

บทที่ 1

คะแนนและการตีความหมาย

I จุดประสงค์การเรียนรู้

หลังจากจบบทนี้แล้ว นักศึกษาสามารถปฏิบัติได้ดังนี้

1. บอกความหมายของคะแนน
2. อธิบายความแตกต่างระหว่างการวัดในแต่ละระดับ
3. อธิบายความหมายของคะแนนจริงและความคลาดเคลื่อนได้
4. อธิบายทฤษฎีการทดสอบในแต่ละทฤษฎี
5. เปรียบเทียบการตีความหมายคะแนนการสอบในแต่ละแบบ

II รายละเอียดของเนื้อหา

1. ความหมายของคะแนน

มีผู้ให้ความหมายของคะแนนไว้หลาย ๆ คน เช่น

อนาสตาซี (Anastasi; 1968:39) กล่าวว่า “คะแนนดิบในการทดสอบนั้นไม่มีความหมายอะไรเลย เช่น ถ้าเด็กคนหนึ่งทำข้อสอบได้ 15 คะแนน ก็หมายความว่าเพียงแค่นั้น ไม่อาจจะมาตีความหมายเป็นอย่างอื่นได้ แม้แต่จะเอามาแปลเป็นคำร้อยละ เช่นจะบอกว่าทำได้ 65% ก็ยังไม่บอกไม่ได้เพราะการทำข้อสอบ 65% ของฉบับหนึ่งอาจจะมีค่าเท่ากับ 30% ของอีกฉบับหนึ่งหรือเท่ากับ 80% ของอีกฉบับหนึ่งก็ได้ การจะตีความหมายจึงจำเป็นจะต้องขึ้นอยู่กับระดับความยากง่ายของข้อสอบแต่ละข้ออีกด้วย”

ไลแมน (Lyman; 1971:6) ให้ความหมายของคะแนนไว้ว่า “คะแนนดิบคือจำนวนข้อที่ตอบถูก คะแนนดิบจะมีมากหรือน้อยเพียงใดย่อมขึ้นอยู่กับจำนวนข้อสอบ และความยากง่ายของข้อสอบ ดังนั้นตัวคะแนนดิบเองจึงไม่มีความหมายอะไรเลย แต่เนื่องจากคะแนนดิบนั้นเรานำมาใช้เป็นพื้นฐานของคะแนนอื่น ๆ หลายชนิด เป็นคะแนนที่ถูกต้อง เชื่อถือได้ หรืออาจกล่าว

“ได้ว่าคะแนนดิบเป็นพื้นฐานของคะแนนที่ไม่มีคะแนนอื่นใดจะถูกตัดและแม่นยำมากไปกว่าคะแนนนี้”

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

จากความหมายของคะแนนที่มีผู้กล่าวไว้นี้ จะเห็นได้ว่าตัวคะแนนดิบจริง ๆ แล้วบอกได้เพียงจำนวนนับของพฤติกรรมที่เราสังเกตได้เท่านั้น แต่อย่างไรก็ตามคะแนนดิบก็เป็นสิ่งสำคัญที่เราจะต้องพยายามทำให้มีคุณภาพและเชื่อถือได้มากที่สุดซึ่งจะเป็นผลให้การตีความหมายคะแนนเกิดความผิดพลาดน้อยที่สุด

2. ระดับของการวัด (Level of Measurement)

ในการนำคะแนนดิบมาตีความหมายนั้นเราจำเป็นต้องอาศัยหลักทางสถิติเข้ามาช่วย ดังนั้นก่อนที่เราจะใช้สถิติอะไรก็จะต้องรู้พื้นฐานของคะแนนก่อนว่ามีคุณสมบัติของการวัดในระดับใด ระดับการวัดซึ่งมีความแตกต่างกันอยู่ 4 ระดับนั้นแต่ละระดับจะมีลักษณะต่อเนื่องกัน ถ้าเราสามารถรู้ได้ว่าคะแนนที่เราจะนำมาตีความหมายนั้นอยู่ในระดับการวัดใดการเลือกใช้สถิติตามระดับการวัดก็จะถูกต้อง ความแตกต่างของระดับการวัดนั้นอยู่ที่ข้อจำกัดของข้อมูลที่จะนำเอาวิธีการทางคณิตศาสตร์มาคำนวณ เช่น การบวก ลบ คูณ หารนั้นเราจะนำมาใช้กับข้อมูลทุกประเภทไม่ได้ ตัวอย่างที่เรามักจะใช้กันผิดพลาดเสมอ เช่น บอกว่าคนที่มี I.Q. 140 จะมีความสามารถทางสติปัญญาสูงกว่าคนที่มี I.Q. 70 อยู่ 2 เท่า นั่น ซึ่งความจริงเราไม่สามารถจะกล่าวเช่นนั้นได้ ตัวอย่างนี้จึงเป็นตัวอย่างที่แสดงให้เห็นการใช้ข้อมูลตีความหมายผิดพลาดจากความเป็นจริง

ระดับของการวัดทั้ง 4 ระดับ มีดังนี้คือ

1. **มาตรานามบัญญัติ (Nominal Scale)** เป็นระดับการวัดที่ต่ำที่สุด ง่ายที่สุด ข้อมูลที่อยู่ในการวัดระดับนี้เป็นเพียงการระบุชื่อ เพศ เช่นเวลาเราจะลงรหัสในคอมพิวเตอร์ กำหนดเพศชาย เป็น 1 เพศหญิง เป็น 2 หรือกลุ่มเก่งเป็นเลข 1 กลุ่มปานกลางเป็นเลข 2 และกลุ่มอ่อนเป็นเลข 3 การกำหนดสัญลักษณ์หรือหมายเลขให้กับค่าวัดระดับนี้จึงทำได้ง่าย แต่เราจะเอาตัวเลขเหล่านี้มาบวก ลบ คูณ หารกันไม่ได้ จะทำได้ก็เพียงนับจำนวนความถี่เท่านั้น

2. **มาตราจัดอันดับ (Ordinal Scale)** ระดับการวัดระดับที่สองนี้คือ การจัดอันดับตามคุณลักษณะบางอย่างที่วัดได้ แต่ยังไม่สามารถบอกช่วงห่างระหว่างระดับแต่ละระดับได้ชัดเจน

เช่นเรียงลำดับนักเรียนตามความถี่ที่เข้าร่วมกิจกรรม การให้ความร่วมมือ ฯลฯ สถิติที่จะนำมาใช้ในการวัดระดับนี้ยังมีข้อจำกัด และมักจะมีผู้นำข้อมูลระดับนี้ไปใช้กันผิด ๆ เสมอ ซึ่งเป็นผลให้การตีความหมายผิดพลาดไปด้วย

3. **มาตราวัดแบบอันตรภาค (Interval Scale)** ข้อมูลในระดับนี้มีช่วงห่างของคะแนนเท่ากัน แต่ข้อมูลระดับนี้ก็ยังไม่รู้ค่าศูนย์ที่แท้จริง (absolute zero) ของคะแนนชุดนั้น คะแนนที่ได้จากการวัดจากแบบทดสอบส่วนใหญ่จะมีระดับคะแนนอยู่ในแบบอันตรภาค ตัวอย่างเช่น เด็ก 2 คน ได้คะแนน 30 และ 60 (หมายความว่าทำข้อสอบถูก 30 ข้อ และ 60 ข้อ ตามลำดับ) จากคะแนนของเด็กคู่นี้เราจะบอกว่าเด็กคนหนึ่งทำข้อสอบถูกมากกว่าอีกคนหนึ่งอยู่ 30 ข้อ แต่เราไม่อาจกล่าวได้ว่าเด็กคนหนึ่งรู้เป็น 2 เท่าของอีกคนหนึ่ง ทั้งนี้เพราะเราไม่รู้ศูนย์แท้จริงของคะแนนชุดนี้ หรือเราไม่อาจกล่าวได้ว่าคนที่ทำข้อสอบผิดหมดไม่มีความรู้อะไรเลย จะบอกได้ก็เพียงว่าเขาทำข้อสอบผิดหมด สมมติว่าจากข้อสอบชุดนี้เราเพิ่มข้อสอบที่ง่าย ๆ เข้าไปอีก 30 ข้อ และข้อสอบ 30 ข้อนี้จะทำให้ทุกคนทำข้อสอบถูกหมด คะแนนของเด็กคู่นั้นก็จะเป็น 60 และ 90 คะแนน ซึ่งก็แสดงว่าเราจะบอกว่าเด็กคนที่เคยได้ 60 คะแนนรู้เป็น 2 เท่าของคนที่ได้ 30 คะแนนไม่ได้ชัดเจนขึ้น

การใช้สถิติสำหรับการวัดระดับนี้สามารถใช้สถิติที่ใช้กับระดับ 1 และระดับ 2 ที่ต่ำกว่าได้ นอกจากนั้นยังสามารถเพิ่มการบวกลบเข้าไปได้ แต่ก็ยังคูณไม่ได้ เพราะฉะนั้นคะแนนจากการสอบตามตัวอย่างที่กล่าวมาแล้ว ถ้าสมมติว่าคำถามแต่ละข้อมีค่าเท่า ๆ กัน เราก็สามารถนำมาบวกลบแล้วหาค่าเฉลี่ยออกมาได้ ดังนั้นการจะนำคะแนนดิบแปลงเป็นคะแนนอื่น ๆ จึงต้องกระทำเฉพาะกับคะแนนที่มีระดับการวัดอยู่ในระดับอันตรภาคชั้นขึ้นไป แต่ถ้าเราไปเอาคะแนนที่มีระดับการวัดต่ำกว่านี้มาบวกลบคูณหารกันโดยสมมติว่าคะแนนเหล่านั้นอยู่ในระดับการวัดแบบอันตรภาคก็จะทำให้การตีความหมายคะแนนผิดพลาดไปได้ ถ้าตามข้อเท็จจริงแล้วข้อมูลที่ได้เป็นเพียงมาตราจัดอันดับหรือนามบัญญัติเท่านั้น

4. **มาตราอัตราส่วน (Ratio Scale)** เป็นระดับการวัดที่สูงที่สุด ที่มีคุณสมบัติครบถ้วนตามแบบมาตรานามบัญญัติ จัดลำดับ และอันตรภาค ซึ่งนอกจากจะสามารถนำมาบวกลบกันได้แล้วยังสามารถนำมาคูณหารกันได้ เพราะการวัดในระดับนี้เรารู้ศูนย์ที่แท้จริงของคะแนน เช่นการวัดน้ำหนัก ส่วนสูง เวลา ฯลฯ ซึ่งข้อมูลเหล่านี้ต่างก็มีศูนย์แท้จริง จึงสามารถนำมาบวก ลบ คูณ หาร กันได้ และมักจะพบเสมอว่า ถึงแม้ข้อมูลที่ได้จากการวัดจะมีคุณสมบัติในระดับนี้ แต่ก็ยังใช้สถิติเช่นเดียวกับการวัดระดับ 1 และระดับ 2 เช่นนำความสูงนักเรียนมาแบ่งเป็นกลุ่มสูงและกลุ่มเตี้ย หรือกลุ่มสูงมาก สูงปานกลาง และเตี้ย เป็นต้น

ข้อมูลจะเป็นระดับอันตรายภาคหรือระดับอัตราส่วนนั้นมักจะไม่ค่อยมีปัญหาในการวัดทางการศึกษามากนัก เพราะส่วนมากมักจะสมมติว่าข้อมูลทางการศึกษาเป็นข้อมูลระดับอันตรายภาค ซึ่งก็สามารถนำมาวิเคราะห์สิ่งที่ต้องการได้แล้ว ถ้าเรายังสงสัยว่าไม่น่าจะใช้สถิติถึงมาตราวัดอัตราส่วนก็ควรจะหลีกเลี่ยงเสีย

3. คะแนนจริง คะแนนความคลาดเคลื่อน และทฤษฎีการทดสอบ

(True Scores, Error Scores and Test Theory)

ในการสอบแต่ละครั้งผลที่ได้ออกมาคือจำนวนข้อที่ตอบถูก จำนวนชิ้นของงานที่ทำได้ ฯลฯ ค่าที่วัดได้นี้เราเรียกว่าคะแนนดิบ (Raw Score) คะแนนดังกล่าวนี้ 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 ประเภทคือ

2.1 ความคลาดเคลื่อนคงที่ (constant error) ความคลาดเคลื่อนชนิดนี้หมายถึงผู้สอบทุก ๆ คนจะได้รับผลกระทบเหมือนกัน เช่น ครูเพิ่มเวลาทำข้อสอบให้ 5 นาที ทุกคนที่เข้าสอบก็จะได้รับผลนี้เหมือนกัน

2.2 ความคลาดเคลื่อนแบบสุ่ม (random error) เป็นความคลาดเคลื่อนที่แต่ละคนจะได้รับไม่เท่ากัน ซึ่งอาจจะมีผลทั้งทางบวกและทางลบ เช่น สุขภาพของผู้สอบขณะสอบสภาพจิตใจก่อนเข้าสอบ ฯลฯ

คะแนนที่ผู้สอบทำได้นั้นเป็นสิ่งที่เรานำมาแปลผล และตัดสินใจ คะแนนดังกล่าวประกอบด้วยคะแนนจริงและคะแนนความคลาดเคลื่อน ซึ่งก็หมายความว่าคะแนนของแต่ละคนอาจจะมากกว่าหรือน้อยกว่าคะแนนจริงที่เขาควรจะได้ คะแนนดังกล่าวเราจึงนำมาเขียนในรูปของสมการได้ดังนี้

