

บทที่ 5

คุณลักษณะของแบบทดสอบที่ดี

นักวัดผลทั้งหลายต่างก็พยายามอย่างยิ่งที่จะอธิบายคำว่า “แบบทดสอบที่ดี” ว่าควรมีคุณลักษณะเช่นไร ทั้งนี้เพราะแบบทดสอบมีความสำคัญเกี่ยวข้องกับชีวิตและอนาคตของคนเป็นอย่างมาก เช่นใช้แบบทดสอบในการประเมินผลการเรียนการสอนว่านักเรียนคนนั้นคนนี้ควรจะผ่านไปได้หรือควรจะเรียนซ้ำ ใช้ในการคัดเลือกนักเรียนเข้าศึกษาต่อในระดับต่าง ๆ ทุกระดับ ใช้ในการคัดเลือกบุคคลเข้าทำงานในหน่วยงานต่าง ๆ จะเห็นได้ว่าจำเป็นจะต้องอาศัยแบบทดสอบเป็นข้อมูลพื้นฐานในการประเมินหรือตัดสินใจใด ๆ ทั้งนี้ถ้าหากแบบทดสอบด้อยคุณภาพ ผลลัพธ์ที่ออกมาจะด้อยคุณภาพและเชื่อถือไม่ได้ ดังนั้นจึงมีความจำเป็นอย่างยิ่งที่ครูและนักวัดผลทั้งหลายควรจะได้ทราบคุณลักษณะที่ดีของแบบทดสอบว่ามีอะไรบ้าง นักวัดผลที่มีชื่อเสียงแต่ละท่านได้ให้ความเห็นเกี่ยวกับเรื่องนี้ไปต่าง ๆ กัน แต่เมื่อสรุปแล้วจะได้ว่าคุณลักษณะที่ดีของแบบทดสอบมี 10 ประการ ดังนี้

1. ความเที่ยงตรง (Validity)
2. ความเชื่อมั่น (Reliability)
3. ความเป็นปรนัย (Objectivity)
4. ความยากง่ายของข้อสอบ (Item Difficulty)
5. อำนาจจำแนก (Discriminating Power)
6. ความยุติธรรม (Fairness)
7. งามลึก (Searching)
8. ชัดเจาะจง (Definite)
9. ยั่วยุ (Exemplary)
10. ประสิทธิภาพ (Efficiency)

1. ความเที่ยงตรง (Validity) หมายถึงคุณลักษณะของแบบทดสอบที่สามารถวัดในสิ่งที่ต้องการวัดได้อย่างถูกต้องสมตามความมุ่งหมาย ถ้าเปรียบเทียบความเที่ยงตรงกับนาฬิกาก็จะได้นาฬิกาที่บอกเวลาได้ตรงกับสัญญาณวิทยุเป็นนาฬิกาที่มีความเที่ยงตรงดี สามารถบอกเวลาได้ถูกต้อง และใช้เป็นเครื่องมือในการวัดเวลาได้ ส่วนนาฬิกาที่เดินไม่ตรงกับสัญญาณวิทยุคือนาฬิกาที่

เพราะบอกเวลาไม่ตรงกับความเป็นจริง และใช้เป็นเครื่องมือวัดเวลาไม่ได้ สำหรับนาฬิกาเก็นนี่ หมายถึงเก็นสำหรับเวลาของประเทศไทย แต่อาจบอกเวลาตรงกับประเทศลาวก็ได้ แต่ก็ยังถือว่าเป็นนาฬิกาเก็นอยู่ดี เพราะมีควัตถุประสงคที่จะวัดเวลาในประเทศไทย ทำนองเดียวกับแบบทดสอบที่มีความเที่ยงตรงก็จะต้องวัดได้ตรงตามความประสงค์ที่ต้องการจะวัด เช่นข้อสอบคณิตศาสตร์ ก็จะต้องวัดความรู้ความสามารถในทางคณิตศาสตร์เท่านั้น ไม่ใช่ไปวัดความสามารถทางภาษาไทย หรือความสามารถทางศิลปะ เพราะถ้าเป็นเช่นนั้นแล้วแบบทดสอบคณิตศาสตร์ดังกล่าวก็จะมี ความเที่ยงตรงสูงในการวัดความสามารถทางด้านภาษาและศิลปะ แต่ไม่มีความเที่ยงตรงในการวัด ความสามารถทางคณิตศาสตร์ ซึ่งคงไม่ใช่แบบทดสอบที่ครูต้องการแน่

ความเที่ยงตรงในที่นี้จะแบ่งออกเป็น 4 ชนิด คือ

1. ความเที่ยงตรงเชิงเนื้อหา (Content Validity)
2. ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity)
3. ความเที่ยงตรงเชิงสภาพ (Concurrent Validity)
4. ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity)

ความเที่ยงตรงแต่ละชนิด หมายถึงอะไร วิธีการในการหาความเที่ยงตรงแต่ละชนิดเป็นอย่างไร จะได้กล่าวอย่างละเอียดในบทต่อไป

2. ความเชื่อมั่น (Reliability) หมายถึงความคงที่สม่ำเสมอของคะแนนจากการวัด แบบทดสอบที่ดีจำเป็นจะต้องมีความเชื่อมั่นสูง กล่าวคือถ้าเป็นข้อสอบก็จะต้องสามารถให้คะแนนหรือ ผลการสอบคงที่แน่นอน ไม่ว่าจะสอบข้อสอบนั้นกี่ครั้งก็ตาม ถ้าสภาพหรือสภาวะการสอบเหมือน เดิม นั่นก็คือจะวัดกี่ครั้ง ๆ ก็ได้คะแนนเหมือนเดิมนั่นเอง แบบทดสอบที่มีความเชื่อมั่นนั้นไม่จำเป็น ต้องมีความเที่ยงตรงเสมอไป เช่นนาฬิกาที่ชี้ 9 น. ทุก ๆ วันที่วิทยุประกาศ 8 น. ก็จัดได้ว่านาฬิกา เรือนนี้มีความเชื่อมั่นดี แต่ขาดความเที่ยงตรง เพราะเวลาที่นาฬิกาเรือนนี้บอกไม่ใช่ความจริง ถ้า จะแกไขนาฬิกาเรือนนี้ให้มีความเที่ยงตรงก็อาจทำได้โดยเลื่อนเวลาออกไป 1 ชั่วโมง นั่นคือ ถ้าเครื่องมือหรือข้อสอบใด ๆ ก็ตาม หากมีความเชื่อมั่นแล้วก็อาจปรับปรุงให้มีความเที่ยงตรงได้ ส่วนวิธีการในการหาความเชื่อมั่นนั้น จะได้กล่าวอย่างละเอียดในบทต่อไป

3. ความเป็นปรนัย (Objectivity) เป็นคุณลักษณะสำคัญของข้อคำถาม ข้อสอบที่จะเป็น ปรนัยไม่ได้ หมายถึงข้อสอบประเภทถูกผิด จับคู่ หรือเลือกตอบแต่อย่างไร ข้อสอบใดจะมีความ เป็นปรนัย จะต้องประกอบด้วยคุณสมบัติ 3 ประการต่อไปนี้

1. มีความชัดเจนในตัวคำถาม ทุกคนอ่านคำถามนั้นแล้วจะเข้าใจความหมายได้ตรงกัน ส่วน

ใครจะตอบถูกหรือไม่นั้นขึ้นอยู่กับความรู้และความสามารถของแต่ละคน ค่าตามที่จะมีความเป็นปรนัย ต้องใช้ภาษาไม่วกวนจนทำให้ผู้สอบหลงผิดคิดไปคนละทิศละทาง นั่นคือตัวค่าตามจะต้องมีจุดที่จะตามแน่นอน ไม่คลุมเครือ

2. มีความคงที่ในการตรวจให้คะแนน สามารถให้คะแนนได้อย่างแน่นอน เชื่อมั่นได้ ไม่ว่าจะให้ผู้ใดตรวจผลการทำงานนั้นก็จะได้คะแนนออกมาตรงกัน หรือจะตรวจเองก็ครั้ง ๆ ก็จะได้คะแนนเท่ากัน การใช้ข้อสอบแบบจำกัดคำตอบ โดยเฉพาะอย่างยิ่งข้อสอบแบบเลือกตอบ จะช่วยให้มีความเป็นปรนัยในการให้คะแนนมาก ทั้งนี้เพราะสามารถตรวจโดยใช้แผ่นเฉลยคำตอบ (key) ได้

3. มีความแจ่มชัดในการแปลความหมายของคะแนน กล่าวคือสามารถแปลความหมายของคะแนนได้ตรงกัน คะแนนนั้น ๆ มีความหมายชัดเจนสามารถนำมาบอกอันดับความสามารถของผู้สอบได้อย่างมั่นใจ ในการที่จะแปลความหมายของคะแนนได้ตรงกันนั้น จำเป็นจะต้องมีเกณฑ์ในการแปลความหมายของคะแนน

ข้อสอบใดที่มีคุณสมบัติครบทั้ง 3 ประการดังกล่าว จึงจะถือว่ามีความเป็นปรนัย ส่วนข้อสอบประเภทถูกผิด จับคู่ หรือเลือกตอบ ซึ่งเรียกว่าข้อสอบแบบปรนัยนั้น เป็นลักษณะของการสร้างข้อสอบที่พยายามจะให้ข้อสอบมีความเป็นปรนัยนั่นเอง แต่ก็ไม่สามารถยืนยันได้ว่าข้อสอบเหล่านี้จะมีความเป็นปรนัยเสมอไป และข้อสอบแบบบรรยายหรือแบบอัตนัยนั้น ถ้าใครสามารถทำให้มีคุณสมบัติครบทั้ง 3 ประการดังกล่าวก็ถือว่าเป็นข้อสอบที่มีความเป็นปรนัยได้เช่นกัน

4. ความยากง่ายของข้อสอบ (Item Difficulty) ค่าความยากของข้อสอบ คือสัดส่วนที่คนทำข้อสอบข้อนั้นถูก กับจำนวนคนทั้งหมดที่เข้าสอบ โดยทั่วไปเขียนแทนค่าความยากง่ายด้วย p และ p นี้สามารถคำนวณได้จากสูตร

$$p = \frac{H + L}{n_H + n_L}$$

เมื่อ p แทน ความยากง่ายของข้อสอบ

H แทนจำนวนนักเรียนในกลุ่มสูงที่ตอบถูก

L แทนจำนวนนักเรียนในกลุ่มต่ำที่ตอบถูก

n_H แทนจำนวนคนทั้งหมดในกลุ่มสูง

n_L แทนจำนวนคนทั้งหมดในกลุ่มต่ำ

ตัวอย่าง ข้อสอบข้อหนึ่งมีคนทำถูก 25 คน ในจำนวนคน 100 คน

$$p = \frac{25}{100}$$

$$= .25$$

ถ้า p น้อย แสดงว่าข้อสอบนั้นยาก เช่น $p = 0$ แสดงว่าข้อสอบนั้นไม่มีคนทำถูกเลย ถ้า $p = 1$ แสดงว่าข้อสอบง่าย นักเรียนทุกคนทำข้อสอบนั้นถูกต้องหมด ถ้า $p = .50$ แสดงว่านักเรียนในกลุ่มนั้นทำถูกครึ่งหนึ่งทำผิดครึ่งหนึ่ง

ในการวัดผลทั่ว ๆ ไป นิยมเลือกใช้ข้อสอบที่มีความยากง่ายปานกลาง กล่าวคือไม่ยากเกินระดับความสามารถของนักเรียน จนนักเรียนส่วนใหญ่ทำไม่ได้ ขณะเดียวกันก็ต้องไม่ง่ายเกินไปจนนักเรียนส่วนใหญ่ตอบถูกต้องหมด ข้อสอบที่ง่ายหรือยากเกินไปจะมีประโยชน์ในการสอบวัดน้อยมาก ข้อสอบที่มีความยากง่ายปานกลาง คือข้อสอบที่มีค่า p อยู่ระหว่าง .20 - .80

6. อำนาจจำแนก (Discriminating Power) เป็นคุณลักษณะที่สำคัญของเครื่องมือในการวัดผลอย่างหนึ่ง ที่สามารถแยกหรือจำแนกบุคคลที่มีความสามารถแตกต่างออกจากกันได้ ซึ่งถือว่าเป็นความไว (sensitivity) ของเครื่องมือ

แบบทดสอบที่มีอำนาจจำแนกสูง หมายถึงแบบทดสอบที่สามารถจำแนกคนเก่งกับคนอ่อนออกจากกันได้อย่างเด็ดขาด คือใครเก่งมากก็ได้คะแนนมาก ใครเก่งน้อยก็ได้คะแนนน้อย ข้อคำถามแต่ละข้อก็จะต้องมีคุณลักษณะในด้านอำนาจจำแนกด้วย คือต้องเป็นข้อคำถามที่คนเก่งหรือคนที่มีความสามารถสูงตอบถูกมากกว่าคนอ่อน หรือคนที่มีความสามารถต่ำ โดยทั่วไปนิยมแทนค่าอำนาจจำแนกด้วย r ค่า r นี้สามารถคำนวณได้จากสูตร

$$r = \frac{H - L}{n_H}$$

เมื่อ r แทน ค่าอำนาจจำแนก

H แทน จำนวนนักเรียนในกลุ่มสูงที่ตอบถูก

L แทน จำนวนนักเรียนในกลุ่มต่ำที่ตอบถูก

n_H แทนจำนวนคนทั้งหมดในกลุ่มสูง ($n_H = n_L$ เสมอ)

ตัวอย่าง มีข้อสอบข้อหนึ่ง นำไปทดสอบนักเรียนจำนวน 30 คน เมื่อนำผลการสอบมาวิเคราะห์แล้วปรากฏผลดังนี้

1. คนเก่งตอบถูกทุกคน คนอ่อนตอบผิดทุกคน ข้อสอบข้อนี้จะมีค่าอำนาจจำแนกสูงสุด หรือมีค่า $r = 1.00$

$$r = \frac{15-0}{15} = 1.00$$

2. คนเก่งและคนอ่อนตอบถูกทุกคน ข้อสอบข้อนี้จะไม่มีความจำแนกเลย หรือมีค่า $r = 0$

$$r = \frac{15-15}{15} = 0$$

3. คนเก่งและคนอ่อนตอบผิดทุกคน ข้อสอบข้อนี้ก็จะไม่มีค่าอำนาจจำแนกเช่นกัน หรือมีค่า $r = 0$

$$r = \frac{0 - 0}{15} = 0$$

4. คนเก่งทุกคนตอบผิด ส่วนคนอ่อนทุกคนตอบถูก ข้อสอบข้อนี้มีอำนาจจำแนกสูงทางลบ ซึ่งเป็นข้อสอบที่ครูไม่ต้องการ หรือมีค่า $r = -1.00$

$$r = \frac{0 - 15}{15} = -1.00$$

5. คนเก่งตอบถูกบ้าง คนอ่อนตอบถูกบ้าง โดยที่คนเก่งตอบถูกมากกว่าคนอ่อน เช่น คนเก่งตอบถูก 10 คน คนอ่อนตอบถูก 3 คน

$$r = \frac{10 - 3}{15} = \frac{7}{15} = .47$$

ข้อสอบข้อนี้ถือว่ามีอำนาจจำแนกปานกลาง

6. คนเก่งตอบถูกบ้าง คนอ่อนตอบถูกบ้าง โดยที่คนอ่อนตอบถูกมากกว่าคนเก่ง เช่น คนเก่งตอบถูก 3 คน คนอ่อนตอบถูก 10 คน

$$r = \frac{3 - 10}{15} = \frac{-7}{15} = -.47$$

ข้อสอบข้อนี้ถือว่ามีอำนาจจำแนกปานกลางทางลบ ซึ่งเป็นข้อสอบที่ครูไม่ต้องการ

โดยทั่วไป ครูมักจะต้องการข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ ยิ่งสูงเท่าไรยิ่งดี ข้อสอบที่ถือว่ามีอำนาจจำแนกสูงคือข้อสอบที่มีค่า r ตั้งแต่ .20 ขึ้นไป

ไพศาล หวังพานิช ได้เสนอแนะวิธีการสร้างคำถามให้มีอำนาจจำแนกสูงดังนี้

1. คำถามต้องมีความชัดเจน หรือมีความเป็นปรนัยสูง
2. คำถามนั้นต้องตามคุณลักษณะที่ใช้ความคิด หรือต้องใช้ดุลยพินิจในการตอบ ไม่ถามผิวเผินเพียงข้อเท็จจริงธรรมดา

3. เป็นข้อคำถามที่เป็นไปตามหลักวิชา หรือมีความสมเหตุสมผล

4. กรณีที่เป็นข้อสอบแบบเลือกตอบ ตัวเลือกที่ใช้ต้องมีความเป็นเอกพันธ์

6. ความยุติธรรม (Fairness) แบบทดสอบที่มีความยุติธรรม หมายถึง แบบทดสอบที่ให้ความเสมอภาคกับผู้สอบทุกคน คือต้องเป็นแบบทดสอบที่ไม่ลำเอียงเข้ากับกลุ่มใด ไม่เปิดโอกาส

ให้คนใดคนหนึ่งได้เปรียบคนอื่น เช่น ในกรณีที่สอนหลายห้อง และใช้วิธีสอนต่างกัน เวลาออกข้อสอบก็ต้องไม่ลำเอียงเข้ากับห้องใดห้องหนึ่ง หรือในการออกข้อสอบวัดความรู้ทางคณิตศาสตร์ ก็จะไม่ออกข้อสอบโดยใช้โจทย์ภาษาอังกฤษ เพราะจะทำให้นักเรียนกลุ่มที่เก่งภาษาอังกฤษได้เปรียบ นอกจากนี้ลักษณะของคำถามที่ใช้จะต้องไม่เปิดโอกาสให้นักเรียนเก่งข้อสอบ หรือใช้ไหวพริบเตาได้ถูกต้อง การที่จะออกข้อสอบให้มีความเสมอภาคเช่นนี้ จะต้องออกข้อสอบให้มาก ๆ ข้อ หรือออกข้อสอบให้ครอบคลุมหลักสูตรนั่นเอง

7. ตามลึก (Searching) ข้อสอบที่ตามลึก หมายถึงข้อสอบที่ตามครอบคลุมพฤติกรรมหลาย ๆ ด้าน เช่น เป็นข้อสอบที่ตามวัดความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า ข้อคำถามที่ดีจะต้องตามให้ผู้สอบได้ใช้ความคิด ไม่ใช่ตามรายละเอียดตามตำรา การตามให้ลึกจะต้องตามตั้งแต่ระดับความเข้าใจขึ้นไป การตามให้ลึกมีลักษณะดังนี้

1. ไม่ตามตรงตามตำรา เช่น ไม่ถามว่า ทักษะสามมิติลูกกี่คน เพราะคำถามประเภทนี้สามารถหาคำตอบได้จากตำรา

2. ไม่ตามตามที่ครูบอก สิ่งที่ครูสอน ครูบอก ครูให้จด เมื่อนักสืบมาถามใหม่ก็จะเป็นการตามให้นักเรียนระลึกถึงสิ่งเดิม ซึ่งมีลักษณะเช่นเดียวกับการตามตามตำรา

3. ไม่ตามสิ่งที่สังเกตเห็นได้จากสังคมหรือสิ่งแวดล้อมโดยตรง ทั้งนี้เพราะคำถามประเภทนี้ไม่สามารถบ่งบอกถึงความสามารถได้เลย นอกจากจดจำสิ่งที่เคยพบมา เช่น ไม่ควรถามว่า คนเรามีกี่ตา? เพราะเด็กที่ไม่เคยเรียนก็สามารถตอบถูกได้ ซึ่งไม่ได้แสดงถึงความสามารถใดๆ ของเด็กเลย

ตัวอย่าง ข้อสอบภาษาอังกฤษที่ไม่ดี

"Tim and Tom are sitting under a tree"

1. Who sits with Tom? (Tim)
2. What are they doing? (sitting)
3. Who are sitting under a tree? (Tim and Tom)
4. Where are they sitting? (under a tree)

จากตัวอย่างนี้ จะเห็นได้ว่าตัวคำถามไม่ได้วัดความเข้าใจอะไรเลย คนที่อ่านข้อความที่กำหนดให้ไม่เข้าใจ ก็อาจตอบคำถามประเภทนี้ถูกได้

ตัวอย่าง ข้อสอบภาษาอังกฤษที่ดีขึ้น

My name is Tom. I am ten years old. Noi is my friend and younger than me one year.

Our school is a girls' school. We go to school on foot everyday because our school is not far from our houses.

1. Who am I?
 - a. a child
 - b. a baby
 - c. a boy
 - d. a girl
2. How old is Noi?
 - a. 7 years old.
 - b. 8 years old.
 - c. 9 years old.
 - d. 10 years old.
3. Where is my house?
 - a. near the street
 - b. near the school
 - c. near the market
 - d. far from school

ตัวอย่าง

A : I don't know why I married you. The house is always dirty.
The children are bad and I never have any clean clothes to wear.

B : You asked me to marry not to work.

1. Who is "A" talking to?
 - a. his girl student
 - b. his mother
 - c. his servant
 - d. his wife

2. How is "B"?

- a. lazy
- b. ugly
- c. silly
- d. fat

8. **จำเพาะเจาะจง (Definite)** ข้อสอบที่ดีต้องไม่ถามหลายแง่หลายมุม ข้อเดียวต้องถามเพียงอย่างเดียว และคำถามที่ใช้ต้องชัดเจน ไม่คลุมเครือ จนเด็กแต่ละคนเข้าใจคำถามไปคนละทิศละทาง คำถามประเภททวน สองแง่สองมุม ควรหลีกเลี่ยงเป็นอย่างยิ่ง ดังตัวอย่าง

ไม่ดี - คำคำทำอะไรกับเมียแก? (คำตอบต้องการ "ทำนา")

9. **ขั้วๆ (Exemplary)** คือเป็นข้อคำถามที่มีลักษณะท้าทายชวนให้เด็กคิดข้อสอบที่มีลักษณะเช่นนี้จะต้องมีการถามต่อ โดยเอาข้อสอบง่าย ๆ ไว้ในตอนแรก ๆ แล้วจึงค่อย ๆ ตามให้ยากขึ้นตามลำดับ ข้อสอบลักษณะดังกล่าวนี้จะเป็นการช่วยให้เด็กเกิดความพยายามที่จะทำข้อสอบให้ได้ทั้งหมด ข้อสอบที่ยากเกินไปและง่ายเกินไปจะไม่ท้าทายความคิดของเด็ก นอกจากนี้ลักษณะคำถามควรเป็นเรื่องที่ใกล้ตัวเด็ก พาดพิงกับชีวิตประจำวันของเด็ก และยังสามารถใช้รูปภาพประกอบ ซึ่งจะช่วยให้ความสนใจของเด็กได้เป็นอย่างดี

10. **ประสิทธิภาพ (Efficiency)** แบบทดสอบที่มีประสิทธิภาพ คือแบบทดสอบที่แสดงถึงการประหยัดทางเศรษฐกิจ เช่น ใช้เวลา แรงงาน และเงินน้อยที่สุด แต่สามารถวัดได้ถูกต้องมากที่สุด ตัวอย่างเช่น แบบทดสอบฉบับหนึ่งมี 20 ข้อ ใช้เวลาทำ 20 นาที กับแบบทดสอบอีกฉบับหนึ่งมี 60 ข้อ ให้เวลาทำ 60 นาที แต่ปรากฏว่า แบบทดสอบทั้งสองฉบับให้ผลเหมือน ๆ กัน ก็ย่อมแสดงว่าแบบทดสอบฉบับแรกมีประสิทธิภาพสูงกว่าแบบทดสอบฉบับหลัง ทั้งนี้เพราะแบบทดสอบฉบับแรกใช้เวลา แรงงาน และเงินน้อยกว่า แต่สามารถวัดได้ถูกต้องพอๆกับแบบทดสอบฉบับหลัง