

## บทที่ 4

### คุณลักษณะที่ดีของเครื่องมือที่ใช้ในการประเมิน

#### เนื้อหา

เนื้อหาที่จะกล่าวถึงในบทที่ 4 มีรายละเอียดดังนี้

- 4.1 ความเชื่อมั่น (Reliability)
- 4.2 ความเที่ยงตรง (Validity)
- 4.3 ความเป็นปรนัย (Objectivity)
- 4.4 ความยาก - ง่าย (Difficulty)
- 4.5 อำนาจจำแนก (Discrimination)

สำหรับแบบทดสอบ และแบบสอบถามนั้น เกณฑ์ที่ใช้ในการพิจารณาคุณภาพของเครื่องมือจะคล้ายคลึงกัน ยกเว้นเรื่อง “ความยาก - ง่าย” ของข้อคำถาม จะเป็นเกณฑ์ที่ใช้พิจารณาคุณภาพเฉพาะแบบทดสอบเท่านั้น

#### 4.1 ความเชื่อมั่น (Reliability)

ความเชื่อมั่นของเครื่องมือ คือ ค่าความคงที่ของคะแนนที่ได้จากการนำเครื่องมือฉบับหนึ่งไปทดสอบกับกลุ่มตัวอย่างกลุ่มเดียวกัน 2 ครั้งหรือหลายครั้ง ความเชื่อมั่นของเครื่องมือจึงเป็นเรื่องที่เกี่ยวข้องกับความคงเส้นคงวาของคะแนนที่ได้จากการทดลองใช้เครื่องมือฉบับนั้นหลายๆ ครั้ง เครื่องมือที่มีค่าความเชื่อมั่นสูง หมายความว่าคะแนนที่ได้จากการตอบเครื่องมือฉบับนั้นเป็นที่เชื่อถือได้ ไม่ว่าผู้ตอบจะตอบกี่ครั้งก็ตาม เขาก็จะได้คะแนนเหมือนเดิมทุกครั้งไป หรือถึงแม้คะแนนจะไม่คงเดิม อาจจะได้สูงหรือต่ำไปบ้างเล็กน้อยก็ตาม แต่ขอให้อันดับที่ของคะแนนในการตอบแต่ละครั้งเหมือนเดิม ก็ถือว่าเครื่องมือฉบับนั้นมีความเชื่อมั่นสูงได้เช่นเดียวกัน ความเชื่อมั่นของเครื่องมือใดๆ มีค่าอยู่ระหว่าง .00 ถึง 1.00

เครื่องมือฉบับใดถ้ามีความเชื่อมั่นเป็น 1.00 หมายความว่าเครื่องมือฉบับนี้มีความเชื่อมั่นสูงสุด นั่นคือคะแนนที่ได้จากเครื่องมือฉบับนี้เชื่อถือได้ ไว้วางใจได้ ไม่ว่าจะนำแบบทดสอบฉบับนี้ไปทดสอบกับผู้สอบคนเดิมกี่ครั้งก็หนักก็ตาม เขาก็จะได้คะแนนหรืออันดับที่เหมือนเดิมทุกทีไป

ส่วนเครื่องมือที่มีความเชื่อมั่นเป็น 0.00 (ศูนย์) หรือใกล้เคียงศูนย์ แสดงว่า เครื่องมือฉบับนั้นขาดความเชื่อมั่น คะแนนที่ได้จากการทดสอบเครื่องมือฉบับนั้นเชื่อถือไม่ได้ สอบครั้งแรกจะได้คะแนนอย่างหนึ่ง แต่พอสอบครั้งต่อมากลับได้คะแนนอีกอย่างหนึ่ง

ตามปกติแล้ว ปริมาณความเชื่อมั่นของแบบทดสอบควรจะสูงต่ำปานใด ขึ้นอยู่กับ จุดมุ่งหมายของการนำแบบทดสอบไปใช้ ถ้าเราต้องการจะใช้ผลการสอบเพื่อเป็นการทบทวนเนื้อหา วิชาที่ครูสอน เราก็อาจจะใช้แบบทดสอบที่ยังไม่รู้ค่าความเชื่อมั่นก็ได้ เพราะการตัดสินของเรา ขึ้นอยู่กับผลการสอบของผู้เข้าสอบทั้งกลุ่ม เราจะไม่ฟังเสียงผลการสอบของผู้เข้าสอบแต่ละคน มากนัก ถึงจะตัดสินผิดพลาดไปบ้างก็คงไม่เกิดผลเสียหายร้ายแรงเท่าใด ผลเสียก็จะมีแต่เพียงว่า ผู้เรียนอาจจะขาดเนื้อหาบางตอนที่ควรทำการทบทวนไป แต่ถ้าเราจะใช้ผลการสอบไปจัดอันดับ ความสามารถของนักเรียนในกลุ่ม ผลจากการวัดก็จะต้องมีความเชื่อมั่นสูงสุด

โดยทั่ว ๆ ไปแล้ว แบบทดสอบหรือเครื่องมืออื่น ๆ ที่ครูสร้างขึ้นมาใช้ในห้องเรียน ควรมีความเชื่อมั่นตั้งแต่ .60 ขึ้นไป แต่สำหรับแบบทดสอบมาตรฐานนั้น จะต้องมีค่าความ เชื่อมั่นตั้งแต่ .80 ขึ้นไปจนถึง 1.00 จึงจะจัดว่าเป็นแบบทดสอบที่ดี

#### 4.2 ความเที่ยงตรง (Validity)

เครื่องมือที่ดีต้องมีความเที่ยงตรง เครื่องมือที่มีความเที่ยงตรงสูง คือ เครื่องมือที่สามารถ วัดได้ตรงตามวัตถุประสงค์ที่ต้องการจะวัด ความเที่ยงตรงของเครื่องมือแบ่งออกเป็นชนิดใหญ่ ๆ ได้ 3 ชนิด คือ

- 1) ความเที่ยงตรงตามเนื้อหา
- 2) ความเที่ยงตรงตามเกณฑ์
- 3) ความเที่ยงตรงตามโครงสร้าง

##### 1) ความเที่ยงตรงตามเนื้อหา

การหาความเที่ยงตามเนื้อหามีจุดประสงค์เพื่อจะศึกษาดูว่าเครื่องมือนั้นสามารถ วัดเนื้อหาวิชาตามที่กำหนดไว้มากน้อยเพียงใด ดังนั้น โดยความหมายทั่ว ๆ ไปแล้ว ความเที่ยงตรง ตามเนื้อหาของแบบทดสอบหมายถึงความสามารถของแบบทดสอบฉบับหนึ่งที่จะมีข้อความที่เป็น ตัวแทนที่แท้จริงของเนื้อหาวิชาซึ่งกำหนดไว้ในหลักสูตร แบบทดสอบที่ควรมีความเที่ยงตรงตาม เนื้อหาวิชาสูงนั้น ควรเป็นแบบทดสอบประเภทวัดผลสัมฤทธิ์ทางการเรียน

## 2) ความเที่ยงตรงตามเกณฑ์

การหาความเที่ยงตรงตามเกณฑ์เน้นหาเฉพาะของแบบทดสอบ มีจุดประสงค์ที่จะดูว่าคะแนนที่ได้จากการทดสอบแบบทดสอบฉบับนั้นสามารถนำมาใช้พยากรณ์ผลการเรียนภาคหน้าหรือนำมาคาดคะเนผลการเรียนในปัจจุบันได้ดีเพียงใด ความเที่ยงตรงตามเกณฑ์มีอยู่ 2 ชนิด คือ ความเที่ยงตรงตามพยากรณ์ (Predictive Validity) กับความเที่ยงตรงตามสภาพ (Concurrent Validity) การหาความเที่ยงตรงทั้ง 2 ชนิดนี้ เป็นการหาความสัมพันธ์ระหว่างคะแนนของแบบทดสอบที่จะหาความเที่ยงตรงกับตัวเกณฑ์ (Criteria) ในการหาความเที่ยงตรงตามเกณฑ์นั้นตัวเกณฑ์ที่จะนำมาใช้เปรียบเทียบคือ คะแนนที่ได้จากผลการเรียน ซึ่งจะต้องติดตามไปในอนาคต ส่วนการหาความเที่ยงตรงตามสภาพนั้น ตัวเกณฑ์จะได้อาจจากการที่ครูจัดอันดับความสามารถที่นักเรียนแสดงออกในห้องเรียนให้ครูสังเกตเห็นได้ ถ้าข้อสอบมีความเที่ยงตรงตามสภาพสูงก็แสดงว่าคะแนนที่ได้จากแบบทดสอบนั้นสอดคล้องไปในทางเดียวกันกับอันดับที่ครูจัด นั่นคือถ้าเขาได้คะแนนจากแบบทดสอบสูงก็ต้องได้อันดับดีด้วย โดยปกติแล้วความเที่ยงตรงทั้ง 2 ชนิดนี้มีหลักการและความหมายเป็นแบบเดียวกัน แต่จะต่างกันตรงระยะเวลาที่จะได้ตัวเกณฑ์มาเปรียบเทียบ

## 3) ความเที่ยงตรงตามโครงสร้าง

ความเที่ยงตรง 2 แบบที่กล่าวมาข้างต้น เกี่ยวข้องโดยตรงกับการใช้ผลจากการทดสอบ ความเที่ยงตรงตามเนื้อหาจะบอกให้ทราบว่าผลที่ได้จากการทดสอบจะเป็นตัวแทนความสัมฤทธิ์ผลในแต่ละวิชาได้ดีเพียงไร ความเที่ยงตรงตามเกณฑ์จะบ่งให้ทราบว่าพยากรณ์หรือคาดคะเนผลการเรียนได้ถูกต้องเพียงไร แต่ความเที่ยงตรงตามโครงสร้างจะบ่งถึงคุณลักษณะบางประการทางจิตวิทยา เช่น สมรรถภาพสมองทางด้านต่าง ๆ เช่น สมรรถภาพทางด้านความมีเหตุผล ความสามารถในการนำทฤษฎีไปใช้ ความสามารถในการวิเคราะห์หรือพฤติกรรมอื่น ๆ ที่ต้องการให้นักเรียนมีหลังจากที่ครูสอนวิชานั้น ๆ ไปแล้ว ดังนั้นข้อสอบวัดผลสัมฤทธิ์ที่มีความเที่ยงตรงตามโครงสร้าง จะเป็นข้อสอบที่วัดสมรรถภาพสมองด้านต่าง ๆ ที่กำหนดไว้ในตารางวิเคราะห์หลักสูตรด้านพฤติกรรมนั่นเอง

### 4.3 ความเป็นปรนัย (Objectivity)

ก่อนอื่นต้องขอทำความเข้าใจเสียก่อนว่า เครื่องมือที่มีความเป็นปรนัยนั้นจะต้องมีคุณสมบัติ 3 อย่าง คือ

- 1) ผู้สอบอ่านโจทก์แล้วต้องเข้าใจคำถามตรงกัน
- 2) การตรวจกระดาษคำตอบนั้น ไม่ว่าจะให้ใครตรวจก็ต้องให้คะแนนเท่ากันทุกคน
- 3) ต้องสามารถนำคะแนนมาแปลความหมายเป็นรูปแบบเดียวกันได้

เครื่องมือใดก็ตามถ้ามีคุณสมบัติครบ 3 ประการนี้จึงจะเรียกว่าเป็นเครื่องมือที่มีความเป็นปรนัย โปรดอย่านึกว่าแบบทดสอบที่มีความเป็นปรนัยหมายถึง ข้อสอบแบบถูก - ผิด จับคู่ เติมคำ หรือเลือกตอบเลยทีเดียว เพราะนั่นเป็นแต่เพียงวิธีการ หรือขบวนการเขียนข้อคำถามซึ่งจะนำข้อสอบไปสู่ความเป็นปรนัยเท่านั้น

### 4.4 ความยากง่าย (Difficulty)

แบบทดสอบที่ดีต้องมีความยากง่ายพอเหมาะ เรื่องนี้มีความสำคัญมาก เพราะคะแนนที่นักเรียนสอบได้นั้นขึ้นอยู่กับความยากง่ายของข้อสอบเป็นสำคัญ มิได้ขึ้นอยู่กับจำนวนข้อคำถามหรือจำนวนเวลาที่กำหนดให้ทำการสอบมากนัก ลงข้อสอบยากแล้วต่อให้มีข้อสอบเป็นร้อย ๆ พัน ๆ ข้อ นักเรียนก็ได้คะแนนต่ำอยู่นั่นเอง เพราะทำข้อสอบไม่ได้ ข้อสอบที่ดีนั้นจะต้องมีคำถามที่ผู้สอบตอบถูกบ้างผิดบ้างสลับกันไป ฉะนั้น ข้อสอบที่ยากที่สุดหรือง่ายที่สุดจึงไม่ประโยชน์ เพราะถ้าข้อสอบยากที่สุด ผู้สอบทุกคนก็จะได้คะแนนศูนย์ หรือถ้าข้อสอบง่ายที่สุด ผู้สอบก็จะได้คะแนนเต็มเท่ากันทุกคน เราเลยไม่สามารถแยกได้ว่าใครเก่งกว่ากัน

ข้อสอบที่ใช้ในการสอบรวม หรือสอบประจำภาคเรียน ควรจะมีอัตราส่วนความยากของข้อคำถามเป็นอัตราส่วน 1 : 2 : 1 คือมีข้อสอบที่ง่าย ๆ 1 ส่วน ข้อสอบที่ยาก ๆ 1 ส่วน และข้อสอบที่ไม่ง่ายไปไม่ยากไปอีก 2 ส่วน ดังนั้น ถ้าข้อสอบฉบับหนึ่งมี 100 ข้อ ก็จะมีข้อสอบง่าย ๆ ประมาณ 25 ข้อ ข้อสอบยากง่ายปานกลางประมาณ 50 ข้อ และมีข้อสอบยาก ๆ อีกประมาณ 25 ข้อ จึงจะดี ความยากง่ายของข้อคำถามเราใช้ตัว  $p$  เป็นสัญลักษณ์แทน ถ้า  $p$  เท่ากับ .50 หมายถึงข้อสอบมีความยากง่ายพอเหมาะดี แต่ถ้า  $p$  น้อยกว่า .50 แปลว่าข้อสอบยาก  $p$  มากกว่า .50 แปลว่าข้อสอบง่าย เราอาจจะท่องเป็นสูตรว่า “น้อยยากมากง่าย” ก็ได้

การเรียงข้อคำถามในแบบทดสอบนั้น โดยทั่ว ๆ ไปแล้วนิยมเรียงจากข้อง่ายไปหาข้อยาก คือเอาข้อง่ายเป็นข้อ 1 แล้วค่อย ๆ ยากขึ้นไปตามลำดับ อย่างไรก็ตามก็ยังมีวิธีเรียงข้อคำถามก็มีหลายวิธี อาจจะสรุปได้เป็นวิธีใหญ่ ๆ ได้ 4 วิธี คือ

### 1. จัดเรียงตามลำดับความยากง่าย

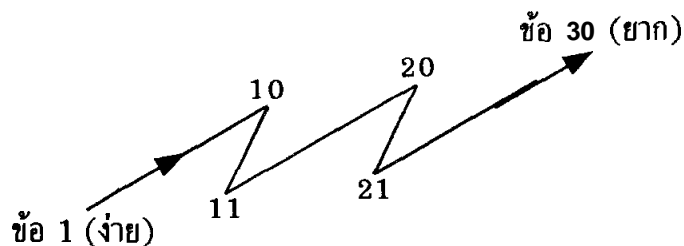
วิธีนี้เป็นการจัดเรียงข้อแบบธรรมดาทั่ว ๆ ไป คือเรียงจากข้อง่ายไปหายาก ประโยชน์ส่วนใหญ่ของการเรียงข้อสอบวิธีนี้ก็คือ ผู้สอบจะได้พบข้อสอบง่าย ๆ ก่อนทำให้เกิดกำลังใจ จึงไม่เกิดความเหนื่อยมากนักเมื่อมาพบข้อสอบยาก ๆ แต่การเรียงข้อสอบแบบนี้ก็มีข้อเสีย 2 ประการ คือ

ก. การเรียงข้อสอบแบบนี้ ทำให้ข้อสอบที่มีเนื้อหาเดียวกันต้องกระจายออกไปตามลำดับความยากง่าย ซึ่งมีผลทำให้ดวงความคิดของผู้เข้าสอบต้องเปลี่ยนแปลงอยู่เสมอ เช่น เดียวคิดเกี่ยวกับเลขบวก ก็ต้องเปลี่ยนมาคิดเลขคูณ แล้วไปคิดเลขลบ แล้วกลับมาคิดเลขคูณอีก ดังนี้ เป็นต้น

ข. การเรียงข้อสอบแบบนี้ มีผลเสียตรงที่ถ้านักเรียนพบข้อยาก ๆ เสียแล้วในตอนต้น ๆ ก็จะทำให้เกิดความท้อใจที่จะทำข้อต่อไป เนื่องจากนักเรียนรู้ว่าข้อสอบจะยิ่งยากขึ้นไปอีก

### 2. จัดเรียงความยากง่ายแบบไซคลิกออร์เดอร์ (Cyclic Order)

วิธีเรียงข้อสอบแบบนี้ต่างจากแบบที่ 1 ตรงที่แทนที่จะเรียงข้อสอบจากง่ายไปหายาก ที่เดียว ก็จะมีการลดความยากง่ายเป็นขั้น ๆ บั่นได้ เช่น จากข้อ 1-10 เรียงจากง่ายไปหายาก พอข้อ 11-14 แทนที่จะยากขึ้นไปอีก กลับลดความยากลงมาเท่ากับข้อ 6-10 พอดี ถึงข้อ 15 ก็จะยากกว่าข้อ 10 ดังรูป



การเรียงข้อสอบแบบนี้ดีตรงที่จะเป็นการกระตุ้นให้ผู้ทำข้อสอบให้หมดทุกข้อ เพราะเขารู้ว่าทำข้อสอบในระยะหนึ่งก็พบข้อง่าย ๆ อีก วิธีนี้ถึงจะดี แต่ก็ยังเสี่ยงตรงที่ผู้สอบก็ยังคงเปลี่ยนดวงความคิดในการแก้ปัญหาเร็วเกินไป

### 3. จัดเรียงตามกลุ่มของเนื้อหาวิชา

การเรียงข้อสอบโดยวิธีนี้เป็นการเรียงโดยรวมข้อสอบที่มีเนื้อหาเดียวกันไว้ด้วยกัน แล้วจัดเรียงข้อตามลำดับความยากง่าย ซึ่งวิธีนี้จะช่วยให้ผู้สอบใช้ความคิดในการแก้ปัญหาหนึ่งก่อนที่จะเปลี่ยนไปคิดเนื้อหาอื่น

### 4. จัดเรียงตามจุดมุ่งหมายของการวัด

ในการจัดเรียงข้อสอบผู้สร้างแบบทดสอบบางคนนิยมจัดเรียงข้อโดยรวมพฤติกรรมที่ต้องการจะวัดประเภทเดียวกันเข้าไว้ด้วยกัน เช่น รวมข้อคำถามเกี่ยวกับความรู้ในเรื่องเรื่องไว้พวกหนึ่ง การเรียงข้อสอบด้วยวิธีนี้มีประโยชน์สำหรับครูมาก ในการที่จะรู้ว่าสามารถวัดได้ตรงตามจุดมุ่งหมายที่ต้องการหรือไม่ แต่มีข้อเสียคือ บางที่ครูไม่สามารถออกข้อสอบได้ตรงตามพฤติกรรมที่ต้องการจะวัด

## 4.5 อำนาจจำแนก (Discrimination)

เครื่องมือที่ดีต้องมีอำนาจจำแนกสูง หมายความว่าเครื่องมือนั้นต้องมีคุณสมบัติในการแจกแจงผู้ตอบออกเป็นประเภทๆ ตามความสามารถของเขาได้ ตั้งแต่เก่งสุดจนถึงอ่อนสุด เห็นด้วยมากที่สุดจนถึงไม่เห็นด้วย แบบทดสอบที่มีอำนาจจำแนกสูงคือ แบบทดสอบที่เด็กเก่งเท่านั้นที่จะตอบถูก ส่วนเด็กอ่อนตอบที่ไรก็ผิดทุกที ถึงจะเดาก็ตอบไม่ถูกข้อสอบใดสามารถเดาได้ ถูกข้อสอบนั้นก็ไม่มีอำนาจจำแนกจึงเป็นข้อสอบที่ดีไม่ได้ ดังนั้นที่มีเสียงพูดกันว่าข้อสอบปรนัยไม่ดีเพราะเดาได้ เผลอๆ บางทีก็พูดกันถึงขนาดว่าให้ลิงขีดตอบก็ได้มันไม่จริง เพราะข้อสอบปรนัยที่เขียนโดยครูที่มีความรู้ทางการวัดผลนั้น ท่านไม่สามารถเดาได้ถูกเลย ข้อสอบฉบับไหนที่เดาได้ ก็แปลว่าเป็นข้อสอบที่ไม่ดี ควรเก็บเข้าลิ้นชักได้แล้ว ไม่ควรเอาออกมาใช้อีก ดังนั้นอำนาจจำแนกจึงเป็นคุณสมบัติที่สำคัญอีกประการหนึ่งของแบบทดสอบ เราใช้ตัว  $r$  เป็นสัญลักษณ์แทนค่าอำนาจจำแนกของข้อสอบที่มีค่า  $r$  ตั้งแต่ .20 ถึง 1.00 ขึ้นไปจึงจะถือว่าเป็นข้อสอบที่ใช้ได้

## คำถามท้ายบทที่ 4

1. เกณฑ์ที่ใช้สำหรับพิจารณาว่าเครื่องมือที่ใช้ในการประเมินเด็กปฐมวัย มีอะไรบ้าง จงอธิบาย
2. ถ้าเครื่องมือชนิดหนึ่งมีค่าความเชื่อมั่น 0.10 หมายความว่าอย่างไร
3. ความเที่ยงตรงตามสภาพ และความเที่ยงตรงตามพยากรณ์แตกต่างกันอย่างไร
4. ถ้าแบบทดสอบฉบับหนึ่งมีอำนาจจำแนกสูงหมายความว่าอย่างไร จงอธิบายพร้อมยกตัวอย่างประกอบ

