

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาวรรณคดีและงานวิจัยที่เกี่ยวข้องโดยแบ่งเป็น 3 ตอน ดังนี้ คือ
ตอนที่ 1 เกี่ยวกับคุณลักษณะของแบบสอบที่ดี ตอนที่ 2 ข้อเสนอแนะและความเห็นเกี่ยวกับการจัดเรียงข้อคำถาม ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 คุณลักษณะของแบบสอบที่ดี

เนื่องจากผลการสอบมีความสำคัญยิ่งต่อผู้สอบจึงมีความจำเป็นที่ผู้สร้างแบบสอบจะต้องทราบถึงคุณลักษณะของแบบสอบที่ดี อันจะเป็นแนวทางในการสร้างแบบสอบ และใช้เป็นเกณฑ์ในการวินิจฉัยว่าแบบสอบที่สร้างขึ้นนั้นมีคุณภาพดี - ไม่คืออย่างไร เพื่อจะได้ปรับปรุงแก้ไขให้เป็นแบบสอบที่มีประสิทธิภาพยิ่งขึ้น

พวงแก้ว ปุณยภนิก ไทกล่าวถึงคุณลักษณะที่ดีของแบบสอบที่สามารถใช้เป็นเกณฑ์สำหรับวินิจฉัยคุณค่าของแบบสอบ และการวัดผลไว้ 10 ประการ (15 : 15) คือ

1. ความตรง (Validity) คือแบบสอบจะต้องมีความแม่นยำของคะแนนในการวัดสิ่งที่ต้องการวัด
2. ความเที่ยง (Reliability) คือต้องมีความคงเส้นคงวาของคะแนนสอบ แบบสอบที่ดีนั้นใช้สอบซ้ำกับผู้สอบคนเดียวกันควรจะได้คะแนนเท่ากัน
3. ความเป็นปรนัย (Objectivity) คือแบบสอบควรจะปราศจากความอคติหรือของคำถาม ปราศจากอคติในการให้คะแนน
4. ความสมดุลย์ (Balance) คือแบบสอบจะมีความตรงตามเนื้อหาได้จะต้องวัดพฤติกรรมที่เกี่ยวข้องกับเนื้อเรื่องที่สอบได้ครอบคลุมเนื้อหา และวัตถุประสงค์ สักส่วนของคำถามเป็นไปตามรายการตารางเฉพาะ
5. ความยุติธรรม (Fairness) คือผู้สอบทุกคนมีโอกาสที่จะตอบถูกถ้ามีความรู้ในเนื้อเรื่องที่ออกข้อสอบนั้น ๆ

6. ประสิทธิภาพในการนำไปใช้ (Efficiency) คือแบบสอบจะต้องประหยัดเวลาในการสร้าง เวลาในการดำเนินการสอบ เวลาในการตรวจข้อสอบ ตลอดจนอุปกรณ์ต่าง ๆ ในการสร้างข้อสอบ

7. ความยาก (Difficulty) ที่เหมาะสม ซึ่งความยากที่เหมาะสมของแบบสอบ คือคะแนนตรงกลางระหว่างคะแนนสูงสุดที่ควรจะได้ กับคะแนนที่ทำได้โดยโอกาสการเดา

8. อำนาจจำแนก (Discrimination Power) คือแบบสอบต้องจำแนกคนเก่งออกจากคนอ่อนได้

9. ความเฉพาะเจาะจง (Specificity) คือผู้ที่มีความรู้ความสามารถเฉพาะเรื่องนั้น ๆ จึงจะตอบข้อสอบนั้นได้ ซึ่งวิธีหาความเฉพาะเจาะจงสามารถทำได้โดยให้ผู้มีความสามารถในเรื่องนั้น ๆ กลุ่มหนึ่งทำแบบสอบนั้นเปรียบเทียบกับคนฉลาดที่ไม่มีความรู้ในเรื่องนั้น ถ้ากลุ่มแรกทำได้ดีกว่ากลุ่มหลัง ถือว่าแบบสอบนั้นมีความเฉพาะเจาะจง

10. ให้ความพอเหมาะ (Speededness) คือแบบสอบที่ไม่มีความจำกัดในเรื่องเวลา แม้ว่าเวลาในการสอบนานเท่าใดคะแนนของผู้สอบก็คงไม่เพิ่มขึ้นอยู่นั่นเอง ดังนั้นแบบสอบวัดผลสัมฤทธิ์ควรให้เวลาพอเพียงที่เด็กเกินกว่า 90% ของจำนวนเด็กทั้งหมดทำแบบสอบนั้นเสร็จ

วิเชียร เกตุสิงห์ ได้กล่าวถึงลักษณะของแบบสอบที่ดีไว้ 10 ประการ

(16 : 27-32) คือ

1. ความตรง (Validity) หมายถึงคุณลักษณะของข้อสอบที่สามารถวัดได้ตรงจุดมุ่งหมายที่ตั้งไว้
2. ความเที่ยง (Reliability) แบบสอบที่มีความเที่ยง หมายถึงแบบสอบที่สามารถวัดได้แน่นอนไม่เปลี่ยนแปลงไปเปลี่ยนมา การวัดครั้งแรกเป็นอย่างไรเมื่อวัดซ้ำอีกครั้งหนึ่งหรือหลาย ๆ ครั้งก็ตาม ผลการวัดก็ยังคงเหมือนเดิม
3. อำนาจจำแนก (Discrimination) คือลักษณะของแบบสอบที่สามารถจะแยก หรือจำแนกเด็กเก่ง-อ่อนได้
4. ความเป็นปรนัย (Objectivity) ข้อสอบที่จะเป็นปรนัยได้ประกอบด้วยคุณสมบัตินี้ 3 ประการ คือ 1) มีความชัดเจนในความหมายของคำถาม 2) มีความคงที่ในการตรวจให้คะแนน 3) มีความแจ่มชัดในการแปลความหมายของคะแนน

5. ความมีประสิทธิภาพ (Efficiency) หมายถึงลักษณะข้อสอบที่มีคุณสมบัติแสดงถึงการประหยัดเศรษฐกิจ เช่น ลงทุนน้อย ราคาถูก ฯลฯ แต่ใช้เวลาสอบน้อยครั้งนี้เป็นต้น
6. การวัดอย่างลึกซึ้ง (Searching) หมายถึงลักษณะของข้อสอบที่ถามครอบคลุมพฤติกรรมหลาย ๆ ด้าน ไม่ใช่วัดแต่ความรู้ความจำอย่างเดียว
7. ข้อสอบที่เที่ยงยุติธรรม (Fair) ไม่เปิดโอกาสให้เด็กคนใดคนหนึ่งได้เปรียบคนอื่น
8. ข้อสอบที่ชัดเจนเฉพาะเจาะจง (Definite) ไม่ถามหลายแง่หลายมุม ข้อเดียวพยายามถามเพียงอย่างเดียว
9. ข้อสอบที่วัดต้องมีความยากง่าย (Difficulty) พอเหมาะ
10. ข้อสอบที่วัดต้องมีการกระตุ้น ยุแหย่ (Exemplary) มีการถามล่อ โดยจับเอาข้อสอบง่าย ๆ ไว้ในตอนแรก ๆ แล้วจึงค่อย ๆ ถามให้ยากขึ้นตามลำดับ

ในการวิจัยครั้งนี้ผู้วิจัยได้เลือกคุณลักษณะที่ดีของแบบสอบมาเพื่อพิจารณาและเปรียบเทียบคุณภาพระหว่างแบบสอบ 4 ชุด รวม 4 ด้าน คือ ความยาก อ่านาจำแนก ความเที่ยง และความตรงของแบบสอบ เนื่องจากสามารถคำนวณเป็นค่าตัวเลข และใช้เทคนิคทางสถิติทดสอบความแตกต่างได้ จึงจะกล่าวรายละเอียดต่อไปนี้

ความยากของแบบสอบ

ความยากของข้อสอบเป็นสัดส่วน หรือเปอร์เซ็นต์ของจำนวนผู้สอบที่ตอบข้อสอบนั้น ๆ ถูกต้อง ถ้าผู้สอบตอบข้อสอบข้อใดถูกกันมากก็แสดงว่าข้อสอบนั้นง่าย ในทางตรงกันข้ามถ้าผู้สอบตอบข้อสอบนั้นถูกกันน้อยก็แสดงว่าข้อสอบนั้นยาก ค่าความยากของข้อสอบไม่ได้เป็นเครื่องชี้ว่าข้อสอบนั้นดีหรือเลว มันเป็นเพียงแต่ระบุว่าข้อสอบนั้นยากหรือง่ายเท่านั้น "อย่างไรก็ตาม คำนี้นี้ความยากของข้อสอบก็ยังสะท้อนให้เห็นความสามารถของกลุ่มผู้เข้ารับการทดสอบด้วย"

(8 : 360) เหตุผลสำคัญในการหาความยากของข้อสอบก็เพื่อที่จะคัดเลือกข้อสอบที่มีความยากไม่เหมาะสมออก ข้อสอบที่ไม่มีผู้ใดทำถูกเลยกับข้อที่ทุกคนทำได้หมดนั้น หากเป็นแบบสอบที่ต้องการจำแนกคนที่มีความสามารถแตกต่างกันแล้วควรตัดออก เพราะข้อสอบทั้งสองประเภทนี้ไม่สามารถแยกความแตกต่างของผู้สอบได้ และข้อสอบแบบนี้ไม่ได้มีผลต่อความแปรปรวนของคะแนนสอบ จึงไม่มีผลต่อค่าความเที่ยงและความตรงของแบบสอบนี้ (17 : 161 - 162)

ในการวัดแบบอิงเกณฑ์ (Criterion Reference) โดยทั่วไปข้อสอบที่มีค่าความยากสูง นักเรียนส่วนมากจะได้คะแนนสูง ซึ่งจะช่วยให้ทราบว่านักเรียนคนใดบ้างที่มีความรู้ และทักษะไม่ถึงเกณฑ์ที่เราคาดหวัง (18 : 257) ส่วนในการวัดแบบอิงกลุ่ม (Norm Reference) นั้น เพื่อให้ทราบความจำแนกระหว่างบุคคลมีค่าสูงสุดจึงเป็นการดีที่สุดที่จะเลือกข้อสอบที่มีความยากประมาณ .50 ซึ่งจะทำให้ได้คะแนนรวมที่มีการกระจายมากที่สุด ซึ่งจะมีผลทำให้ค่าสัมประสิทธิ์ความเที่ยงมีค่าสูงสุดด้วย (17 : 162) นั้นแสดงว่าค่าความยากย่อมมีผลถึงความเที่ยง และความตรงของแบบสอบด้วย

อย่างไรก็ตามค่าความยากนี้มีความสัมพันธ์กับค่าอำนาจจำแนก ค่าความยากที่ทำให้ค่าอำนาจจำแนกมากที่สุดคือ ค่าความยากเท่ากับ .50 แต่ข้อพึงระวังอีกอย่างก็คือ การที่ข้อสอบมีค่าความยากเท่ากับ .50 อาจจะทำให้ข้อสอบนั้นมีค่าอำนาจจำแนกสูงสุด หรือไม่มีเลยก็ได้ เพราะถ้าครึ่งหนึ่งของจำนวนคนในกลุ่มสูงตอบถูก และครึ่งหนึ่งของจำนวนคนในกลุ่มต่ำตอบถูก ก็ทำให้ค่าอำนาจจำแนกเท่ากับ 0 (18 : 257) นั้นแสดงว่าค่าความยากเท่ากับ .50 นั้นอาจทำให้ข้อสอบมีอำนาจจำแนกสูงสุดหรือไม่มีอำนาจจำแนกเลยก็ได้

องค์ประกอบที่มีอิทธิพลต่อความยากของแบบสอบ

องค์ประกอบที่มีอิทธิพลต่อความยากของแบบสอบมี 2 ประเภท (19 : 899-913) คือ

1. องค์ประกอบภายใน (Intrinsic Factors) ประกอบด้วย
 - ก. เนื้อหาของข้อสอบแต่ละข้อ (Item Content) ในด้านความซับซ้อน ความเป็นนามธรรม และความแปลกใหม่
 - ข. โครงสร้างของข้อสอบ (Item Structure) หมายถึงวิธีการแสดงออกซึ่งเนื้อหานั้น
2. องค์ประกอบภายนอก (Extrinsic Factors) ประกอบด้วย
 - ก. ความไม่คุ้นเคยต่อเนื้อหา (Unfamiliarity) คืออยู่นอกเหนือประสบการณ์ของผู้สอบ
 - ข. สิ่งที่สัมพันธ์กับข้อสอบ (Item Context) เช่นข้อสอบที่อยู่ใกล้เคียงกัน
 - ค. ตัวแปรด้านบุคลิกภาพ ได้แก่ สภาพร่างกาย ลักษณะนิสัย และความตั้งใจของผู้สอบ

นอกจากนี้ธรรมชาติของเนื้อหา พฤติกรรมที่ต้องการ วัด และตัวการที่สลับซับซ้อนอื่น ๆ เช่น ภาษาที่ใช้ในข้อสอบ รูปแบบของคำถาม และคำชี้แจงต่าง ๆ ยังเป็นองค์ประกอบที่อาจจะ มีอิทธิพลต่อความยากของแบบสอบที่ควรคำนึงถึงด้วย

อำนาจจำแนกของแบบสอบ

อำนาจจำแนก หมายถึงประสิทธิภาพของข้อคำถามในการแบ่งเด็กออกเป็นกลุ่มคน เก่ง และอ่อน อำนาจจำแนกมีค่าตั้งแต่ -1.00 ถึง 1.00 ข้อสอบที่มีค่าอำนาจจำแนก เข้าใกล้ -1.00 หรือ 1.00 แสดงว่ามีอำนาจจำแนกสูง แต่ข้อสอบที่มีคุณค่าตรงตาม วัตถุประสงค์ของการ สอบคือ ข้อสอบที่มีค่าอำนาจจำแนกเป็นบวก ซึ่งหมายความว่านักเรียนใน กลุ่มสูงตอบข้อสอบได้ถูกต้องมากกว่านักเรียนในกลุ่มต่ำ และแสดงว่าข้อสอบสามารถจำแนกไป ในทางเดียวกับคะแนนรวมที่ได้จากการสอบ ส่วนข้อสอบที่มีค่าอำนาจจำแนกเป็นทางลบถือว่า จำแนกผิดทิศทาง ไม่ตรงตามวัตถุประสงค์ที่ต้องการ และข้อสอบที่มีอำนาจจำแนกเป็น 0 ถือว่า จำแนกความแตกต่างระหว่างคน 2 กลุ่มไม่ได้เลย

ในการประเมินคุณภาพของข้อสอบโดยอาศัยค่าอำนาจจำแนกควรถือเกณฑ์ดังนี้(8 : 364)

ค่าอำนาจจำแนก	คุณภาพของข้อสอบ
.40 ขึ้นไป	ดีมาก
.30 ถึง .39	ดี
.20 ถึง .29	ใช้ได้
ต่ำกว่า .20	ไม่ดีควรตัดทิ้ง

อำนาจจำแนกของข้อสอบเป็นสิ่งสำคัญอีกอย่างหนึ่งที่ควรคำนึงถึงในการประเมินคุณภาพของข้อสอบควบคู่ไปกับความยากของข้อสอบ เนื่องจากข้อสอบที่มีความยากเท่ากันยัง ต่างกันในค่าอำนาจจำแนกได้ด้วย ดังที่ได้อธิบายแล้วว่าข้อสอบที่มีความยาก .50 อาจจะทำให้ค่าอำนาจจำแนกสูงที่สุดหรืออาจจะไม่มีเลยก็ได้ ฉะนั้นข้อสอบที่มีความยาก .50 มีโอกาสทำให้ความเที่ยงของแบบสอบต่ำได้เมื่อข้อสอบนั้นมีอำนาจจำแนกเป็น 0 แต่ข้อสอบที่มีค่าอำนาจ จำแนกสูงจะมีผลทำให้ความเที่ยงของแบบสอบสูงด้วยอย่างแน่นอน เนื่องจากค่าอำนาจจำแนก ของข้อสอบเกี่ยวข้องกับส่วน เบี่ยงเบนมาตรฐานของคะแนนซึ่งเป็นองค์ประกอบสำคัญที่มีอิทธิพล ต่อค่าความเที่ยงของแบบสอบ (8 : 302)

องค์ประกอบที่มีอิทธิพลต่ออำนาจจำแนกของแบบสอบ

องค์ประกอบที่สำคัญที่มีอิทธิพลต่ออำนาจจำแนกของแบบสอบได้แก่

1. คุณภาพของตัวลวง และจำนวนตัวเลือก ข้อสอบที่สร้างตัวลวงได้ที่มีความใกล้เคียงกับตัวเลือกที่เป็นคำตอบถูกหลาย ๆ ตัวจะทำให้จำนวนตัวเลือกเพิ่มขึ้น ซึ่งจะทำให้อำนาจจำแนกของข้อสอบสูงขึ้นด้วย (8 : 165) เนื่องจากผู้ที่มีความรู้จริงเท่านั้นสามารถทำข้อสอบได้ถูกต้อง แต่ผู้ที่ไม่มีความรู้จริงจะเลือกตอบตัวลวงที่มีความใกล้เคียงกับคำตอบถูกกระจายกันออกไปทุกตัว

2. ความยากของข้อสอบ การจำแนกความแตกต่างของผลสัมฤทธิ์ระหว่างกลุ่มขึ้นอยู่กับความยากของข้อสอบ (20 : 154) คือข้อสอบที่มีความยาก .50 มีโอกาสทำให้ค่าอำนาจจำแนกสูงสุดได้ เมื่อนักเรียนในกลุ่มสูงทุกคนตอบข้อสอบได้ถูกต้อง ในขณะที่นักเรียนในกลุ่มต่ำทั้งหมดตอบข้อสอบผิด และหากข้อสอบมีค่าความยากสูงหรือต่ำกว่านี้ค่าอำนาจจำแนกของข้อสอบก็จะต่ำลง (20 : 270) ซึ่งหมายความว่าข้อสอบที่ง่ายหรือยากเกินไปจะเป็นข้อสอบที่มีค่าอำนาจจำแนกต่ำนั่นเอง

ความเที่ยงของแบบสอบ

ได้มีผู้ให้ความหมายของคำว่า ความเที่ยงมากมาย เช่น

ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการวัด (20 : 105)

ความเที่ยง หมายถึง ความคงที่ของผลที่ได้รับจากการใช้เครื่องมือชนิดเดียวกัน สอบมากกว่า 1 ครั้ง (21 : 163)

ความเที่ยง หมายถึง ระดับ (degree) ของความคงที่ระหว่างการวัดสิ่งเดียวกัน 2 ครั้ง (5 : 88)

ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการสอบบุคคลกลุ่มเดียวกัน ด้วยแบบสอบชุดเดียวกัน ในเวลาที่ต่างกัน หรือสอบด้วยแบบสอบคนละชุดที่เทียบเท่ากัน หรือภายใต้สภาพการสอบที่ต่างกัน (17 : 73)

จากความหมายดังกล่าวข้างต้นก็จะสรุปได้ว่า ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการวัดด้วยเครื่องมือชุดเดียวกันในโอกาสที่ต่างกัน

ทฤษฎีของความเที่ยง

ในการวัดทางค่านิจภาพ ถึงแม้เครื่องมือที่ใช้จะมีความเที่ยงสูงก็ไม่มีเครื่องมือชนิดใดที่วัดได้ถูกต้องแม่นยำโดยสมบูรณ์ ฉะนั้นคะแนนจากการสอบที่นักเรียนทำได้

(Observed Scores) จะประกอบด้วยส่วนที่เป็นคะแนนจริง (True Scores) และส่วนของคะแนนจากความคลาดเคลื่อน (Error Scores) ซึ่งเขียนเป็นสมการ (22 : 408) ดังนี้

$$X_o = X_t + X_e$$

เมื่อ X_o แทนคะแนนที่ได้จากการวัด

X_t แทนคะแนนจริง

X_e แทนคะแนนความคลาดเคลื่อน

คะแนนจริง คือคะแนนเฉลี่ยที่ผู้สอบได้รับจากการทำแบบสอบเดิมหลายครั้งโดยปราศจากอิทธิพลของการเรียนรู้จากการสอบครั้งแรก แรงงูใจ หรือสภาพทางอารมณ์ เป็นต้น

คะแนนความคลาดเคลื่อน คือคะแนนที่เกิดจากความผิดพลาดในการวัด ซึ่งแบ่งเป็น 2 ประเภท คือ ความคลาดเคลื่อนที่เกิดจากการสุ่ม (Random Error) กับความคลาดเคลื่อนอย่างมีระบบ (Systematic Error) ซึ่งความคลาดเคลื่อนประเภทหลังนี้ไม่มีผลต่อค่าความเที่ยง แต่ความคลาดเคลื่อนที่เกิดจากการสุ่ม ซึ่งเกิดจากตัวข้อสอบ การกำเนินการสอบ การตรวจให้คะแนน และจากตัวผู้สอบเอง ซึ่งเป็นไปได้ทั้งทางบวก และทางลบ และจะหักล้างกันหมดหากคะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง หรือขนาดของคะแนนความคลาดเคลื่อนไม่เกี่ยวข้องกับขนาดของคะแนนจริงแล้ว ความแปรปรวนของคะแนนที่วัดได้จะมีค่าเท่ากับผลบวกของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนความคลาดเคลื่อน (8 : 331) ซึ่งเขียนเป็นสมการ (22 : 410) ดังนี้



$$s_o^2 = s_t^2 + s_e^2 \dots\dots\dots (1)$$

เมื่อ s_o^2 แทนความแปรปรวนของคะแนนที่วัดได้

s_t^2 แทนความแปรปรวนของคะแนนจริง

s_e^2 แทนความแปรปรวนของคะแนนความคลาดเคลื่อน

ความเที่ยง หมายถึงอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่วัดได้ เขียนเป็นสมการ (22 : 410) ได้ว่า

$$r_{tt} = \frac{s_t^2}{s_o^2} \dots\dots\dots (2)$$

จากสมการ (1) และ(2) จะได้สมการพื้นฐานในการหาค่าสัมประสิทธิ์สหสัมพันธ์ของความเที่ยง (22 : 410) ดังนี้

$$r_{tt} = 1 - \frac{s_e^2}{s_o^2}$$

วิธีประมาณค่าความเที่ยง

โดยทั่วไปวิธีประมาณค่าความเที่ยงทำได้ 4 แบบ (5 : 109) คือ

1. แบบวัดซ้ำ (Measure of Stability)
2. แบบใช้ข้อสอบคล้าย (Measure of Equivalence)
3. แบบใช้ข้อสอบคล้าย และวัดซ้ำ (Measure of Equivalence and Stability)
4. แบบวัดความคงที่ภายใน (Measure of Internal Consistency)
 - 4.1 แบบแบ่งครึ่งข้อสอบ (Split - Half)
 - 4.2 แบบคูเดอร์ ริชาร์ดสัน (Kuder - Richardson Estimates)
 - 4.3 แบบวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyts Analysis of

Variance Procedure)

ในที่นี้ของกล่าวถึงเฉพาะความเที่ยงแบบคูเคอร์-ริชาร์ดสัน ซึ่งเป็นวิธีการประมาณค่าที่ใช้ในการวิจัยครั้งนี้

ความเที่ยงแบบคูเคอร์-ริชาร์ดสัน

ความเที่ยงแบบนี้เป็นการหาความสอดคล้องกันระหว่างข้อสอบโดยดำเนินการสอบครั้งเดียว และใช้แบบสอบซุกเดียว ซึ่งความสอดคล้องกันระหว่างข้อนี้ได้รับอิทธิพลจากแหล่งความแปรปรวนคลาดเคลื่อน 2 ชนิด (17 : 84) คือ

1. เนื้อหาที่สุ่ม
2. ความเป็นวิวิธพันธ์ของพฤติกรรมที่สุ่ม ข้อสอบที่มีความเป็นเอกพันธ์มากขึ้น ความสอดคล้องกันระหว่างข้อก็ยิ่งสูงขึ้น

วิธีการหาความเที่ยงแบบคูเคอร์-ริชาร์ดสัน

ในปี 1937 คูเคอร์และริชาร์ดสัน (Kuder & Richardson) ได้เสนอสูตรสำหรับประมาณค่าความเที่ยงแบบนี้ คือ สูตรคูเคอร์-ริชาร์ดสัน 20 (23 : 53)

$$r_{tt} = \left(\frac{k}{k-1} \right) \left(\frac{s_t^2 - \sum p_i q_i}{s_t^2} \right)$$

เมื่อ	r_{tt}	แทนค่าความเที่ยงของแบบสอบ
	k	แทนจำนวนข้อสอบในแบบสอบ
	s_t^2	แทนความแปรปรวนของคะแนนทั้งหมด
	p_i	แทนสัดส่วนของผู้ตอบถูกในแต่ละข้อ
	q_i	แทนสัดส่วนของผู้ตอบผิดในแต่ละข้อ

ค่าความเที่ยงแบบคูเคอร์-ริชาร์ดสัน ประมาณได้กับ

1. ค่าสหสัมพันธ์เฉลี่ยระหว่างข้อสอบ k ข้อในแบบสอบ
2. ค่าสัมประสิทธิ์ความเที่ยงของข้อสอบในแบบสอบ

3. ค่าเฉลี่ยของสัมประสิทธิ์ความเที่ยงที่ประมาณค่าจากสูตรสเปียร์แมน บราวน์ (Spearman - Brown Formula) ซึ่งแบ่งครึ่งข้อสอบด้วยวิธีต่าง ๆ (9 : 191)
- ข้อตกลงเบื้องต้นของวิธีหาความเที่ยงแบบคูเคอร์-ริชาร์ดสัน 20
1. การตรวจให้คะแนนข้อสอบแต่ละข้อเป็นแบบ 0-1 คือตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน (8 : 326)
 2. ข้อสอบแต่ละข้อในแบบสอบจะต้องมีลักษณะเป็นเอกพันธ์ คือวัดในเรื่องเดียวกัน (23 : 78)

องค์ประกอบที่มีอิทธิพลต่อความเที่ยง

นอกจากวิธีการที่ใช้ประมาณค่าความเที่ยงจะมีอิทธิพลต่อค่าความเที่ยงของแบบสอบแล้ว ยังมีองค์ประกอบอื่น ๆ (5 : 100-103) ต่อไปนี้มีอิทธิพลต่อความเที่ยงของแบบสอบด้วย

1. ความยาวของแบบสอบ (Test Length) การเพิ่มจำนวนข้อสอบที่มีคุณภาพดีเท่าเทียมกับข้อสอบเดิม จะทำให้แบบสอบมีค่าความเที่ยงสูงขึ้น นั่นคือแบบสอบที่มีจำนวนข้อมากจะมีความเที่ยงสูงกว่าแบบสอบที่มีจำนวนข้อน้อย สูตรที่แสดงให้เห็นความสัมพันธ์ระหว่างความยาวของแบบสอบกับความเที่ยง คือ สูตรของ สเปียร์แมน-บราวน์ (Spearman - Brown Formula)

2. ความเป็นเอกพันธ์ของกลุ่มผู้รับการทดสอบ (Group Homogeneity) กลุ่มที่มีความเป็นเอกพันธ์ จะมีพิสัยของคะแนนความสามารถของกลุ่มน้อยกว่ากลุ่มที่มีความเป็นวิวิธพันธ์ (Group Heterogeneous) ซึ่งจะทำให้ค่าความเที่ยงของกลุ่มที่มีความเป็นเอกพันธ์ต่ำกว่ากลุ่มที่มีความเป็นวิวิธพันธ์

3. ความยากของข้อสอบ (Item Difficulty) เนื่องจากค่าความเที่ยงนั้นประมาณจากความแปรปรวนของคะแนนผลการสอบ ถ้าข้อสอบยากเกินไปทุกคนทำผิดหมด หรือข้อสอบง่ายเกินไปทุกคนทำถูกหมดก็ทำให้พิสัยของคะแนนลดลง ค่าความแปรปรวนของคะแนนมีน้อย ซึ่งจะทำให้ค่าความเที่ยงลดลงด้วย ดังนั้นถ้าค่าความยากของข้อสอบทำให้คะแนนกระจายมาก ๆ ก็จะทำให้ความเที่ยงมีค่าสูงขึ้นด้วย

ความเที่ยงจะเพิ่มขึ้น 70% เมื่อจำนวนข้อสอบเพิ่มขึ้น 1 เท่า (Spearman-Brown Formula)

4. ความเร็ว (Speed) การประมาณค่าความเที่ยงของแบบสอบความเร็ว (Speed Test) ด้วยวิธีวัดความคงที่ภายในเป็นวิธีไม่เหมาะสม เพราะแบบสอบชนิดนี้เป็นแบบสอบที่ง่ายแต่ใช้เวลาทำน้อย ซึ่งนักเรียนไม่สามารถทำได้ครบทุกข้อ และจะทำให้ค่าความเที่ยงสูงกว่าที่ควรจะเป็นจริง

5. ความเป็นปรนัย (Objective) ถ้าแบบสอบมีความเป็นปรนัยสูง ค่าความเที่ยงของแบบสอบก็สูงขึ้นด้วย

นอกจากนั้นการบริหารแบบสอบ เช่น การจับเวลา การให้คำชี้แจง การป้องกันการทุจริต เป็นต้น หรือองค์ประกอบในค่านเกี่ยวกับตัวผู้สอบเอง เช่น แรงงูใจ ความวิตกกังวล ประสบการณ์ในการสอบ การเคา เป็นต้น สิ่งเหล่านี้อาจเป็นแหล่งความคลาดเคลื่อนที่ทำให้ค่าความเที่ยงของแบบสอบเปลี่ยนแปลงได้

ความตรงของแบบสอบ

ได้มีผู้ให้ความหมายของความตรงไว้มากมาย เช่น

ความตรง หมายถึง ขอบเขตของผลการประเมินที่สอดคล้องกับจุดประสงค์นั้น

(20 : 79)

ความตรงของแบบสอบ หมายถึง ความสามารถที่แบบสอบนั้นบรรลุถึงจุดมุ่งหมายอย่างใดอย่างหนึ่ง (5 : 109)

ความตรง หมายถึง ความถูกต้องแน่นอนในการวัดสิ่งที่ต้องการวัด (24 : 213)

ความตรง หมายถึง คุณสมบัติที่แบบสอบสามารถวัดในสิ่งที่ต้องการจะวัด (8 : 468)

ความตรงของแบบสอบ คือ ค่าสหสัมพันธ์ระหว่างแบบสอบกับเกณฑ์ (25 : 88)

จากความหมายของความตรงดังกล่าว สรุปได้ว่า ความตรง หมายถึง ความแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด หรือสิ่งที่เครื่องมือควรจะวัด เช่น แบบสอบคณิตศาสตร์ ต้องการวัดความสามารถทางคณิตศาสตร์ ถ้ามีการให้คะแนนความถูกต้องในการสะกดคำ คะแนนที่ได้ย่อมขาดความตรง คือ คะแนนที่ได้ให้ความหมายไม่ตรงจุดมุ่งหมายของการวัดนั่นเอง

ชนิดของความทรง

ความทรงของแบบสอบจำแนกได้เป็น 3 ชนิดใหญ่ ๆ (20 : 81) คือ

1. ความทรงตามเนื้อหา (Content Validity)
2. ความทรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity)
3. ความทรงตามคุณลักษณะ (Construct Validity)

ในที่นี้จะกล่าวถึงความทรงที่เกี่ยวข้องกับการวิจัยครั้งนี้ คือ ความทรงตามเนื้อหา และความทรงร่วมสมัย ซึ่งเป็นแบบหนึ่งของความทรงตามเกณฑ์สัมพันธ์

ความทรงตามเนื้อหา

ความทรงตามเนื้อหา คือ ความทรงที่เกี่ยวข้องกับการสุ่ม (Sampling) เนื้อหาวิชาที่ใช้ในการทดสอบ โดยเฉพาะอย่างยิ่งการทดสอบวัดผลสัมฤทธิ์ทางการเรียน มักจะเน้นในเรื่องความทรงในเนื้อหาวิชา การที่จะตัดสินว่าแบบสอบมีความทรงในเนื้อหาวิชามากน้อยเพียงใดมักจะพิจารณาทั้งหัวข้อเนื้อหาวิชา (Subject Matter) และชนิดของพฤติกรรมที่จะวัดไปพร้อมกัน เราจะกำหนดสองสิ่งนี้ในตารางวิเคราะห์หลักสูตร หรือผังการสร้างข้อสอบของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนซึ่งเป็นตาราง 2 มิติ ที่จะช่วยให้การสุ่มหัวข้อเนื้อหาวิชาและพฤติกรรมที่จะนำมาสร้างเป็นข้อสอบ เป็นตัวแทนที่ดีของการถามเนื้อหาวิชา และพฤติกรรมที่มีอยู่ทั้งหมด (26 : 69-70)

เราไม่สามารถแสดงค่าความทรงในเนื้อหาวิชาของแบบสอบออกมาเป็นตัวเลขได้ โดยทั่วไปมักจะพิจารณาข้อสอบในแบบสอบ โดยเทียบกับผังการสร้างข้อสอบแล้วประมาณออกมาว่ามีความทรงตามเนื้อหาวิชาเพียงใด ผู้ที่จะตัดสินความทรงตามเนื้อหาได้ที่ดีที่สุดก็คือผู้เชี่ยวชาญในสาขาวิชานั้น ๆ (5 : 112) สำหรับแบบสอบวัดผลสัมฤทธิ์ในชั้นเรียนผู้ที่ตัดสินความทรงตามเนื้อหาได้ที่ดีที่สุดก็คือ ผู้สอนในวิชาที่จะสอบนั้น

ความทรงร่วมสมัย (Concurrent Validity)

ความทรงร่วมสมัยของแบบสอบ หมายความว่าแบบสอบสามารถช่วยในการคาดคะเนสภาพที่แท้จริงของผู้สอบในปัจจุบันได้ถูกต้องเพียงใด (27 : 11) การหาความทรงร่วมสมัยของแบบสอบทำได้โดยการหาความสัมพันธ์ระหว่างคะแนนจากการสอบด้วยแบบสอบ

ฉบับนั้น กับ เหตุที่่เกิดในเวลาเดียวกัน หรือใกล้เคียงกันว่าสอดคล้องกันเพียงใด หรือโดย การตรวจดูว่าแบบสอบฉบับนั้นสามารถให้คะแนนกระจายมากน้อยเพียงใด และคะแนนเหล่านั้น สอดคล้องกับความสามารถที่แท้จริงของผู้สอบที่ปรากฏในปัจจุบันหรือไม่ หากแบบสอบใด สามารถให้คะแนนกระจายมาก ถือว่าแบบสอบนั้นมีความตรงร่วมสมัยได้ นอกจากนี้ยังสามารถ หากความตรงร่วมสมัยได้โดยใช้สภาพความจริงในปัจจุบันที่สังเกตเห็นเป็นตัวเกณฑ์ คือการนำ คะแนนของผู้สอบแต่ละคนมาจัดอันดับ และเปรียบเทียบกับอันดับความสามารถที่แท้จริง ว่า สอดคล้องกันมากน้อยเพียงใด (27 : 11)

องค์ประกอบที่มีอิทธิพลต่อความตรงของแบบสอบ

องค์ประกอบต่าง ๆ ที่มีอิทธิพลต่อความตรง (20 : 98-102) คือ

1. องค์ประกอบภายในแบบสอบ ได้แก่

1.1 ความไม่ชัดเจนของคำชี้แจงเกี่ยวกับวิธีตอบข้อสอบ การบันทึกคำตอบ เป็นต้น จะทำให้ค่าความตรงลดลง

1.2 โครงสร้างของคำและประโยคยากเกินไป ทำให้นักเรียนสับสนแทนที่ จะวัดได้ตรงตามวัตถุประสงค์ที่ต้องการ ก็จะกลายเป็นการวัดความสามารถทางภาษา

1.3 ระดับความยากที่ไม่เหมาะสมของข้อสอบ ทำให้ไม่สามารถจำแนก นักเรียนได้ จะทำให้ความตรงลดลง

1.4 ข้อสอบที่สร้างไม่ครบข้อ เป็นการแนะนำคำตอบให้นักเรียน

1.5 ความกำกวมของภาษาที่ใช้ในแบบสอบทำให้นักเรียนสับสน

1.6 ความไม่เหมาะสมของแบบสอบที่จะวัดสิ่งที่ต้องการวัด

1.7 แบบสอบสั้นเกินไป ไม่สามารถวัดได้ครอบคลุมพฤติกรรมที่ต้องการวัด

1.8 การเรียงข้อสอบที่ไม่เหมาะสม ข้อสอบที่เรียงจากยากไปหาง่ายทำให้นักเรียนใช้เวลาคิดหาคำตอบนาน อาจจะไม่มีความเวลาทำข้อสอบที่ง่าย ๆ ทอนหลัง การเรียง ข้อสอบแบบนี้มีผลต่อความตรงโดยมีผลต่อแรงจูงใจในการทำแบบสอบของนักเรียน

1.9 การเรียงคำตอบอย่างเป็นระบบ จะช่วยให้นักเรียนเดาคำตอบได้ง่าย

2. เนื้อหาและวิธีดำเนินการสอน การสอนในเนื้อหาที่ตรงกับเนื้อหาที่ต้องการ วัดในข้อสอบ จะทำให้เครื่องมือวัดความตรงในการวัดขบวนการทางสมองที่ซับซ้อนกว่า

3. การบริหารแบบสอบและการให้คะแนน เช่น เวลาในการสอบน้อยเกินไป การชี้แนะนักเรียนบางคนในการตอบ การให้คะแนนที่ไม่คงที่ เหล่านี้ทำให้ค่าความตรงลดลงได้

4. องค์ประกอบในการตอบแบบสอบของนักเรียนเอง รูปแบบการตอบแบบสอบของนักเรียนแต่ละคนแตกต่างกันออกไป ซึ่งทำให้ผลการสอบไม่ตรงกับความสามารถที่แท้จริง ตลอดจนถึงความวิตกกังวล ความเปลี่ยนแปลงทางอารมณ์ เป็นต้น ทำให้ค่าความตรงลดลง

5. ธรรมชาติของกลุ่มและเกณฑ์ที่ใช้ ตามปกติความตรงของแบบสอบจะมีค่าสูงเฉพาะในบางกลุ่ม เนื่องจากในการสอบวันนั้นมีอิทธิพลจากองค์ประกอบอื่น เช่น อายุ เพศ ระดับความสามารถ ภูมิหลังทางการศึกษา และภูมิหลังทางวัฒนธรรม เป็นต้น ดังนั้นในการหาความตรงควรคำนึงถึงกลุ่ม และเกณฑ์ที่ใช้ให้เหมาะสมสมควร

ตอนที่ 2 ข้อเสนอแนะและความเห็นเกี่ยวกับการจัดเรียงข้อคำถาม

ในเรื่องการจัดเรียงข้อคำถามของแบบสอบแต่ละฉบับนั้น ได้มีผู้เสนอแนะ และให้ความคิดเห็นไว้หลายท่านด้วยกัน ดังจะกล่าวต่อไปนี้

ชวาล แพร์ทกุล ได้ชี้ให้เห็นถึงความสำคัญของการจัดระเบียบข้อสอบว่า การพิมพ์ การจัดลำดับข้อ การเขียนคำชี้แจง หรือ คำอธิบายในการทำแบบสอบ การตรวจให้คะแนน วิธีการเหล่านี้มีความสำคัญไม่ยิ่งหย่อนไปกว่าตัวข้อสอบเองเหมือนกัน สำหรับในแบบสอบแต่ละฉบับควรเรียงคำถามเริ่มจากง่ายไปหายากเสมอ ไม่ควรเรียงตามความยากง่ายของบทเรียน (3 : 2-4) ซึ่ง อนันต์ ศรีโสภา ได้เสนอแนะไว้ท่านเองเกี่ยวกับ การจัดเรียงข้อสอบที่สั้นๆ ควรเรียงจากข้อง่ายไปหาข้อยาก (26 : 141) เช่นเดียวกับ สแตนเลย์ และ ฮอปกินส์ (Stanley and Hopkins) ก็มีความเห็นตรงกันว่า การจัดเรียงลำดับข้อสอบเป็นเรื่องสำคัญ ควรจะเริ่มต้นจากข้อง่ายที่สุดไปหาข้อยากที่สุด เพราะตามหลักจิตวิทยา ถ้านักเรียนพบข้อยาก ๆ ในตอนแรกแล้วอาจจะทำให้เกิดความท้อใจ โดยเฉพาะในกลุ่มนักเรียนที่มีความสามารถในการเรียนระดับปานกลางและระดับต่ำ (9 : 191)

ทราเวอร์ส (Travers) ได้เสนอวิธีเรียงข้อสอบไว้ 4 วิธี (10 : 127-128)

1. เรียงตามลำดับความยาก (Arrangement in Order of Difficulty)
 วิธีนี้เป็นการเรียงแบบธรรมดา ซึ่งความยากง่ายของข้อคำถามนั้นอาจจะได้จากการวิเคราะห์ทางสถิติ หรือโดยการประมาณค่า ข้อดีของวิธีนี้คือ ผู้สอบสามารถทำข้อสอบจากง่ายขึ้นไปตามลำดับและไม่ท้อใจเพราะเจอปัญหายาก ๆ ก่อน ข้อเสียมี 2 ประการคือ ประการแรก ข้อสอบที่มีเนื้อหาเดียวกันต้องกระจายไปตามลำดับความยากง่าย ซึ่งมีผลทำให้ความคิดของนักเรียนเปลี่ยนแปลงอยู่เสมอ ประการที่สอง เมื่อนักเรียนทำข้อสอบถึงข้อที่ยากแล้วอาจจะเกิดท้อใจ เพราะรู้ว่าข้อต่อ ๆ ไปจะยิ่งยากขึ้นไปอีก X

2. เรียงความยากแบบหมุนเวียน (Arrangement in Cyclic Order of Difficulty) เป็นการเรียงจากข้อง่ายไปหาข้อยากในกลุ่มหมุนเวียนสลับกันโดยไม่คำนึงถึงเนื้อหา วิธีนี้เป็นการหลีกเลี่ยงข้อบกพร่องประการที่สองของวิธีเรียงตามลำดับความยาก การเรียงแบบนี้มีข้อดี คือ กระตุ้นให้นักเรียนอ่านข้อสอบทุกข้อ เพราะนักเรียนรู้ว่าถ้าอ่านต่อไปจะพบข้อง่าย ๆ อีก จุดอ่อนของวิธีนี้ คือ นักเรียนต้องเปลี่ยนความคิดเร็วเกินไป เนื่องจากเนื้อหายังกระจายอยู่

3. เรียงตามกลุ่มเนื้อหาวิชา (Arrangement According to Subject Matter Area) วิธีนี้มีข้อดีที่เปิดโอกาสให้นักเรียนสามารถใช้ความคิดที่จะตอบปัญหาในเนื้อหาหนึ่ง ๆ ตามลำดับ ก่อนที่จะตอบคำถามในเนื้อหาต่อ ๆ ไป เพราะการเรียงแบบนี้เป็นการเรียงโดยรวมข้อสอบที่มีเนื้อหาเดียวกันเข้าไว้ด้วยกันแล้วจึงเรียงข้อสอบในแต่ละเนื้อหาวิชาจากง่ายไปหายาก

4. เรียงตามจุดมุ่งหมายของการวัด (Arrangement According to the Goals Measured) ในการเรียงข้อสอบ บางครั้งผู้เขียนข้อสอบนิยมเรียงคำถามโดยจัดพฤติกรรมที่ต้องการวัดประเภทเดียวกันไว้ด้วยกันเป็นกลุ่ม ๆ ตามจุดมุ่งหมายที่ต้องการวัดนั้น ๆ ข้อดีของวิธีนี้ คือ มีประโยชน์สำหรับครูผู้สอนสามารถตรวจสอบการวัดว่าตรงตามจุดมุ่งหมายที่ต้องการหรือไม่ ส่วนข้อเสีย คือ ไม่สามารถรวบรวมเนื้อหาเดียวกันหรือที่คล้ายกันไว้ด้วยกัน

ลิงควิสต์ (Lindquist) ก็มีความเห็นในด้านการเรียงข้อสอบคล้ายกับข้อเสนอแนะของทราเวอร์ส กล่าวคือ ได้เสนอแนะวิธีจัดกลุ่มข้อคำถามเข้าไว้ด้วยกันว่า ถ้าข้อสอบเหล่านั้นอยู่ในเนื้อหาเดียวกัน และมีความยากง่ายพอ ๆ กันแล้วก็อาจสลับข้อกันได้

ไม่จำเป็นต้องเรียงลำดับกันตามความยาก แต่ถ้าข้อสอบเหล่านั้นอยู่ในเนื้อหาที่ต่างกัน และให้คะแนนแต่ละข้อเท่ากัน ก็จะต้องแยกกลุ่มข้อสอบออกเป็นฉบับย่อย ๆ (Subtest) และแยกเวลาในการทดสอบในแต่ละฉบับด้วย การเรียงลำดับข้อสอบในแต่ละฉบับนั้นอาจจะขึ้นอยู่กับความสัมพันธ์ของเนื้อหาวิชา คือ อาจให้สอบเนื้อหาทั่ว ๆ ไปก่อนแล้วจึงสอบเนื้อหาที่จำเพาะเจาะจงในตอนหลัง (28 : 179)

นูนนาลลี (Nunnally) ได้เสนอแนะว่าไม่ว่าจะเป็นแบบสอบเพื่อใช้ในการทำนายหรือวัดผลสัมฤทธิ์ทางการเรียน อาจเรียงลำดับข้อคำถามจากง่ายไปหายากเพื่อป้องกันไม่ให้นักเรียนเกิดความท้อใจ สำหรับแบบสอบที่ควรเรียงข้อคำถามตามระดับความยากนั้นจะต้องมีการทดลองสอบ และวิเคราะห์หาคุณภาพแล้ว ส่วนแบบสอบที่ยังไม่ได้ทดลองสอบและวิเคราะห์ ควรเรียงลำดับข้อคำถามแบบสุ่ม (29 : 159)

กรอนลันด์ (Gronlund) ได้เสนอแนวทางที่เป็นประโยชน์ในการจัดเรียงข้อคำถาม (30 : 104-106) ดังต่อไปนี้

1. ควรรวมกลุ่มข้อคำถามที่ต้องการวัดจุดมุ่งหมายเดียวกันไว้ด้วยกัน
2. คำถามที่เป็นแบบเดียวกันควรจัดเรียงไว้ด้วยกัน คือถ้าเป็นคำถามแบบเลือกคำตอบ คำตอบสั้น ข้อสอบความเรียง และอื่น ๆ ก็จัดรวมแบบเดียวกันไว้ด้วยกันเพราะการเรียงแบบนี้มีประโยชน์ต่อนักเรียนในการทำข้อสอบในแต่ละรูปแบบ และง่ายต่อการตรวจให้คะแนน
3. ข้อคำถามควรจัดเรียงตามลำดับความยากง่าย โดยควรเรียงจากข้อง่ายขึ้นไปหาข้อยากตามลำดับ ซึ่งวิธีนี้มีผลในค่านแรงจูงใจนักเรียน

ดูโรสต์ (Durost) ก็มีความเห็นทำนองเดียวกับ กรอนลันด์ คือ คำถามแต่ละรูปแบบควรรวมเป็นกลุ่ม ๆ กันไป ทั้งนี้เพื่อสะดวกในการทำของนักเรียน และการตรวจให้คะแนน ส่วนข้อคำถามนั้นควรเรียงตามลำดับความยากง่าย (31 : 39-40) ธอร์นไคค์ (Thorndike) ก็มีข้อเสนอแนะที่คล้ายกันว่า ข้อคำถามในรูปแบบเหมือนกัน (แบบถูกผิด แบบเลือกตอบ ฯลฯ) ควรรวมกลุ่มเดียวกัน ส่วนคำสั่งชี้แจงแยกตามรูปแบบของคำถามนั้น ๆ นอกจากนี้ควรเรียงลำดับคำถามจากง่ายไปหายาก และข้อคำถามที่เนื้อหาใกล้เคียงกันควรจะไว้ด้วยกันดีกว่าให้เนื้อหากระจายกันอยู่ (32 : 86)

จากข้อเสนอแนะเกี่ยวกับการจัดเรียงข้อความดังกล่าวข้างต้น สรุปได้ว่าวิธีการจัดเรียงข้อความในแบบทดสอบมีหลายวิธี เช่น เรียงจากง่ายไปหายาก เรียงความยากแบบหมุนเวียน เรียงตามเนื้อหาวิชา เรียงตามจุดมุ่งหมาย และเรียงแบบสุ่ม

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

งานวิจัยเกี่ยวกับการเรียงลำดับข้อความในต่างประเทศมีดังนี้

เบรนเนอร์ (Brenner) ได้ศึกษาการจัดเรียงลำดับความยากของข้อความวิชาจิตวิทยาการศึกษา โดยใช้ข้อความทั้งหมด 40 ข้อ กลุ่มตัวอย่างเป็นนักศึกษาของมหาวิทยาลัยโอไฮโอ (Ohio State University) ซึ่งเรียนวิชาเอกทางการศึกษา เขาได้แบ่งการศึกษาเป็น 4 ครั้ง ดังนี้คือ

ครั้งที่ 1 จัดเรียงข้อความ 3 วิธี คือ เรียงจากง่ายไปหายาก เรียงแบบสุ่ม และเรียงจากยากไปหาง่าย

ครั้งที่ 2 จัดเรียงข้อความ 2 วิธี คือ วิธีแรกเรียงจากข้อง่ายที่สุดไปหาข้อความลำดับ 10 ข้อแรก อีก 30 ข้อเรียงแบบสุ่ม วิธีที่สองเรียงจากข้อยากที่สุดไปหาข้อความลำดับ 10 ข้อแรก อีก 30 ข้อเรียงแบบสุ่ม

ครั้งที่ 3 จัดเรียงข้อความ 2 วิธี คือ เรียงจากง่ายไปหายาก และเรียงจากยากไปหาง่าย

ครั้งที่ 4 จัดเรียงข้อความ 2 วิธี คือ เรียงจากง่ายไปหายาก และเรียงจากยากไปหาง่าย

ผลการศึกษาพบว่า การสอบทั้ง 4 ครั้ง การจัดเรียงลำดับข้อความต่างกันไม่มีผลต่อความยาก และความเที่ยงของแบบสอบ ส่วนค่าอำนาจจำแนกในการสอบครั้งที่ 2 แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 (33 : 98-100)

คลอสเนอร์ และเกลดแมน (Klosner and Gellman) ได้ศึกษาเปรียบเทียบการเรียงข้อสอบในชั้นเรียน 3 วิธี คือ

1. เรียงตามหัวเรื่อง
2. เรียงจากง่ายไปหายาก
3. เรียงจากง่ายไปหายากในแต่ละหัวเรื่อง

กลุ่มตัวอย่าง เป็นนักศึกษาที่กำลังเรียนวิชาหลักการศึกษาระดับชั้น จำนวน 54 คน ผลการทดลองชี้ให้เห็นว่า โดยทั่วไปแล้วการเรียงข้อสอบวิธีต่าง ๆ กัน ไม่มีผลต่อการสอบของนักเรียน แต่สำหรับในกลุ่มนักเรียนที่มีความสามารถต่ำแล้ว การเรียงข้อคำถามมีผลทำให้ผลการสอบเปลี่ยนไปด้วย และค่าความเที่ยงของแบบสอบที่เรียงตามหัวเรื่อง และเรียงตามลำดับความยากง่ายมีค่าเท่ากัน ส่วนการเรียงจากง่ายไปหายากในแต่ละหัวเรื่องนั้นมีค่าความเที่ยงต่ำกว่า 2 วิธีแรก (12 : 413-418)

สเมาส์ และมันซ์ (Smouse and Munz) ได้ศึกษาการเรียงความยากง่ายของข้อคำถาม โดยจัดเรียงข้อคำถาม 3 แบบคือ เรียงจากง่ายไปหายาก เรียงจากยากไปหาง่าย และเรียงแบบสุ่ม กลุ่มตัวอย่าง เป็นนักศึกษามหาวิทยาลัยโอคลาโฮมาที่เรียนวิชาฟิสิกส์เบื้องต้นจำนวน 120 คน ผลการศึกษาพบว่า คะแนนการสอบจากการเรียงข้อคำถามต่างกันไม่แตกต่างกัน (34 : 370-374)

นอกจากนี้ เฟลค (Flake) ได้ทำการวิจัยเกี่ยวกับการจัดเรียงข้อคำถาม 3 แบบ คือ เรียงจากง่ายไปหายาก เรียงความยากแบบหมุนเวียน (Spiral Cyclical) และเรียงแบบสุ่ม กลุ่มตัวอย่าง เป็นนักเรียนพยาบาลจำนวน 104 คน ปรากฏผลว่าคะแนนสอบจากการจัดเรียงทั้ง 3 แบบไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 (35 : 56-58)

แต่จากการศึกษาของ แฟล็กเกอร์ เมลตัน และไมเออร์ส (Flaughter Melton and Myers) ได้ศึกษาว่าการจัดเรียงข้อสอบเสียใหม่จะมีผลทำให้คะแนนในการสอบเปลี่ยนแปลงหรือไม่ ข้อสอบที่ใช้ในการวิจัย คือ ข้อสอบด้านภาษา 40 ข้อ และข้อสอบคณิตศาสตร์ 25 ข้อ ข้อสอบทั้ง 2 ชนิดเป็นส่วนหนึ่งของแบบสอบ ซีอีบี สกอลาสติก แอปติจูด (CEEB Scholastic Aptitude Test) ซึ่งจัดแบ่งข้อสอบไว้เป็นตอน ๆ (blocks) ตอนหนึ่งมี 5 ข้อ ข้อสอบที่ใช้มีการจัดเรียง 4 แบบ คือ

1. เรียงตามมาตรฐานเดิม คือเรียงเป็นตอน ๆ (blocks) แต่ละตอนมี 5 ข้อ เรียงจากง่ายไปหายาก
2. จัดเรียงภายในแต่ละตอนใหม่ ส่วนการเรียงระหว่างตอนเหมือนมาตรฐานเดิม
3. จัดเรียงระหว่างตอนใหม่ ส่วนการเรียงภายในแต่ละตอนเหมือนมาตรฐานเดิม

4. จัดเรียงใหม่ทั้งภายในแต่ละตอน และระหว่างตอน

ผู้วิจัยทำการทดลองกับนักเรียนวิทยาลัย จำนวน 495 คน แบ่งเป็น 2 กลุ่ม ปรากฏผลว่า การจัดเรียงข้อสอบด้วยวิธีต่างกัน จะมีผลทำให้คะแนนจากแบบสอบคำณภาษาแตกต่างกัน แต่ไม่มีผลทำให้คะแนนจากแบบสอบคณิตศาสตร์แตกต่างกันเลยไม่ว่าจะเรียงข้อสอบด้วยวิธีใด แสดงว่าการจัดเรียงข้อสอบต่างวิธีกันจะทำให้ข้อสอบยากและง่ายแตกต่างกันในบางวิชา (36 : 813-824) ผลการวิจัยนี้สอดคล้องกับผลการวิจัยของ แมคนิคอลล (Mac Nicoll) ซึ่งเปรียบเทียบลักษณะการเรียงลำดับข้อสอบของแบบสอบ ซึ่งแต่เดิมเรียงจากง่ายไปหายาก ไปเป็นการเรียงแบบสุ่ม กับการเรียงจากข้อยากไปหาง่าย ปรากฏผลว่า การเรียงลำดับข้อสอบจากยากไปหาง่ายจะทำให้ข้อสอบมีค่าความยากขึ้นกว่าเดิมอย่างมีนัยสำคัญทางสถิติ (11 : 1-21)

นอกจากนี้ แฮมเบิลตัน และทรา็บ (Hambleton and Traub) ได้ศึกษาเกี่ยวกับการจัดเรียงลำดับแบบสอบคณิตศาสตร์ 2 วิธี คือ เรียงจากง่ายไปหายาก และเรียงจากยากไปหาง่าย ใช้กลุ่มตัวอย่าง 106 คน ปรากฏว่า การเรียงจากยากไปหาง่ายมีผลทำให้คะแนนเฉลี่ยต่ำกว่าเรียงจากง่ายไปหายากอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (37 : 40-46)

ส่วนการวิจัยของ คลิงเก (Kleinke) ซึ่งได้ทดลองศึกษากับเด็กเกรด 4 พบว่า การเรียงลำดับข้อสอบจากง่ายไปหายากมีผลทำให้คะแนนสอบของนักเรียนสูงกว่าการเรียงข้อสอบแบบสุ่ม (38 : 225-228)

สำหรับงานวิจัยในประเทศนั้นมีดังนี้

เตือนใจ เศรษฐสุกโก ได้ศึกษาถึงการจัดระเบียบข้อสอบวิชาภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ และภาษาไทย 4 วิธี คือ เรียงจากฉบับที่ง่ายไปหาฉบับที่ยาก และภายในแต่ละฉบับเรียงจากข้อง่ายไปยาก เรียงจากฉบับที่ยากไปฉบับที่ง่าย และภายในแต่ละฉบับเรียงจากข้อง่ายไปยาก เรียงฉบับที่ง่ายไปฉบับที่ยาก และภายในแต่ละฉบับเรียงข้อแบบสุ่ม เรียงข้อทดสอบแบบผสม แต่ละฉบับมีรวมทั้ง 4 วิชา วิชาละ 6 ข้อ เรียงจากวิชาที่ง่ายไปหายาก และภายในแต่ละวิชาเรียงจากข้อง่ายไปหายากรวม 5 ฉบับ แล้วเรียงแบบสอบจากฉบับที่ง่ายไปฉบับที่ยาก กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 7 จำนวน 240 คน ผลการวิจัยสรุปได้ว่า การเรียงข้อสอบด้วยวิธีต่างกันไม่มีผลทำให้ค่าความตรง

ตามสภาพของข้อสอบเปลี่ยนแปลง แต่การเรียงข้อสอบต่างกันในวิชาภาษาอังกฤษ และคณิตศาสตร์ ค่าความเที่ยงเปลี่ยนไปอย่างมีนัยสำคัญที่ระดับ .05 ส่วนผลสัมฤทธิ์ในการสอบของนักเรียนเมื่อจัดเรียงข้อสอบต่างกันแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 การเรียงข้อสอบจากฉบับที่ง่ายไปหาฉบับที่ยาก และภายในแต่ละฉบับเรียงจากข้อง่ายไปหาข้อยาก ให้ผลสัมฤทธิ์ในการสอบดีกว่าการเรียงข้อสอบจากฉบับที่ง่ายไปฉบับที่ยาก และภายในแต่ละฉบับเรียงแบบสุ่ม (39 : 45-49)

ศศิธร สุวรรณสุข ได้ศึกษาการจัดเรียงข้อคำถามวิชาคณิตศาสตร์ที่วัดทักษะการแก้ปัญหา และการหาเหตุผล ระดับชั้นประถมศึกษาปีที่ 7 ว่าการจัดเรียงแบบใดจะทำให้ค่าความตรง ค่าความเที่ยงของแบบสอบสูงที่สุด โดยจัดเรียงข้อคำถาม 4 วิธี คือ เรียงจากข้อง่ายไปหาข้อยาก เรียงจากข้อยากไปหข้อง่าย เรียงตามหัวข้อเนื้อหาวิชา และ เรียงจากข้อง่ายไปหาข้อยากในแต่ละหัวข้อเนื้อหาวิชา พบว่า การจัดเรียงข้อคำถามต่างกันไม่มีผลทำให้ค่าความตรงของแบบสอบวัดทักษะ การแก้ปัญหา และการหาเหตุผลเปลี่ยนแปลงอย่างมีนัยสำคัญที่ระดับ .05 การจัดเรียงทั้ง 4 วิธีทำให้ค่าความเที่ยงของแบบสอบวัดทักษะเปลี่ยนแปลงอย่างมีนัยสำคัญที่ระดับ .05 ส่วนค่าความเที่ยงของแบบสอบการแก้ปัญหา และการหาเหตุผลไม่เปลี่ยนแปลงอย่างมีนัยสำคัญที่ระดับ .05 และการจัดเรียงทั้ง 4 วิธี ทำให้คะแนนการสอบจากแบบสอบวัดทักษะ และการแก้ปัญหาคงกันอย่างมีนัยสำคัญที่ระดับ .05 ส่วนแบบสอบการหาเหตุผลคะแนนการสอบไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 (40 : 90-92)

แนนน้อย เพชรรัตน์ ได้ศึกษาเกี่ยวกับการเรียงลำดับข้อคำถามต่างกัน 3 วิธี คือ เรียงจากง่ายไปหายาก เรียงตามลำดับความยากแบบหมุนเวียน และเรียงแบบสุ่ม โดยมีได้คำนึงถึงเนื้อหาของคำถามนั้นว่าจะมีผลต่อคุณภาพของแบบสอบวิชาวิทยาศาสตร์ และแบบสอบทักษะขบวนการทางวิทยาศาสตร์อย่างไร กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2520 ในเขตจังหวัดนครปฐม จำนวน 300 คน ผลการศึกษาพบว่า การเรียงลำดับข้อคำถามต่างกันไม่ทำให้คุณภาพของแบบสอบด้าน ความตรง ความเที่ยง ความยาก และอำนาจจำแนกแตกต่างกันในแบบสอบทั้งสองฉบับ เว้นแต่ในแบบสอบทักษะขบวนการทางวิทยาศาสตร์ พบว่า การเรียงแบบสุ่มส่งผลต่อความเที่ยงสูงกว่าการเรียงคำถามอีก 2 วิธี (13 : 36-40)

สิริพร พุ่มแก้ว ได้ศึกษาการเรียงข้อความต่างกัน 4 วิธี ได้แก่ เรียงจากง่ายไปหายาก เรียงจากยากไปหาง่าย เรียงตามเนื้อเรื่อง และเรียงแบบสุ่ม ใช้แบบสอบวิชาภาษาไทย และคณิตศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชายและหญิง ชั้นมัธยมศึกษาปีที่ 3 จากโรงเรียนในกรุงเทพมหานคร จำนวน 120 คน ผลการวิจัยพอสรุปได้ดังนี้ การจัดเรียงข้อสอบทั้ง 4 แบบ มีผลต่อความเที่ยงและความตรงของแบบสอบเล็กน้อย คะแนนสอบวิชาภาษาไทยที่จัดเรียงตามเนื้อเรื่องแตกต่างจากคะแนนสอบจากแบบสอบที่เรียงวิธีอื่น ๆ อย่างมีนัยสำคัญที่ระดับ .01 นอกจากนี้คะแนนสอบวิชาคณิตศาสตร์ที่จัดเรียงจากง่ายไปหายาก แตกต่างจากการเรียงจากยากไปหาง่าย และการเรียงแบบสุ่มอย่างมีนัยสำคัญที่ระดับ .01 และ คะแนนสอบวิชาคณิตศาสตร์ที่จัดเรียงตามเนื้อเรื่องแตกต่างจากที่เรียงแบบสุ่มอย่างมีนัยสำคัญที่ระดับ .01 (41 : 30-36)

โสภณา เหลืองวิลาวงษ์ ได้ศึกษาเปรียบเทียบความยาก ความเที่ยง และความตรงของแบบสอบผลสัมฤทธิ์วิชาคณิตศาสตร์ ที่เรียงลำดับข้อแตกต่างกัน 4 แบบ คือ

1. เรียงลำดับข้อความจากข้อง่ายที่สุดไปหาข้อยากที่สุด โดยไม่คำนึงถึงเนื้อหา
2. เรียงลำดับข้อความตามลำดับเนื้อหาของหัวข้อเรื่องที่เรียน โดยความยากง่ายเรียงแบบสุ่ม
3. เรียงลำดับข้อความตามลำดับเนื้อหาของหัวข้อเรื่องที่เรียน และเรียงลำดับข้อในแต่ละเนื้อหาจากง่ายไปหายาก
4. เรียงลำดับข้อความแบบสุ่ม โดยไม่คำนึงถึงเนื้อหา หรือความยากง่ายของแต่ละข้อ

กลุ่มตัวอย่างได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 3 (ม.3) โรงเรียนพระโขนงพิทยาลัย กรุงเทพมหานคร จำนวน 320 คน และโรงเรียนพิบูลวิทยาลัย ลพบุรี จำนวน 320 คน ผลการวิจัยพบว่า การเรียงลำดับข้อความต่างกัน 4 วิธี ไม่ทำให้ค่าความเที่ยง ความตรง และความยาก แตกต่างกันอย่างมีนัยสำคัญ (14 : 58-100)

จากรายงานการวิจัยเกี่ยวกับการเรียงลำดับข้อความของแบบสอบที่กล่าวมาทั้งหมด จะเห็นได้ว่า ผลการวิจัยยังแยกออกเป็น 2 กลุ่ม กลุ่มหนึ่งพบว่า การเรียงลำดับข้อความที่ต่างกัน มีผลต่อคุณภาพของแบบสอบในด้าน ความยาก อำนาจจำแนก ความเที่ยง ()

และความตรง ตลอดจนคะแนนสอบ แต่อีกกลุ่มหนึ่งกลับพบว่า การเรียงลำดับข้อคำถามที่ต่างกันไม่มีผลต่อคุณภาพของแบบสอบ ตลอดจนคะแนนสอบ เป็นที่น่าสังเกตว่างานวิจัยที่ผ่านมาทั้งในประเทศ และต่างประเทศ ส่วนใหญ่เป็นการศึกษาในระดับชั้นที่ต่างกัน เนื้อหาวิชาที่ต่างกัน และงานวิจัยเหล่านี้มักจะศึกษาถึงผลของการเรียงลำดับข้อคำถามต่อคุณภาพด้านใดด้านหนึ่งของแบบสอบ หรือ คะแนนสอบ เช่น เบรนนเนอร์ ศึกษาถึงผลของการจัดเรียงข้อคำถามที่ต่างกันต่อค่าความยาก ความเที่ยง และอำนาจจำแนก เฟลค ศึกษาผลของการจัดเรียงข้อคำถามต่างกัน ต่อ คะแนนสอบของนักเรียน ศศิธร สุวรรณสุข ศึกษาถึงการจัดเรียงข้อคำถามต่างกันว่ามีผลต่อค่าความเที่ยง ความตรง และคะแนนสอบของนักเรียนหรือไม่ ส่วนเน่งน้อย เพชรรัตน์ ศึกษาถึงการจัดเรียงข้อคำถามที่แตกต่างกันต่อคุณภาพของแบบสอบในด้านความตรง ความเที่ยง ความยาก และอำนาจจำแนก ผู้วิจัยจึงมีความสนใจที่จะศึกษาถึงผลของการเรียงลำดับข้อคำถามที่แตกต่างกันต่อคุณภาพของแบบสอบในด้าน ความยาก อำนาจจำแนก ความเที่ยง และความตรง ตลอดจนคะแนนสอบของนักเรียน โดยอาศัยเอกสารและงานวิจัยเท่าที่ผ่านมาเป็นแนวทางในการศึกษาเปรียบเทียบครั้งนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย