



วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้อง จะแบ่งออกเป็น 2 ตอน คือ ตอนที่ 1 เป็นมโนทัศน์เบื้องต้นเกี่ยวกับแบบสอบแบบเลือกตอบและแบบถูกผิด ความเที่ยง ความยาก ความตรง และตอนที่ 2 เป็นงานวิจัยเกี่ยวกับการเปรียบเทียบค่าความเที่ยง ความยาก และความตรงของแบบเลือกตอบและแบบถูกผิด ซึ่งผู้วิจัยจะขอลาวรายละเอียดย่อยเป็นตอน ๆ ดังนี้

ตอนที่ 1 มโนทัศน์เบื้องต้นเกี่ยวกับแบบสอบแบบเลือกตอบและแบบถูกผิด ความเที่ยง ความยาก ความตรง

แบบสอบแบบเลือกตอบและแบบถูกผิด

ในปัจจุบันนี้แบบสอบแบบเลือก เป็นแบบสอบที่ได้รับความนิยมสูงและเป็นที่นิยมใช้กันมาก เพราะแบบสอบแบบเลือกตอบสามารถวัดความรู้ความสามารถได้ตั้งแต่ระดับความรู้ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์จนถึงการประเมินผล แต่การร่างแบบเลือกตอบที่ดีทำได้ยากมากและใช้เวลานาน ด้วยเหตุนี้จึงมีผู้นิยมใช้แบบสอบแบบถูกผิด เนื่องจากแบบถูกผิดสามารถตรวจให้คะแนนได้ง่าย สะดวก รวดเร็ว และเป็นปรนัยเช่นเดียวกับแบบเลือกตอบ แต่อย่างไรก็ตามก็วัดผลบางท่านได้กล่าวถึงข้อเสียของแบบถูกผิด ไว้ดังนี้ กรอนลันด์ (Gronlund) ได้กล่าวว่า แบบสอบแบบถูกผิดวัดสมรรถภาพทางสมองชั้นความรู้ความจำเท่านั้น ไม่สามารถวัดความคิดเห็น และการให้เหตุผลได้¹ ส่วนอาทแมนน์ และคอก (Ahmann and Glock) ได้กล่าวว่า แบบสอบแบบถูกผิดนั้นมีข้อบกพร่องในการใช้จำกัด และสแตนเลย์ (Stanley) ได้กล่าวว่า แบบสอบแบบถูกผิดควรนำมาใช้ เมื่อแบบสอบแบบอื่นไม่สามารถนำมาใช้ได้²

¹Gronlund, Measurement and Evaluation in Teaching, p.175.

²Mehrens and Lehmann, Measurement and Evaluation Educational and Psychology, p.261.

อ็อบเด เป็นผู้ที่ตั้งนิยามการไร้แบบสอบแบบถูกผิด โกลด์ลาวา จุดอ่อนของแบบสอบแบบถูกผิดในค่านความกำกวมมีได้เนื่องมาจากรูปแบบของข้อสอบ แต่เนื่องมาจากผู้สร้างข้อสอบขาดทักษะและความชำนาญ การที่หลายคนกล่าวว่าข้อสอบแบบถูกผิดสามารถวัดได้ในระดับความรู้ความจำเท่านั้น ไม่สามารถวัดกระบวนการทางสมองที่สูง ๆ ได้ เป็นข้อเสียที่เกิดจากผู้สร้างข้อสอบมากกว่าจะเกิดจากรูปแบบของข้อสอบ เจนกิน และดีโน (Jenkin and Deno) กล่าวว่า ข้อสอบแบบถูกผิดไม่จำเป็นต้องวัดความรู้ในระดับความรู้ความจำ ข้อสอบนี้สามารถเขียนให้วัดในระดับความเข้าใจ การนำไปใช้สรุปและการแก้ปัญหาได้ นอกจากนี้ เมเรนส์และเลแมนน์ (Mehrens and Lehmann) กล่าวว่า ในเวลาที่เท่ากัน ข้อสอบแบบถูกผิดสามารถวัดเนื้อหาได้มากกว่าแบบปรนัยชนิดอื่น ๆ และสามารถปรับให้เข้ากับเนื้อหาวิชาใดมาก

การเคาในข้อสอบแบบถูกผิด

ปัญหาที่สำคัญอีกอย่างหนึ่งในแบบถูกผิด คือ การเคา เป็นที่ถกเถียงกันว่าข้อสอบแบบถูกผิดจะมีโอกาสให้นักเรียนเคาถูก 50 % ซึ่งเป็นการกระตุ้นให้นักเรียนเคา แต่อย่างไรก็ดี อ็อบเด ได้ให้ความคิดเห็นเกี่ยวกับการเคาในข้อสอบแบบถูกผิด ดังนี้

ประการแรก คือ ความแตกต่างระหว่างการเคาอย่างมีความรู้และการเคาสุ่ม การเคาสุ่ม หมายถึง การเคาที่ไม่ได้ใช้ความรู้อะไรเลย และทำให้คะแนนเกิดความคลาดเคลื่อน ส่วนการเคาอย่างมีความรู้ซึ่งมีความตรงในการวัดผลสัมฤทธิ์ เพราะการที่นักเรียนมีความรู้มาก จะทำให้โอกาสเคาถูกมีมากด้วย

ประการที่สอง คือ การดึงดูดความสนใจของนักเรียน การกำหนดให้แบบสอบมีความยากและเวลาที่เหมาะสมจะทำให้นักเรียนลดการเคาสุ่มน้อยลงไป นักเรียนเหล่านั้นยอมจะทราบดีว่าการคิดขอมทำให้เกิดความมั่นใจในการทำมากกว่าการเคาสุ่มในการวิจัยของอ็อบเด เรื่องการเคาสุ่มในแบบสอบสัมฤทธิ์ผลแบบปรนัย พบว่า โดยเฉลี่ยจะมีข้อสอบ 1 ข้อ ในจำนวน 20 ข้อ

¹Ibid., pp.264-266.

ที่นักเรียนจะคาดเดา

ประการที่สาม คือ อิทธิพลของการคาดเดาจะลดลงเมื่อความยาวของแบบสอบเพิ่มขึ้น ข้อสอบแบบถูกผิดหนึ่งข้อมีโอกาสเดาได้ถูกต้อง 50% แต่โอกาสที่จะทำถูก 2 ข้อมีเพียง 25% และ 10 ข้อมีเพียง 0.1% และในข้อสอบ 100 ข้อ จะมีโอกาสเดาถูกหมดคนน้อยมาก และโอกาสที่จะเดาข้อสอบแบบถูกผิดได้ถูกต้องเป็นจำนวน 70 ข้อ จากข้อสอบทั้งหมด 100 ข้อนั้นมีโอกาสน้อยกว่า 1 ใน 1,000

ประการที่สี่ คือ ถ้ามีความเชื่อว่าคะแนนที่ได้จากแบบสอบแบบถูกผิดมาจากการเดา คะแนนที่ได้เหล่านั้นก็จะเป็นความบังเอิญ แต่ความจริงแล้วค่าความเที่ยงของแบบสอบแบบถูกผิดจำนวน 100 ข้อ มีค่าความเที่ยงสูงถึง .85 ถึง .95 ซึ่งเป็นค่าที่ค่อนข้างสูง และเป็นค่าที่เรากำหนดว่าแบบสอบที่วัดผลสัมฤทธิ์ในห้องเรียนที่ควรเป็น ดังนั้นการคำนึงถึงเรื่องการเดามากเกินไปโดยไม่นิยมนำแบบสอบแบบถูกผิดที่ตีไปใช้ก็เป็นสิ่งที่น่าเสียดาย

บางคนมีความเชื่อว่า ปัญหาการเดาสามารถแก้ไขได้โดยการปรับคะแนนการเดา แต่อีเบล มีความคิดเห็นว่าเป็นความเชื่อที่ผิด เขากล่าวว่า การที่แบบสอบต้องแก้การเดาเพื่อป้องกันการเดานั้นเป็นเพียงแสดงให้เห็นถึงความแตกต่างระหว่างผู้เดาที่โชคดีกับผู้เดาที่โชคร้ายเท่านั้น²

การเขียนข้อสอบแบบถูกผิดที่มีอำนาจจำแนก

หน้าที่ของข้อสอบที่สำคัญคือ ต้องจำแนกเด็กที่มีความรู้และเด็กที่ไม่มีความรู้ ดังนั้นการเขียนข้อสอบแบบถูกผิดใหม่มีอำนาจจำแนกจึงเป็นสิ่งสำคัญและมีวิธีการ ดังนี้คือ

1. ให้มีจำนวนข้อผิดมากกว่าข้อถูกในแบบสอบแบบถูกผิด

¹Robert L. Ebel, "Blind Guessing on Objective Achievement Tests" Journal of Educational Measurement, 5 (1968): 321 - 325.

²Ebel, Essentials of Educational Measurement, pp. 116 - 117.

เมื่อนักเรียนสงสัยสิ่งใดนักเรียนจะยอมรับประโยชน์นั้นมากกว่าที่จะปฏิเสธ นักวิชาการหลายคนได้พบว่า ข้อดีมีอำนาจจำแนกเด็กเก่งและเด็กอ่อนไวกว่าข้อด้อย สิ่งนี้อาจเป็นผลเนื่องมาจากสิ่งที่เรียกว่า "acquiescent response set" กล่าวคือ ในกรณีที่นักเรียนขาดความรู้ในเรื่องนั้น นักเรียนจะยอมรับว่าประโยชน์นั้นจริง นักวิจัยบางคนแนะนำว่า ในการสร้างข้อสอบแบบถูกผิดควรประกอบด้วยจำนวนข้อถูกและข้อผิดเท่า ๆ กัน แต่หาข้อดีมีแนวโน้มที่จะมีอำนาจจำแนกสูง จึงเป็นประโยชน์ที่จะมีข้อดีมากกว่าข้อด้อยในแบบถูกผิด และอีเบลเสนอว่า ควรมีจำนวนข้อดีประมาณ 67%

2. ข้อสอบแบบถูกผิดที่เป็นข้อดีควรสร้างให้มีสภาพการผสมผสานผล
3. ควรใช้ข้อสอบแบบถูกผิดที่เป็นข้อดีวัดคุณโน้ตที่ผิด ๆ หรือความเชื่อ

ที่ผิด¹

ความเที่ยง

นิยามของความเที่ยง

ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการทดสอบไม่ว่าจะทำการวัดเมื่อใดก็ตาม²

ทฤษฎีของความเที่ยง

ทฤษฎีของความเที่ยงเป็นทฤษฎีที่กล่าวถึง ความคงที่ของคะแนนจริง ทฤษฎีของความเที่ยงสามารถอธิบายได้โดยเริ่มต้นจากคะแนนสอบที่นักเรียนทำได้ (Observed Score) ซึ่งคะแนนที่นักเรียนทำได้นั้นประกอบด้วย คะแนนจริง (True Score) กับคะแนนที่เกิดจากความคลาดเคลื่อน

¹Ibid., pp.129 - 130.

²Ibid., p.310.

(Error Score) ซึ่งเขียนเป็นสมการได้ ดังนี้¹

$$X = T + E$$

เมื่อ X = คะแนนที่นักเรียนทำได้

T = คะแนนจริง

E = คะแนนที่เกิดจากความคลาดเคลื่อน

คะแนนจริง (True Score) หมายถึง คะแนนที่ผู้สอบพึงได้รับจากการวัดด้วยเครื่องมือที่มีคุณลักษณะสมบูรณ์ปราศจากความคลาดเคลื่อน หรือหมายถึง คะแนนเฉลี่ยของผู้สอบซึ่งได้จากการทำแบบสอบเกมหลาย ๆ ครั้ง โดยมีข้อตกลงว่า ไม่มีอิทธิพลจากการฝึกฝน ความเมื่อยล้า และการเรียนรู้ในการทดสอบ²

คะแนนที่เกิดจากความคลาดเคลื่อน (Error Score) หมายถึง ค่าความผิดพลาดที่เกิดขึ้นจากการวัด ซึ่งเป็นความคลาดเคลื่อนที่เกิดขึ้นในลักษณะสุ่ม (Random Error) เช่น ความคลาดเคลื่อนที่เกิดขึ้นจากตัวแบบสอบ การบริหารการสอบ การให้คะแนน และสิ่งต่าง ๆ ที่เกิดจากตัวผู้เขาสอบ เป็นต้น โดยทฤษฎีความคลาดเคลื่อนที่เกิดขึ้นในลักษณะสุ่มนี้ อาจเป็นได้ทั้งทางบวกและทางลบ ซึ่งจะหักล้างกันหมดไป การแจกแจงของคะแนนความคลาดเคลื่อนเป็นการแจกแจงแบบปกติที่มีค่าเฉลี่ย (Mean) เท่ากับ 0 และค่าความเบี่ยงเบนมาตรฐาน (Standard Deviation) เท่ากับ 1 และถ้าความคลาดเคลื่อนเกิดขึ้นในลักษณะสุ่มแล้วก็จะไม่มีความสัมพันธ์กับคะแนนจริง นั่นคือสัมประสิทธิ์สหสัมพันธ์ (Correlation) ระหว่างคะแนนจริงกับคะแนนความคลาดเคลื่อนมีค่าเท่ากับ 0

¹Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p.90.

²Frederick G. Brown, Principle of Educational and Psychological Testing, 2d ed. (New York: Holt, Rinehart and Winston, 1976), p.53.

ดังนั้นความแปรปรวนของคะแนนที่วัดได้ สามารถเขียนเป็นสมการได้ ดังนี้¹

$$s_x^2 = s_t^2 + s_e^2$$

เมื่อ s_x^2 = ความแปรปรวนของคะแนนที่วัดได้

s_t^2 = ความแปรปรวนของคะแนนจริง

s_e^2 = ความแปรปรวนของคะแนนความคลาดเคลื่อน

โดยทฤษฎี ความเที่ยง หมายถึง อัตราส่วนระหว่างความแปรปรวนของคะแนนจริง กับความแปรปรวนของคะแนนที่วัดได้ นั่นคือ

$$r_{tt} = \frac{s_t^2}{s_x^2} \quad (1)$$

เมื่อ r_{tt} = ค่าความเที่ยง

จากสมการ

$$s_x^2 = s_t^2 + s_e^2$$

หารตลอดด้วย s_x^2 จะได้

$$\frac{s_x^2}{s_x^2} = \frac{s_t^2}{s_x^2} + \frac{s_e^2}{s_x^2}$$

$$1 = r_{tt} + \frac{s_e^2}{s_x^2}$$

$$\therefore r_{tt} = 1 - \frac{s_e^2}{s_x^2} \quad (2)$$

จากสมการ(2) จะพบว่า ค่าความเที่ยงที่สูงที่สุดจะเท่ากับ 1

ซึ่งมันเป็นเพียงมีโนทัศน์ของความเที่ยงและไม่สามารถคำนวณหาความเที่ยงได้

¹Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p.91.

ดังนั้นในการหาความเที่ยงในทางปฏิบัติ จึงประมาณค่าของความเที่ยงโดยอาศัยสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของนักเรียนกลุ่มเดียวกัน ดังรายละเอียดต่อไปนี้¹

วิธีประมาณค่าความเที่ยง

วิธีประมาณค่าความเที่ยงสามารถทำได้หลายวิธี แต่ละวิธีก็ทำให้เกิดความคลาดเคลื่อนแตกต่างกัน การที่จะเลือกใช้วิธีใดย่อมขึ้นอยู่กับความมุ่งหมาย ลักษณะของคะแนน และชนิดของแบบสอบ โดยทั่วไปวิธีประมาณค่าความเที่ยงมี 4 แบบ² คือ

1. แบบสอบซ้ำ (Measure of Stability)
2. แบบไขว้ขอสอบคล้ายกัน (Measure of Equivalence)
3. แบบไขว้ขอสอบคล้ายกันและสอบซ้ำ (Measure of Equivalence and Stability)
4. แบบวัดความคงที่ภายใน (Measure of Internal Consistency)

ซึ่งแบ่งออกเป็น

- 4.1 แบบแบ่งครึ่งจำนวนขอสอบ (Split-half)
- 4.2 แบบคูเคอร์ ริชาร์ดสัน (Kuder-Richardson Estimates)
- 4.3 แบบวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's Analysis of Variance Procedure)

ในที่นี้จะกล่าวเฉพาะความเที่ยงแบบคูเคอร์ ริชาร์ดสัน ซึ่งเป็นวิธีประมาณค่าความเที่ยงที่ใช้ในการวิจัยครั้งนี้

¹Fred N. Kerlinger, Foundations of Behavioral Research, (New York: Holt, Rinehart and Winston, 1979), pp.444-446.

²Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p.109.

ความเที่ยงแบบคูเคอร์ ริชาร์ดสัน

ในปี ค.ศ. 1937 คูเคอร์ และริชาร์ดสัน (Kuder and Richardson) ได้เสนอสูตรสำหรับประมาณค่าความเที่ยง ซึ่งเป็นที่รู้จักกันและนิยมใช้กันอย่างแพร่หลาย คือ สูตรคูเคอร์ ริชาร์ดสัน 20 (Kuder-Richardson 20)¹

$$r = \frac{k}{k-1} \left[1 - \frac{\sum pq}{\sigma^2} \right]$$

เมื่อ r = ค่าความเที่ยงของแบบสอบ

k = จำนวนข้อในแบบสอบ

p = สัดส่วนของผู้ตอบถูกในแต่ละข้อ

q = สัดส่วนของผู้ตอบผิดในแต่ละข้อ

σ^2 = ความแปรปรวนของคะแนนทั้งหมด

การหาค่าความเที่ยงแบบคูเคอร์ ริชาร์ดสันนี้ เป็นการหาความสอดคล้องกันระหว่างข้อสอบ ซึ่งดำเนินการสอบครั้งเดียวและใช้แบบสอบชุดเดียว ความสอดคล้องกันระหว่างข้อนี้ได้รับอิทธิพลจากแหล่งความแปรปรวนคลาดเคลื่อน 2 แหล่ง คือ

1. เนื้อหาที่สุ่ม

2. ความเป็นวิวิธพันธ์ของพฤติกรรมที่สุ่ม ยิ่งข้อสอบมีความเป็นเอกพันธ์มาก

ความสอดคล้องกันระหว่างข้อก็ยิ่งสูง

ค่าความเที่ยงแบบคูเคอร์ ริชาร์ดสัน มีค่าประมาณได้กับค่าเฉลี่ยของสัมประสิทธิ์ความเที่ยงที่ได้จากการแบ่งครึ่งแบบสอบควยวิธีต่าง ๆ กัน²

¹Ebel, Measuring Educational Achievement, p.318.

²แอน อนาสตาซี, การตรวจสอบเชิงจิตวิทยา แปลโดย ประชุมสุข อชาวอำรุงและคณะ (กรุงเทพมหานคร: สำนักพิมพ์ไทยวัฒนาพานิช, 2519), หน้า 84.

ข้อตกลงเบื้องต้นของสูตรคูเคอร์ ริชาร์ดสัน 20 คือ

1. การตรวจให้คะแนนแต่ละข้อเป็นแบบ 0 - 1 คือ ตอบถูกให้ 1 คะแนน
ตอบผิดให้ 0 คะแนน
2. ข้อสอบในแบบสอบจะต้องมีลักษณะ เป็นเอกพันธ์ คือ วัตถุประสงค์
เดียวกัน¹

สูตรคูเคอร์ ริชาร์ดสัน 20 เหมาะสำหรับประมาณค่าความเที่ยงของแบบสอบ
ความสามารถ (Power Test) เท่านั้น ไม่เหมาะสำหรับแบบสอบความเร็ว (Speed Test)
เพราะค่า p และ q ของแต่ละข้อจะต้องเป็นค่าที่ได้จากการที่ผู้สอบทุกคนมีโอกาสทำข้อนั้นแล้ว ซึ่ง
เป็นไปได้ที่ทุกคนจะทำถึงข้อสุดท้ายในข้อสอบแบบความเร็ว²

เนื่องจากสูตรคูเคอร์ ริชาร์ดสัน 20 ไม่สะดวกในการคำนวณ คูเคอร์และ
ริชาร์ดสัน จึงเสนอสูตรที่สามารถคำนวณได้ง่ายและสะดวกกว่า คือ สูตรคูเคอร์ ริชาร์ดสัน 21
(Kuder-Richardson 21) ซึ่งสูตรนี้ข้อตกลงว่าข้อสอบแต่ละข้อมีความยากโดยเฉลี่ยเท่ากัน
หรือกำหนดให้ค่า p คงที่ และค่าที่ได้จากสูตรนี้จะต่ำกว่าสูตรคูเคอร์ ริชาร์ดสัน 20

ความยาก

นิยามของความยาก

ความยากของข้อสอบ หมายถึง สัดส่วนของนักเรียนที่ตอบข้อสอบข้อนั้นถูก นั่นคือ
ค่าความยากสามารถคำนวณได้จากจำนวนนักเรียนที่ตอบข้อสอบนั้นถูกหารด้วยจำนวนนักเรียนที่ตอบ
ข้อสอบนั้นทั้งหมด³

¹Ebel, Measuring Educational Achievement, p.318.

²Brown, Principle of Educational Achievement, p.326.

³Donald L. Beggs and Ernest L. Lewis, Measurement and Evaluation in the Schools (Boston: Houghton Mifflin Co., 1975), p.195.

จากคำนิยามที่กำหนดให้ แสดงให้เห็นว่า ความยากของข้อสอบมีค่ามากที่สุดเท่ากับ -1 และน้อยที่สุดเท่ากับ 0 ค่าความยากของข้อสอบที่ใกล้ 0 แสดงว่า ข้อสอบขงนั้นยากมาก และถาคค่าความยากของข้อสอบที่ใกล้ $+1$ ย่อมแสดงว่า ข้อสอบขงนั้นง่าย เนื่องจาก แบบสอบมีจุดมุ่งหมายที่จะจำแนกนักเรียนดังนั้นข้อสอบที่มีค่าความยากใกล้ $+1$ จึงไม่มีประโยชน์ เพราะข้อสอบขงนั้นนักเรียนทุกคนทำได้ และข้อสอบนั้นมิได้จำแนกนักเรียนได้ว่ามีความรู้ในระดับใด ตามการวัดผลแบบอิงกลุ่ม ถ้าข้อสอบมีค่าความยากเท่ากับ 0 ข้อสอบนั้นก็มิได้จำแนกนักเรียนเช่นเดียวกัน เพราะทุกคนทำข้อสอบนั้นไม่ได้ ดังนั้นข้อสอบจึงควรมีค่าความยากมากกว่า 0 และน้อยกว่า $+1$ ¹ นอกจากนี้ ความยากของข้อสอบยังขึ้นอยู่กับจุดมุ่งหมายของการวัด สำหรับในการวัดแบบอิงเกณฑ์ (Criterion Reference) ข้อสอบจะมีค่าความยากสูง ทำให้การกระจายของคะแนน มีลักษณะ เบ้ชาย กล่าวคือ คะแนนของนักเรียนส่วนใหญ่จะสูง ซึ่งจะช่วยให้เราทราบได้ว่ามีนักเรียนคนใดบ้างที่มีความรู้และทักษะไม่ถึง เกณฑ์ที่กำหนด ส่วนในการวัดแบบอิงกลุ่ม (Norm Reference) นั้น ต้องการให้มีการกระจายของคะแนนกว้างมากและมีค่าความเที่ยงมากที่สุดเท่าที่เป็นไปได้ นั่นคือ ต้องการทำให้การกระจายของคะแนนมีมากที่สุด ซึ่งสามารถทำได้โดยกำหนดค่าความยากของข้อสอบเท่ากับ $.50$ ข้อสอบที่มีค่าความยากระดับนี้ จะทำให้มีการกระจายของคะแนนมากที่สุด ทั้งนี้ในการกำหนดระดับความยากของแบบสอบนั้น จึงขึ้นอยู่กับจุดมุ่งหมายหรือการนำผลการสอบไปใช้ นอกจากนี้ยังขึ้นอยู่กับระดับความสามารถของนักเรียน อายุ และชั้นเรียนอีกด้วย²

ค่าความยากมีความสัมพันธ์กับค่าอำนาจจำแนก กล่าวคือ ค่าอำนาจจำแนกของข้อสอบจะมีค่าสูงสุดเมื่อข้อสอบมีค่าความยากเท่ากับ $.50$ แต่การที่ข้อสอบมีค่าความยากเท่ากับ $.50$

¹ Ibid.

² Gronlund, Measurement and Evaluation in Teaching, pp.120-121.

อาจทำให้ข้อสอบนั้นมีค่าอำนาจจำแนกสูงสุดหรือไม่เลยก็ได้ เพราะถาครั้งหนึ่งของจำนวนคนในกลุ่ม
สูงตอบถูก และครั้งหนึ่งของจำนวนคนในกลุ่มทำตอบถูก ก็จะทำให้ค่าอำนาจจำแนกเท่ากับ 0

ความตรง

นิยามของความตรง

ความตรง หมายถึง ความสามารถที่จะวัดในสิ่งที่ต้องการวัดได้ถูกต้องแม่นยำ หรือ
คือ ค่าสหสัมพันธ์ระหว่างแบบสอบกับเกณฑ์²

ชนิดของความตรง

การแบ่งประเภทของความตรงมีหลายวิธี ซึ่งขึ้นอยู่กับจุดมุ่งหมายของการวัด
โดยทั่วไปในการวัดผลทางการศึกษาและทางจิตวิทยาได้แบ่งความตรงออกเป็น 3 ประเภท³ คือ

1. ความตรงตามเนื้อหา (Content Validity)
2. ความตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity)
3. ความตรงตามทฤษฎี (Construct Validity)

ในที่นี้จะกล่าวถึงความตรงตามเนื้อหา และความตรงตามเกณฑ์สัมพันธ์ ซึ่งเกี่ยวข้อง
กับการวิจัยนี้เท่านั้น

¹E.F. Linquist, A First Course in Statistic, (Boston: Houghton
Mittlin, 1942), p.213.

²H. Gullisen, Theory of Mentals Tests, (New York: John Willey
& Son, 1950), p.88.

³Mehrens and Lehmann, Measurement and Evaluation Education and
Psychology, . pp.109-110.

ความตรงตามเนื้อหา

ความตรงตามเนื้อหา คือ ความตรงที่เกี่ยวกับการสุ่ม (Sampling) เนื้อหาวิชาที่ใช้ในการทดสอบ โดยเฉพาะอย่างยิ่งการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียน (Achievement Test) มักจะเน้นในเรื่องความเที่ยงในเนื้อหาวิชา การที่จะตัดสินว่าแบบสอบนั้นมีความตรงในเนื้อหาวิชามากน้อยเพียงใด จะพิจารณาทั้งหัวข้อเนื้อหาวิชา (Subject Matter) และชนิดของพฤติกรรมที่เราต้องการจะวัดไปพร้อมกัน¹ อย่างไรก็ตาม เราไม่สามารถแสดงค่าความตรงในเนื้อหาวิชาของแบบสอบออกมาเป็นตัวเลขได้ โดยทั่วไปมักจะพิจารณาข้อสอบในแบบสอบโดยเทียบกับผังการสร้างข้อสอบ แล้วประมาณออกมาว่ามีความตรงตามเนื้อหา มากน้อยเพียงใด ผู้ที่จะตัดสินความตรงตามเนื้อหาได้คือ ผู้เชี่ยวชาญในสาขาวิชานั้น ๆ²

ความตรงตามเกณฑ์สัมพัทธ์

ความตรงตามเกณฑ์สัมพัทธ์ เป็นวิธีการศึกษาความสัมพันธ์ระหว่างคะแนนการทดสอบและเกณฑ์ภายนอก (External Criteria) ความตรงตามเกณฑ์สัมพัทธ์จำแนกได้เป็น 2 ชนิด³ คือ

1. ความตรงร่วมสมัย (Concurrent Validity) หมายถึง ความตรงของแบบสอบที่จะบอกสิ่งที่วัดได้ถูกต้องตามสภาพที่เป็นอยู่ในปัจจุบัน โดยหาความสัมพันธ์ระหว่างคะแนนของแบบสอบกับคะแนนเกณฑ์ที่เกิดขึ้นในระยะเวลาใกล้เคียงกัน เช่น ต้องการทราบความตรงของแบบสอบวัดการวัดกลมมาตราเงินกับตัวแปร เกณฑ์อื่น ๆ ก็อาจหาได้โดยหาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบกับการสังเกตการทอนสตางค์จริงของนักเรียนในขณะนั้น ถ้าคะแนนที่นักเรียนได้จากแบบสอบให้ผลสอดคล้องกับ เกณฑ์สัมพัทธ์ก็แสดงว่าแบบสอบนั้นมีความตรงร่วมสมัย

¹อนันต์ ศรีโสภณ, การวัดและประเมินผลการศึกษา, หน้า 69-70.

²Mehrens and Lehman, Measurement and Evaluation and Psychology, p. 112.

³อนันต์ ศรีโสภณ, การวัดและประเมินผลการศึกษา, หน้า 70.

2. ความตรงเชิงทำนาย (Predictive Validity) หมายถึง ความตรงของแบบสอบที่จะบอกสิ่งที่วัดได้ในขณะนั้นได้ถูกต้องตามสภาพที่เป็นอยู่ในอนาคต โดยหาความสัมพันธ์ระหว่างคะแนนของแบบสอบ กับคะแนนเกณฑ์ที่เกิดขึ้นในอนาคต เช่น การใช้แบบสอบวัดความถนัดทางการเรียนทำนายผลการเรียน ก็อาจใช้คะแนนเฉลี่ยสะสม (G.P.A.) ปีสุดท้ายเป็นเกณฑ์ การหาความตรงชนิดนี้ต้องอาศัยเวลาเพราะต้องหาความสัมพันธ์ระหว่างคะแนนสอบความถนัดกับเกณฑ์ซึ่งจะเกิดขึ้นภายหลัง ความตรงเชิงทำนายเหมาะสำหรับใช้กับแบบสอบเพื่อทำนายผล

ความแตกต่างระหว่างความตรงรวมสัมพันธ์กับความตรงเชิงทำนาย มีสาระสำคัญอยู่ 2 ประการ¹ คือ

1. ช่วงระยะเวลาของการเก็บข้อมูลที่เป็นเกณฑ์ ถ้าเก็บรวบรวมข้อมูลที่เป็นเกณฑ์ และคะแนนสอบในเวลาเดียวกันก็จะเป็นความตรงรวมสัมพันธ์ แต่ถ้าเก็บคะแนนที่เป็นเกณฑ์ล่าออกไปหรือหลังจากการเก็บคะแนนสอบ จะเป็นการตรงเชิงทำนาย
2. ความมุ่งหมายของการสอบ ความตรงเชิงทำนายเป็นการใช้คะแนนสอบในการพยากรณ์ความสามารถหรือการกระทำในอนาคต ส่วนความตรงรวมสัมพันธ์เป็นการนำคะแนนสอบมาเทียบกับความสามารถในปัจจุบัน

ในการหาความตรงตามเกณฑ์สัมพันธ์นั้น สิ่งที่สำคัญที่สุดคือ การวัดเกณฑ์ในการวัดเกณฑ์นั้นมีสิ่งต้องคำนึงอยู่ 3 ประการ² คือ

1. ความตรงประเด็น (Relevance) สิ่งที่ใช้เป็นเกณฑ์ต้องมีความตรงประเด็น หรือสะท้อนถึงคุณลักษณะที่สำคัญของเรื่องที่จะศึกษา โดยทั่วไปการพิจารณาความตรงประเด็นของ เกณฑ์ขึ้นอยู่กับ การตัดสินใจคุณค่า

¹อนันต์ ศรีโสภกา, การวัดและประเมินผลการศึกษา, หน้า 70.

²Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, pp.113-114.

2. ความเที่ยง (Reliability) เนื่องจากความเที่ยงของแบบสอบมีอิทธิพลต่อความสัมพันธ์ระหว่างแบบสอบกับเกณฑ์ จากทฤษฎีความสัมพันธ์กล่าวว่า ความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรจะเท่ากับหรือน้อยกว่ารากกำลังที่สองของผลคูณระหว่างค่าความเที่ยงของตัวแปรทั้งสองตัวนั้น นั่นคือ

$$r_{xy} \leq \sqrt{(r_{xx})(r_{yy})}$$

เมื่อ r_{xy} แทน ค่าสหสัมพันธ์ระหว่างตัวทำนาย (X) กับเกณฑ์ (Y)
 r_{xx} แทน ค่าความเที่ยงของแบบสอบที่เป็นตัวทำนาย
 r_{yy} แทน ค่าความเที่ยงของเกณฑ์

นั่นคือ ค่าความเที่ยงของเกณฑ์มีอิทธิพลต่อค่าความตรงตามเกณฑ์สัมพันธ์ พอ ๆ กับค่าความเที่ยงของตัวทำนาย (r_{xx})

3. ความเป็นอิสระปราศจากอคติ (Free from Bias) คือ เกณฑ์ที่จะวัดควรเป็นอิสระและไม่มีอิทธิพลจากการสอบที่นำมาใช้เป็นตัวทำนาย เช่น ครูนำแบบสอบความถนัดทางวิชาการมาทดสอบนักเรียนคนหนึ่ง ถ้าครูพบว่านักเรียนคนใดใคร่คะแนนความถนัดต่ำ ก็จะประเมินผลการเรียนให้นักเรียนผู้นั้นต่ำไปด้วย นั่นคือการรู้คะแนนตัวทำนายมีอิทธิพลต่อคะแนนเกณฑ์ อิทธิพลเช่นนี้จะมีผลทำให้สหสัมพันธ์ระหว่างคะแนนสอบกับ เกณฑ์สูงขึ้น อย่างไรก็ตามปัญหานี้อาจแก้ไขได้โดยไม่ใหญ่หนาที่ตีค่าเกณฑ์ (Rater) รู้คะแนนสอบที่เป็นตัวทำนาย

วิธีประมาณค่าความตรงตามเกณฑ์สัมพันธ์

การประมาณค่าความตรงตามเกณฑ์สัมพันธ์สามารถวิเคราะห์ได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบกับเกณฑ์ สูตรที่ใช้ในการคำนวณคือสูตรสหสัมพันธ์แบบต่าง ๆ โดยทั่วไปนิยมใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's Product Moment Correlation Coefficient)

ตอนที่ 2 งานวิจัยเกี่ยวกับการเปรียบเทียบค่าความเที่ยง ความยากและความตรงของแบบสอบแบบเลือกตอบและแบบถูกผิด

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบค่าความเที่ยง และความยากของแบบสอบแบบเลือกตอบและแบบถูกผิด พบว่า ผลการวิจัยเกี่ยวกับความเที่ยงแยกออกเป็น 2 ตอน คือ พวกหนึ่ง พบว่า ความเที่ยงของแบบสอบทั้งสองชนิดไม่แตกต่างกัน และอีกพวกหนึ่ง พบว่า ความเที่ยงของแบบสอบทั้งสองแตกต่างกัน ทั้งรายละเอียดต่อไปนี้

ผลการวิจัยที่พบว่าความเที่ยงของแบบสอบทั้งสองไม่แตกต่างกัน

รินส์แลนด์ (Rinsland) ได้รวบรวมผลการทดลองที่เกี่ยวข้องกับแบบสอบต่าง ๆ และเขาได้สรุปไว้ว่า แบบสอบแบบปรนัยทั้งหมดยกเว้นแบบถูกผิด มีค่าความตรงพอ ๆ กันกับแบบอัตนัยหรือแบบกาเลือกตอบ ส่วนในกรณีความเที่ยง ปรากฏว่า แบบสอบแบบปรนัยทุกชนิดให้ค่าความเที่ยงพอ ๆ กันเมื่อใช้เวลาในการสอบเท่ากัน แต่การที่จะมีความเที่ยงต่างกันนั้นอาจเป็นเพราะการใช้คำหรือภาษาที่แตกต่างกันมากกว่าจะเป็นเพราะรูปแบบของคำถามหรือข้อสอบ

เมเยอร์ (Mayer) ได้ศึกษาผลของการใช้แบบสอบแบบอัตนัย แบบถูกผิด แบบเลือกตอบ และแบบเติมคำ โดยใช้กลุ่มตัวอย่างเป็นนักเรียน จำนวน 124 คน แบ่งเป็น 4 กลุ่ม แล้วทดสอบ ดังนี้

- กลุ่มที่ 1 สอบแบบสอบแบบอัตนัย
- กลุ่มที่ 2 สอบแบบสอบแบบถูกผิด
- กลุ่มที่ 3 สอบแบบสอบแบบเลือกตอบ
- กลุ่มที่ 4 สอบแบบสอบแบบเติมคำ

¹Rinsland, Constructing Tests and Grading in Elementary and High School Subjects (New York: Prentice-Hall Inc., 1938), pp.295 - 299.

ผลจากการศึกษา พบว่า แบบสอบทั้ง 4 แบบให้แรงกระตุ้นในการเรียนแก่เด็กเท่า ๆ กัน และหลังจากสอบเสร็จแล้ว 5 สัปดาห์ ให้นักเรียนทำแบบสอบเดิมซ้ำอีก ปรากฏว่ากลุ่มที่สอบแบบสอบแบบอัตนัย และแบบเติมคำหำกะแผนโคสูงขึ้นไป ส่วนกลุ่มที่สอบแบบสอบแบบเลือกตอบและแบบถูกผิด คะแนนไม่สูงขึ้น แสดงว่า แบบสอบ 2 แบบหลังนี้มีความเที่ยงมากกว่า 2 แบบแรก¹

อีเบล (Ebel) ได้ศึกษาเปรียบเทียบค่าความเที่ยง และความทรงของแบบสอบแบบถูกผิดและแบบเลือกตอบ โดยใช้อัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบ ให้นักเรียนทำได้ในเวลาเท่ากันซึ่งในที่นี้เท่ากัน 2 : 1 เป็นตัวปรับค่าความเที่ยงของแบบสอบแบบถูกผิด เพื่อให้เป็นค่าความเที่ยงของแบบสอบแบบถูกผิดที่ใช้เวลาในการสอบเท่ากับแบบสอบแบบเลือกตอบ ผลการศึกษาพบว่า ค่าความเที่ยงของแบบสอบแบบเลือกตอบและแบบถูกผิด ไม่แตกต่างกัน และแบบสอบแบบเลือกตอบและแบบถูกผิดสามารถวัดได้ในสิ่งเดียวกัน²

เออร์วิน, ฮัลเปอร์น และ แลนด์แมน (Irvin, Halpern and Landman) ได้ประเมินผลสัมฤทธิ์ของเด็กที่เรียนซ้ำด้วยแบบสอบมาตรฐานแบบถูกผิด และแบบเลือกตอบ เครื่องมือที่ใช้ได้แก่ แบบสอบ 3 ฉบับ ดังนี้ คือ

1. แบบสอบฉบับที่ 1 เป็นแบบสอบมาตรฐานแบบถูกผิดเกี่ยวกับสังคมและอาชีพ

¹Mayer, G., "An Experimental Study of the Old and New Types of Examination" Journal of Educational Psychology 25 (1934) quoted in H.H. Remmer and N.L. Gage, Educational Measurement and Evaluation (New York: Harper & Brother Publisher, 1955), p.59.

²Ebel, R.L., "The Comparative Effectiveness of True-False and Multiple Choice Achievement Test Items," quoted in Frisbie, "Multiple Choice Versus True-False: A Comparison of Reliabilities and Concurrent Validities," p.297.

(SPIB) ซึ่งประกอบด้วยแบบสอบ 3 ชุด ได้แก่ การธนาคาร การซื้อขาย และทักษะในการทำงาน

2. แบบสอบฉบับที่ 2 เป็นแบบสอบแบบเลือกตอบชนิด 3 ตัวเลือกที่สร้างขึ้นเพื่อวัดเนื้อหาเดียวกันกับฉบับที่ 1

3. แบบสอบฉบับที่ 3 เป็นแบบสอบแบบเลือกตอบชนิด 2 ตัวเลือก สร้างขึ้นโดยการคัดตัวดวงพิมพ์คำอ่านจากจำแนกคำสุดท้ายของข้อสอบในฉบับที่ 2 ออกข้อละ 1 คำ

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนที่เรียนวิชาเกรด 9 - 12 จำนวน 425 คน ซึ่งแบ่งออกเป็น 3 กลุ่ม แต่ละกลุ่มทำแบบสอบฉบับใดฉบับหนึ่งใน 3 ฉบับ ผลการวิจัยพบว่า ความเที่ยงและความตรงของแบบสอบแบบเลือกตอบทั้งสองฉบับ และแบบถูกผิดมีค่าสูงและใกล้เคียงกัน

กรีน (Green) ได้ศึกษาเปรียบเทียบความเที่ยงและความตรงของแบบสอบแบบเลือกตอบกับแบบถูกผิด โดยใช้แบบสอบสัมฤทธิ์ผลวิชาการทดสอบและการวัดผลที่ครูสร้างขึ้น มีลักษณะเป็นแบบเลือกตอบชนิด 3 ตัวเลือก จำนวน 40 ข้อ ขอบข่ายของแบบเลือกตอบแต่ละข้อมีลักษณะที่ตัวคำถาม (Stem) อยู่ร่วมกับตัวเลือก (Option) การแปลงแบบสอบแบบเลือกตอบเป็นแบบถูกผิด กระทำโดยแปลงขอบข่ายแบบเลือกตอบ 1 ข้อ เป็นขอบข่ายแบบถูกผิด 3 ข้อ ซึ่งจะมีข้อที่ถูกต้อง 1 ข้อ และข้อที่ผิด 2 ข้อ แบบถูกผิดที่แปลงได้จะมีจำนวน 120 ข้อ แลแบ่งขอบข่ายทั้งหมดออกเป็น 3 ส่วนเท่า ๆ กันควยวิธีนำข้อขอบข่ายทุก ๆ 3 ข้อ ที่แปลงมาจากขอบข่ายแบบเลือกตอบแต่ละข้อมาแยกไปอยู่ข้อละส่วนโดยวิธีสุ่ม หลังจากนั้นจัดลำดับของข้อขอบข่ายในแต่ละส่วนโดยวิธีสุ่มเช่นเดียวกัน กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือ นิสิตระดับปริญญาตรี จำนวน 50 คน แบ่งออกเป็น 2 กลุ่ม ๆ ละ 25 คน ให้แต่ละกลุ่มทำแบบสอบฉบับใดฉบับหนึ่งจาก 2 ฉบับ โดยให้ทำที่ละส่วนเสร็จแล้วจึงให้ทำส่วนต่อไป และเมื่อหมดเวลา 12 นาทีแรกของการสอบ ให้นิสิตบันทึกจำนวนข้อที่ทำได้ไว้วางย ผลจากการศึกษา พบว่า

1. อัตราส่วนของจำนวนข้อขอบข่ายแบบถูกผิด ต่อแบบเลือกตอบที่นิสิตทำได้ใน

¹ Larry K. Irvin, Andrew S. Halpern and Janet T. Landman, "Assessment of Retarded Student Achievement with Standardized True-False and Multiple-Choice Tests," Journal of Educational Measurement 17 (Spring 1980), pp.51-58.

เวลาเท่ากัน มีค่าเท่ากับ 2.4 : 1

2. ความเที่ยงของแบบสอบแบบถูกผิดกับแบบสอบแบบเลือกตอบ หลังจากปรับแล้ว โดยเพิ่มความยาวของแบบเลือกตอบเป็น 1.25 เท่าของจำนวนข้อเดิม เพื่อให้ใช้เวลาในการสอบเท่ากันแบบสอบแบบถูกผิด มีค่าไม่แตกต่างกัน

3. ความตรงของแบบสอบแบบเลือกตอบ และแบบถูกผิด ไม่แตกต่างกัน¹

ผลการวิจัยที่พบหาความเที่ยงของแบบสอบทั้ง สอง แยกต่างหาก

สุธรรม จันทรหอม ศึกษาเปรียบเทียบการใช้แบบสอบแบบถูกผิด แบบเลือกตอบ และแบบเติมคำ วัตถุประสงค์ทางการเรียนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 7 กลุ่มตัวอย่างจำนวน 1,350 คน เครื่องมือที่ใช้ศึกษาคือ แบบสอบวิชาคณิตศาสตร์ ซึ่งแยกเป็น 3 ฉบับ คือ คณิตศาสตร์ทักษะ คณิตศาสตร์ปัญหาและคณิตศาสตร์ความเข้าใจ การเก็บรวบรวมข้อมูล ให้นักเรียนสอบแต่ละ 3 ฉบับ โดยนักเรียนแต่ละคนไม่ได้รับรูปแบบของข้อสอบซ้ำกัน การศึกษาครั้งนี้มีจำนวนของแบบสอบทั้ง 3 แบบเท่ากัน และวัดเนื้อหาเดียวกัน ผลการศึกษา พบว่า

1. ค่าความเที่ยงของแบบสอบแบบเลือกตอบ แบบถูกผิด และแบบเติมคำ แยกต่างหากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบเติมคำมีความเที่ยงสูงสุด และแบบถูกผิดมีค่าต่ำสุด

2. ค่าความตรงของแบบสอบแบบเลือกตอบ แบบถูกผิด และแบบเติมคำของวิชาคณิตศาสตร์ทักษะ และคณิตศาสตร์ปัญหาไม่แตกต่างกัน ส่วนวิชาคณิตศาสตร์ความเข้าใจแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

3. ความยากของแบบสอบแบบเลือกตอบ แบบถูกผิด และแบบเติมคำ แยกต่างหากอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบสอบแบบถูกผิดเป็นแบบที่ง่ายที่สุด แบบเลือกตอบมีความยากปานกลาง และแบบเติมคำยากที่สุด²

¹Green, "Multiple Choice and True-False: Reliability and Validity Compared," pp.42-44.

²สุธรรม จันทรหอม, "การศึกษาเปรียบเทียบการใช้แบบสอบแบบถูกผิด แบบเลือกตอบ และแบบเติมคำ วัตถุประสงค์ทางการเรียนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 7," หน้า 92-95.

องุ่น มณีคำ ได้ศึกษาการเปรียบเทียบความเที่ยงและความตรงของแบบสอบสัมฤทธิ์ผล แบบเลือกตอบและแบบถูกผิด กลุ่มตัวอย่างเป็นจำนวน 570 คน เครื่องมือที่ใช้ คือ แบบสอบสัมฤทธิ์ผล วิชาวิทยาศาสตร์ที่ผู้วิจัยสร้างขึ้น 3 แบบ คือ แบบเลือกตอบ แบบถูกผิดที่แปลงมาจากแบบเลือกตอบโดยวิธีพิจารณาภาคอำนาจจำแนกของตัวลง และแบบถูกผิดที่แปลงมาจากแบบเลือกตอบโดยวิธีสุ่มแบบสอบ แบบถูกผิดทั้งสองฉบับที่แปลงได้ มีข้อผิดพลาดประมาณ 50% ในการเก็บรวบรวมข้อมูล ไขการสุ่มอย่างง่าย คือ ให้นักเรียนแต่ละคนที่แบ่งกลุ่มตัวอย่างก่อนแบบสอบเพียงฉบับเดียวเท่านั้น ผลการวิจัย ปรากฏว่า

1. อัตราส่วนของจำนวนข้อที่ตรงแบบถูกผิดที่แปลงมาจากแบบเลือกตอบ โดยวิธีพิจารณาภาคอำนาจจำแนกของตัวลงต่อแบบเลือกตอบ และอัตราส่วนแบบถูกผิดที่แปลงมาจากแบบเลือกตอบโดยวิธีสุ่มต่อแบบเลือกตอบ มีค่าเป็น 1.41: 1 และ 1.5: 1 ตามลำดับ
2. ค่าความเที่ยงของแบบสอบแบบเลือกตอบกับแบบถูกผิดซึ่งจากปรับให้ใช้เวลาในการสอบเท่ากัน แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01
3. ไม่มีความแตกต่างอย่างมีนัยสำคัญทาง สถิติระหว่าง ค่าความตรงของแบบสอบแบบเลือกตอบ กับแบบสอบแบบถูกผิดที่แปลงมาจากแบบเลือกตอบ โดยวิธีพิจารณาภาคอำนาจจำแนกของตัวลง ($p > .05$) แต่มีความแตกต่างอย่างมีนัยสำคัญทาง สถิติที่ระดับ .05 ระหว่างค่าความตรงของแบบสอบแบบเลือกตอบกับแบบสอบแบบถูกผิดที่แปลงมาจากแบบเลือกตอบ โดยวิธีสุ่ม

ฟรีสบี (Frisbie, 1973) ได้ศึกษาเปรียบเทียบค่าความเที่ยงและความตรงของแบบสอบแบบเลือกตอบและแบบถูกผิด กลุ่มตัวอย่างที่ใช้เป็นนักเรียนจำนวน 1,018 คน เครื่องมือที่ใช้เป็นแบบสอบสัมฤทธิ์ผลวิชาสังคมศึกษา (Social Study) และวิชาธรรมชาติวิทยา (Natural Science) ทั้ง 2 วิชาสร้างเป็นแบบเลือกตอบก่อน แบบสอบวิชาสังคมศึกษาเรียกว่า ฟอรัม SM วิชาธรรมชาติวิทยาเรียกว่า ฟอรัม NM หลังจากนั้นแปลงข้อที่ตรงแบบเลือกตอบให้เป็นแบบถูกผิด ซึ่งมีวิธีแปลง 2 วิธี

¹ องุ่น มณีคำ, "การเปรียบเทียบความเที่ยงและความตรงของแบบสอบสัมฤทธิ์ผลแบบเลือกตอบและแบบถูกผิด," หน้า 60-63.

1. ใ้ครูผู้สอนวิชาละ 5 คน เป็นผู้คัดเลือกคุณภาพของตัวดวงของข้อกระทงแต่ละข้อ จากครูอย่างน้อย 4 คน คัดเห็นว่าตัวดวงใดที่คิดที่สุดตรงกันแล้วขณนั้นจะแปลง เป็นขอมติตามตัวดวงนั้น ถ้าไม่เป็นไปตามเกณฑ์ดังกล่าวจะแปลง เป็นขอมติ เมื่อแปลงแล้ว แบบสอบวิชาสังคมศึกษา เรียกว่า ฟอรัม SJ และวิชาวิทยาศาสตร์ธรรมชาติ เรียกว่า ฟอรัม NJ แบบสอบแบบถูกผิดทั้งสองฉบับที่แปลง ได้มีขอมติประมาณ 60%

2. วิธีพิจารณาคำอำนาจจำแนกของตัวดวงของข้อกระทงแต่ละข้อ ตัวดวงที่ดีที่สุดของข้อใดมีค่าอำนาจจำแนกตั้งแต่ -0.20 ขึ้นไป ขอนั้นจะแปลง เป็นขอมติตามตัวดวงนั้น ถ้าไม่เป็นไปตามเกณฑ์จะแปลง เป็นขอมติตามคำตอบที่ถูกของที่สุด เมื่อแปลงแล้วแบบสอบวิชาสังคมศึกษา เรียกว่า ฟอรัม SD และวิชาวิทยาศาสตร์ธรรมชาติ เรียกว่า ฟอรัม ND แบบสอบแบบถูกผิดทั้งสองฉบับที่แปลง ได้มีขอมติประมาณ 50%

หลังจากนั้นนำแบบสอบทั้ง 6 ฟอรัม คือ SM, NM, NJ, SJ, SD และ ND มาจัดใหม่เป็น 8 ฟอรัม ซึ่งแตกต่างกันตามเนื้อหาวิชาและวิธีแปลง การเก็บรวบรวมขอมติ ให้นักเรียนทำแบบสอบคนละ 1 ฟอรัม โดยแบ่งเวลาเป็น 2 ช่วง ๆ แรกทำ 8 นาที แล้วให้นักเรียนบันทึกจำนวนข้อที่ทำได้ หลังจากนั้นให้ทำข้ออีก 50 นาที ผลการศึกษา พบว่า

1. อัตราส่วนจำนวนข้อกระทงแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้ในเวลาเท่ากันของวิชาวิทยาศาสตร์ธรรมชาติ เป็น 1.44: 1 และวิชาสังคมศึกษาเป็น 1.53: 1
2. ค่าความเที่ยงของแบบสอบแบบเลือกตอบกับแบบถูกผิดหลังจากปรับให้เวลาในการสอบเท่ากัน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ
3. แบบสอบแบบเลือกตอบและแบบถูกผิดมีแนวโน้มที่จะวัดในสิ่งเดียวกัน¹

¹ Frisbie, "Multiple Choice Versus True-False: A Comparison of Reliabilities and Concurrent Validities," pp.297-304.

ต่อมาในปี ค.ศ. 1974 ฟรีสบี ได้ศึกษาอิทธิพลของขอกระพงแบบเลือกตอบและแบบ
 ถูกผิดที่มีค่าความเที่ยงของแบบสอบอีกครั้งหนึ่ง โดยมีวัตถุประสงค์เพื่อเปรียบเทียบผลการวิจัยครั้งแรก
 กลุ่มตัวอย่างเป็นนักเรียนจำนวน 529 คน แบบสอบที่ใช้ศึกษา คือ แบบสอบสัมฤทธิ์ผลที่ใช้ในการศึกษา
 ครั้งก่อน แต่แปลงให้เป็นแบบสอบแบบถูกผิดโดยวิธีพิจารณาจากอำนาจจำแนกของตัวดวงวิธีเดียว
 เท่านั้น หลังจากนั้นจัดแบบสอบใหม่เป็น 4 ฟอร์ม โดยให้แต่ละฟอร์มประกอบด้วยขอกระพงแบบถูกผิด
 และแบบเลือกตอบ และให้แบบถูกผิดมีความยาวเป็น 1.5 เท่าของแบบเลือกตอบ วิธีการเก็บรวบรวม
 ข้อมูลเหมือนครั้งแรกทุกประการ ส่วนเวลาในการทำแบบสอบใช้ตามผลการศึกษาก่อน ปกติ
 ผลการศึกษาก่อนนี้สอดคล้องกับผลการศึกษาก่อนทุกประการ¹

ในปีเดียวกัน อูสเตอร์ฮอฟและกลาสแนพ (Oosterhof and Glasnapp) ได้
 ศึกษาเปรียบเทียบความเที่ยงและความยากของแบบสอบแบบเลือกตอบ และแบบถูกผิด โดยใช้แบบ
 สอบสัมฤทธิ์ผลที่ครูสร้างขึ้น 2 ฉบับ ฉบับที่ 1 เป็นแบบเลือกตอบ 4 ตัวเลือก จำนวน 40 ขอ
 ฉบับที่ 2 เป็นแบบถูกผิดที่แปลงมาจากฉบับที่ 1 ทวิวิธีรวมขอคำตามกับตัวเลือกที่เป็นตัวถูกเป็นขอ
 กระพง 1 ขอ และรวมขอคำตามกับตัวดวงที่ปรากฏอำนาจจำแนกสูงสุดเป็นขอกระพง 1 ขอ นำขอกระพง
 ทั้ง 80 ขอ มาแยกเป็น 2 ส่วน โดยแยกขอกระพงแบบถูกผิดที่ละคู่ไปอยู่ขอละส่วน โดยวิธีสุ่ม แล้ว
 จัดลำดับของขอกระพงในแต่ละส่วนทวิวิธีสุ่มเช่นเดียวกัน กลุ่มตัวอย่างของการวิจัย คือ นิสิตปริญญา
 ตรีจำนวน 101 คน แบ่งออกเป็น 2 กลุ่ม ๆ แรกมี 51 คน ให้ทำแบบสอบฉบับที่ 1 ก่อนแล้วจึง
 ทำฉบับที่ 2 ส่วนกลุ่มที่ 2 มี 50 คน ให้ทำแบบสอบฉบับที่ 2 ก่อนแล้วจึงทำฉบับที่ 1 ในการทำ
 แบบสอบใหม่ทั้งใช้เวลาที่ใช้เมื่อทำเสร็จ 40 ขอ 80 ขอ และ 120 ขอ ผลการศึกษาได้สรุป
 ดังนี้

¹Frisbie, "The Effect of Item Format on Reliability and Validity: A Study of Multiple Choice and True-False Achievement Test," Educational and Psychological Measurement 34 (1974): 385-392.

1. อัตราส่วนของจำนวนข้อกระทงแบบถูกผิดกับแบบเลือกตอบที่กลุ่มตัวอย่างทำได้ในเวลาเท่ากัน เป็น 1.73 : 1
2. ค่าความเที่ยงของแบบสอบแบบถูกผิดหลังจากปรับแก้ค่าความเที่ยงของแบบสอบแบบเลือกตอบ และถ้าจะให้แบบสอบแบบถูกผิดมีความเที่ยงเท่ากับแบบสอบแบบเลือกตอบ แล้วแบบสอบแบบถูกผิดจะของยาวเป็น 3.75 เท่าของแบบสอบแบบเลือกตอบ
3. ค่าความยากของแบบสอบแบบเลือกตอบหลังจากปรับแก้ค่าแล้ว ปรากฏว่า แบบสอบแบบเลือกตอบง่ายที่สุด แบบถูกผิดทั้งหมดจะง่ายกว่าแบบถูกผิดที่มีข้อผิดทั้งหมด¹

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

¹Oosterhof and Glasnapp, "Comparative Reliabilities and Difficulties of the Multiple Choice and True - False Formats," pp.62-64.