

## บทที่ 1

# คุณภาพและการตีความหมาย

## I จุดประสงค์การเรียนรู้

หลังจากจบบทนี้แล้ว นักศึกษาสามารถปฏิบัติได้ดังนี้

1. บอกความหมายของคุณภาพ
2. อธิบายความแตกต่างระหว่างการวัดในแต่ละระดับ
3. อธิบายความหมายของคุณภาพจริงและความคลาดเคลื่อนได้
4. อธิบายทฤษฎีการทดสอบในแต่ละทฤษฎี
5. เปรียบเทียบการตีความหมายคุณภาพและการทดสอบในแต่ละแบบ

## II รายละเอียดของเนื้อหา

### 1. ความหมายของคุณภาพ

มีผู้ให้ความหมายของคุณภาพไว้หลาย ๆ คน เช่น

アナ斯塔ซี (Anastasi; 1968:39) กล่าวว่า “คุณภาพดิบในการทดสอบนั้นไม่มีความหมายอะไรเลย เช่น ถ้าเด็กคนหนึ่งทำข้อสอบได้ 15 คะแนน ก็หมายความเพียงแค่นั้น ไม่อาจจะมาตีความหมายเป็นอย่างอื่นได้ แม้แต่จะนำมาแปลเป็นค่าว้อยลง เช่นจะบอกว่าทำได้ 65% ก็ยังบอกไม่ได้ เพราะการทำข้อสอบ 65% ของฉบับหนึ่งอาจจะมีค่าเท่ากับ 30% ของอีกฉบับหนึ่ง หรือเท่ากับ 80% ของอีกฉบับหนึ่งก็ได้ การตีความหมายจึงจำเป็นจะต้องขึ้นอยู่กับระดับความยากง่ายของข้อสอบแต่ละข้ออีกด้วย”

ไลแมน (Lyman; 1971:6) ให้ความหมายของคุณภาพไว้ว่า “คุณภาพดิบคือจำนวนข้อที่ตอบถูก คุณภาพดิบจะมีมากหรือน้อยเพียงใดຍ่อมขึ้นอยู่กับจำนวนข้อสอบ และความยากง่ายของข้อสอบ ดังนั้นตัวคุณภาพดิบเองจึงไม่มีความหมายอะไรเลย แต่เนื่องจากคุณภาพดิบนั้นเรานำมาใช้เป็นพื้นฐานของคุณภาพอื่น ๆ หลายชนิด เป็นคุณภาพที่ถูกต้อง เชื่อถือได้ หรือจากล่า

“ได้ว่าคะแนนดิบเป็นพื้นฐานของคะแนนที่ไม่มีคะแนนอื่นใดจะถูกต้องและแม่นยำมากไปกว่าคะแนนนี้”

บราวน์ (Brown; 1980:122) ให้ความหมายของคะแนนดิบว่า “เป็นปริมาณค่าที่เกิดจากการแสดงความสามารถในการทำแบบทดสอบสเกลต่าง ๆ หรือกลุ่มของคำถาม หรืออาจจะหมายถึงคะแนนที่นับได้จากการตอบแบบทดสอบได้ถูกต้อง หรือหมายถึงผลรวมของแต้มที่ได้รับ”

จากความหมายของคะแนนที่มีผู้กล่าวไว้นี้ จะเห็นได้ว่าตัวคะแนนดิบจริง ๆ แล้วออกได้เพียงจำนวนนับของพฤติกรรมที่เราสังเกตได้เท่านั้น แต่อย่างไรก็ตามคะแนนดิบก็เป็นสิ่งสำคัญที่เราจะต้องพยายามทำให้มีคุณภาพและเชื่อถือได้มากที่สุดซึ่งจะเป็นผลให้การตีความหมายคะแนนเกิดความผิดพลาดน้อยที่สุด

## 2. ระดับของการวัด (Level of Measurement)

ในการนำคะแนนดิบมาตีความหมายนั้นเราจำเป็นจะต้องอาศัยหลักทางสถิติเข้ามาช่วย ดังนั้นก่อนที่เราจะใช้สถิติอะไรก็จะต้องรู้พื้นฐานของคะแนนก่อนว่ามีคุณสมบัติของการวัดในระดับใด ระดับการวัดซึ่งมีความแตกต่างกันอยู่ 4 ระดับนั้นแต่ละระดับจะมีลักษณะต่อเนื่องกัน ถ้าเราสามารถรู้ได้ว่าคะแนนที่เราจะนำมาตีความหมายนั้นอยู่ในระดับการวัดใดการเลือกใช้สถิติตามระดับการวัดก็จะถูกต้อง ความแตกต่างของระดับการวัดนั้นอยู่ที่ข้อจำกัดของข้อมูล ที่จะนำเอารวบกับทางคณิตศาสตร์มาคำนวณ เช่น การบวก ลบ คูณ หารนั้นเราจะนำมาใช้กับข้อมูลทุกประเภทไม่ได้ ตัวอย่างที่เรามักจะใช้กันผิดพลาดเสมอ เช่น บอกว่าคนที่มี I.Q. 140 จะมีความสามารถทางสถิติปัญญาสูงกว่าคนที่มี I.Q. 70 อยู่ 2 เท่ากัน ซึ่งความจริงเรามิ่งสามารถจะกล่าวเช่นนั้นได้ ตัวอย่างนี้จึงเป็นตัวอย่างที่แสดงให้เห็นการใช้ข้อมูลตีความหมายผิดพลาด จากความเป็นจริง

ระดับของการวัดทั้ง 4 ระดับ มีดังนี้คือ

1. มาตรานามบัญญัติ (Nominal Scale) เป็นระดับการวัดที่ต่ำที่สุด ง่ายที่สุด ข้อมูลที่อยู่ในการวัดระดับนี้เป็นเพียงการระบุชื่อ เพศ เช่นเวลาเราจะลงรหัสในคอมพิวเตอร์ กำหนดเพศชาย เป็น 1 เพศหญิง เป็น 2 หรือกลุ่มเก่งเป็นเลข 1 กลุ่มปานกลางเป็นเลข 2 และกลุ่มอ่อน เป็นเลข 3 การกำหนดสัญลักษณ์หรือหมายเลขอ้างกับค่าวัดระดับนี้จึงทำได้ง่าย แต่เราจะตัวเลขเหล่านี้มาบวก ลบ คูณ หารกันไม่ได้ จะทำได้ก็เพียงนับจำนวนความถี่เท่านั้น

2. มาตราจัดอันดับ (Ordinal Scale) ระดับการวัดระดับที่สองนี้คือ การจัดอันดับตามคุณลักษณะบางอย่างที่วัดได้ แต่ยังไม่สามารถบอกอันดับของห่างระหว่างระดับแต่ละระดับได้ชัดเจน

เช่นเรียงลำดับนักเรียนตามความถี่ที่เข้าร่วมกิจกรรม การให้ความร่วมมือ ฯลฯ สกัดที่จะนำมาใช้ในการวัดระดับนี้ยังมีข้อจำกัด และมักจะมีผู้นำข้อมูลระดับนี้ไปใช้กันผิด ๆ เช่น ซึ่งเป็นผลให้การตีความหมายผิดพลาดไปด้วย

**3. มาตรวัดแบบอันตรภาค (Interval Scale)** ข้อมูลในระดับนี้มีช่วงห่างของคะแนนเท่ากัน แต่ข้อมูลระดับนี้ก็ยังไม่รู้ค่าศูนย์ที่แท้จริง (absolute zero) ของคะแนนชุดนั้น คะแนนที่ได้จากการวัดจากแบบทดสอบส่วนใหญ่จะมีระดับคะแนนอยู่ในแบบอันตรภาค ตัวอย่างเช่น เด็ก 2 คน ได้คะแนน 30 และ 60 (หมายความว่าทำข้อสอบถูก 30 ข้อ และ 60 ข้อ ตามลำดับ) จากคะแนนของเด็กคนนี้เราจะบอกว่าเด็กคนหนึ่งทำข้อสอบถูกมากกว่าเด็กคนหนึ่งอยู่ 30 ข้อ แต่เราไม่อาจกล่าวได้ว่าเด็กคนหนึ่งรู้เป็น 2 เท่าของเด็กคนหนึ่ง ทั้งนี้เพราะเรามิรู้ค่าศูนย์แท้จริงของคะแนนชุดนี้ หรือเรามิอาจกล่าวได้ว่าคนที่ทำข้อสอบผิดหมวดไม่มีความรู้อะไรเลย จะบอกได้ก็เพียงว่าเขาทำข้อสอบผิดหมวด สมมติว่าจากข้อสอบชุดนี้เราเพิ่มข้อสอบที่ง่าย ๆ เข้าไปอีก 30 ข้อ และข้อสอบ 30 ข้อนี้จะทำให้ทุกคนทำข้อสอบถูกหมวด คะแนนของเด็กคนหนึ่งก็จะเป็น 60 และ 90 คะแนน ซึ่งก็แสดงว่าเราจะบอกว่าเด็กคนที่เคยได้ 60 คะแนนรู้เป็น 2 เท่าของคนที่ได้ 30 คะแนนไม่ได้ชัดเจนขึ้น

การใช้สถิติสำหรับการวัดระดับนี้สามารถใช้สถิติที่ใช้กับระดับ 1 และระดับ 2 ที่ต่ำกว่าได้ นอกจากนี้ยังสามารถเพิ่มการบวกลบเข้าไปได้ แต่ก็ยังคุณไม่ได้ เพราะฉะนั้นคะแนนจากการสอบตามตัวอย่างที่กล่าวมาแล้ว ถ้าสมมติว่าคำถามแต่ละข้อมีค่าเท่า ๆ กัน เราถึงสามารถนำมารวบรวมแล้วหาค่าเฉลี่ยออกมาได้ ดังนั้นการจะนำคะแนนดิบแปลงเป็นคะแนนอื่น ๆ จึงต้องกระทำการเฉพาะกับคะแนนที่มีระดับการวัดอยู่ในระดับอันตรภาคชั้นขึ้นไป แต่ถ้าเราไปเอาคะแนนที่มีระดับการวัดต่ำกว่านี้มาบวกกับคุณหารกันโดยสมมติว่าคะแนนเหล่านั้นอยู่ในระดับการวัดแบบอันตรภาคก็จะทำให้การตีความหมายคะแนนผิดพลาดไปได้ ถ้าตามข้อเท็จจริงแล้วข้อมูลที่ได้เป็นเพียงมาตราจัดอันดับหรือนามบัญญัติเท่านั้น

**4. มาตรอัตราส่วน (Ratio Scale)** เป็นระดับการวัดที่สูงสุด ที่มีคุณสมบัติครบถ้วน ตามแบบมาตรฐานนามบัญญัติ จัดลำดับ และอันตรภาค ซึ่งนอกจากจะสามารถนำมารวบรวมกันแล้วยังสามารถนำมาคูณหารกันได้ เพราะการวัดในระดับนี้เรารู้ค่าศูนย์ที่แท้จริงของคะแนน เช่นการวัดน้ำหนัก ส่วนสูง เวลา ฯลฯ ซึ่งข้อมูลเหล่านี้ต่างก็มีค่าศูนย์แท้จริง จึงสามารถนำมารวบ ลบ คูณ หาร กันได้ และมักจะพบเสมอว่า ถ้าแม้ข้อมูลที่ได้จากการวัดจะมีคุณสมบัติในระดับนี้ แต่ก็ยังใช้สถิติเช่นเดียวกับการวัดระดับ 1 และระดับ 2 เช่นนำความสูงนักเรียนมาแบ่งเป็นกลุ่มสูงและกลุ่มต่ำ หรือกลุ่มสูงมาก สูงปานกลาง และต่ำ เป็นต้น

ข้อมูลที่เราได้จากการวัดนั้นมองผิวน ๆ อาจจะคิดว่าสามารถบอกระดับการวัดได้ง่าย แต่ที่จริงแล้วไม่ใช่เรื่องง่ายเลย ดังนั้นถ้าเราเกิดสงสัยว่าจะใช้การวัดระดับใดก็ลองใช้สถิติขั้นสูง ๆ ดูก่อนถ้าผลออกมากขัดแย้งกับสิ่งที่ควรจะเป็นก็ค่อย ๆ ลดระดับลงมา ข้อโต้เถียงที่ว่า ข้อมูลจะเป็นระดับอันตรภาคหรือระดับอัตราส่วนนั้nm ก็จะไม่ค่อยมีปัญหาในการวัดทางการศึกษามากนัก เพราะส่วนมากมักจะสมมติว่าข้อมูลทางการศึกษาเป็นข้อมูลระดับอันตรภาค ซึ่งก็สามารถนำมาวิเคราะห์สิ่งที่ต้องการได้แล้ว ถ้าเรายังสงสัยว่าไม่น่าจะใช้สถิติถึงมาตรฐานวัดอัตราส่วนก็ควรจะหลีกเลี่ยงเสีย

### 3. คะแนนจริง คะแนนความคลาดเคลื่อน และทฤษฎีการทดสอบ

(True Scores, Error Scores and Test Theory)

ในการสอบแต่ละครั้งผลที่ได้ออกมาคือจำนวนข้อที่ตอบถูก จำนวนชั้นของงานที่ทำได้ ฯลฯ ค่าที่วัดได้นี้เรียกว่าคะแนนดิบ (Raw Score) คะแนนดังกล่าววนนี้ แบ่งเป็น 2 ส่วนคือ

1. คะแนนจริง (True Score) หมายถึงคะแนนที่แสดงความสามารถหรือความรู้ที่แท้จริงของผู้สอบ ถ้าเอาข้อสอบฉบับเดิมไปสอบกับผู้สอบคนเดิม คะแนนที่ได้ในแต่ละครั้งจะคงที่เสมอ

2. คะแนนความคลาดเคลื่อน (Error Score) หมายถึงคะแนนที่เกิดจากความผิดพลาดต่าง ๆ โดยบังเอิญ เป็นความคลาดเคลื่อนอย่างไม่มีระบบ ไม่สามารถป้องกันได้ เป็นความคลาดเคลื่อนที่ทำให้ผู้สอบทำข้อสอบได้ถูกต้องทั้ง ๆ ที่ไม่มีความรู้จริง หรือผู้สอบทำข้อสอบผิดพลาดทั้ง ๆ ที่สามารถตอบข้อสอบได้ถูกต้อง คะแนนความคลาดเคลื่อน มีอยู่ 2 ประเภท คือ

2.1 ความคลาดเคลื่อนคงที่ (constant error) ความคลาดเคลื่อนชนิดนี้หมายถึงผู้สอบทุก ๆ คนจะได้รับผลกระทบเหมือน ๆ กัน เช่นครูเพิ่มเวลาทำข้อสอบให้ 5 นาที ทุกคนที่เข้าสอบก็จะได้รับผลนี้เหมือนกัน

2.2 ความคลาดเคลื่อนแบบสุ่ม (random error) เป็นความคลาดเคลื่อนที่แต่ละคนจะได้รับไม่เท่ากัน ซึ่งอาจจะมีผลทั้งทางบวกและทางลบ เช่น สุภาพของผู้สอบขณะสอบ สภาพจิตใจก่อนเข้าสอบ ฯลฯ

คะแนนที่ผู้สอบทำได้นั้นเป็นสิ่งที่เราสามารถเปลี่ยน แต่ตัดสินใจ คะแนนดังกล่าวประกอบด้วยคะแนนจริงและคะแนนความคลาดเคลื่อน ซึ่งก็หมายความว่าคะแนนของแต่ละคนอาจจะมากกว่าหรือน้อยกว่าคะแนนจริงที่เข้าควรจะได้ คะแนนดังกล่าวเราจึงนำมาเขียนในรูปของสมการได้ดังนี้

$$X = T + E$$

เมื่อ

$$X = \text{คะแนนดิบ}$$

$$T = \text{คะแนนจริง}$$

$$E = \text{คะแนนความคลาดเคลื่อน}$$

คะแนนความคลาดเคลื่อนนี้ มีข้อตกลงเบื้องต้นว่า

1. ค่าเฉลี่ยของคะแนนความคลาดเคลื่อนมีค่าเป็นศูนย์

$$M_E = 0$$

**เหตุผล** เนื่องจากคะแนนความคลาดเคลื่อนอาจจะมีต่าง ๆ กันมากบ้างน้อยบ้าง เป็นบวกบ้างเป็นลบบ้าง เมื่อนำผลของความคลาดเคลื่อนเหล่านี้มารวมกัน แล้วหาค่ารายเฉลี่ย จึงมีค่า = 0

2. ความสัมพันธ์ระหว่างคะแนนจริงและคะแนนความคลาดเคลื่อนเท่ากับศูนย์

$$r_{TE} = 0$$

**เหตุผล** การที่ผู้เข้าสอบแต่ละคนมีความคลาดเคลื่อนของคะแนนเท่ากับเท่าไรนั้นเราไม่อาจวัดได้ คนที่ได้คะแนนจริงสูงไม่จำเป็นว่าจะมีความคลาดเคลื่อนของคะแนนสูงด้วย ในขณะเดียวกันคนที่ได้คะแนนจริงต่ำไม่จำเป็นว่าจะได้คะแนนความคลาดเคลื่อนต่ำไปด้วย ดังนั้นความสัมพันธ์ระหว่างคะแนนจริงและคะแนนความคลาดเคลื่อนจึงไม่เกี่ยวข้องกันค่าสหสัมพันธ์จึงเป็นศูนย์

3. ความคลาดเคลื่อนของคะแนนจากแบบทดสอบ 2 ฉบับที่เป็นคู่ขานกันมีค่าเท่ากับศูนย์

$$r_{E_1 E_2} = 0$$

**เหตุผล** เนื่องจากความคลาดเคลื่อนของแบบทดสอบฉบับที่ 1 เกิดขึ้นอย่างไม่มีระบบ เช่นเดียวกับคะแนนความคลาดเคลื่อนของแบบทดสอบฉบับที่ 2 ซึ่งวัดเรื่องเดียวกันก็มีความคลาดเคลื่อนไม่มีระบบเหมือนกัน ค่าสหสัมพันธ์ของคะแนนในแบบทดสอบทั้ง 2 ฉบับจึงไม่เกี่ยวข้องกัน ค่าสหสัมพันธ์จึงมีค่าเป็นศูนย์

## ทฤษฎีการทดสอบ

จากความคลาดเคลื่อนของคะแนนและข้อตกลงเบื้องต้นของความคลาดเคลื่อน จึงกำหนดเป็นทฤษฎีดังนี้

**ทฤษฎีที่ 1** รายเฉลี่ยของคะแนนจริงเท่ากับคะแนนดิบ ซึ่งหมายความว่า จากการสอบในครั้งหนึ่ง ๆ ค่าเฉลี่ยของคะแนนดิบที่ผู้เข้าสอบทำได้ ถือเป็นค่าเฉลี่ยของคะแนนจริงของผู้เข้าสอบ ซึ่งเขียนเป็นสมการได้ดังนี้

$$M_x = M_T$$

**ทฤษฎีที่ 2** ความแปรปรวนของคะแนนดิบเท่ากับผลรวมของความแปรปรวนของคะแนนจริงกับผลรวมของความแปรปรวนของคะแนนความคลาดเคลื่อน หรือความเบี่ยงเบนมาตรฐานของคะแนนดิบเท่ากับรากที่สองของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนความคลาดเคลื่อน

เขียนเป็นสมการได้ดังนี้

$$\begin{aligned} S_x^2 &= S_T^2 + S_E^2 \\ \text{หรือ} \quad S_x &= \sqrt{S_T^2 + S_E^2} \\ \text{เมื่อ} \quad S_x &= \text{ความเบี่ยงเบนมาตรฐานของคะแนนดิบ} \\ S_x^2 &= \text{ความแปรปรวนของคะแนนดิบ} \\ S_T^2 &= \text{ความแปรปรวนของคะแนนจริง} \\ S_E^2 &= \text{ความแปรปรวนของคะแนนความคลาดเคลื่อน} \end{aligned}$$

**ทฤษฎีที่ 3** ค่าเฉลี่ยของคะแนนจริงจากการสอบด้วยแบบทดสอบคู่ขนานกันย่อไม่เท่ากัน

$$\begin{aligned} M_{T_g} &= M_{T_h} \\ \text{เมื่อ} \quad M_{T_g} &= \text{คะแนนจริงของแบบทดสอบคู่ขนานฉบับ g} \\ M_{T_h} &= \text{คะแนนจริงของแบบทดสอบคู่ขนานฉบับ h} \end{aligned}$$

**ทฤษฎีที่ 4** ความเบี่ยงเบนมาตรฐานของคะแนนจริงจากการสอบด้วยแบบทดสอบคู่ขนานกันย่อไม่เท่ากัน

$$\begin{aligned} S_{T_g} &= S_{T_h} \\ \text{เมื่อ} \quad S_{T_g} &= \text{ความเบี่ยงเบนมาตรฐานของคะแนนจริง} \\ &\text{ของแบบทดสอบคู่ขนาน g} \\ S_{T_h} &= \text{ความเบี่ยงเบนมาตรฐานของคะแนนจริง} \\ &\text{ของแบบทดสอบคู่ขนาน h} \end{aligned}$$

**ทฤษฎีที่ 5** สมมติว่าระหว่างคะแนนจริงของแบบทดสอบคู่ขนาน มีค่าเท่ากับ 1

$$\begin{array}{lcl} r_{T_g T_h} & = & 1 \\ \text{เมื่อ} & & r_{T_g T_h} = \text{ค่าสมมติของคะแนนจริง} \\ & & \text{ของแบบทดสอบคู่ขนาน } g \text{ และ } h \end{array}$$

**ทฤษฎีที่ 6** รายเฉลี่ยของคะแนนดิบจากการสอบด้วยแบบทดสอบคู่ขนาน มีค่าเท่ากัน

$$\begin{array}{lcl} M_{X_g} & = & M_{X_h} \\ \text{เมื่อ} & & \\ M_{X_g} & = & \text{ค่าเฉลี่ยของคะแนนดิบจากแบบ} \\ & & \text{ทดสอบคู่ขนานฉบับ } g \\ M_{X_h} & = & \text{ค่าเฉลี่ยของคะแนนดิบจากแบบ} \\ & & \text{ทดสอบคู่ขนานฉบับ } h \end{array}$$

**ทฤษฎีที่ 7** ความแปรปรวนหรือความเบี่ยงเบนมาตรฐานของคะแนนดิบจากการสอบด้วยแบบทดสอบคู่ขนานมีค่าเท่ากับ

$$\begin{array}{lcl} S_{X_g}^2 & = & S_{X_h}^2 \\ \text{หรือ} & & S_{X_g} = S_{X_h} \\ \text{เมื่อ} & & \\ S_{X_g}^2 & = & \text{ความแปรปรวนของคะแนนดิบของ} \\ & & \text{แบบทดสอบคู่ขนานฉบับ } g \\ S_{X_h}^2 & = & \text{ความแปรปรวนของคะแนนดิบของ} \\ & & \text{แบบทดสอบคู่ขนานฉบับ } h \\ S_{X_g} & = & \text{ความเบี่ยงเบนมาตรฐานของคะแนนดิบของ} \\ & & \text{แบบทดสอบคู่ขนานฉบับ } g \\ S_{X_h} & = & \text{ความเบี่ยงเบนมาตรฐานของคะแนนดิบของ} \\ & & \text{แบบทดสอบคู่ขนานฉบับ } h \end{array}$$

**ทฤษฎีที่ 8** สมมติว่าระหว่างคะแนนดิบจากการสอบด้วยแบบทดสอบคู่ขนาน คือ อัตราส่วนของความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนดิบ

$$r_{X_g X_h} = \frac{S_T^2}{S_{X_g}^2}$$

เมื่อ

$r_{x_g x_h}$  = ค่าสหสัมพันธ์ของคะแนนดิบของแบบทดสอบคู่ขานาน  $g$  และ  $h$

$S_T^2$  = ความแปรปรวนของคะแนนจริง

$S_{x_g}^2$  = ความแปรปรวนของคะแนนดิบ

ทฤษฎีที่ ๑ ถ้ามีแบบทดสอบคู่ขานานหลายฉบับสหสัมพันธ์กันในระหว่างแบบทดสอบคู่ขานานเหล่านั้นย่อมเท่ากัน

$$r_{x_1 x_2} = r_{x_1 x_3} = r_{x_2 x_3} = r_{x_3 x_3} = \dots$$

เมื่อ

$r_{x_1 x_2}$  = ค่าสหสัมพันธ์ระหว่างคะแนนดิบของแบบทดสอบฉบับที่ ๑ และ ๒

$r_{x_1 x_3}$  = ค่าสหสัมพันธ์ระหว่างคะแนนดิบของแบบทดสอบฉบับที่ ๑ และ ๓

$r_{x_2 x_3}$  = ค่าสหสัมพันธ์ระหว่างคะแนนดิบของแบบทดสอบฉบับที่ ๒ และ ๓

$r_{x_3 x_4}$  = ค่าสหสัมพันธ์ระหว่างคะแนนดิบของแบบทดสอบฉบับที่ ๓ และ ๔

#### 4. รูปแบบการตีความหมายคะแนนการสอบ

(Types of Test Score Interpretation)

บราวน์ (Brown, 1980:74) ได้เสนอคะแนนการสอบไว้ ๓ รูปแบบคือ

1. คะแนนแบบอิงกลุ่ม (Norm - referenced scores)
2. คะแนนแบบอิงเนื้อหา (Content - referenced scores)
3. คะแนนแบบอิงเกณฑ์ (Criterion - referenced scores)

ซึ่งอธิบายรายละเอียดได้ดังนี้

1. คะแนนแบบอิงกลุ่ม เป็นคะแนนที่รู้จักและใช้กันแพร่หลาย การตีความหมายคะแนนชนิดนี้จะต้องเข้าไปเกี่ยวข้องกับคะแนนของคนอื่น ๆ ในกลุ่มเดียวกันด้วย เช่นคะแนนแบบเปอเซ็นต์ไทล์ ซึ่งค่าของคะแนนแบบนี้จะบอกถึงจำนวนร้อยละของคนที่ได้คะแนนต่ำกว่า

**2. คะแนนแบบอิงเนื้อหา** เป็นการตีความหมายคะแนนตามระดับของพฤติกรรมที่ตอบวัดได้จากแบบทดสอบวัดผลสัมฤทธิ์ คะแนนแบบนี้แตกต่างจากคะแนนแบบอิงกลุ่มคือกรอบที่นำมาตีความ นั่นคือคะแนนแบบอิงเนื้อหาใช้กรอบของเนื้อหาในแบบทดสอบ ในขณะที่คะแนนแบบอิงกลุ่มกรอบในการตีความหมายคะแนนคือคะแนนของผู้สอบคนอื่น ๆ เช่น บอกว่าผู้เรียนสามารถสะกดคำได้ถูกต้อง 75% ของคำมาตรฐานที่กำหนดไว้ การตีความหมายคะแนนแบบนี้ถือว่าเป็นการตีความหมายคะแนนแบบอิงเนื้อหา

**3. คะแนนแบบอิงเกณฑ์** วิธีการให้คะแนนแบบนี้ก็มีใช้กันอยู่บ้าง วิธีการคือนำคะแนนสอบไปตีความหมายกับตัวแปรภายนอก เช่น การตีความหมายคะแนนสอบในมหาวิทยาลัยว่าผู้สอบคนนี้ควรจะได้เกรด B หรือไม่ ระดับเกรด B เป็นค่าคะแนนเฉลี่ยของผู้ที่สมควรได้รับเกรดนั้น ดังนั้นการเอาคะแนนไปตีความหมายกับเกณฑ์ภายนอกนี้ อาจจะต้องใช้ตารางพยากรณ์ (Expectancy Table) ซึ่งก็เป็นตัวอย่างการให้คะแนนแบบนี้ หรือเป็นการนำคะแนนไปเปรียบเทียบกับเกณฑ์ที่โรงเรียนกำหนดว่า คนที่จะได้เกรด A จะต้องทำคะแนนได้ 80% ขึ้นไป เป็นต้น

คะแนนทั้ง 3 แบบนี้ถึงแม้จะมีการตีความหมายแตกต่างกันไป แต่ในตัวคะแนนจริง ๆ นั้นก็ยังเหมือนกันไม่มีอะไรต่างกัน เช่น คะแนนการสอบวิชาหนึ่งในระดับมหาวิทยาลัยของนักศึกษาคนหนึ่ง ถ้าเรานำคะแนนนั้นไปตีความหมายโดยเอาคะแนนของนักศึกษาคนนั้นไปเปรียบเทียบกับคะแนนของคนอื่น ๆ ก็เป็นการตีความหมายคะแนนแบบอิงกลุ่ม ถ้านำคะแนนนี้ไปตีความหมายว่านักศึกษาคนนี้มีความสามารถในระดับใด สำหรับการศึกษาในระดับปฐมวัย ตรีก็เป็นการตีความหมายคะแนนแบบอิงเนื้อหา แต่ถ้านำคะแนนไปเปรียบเทียบกับข้อมูลอื่นเพื่อทำนายว่าผู้เข้าสอบผู้นั้นจะประสบความสำเร็จตามเกณฑ์เพียงใด ก็เป็นการตีความหมายคะแนนแบบอิงเกณฑ์ ตัวอย่างที่ยกมาได้แสดงให้เห็นว่าที่จริงแล้วตัวคะแนนดังเดิมไม่มีอะไรต่างกัน แต่การตีความหมายทำให้คะแนนต่างกัน

การใช้คำว่าอิงเกณฑ์และอิงเนื้อหาเป็นสองครั้งอาจเกิดความสับสนได้แต่ก็มีผู้ใช้คำว่าอิงเกณฑ์ในการทดสอบกันหลายคน เช่นแกรเชอร์ โพแฟม (Glaser, 1963, Popham & Husek 1969) ทั้งนี้อาจเนื่องจากคำว่า “เกณฑ์” นั้นก็จะนึกไปถึงความเที่ยงตรง จึงมีการใช้คำแยกออกมา (Brown, 1980 : 75)

### การตีความหมายคะแนนแบบอิงกลุ่ม

โดยปกติคะแนนที่ได้จากการสอบทางการศึกษาและจิตวิทยามักจะใช้วิธีการตีความหมายคะแนนแบบอิงกลุ่ม คือนำคะแนนของผู้สอบไปเปรียบเทียบกับคนอื่น ๆ ซึ่งกลุ่มที่

เรานำไปเปรียบเทียบนี้เรียกว่า กลุ่มเกณฑ์ปกติ (Normal Group) การตีความหมายคะแนนแบบ อิงกลุ่มนี้จะนำมาใช้เมื่อต้องการจะเปรียบเทียบกัน เช่น การสอบเข้ามหาวิทยาลัย กลุ่มคนที่ เป็น Normal Group ก็คือคนที่สอบเข้ามหาวิทยาลัยในสาขาวิชาเดียวกัน เรายังสามารถของ ผู้เข้าสอบไปเปรียบเทียบกันว่าคร่าจะมีคะแนนอยู่ลำดับเท่าไร สูงต่ำกว่ากันอย่างไร เพื่อจะ ได้คัดเลือกคนที่มีคะแนนต่ำกว่าคนอื่น ๆ เข้าเรียน เป็นต้น การตัดสินใจที่จะใช้วิธีการตีความ หมายแบบอิงกลุ่ม ควรจะพิจารณาประเด็น 2 ประการคือ

1. ในบรรดากลุ่มเกณฑ์ปกติที่จะนำมาใช้เปรียบเทียบพฤติกรรมกันนั้นควรจะมีความ คล้ายคลึงกันในด้าน อายุ เพศ เชื้อชาติ ระดับการศึกษา ประเภทของโรงเรียน ภูมิหลังทาง สังคมและเศรษฐกิจ ภาษา ลักษณะชุมชน สภาพภูมิศาสตร์ และในบางครั้งอาจจำเป็นจะต้อง ให้นิยามผู้ที่จะนำมาเปรียบเทียบกันให้ชัดเจน เช่น การสอบคัดเลือกเข้ามหาวิทยาลัย ควรจะ ระบุให้ชัดว่าผู้ที่จะนำมาเปรียบเทียบกันนั้นคือผู้ที่สมัครสอบเข้ามหาวิทยาลัยเดียวกัน และ สาขาวิชาเอกเดียวกัน เป็นต้น

2. วิธีการรายงานผลการสอบ เป็นการแสดงผลของการเปรียบเทียบคะแนนของผู้ เข้าสอบกับกลุ่มเกณฑ์ปกติ การรายงานจึงควรทำในรูปของเปอร์เซนต์ไทล์ คะแนนมาตรฐาน ทุกชนิด รวมทั้ง Grade equivalent และพวก I.Q. ต่าง ๆ ด้วย

### การตีความหมายคะแนนแบบอิงเนื้อหา

คะแนนแบบอิงเนื้อหานี้เพื่อจะเริ่มนำมาใช้เมื่อไม่นานมานี้เองเป็นการตีความหมายโดย อ้างอิงถึงขอบเขต (domain) ของงานที่ให้ทำ รวมถึงทักษะที่จะต้องมีด้วย ดังนั้นความสำคัญ ของการตีความหมายคะแนนแบบนี้จึงขึ้นอยู่กับ คำอธิบายขอบเขตของเนื้อหาและทักษะที่นำ มาสร้างเป็นข้อสอบ การเขียนคำอธิบายขอบเขตนั้นจึงควรจะเขียนเป็นนิยามเชิงปฏิบัติการ (Operational terms) และถ้าจะใช้แบบทดสอบรอบรู้ (Mastery test) ด้วยแล้วก็จะต้องอธิบาย ถึงจุดตัดของคะแนนที่กำหนดให้ผู้เรียนจะต้องรอบรู้ไว้ด้วย พร้อมทั้งอธิบายเหตุผลในการ กำหนดจุดตัดนั้น ๆ รวมทั้งขบวนการที่ผู้ใช้แบบทดสอบจะต้องดำเนินการให้ผู้เรียนรอบรู้ถึง ระดับที่กำหนดก็จะต้องเขียนอธิบายไว้ด้วย

การกำหนดขอบเขตหรือมาตรฐานต่าง ๆ เหล่านี้อาจจะเขียนได้ง่าย แต่การนำไปใช้ ค่อนข้างจะลำบาก และยังไม่กราจั่งชัด ทั้งนี้เนื่องจากการให้นิยามขอบเขตของเนื้อหาและ ทักษะมักจะเขียนกันง่าย ๆ เช่น

ความสามารถในการอ่าน  
การแก้ปัญหาโจทย์คณิตศาสตร์

จากการเขียนข้างต้นนี้จะเห็นได้ว่าข้อความนี้สามารถเขียนคำถ้ามต่าง ๆ ได้มากmany เช่น ความสามารถในการอ่านหมายความว่าอย่างไร? เป็นการวัดอัตราเร็วในการอ่าน หรือ วัดความเข้าใจในการอ่าน? การอ่านแบบไหนที่ถือว่าถูกต้อง? จะอ่านภาษาไทยเงื่อนไขหรือ สถานการณ์อย่างไร? คำถ้ามต่าง ๆ เหล่านี้มีมากmany และไม่คงที่ จะเปลี่ยนไปตลอดเวลา ดังนั้น ปัญหาจึงอยู่ที่ว่าจะทำอย่างไรจึงจะเขียนข้อมูลของเนื้อหาและทักษะได้ชัดเจน เนพาะ เจาะจงไม่แคนจนขยายให้ความไม่ได้

การตีความหมายแบบอิงเนื้อหานี้ อาจทำได้ 2 วิธีคือ

1. อธิบายพฤติกรรมต่าง ๆ เป็นเชิงต่อเนื่องกัน เช่นเป็นร้อยละ
2. ใช้มาตรฐานเด็ดสินความรอบรู้ของผู้สอบไว้ล่วงหน้าว่าป่างน้อยที่สุดผู้สอบควรจะได้คะแนนเท่าใดจึงจะยอมรับว่ารอบรู้ในเนื้อหาและทักษะแล้ว ตัวอย่างเช่นนักเรียนคนหนึ่ง สามารถ

- แก้สมการกำลังสองได้ถูกต้อง 8 ใน 10 ข้อ
- เขียนบรรยายโดยใช้คำประมาณ 200 คำ โดยผิดไวยากรณ์ไม่เกิน 3 แห่ง
- พิมพ์ข้อความ 150 คำ ภายใน 5 นาที โดยไม่มีคำผิดเลย

การกำหนดความรอบรู้ตามตัวอย่างข้างบนนี้จะต้องกำหนดไว้ล่วงหน้า

### การตีความหมายคะแนนแบบอิงเกณฑ์

เป็นการตีความหมายพฤติกรรมโดยเปรียบเทียบกับตัวแปรภายนอก ปกติมักจะไม่ใช้ กับแบบทดสอบ วิธีการตีความหมายแบบนี้โดยทั่ว ๆ ไปจะใช้ตารางพยากรณ์ (Expectancy Table)

วิธีการตีความหมายคะแนนแบบอิงเกณฑ์ คือการนำคะแนนสอบและคะแนนตัวเกณฑ์ มาหารความสัมพันธ์กัน และเปลี่ยนลงในตารางพยากรณ์ คะแนนที่ได้จึงจะสามารถอธิบายพฤติกรรม ตามเกณฑ์ที่คาดหวัง แต่ก็จะยังคงมีความคลาดเคลื่อนในการพยากรณ์เช่นเดียวกับการตี ความหมายคะแนนอื่น ๆ

การตีความหมายคะแนนแบบอิงเกณฑ์มักนิยมใช้ในการทำนาย หรือการตัดสินใจที่ จะเลือกใช้แบบทดสอบ

### III บทสรุป

1. คะแนนในตัวของมันเองไม่มีความหมายอย่างอื่น นอกไปจากหมายถึงจำนวนข้อที่ทำข้อสอบถูก จำนวนข้อของพฤติกรรมที่สังเกตได้ แต่คะแนนก็เป็นสิ่งสำคัญที่เป็นพื้นฐานในการตีความหมาย หรือการแปลงเป็นคะแนนอื่น ๆ ซึ่งจำเป็นจะต้องทำให้คะแนนนั้นมีความแม่นยำ และมีความคลาดเคลื่อนน้อยที่สุด

2. การตีความหมายคะแนนจะต้องพิจารณาคุณลักษณะของคะแนนก่อนว่ามีการวัดในระดับใด ตั้งแต่มาตรฐานบัญญัติ จัดอันดับ อันตรภาค และอัตราส่วน ทั้งนี้เพราะภารตีความหมายคะแนนจำเป็นจะต้องใช้สถิติข้ามช่วง การจะนำสถิติอะไรมาใช้ก็ต้องดูข้อมูลก่อน การพิจารณาว่าคะแนนที่ได้มีการวัดอยู่ในระดับใดนั้นอาจจะทำได้ไม่ง่ายนัก จะนั้นจึงควรทดลองใช้สถิติสูง ๆ ศูนย์กลางไม่ตรงกับความท่าจะเป็นของข้อมูลก็คือ ฯ ลดระดับของสถิติลงมา อย่างไรก็ตามการวัดทางการศึกษานั้นมักจะมีระดับการวัดไม่เกินระดับอันตรภาค จึงควรหลีกเลี่ยงสถิติในระดับอัตราส่วน

3. คะแนนในการสอบแต่ละครั้ง จะมีคะแนนจริงผสมอยู่กับคะแนนความคลาดเคลื่อน การตีความหมายคะแนนจึงต้องระมัดระวังให้เกิดความคลาดเคลื่อนน้อยที่สุด ความคลาดเคลื่อน มีทั้งแบบคงที่คือทุกคนได้รับเท่า ๆ กัน กับความคลาดเคลื่อนแบบสูงที่แต่ละคนได้รับไม่เท่ากัน คะแนนความคลาดเคลื่อนเมื่อนำมาหาค่าเฉลี่ยแล้วจะเป็นศูนย์ ความสัมพันธ์ของคะแนนจริง และคะแนนความคลาดเคลื่อนก็เป็นศูนย์ ความสัมพันธ์ของความคลาดเคลื่อนของแบบทดสอบ คู่หน้างันก็เป็นศูนย์ จากคุณสมบัติดังกล่าว�ังสามารถสรุปเป็นทฤษฎีการทดสอบได้ ๙ ทฤษฎี ซึ่งในทฤษฎีดัง ๆ นี้จะช่วยในการตีความหมายของคะแนนได้แม่นยำขึ้น

4. คะแนนการสอบที่จะนำมาตีความหมายนั้น สามารถจะจำแนกเป็นคะแนนแบบอิงกลุ่ม แบบอิงเนื้อหา และแบบอิงเกณฑ์ การตีความหมายคะแนนแบบอิงกลุ่มเป็นการตีความหมายคะแนนสอบของผู้สอบโดยเปรียบเทียบกับคะแนนของกลุ่มเกณฑ์ปกติ และมีการใช้กันแพร่หลายทั่วไป การตีความหมายคะแนนแบบอิงเนื้อหาเป็นการนำคะแนนไปเปรียบเทียบกับขอบเขตของเนื้อหาหรือทักษะ ซึ่งอาจกำหนดเป็นเกณฑ์ขั้นต่ำของการตอบรู้ การตีความหมายแบบอิงเนื้อหามักมีปัญหาการเขียนขอบเขตของทักษะและเนื้อหา yang ไม่ค่อยเด่นชัด จึงต้องเขียนให้ชัดเจน เนพาะเจาะจง และควรกำหนดระดับการตอบรู้ไว้ล่วงหน้าด้วย การตีความหมายแบบอิงเกณฑ์นั้นเป็นการตีความหมายคะแนนเพื่อพยากรณ์โดยเปรียบเทียบกับตัวแปรภายนอก มักจะทำในรูปของตารางพยากรณ์เพื่อใช้คำนายนและตัดสินใจเลือกใช้แบบทดสอบ

## **คำ답นท์ไทยบท**

1. อธิบายความหมายของคะแนนดิบ รวมทั้งความสำคัญของคะแนนดิบ
2. วิเคราะห์สาเหตุที่จะต้องรู้ระดับมาตรฐานการวัดของคะแนนในแต่ละระดับในส่วนที่เกี่ยวกับการตีความหมายคะแนน
3. คะแนนจริงและคะแนนความคลาดเคลื่อนเกี่ยวข้องกันอย่างไร และคะแนนความคลาดเคลื่อน มีอะไรบ้าง จยกตัวอย่างให้เห็นความแตกต่าง
4. ทฤษฎีการทดสอบบางทฤษฎีเกี่ยวข้องกับการตีความหมายคะแนนอย่างไร
5. อธิบายเปรียบเทียบรูปแบบการตีความหมายคะแนนในแต่ละแบบพร้อมทั้งยกตัวอย่างประกอบ