรือดูภาพรวมอย่างรวดเร็ว ไม่ควรใช้ตัดสินคุณภาพการกระจายแทนส่วน

เบี่ยงเบนมาตรฐาน และต้องระวังการสับสนระหว่างสูตรของประชากร (หารด้วย  $N$ ) กับสูตรของกลุ่มตัวอย่าง (หารด้วย  $N-1$ ) ซึ่งให้ค่าต่างกันเล็กน้อยโดยเฉพาะเมื่อข้อมูลมีจำนวนน้อย

#### 12.5.5 การวัดตำแหน่ง คะแนนมาตรฐาน และโค้งปกติ

คะแนนดิบ (raw score) เพียงอย่างเดียวยังบอกตำแหน่งของผู้เรียนในกลุ่มไม่ได้ เพราะความยากง่ายของข้อสอบแต่ละชุดไม่เท่ากัน การได้ 30 คะแนนจากข้อสอบที่ยากมากอาจหมายถึงผลงานที่ดีกว่าการได้ 30 คะแนนจากข้อสอบที่ง่าย การวัดตำแหน่ง (measures of position) จึงช่วยบอกว่าผู้เรียนอยู่ ณ ตำแหน่งใดเมื่อเทียบกับกลุ่ม ค่าที่ใช้บ่อยคือ เปอร์เซ็นไทล์ (percentile) และอันดับที่เปอร์เซ็นไทล์ (percentile rank, PR) เช่น PR70 หมายถึงมีผู้ได้คะแนนต่ำกว่าผู้นี้อยู่ร้อยละ 70 ของกลุ่ม การรายงานผลด้วย PR จึงสื่อสารกับผู้ปกครองได้ชัดเจนกว่าคะแนนดิบ (พัชรินทร์ ชมภูวิเศษ, 2559; พิชิต ฤทธิ์จรูญ, 2564)

เพื่อเปรียบเทียบผลข้ามวิชาหรือข้ามแบบทดสอบอย่างเป็นธรรม เราแปลงคะแนนดิบให้เป็นคะแนนมาตรฐาน (standard score) คะแนนมาตรฐานที่นิยมใช้คือคะแนนมาตรฐาน  $Z$  ซึ่งบอกว่าคะแนนดิบอยู่ห่างจากค่าเฉลี่ยกี่เท่าของส่วนเบี่ยงเบนมาตรฐาน คำนวณจาก  $Z = (X - \bar{X}) \div S.D.$  มีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 และมักมีค่าอยู่ระหว่าง  $-3$  ถึง  $+3$  เนื่องจากคะแนน  $Z$  อาจติดลบและมีทศนิยมจึงสื่อสารยาก จึงแปลงต่อเป็นคะแนนมาตรฐาน  $T$  ด้วยสูตร  $T = 50 + 10Z$  ซึ่งมีค่าเฉลี่ย 50 ส่วนเบี่ยงเบนมาตรฐาน 10 และมีค่าอยู่ระหว่างประมาณ 20–80 ทำให้ไม่มีค่าติดลบและเข้าใจง่ายขึ้น (พัชรินทร์ ชมภูวิเศษ, 2559; ศิริชัย กาญจนวาสี, 2552; Anastasi & Urbina, 1997)

##### กล่องที่ 12.5 ตัวอย่างการคำนวณและการใช้คะแนนมาตรฐาน

กรณีที่ 1: นาย ก สอบได้ 80 คะแนน ( $\bar{X} = 70$ ,  $S.D. = 5$ )

$$Z = (80 - 70) \div 5 = +2.0 \rightarrow T = 50 + 10(2.0) = 70$$

แปลผล: นาย ก อยู่สูงกว่าค่าเฉลี่ย 2 หน่วยส่วนเบี่ยงเบนมาตรฐาน (ราว PR98)

กรณีที่ 2: เปรียบเทียบข้ามวิชา — นาย ข ได้ คะแนนมาตรฐานมีความหมายชัดเจนเมื่ออ้างอิงกับ “โค้งปกติ” (normal curve) ซึ่งเป็นเส้นโค้งรูประฆังคว่ำที่สมมาตรรอบค่าเฉลี่ย ภายใต้โค้งปกติ คะแนนประมาณร้อยละ 68.26 จะอยู่ในช่วง  $\bar{X} \pm 1 S.D.$  ประมาณร้อยละ 95.44 อยู่ในช่วง  $\bar{X} \pm 2 S.D.$  และประมาณร้อยละ 99.74 อยู่ในช่วง  $\bar{X} \pm 3 S.D.$  ดังนั้นผู้ที่ได้  $Z = +1$  (หรือ  $T = 60$ ) จะอยู่สูงกว่าเพื่อนประมาณร้อยละ 84 ของกลุ่ม ความสัมพันธ์ระหว่างพื้นที่ใต้โค้ง คะแนน  $Z$  คะแนน  $T$  และอันดับที่เปอร์เซ็นไทล์ แสดงดังภาพที่ 12.1

15 คะแนนทั้งสองวิชา

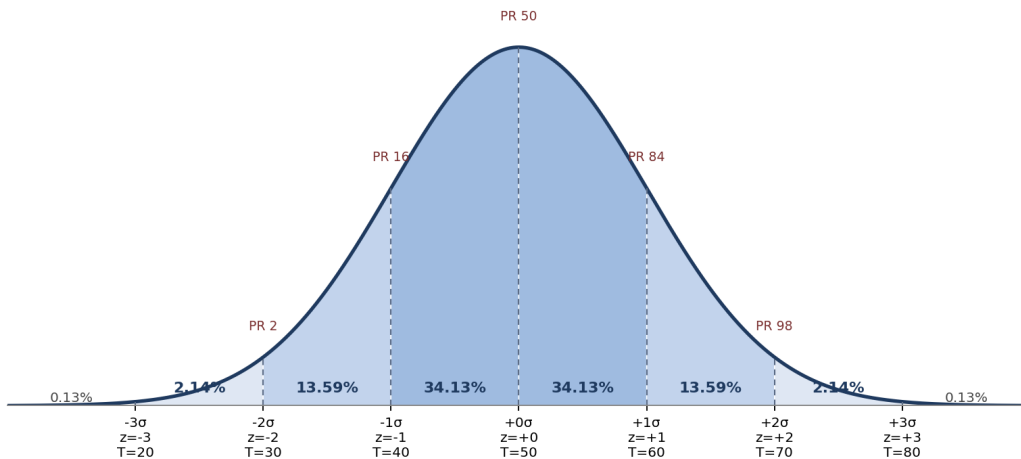
คณิตศาสตร์ ( $\bar{X} = 10$ ,  $S.D. = 2$ ):  $Z = (15 - 10) \div 2 = +2.5$

ภาษาไทย ( $\bar{X} = 14$ ,  $S.D. = 1$ ):  $Z = (15 - 14) \div 1 = +1.0$

สรุป: แม้คะแนนดิบเท่ากัน แต่ผลงานวิชาคณิตศาสตร์โดดเด่นกว่าอย่างชัดเจน

คะแนนมาตรฐานมีความหมายชัดเจนเมื่ออ้างอิงกับ “โค้งปกติ” (normal curve) ซึ่งเป็นเส้นโค้งรูประฆังคว่ำที่สมมาตรรอบค่าเฉลี่ย ภายใต้โค้งปกติ คะแนนประมาณร้อยละ 68.26 จะอยู่ในช่วง  $\bar{X} \pm 1$  S.D. ประมาณร้อยละ 95.44 อยู่ในช่วง  $\bar{X} \pm 2$  S.D. และประมาณร้อยละ 99.74 อยู่ในช่วง  $\bar{X} \pm 3$  S.D. ดังนั้นผู้ที่ได้  $Z = +1$  (หรือ  $T = 60$ ) จะอยู่สูงกว่าเพื่อนประมาณร้อยละ 84 ของกลุ่ม ความสัมพันธ์ระหว่างพื้นที่ใต้โค้งคะแนน  $z$  คะแนน  $T$  และอันดับที่เปอร์เซ็นต์ไทล์ แสดงดังภาพที่ 12.1

The Normal Curve: Areas, z-scores, and T-scores



ภาพที่ 12.1 โค้งปกติกับพื้นที่ใต้โค้ง คะแนน  $z$  คะแนน  $T$  และอันดับที่เปอร์เซ็นต์ไทล์

ที่มา: ผู้เขียน (อ้างอิงสัดส่วนพื้นที่ใต้โค้งปกติจาก พชรินทร์ ชมภูวิเศษ, 2559)

ตารางที่ 12.5 สรุปการเทียบตำแหน่งของผู้เรียนใต้โค้งปกติด้วยคะแนนมาตรฐานหลายแบบ ซึ่งช่วยให้ครูแปลความหมายตำแหน่งของผู้เรียนได้อย่างสอดคล้องกัน

ตารางที่ 12.5 การเทียบคะแนนมาตรฐานกับตำแหน่งใต้โค้งปกติ (โดยประมาณ)

| ตำแหน่งใต้โค้งปกติ  | $z$ | $T$ | Stanine | PR |
|---------------------|-----|-----|---------|----|
| ต่ำกว่าค่าเฉลี่ยมาก | -2  | 30  | 2       | 2  |
| ต่ำกว่าค่าเฉลี่ย    | -1  | 40  | 3       | 16 |
| ที่ค่าเฉลี่ย        | 0   | 50  | 5       | 50 |
| สูงกว่าค่าเฉลี่ย    | +1  | 60  | 7       | 84 |
| สูงกว่าค่าเฉลี่ยมาก | +2  | 70  | 8       | 98 |

ที่มา: ดัดแปลงจาก พชรินทร์ ชมภูวิเศษ (2559); Anastasi & Urbina (1997)

การประยุกต์ใช้คะแนนมาตรฐานที่พบมากในวงการศึกษไทยคือการสร้าง “เกณฑ์ปกติ” (norms) ของแบบทดสอบ ซึ่งมักรายงานในรูปคะแนนที่ปกติ (normalized T-score) เพื่อให้โรงเรียนต่าง ๆ ใช้แปลผล

ตำแหน่งของผู้เรียนได้บนมาตรฐานเดียวกัน ตัวอย่างเช่น งานพัฒนาแบบทดสอบวัดความเข้าใจในการอ่านภาษาอังกฤษสำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ที่สร้างเกณฑ์ปกติเป็นคะแนนที่ตั้งตั้งแต่ T13 ถึง T90 (นพเก้า ณ พัทลุง, 2558) ข้อควรระวังสำคัญคือ คะแนนมาตรฐานเปรียบเทียบกับกันได้เฉพาะภายในกลุ่มอ้างอิงเดียวกันเท่านั้น การนำคะแนน T ของนักเรียนคนละกลุ่มอ้างอิงมาเทียบกันโดยตรงจะทำให้สรุปผิด และเนื่องจากคะแนน Z ที่ติดลบอาจสร้างความเข้าใจผิดแก่ผู้ปกครอง ครูจึงควรเลือกใช้คะแนน T หรือ PR ในการรายงานผล

#### เชื่อมโยงกับผลลัพธ์การเรียนรู้ (CLO4)

เนื้อหา 12.5 ทั้งหมดรองรับ CLO4: ปฏิบัติการคำนวณค่าสถิติพื้นฐานและแปลผลเพื่อการตัดสินใจ

วัดด้วย: ใบงานที่ 12.1 (ค่ากลาง/การกระจาย) และใบงานที่ 12.2 (แปลงคะแนนมาตรฐาน)

ประเมินด้วย: RUBRIC การคำนวณและแปลผลสถิติ (หัวข้อ 12.10) เกณฑ์ผ่านร้อยละ 70

## 12.6 กิจกรรมการเรียนรู้การสอนและใบงาน

### ใบงานที่ 12.1 คำนวณค่ากลางและการกระจาย (รายบุคคล)

คำชี้แจง: กำหนดชุดคะแนนสอบของผู้เรียน 10 คน ให้คำนวณค่าเฉลี่ย มัธยฐาน ฐานนิยม พิสัย

และส่วนเบี่ยงเบนมาตรฐาน พร้อมแปลผลความหมายของค่าที่ได้ (มีเฉลยในหัวข้อ 12.10) — เชื่อมโยง CLO4

### ใบงานที่ 12.2 แปลงคะแนนดิบเป็นคะแนนมาตรฐาน (รายบุคคล)

คำชี้แจง: จากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่กำหนด ให้แปลงคะแนนดิบของผู้เรียนเป็นคะแนน z และ T

แล้วเปรียบเทียบ ตำแหน่งของผู้เรียนข้ามวิชา 2 วิชา พร้อมสรุปว่าผู้เรียนทำได้ดีกว่าในวิชาใด — เชื่อมโยง CLO4

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป

ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทร่วมกับผู้เรียน

ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน

เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 12.7 สรุปบทเรียน

ผลการวัดทางการศึกษาเป็นตัวเลข จึงต้องใช้สถิติเชิงบรรยายช่วยสรุปและแปลความหมาย สถิติพื้นฐานที่จำเป็นมี 3 กลุ่ม คือ การวัดแนวโน้มเข้าสู่ส่วนกลาง (ค่าเฉลี่ย มัธยฐาน ฐานนิยม) ที่บอกตำแหน่งศูนย์กลาง การวัดการกระจาย (พิสัย ส่วนเบี่ยงเบนมาตรฐาน) ที่บอกความแตกต่างของข้อมูล และการวัดตำแหน่ง/คะแนนมาตรฐาน (เปอร์เซ็นต์ไทล์ คะแนน  $z = (X - \bar{X}) / S.D.$  คะแนน  $T = 50 + 10z$ ) ที่ช่วยระบุตำแหน่งและเปรียบเทียบผลข้ามวิชาได้ การเลือกสถิติต้องสอดคล้องกับระดับ การวัด เมื่อข้อมูลกระจายปกติ ค่าเฉลี่ย มัธยฐาน และฐาน

นิยมจะใกล้เคียงกัน และคะแนนประมาณ 68% อยู่ในช่วง  $\bar{X} \pm 1$  S.D. ความเข้าใจสถิติพื้นฐานนี้เป็นรากฐานสำคัญสู่การตัดสินใจและการรายงานผลในบทถัดไป

## 12.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

## 12.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

- ฝ่ายวิชาการขอให้ครูเปรียบเทียบผลสอบกลางภาคของนักเรียนสามห้องที่ใช้ข้อสอบต่างชุดกัน จงวิเคราะห์ว่าการเทียบคะแนนดิบโดยตรงเหมาะสมหรือไม่ เพราะเหตุใด และควรใช้สถิติหรือคะแนนมาตรฐานใดจึงจะเปรียบเทียบได้อย่างเป็นธรรม
- คำนวณค่าเฉลี่ย มัธยฐาน และฐานนิยมของชุดคะแนน: 12, 15, 15, 18, 20
- คำนวณพิสัยและส่วนเบี่ยงเบนมาตรฐานของชุดคะแนน: 2, 4, 6, 7, 8, 9 พร้อมแปลผล
- อธิบายความต่างของคะแนน  $z$  กับคะแนน  $T$  และประโยชน์ของคะแนนมาตรฐาน
- นาย ข สอบได้ 80 คะแนน ( $\bar{X}=70$ ,  $S.D.=5$ ) คำนวณ  $z$  และ  $T$  พร้อมแปลผลตำแหน่ง
- อธิบายว่าเหตุใดมัธยฐานจึงเหมาะกับข้อมูลที่มีค่าผิดปกติมากกว่าค่าเฉลี่ย

## 12.10 เครื่องมือวัดและประเมินผลประจำบท

### ก) เฉลยแนวทางใบงาน 12.1–12.2 (สูตรหลัก)

$\bar{X} = \Sigma X/N$  • Median = ค่าตำแหน่ง  $(N+1)/2$  • Mode = ค่าที่ถี่สุด • Range = สูงสุด–ต่ำสุด • S.D. =  $\sqrt{[\Sigma(X-\bar{X})^2/(n-1)]}$  •  $z = (X-\bar{X})/S.D.$  •  $T = 50+10z$  (เช่น  $X=80$ ,  $\bar{X}=70$ ,  $S.D.=5 \rightarrow z=2$ ,  $T=70$ )

### ข) รุบริคประเมินการคำนวณและแปลผลสถิติ (Analytic) — เชื่อมโยง CLO4

| ประเด็น                     | ดีมาก (4)          | ดี (3)          | พอใช้ (2)      | ปรับปรุง (1) |
|-----------------------------|--------------------|-----------------|----------------|--------------|
| ความถูกต้องของการคำนวณ      | ถูกต้องทุกค่า      | ถูกส่วนใหญ่     | มีข้อผิดพลาด   | ผิดมาก       |
| การเลือกสถิติเหมาะกับข้อมูล | เหมาะสมทุกกรณี     | เหมาะสมส่วนใหญ่ | เหมาะสมบางส่วน | ไม่เหมาะสม   |
| การแปลผล/สื่อความหมาย       | แปลผลถูกต้องชัดเจน | แปลผลได้        | ไม่ชัดเจน      | แปลผลผิด     |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 8$  จาก 12 (ร้อยละ 70 ขึ้นไป)

## 12.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

ชูศรี วงศ์รัตนะ. (2552). เทคนิคการใช้สถิติเพื่อการวิจัย (พิมพ์ครั้งที่ 11).

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

นพเก้า ณ พัทลุง. (2558). การพัฒนาแบบทดสอบวัดความเข้าใจในการอ่านภาษาอังกฤษ สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 ใน 5 จังหวัดชายแดนภาคใต้. วารสารนวัตกรรมการศึกษา, 15(1), 42–47.

บุญชม ศรีสะอาด. (2560). การวิจัยเบื้องต้น (พิมพ์ครั้งที่ 10). สุวีริยาสาส์น.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการศึกษา. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จรรยา. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้.

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6).

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมชาย วรกิจเกษมสกุล. (2559). การวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 7).

โรงพิมพ์อักษรศิลป์.

## ภาษาอังกฤษ

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). Macmillan.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.

Ziegler, L., & Garfield, J. (2018). Developing a statistical literacy assessment for the modern introductory statistics course. *Statistics Education Research Journal*, 17(2), 161–178.

## บทที่ 13

### การตัดสินผล การให้ระดับผลการเรียน และการรายงานผล

### Scoring, Grading and Reporting of Results

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 14 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

เมื่อเก็บคะแนนและวิเคราะห์ทางสถิติแล้ว (บทที่ 12) ขั้นตอนสุดท้ายคือการตัดสินผลและรายงานผล บทที่ 13 มุ่งให้ ผู้เรียนเข้าใจการแปลความหมายคะแนนแบบอิงเกณฑ์และอิงกลุ่ม ระบบและสัญลักษณ์การให้คะแนน วิธีการตัดเกรดที่ ยุติธรรม การคำนวณระดับผลการเรียนเฉลี่ย (GPA) และการรายงานผลด้วยเอกสารหลักฐานการศึกษา เพื่อให้ปฏิบัติการ ตัดสินและรายงานผลได้อย่างถูกต้อง โปร่งใส และเป็นธรรม

### แผนบริหารการสอนประจำบทที่ 13

| รายการ                | รายละเอียด                                                                                                                                                                                                                                 |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) ความหมายและความสำคัญของการตัดสินผล 2) การแปลความหมายคะแนน (อิงเกณฑ์/อิงกลุ่ม) 3) ระบบและสัญลักษณ์การให้คะแนน 4) วิธีการตัดเกรด 5) การคำนวณระดับผลการเรียนเฉลี่ย (GPA) และการรายงานผล                                                    |
| ความคิดรวบยอด         | การตัดสินผลคือการนำคะแนนมาเทียบเกณฑ์/กลุ่มแล้วให้ระดับผลการเรียน ครูต้องเลือกวิธีตัดเกรดที่เหมาะสม ยุติธรรม และโปร่งใส แล้วรายงานผลอย่างถูกต้องด้วยเอกสารหลักฐานการศึกษา                                                                   |
| จุดประสงค์การเรียนรู้ | ปฏิบัติการตัดสินผล ให้ระดับผลการเรียน (อิงเกณฑ์/อิงกลุ่ม) คำนวณ GPA และจัดทำรายงานผลได้อย่างถูกต้อง (CLO4)                                                                                                                                 |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 13.1–13.5 ของบท)                                                                                                                                                                                |
| กิจกรรมการเรียนการสอน | 1) ทบทวน + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + สาธิตการตัดเกรด 3) ใบงาน 13.1 ตัดเกรดอิงเกณฑ์และอิงกลุ่มจากชุดคะแนน 4) ใบงาน 13.2 คำนวณ GPA + เขียนรายงานผล 5) อภิปรายความยุติธรรมในการตัดสินผล 6) สรุป + แบบทดสอบหลังเรียน + แบบฝึกหัด |
| สื่อการเรียนการสอน    | เอกสารประกอบการสอนบทที่ 13 • สไลด์นำเสนอ • ตัวอย่างการตัดเกรด/แบบรายงานผล (ปพ.) • เครื่องคิดเลข • ใบงาน 13.1–13.2 • ตำราที่อ้างอิง                                                                                                         |

|                    |                                                                                                                                                            |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| การวัดและประเมินผล | แบบทดสอบก่อน-หลังเรียน • ใบงาน 13.1 การตัดเกรด (เฉลย) • ใบงาน 13.2 GPA+รายงานผล (รูปรีดประเมินรายงาน) • การอภิปราย • แบบฝึกหัดท้ายบท (เกณฑ์ผ่าน ร้อยละ 70) |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|

### 13.1 ความคิดรวบยอด

การตัดสินผลคือกระบวนการนำคะแนนหรือสารสนเทศที่รวบรวมจากผู้เรียนมาเทียบกับเกณฑ์หรือกลุ่ม แล้วให้ระดับผลการเรียน (Grade) ที่สะท้อนความรู้ความสามารถที่แท้จริงตามมาตรฐานของรายวิชา (สมชาย วรกิจเกษมสกุล, 2559) การแปล ความหมายคะแนนทำได้ 2 แบบ คือ แบบอิงเกณฑ์ (Criterion-referenced) ที่เทียบกับเกณฑ์/จุดประสงค์ และแบบอิงกลุ่ม (Norm-referenced) ที่เทียบกับกลุ่มผู้สอบ ระบบการให้คะแนนมีทั้งระบบสมบูรณ์และสัมพัทธ์ และใช้สัญลักษณ์เป็นตัวเลข หรือตัวอักษร (เช่น A-F, 4-0) วิธีตัดเกรดมีหลายวิธีให้เลือกตามลักษณะรายวิชา เมื่อได้ระดับผลการเรียนแล้วจึงคำนวณ ระดับผลการเรียนเฉลี่ย (GPA) และ รายงานผลด้วยเอกสารหลักฐานการศึกษา ครูต้องตัดสินผลด้วยความรอบคอบ ยุติธรรม และปราศจากอคติ (พิชิต ฤทธิ์จัญญ, 2564; สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553)

### 13.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 13 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                      | เชื่อมโยง CLO              |
|------------------------------------------------------------------|----------------------------|
| 1) อธิบายความหมายและความสำคัญของการตัดสินผลได้                   | CLO4 (เข้าใจ)              |
| 2) เปรียบเทียบการแปลความหมายคะแนนแบบอิงเกณฑ์กับอิงกลุ่มได้       | CLO4 (เข้าใจ-วิเคราะห์)    |
| 3) ให้ระดับผลการเรียน (ตัดเกรด) จากชุดคะแนนด้วยวิธีที่เหมาะสมได้ | CLO4 (ประยุกต์)            |
| 4) คำนวณระดับผลการเรียนเฉลี่ย (GPA) ได้                          | CLO4 (ประยุกต์)            |
| 5) จัดทำรายงานผลการเรียนอย่างถูกต้องและเป็นธรรมได้               | CLO4 (ประยุกต์-สร้างสรรค์) |

### 13.3 คำสำคัญ

การตัดสินผล (Grading) • การแปลความหมายอิงเกณฑ์ (Criterion-referenced) • อิงกลุ่ม (Norm-referenced) • ระบบสมบูรณ์/สัมพัทธ์ (Absolute/Relative) • ระดับผลการเรียน (Grade) • การตัดเกรด • ระดับผลการเรียนเฉลี่ย (GPA/GPAX) • การรายงานผล (Reporting) • เอกสารหลักฐานการศึกษา (ปพ.)

### 13.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงลิงก์ที่ผู้สอน ประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยและสะท้อนผลการทดสอบ พร้อมให้ข้อมูลย้อนกลับแก่ผู้เรียนในสัปดาห์ถัดไป

### 13.5 เนื้อหาวิชา

#### 13.5.1 ความหมายและความสำคัญของการตัดสินผล

การกำหนดระดับผลการเรียนหรือการตัดเกรด (Grading) หมายถึง กระบวนการพิจารณาตัดสินผลการเรียนรู้ของผู้เรียน โดยนำผลที่ได้จากการวัดตลอดภาคเรียนมาเทียบกับเกณฑ์หรือกับกลุ่ม แล้วกำหนดออกมาเป็นระดับคุณภาพหรือสัญลักษณ์ ที่สื่อความหมายได้ เช่น A B C หรือ ผ่าน/ไม่ผ่าน (พัชรินทร์ ชมภูวิเศษ, 2559; พิชิต ฤทธิ์จรูญ, 2564) ทั้งนี้ควรแยกแนวคิดสามคำที่มักสับสนออกจากกันให้ชัดเจน ได้แก่ การวัด (Measurement) คือการกำหนดตัวเลขแทนปริมาณของคุณลักษณะตามกฎเกณฑ์ การประเมินค่า (Evaluation) คือการตัดสินคุณค่าโดยนำผลการวัดไปเทียบกับเกณฑ์ และการตัดสินผลการเรียน (Grading) คือการสรุปคุณค่านั้นออกมาเป็นระดับผลการเรียนเพื่อการรายงานและการตัดสินใจทางการศึกษา ความสัมพันธ์ของทั้งสามแสดงไว้ในตารางที่ 13.1

ตารางที่ 13.1 ความสัมพันธ์ระหว่างการวัด การประเมินค่า และการตัดสินผล

| ประเด็น   | การวัด (Measurement)  | การประเมินค่า (Evaluation) | การตัดสินผล (Grading)  |
|-----------|-----------------------|----------------------------|------------------------|
| คำถามหลัก | ได้เท่าไร (ปริมาณ)    | ดี/พอใช้/ไม่ผ่าน เพียงใด   | ควรได้ระดับใด          |
| ผลลัพธ์   | คะแนนดิบ/คะแนนแปลงรูป | การตัดสินคุณค่าเทียบเกณฑ์  | ระดับผลการเรียน (เกรด) |
| ตัวอย่าง  | สอบได้ 78 คะแนน       | อยู่ในระดับดี              | ได้เกรด B+             |

ในเชิงลึก คะแนนดิบ (Raw Score) ที่ครูเก็บได้ไม่ได้สะท้อนความสามารถจริงทั้งหมด เพราะประกอบด้วยคะแนนความสามารถจริงและคะแนนความคลาดเคลื่อนเสมอ ( $X_i = T_i + E_i$ ) การตัดสินผลจึงต้องตั้งอยู่บนคะแนนที่ได้จากเครื่องมือที่มีคุณภาพ มิฉะนั้นเกรดที่ได้ย่อมบิดเบือน (พัชรินทร์ ชมภูวิเศษ, 2559) การตัดสินผลทำหน้าที่เสมือน “ผู้รักษาประตูคุณภาพ” (gatekeeper) ของระบบการศึกษา หากผิดพลาดจะเกิดผลกระทบลุกล้ำต่อความเชื่อถือได้ของวุฒิการศึกษาและอนาคตของผู้เรียน โดยมีจุดมุ่งหมายสำคัญสามประการ คือ การวินิจฉัยจุดเด่น-จุดบกพร่อง การปรับปรุงการเรียนการสอน และการตัดสิน-รับรองผลเพื่อเลื่อนชั้นหรือสำเร็จการศึกษา (สมชาย วรภิเษมสกุล, 2559)

### กล่องที่ 13.1 หลักการตัดสินผลที่ดี 4 ประการ

- 1) เป็นระบบ — ใช้หลักฐานและขั้นตอนที่กำหนดไว้ล่วงหน้า ไม่ตัดสินตามอารมณ์
- 2) ยุติธรรม — ใช้เกณฑ์เดียวกันกับผู้เรียนทุกคน ลดความคลาดเคลื่อนของการวัด
- 3) โปร่งใส — เปิดเผยเกณฑ์/น้ำหนักคะแนนให้ผู้เรียนทราบตั้งแต่ต้นภาคเรียน
- 4) ใช้วิจารณ์อย่างมีคุณธรรม — รับผิดชอบต่อผลและอธิบายได้ด้วยเหตุผล

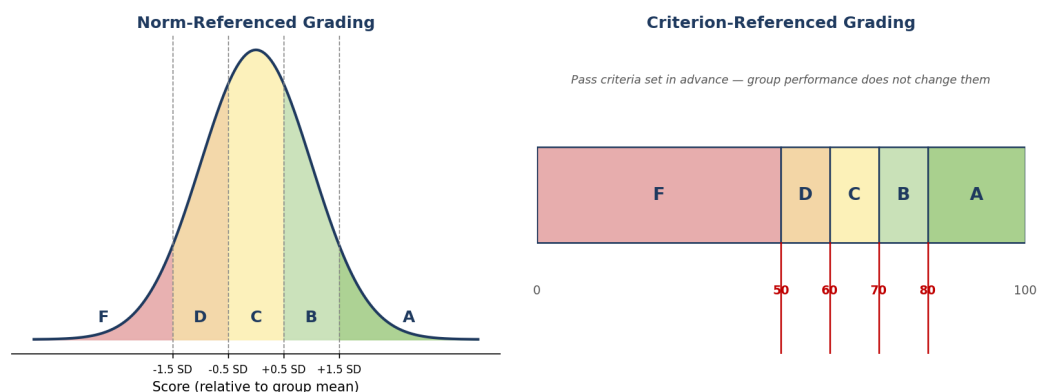
ที่มา: ดัดแปลงจาก พัทธจันทร์ ชมภูวิเศษ (2559, น. 302)

ตัวอย่างในชั้นเรียนไทย ครูผู้สอนรายวิชาหนึ่งกำหนดสัดส่วนคะแนนระหว่างภาคต่อปลายภาคเป็น 60 : 40 เมื่อสิ้นภาคจึงรวมคะแนนเก็บ คะแนนกลางภาค และปลายภาคตามน้ำหนักที่ประกาศไว้ แล้วนำคะแนนรวมไปเทียบเกณฑ์เพื่อให้เกรด กรณีที่ต้องระวังคือการ “ปิดคะแนน” ให้ผู้เรียนที่ยันหรือมีจิตอาสาเพียงเพราะความเห็นใจ ซึ่งถือเป็นอคติ (Halo Effect) และทำให้เกรดไม่สะท้อนความสามารถจริง แนวทางที่ถูกต้องคือกำหนดน้ำหนักคะแนนด้านความตั้งใจ/การมีส่วนร่วมไว้ในโครงสร้างคะแนนตั้งแต่ต้น แล้วตัดสินตามหลักฐานเชิงประจักษ์เท่านั้น

### 13.5.2 การแปลความหมายคะแนน: อิงเกณฑ์ กับ อิงกลุ่ม

การแปลความหมายคะแนนมีสองกรอบหลักที่ครูต้องเลือกใช้ให้เหมาะสมกับธรรมชาติของการวัด ได้แก่ การแปลความหมายแบบอิงเกณฑ์ (Criterion-referenced) คือการเทียบคะแนนกับเกณฑ์หรือจุดมุ่งหมายที่กำหนดไว้ล่วงหน้า เพื่อบอกว่าผู้เรียนรู้/ทำอะไรได้ตามมาตรฐานหรือไม่ และการแปลความหมายแบบอิงกลุ่ม (Norm-referenced) คือการเทียบคะแนนของผู้เรียนกับกลุ่มผู้สอบด้วยกัน เพื่อบอกตำแหน่งสัมพัทธ์ว่าเก่ง-อ่อนกว่าผู้อื่นเพียงใด (ศิริชัย กาญจนวาสี, 2552; ระพินทร์ โพธิ์ศรี, สุกัญญา รุจิเมธภาส, และสุรเชษฐ์ บุญรักษ์, 2567) มโนทัศน์พื้นฐานของการวัดแบบอิงกลุ่มและอิงเกณฑ์ ทบทวนได้ในบทที่ 2 (หัวข้อ 2.5.4)

ระพีพร โพธิ์ศรี และคณะ (2567) อธิบายความต่างนี้ด้วยอุปมา “การชิมแกง” ไว้อย่างเข้าใจง่ายว่า เมื่อกรอบเนื้อหาลึกและเป็นเนื้อเดียวกัน การสุ่มเนื้อหามาทดสอบเพียงส่วนน้อยก็แทนทั้งหมดได้ เปรียบเหมือนชิมแกงเพียงหนึ่งช้อนก็ตัดสินรสแกงทั้งหมดได้ จึงแปลความหมายคะแนนเป็นร้อยละแล้วเทียบเกณฑ์ได้ (อิงเกณฑ์) แต่เมื่อกรอบเนื้อหากว้างและไม่เป็นเนื้อเดียวกัน คะแนนที่ได้ไม่ใช่คะแนนเสมือนจริง จึงต้องแปลความหมายโดยอิงกลุ่มผู้สอบแทน หลักการสองกรอบนี้สอดคล้องกับงานของ Nitko และ Brookhart (2019) ที่ย้ำว่ากรอบการแปลผลต้องเลือกตามวัตถุประสงค์ของการสอบ ความแตกต่างเชิงปฏิบัติเปรียบเทียบได้ดังภาพที่ 13.1 และสรุปไว้ในตารางที่ 13.2



ภาพที่ 13.1 การแปลความหมายคะแนนแบบอิงกลุ่ม (เทียบตำแหน่งในกลุ่ม) เทียบกับแบบอิงเกณฑ์ (เทียบจุดตัดคงที่)

ตารางที่ 13.2 เปรียบเทียบการแปลความหมายคะแนนแบบอิงเกณฑ์กับอิงกลุ่ม

| ประเด็น     | อิงเกณฑ์ (Criterion-referenced)       | อิงกลุ่ม (Norm-referenced)           |
|-------------|---------------------------------------|--------------------------------------|
| เทียบกับ    | เกณฑ์/จุดมุ่งหมายที่กำหนดล่วงหน้า     | คะแนนของกลุ่มผู้สอบด้วยกัน           |
| ปรัชญา      | ทุกคนรอบรู้ได้หากถึงเกณฑ์             | ความสามารถกระจายตามโค้งปกติ          |
| จุดมุ่งหมาย | ตรวจสอบความรู้ตามมาตรฐาน              | จำแนก/คัดเลือกบุคคล                  |
| ข้อดี       | ส่งเสริมความร่วมมือ สื่อสารมาตรฐานชัด | กระตุ้นการแข่งขัน สะดวกในการคัดเลือก |
| ข้อจำกัด    | เกณฑ์ต่างคราต่างกัน/เกรดเพ้อได้       | มาตรฐานผูกกับกลุ่ม เทียบข้ามรุ่นยาก  |
| เหมาะกับ    | การเรียนรายวิชา/ประเมินสมรรถนะ (OBE)  | การสอบคัดเลือก/ทุนจำกัด              |

ที่มา: ดัดแปลงจาก ศิริชัย กาญจนวาสี (2552) และ ระพีพร โพธิ์ศรี และคณะ (2567)

ตัวอย่างในบริบทไทย หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ใช้แนวทางอิงเกณฑ์เป็นหลัก โดยกำหนดเกณฑ์ผ่านแต่ละรายวิชาที่ผู้เรียนต้องได้ไม่น้อยกว่าร้อยละ 50 (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553) ขณะที่การสอบคัดเลือกเข้าศึกษาต่อหรือการสอบชิงทุนซึ่งมีที่นั่งจำกัด มักใช้การอิงกลุ่มเพื่อจัดอันดับผู้สมัคร ข้อควรระวังคือ ในระบบที่เน้นผลลัพธ์การเรียนรู้ (OBE) การอิงกลุ่มมีความอ่อนแอ

เชิงปรัชญา เพราะหากผู้สอนจัดการเรียนรู้ได้ดีจนผู้เรียนทุกคนบรรลุจุดมุ่งหมาย การยึดโค้งงอติจะ “บังคับให้มีผู้สอบตก” ทั้งที่ไม่ควรมี การเลือกกรอบแปลผลจึงต้องสอดคล้องกับเจตนารมณ์ของรายวิชา

### 13.5.3 ระบบและสัญลักษณ์การให้คะแนน

ก่อนตัดเกรด ครูต้องเข้าใจชนิดของคะแนนและระบบสัญลักษณ์ที่ใช้ คะแนนแบ่งเป็นคะแนนดิบ (Raw Score) คือคะแนนที่ได้โดยตรงจากเครื่องมือวัด และคะแนนแปลงรูป (Derived Score) ที่ปรับให้สื่อความหมายได้ชัดเจน เช่น คะแนนร้อยละ อันดับเปอร์เซ็นต์ไทล์ และคะแนนมาตรฐาน (Z, T) คะแนนดิบจากต่างวิทยายังนำมาเทียบกันโดยตรงไม่ได้จนกว่าจะแปลงให้เป็นหน่วยเดียวกันเสียก่อน (Gronlund & Linn, 1990; พิชรินทร์ ชมภูวิเศษ, 2559) เมื่อจะรายงานผล ครูจึงนำคะแนนเหล่านี้มากำหนดเป็นระดับผลการเรียนด้วยระบบสัญลักษณ์ที่เลือกไว้

ระบบการให้คะแนนจำแนกได้สองแกน คือ จำแนกตามจำนวนระดับ (เช่น 2, 4, 5 หรือ 8 ระดับ) และจำแนกตามชนิดของสัญลักษณ์ (ตัวเลข เช่น 4 3 2 1 0 ตัวอักษร เช่น A B C D F คำบรรยาย เช่น ดีเยี่ยม-ดี-ผ่าน หรือระบบผ่าน/ไม่ผ่าน S/U) การเลือกจำนวนระดับและสัญลักษณ์ควรเหมาะสมกับธรรมชาติของวิชาและวัตถุประสงค์ของการรายงาน (Airasian, 2000; สมชาย วรภิเกษมสกุล, 2559) ดังสรุปในตารางที่ 13.3

ตารางที่ 13.3 ระบบและสัญลักษณ์การให้คะแนนจำแนกตามจำนวนระดับ

| จำนวนระดับ | สัญลักษณ์ที่นิยม                        | ตัวอย่างการใช้                                    |
|------------|-----------------------------------------|---------------------------------------------------|
| 2 ระดับ    | ผ่าน (P) / ไม่ผ่าน (F), S/U             | รายวิชาที่ตัดสินผ่าน-ไม่ผ่าน กิจกรรมพัฒนาผู้เรียน |
| 4-5 ระดับ  | ดีเยี่ยม/ดี/พอใช้/ปรับปรุง หรือ ก ข ค ง | ประเมินคุณลักษณะ การอ่านคิดวิเคราะห์เขียน         |
| 5 ระดับ    | A B C D F (4-0)                         | บางรายวิชา/บางสถาบัน                              |
| 8 ระดับ    | A B+ B C+ C D+ D F (4.0-0.0)            | ระดับมัธยมศึกษาและอุดมศึกษาโดยทั่วไป              |

ที่มา: ดัดแปลงจาก สมชาย วรภิเกษมสกุล (2559)

ระบบ 8 ระดับที่ใช้กันแพร่หลายในระดับอุดมศึกษากำหนดช่วงคะแนนและค่าระดับไว้ดังตารางที่ 13.4 โดยมีสัญลักษณ์พิเศษ I (Incomplete) สำหรับผู้เรียนที่ส่งงานไม่ครบ ขาดสอบ หรือไม่เข้าร่วมกิจกรรมตามที่กำหนด ซึ่งต้องดำเนินการให้สมบูรณ์ภายในเวลาที่ระเบียบกำหนด (พิชรินทร์ ชมภูวิเศษ, 2559)

### ตารางที่ 13.4 เกณฑ์ระดับผลการเรียน 8 ระดับแบบอิงเกณฑ์

| ช่วงคะแนน (ร้อยละ)  | ระดับผลการเรียน | ค่าระดับคะแนน |
|---------------------|-----------------|---------------|
| 80–100              | A               | 4.0           |
| 75–79               | B+              | 3.5           |
| 70–74               | B               | 3.0           |
| 65–69               | C+              | 2.5           |
| 60–64               | C               | 2.0           |
| 55–59               | D+              | 1.5           |
| 50–54               | D               | 1.0           |
| 0–49                | F               | 0.0           |
| ส่งงานไม่ครบ/ขาดสอบ | I               | —             |

ที่มา: พัชรินทร์ ชมภูวิเศษ (2559, น. 300)

ตัวอย่างในชั้นเรียนไทย ระดับอุดมศึกษาส่วนใหญ่รายงานผลด้วยระบบ 8 ระดับ (A ถึง F) ส่วนระดับการศึกษาขั้นพื้นฐานใช้ 8 ระดับสำหรับผลสัมฤทธิ์ 8 กลุ่มสาระ แต่ใช้ระบบคำบรรยาย (ดีเยี่ยม/ดี/ผ่าน) สำหรับการอ่านคิดวิเคราะห์เขียนและคุณลักษณะอันพึงประสงค์ และใช้ระบบผ่าน/ไม่ผ่านสำหรับกิจกรรมพัฒนาผู้เรียน ข้อควรระวังคือไม่ควรใช้จำนวนระดับที่ละเอียดเกินกว่าความเที่ยงของเครื่องมือจะรองรับได้ เพราะจะทำให้การจำแนกระดับขาดความน่าเชื่อถือ วิชาที่เน้นการฝึกปฏิบัติบางลักษณะจึงเหมาะกับระบบ S/U มากกว่าการตัดเกรดละเอียด

### 13.5.4 วิธีการตัดเกรด (Grading Methods)

วิธีการตัดเกรดหรือการกำหนดระดับผลการเรียนมีสามแนวทางหลัก (ศิริชัย กาญจนวาสี, 2552 อ้างถึงใน พัชรินทร์ ชมภูวิเศษ, 2559) ครูต้องเลือกวิธีที่เหมาะสมและยุติธรรมที่สุดกับลักษณะของรายวิชา ดังเปรียบเทียบในตารางที่ 13.5

ตารางที่ 13.5 เปรียบเทียบวิธีการตัดเกรดสามแนวทาง

| วิธี                 | หลักการ                                     | วิธีย่อย/ขั้นตอนสำคัญ                                                                                 | เหมาะกับ                |
|----------------------|---------------------------------------------|-------------------------------------------------------------------------------------------------------|-------------------------|
| อิงกลุ่ม (Norm)      | เทียบคะแนนภายในกลุ่ม<br>จัดตำแหน่งเก่ง-อ่อน | กำหนดสัดส่วนล่วงหน้า · แบ่งช่วงจากพิสัย ·<br>ขีดจำกัดกลางตามความสามารถกลุ่ม ·<br>แปลงเป็นคะแนนที่ปกติ | การคัดเลือก/จัดอันดับ   |
| อิงเกณฑ์ (Criterion) | เทียบคะแนนกับเกณฑ์ที่กำหนดล่วงหน้า          | ใช้เกณฑ์จุดตัดเป็นร้อยละ ·<br>ใช้เกณฑ์การบรรลุจุดมุ่งหมายเชิงพฤติกรรม                                 | การเรียนรายวิชา/สมรรถนะ |

|                |                                              |                                                |                                           |
|----------------|----------------------------------------------|------------------------------------------------|-------------------------------------------|
| ผสม<br>(Stuit) | ใช้ทั้งสองแบบร่วมกัน<br>ช่วงระดับห่างเท่ากัน | สอบย่อยเทียบเกณฑ์<br>แล้วใช้คะแนนรวมเทียบกลุ่ม | วิชาที่ต้องการทั้งมาตรฐานและ<br>ภาระจำแนก |
|----------------|----------------------------------------------|------------------------------------------------|-------------------------------------------|

ที่มา: ดัดแปลงจาก พัทรินทร์ ชมภูวิเศษ (2559, น. 302–306)

วิธีที่สอดคล้องกับระบบการศึกษาอิงผลลัพธ์มากที่สุดคือการตัดเกรดอิงเกณฑ์ด้วยเกณฑ์จุดตัดเป็นร้อยละ ซึ่งกำหนดช่วงคะแนนของแต่ละระดับไว้ล่วงหน้า เช่นเดียวกับเกณฑ์ 8 ระดับในตารางที่ 13.4 การกำหนดจุดตัด (cut score) ที่สมเหตุสมผลควรพิจารณาจากความสำเร็จของจุดมุ่งหมาย ความยากของข้อสอบ และความคาดหวังเชิงสมรรถนะของรายวิชา มิใช่ตั้งเกณฑ์ตามใจหรือคงเดิมโดยไม่ทบทวน ดังตัวอย่างการคำนวณในกล่องที่ 13.2

### กล่องที่ 13.2 ตัวอย่าง: การตัดเกรดอิงเกณฑ์ด้วยเกณฑ์จุดตัดร้อยละ

นักเรียน 5 คน มีคะแนนรวม (เต็ม 100): สมชาย 83, มะลิ 77, ก้อง 71, ฟ้า 58, ปอ 47

เทียบกับเกณฑ์ 8 ระดับ (ตารางที่ 13.4):

สมชาย 83 → A · มะลิ 77 → B+ · ก้อง 71 → B · ฟ้า 58 → D+ · ปอ 47 → F

ข้อสังเกต: เกรดของแต่ละคนขึ้นกับเกณฑ์คงที่ ไม่ขึ้นกับว่าเพื่อนได้คะแนนเท่าใด

ที่มา: ผู้เขียน (ใช้เกณฑ์ตาม พัทรินทร์ ชมภูวิเศษ, 2559)

แนวคิดร่วมสมัยที่เรียกว่า การให้เกรดอิงมาตรฐาน (Standards-based Grading) เสนอให้แยกการรายงานผลสัมฤทธิ์ทางวิชาการออกจากปัจจัยที่ไม่ใช่วิชาการ เช่น ความพยายาม การเข้าเรียน หรือความประพฤติ และให้ผู้เรียนมีโอกาสดูแลความรอบรู้ได้หลายครั้งโดยอิงเกณฑ์ที่ชัดเจน (Townsend, 2019) แนวคิดนี้ตอบโจทยปัญหาที่ Brookhart และคณะ (2016) สรุปจากการทบทวนงานวิจัยด้านการให้เกรดกว่าหนึ่งศตวรรษว่า เกรดมักเป็นส่วนผสมระหว่างพฤติกรรมและผลสัมฤทธิ์ ทำให้สื่อความหมายไม่ชัดเจน

ตัวอย่างในชั้นเรียนไทย ครูคณิตศาสตร์ที่ตัดเกรดอิงเกณฑ์จะประกาศช่วงคะแนนไว้ล่วงหน้า ส่วนครูศิลปะหรือดนตรีอาจให้ระดับคุณภาพโดยตรงจากrubricที่กำหนดเกณฑ์ของชิ้นงานแต่ละระดับ ทั้งสองแบบจัดเป็นการอิงเกณฑ์ ข้อควรระวังที่สำคัญคือปัญหา “เกรดเฟ้อ” (grade inflation) ซึ่งเกิดเมื่อเกณฑ์ต่ำเกินไปหรือครูปล่อยคะแนนง่าย ทำให้เกรดสูงขึ้นโดยไม่สะท้อนความสามารถจริง และอีกข้อคือการนำคะแนนการเข้า

เรียนหรือความประพฤติไปปนกับคะแนนผลสัมฤทธิ์จนเกิดสื่อความหมายคลาดเคลื่อน (Townsend, 2019; Gronlund & Linn, 1990)

### 13.5.5 การคำนวณระดับผลการเรียนเฉลี่ย (GPA) และการรายงานผล

ระดับผลการเรียนเฉลี่ย (Grade Point Average: GPA) เป็นการรายงานผลภาพรวมของหลายรายวิชา ในหนึ่งภาคเรียนหรือปีการศึกษา คำนวณจากผลรวมของผลคูณระหว่างค่าระดับเกรดกับหน่วยกิตของแต่ละวิชาหารด้วยผลรวมของหน่วยกิตทั้งหมด ส่วนค่าเฉลี่ยสะสมตลอดหลักสูตรเรียกว่า GPAX (พัชรินทร์ ชมภูวิเศษ, 2559) วิธีคำนวณและตัวอย่างแสดงในกล่องที่ 13.3

#### กล่องที่ 13.3 สูตรและตัวอย่างการคำนวณ GPA

$$\text{GPA} = \frac{\sum(\text{ค่าระดับเกรด} \times \text{หน่วยกิต})}{\sum \text{หน่วยกิต}}$$

ตัวอย่าง (ด.ช. สายฟ้า): คณิตศาสตร์ 2.5 นก. เกรด 3.5 · วิทยาศาสตร์ 2 นก. เกรด 4 · สังคมศึกษา 1 นก. เกรด 3.5 · ภาษาอังกฤษ 2 นก. เกรด 3

$$\text{ผลคูณ} = (3.5 \times 2.5) + (4 \times 2) + (3.5 \times 1) + (3 \times 2) = 8.75 + 8 + 3.5 + 6 = 26.25$$

$$\text{หน่วยกิตรวม} = 2.5 + 2 + 1 + 2 = 7.5 ; \text{GPA} = 26.25 / 7.5 = 3.50$$

ที่มา: พัชรินทร์ ชมภูวิเศษ (2559, น. 311–312)

การรายงานผล (Reporting) คือการสื่อสารผลการเรียนและการพัฒนาของผู้เรียนต่อผู้เรียน ผู้ปกครอง และหน่วยงาน อย่างน้อยภาคเรียนละหนึ่งครั้ง เพื่อใช้ปรับปรุงการเรียนและออกเอกสารหลักฐานการศึกษา (นงลักษณ์ วิรัชชัย, 2546; พัชรินทร์ ชมภูวิเศษ, 2559) วิธีการรายงานผลมีหลายรูปแบบให้เลือกตามวัตถุประสงค์ ดังตารางที่ 13.6

ตารางที่ 13.6 วิธีการรายงานผลการประเมินผลการเรียน 5 รูปแบบ

| วิธีการรายงานผล       | สาระสำคัญ                                                              |
|-----------------------|------------------------------------------------------------------------|
| 1) เขียนบรรยาย        | บรรยายพฤติกรรม/ความรู้และข้อบกพร่องเพื่อการวินิจฉัยและปรับปรุง         |
| 2) ระดับอันดับในกลุ่ม | จัดอันดับเก่ง-อ่อนภายในกลุ่ม กระตุ้นการพัฒนา (แต่ก่อให้เกิดการแข่งขัน) |
| 3) ระบุพัฒนาการ       | ใช้คะแนนมาตรฐาน (คะแนนที่) เทียบเกณฑ์ปกติ บอกพัฒนาการชัดเจน            |
| 4) อัตราเจริญงอกงาม   | เทียบผลก่อนเรียน-หลังเรียนด้วยแบบทดสอบคู่ขนาน                          |
| 5) เส้นภาพ (Profile)  | แสดงผลหลายวิชาด้วยคะแนนมาตรฐานเดียวกันในรูปเส้นภาพ                     |

ที่มา: ดัดแปลงจาก พัชรินทร์ ชมภูวิเศษ (2559, น. 312–313)

ตัวอย่างในบริบทไทย การรายงานผลตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ครอบคลุมสี่องค์ประกอบ ได้แก่ ผลสัมฤทธิ์ 8 กลุ่มสาระ (รายงานเป็น 8 ระดับ) การอ่าน-คิดวิเคราะห์-เขียน คุณลักษณะอันพึงประสงค์ และกิจกรรมพัฒนาผู้เรียน (ผ่าน/ไม่ผ่าน) โดยบันทึกในเอกสารหลักฐานการศึกษา เช่น ปพ.1 (ระเบียบแสดงผลการเรียนรู้) ปพ.5 (บันทึกผลการพัฒนาคุณภาพผู้เรียนรายวิชา) และ ปพ.6 (รายงานผลต่อผู้ปกครอง) ตามที่กำหนด (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2553)

#### กล่องที่ 13.4 เชื่อมโยงสู่ CLO4 และแนวโน้มการรายงานผล

เนื้อหา 13.5 ทั้งหมดรับใช้ CLO4: ผู้เรียนปฏิบัติการตัดเกรด (อิงเกณฑ์/อิงกลุ่ม) คำนวณ GPA และจัดทำรายงานผลการรายงานควรอยู่บนฐานหลักฐาน ใช้ภาษาสร้างสรรค์เชิงพัฒนา และเป็นธรรมต่อผู้เรียน (Nitko & Brookhart, 2019)

แนวโน้มศตวรรษที่ 21: รายงานผลเชิงสมรรถนะและใช้แดชบอร์ดข้อมูลการเรียนรู้เพื่อให้ข้อมูลย้อนกลับเชิงพัฒนา  
ที่มา: ผู้เขียน

ข้อควรระวังในการคำนวณและรายงาน คือ ต้องตรวจสอบหน่วยกิตและค่าระดับให้ถูกต้องก่อนหาค่าเฉลี่ย ระวังการปัดเศษที่ทำให้ผลคลาดเคลื่อน และต้องไม่เปิดเผยผลของผู้เรียนต่อสาธารณะในลักษณะที่ละเมิดความเป็นส่วนตัว การรายงานที่ดีจึงผานความถูกต้องเชิงสถิติเข้ากับจริยธรรมในการสื่อสารผล (พิชิต ฤทธิ์จรูญ, 2564)

## 13.6 กิจกรรมการเรียนรู้การสอนและใบงาน

---

### ใบงานที่ 13.1 ฝึกตัดเกรดอิงเกณฑ์และอิงกลุ่ม (รายบุคคล)

คำชี้แจง: กำหนดชุดคะแนนของผู้เรียน 15 คน ให้ตัดเกรด 2 แบบ คือ (1) อิงเกณฑ์ (ตามเกณฑ์ร้อยละที่กำหนด) และ (2) อิงกลุ่ม (ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน) แล้วเปรียบเทียบผลและอภิปรายข้อดี-ข้อจำกัด (มีเฉลยในหัวข้อ 13.10) — เชื่อมโยง CLO4

### ใบงานที่ 13.2 คำนวณ GPA และเขียนรายงานผล (รายบุคคล)

คำชี้แจง: จากตารางผลการเรียน (รายวิชา-เกรด-หน่วยกิต) ให้คำนวณ GPA และเขียนรายงานผลการเรียนเชิงพัฒนา 1 ฉบับ ที่ระบุจุดเด่น จุดที่ควรพัฒนา และข้อเสนอแนะ (ประเมินด้วยรูปรีในหัวข้อ 13.10) — เชื่อมโยง CLO4

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงานและกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

## 13.7 สรุปบทเรียน

การตัดสินผลคือการนำคะแนนหรือสารสนเทศมาเทียบกับเกณฑ์หรือกลุ่ม แล้วให้ระดับผลการเรียนที่สะท้อนความสามารถจริง การแปลความหมายคะแนนทำได้แบบอิงเกณฑ์ (เทียบเกณฑ์/จุดประสงค์) และอิงกลุ่ม (เทียบกลุ่มผู้สอบ) โดยหลักสูตรแกนกลางๆ เน้นอิงเกณฑ์ที่เกณฑ์ผ่านร้อยละ 50 ระบบการให้คะแนนมีทั้งระบบสมบูรณ์/สัมพัทธ์ และใช้สัญลักษณ์ตัวเลข/ตัวอักษร (เช่น 2, 3, 5, 8 ระดับ) วิธีตัดเกรดมีหลายวิธี เช่น โดยตรง ร้อยละ อันดับที่ การกระจาย คะแนนมาตรฐาน และแบบผสม ครูต้องเลือกวิธีที่เหมาะสมและยุติธรรม

เมื่อได้เกรดแล้วจึงคำนวณ  $GPA = \frac{\Sigma(\text{เกรด} \times \text{หน่วยกิต})}{\Sigma \text{หน่วยกิต}}$  และรายงานผล ด้วยเอกสารหลักฐาน การศึกษาอย่างถูกต้อง ชัดเจน ทันเวลา และเป็นธรรม โดยมุ่งให้ข้อมูลที่เป็นประโยชน์ต่อการพัฒนาผู้เรียน

### 13.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

หมายเหตุ: ผู้สอนจะเฉลยพร้อมให้ข้อมูลย้อนกลับรายจุดประสงค์ และสะท้อนผลการเรียนรู้ร่วมกับผู้เรียนในสัปดาห์ถัดไป

### 13.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

---

1. ปลายภาคเรียน นักเรียนคนหนึ่งได้คะแนนรวม 79.4 ขณะที่เกณฑ์เกรด A กำหนดไว้ที่ 80 คะแนน จงวิเคราะห์ว่าควรตัดสินใจอย่างไรจึงยุติธรรม โปร่งใส และตรวจสอบได้ โดยอ้างหลักการอิงเกณฑ์ อิงกลุ่ม และแนวปฏิบัติเรื่องการปิดคะแนนที่ประกาศล่วงหน้า
2. เปรียบเทียบการแปลความหมายคะแนนแบบอิงเกณฑ์กับอิงกลุ่ม พร้อมยกตัวอย่าง
3. อธิบายระบบและสัญลักษณ์การให้คะแนน พร้อมตัวอย่างระดับเกรดแบบต่าง ๆ
4. เปรียบเทียบวิธีตัดเกรดอย่างน้อย 3 วิธี ว่าเหมาะกับรายวิชาแบบใด
5. กำหนดตารางผลการเรียน (วิชา-เกรด-หน่วยกิต) แล้วคำนวณ GPA
6. เหตุใดการตัดสินผลและรายงานผลจึงต้องยุติธรรมและให้ข้อมูลเชิงพัฒนา อภิปราย

## 13.10 เครื่องมือวัดและประเมินผลประจำบท

### ก) เฉลยแนวทางใบงาน 13.1–13.2

ตัดเกรดอิงเกณฑ์: เทียบร้อยละกับเกณฑ์ที่กำหนด • ตัดเกรดอิงกลุ่ม: ใช้  $\bar{X}$  และ S.D. กำหนดช่วงเกรด • GPA  
$$= \frac{\sum(\text{เกรด} \times \text{หน่วยกิต})}{\sum \text{หน่วยกิต}} \text{ (เช่น } (4 \times 3 + 3 \times 3 + 2 \times 2) / 8 = 25/8 = 3.13)$$

### ข) rubric ประเมินรายงานผลการเรียน ใบงาน 13.2 (Analytic) — เชื่อมโยง CLO4

| ประเด็น                         | ดีมาก (4)             | ดี (3)      | พอใช้ (2)    | ปรับปรุง (1) |
|---------------------------------|-----------------------|-------------|--------------|--------------|
| ความถูกต้องของการคำนวณ/ตัดเกรด  | ถูกต้องทุกค่า         | ถูกส่วนใหญ่ | มีข้อผิดพลาด | ผิดมาก       |
| ความชัดเจนของรายงาน             | ชัดเจน อ่านเข้าใจง่าย | ค่อนข้างชัด | ไม่ชัด       | สับสน        |
| ข้อมูลเชิงพัฒนา (จุดเด่น/พัฒนา) | ครบและสร้างสรรค์      | ครบส่วนใหญ่ | ไม่ครบ       | ไม่มี        |
| ความเป็นธรรม/ปราศจากอคติ        | ยึดเกณฑ์ชัดเจน        | ค่อนข้างดี  | พอใช้        | ลำเอียง      |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 11$  จาก 16 (ร้อยละ 70 ขึ้นไป)

## 13.11 เอกสารอ้างอิงประจำบท

### ภาษาไทย

นงลักษณ์ วิรัชชัย. (2546). การวิเคราะห์และการรายงานผลการประเมิน.

สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

พัชรินทร์ ชมภูวิเศษ. (2559). เอกสารประกอบการสอนรายวิชาการวัดและประเมินผลการศึกษา.

คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์จัญญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้.

ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ระพีพันธ์ โพธิ์ศรี, สุกัญญา รุจิเมธธาส, และสุรเชษฐ์ บุญรักษ. (2567).

การวัดแบบอิงกลุ่มและการวัดแบบอิงเกณฑ์: ทฤษฎีและแนวปฏิบัติ. วารสารมหาวิทยาลัยราชภัฏมหาสารคาม, 18(1), 13–22.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6).

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมชาย วรกิจเกษมสกุล. (2559). การวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 7).

โรงพิมพ์อักษรศิลป์การพิมพ์.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2553).

แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.

โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

## ภาษาอังกฤษ

Airasian, P. W. (2000). Classroom assessment: Concepts and applications (4th ed.).

McGraw-Hill.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). Macmillan.

Nitko, A. J., & Brookhart, S. M. (2019). *Educational assessment of students* (8th ed.). Pearson.

Townsley, M. (2019). Considering standards-based grading: Challenges for secondary school leaders. *Journal of School Administration Research and Development*, 4(1), 35–38.

## บทที่ 14

### การใช้เทคโนโลยีและจริยธรรมในการวัดและประเมินผล

### Technology and Ethics in Measurement and Evaluation

รายวิชา ED1006 การวัดและประเมินผลการเรียนรู้ • 3(2-2-5) หน่วยกิต

หมวดวิชาเฉพาะ กลุ่มวิชาชีพครู หลักสูตรศึกษาศาสตรบัณฑิต (4 ปี)

คณะศึกษาศาสตร์ มหาวิทยาลัยมหาจุฬาราชวิทยาลัย

สัปดาห์ที่ 15 • จำนวน 4 คาบ (บรรยาย 2 / ปฏิบัติ 2)

#### ขอบเขตของบทเรียน

บทสุดท้ายบูรณาการความรู้ทั้งรายวิชาสู่บริบทปัจจุบัน มุ่งให้ผู้เรียนเข้าใจบทบาทของเทคโนโลยีในการวัดและประเมินผล (การทดสอบด้วยคอมพิวเตอร์ การประเมินออนไลน์ การทดสอบปรับเหมาะ ระบบให้คะแนนอัตโนมัติ การวิเคราะห์การเรียนรู้ และปัญญาประดิษฐ์) สามารถออกแบบแบบประเมินออนไลน์อย่างมีคุณภาพ และที่สำคัญที่สุดคือยึดมั่นในจริยธรรมและ จรรยาบรรณวิชาชีพ ทั้งความเป็นธรรม การรักษาความลับ และจริยธรรมดิจิทัล (ความเป็นส่วนตัวของข้อมูล การลอคคิตี เชิงอัลกอริทึม และการใช้ AI อย่างรับผิดชอบ)

#### แผนบริหารการสอนประจำบทที่ 14

| รายการ                | รายละเอียด                                                                                                                                                                                                                                                     |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อสาระ            | 1) บทบาทของเทคโนโลยีในการวัดและประเมินผล 2) รูปแบบการประเมินด้วยเทคโนโลยี 3) การออกแบบแบบประเมินออนไลน์ที่มีคุณภาพ 4) จริยธรรมและจรรยาบรรณวิชาชีพในการวัดและประเมินผล 5) จริยธรรมดิจิทัล (ความเป็นส่วนตัวของข้อมูล เชิงอัลกอริทึม และการใช้ AI อย่างรับผิดชอบ) |
| ความคิดรวบยอด         | เทคโนโลยีช่วยให้การประเมินสะดวก รวดเร็ว และหลากหลายขึ้น แต่ต้องคงไว้ซึ่งความตรง ความเที่ยง และความยุติธรรม ครูต้องใช้เทคโนโลยีอย่างมีจริยธรรม รักษาความเป็นส่วนตัวของข้อมูล และใช้ AI อย่างโปร่งใสและรับผิดชอบ                                                 |
| จุดประสงค์การเรียนรู้ | ออกแบบและทดลองใช้แบบประเมินออนไลน์ได้ (CLO4)<br>ตระหนักในจริยธรรมและความเที่ยงธรรมของการประเมิน (CLO5)<br>และใช้ผล/เทคโนโลยีเพื่อพัฒนาผู้เรียนอย่างสร้างสรรค์ (CLO6)                                                                                           |
| เนื้อหาวิชา           | ตามหัวข้อสาระ 5 หัวข้อ (รายละเอียดในหัวข้อ 14.1–14.5 ของบท)                                                                                                                                                                                                    |

|                       |                                                                                                                                                                                                                                                                              |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| กิจกรรมการเรียนการสอน | 1) ทบทวนภาพรวมรายวิชา + แบบทดสอบก่อนเรียน 2) บรรยายเชิงโต้ตอบ + สาธิตเครื่องมือประเมินออนไลน์ 3) ใบงาน 14.1 ออกแบบแบบประเมินออนไลน์ 4) ใบงาน 14.2 วิเคราะห์กรณีจริยธรรมดิจิทัลในการประเมิน 5) อภิปรายการใช้ AI อย่างรับผิดชอบ 6) สรุปรายวิชา + แบบทดสอบหลังเรียน + แบบฝึกหัด |
| สื่อการเรียนการสอน    | เอกสารประกอบการสอนบทที่ 14 • สไลด์นำเสนอ • เครื่องมือประเมินออนไลน์ (Google Forms/LMS/Quiz) • กรณีศึกษาจริยธรรม • ใบงาน 14.1–14.2 • ตำราที่อ้างอิง                                                                                                                           |
| การวัดและประเมินผล    | แบบทดสอบก่อน–หลังเรียน • ใบงาน 14.1 แบบประเมินออนไลน์ (รูปрик) • ใบงาน 14.2 การวิเคราะห์จริยธรรม • แบบสังเกตจริยธรรมดิจิทัล (CLO5) • แบบฝึกหัดท้ายบท (เกณฑ์ผ่านร้อยละ 70)                                                                                                    |

### 14.1 ความคิดรวบยอด

เทคโนโลยีเข้ามามีบทบาทสำคัญในการวัดและประเมินผล ตั้งแต่การทดสอบด้วยคอมพิวเตอร์ (Computer-Based Testing) การประเมินออนไลน์ การทดสอบปรับเหมาะ (Adaptive Testing) ระบบให้คะแนนอัตโนมัติ การวิเคราะห์การเรียนรู้ (Learning Analytics) จนถึงปัญญาประดิษฐ์ (AI) ซึ่งช่วยให้การประเมินสะดวก รวดเร็ว ปรับเหมาะกับผู้เรียน และให้ข้อมูลย้อนกลับได้ทันที อย่างไรก็ตาม เทคโนโลยีไม่ได้ลดความสำคัญของหลักการพื้นฐาน เครื่องมือยังต้องมีความตรง ความเที่ยง และความยุติธรรม ขณะเดียวกันก็เกิดประเด็นจริยธรรมใหม่ ๆ ได้แก่ ความเป็นส่วนตัวของข้อมูล อคติเชิงอัลกอริทึม (Algorithmic Bias) ความโปร่งใส และการตรวจสอบได้ ครูในฐานะนักวัดและประเมินผลจึงต้องยึดมั่นทั้งจรรยาบรรณวิชาชีพ ตามมาตรฐานสากล (AERA, APA, & NCME, 2014) และจริยธรรมดิจิทัล ใช้เทคโนโลยีและ AI อย่างโปร่งใสและรับผิดชอบต่อประโยชน์ในการพัฒนาผู้เรียนเป็นสำคัญ (UNESCO, 2021; Nitko & Brookhart, 2019)

### 14.2 จุดประสงค์การเรียนรู้ประจำบท

เมื่อศึกษาบทที่ 14 จบแล้ว ผู้เรียนสามารถ

| จุดประสงค์การเรียนรู้ (Learning Objectives)                            | เชื่อมโยง CLO              |
|------------------------------------------------------------------------|----------------------------|
| 1) อธิบายบทบาทและรูปแบบของเทคโนโลยีในการวัดและประเมินผลได้             | CLO4 (เข้าใจ)              |
| 2) ออกแบบและทดลองใช้แบบประเมินออนไลน์ที่มีคุณภาพได้                    | CLO4 (ประยุกต์–สร้างสรรค์) |
| 3) อธิบายจรรยาบรรณวิชาชีพในการวัดและประเมินผลได้                       | CLO5 (เข้าใจ)              |
| 4) ตระหนักและแสดงออกถึงจริยธรรมดิจิทัลและความเที่ยงธรรมในการประเมิน    | CLO5 (จิตพิสัย)            |
| 5) เสนอแนวทางใช้เทคโนโลยี/ผลการประเมินเพื่อพัฒนาผู้เรียนอย่างรับผิดชอบ | CLO6 (ประยุกต์–ประเมินค่า) |

### 14.3 คำสำคัญ

การทดสอบด้วยคอมพิวเตอร์ (Computer-Based Testing) • การประเมินออนไลน์ (Online Assessment) • การทดสอบปรับเหมาะ (Adaptive Testing) • การให้คะแนนอัตโนมัติ (Automated Scoring) • การวิเคราะห์การเรียนรู้ (Learning Analytics) • ปัญญาประดิษฐ์ (AI) • จริยธรรมดิจิทัล (Digital Ethics) • ความเป็นส่วนตัวของข้อมูล (Data Privacy) • อคติเชิงอัลกอริทึม (Algorithmic Bias) • จรรยาบรรณวิชาชีพ

### 14.4 แบบทดสอบก่อนเรียน

แบบทดสอบก่อนเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบ จัดทำในระบบทดสอบออนไลน์ แทนการพิมพ์ไว้ในเล่ม เพื่อป้องกันการเปิดดูเฉลยล่วงหน้าและเพื่อบันทึกผลการทดสอบรายบุคคลตามรหัส นักศึกษา ให้นักศึกษาทำแบบทดสอบก่อนเริ่มเรียนเนื้อหาบทนี้ โดยสแกน QR Code หรือเข้าถึงที่ผู้สอน ประกาศในชั้นเรียน

### 14.5 เนื้อหาวิชา

#### 14.5.1 บทบาทของเทคโนโลยีในการวัดและประเมินผล

การวัดและประเมินผลด้วยเทคโนโลยี หรือการประเมินดิจิทัล (Digital Assessment หรือ Technology-Enhanced Assessment) หมายถึง การนำเทคโนโลยีสารสนเทศและการสื่อสารมาใช้นับสนุน กระบวนการวัดและประเมินผลตลอดทั้งวงจร ตั้งแต่การสร้างเครื่องมือ การจัดสอบ การให้คะแนน การวิเคราะห์ คุณภาพ ไปจนถึงการรายงานและการนำผลไปใช้ จุดสำคัญคือเทคโนโลยีไม่ได้เปลี่ยน “หลักการ” ของการวัด แต่เปลี่ยน “วิธีการและเครื่องมือ” ให้มีความสะดวก รวดเร็ว และละเอียดขึ้น การเปลี่ยนจากกระดาษเป็น หน้าจอจึงไม่ได้ทำให้ข้อสอบที่ขาดคุณภาพกลายเป็นข้อสอบที่ดีขึ้นโดยอัตโนมัติ (พิชิต ฤทธิ์จรูญ, 2564)

บทบาทของเทคโนโลยีต่อการวัดและประเมินผลครอบคลุมอย่างน้อยหกด้าน ได้แก่ (1) คลังข้อสอบ ดิจิทัล (item bank) ที่จัดเก็บ ปรับปรุง และสุ่มข้อสอบได้สะดวก ลดการรั่วไหลและการใช้ข้อสอบซ้ำ (2) การจัด สอบและการตรวจให้คะแนนที่รวดเร็ว ลดภาระงานประจำของครู (3) การวิเคราะห์คุณภาพข้อสอบด้วย โปรแกรมสำเร็จรูป เช่น การหาค่าความยาก อำนาจจำแนก และความเที่ยง โดย พิชรินทร์ ชมภูวิเศษ (2559) ระบุว่าปัจจุบันมีการนำคอมพิวเตอร์มาช่วยวิเคราะห์ข้อสอบอย่างแพร่หลาย ทั้งตามทฤษฎีการทดสอบแบบ ดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ผ่านโปรแกรมต่าง ๆ เช่น BILOG และ ITEM (4) การให้ข้อมูลย้อนกลับทันที (immediate feedback) (5) การรายงานผลอัตโนมัติ และ (6) การ

ประเมินที่ขับเคลื่อนด้วยข้อมูล (data-driven assessment) ซึ่งใช้ร่องรอยการเรียนรู้จำนวนมากประกอบการตัดสินใจ

แม้เทคโนโลยีจะทรงพลัง แต่ต้องย้ำว่าเทคโนโลยีเป็นเพียง “เครื่องมือ” ไม่ใช่ “เป้าหมาย” หลักการพื้นฐานเรื่องความตรง (validity) ความเที่ยง (reliability) ความยุติธรรม (fairness) และความสอดคล้องเชิงสร้างสรรค์ (constructive alignment) ยังคงเป็นแกนกลางที่ต้องรักษาไว้เสมอ (ศิริชัย กาญจนวาสี, 2552; สมชาย วรกิจเกษมสกุล, 2559) ตัวอย่างในบริบทไทย เช่น โรงเรียนสร้างคลังข้อสอบในระบบ Google Forms หรือ Moodle เพื่อสุ่มข้อสอบให้นักเรียนแต่ละคนไม่ซ้ำกัน สถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) ใช้ระบบดิจิทัลประมวลผลการสอบระดับชาติ และครูใช้โปรแกรมตารางคำนวณวิเคราะห์ข้อสอบหลังสอบเพื่อปรับปรุงข้อที่ด้อยคุณภาพก่อนนำกลับมาใช้ใหม่

ข้อพึงระวังคือ ความเหลื่อมล้ำในการเข้าถึงเทคโนโลยี (digital divide) ที่อาจทำให้ผู้เรียนบางกลุ่มเสียเปรียบ และการพึ่งพาความสะดวกของเทคโนโลยีจนละเลยการตรวจสอบคุณภาพเครื่องมือ ครูจึงต้องใช้เทคโนโลยีอย่างรู้เท่าทัน โดยเลือกใช้ให้เหมาะกับจุดประสงค์และบริบทของผู้เรียน ดังสรุปบทบาท ประโยชน์ และข้อพึงระวังในตารางที่ 14.1

ตารางที่ 14.1 บทบาทของเทคโนโลยีในการวัดและประเมินผล: ประโยชน์และข้อพึงระวัง

| บทบาท                    | ประโยชน์หลัก                               | ข้อพึงระวัง                                              |
|--------------------------|--------------------------------------------|----------------------------------------------------------|
| คลังข้อสอบดิจิทัล        | จัดเก็บ ปรับปรุง สุ่มข้อสอบได้สะดวก        | ต้องควบคุมสิทธิ์เข้าถึงและความปลอดภัย                    |
| จัดสอบ-ตรวจอัตโนมัติ     | รวดเร็ว<br>ลดภาระและความคลาดเคลื่อนของครู  | เหมาะข้อปรนัย<br>ส่วนอัตนัยยังต้องครูร่วมตรวจ            |
| วิเคราะห์คุณภาพข้อสอบ    | หาความยาก-อำนาจจำแนก-<br>ความเที่ยงได้เร็ว | ต้องเข้าใจความหมายค่าสถิติ<br>ไม่ใช่เครื่องแทนวิจารณ์ญาณ |
| ข้อมูลย้อนกลับทันที      | ผู้เรียนรู้ผลและปรับปรุงได้เร็ว            | ระวังผลที่ผิดพลาดจากระบบทำให้เข้าใจคลาดเคลื่อน           |
| รายงานผล-วิเคราะห์ข้อมูล | เห็นภาพรวมและแนวโน้มเพื่อพัฒนาการสอบ       | ต้องคุ้มครองข้อมูลส่วนบุคคลของผู้เรียน                   |

ที่มา: ผู้เขียน สังเคราะห์จาก พิชิต ฤทธิจรูญ (2564) และ พัทธรินทร์ ชมภูวิเศษ (2559)

### 14.5.2 รูปแบบการประเมินด้วยเทคโนโลยี

รูปแบบการประเมินด้วยเทคโนโลยีมีหลากหลายและพัฒนาอย่างรวดเร็ว Heil และ Ifenthaler (2023) สังเคราะห์งานวิจัยเกี่ยวกับการประเมินออนไลน์ในระดับอุดมศึกษาจำนวน 114 เรื่อง (คัดจากกว่า 4,290 เรื่อง) พบว่าการประเมินออนไลน์จำแนกตามผู้ประเมินได้เป็นสี่ลักษณะ คือ การประเมินโดยเพื่อน (peer) โดยครู (teacher) โดยระบบอัตโนมัติ (automated) และการประเมินตนเอง (self) และมีแนวโน้มเคลื่อนไปสู่การประเมินที่ขับเคลื่อนด้วยการวิเคราะห์ข้อมูล (analytics-driven assessment) ที่ใช้ข้อมูลแบบเรียลไทม์เพื่อช่วยการตัดสินใจของผู้สอน

เมื่อพิจารณาตามลักษณะของระบบ รูปแบบที่พบบ่อยมีหกประเภท ได้แก่ การทดสอบด้วยคอมพิวเตอร์ (Computer-Based Testing: CBT) ที่ผู้สอบทำข้อสอบผ่านเครื่องและตรวจให้คะแนนทันที การประเมินออนไลน์ (Online Assessment) ผ่านเว็บหรือระบบจัดการเรียนรู้ (LMS) การทดสอบปรับเหมาะ (Adaptive Testing) ที่ปรับความยากของข้อถัดไปตามคำตอบของผู้สอบ ช่วยลดจำนวนข้อและวัดได้แม่นยำขึ้น ซึ่ง พระครูปลัดจักรพล สิริโร (2567) อธิบายว่าปัญญาประดิษฐ์ช่วยวิเคราะห์ข้อมูลผู้เรียนเพื่อปรับเนื้อหาและระดับความยากให้เหมาะกับแต่ละบุคคล การให้คะแนนอัตโนมัติ (Automated Scoring) ที่ตรวจข้อปรนัยได้ทันทีและเริ่มมี AI ช่วยตรวจข้อเขียน การวิเคราะห์การเรียนรู้ (Learning Analytics) ที่ใช้ร่องรอยพฤติกรรมเพื่อพัฒนาการสอน และปัญญาประดิษฐ์ (AI) ที่ช่วยสร้างข้อสอบ ให้ข้อมูลย้อนกลับ และวิเคราะห์เชิงลึก ดังตารางที่ 14.2

ตารางที่ 14.2 รูปแบบการประเมินด้วยเทคโนโลยีและตัวอย่างการใช้งาน

| รูปแบบ                                       | ลักษณะสำคัญ                                   | ตัวอย่างการใช้ในชั้นเรียน                    |
|----------------------------------------------|-----------------------------------------------|----------------------------------------------|
| การทดสอบด้วยคอมพิวเตอร์ (CBT)                | สอบผ่านเครื่อง ตรวจให้คะแนนอัตโนมัติ          | สอบเก็บคะแนนปรนัยในห้องคอมพิวเตอร์           |
| การประเมินออนไลน์ (Online)                   | ใช้แบบทดสอบ/ใบงานผ่านเว็บหรือ LMS             | มอบหมายแบบฝึกผ่าน Google Forms/Moodle        |
| การทดสอบปรับเหมาะ (Adaptive)                 | ปรับความยากข้อถัดไปตามคำตอบ                   | แบบฝึกที่ปรับโจทย์ตามระดับผู้เรียน           |
| การให้คะแนนอัตโนมัติ (Automated)             | ตรวจปรนัยทันที เริ่มมี AI ช่วยตรวจอัตนัย      | ตรวจ Quiz และให้คะแนนทันทีหลังส่ง            |
| การวิเคราะห์การเรียนรู้ (Learning Analytics) | วิเคราะห์ร่องรอยพฤติกรรมและผลการเรียน         | ดู Dashboard เพื่อหาผู้เรียนที่ต้องช่วยเหลือ |
| ปัญญาประดิษฐ์ (AI)                           | ช่วยสร้างข้อสอบ ให้ feedback วิเคราะห์เชิงลึก | ใช้ AI ร่างข้อสอบแล้วครูตรวจปรับ             |

การเลือกเครื่องมือดิจิทัลควรยึดจุดประสงค์เป็นหลัก เครื่องมียอดนิยมแต่ละชนิดมีจุดเด่นและบริบทที่เหมาะสมต่างกัน เช่น Google Forms เหมาะกับการประเมินย่อยเพื่อพัฒนา (formative) เพราะใช้ง่ายและเชื่อมกับตารางคำนวณ Microsoft Forms มีความปลอดภัยระดับองค์กรจึงเหมาะกับการสอบเก็บคะแนนอย่างเป็นทางการ ระบบ Moodle หรือ Thai MOOC รองรับคลังข้อสอบและสถิติขั้นสูงจึงเหมาะกับรายวิชาขนาดใหญ่ ขณะที่ Kahoot! และ Quizizz ใช้กลไกเกม (gamification) สร้างแรงจูงใจจึงเหมาะกับการทบทวนก่อนเรียน และ Google Classroom เหมาะกับการจัดการส่งงานและการให้ข้อมูลย้อนกลับต่อชิ้นงาน ดังเปรียบเทียบในตารางที่ 14.3

ตารางที่ 14.3 เปรียบเทียบเครื่องมือดิจิทัลเพื่อการประเมินและบริบทที่เหมาะสม

| เครื่องมือ         | จุดเด่น                                      | บริบทที่เหมาะสม                           |
|--------------------|----------------------------------------------|-------------------------------------------|
| Google Forms       | ใช้ง่าย เชื่อมต่อตารางคำนวณ<br>ตรวจอัตโนมัติ | การประเมินย่อย (formative)<br>และแบบสำรวจ |
| Microsoft Forms    | สร้างแบบทดสอบพร้อมเฉลย<br>ปลอดภัยระดับองค์กร | การสอบเก็บคะแนนอย่างเป็นทางการ            |
| Moodle / Thai MOOC | รองรับคลังข้อสอบและสถิติขั้นสูง              | รายวิชาขนาดใหญ่หรือระดับสถาบัน            |
| Kahoot! / Quizizz  | กลไกเกม สร้างแรงจูงใจ รายงานผลละเอียด        | การทบทวนก่อนเรียนและฝึกทักษะ              |
| Google Classroom   | จัดการการส่งงานและให้ข้อมูลย้อนกลับ          | การประเมินงานเขียนหรือโครงงาน             |

ที่มา: ผู้เขียน สังเคราะห์จากการทบทวนเครื่องมือที่ใช้แพร่หลายในสถานศึกษาไทย

ข้อควรระวังคือ ครูควรเลือกเครื่องมือตามจุดประสงค์การวัด ไม่ใช่ตามความนิยม เครื่องมือประเภทเกมเหมาะกับการประเมินเพื่อพัฒนามากกว่าการสอบตัดสินที่มีเดิมพันสูง (high-stakes) และเมื่อสอบนอกห้องเรียนต้องออกแบบให้ลดโอกาสการทุจริต เช่น สุ่มข้อ จำกัดเวลา หรือใช้ภาระงานที่ลอกกันได้ยาก

### 14.5.3 การออกแบบแบบประเมินออนไลน์ที่มีคุณภาพ

การออกแบบแบบประเมินออนไลน์ที่มีคุณภาพ หมายถึง การออกแบบเครื่องมือประเมินบนสื่อดิจิทัลที่ยังคงความตรง ความเที่ยง ความยุติธรรม และความสอดคล้องกับจุดประสงค์ พร้อมกับคำนึงถึงข้อพิจารณาเฉพาะของสื่อดิจิทัล เช่น การเข้าถึง ความปลอดภัย และความสุจริต หลักการสำคัญยังคงเป็นหลักเดียวกับการสร้างเครื่องมือทั่วไป กล่าวคือ ต้องมีหลักฐานสนับสนุนการตีความผลที่เหมาะสมกับจุดประสงค์ของการใช้ (Nitko & Brookhart, 2019)

หลักการออกแบบที่ควรยึดมีห้าประการ ประการแรก คือความสอดคล้องเชิงสร้างสรรค์ โดยข้อสอบและภาระงานต้องตรงกับจุดประสงค์และระดับพฤติกรรมที่ต้องการวัด ประการที่สอง คือการคงคุณภาพเครื่องมือ โดยตรวจสอบความตรงเชิงเนื้อหาด้วยค่าดัชนีความสอดคล้อง (IOC) และตรวจสอบความเที่ยงก่อนนำไปใช้จริง (ศิริชัย กาญจนวาสี, 2552) ประการที่สาม คือการเข้าถึงและความเป็นธรรม โดยออกแบบตามหลักการออกแบบที่เป็นสากลเพื่อการเรียนรู้ (Universal Design for Learning: UDL) ให้รองรับผู้เรียนที่หลากหลายและลดความเหลื่อมล้ำ ซึ่งสอดคล้องกับมาตรฐานการทดสอบที่เน้นความเป็นธรรมต่อผู้สอบทุกกลุ่ม (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) ประการที่สี่ คือความปลอดภัยและความสุจริต (test security and integrity) เช่น การสุ่มข้อ สุ่มตัวเลือก จำกัดเวลา และล็อกหน้าจอ และประการที่ห้า คือการคุมสอบทางไกล (remote proctoring) ที่ต้องทำอย่างเคารพสิทธิและความเป็นส่วนตัวของผู้เรียน

งานทบทวนวรรณกรรมของ Gikandi, Morrow และ Davis (2011) เกี่ยวกับการประเมินเพื่อพัฒนาแบบออนไลน์ (online formative assessment) พบว่า การใช้ภาระงานตามสภาพจริงอย่างต่อเนื่อง (ongoing authentic tasks) ร่วมกับการให้ข้อมูลย้อนกลับเชิงโต้ตอบ (interactive feedback) เป็นคุณลักษณะสำคัญที่ช่วยลดภัยคุกคามต่อความตรงและความเที่ยงของการประเมินในบริบทออนไลน์ และเพิ่มความผูกพันของผู้เรียน ข้อค้นพบนี้ชี้ว่าการออกแบบที่ดีไม่ได้เน้นเพียงการป้องกันการทุจริต แต่เน้นการออกแบบภาระงานที่มีความหมายและให้ข้อมูลย้อนกลับที่เป็นประโยชน์ต่อการเรียนรู้

ตัวอย่างในชั้นเรียนไทย เช่น ครูออกแบบแบบทดสอบใน Google Forms จำนวน 20 ข้อ เปิดการสลับข้อและสลับตัวเลือก จำกัดเวลา 30 นาที แนบคำชี้แจงและนโยบายความเป็นส่วนตัวของผู้สอบ และใช้ภาระงานเขียนสะท้อนคิดประกอบเพื่อลดการลอกในการสอบนอกห้องเรียน ข้อควรระวังที่พบบ่อย ได้แก่ ข้อสอบรั่วไหล ผู้เรียนบางคนเข้าไม่ถึงอินเทอร์เน็ตหรืออุปกรณ์ และการคุมสอบด้วยกล้องที่อาจละเมิดความเป็นส่วนตัว ครูจึงควรเตรียมทางเลือกสำรองและแจ้งเงื่อนไขล่วงหน้า ดังเช็กलिस्टในกล่องที่ 14.1

#### กล่องที่ 14.1 เช็กलिस्टออกแบบแบบประเมินออนไลน์คุณภาพ

- 1) ข้อสอบ/ภาระงานทุกข้อสอดคล้องกับจุดประสงค์และระดับพฤติกรรม (constructive alignment)
- 2) ตรวจสอบความตรงเชิงเนื้อหาด้วย IOC และประมาณความเที่ยงก่อนใช้
- 3) คำชี้แจงชัดเจน ส่วนติดต่อใช้งานง่าย เข้าถึงได้ตามหลัก UDL
- 4) ออกแบบให้เป็นธรรมต่อผู้เรียนทุกกลุ่ม มีทางเลือกสำรองเมื่อเข้าไม่ถึง
- 5) มีมาตรการความปลอดภัย: สุ่มข้อ สุ่มตัวเลือก จำกัดเวลา

- 6) แผนนโยบายความเป็นส่วนตัวของข้อมูลผู้สอบ และขอความยินยอมเมื่อจำเป็น
- 7) ให้ข้อมูลย้อนกลับที่เป็นประโยชน์ ไม่ใช่เพียงคะแนน
- 8) ทดลองใช้ (pilot) และเก็บหลักฐานเพื่อปรับปรุงเครื่องมือ

ที่มา: ผู้เขียน สังเคราะห์จาก Nitko และ Brookhart (2019), Gikandi และคณะ (2011) และ AERA, APA และ NCME (2014)

#### 14.5.4 จริยธรรมและจรรยาบรรณวิชาชีพในการวัดและประเมินผล

จรรยาบรรณในการวัดและประเมินผล (assessment ethics) หมายถึง หลักความประพฤติที่ครูในฐานะนักวัดและประเมินผลพึงยึดถือ เพื่อให้การประเมินมีความเป็นธรรม ถูกต้องตามหลักวิชา และเป็นประโยชน์ต่อผู้เรียน หลักเหล่านี้อ้างอิงได้ทั้งจากมาตรฐานสากลอย่าง Standards for Educational and Psychological Testing ของ AERA, APA และ NCME (2014) และจากข้อบังคับคุรุสภาว่าด้วยจรรยาบรรณของวิชาชีพ (สำนักงานเลขาธิการคุรุสภา, 2556) ซึ่งกำหนดให้ครูปฏิบัติต่อศิษย์ด้วยความเป็นธรรมและความเมตตา พิชรินทร์ ชมภูวิเศษ (2559, น. 20–21) ได้สรุปคุณลักษณะด้านคุณธรรมของผู้ทำหน้าที่วัดและประเมินผลไว้อย่างชัดเจน ประกอบด้วย ความรับผิดชอบ ความซื่อสัตย์สุจริต การมีวินัยในตนเอง ความซื่อสัตย์ ความยุติธรรม (พิจารณาตัดสินโดยไม่เห็นแก่ประโยชน์ของตนเองหรือของผู้สอบคนใดคนหนึ่ง) การมีเจตคติที่ดีต่อการวัดผลในฐานะเครื่องมือพัฒนา การมีใจกว้างยอมรับความเห็นต่าง การมีมนุษยสัมพันธ์ การแสวงหาความรู้ใหม่ ๆ และการรักษาความลับของเครื่องมือและผลการประเมินของผู้เรียน คุณลักษณะข้อสุดท้ายนี้มีความสำคัญอย่างยิ่งในยุคดิจิทัล เพราะข้อมูลผลการประเมินถูกจัดเก็บและส่งต่อทางอิเล็กทรอนิกส์ได้ง่าย

เมื่อสรุปเป็นหลักปฏิบัติ จริยธรรมสำคัญในการวัดและประเมินผลมีห้าด้าน ได้แก่ ความเป็นธรรมและปราศจากอคติ คือประเมินทุกคนด้วยเกณฑ์เดียวกันโดยไม่เลือกปฏิบัติ ความตรงและความเที่ยง คือใช้เครื่องมือที่มีคุณภาพและตีความผลอย่างเหมาะสม การรักษาความลับ คือปกป้องข้อมูลและผลการประเมินของผู้เรียน ความรับผิดชอบและความสุจริต คือดำเนินการสอบและให้คะแนนอย่างซื่อสัตย์โปร่งใส และการใช้ผลเพื่อพัฒนา คือมุ่งพัฒนาผู้เรียนเป็นสำคัญ ไม่ใช่เพื่อตีตราหรือลงโทษ ซึ่งสอดคล้องกับแนวปฏิบัติของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (2553) ที่ย้ำว่าการวัดและประเมินผลต้องมุ่งพัฒนาผู้เรียนเป็นหลัก

ตัวอย่างในชั้นเรียนไทย เช่น ครูไม่ประกาศคะแนนรายบุคคลต่อหน้าชั้นเรียนเพื่อรักษาศักดิ์ศรีของผู้เรียน ให้คะแนนตามเกณฑ์เดียวกันกับทุกคนแม้จะคุ้นเคยกับบางคนมากกว่า เก็บกระดาษคำตอบและไฟล์คะแนนเป็นความลับ และนำผลไปใช้ออกแบบการสอนซ่อมเสริมแทนการประจาน ข้อควรระวังที่พบบ่อย คืออคติจากความคุ้นเคยหรือความประทับใจแรก (halo effect) และการให้คะแนนตามความชอบส่วนตัว ดังสรุปหลักจรรยาบรรณและแนวปฏิบัติในตารางที่ 14.4

#### ตารางที่ 14.4 หลักจรรยาบรรณการวัดและประเมินผลกับพฤติกรรมพึงปฏิบัติและพึงเลี่ยง

| หลักจรรยาบรรณ        | พฤติกรรมพึงปฏิบัติ                             | พฤติกรรมพึงเลี่ยง                        |
|----------------------|------------------------------------------------|------------------------------------------|
| ความเป็นธรรม/ไร้อคติ | ใช้เกณฑ์เดียวกันกับผู้เรียนทุกคน               | ให้คะแนนตามความชอบหรือความคุ้นเคย        |
| ความตรง-ความเที่ยง   | ใช้เครื่องมือมีคุณภาพ ตีความผลอย่างระมัดระวัง  | ตัดสินจากข้อสอบที่ไม่ผ่านการตรวจสอบ      |
| การรักษาความลับ      | เก็บคะแนน/ข้อมูลเป็นความลับ จำกัดสิทธิ์เข้าถึง | ประกาศคะแนนรายบุคคลหรือเปิดเผยต่อสาธารณะ |
| ความรับผิดชอบ/สุจริต | ดำเนินการสอบและให้คะแนนอย่างโปร่งใส            | ปล่อยผลการคุมสอบหรือแก้คะแนนโดยมิชอบ     |
| ใช้ผลเพื่อพัฒนา      | นำผลไปสอนซ่อมเสริมและให้ข้อมูลย้อนกลับ         | ใช้ผลตีตรา จัดอันดับ หรือลงโทษผู้เรียน   |

ที่มา: ดัดแปลงจาก พัทธินทร์ ชมภูวิเศษ (2559, น. 20–21) และ สำนักงานเลขาธิการคุรุสภา (2556)

#### 14.5.5 จริยธรรมดิจิทัล: ความเป็นส่วนตัว อคติเชิงอัลกอริทึม และการใช้ AI อย่างรับผิดชอบ

จริยธรรมดิจิทัล (digital ethics) ในการประเมิน หมายถึง การประยุกต์หลักจริยธรรมเข้ากับบริบทของการใช้เทคโนโลยีและปัญญาประดิษฐ์ ซึ่งก่อให้เกิดประเด็นใหม่ที่ต้องตระหนักเพิ่มเติมจากจรรยาบรรณเดิม องค์การยูเนสโก (UNESCO, 2021) ได้ออกข้อเสนอแนะว่าด้วยจริยธรรมของปัญญาประดิษฐ์ โดยเน้นหลักความโปร่งใส ความเป็นธรรม ความรับผิดชอบ และการกำกับดูแลโดยมนุษย์ ซึ่งใช้เป็นกรอบสำหรับการนำ AI มาใช้ในการวัดและประเมินผลได้เป็นอย่างดี ประเด็นจริยธรรมดิจิทัลที่สำคัญมีอย่างน้อยสี่เรื่อง

เรื่องแรก คือความเป็นส่วนตัวของข้อมูล (data privacy) ข้อมูลผลการประเมินของผู้เรียนถือเป็นข้อมูลส่วนบุคคลที่ได้รับการคุ้มครองตามพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 (Personal Data Protection Act: PDPA) ครูจึงต้องเก็บข้อมูลเท่าที่จำเป็น (data minimization) ขอความยินยอมเมื่อจำเป็น จำกัดสิทธิ์การเข้าถึง และไม่เปิดเผยข้อมูลโดยไม่ได้รับอนุญาต ซึ่งสอดคล้องโดยตรงกับหลัก “การรักษาความลับ” ที่ พัทธินทร์ ชมภูวิเศษ (2559) ระบุไว้ เรื่องที่สอง คืออคติเชิงอัลกอริทึม (algorithmic bias) Baker และ Hawn (2022) ทบทวนงานวิจัยพบว่าอัลกอริทึมทางการศึกษาอาจให้ผลที่ลำเอียงต่อผู้เรียนตามเชื้อชาติ เพศ ภาษาแม่ หรือฐานะทางเศรษฐกิจ เพราะระบบสืบทอดอคติจากข้อมูลที่ใช้ฝึก ตัวอย่างที่พบ เช่น ระบบตรวจเรียงความอัตโนมัติให้คะแนนต่ำกว่าผู้เรียนที่ใช้ไวยากรณ์ภาษาถิ่น แนวทางแก้ไขคือการตรวจสอบความเป็นธรรมของระบบและให้ครูเป็นผู้กำกับการตัดสิน (human-in-the-loop)

เรื่องที่สาม คือความโปร่งใสและการตรวจสอบได้ (transparency) ระบบ AI ที่ซับซ้อนจนไม่สามารถอธิบายที่มาของคะแนนได้ หรือที่เรียกว่าปัญหา “กล่องดำ” (black box) ขัดต่อหลักการที่ว่า การประเมินต้องมี

ร่องรอยหลักฐานที่ตรวจสอบได้ พระครูปลัดจกักรพล สิริธโร (2567) ชี้ว่าแนวโน้มสำคัญในอนาคตคือการพัฒนาปัญญาประดิษฐ์ที่อธิบายการตัดสินใจของตนเองได้ (explainable AI) เพื่อให้ผู้เรียนเข้าใจเหตุผลของคะแนนและสามารถอุทธรณ์ผลได้ และเรื่องที่สี่ คือการพึ่งพา AI เกินควร ซึ่งอาจลดบทบาทดุลยพินิจของครู หลักสำคัญคือการใช้ AI เป็นผู้ช่วย โดยมีครูเป็นผู้ตัดสินสุดท้ายเสมอ ดังสรุปประเด็น ความเสี่ยง และแนวปฏิบัติในตารางที่ 14.5 ตารางที่ 14.5 ประเด็นจริยธรรมดิจิทัล ความเสี่ยง และแนวปฏิบัติสำหรับครู

| ประเด็น                  | ความเสี่ยง                                     | แนวปฏิบัติ                                         |
|--------------------------|------------------------------------------------|----------------------------------------------------|
| ความเป็นส่วนตัวของข้อมูล | ข้อมูลผู้เรียนรั่วไหลหรือถูกใช้ผิดวัตถุประสงค์ | เก็บเท่าที่จำเป็น ขอความยินยอม จำกัดสิทธิ์ตาม PDPA |
| อคติเชิงอัลกอริทึม       | AI สืบทอดอคติจากข้อมูลฝึกจึงตัดสินลำเอียง      | ตรวจสอบความเป็นธรรม ใช้ครูกำกับการตัดสินใจ         |
| ความโปร่งใส/ตรวจสอบได้   | ระบบเป็นกล่องดำ อธิบายที่มาของคะแนนไม่ได้      | เลือกระบบที่อธิบายได้ เปิดให้อุทธรณ์ผล             |
| การพึ่งพา AI เกินควร     | ลดบทบาทดุลยพินิจและความรับผิดชอบของครู         | ใช้ AI เป็นผู้ช่วย ครูเป็นผู้ตัดสินสุดท้าย         |
| ความเหลื่อมล้ำดิจิทัล    | ผู้เรียนบางกลุ่มเข้าไม่ถึงอุปกรณ์/เครือข่าย    | จัดทางเลือกสำรอง ออกแบบตามหลัก UDL                 |

ที่มา: ผู้เขียน สังเคราะห์จาก UNESCO (2021), Baker และ Hawn (2022) และพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

#### กล่องที่ 14.2 หลักการใช้ AI อย่างรับผิดชอบในการวัดและประเมินผล

- ความโปร่งใส (Transparency): เลือกระบบที่อธิบายเหตุผลของผลได้ และแจ้งผู้เรียนเมื่อใช้ AI
- ความเป็นธรรม (Fairness): ตรวจสอบและลดอคติของระบบต่อผู้เรียนทุกกลุ่ม
- ความรับผิดชอบ (Accountability): ครูรับผิดชอบผลการตัดสินใจ ไม่โยนให้ระบบ
- การกำกับโดยมนุษย์ (Human-in-the-loop): ครูตรวจสอบซ้ำและเป็นผู้ตัดสินสุดท้ายเสมอ
- ความเป็นส่วนตัว (Privacy): คุ้มครองข้อมูลผู้เรียนตาม PDPA และหลักการรักษาความลับ

ที่มา: ดัดแปลงจาก UNESCO (2021) และ พระครูปลัดจกักรพล สิริธโร (2567)

#### เชื่อมโยง CLO4, CLO5 และ CLO6

CLO4: ออกแบบแบบประเมินออนไลน์ที่มีคุณภาพ • CLO5: ตระหนักในจริยธรรมดิจิทัลและความเที่ยงธรรม • CLO6: ใช้เทคโนโลยีและผลการประเมินเพื่อพัฒนาผู้เรียนอย่างรับผิดชอบ ประเมินด้วยรูปแบบประเมินออนไลน์ (ใบงาน 14.1) และแบบสังเกตจริยธรรมดิจิทัล (ใบงาน 14.2) ในหัวข้อ 14.10

#### บทส่งท้ายรายวิชา

รายวิชา ED1006 ได้พาผู้เรียนเดินทางจากความรู้พื้นฐาน (บทที่ 1–3) สู่การสร้างเครื่องมือ (บทที่ 4–6) การประกันคุณภาพเครื่องมือ (บทที่ 7–9) การประเมินตามสภาพจริงและการใช้ผลพัฒนาผู้เรียน (บทที่ 10–11) การวิเคราะห์ทางสถิติ ตัดสินผล และรายงานผล (บทที่ 12–13) จนถึงการใช้อย่างมีจริยธรรม (บทที่ 14) ทั้งหมดยึดหลักความสอดคล้องเชิงสร้างสรรค์และความเป็นธรรม เพื่อให้การวัดและประเมินผลเป็นเครื่องมือพัฒนาผู้เรียนอย่างแท้จริง

#### **กรณีศึกษา: ปัญญาประดิษฐ์กับการวัดและประเมินผลอย่างรับผิดชอบ**

กฤติน ชันละ (2569) สังเคราะห์นโยบายและปรากฏการณ์การใช้ปัญญาประดิษฐ์ในระบบการศึกษาไทยช่วง พ.ศ. 2566–2568 ด้วยวิธีภาษาวีธี พบว่าด้านหนึ่ง AI ช่วยยกระดับการเรียนรู้เฉพาะบุคคลและลดภาระงานซ้ำซ้อนของครู แต่อีกด้านหนึ่งก็มีความเสี่ยงเชิงโครงสร้าง ได้แก่ การโอนถ่ายภาระทางปัญญาของผู้เรียน ข้อมูลสมจริงแต่เท็จที่ AI สร้างขึ้น และความเหลื่อมล้ำทางดิจิทัล ผู้เขียนเสนอหลัก "มนุษย์นำเทคโนโลยี" ยกระดับความฉลาดรู้ด้าน AI ให้เป็นการคิดเชิงวิพากษ์ และกำกับดูแลสามระดับ คือ รัฐ สถานศึกษา และ ผู้เรียนสำหรับการวัดและประเมินผล ข้อสรุปสำคัญคือ ครูใช้ AI ช่วยตรวจงาน วิเคราะห์คะแนน หรือให้ข้อมูลย้อนกลับเบื้องต้นได้ แต่การตัดสินผลขั้นสุดท้ายต้องอยู่ที่ครู พร้อมกลไกความโปร่งใสและการอุทธรณ์ของผู้เรียน

**คำถามชวนคิด:** หากสถานศึกษาของท่านนำระบบ AI มาช่วยตรวจข้อเขียน ท่านจะออกแบบมาตรการกำกับก่อนสอบ ระหว่างสอบ และหลังสอบอย่างไร เพื่อลดอคติเชิงอัลกอริทึมและรักษาความเป็นธรรมแก่ผู้เรียนทุกกลุ่ม

## **14.6 กิจกรรมการเรียนรู้การสอนและใบงาน**

### **ใบงานที่ 14.1 ออกแบบและทดลองใช้แบบประเมินออนไลน์ (รายกลุ่ม)**

คำชี้แจง: ให้กลุ่มออกแบบแบบทดสอบ/แบบประเมินออนไลน์ 1 ชุด (เช่น ด้วย Google Forms หรือเครื่องมือใน LMS) ที่สอดคล้องกับจุดประสงค์ มีคำชี้แจงชัดเจน คำนึงถึงความปลอดภัยและการเข้าถึง แล้วทดลองใช้กับเพื่อนและสะท้อนผล (ประเมินด้วยรูปรีในหัวข้อ 14.10) — เชื่อมโยง CLO4

## ใบงานที่ 14.2 วิเคราะห์กรณีจริยธรรมดิจิทัลในการประเมิน (รายกลุ่ม)

คำชี้แจง: อ่านกรณีศึกษา (เช่น ‘ระบบ AI ให้คะแนนต่ำกับผู้เรียนบางกลุ่มอย่างเป็นระบบ’ หรือ ‘การเปิดเผยผลสอบออนไลน์’) ให้กลุ่มวิเคราะห์ว่าขัดต่อจริยธรรม/จรรยาบรรณข้อใด และเสนอแนวปฏิบัติที่ถูกต้อง พร้อมนำเสนอ (ประเมินด้วยแบบสังเกตจริยธรรมดิจิทัล) — เชื่อมโยง CLO5/CLO6

การสะท้อนคิดและการให้ข้อมูลย้อนกลับ: ในคาบแรกของสัปดาห์ถัดไป ผู้สอนเฉลยและอภิปรายผลใบงาน และกิจกรรมของบทนี้ร่วมกับผู้เรียน ควบคู่กับการสะท้อนผลแบบทดสอบออนไลน์ก่อนเรียน-หลังเรียน เพื่อให้ผู้เรียนตรวจสอบความเข้าใจของตนเองและเห็นแนวทางพัฒนา ก่อนเริ่มเนื้อหาบทถัดไป

### 14.7 สรุปบทเรียน

เทคโนโลยีมีบทบาทสำคัญในการวัดและประเมินผล ตั้งแต่การทดสอบด้วยคอมพิวเตอร์ การประเมินออนไลน์ การทดสอบปรับเหมาะ การให้คะแนนอัตโนมัติ การวิเคราะห์การเรียนรู้ จนถึงปัญญาประดิษฐ์ ซึ่งช่วยให้การประเมินสะดวก รวดเร็ว และปรับเหมาะกับผู้เรียน แต่ไม่ได้ลดความสำคัญของหลักการพื้นฐานเรื่องความตรง ความเที่ยง และความยุติธรรม การออกแบบแบบประเมิน ออนไลน์ที่ดีต้องยึดความสอดคล้องเชิงสร้างสรรค์ คุณภาพเครื่องมือ และคำนึงถึงความปลอดภัยและการเข้าถึง ครูต้องยึด จรรยาบรรณวิชาชีพตามมาตรฐานสากลและคุรุสภา ทั้งความเป็นธรรม การรักษาความลับ และการใช้ผลเพื่อพัฒนา ในยุคดิจิทัล ยังต้องตระหนักถึงความเป็นส่วนตัวของข้อมูล อคติเชิงอัลกอริทึม ความโปร่งใส และใช้ AI อย่างรับผิดชอบโดยมีครูเป็นผู้กำกับ การตัดสินใจสุดท้าย ทั้งหมดนี้เพื่อให้การวัดและประเมินผลเป็นเครื่องมือพัฒนาผู้เรียนอย่างแท้จริง

### 14.8 แบบทดสอบหลังเรียน

---

แบบทดสอบหลังเรียนประจำบทเป็นแบบทดสอบปรนัยชนิดเลือกตอบชุดคู่ขนานกับแบบทดสอบก่อนเรียน จัดทำในระบบทดสอบออนไลน์ ให้นักศึกษาทำหลังเสร็จสิ้นกิจกรรมการเรียนรู้ประจำบท ระบบจะบันทึกคะแนนรายบุคคลเพื่อเปรียบเทียบพัฒนาการกับผลการทดสอบก่อนเรียน โดยสแกน QR Code หรือเข้าถึงที่ผู้สอนประกาศในชั้นเรียน

## 14.9 แบบฝึกหัดท้ายบทและคำถามทบทวน

1. โรงเรียนแห่งหนึ่งจะเปลี่ยนการสอบปลายภาคทุกรายวิชาเป็นการสอบด้วยคอมพิวเตอร์ จงวิเคราะห์ประเด็นที่ต้องเตรียมทั้งด้านคุณภาพเครื่องมือ ความปลอดภัยของระบบ และความเหลื่อมล้ำในการเข้าถึงเทคโนโลยีของผู้เรียน พร้อมเสนอแผนรองรับกรณีระบบขัดข้อง
2. เปรียบเทียบการทดสอบด้วยคอมพิวเตอร์ทั่วไปกับการทดสอบปรับเหมาะ (Adaptive)
3. อธิบายหลักการออกแบบแบบประเมินออนไลน์ที่มีคุณภาพและปลอดภัย
4. ระบุหลักจรรยาบรรณวิชาชีพในการวัดและประเมินผลอย่างน้อย 4 ข้อ
5. อธิบายประเด็นจริยธรรมดิจิทัล (ความเป็นส่วนตัว อคติเชิงอัลกอริทึม ความโปร่งใส) พร้อมแนวปฏิบัติ
6. อภิปรายว่าควรใช้ AI ในการวัดและประเมินผลอย่างไรจึงจะรับผิดชอบและเป็นธรรม

### 14.10 เครื่องมือวัดและประเมินผลประจำบท

#### ก) รوبرิคประเมินแบบประเมินออนไลน์ ใบงาน 14.1 (Analytic) — เชื่อมโยง CLO4

| ประเด็น                   | ดีมาก (4)           | ดี (3)      | พอใช้ (2)    | ปรับปรุง (1) |
|---------------------------|---------------------|-------------|--------------|--------------|
| ความสอดคล้องกับจุดประสงค์ | ทุกข้อตรงจุดประสงค์ | ตรงส่วนใหญ่ | ตรงบางส่วน   | ไม่สอดคล้อง  |
| คุณภาพข้อสอบ/ความชัดเจน   | ถูกหลักและชัดเจน    | ดีส่วนใหญ่  | มีข้อบกพร่อง | บกพร่องมาก   |
| การใช้งาน/การเข้าถึง      | ใช้ง่าย เข้าถึงได้  | ค่อนข้างดี  | พอใช้        | ใช้ยาก       |
| ความปลอดภัย/ความสุจริต    | มีมาตรการเหมาะสม    | มีบ้าง      | น้อย         | ไม่มี        |

เกณฑ์ผ่าน: ได้คะแนนรวม  $\geq 11$  จาก 16 (ร้อยละ 70 ขึ้นไป)

#### ข) แบบสังเกตจริยธรรมดิจิทัล ใบงาน 14.2 — เชื่อมโยง CLO5/CLO6

| พฤติกรรมบ่งชี้                                          | ปฏิบัติสม่ำเสมอ (3) | บางครั้ง (2) | ควรปรับปรุง (1) |
|---------------------------------------------------------|---------------------|--------------|-----------------|
| ตระหนักและเคารพความเป็นส่วนตัวของข้อมูลผู้เรียน         |                     |              |                 |
| ระบุและเสนอแนวทางลดอคติเชิงอัลกอริทึม                   |                     |              |                 |
| ยืนยันความโปร่งใสและการตรวจสอบได้ของการประเมิน          |                     |              |                 |
| เสนอการใช้เทคโนโลยี/AI เพื่อพัฒนาผู้เรียนอย่างรับผิดชอบ |                     |              |                 |

เกณฑ์: ดี = 10–12, ปานกลาง = 7–9, ควรปรับปรุง = ต่ำกว่า 7 (เต็ม 12)

### 14.11 เอกสารอ้างอิงประจำบท

ภาษาไทย

กฤตินี ชันละ. (2569). ปรัชญาการแก้ปัญหาประดิษฐ์ในระบบการศึกษาไทย: บทสังเคราะห์เชิงวิภาษวิธี พ.ศ. 2566–2568. วารสารวิจัยสังคมและปริทัศน์, 49(1).

พระครูปลัดจรรพ สิริธโร. (2567). การใช้เทคโนโลยี AI ในการเรียนการสอนแบบปรับเหมาะ (Adaptive Learning). วารสารวิชาการจินตาสี, 2(3).

พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562. (2562, 27 พฤษภาคม). ราชกิจจานุเบกษา, เล่ม 136 ตอนที่ 69 ก, หน้า 52–95.

พัชรินทร์ ชมภูวิเศษ. (2559). การวัดและประเมินผลการเรียนรู้. คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.

พิชิต ฤทธิ์เจริญ. (2564). เทคนิคการวัดและประเมินผลการเรียนรู้. ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (พิมพ์ครั้งที่ 6). โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมชาย วรภิเษมสกุล. (2559). การวัดและประเมินผลการศึกษา (พิมพ์ครั้งที่ 7). โรงพิมพ์อักษรศิลป์.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2553). แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย. สำนักงานเลขาธิการคุรุสภา. (2556). ข้อบังคับคุรุสภาว่าด้วยจรรยาบรรณของวิชาชีพ พ.ศ. 2556. คุรุสภา.

## ภาษาอังกฤษ

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>

Heil, J., & Ifenthaler, D. (2023). Online assessment in higher education: A systematic review. *Online Learning*, 27(1), 187–218.

รายนามคณะกรรมการวิชา ED 1006 การวัดและประเมินผลการเรียนรู้

- |                            |                                      |
|----------------------------|--------------------------------------|
| 1. ดร.ปิยะสุตา เพชรเวช     | วิทยาเขตร้อยเอ็ด                     |
| 2. นางสาวนัยน์พร สุภา      | วิทยาเขตอีสาน                        |
| 3. นายชนมกร ประไกร         | มหาวิทยาลัยมหามกุฏราชวิทยาลัย ศาลายา |
| 4. นายแสงสุรีย์ ทองขาว     | วิทยาเขตศรีธรรมมาโคกราช              |
| 5. นายอาคม อุโคตร          | วิทยาเขตล้านนา                       |
| 6. ดร.สุรีพร ขาบุตรบุญทริก | วิทยาเขตศรีล้านช้าง                  |
| 7. นายศุภวัฒน์ บุญนาดี     | วิทยาเขตอีสาน                        |
| 8. นายอานันท์ กรมน้อย      | วิทยาเขตอีสาน                        |