



วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยได้แบ่งวรรณคดีที่เกี่ยวข้องออกเป็น 2 ตอน ตอนที่ 1 เป็นมโนทัศน์เบื้องต้นเกี่ยวกับความเที่ยงและความตรงของแบบสอบ และตอนที่ 2 เป็นงานวิจัยที่เกี่ยวข้อง

มโนทัศน์เบื้องต้นเกี่ยวกับความเที่ยงและความตรงของแบบสอบ

ความเที่ยงของแบบสอบ

มีผู้ให้ความหมายของ ความเที่ยง เป็นจำนวนมาก ซึ่งส่วนใหญ่มีความคล้ายคลึงกัน ดังนี้

เมเรนส์และเลแมนน์ (Mehrens and Lehmann) ให้ความหมายว่า ความเที่ยง หมายถึง ระดับ (Degree) ของความคงที่ระหว่างการวัดสิ่งเดียวกัน 2 ครั้ง¹

กรอนลันด์ (Gronlund) ให้ความหมายสั้น ๆ ว่า ความเที่ยงหมายถึง ความคงเสถียรของการวัด²

¹Mehrens and Lehmann, Measurement and Evaluation Education and Psychology, p. 88.

²Norman E. Gronlund, Measurement and Evaluation in Teaching, 3d ed. (New York: Macmillan Publishing Co., 1976), p. 105.

อีเบล (Ebel) ให้ความหมายคล้ายกับกรอนสันคัวว่า ความเที่ยง หมายถึง ความคงเส้นคงวาของชุดของคะแนนจากแบบสอบที่วัดในสิ่งที่ต้องการวัด¹

แอน อนาสตาซี ให้ความหมายว่า ความเที่ยง หมายถึงความคงที่ของคะแนนที่ได้จากการทดสอบคนกลุ่มเดียวกันสองครั้งด้วยแบบสอบเดิมในเวลาที่แตกต่างกัน หรือทดสอบคนกลุ่มเดียวกันด้วยข้อสอบต่างชุดที่มีข้อเทียบเท่ากัน หรือภายใต้สภาพการสอบแตกต่างกัน²

ทฤษฎีของความเที่ยง

การอธิบายทฤษฎีของความเที่ยงเริ่มต้นจากคะแนนที่วัดได้ (Observe scores) ซึ่งประกอบด้วยคะแนนจริง (True score) กับคะแนนความคลาดเคลื่อน (Error scores) เขียนเป็นสมการ³ ได้ดังนี้

$$X = T + E$$

เมื่อ $X =$ คะแนนที่วัดได้

$T =$ คะแนนจริง

$E =$ คะแนนความคลาดเคลื่อน

¹Robert L. Ebel, Measuring Educational Achievement (Englewood Cliffs, N.J.: Prentice-Hall, 1965), p. 310.

²แอน อนาสตาซี, การตรวจสอบเชิงจิตวิทยา แปลโดย ประชุมสุข อาชาวำรุง และคณะ (กรุงเทพมหานคร: สำนักพิมพ์ไทยวัฒนาพานิช, 2519), หน้า 73.

³Mehrens and Lehmann, Measurement and Evaluation Education and Psychology, p. 90.

คะแนนจริง (True score) หมายถึง คะแนนที่บุคคลพึงได้รับจากการวัด
ด้วยเครื่องมือ (แบบสอบ) ที่มีคุณลักษณะสมบูรณ์ปราศจากความคลาดเคลื่อน หรือหมายถึง
คะแนนเฉลี่ยของบุคคลที่ได้จากการทำแบบสอบเดิม หรือแบบสอบคู่ขนานหลาย ๆ ครั้ง¹

คะแนนความคลาดเคลื่อน (Error score) หมายถึง ค่าความผิดพลาดที่
เกิดจากการวัด ซึ่งเป็นความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญหรือไม่เป็นระบบเท่านั้น กล่าว
ได้ว่าเป็นตัวแปรที่ไม่เกี่ยวข้องกับวัตถุประสงค์ของการสอบแล้วทำให้เกิดความไม่คงที่ในการ
วัด คะแนนความคลาดเคลื่อนอาจเกิดขึ้นเนื่องจากตัวแบบสอบ การดำเนินการสอบ การ
ให้คะแนน หรือความไม่พร้อมของผู้สอบ โดยทฤษฎีความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญนี้
อาจเป็นได้ทั้งทางบวกและทางลบซึ่งจะหักล้างกันหมดไป นั่นคือค่าเฉลี่ยของความคลาดเคลื่อน
จะเท่ากับ 0 และความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญนี้จะไม่มีความสัมพันธ์กับคะแนนจริง
เลย นั่นคือสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงกับคะแนนความคลาดเคลื่อนมีค่าเท่ากับ
0 ดังนั้นความแปรปรวนของคะแนนที่วัดได้ สามารถเขียนได้ดังสมการ²

$$\begin{aligned} s_x^2 &= s_t^2 + s_e^2 \\ \text{เมื่อ } s_x^2 &= \text{ความแปรปรวนของคะแนนที่วัดได้} \\ s_t^2 &= \text{ความแปรปรวนของคะแนนจริง} \\ s_e^2 &= \text{ความแปรปรวนของคะแนนความคลาดเคลื่อน} \end{aligned}$$

โดยทฤษฎีความเที่ยงหมายถึง อัตราส่วนระหว่างความแปรปรวนของคะแนนจริง
ต่อความแปรปรวนของคะแนนที่วัดได้³ ดังนั้น

¹ Frederick G. Brown, Principle of Educational and Psychological Testing, 2d ed (New York: Holt, Rinehart and Winston, 1976), p. 53.

² Mehrens and Lehmann, Measurement and Evaluation Education and Psychology, p. 91.

³ Ibid.

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

เมื่อ r_{xx} = ค่าความเที่ยง

จากสมการค่าความเที่ยงจึงบอกให้เราทราบถึงความแปรปรวนของคะแนนจริงว่าเป็นสัดส่วนเท่าใดเมื่อเทียบกับความแปรปรวนของคะแนนที่วัดได้ และเนื่องจาก $s_t^2 = s_x^2 - s_e^2$ ดังนั้นสมการ (1) จึงเขียนใหม่ได้เป็น

$$r_{xx} = 1 - \frac{s_e^2}{s_x^2}$$

003996

ซึ่งเป็นสมการที่นิยมใช้ประมาณค่าความเที่ยงของแบบสอบโดยทั่วไป

วิธีประมาณค่าความเที่ยง (Estimates of Reliability)

วิธีประมาณค่าความเที่ยงมีอยู่ด้วยกันหลายวิธี แต่ละวิธีทำให้เกิดความคลาดเคลื่อนแตกต่างกัน การที่จะเลือกใช้วิธีใดนั้นขึ้นอยู่กับความมุ่งหมาย และลักษณะของคะแนนที่ได้จากการสอบนั้น ๆ วิธีประมาณค่าความเที่ยงโดยทั่วไปมี 4 แบบ¹ คือ

1. แบบสอบซ้ำ (Measure of stability)
2. แบบไขข้อสอบคล้ายกัน (Measure of equivalence)
3. แบบไขข้อสอบคล้ายกันและสอบซ้ำ (Measure of equivalence and stability)
4. แบบวัดความคงที่ภายใน (Measure of internal consistency)

ซึ่งแบ่งย่อยออกได้เป็น

- 4.1 แบบแบ่งครึ่งจำนวนข้อสอบ (Split-half)
- 4.2 แบบคูเคอร์ริชาร์ดสัน (Kuder-Richardson estimates)

¹William A. Mehrens and Irvin J. Lehmann, Measurement and Evaluation in Education and Psychology (New York: Holt Rinehart and Winston, 1973), p. 109.

4.3 แบบวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's analysis of variance procedure)

ในที่นี้จะกล่าวเฉพาะความเที่ยงแบบคูเคอร์ริชาร์ดสัน ซึ่งเป็นวิธีประมาณค่าความเที่ยงที่ใช้ในการวิจัยนี้เท่านั้น

ความเที่ยงแบบคูเคอร์ริชาร์ดสัน

ในปี ค.ศ. 1937 คูเคอร์และริชาร์ดสัน (Kuder and Richardson) เสนอสูตรสำหรับการประมาณค่าความเที่ยงของแบบสอบหลายสูตร แต่สูตรที่เป็นที่นิยมและใช้กันอย่างกว้างขวางคือสูตร คูเคอร์ริชาร์ดสัน 20^1 (Kuder-Richardson 20)

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

เมื่อ r แทน	ค่าความเที่ยงของแบบสอบ
k แทน	จำนวนข้อในแบบสอบ
p แทน	สัดส่วนของผู้ตอบถูกในแต่ละข้อ
q แทน	สัดส่วนของผู้ตอบผิดในแต่ละข้อ
σ^2 แทน	ความแปรปรวนของคะแนนทั้งหมด

การหาค่าความเที่ยงแบบคูเคอร์ริชาร์ดสัน เป็นการหาความสัมพันธ์กันระหว่างข้อสอบ ซึ่งดำเนินการสอบครั้งเดียว และใช้แบบสอบชุดเดียว ความสอดคล้องกันระหว่างข้อนี้ ได้รับอิทธิพลจากแหล่งความแปรปรวนคลาดเคลื่อน 2 แหล่ง คือ

1. เนื้อหาที่สุ่ม
2. ความเป็นวิวิธพันธ์ของพฤติกรรมที่สุ่ม ยิ่งข้อสอบมีความเป็นเอกพันธ์มาก ความสอดคล้องกันระหว่างข้อก็ยิ่งสูง²

¹ Ebel, Measuring Educational Achievement, p. 318.

² แอน อนุสตาซี, การตรวจสอบเชิงจิตวิทยา, หน้า 84.

ค่าความเที่ยงแบบคูเคอร์ริชาร์ดสัน มีค่าประมาณใกล้เคียงกับค่าเฉลี่ยของสัมประสิทธิ์ความเที่ยงที่ได้จากการแบ่งครึ่งแบบสอควัยวิธีต่าง ๆ กัน¹

ข้อตกลงเบื้องต้นของสูตรคูเคอร์ริชาร์ดสัน 20 คือ

1. การตรวจให้คะแนนแต่ละข้อ ให้ 1 คะแนนเมื่อตอบถูกและให้ 0 คะแนนเมื่อตอบผิด²
2. ข้อกระทงในแบบสอบจะต้องมีลักษณะเป็นเอกพันธ์ คือวัดคุณลักษณะเดียวกัน³

สูตรคูเคอร์ริชาร์ดสัน 20 นี้เหมาะที่จะหาค่าความเที่ยงของแบบสอบความสามารถ (Power Test) เท่านั้น ไม่เหมาะที่จะหาค่าความเที่ยงของแบบสอบความเร็ว (Speed Test) เพราะค่า p และ q ของแต่ละข้อ จะต้องเป็นค่าที่ได้จากการที่ผู้สอบทุกคนมีโอกาสทำข้อนั้นแล้ว ซึ่งเป็นไปไม่ได้สำหรับแบบสอบความเร็ว เพราะผู้สอบทุกคนไม่มีโอกาสทำถึงข้อสุดท้าย⁴

เนื่องจากสูตรคูเคอร์ริชาร์ดสัน 20 ไม่สะดวกในการคำนวณด้วยมือ คูเคอร์และริชาร์ดสัน (Kuder and Richardson) จึงเสนอสูตรที่สามารถคำนวณได้ง่ายกว่า โดยมีข้อสมมติว่า ข้อกระทงแต่ละข้อมีความยากเท่ากัน หรือกำหนดให้ค่า p คงที่ สูตรใหม่นี้เรียกว่า คูเคอร์ริชาร์ดสัน 21 ($K - R 21$)⁵

¹ เรื่องเดียวกัน, หน้า 86.

² Gronlund, Measurement and Evaluation in Teaching, p. 112.

³ Ebel, Measuring Educational Achievement, p. 318.

⁴ Brown, Principle of Educational and Psychological Testing, p. 88.

⁵ Ebel, Measuring Educational Achievement, p. 319.



$$= \frac{k}{k-1} \left[1 - \frac{m(1 - \frac{m}{k})}{\sigma^2} \right]$$

เมื่อ	r	=	ค่าความเที่ยง
	k	=	จำนวนข้อในแบบสอบ
	m	=	ค่าเฉลี่ยของคะแนนที่นักเรียนทำได้
	σ^2	=	ความแปรปรวนของคะแนนของนักเรียนทั้งหมด

การประมาณค่าความเที่ยงของแบบสอบโดยใช้สูตรคูเคอร์ริชาร์ดสัน 21 ค่าที่ได้มักต่ำกว่าที่ประมาณด้วยสูตรคูเคอร์ริชาร์ดสัน 20 เนื่องจากว่า โดยทั่วไปข้อกระทงแต่ละข้อในแบบสอบจะมีระดับความยากแตกต่างกัน ซึ่งไม่เป็นไปตามข้อสมมติดังกล่าว

องค์ประกอบที่มีอิทธิพลต่อความเที่ยง

นอกจากวิธีการที่ใช้ประมาณค่าความเที่ยงจะมีอิทธิพลต่อค่าความเที่ยงของแบบสอบแล้ว องค์ประกอบอื่น ๆ¹ ต่อไปนี้ มีอิทธิพลต่อความเที่ยงของแบบสอบด้วย

1. ความยาวของแบบสอบ (Test length) การเพิ่มแบบสอบที่มีคุณภาพเทียบเท่ากับแบบสอบเดิมที่มีอยู่ จะทำให้แบบสอบมีค่าความเที่ยงสูงขึ้น นั่นคือแบบสอบที่ยาวจะมีความเที่ยงสูงกว่าแบบสอบที่สั้น สูตรแสดงให้เห็นความสัมพันธ์ระหว่างความยาวของแบบสอบกับความเที่ยงคือสูตรของสเปียร์แมนบราวน์ (Spearman-Brown)

2. ความเร็ว (Speed) การหาค่าความเที่ยงของแบบสอบความเร็ว (Speed test) ด้วยวิธีวัดความคงที่ภายในเป็นวิธีที่ไม่เหมาะสม เพราะแบบสอบความเร็วเป็นแบบสอบที่ง่าย แต่ให้เวลาสอบน้อยจึงไม่มีผู้ใดทำได้ครบทุกข้อ นักเรียนสามารถที่จะตอบถูกทุกข้อที่ทำถึง ดังนั้นถ้าแบ่งเป็นข้อคู่ ข้อคี่ นักเรียนจะสามารถทำทั้ง 2 ส่วนถูกเท่า ๆ กัน ดังนั้นค่าสหสัมพันธ์ระหว่างทั้ง 2 ส่วน จะมีค่าสูงมาก และการใช้สูตรคูเคอร์ริชาร์ดสัน 20 ก็ไม่เหมาะสมดังเหตุผลที่กล่าวข้างต้น ดังนั้นการหาค่าความเที่ยงของแบบ

¹ Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, pp. 100-103.

สอบความเร็ว ควรหาโดยการสอบ 2 ครั้ง หรือใช้แบบสอบคล้ายกันแล้วคำนวณค่าความเที่ยงโดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Product moment correlation)

3. ความยากของข้อกระทง (Item difficulty) ความยากของข้อกระทงมีอิทธิพลต่อค่าความเที่ยงของแบบสอบ เพราะค่าความเที่ยงขึ้นอยู่กับความแปรปรวนของคะแนนผลการสอบ ถ้าความแปรปรวนระหว่างคะแนนมีน้อย ย่อมได้ค่าความเที่ยงต่ำกว่าคะแนนที่มีความแปรปรวนมาก เช่น แบบสอบง่ายมากนักเรียนทุกคนทำถูกต้องเกือบหมด หรือแบบสอบยากมากนักเรียนทุกคนทำผิดเกือบหมด กรณีเช่นนี้ความแปรปรวนของคะแนนการสอบจะมีน้อย ทำให้ค่าความเที่ยงต่ำ

4. ความเป็นเอกพันธ์ของกลุ่ม (Group homogeneous) กลุ่มที่มีความเป็นเอกพันธ์ จะมีพิสัยของคะแนนความสามารถของกลุ่มน้อยกว่ากลุ่มที่มีลักษณะเป็นวิวิธพันธ์ (Group heterogeneous) ดังนั้นจะทำให้ค่าความเที่ยงน้อยกว่าความเที่ยงจากกลุ่มที่มีลักษณะวิวิธพันธ์

5. ความเป็นปรนัย (Objective) แบบสอบที่มีความเป็นปรนัยสูง ค่าความเที่ยงของแบบสอบก็สูงด้วย

นอกจากนั้นการบริหารแบบสอบ เช่น การจับเวลา การให้คำชี้แจง การป้องกัน การทุจริต เป็นต้น หรือองค์ประกอบในตัวผู้สอบเอง เช่น แรงจูงใจ ความวิตกกังวล ประสิทธิภาพในการสอบ การเคา เป็นต้น สิ่งเหล่านี้อาจเป็นแหล่งความคลาดเคลื่อนที่ทำให้ค่าความเที่ยงของแบบสอบลดลงได้

ความตรงของแบบสอบ

มีผู้ให้ความหมายของ ความตรง (Validity) ไว้ต่าง ๆ กันดังนี้

กัลลิกเสน (Gulliksen) "ความตรงของแบบสอบคือค่าสหสัมพันธ์ระหว่างแบบสอบกับเกณฑ์"¹

เคอตัน (Cureton) "ความตรงของแบบสอบ คือ ค่าประมาณของสหสัมพันธ์ระหว่างแบบสอบกับเกณฑ์"²

ลินควิส (Lindquist) "ความตรงของแบบสอบ คือ ความสามารถวัดในสิ่งที่ต้องการวัดได้ถูกต้องแม่นยำ หรือคือความสามารถในการให้ความหมายในสิ่งที่วัดได้อย่างไม่ผิดพลาด"³

สแตนเลย์และฮอปกินส์ (Stanley and Hopkins) "ความตรงของแบบสอบ คือ ความถูกต้องแม่นยำของคะแนนในการพยากรณ์ที่จำเพาะเจาะจง"⁴

กรอนลันด์ (Gronlund) "ความตรง คือ ขอบเขตของผลการประเมินที่สอดคล้องกับจุดประสงค์"⁵

จากความหมายต่าง ๆ ดังกล่าวข้างต้น สามารถสรุปความหมายของความตรง

¹ Gulliksen, H. Theory of Mental Tests, (New York: John Wiley & Son, 1950), p. 88.

² Cureton, E.E., "Validity" quoted in E.E. Linquist, Educational Measurement (Washington D.C.: American Council on Education, 1951), p. 625.

³ Linquist E.E., A First Course in Statistics (Boston: Houghton Mifflin, 1942), p. 213.

⁴ Julian C. Stanley & Kenneth D. Hopkins, Educational & Psychological Measurement & Evaluation (New Delhi: Prentice-Hall of India, 1978), p. 101.

⁵ Gronlund, Measurement and Evaluation in Teaching, p.79.

ไต่ 2 ประการ ดังนี้

1. ความตรงของแบบสอบ หมายถึง ค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบกับคะแนนเกณฑ์
2. ความตรงของแบบสอบ หมายถึง ความสามารถของแบบสอบที่วัดได้ในสิ่งที่ต้องการวัด หรือวัดได้สอดคล้องกับวัตถุประสงค์ และคะแนนที่ได้จากแบบสอบที่มีความตรงสูงสามารถบอกถึงสภาพที่แท้จริงและพยากรณ์ได้ถูกต้องแม่นยำ

ชนิดของความตรง

จากการให้ความหมายของความตรงที่แตกต่างกันทำให้จำแนกความตรงออกได้เป็นหลายชนิดและต่าง ๆ กันไป ต่อมาสมาคมจิตวิทยาอเมริกัน (American Psychological Association) ได้จำแนกความตรงออกเป็น 3 ชนิดใหญ่ ๆ¹ ดังนี้

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

ในที่นี้จะกล่าวถึงความตรงตามเนื้อหาและความตรงตามเกณฑ์สัมพันธ์ซึ่งเกี่ยวข้องกับการวิจัยนี้เท่านั้น

ความตรงตามเนื้อหา (Content Validity)

ความตรงตามเนื้อหาจะเกี่ยวข้องกับความตรงตามเนื้อหาของข้อคำถามในแบบสอบ ซึ่งข้อคำถามนั้นต้องเป็นตัวแทนของเนื้อหาสาระ และครอบคลุมขอบเขตของพฤติกรรมที่ต้องการวัด ความตรงตามเนื้อหามีความสำคัญอย่างยิ่งในการสอบวัดผลสัมฤทธิ์ทางการเรียน การที่จะตัดสินว่าแบบสอบนั้นมีความตรงตามเนื้อหามากน้อยเพียงใด จะต้องพิจารณาทั้งหัวข้อ

¹ Mehrens and Lehmann, Measurement and Evaluation Education and Psychology, pp. 109-110.

เนื้อหาวิชาและชนิดของพฤติกรรมที่ต้องการวัดไปพร้อม ๆ กัน

อย่างไรก็ตามเราไม่สามารถแสดงค่าความตรงตามเนื้อหาออกมาเป็นตัวเลขได้ โดยทั่วไปมักจะพิจารณาข้อกระทงในแบบสอบ โดยเทียบกับตารางวิเคราะห์หลักสูตร (Table of specifications) แล้วประมาณออกมามีความตรงตามเนื้อหาเล็กน้อย เพียงใด ผู้ที่จะตัดสินความตรงตามเนื้อหาได้คือ ผู้เชี่ยวชาญทางสาขาวิชานั้น ๆ¹

ความตรงตามเกณฑ์สัมพันธ์ (Criterion - Related Validity)

ความตรงตามเกณฑ์สัมพันธ์ เป็นเรื่องที่เกี่ยวข้องกับการศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบกับคะแนนการวัดจากเกณฑ์ภายนอกที่เป็นอิสระ²

เกณฑ์ (Criterion) คือมาตรฐานในการตัดสิน ในการพัฒนาแบบสอบเกณฑ์ คือคุณลักษณะ หรือกลุ่มของคุณลักษณะที่ใช้เป็นพื้นฐานสำหรับตัดสินความตรงของแบบสอบหรือกระบวนการวัดอื่น ๆ³

ความตรงตามเกณฑ์สัมพันธ์จำแนกได้เป็น 2 ชนิด⁴ คือ

1. ความตรงร่วมสมัย (Concurrent Validity) หมายถึงความตรงของแบบสอบที่จะบ่งบอกสิ่งที่วัดได้ถูกต้องตามสภาพที่แท้จริงในปัจจุบัน โดยอาศัยความสัมพันธ์ระหว่างคะแนนของแบบสอบ กับคะแนนเกณฑ์ ซึ่งกำหนดขึ้นในขณะนั้น เช่น เกณฑ์สัมพันธ์ของแบบวัดทักษะในการคำนวณ ก็อาจใช้วิธีการสังเกตการทำแบบฝึกหัดเลขในชั่วโมงสอนเป็นเวลาสองเดือน ถ้าคะแนนที่นักเรียนได้จากแบบสอบให้ผลสอดคล้องกับเกณฑ์สัมพันธ์

¹ Ibid., pp. 110-112.

² Ibid., p. 112.

³ Ebel, Measuring Educational Achievement, p. 448.

⁴ บุญเชิด ภิญโญอนันตพงษ์ การวัดและประเมินผลการศึกษา, หน้า 337.

แสดงว่าแบบสอบถามสามารถวัดสภาพปัจจุบันของนักเรียนได้

2. ความตรงเชิงทำนาย (Predictive Validity) หมายถึงความสามารถของแบบสอบถามที่จะบ่งบอกผลที่วัดในขณะนั้นได้ถูกต้องตามสภาพที่แท้จริงในอนาคต โดยอาศัยความสัมพันธ์ระหว่างคะแนนของแบบสอบถามกับคะแนนเกณฑ์สัมพัทธ์ ซึ่งจะปรากฏในอนาคต เช่น เกณฑ์สัมพัทธ์ของแบบสอบถามความถนัดทางวิชาการ เพื่อทำนายผลการเรียน ก็อาจใช้คะแนนเฉลี่ยสะสมปีสุดท้ายเป็นเกณฑ์สัมพัทธ์ ซึ่งการคำนวณหาความตรงเชิงพยากรณ์นี้จะต้องอาศัยเวลารอคอย เพราะคะแนนของแบบสอบถามกับเกณฑ์สัมพัทธ์ได้มาคนละเวลาด้วยกัน ความตรงเชิงทำนายเหมาะสำหรับที่จะใช้กับแบบสอบถามความถนัดชนิดต่าง ๆ และแบบสอบถามที่ต้องการวัดเพื่อทำนายผล

ข้อแตกต่างระหว่างความตรงร่วมสมัยกับความตรงเชิงทำนาย ขึ้นอยู่กับวัตถุประสงค์ของการสอบ ความตรงร่วมสมัยใช้กับแบบสอบถามสำหรับวินิจฉัยสถานภาพที่เป็นอยู่ ส่วนความตรงเชิงทำนายใช้กับแบบสอบถามสำหรับพยากรณ์เหตุการณ์ในอนาคต¹ นอกจากนี้ เมเรนต์และเลแมนน์ (Mehrens and Lehmann) ได้กล่าวอีกว่า ข้อแตกต่างระหว่างความตรง 2 ชนิดนี้ คือช่วงระยะเวลาของการเก็บข้อมูลที่เป็นเกณฑ์ ถ้าเก็บรวบรวมข้อมูลที่เป็นเกณฑ์และคะแนนการสอบในเวลาเดียวกันจะเป็นความตรงร่วมสมัย แต่ถ้าเก็บคะแนนที่เป็นเกณฑ์ล่าออกไปหรือหลังการเก็บคะแนน จะเป็นความตรงเชิงทำนาย²

การวัดเกณฑ์

การที่จะได้ข้อมูลของเกณฑ์ที่เหมาะสมและถูกต้องนั้นทำได้ยาก การวัดเกณฑ์ก็เหมือนกับการวัดสิ่งอื่น ๆ คือผู้วัดจะต้องทราบคุณลักษณะของสิ่งที่จะเป็นเกณฑ์อย่างถี่ถ้วน การพิจารณาเกณฑ์นั้น เมเรนต์และเลแมนน์ (Mehrens and Lehmann)

¹แอน อนาสตาซี การตรวจสอบเชิงจิตวิทยา, หน้า 105.

²Mehrens and Lehmann, Measurement and Evaluation Education and Psychology, p. 112.

เสนอว่าจะต้องพิจารณาในสิ่งต่อไปนี้

1. ความเกี่ยวข้อง (Relevance) สิ่งที่จะใช้เป็นเกณฑ์จะต้องมีความเกี่ยวข้องกับเรื่องที่ศึกษา นั่นคือเกณฑ์นั้นจะต้องสะท้อนถึงคุณลักษณะที่สำคัญของเรื่องที่จะศึกษา โดยทั่วไปการพิจารณาความเกี่ยวข้องของเกณฑ์ขึ้นอยู่กับการตัดสินคุณค่า ไม่ได้วัดออกมาเป็นตัวเลข

2. ความเที่ยง (Reliability) เกณฑ์นั้นจะต้องมีความเที่ยงพอสมควร เนื่องจากความเที่ยงของแบบสอบถามมีอิทธิพลต่อค่าสหสัมพันธ์ระหว่างแบบสอบถามกับเกณฑ์ ซึ่งทฤษฎีความสัมพันธ์กล่าวไว้ว่า ความสัมพันธ์สูงสุดระหว่าง 2 ตัวแปร จะเท่ากับหรือน้อยกว่ารากกำลังที่สองของผลคูณระหว่างความเที่ยงของตัวแปรทั้งสองนั้น นั่นคือ

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

เมื่อ r_{xy} = ค่าสหสัมพันธ์ระหว่างตัวทำนาย (x) กับเกณฑ์ (y)

r_{xx} = ค่าความเที่ยงของแบบสอบถามที่เป็นตัวทำนาย

r_{yy} = ค่าความเที่ยงของเกณฑ์

จะเห็นว่า ค่าความตรงตามเกณฑ์สัมพันธ์ (r_{xy}) มีค่ามากหรือน้อยขึ้นอยู่กับทั้งความเที่ยงของแบบสอบถามเอง (r_{xx}) และความเที่ยงของเกณฑ์ (r_{yy})

3. ความเป็นอิสระจากอคติ (Free from bias) เกณฑ์ที่จะใช้จะต้องเป็นอิสระ ต้องแน่ใจว่าจะแนะนำจากแบบสอบถามที่จะหาความตรงนั้น ไม่มีอิทธิพลต่อสถานภาพใด ๆ ของเกณฑ์สัมพันธ์ที่ใช้เทียบ นั่นคือต้องไม่ให้เป็นเกณฑ์มลทิน (Criteria Contamination) ตัวอย่างเช่น นำแบบสอบถามถนัดทางวิชาการไปสอบถามนักเรียนตอนต้นปีเพื่อพยากรณ์ผลการเรียนของนักเรียนตอนปลายปี ถ้าปรากฏว่าความรู้ที่ครูได้รับจากการสอบครั้งนี้ไปส่งเสริมให้ประเมินคะแนนการสอบปลายปีของนักเรียนบางคนสูงขึ้นหรือต่ำลง ลักษณะเช่นนี้แสดงว่าคะแนนของเกณฑ์ไม่มีความอิสระ หรือเป็นเกณฑ์มลทิน ซึ่งอิทธิพลเช่นนี้ทำให้ค่าความตรงตามเกณฑ์สัมพันธ์ของแบบสอบถามสูงขึ้น โดยมีสาเหตุจากคุณภาพของแบบสอบถามจริง ๆ การป้องกันความคลาดเคลื่อนเช่นนี้อาจทำได้โดยหลีกเลี่ยงไม่ให้ผู้ประเมิน

ทราบคะแนนการสอบที่จะใช้พยากรณ์¹

วิธีประมาณค่าความตรงตามเกณฑ์สัมพัทธ์

การประมาณค่าความตรงตามเกณฑ์สัมพัทธ์สามารถวิเคราะห์ได้จาก สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบกับเกณฑ์ ดังนั้นสูตรที่ใช้จึงเป็นสูตรคำนวณสัมประสิทธิ์สหสัมพันธ์แบบต่าง ๆ โดยทั่วไปนิยมใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's product moment correlation coefficient)

องค์ประกอบที่มีอิทธิพลต่อความตรง

องค์ประกอบต่อไปนี้² ทำให้ค่าความตรงของแบบสอบลดลง

1. องค์ประกอบในแบบสอบ ได้แก่

1.1 คำสั่งไม่ชัดเจน คำสั่งไม่ชัดเจนจะทำให้ให้นักเรียนปฏิบัติตามคำสั่ง

ผิดพลาดได้

1.2 การใช้คำและประโยคที่ยากเกินไป การใช้คำและประโยคที่ยากเกินไปจะทำให้ให้นักเรียนเกิดความสับสนในการสอบ ผลที่ได้กลายเป็นนักเรียนที่ทำข้อสอบได้ถูกต้องคือผู้ที่มีความสามารถในการอ่านสูง ดังนั้นแบบสอบนี้จึงวัดความเข้าใจในการอ่านของนักเรียน แทนที่จะวัดพฤติกรรมที่ต้องการ

1.3 ระดับความยากของข้อกระทงไม่เหมาะสม ข้อกระทงที่ยากหรือง่ายเกินไปไม่สามารถจำแนกนักเรียนได้

¹Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, pp. 113-114.

²Gronlund, Measurement and Evaluation in Teaching, pp. 98-102.

1.4 การสร้างข้อระงทงไม่รัดกุม ข้อสอบที่สร้างอย่างไม่ระมัดระวัง อาจชี้แนะคำตอบแก่นักเรียน

1.5 ภาษาในข้อระงทงกำกวม ภาษาในข้อระงทงที่กำกวมจะทำให้ นักเรียนตีความผิดและสับสน

1.6 แบบสอบที่ใช้ไม่เหมาะจะใช้วัดพฤติกรรมที่ต้องการวัด

1.7 แบบสอบสั้นเกินไป แบบสอบที่สั้นเกินไปจะถามได้ไม่ครอบคลุม วัตถุประสงค์ของการวัด

1.8 การจัดเรียงข้อระงทงไม่เหมาะสม ข้อระงทงควรเรียงจากง่าย ไปหายาก ถ้าเรียงจากยากก่อนนักเรียนอาจใช้เวลาคิดนานจนกระทั่งไม่มีเวลาทำข้อง่าย ๆ ซึ่งอยู่ตอนท้ายของแบบสอบ จึงวัดได้ไม่ตรงตามความสามารถที่แท้จริง

1.9 คำตอบที่ถูกเรียงกันอย่างเป็นระบบ การมีระบบของคำตอบที่ถูก ต้องทำให้นักเรียนสามารถเดาคำตอบข้อหลัง ๆ ได้

นอกจากนั้น ค่าความเที่ยงของแบบสอบก็มีส่วนสำคัญที่จะกำหนดค่าความตรง ของแบบสอบ¹ คังสมการ

$$r_{xy} \leq \sqrt{r_{xx}}$$

เมื่อ r_{xy} = ค่าความตรงของแบบสอบ

r_{xx} = ค่าความเที่ยงของแบบสอบ

จะเห็นได้ว่าค่าความตรงของแบบสอบมีค่าไม่เกินรากกำลังที่สองของค่าความ เที่ยงของแบบสอบนั้น นั่นคือ ค่าความตรงของแบบสอบจะถูกจำกัดโดยค่าความเที่ยงของ มันเอง อย่างไรก็ตาม ความเที่ยงเป็นสิ่งจำเป็นแต่ก็ยังไม่เพียงพอสำหรับความตรงของ แบบสอบ²

¹Frederick G. Brown, Principle of Education and Psychological Testing, p. 99.

²Stanley and Hopkins, Educational and Psychological Measurement and Evaluation, p. 280.

2. การบริหารแบบสอบและการให้คะแนน

การบริหารแบบสอบและการให้คะแนน ทำให้ความตรงของแบบสอบลดลงได้ เช่น ให้ความเวลาในการสอบน้อยเกินไป การอธิบายข้อสอบให้แก่แก่นักเรียนบางคนที่สูงสัยหรือขอร้องให้ช่วย การทุจริตของนักเรียน การให้คะแนนที่ขาดความเชื่อถือได้ เมื่อใช้ข้อสอบอัตนัย สิ่งเหล่านี้ทำให้ความตรงของแบบสอบที่ครูสร้างลดลง ในกรณีของแบบสอบมาตรฐาน ถ้าผู้ดำเนินการสอบไม่ปฏิบัติตามคำสั่ง เช่น ไม่จับเวลาตามที่กำหนด ตรวจให้คะแนนผิด หรือให้ความช่วยเหลือแก่บางคน การปฏิบัติเหล่านี้ทำให้ความตรงของแบบสอบลดลงเช่นเดียวกัน

3. องค์ประกอบในคำตอบของนักเรียน

นักเรียนบางคนอาจมีสภาพจิตใจไม่ปกติ เช่น มีความวิตกกังวลในการสอบ จึงไม่สามารถทำแบบสอบได้เต็มตามความสามารถ หรือไม่มีแรงบันดาลใจให้ทำแบบสอบ สิ่งเหล่านี้ก็มีส่วนทำให้ความตรงของแบบสอบต่ำ นอกจากนี้อาจเกี่ยวกับลักษณะการตอบของนักเรียน เช่น นักเรียนบางคนเลือกตอบถูกเสมอในกรณีไม่แน่ใจเมื่อทำแบบสอบแบบถูกผิด ในขณะที่บางคนเลือกตอบผิดเสมอเมื่อไม่แน่ใจ และถ้าการเลือกโดยการเดาสุ่มนี้บังเอิญไปตรงกับคำตอบที่ถูกต้องของข้อกระทงเป็นส่วนใหญ่ คะแนนที่ได้จะไม่ตรงกับความสามารถที่แท้จริงของนักเรียน ทำให้ความตรงของแบบสอบต่ำ

4. ลักษณะของกลุ่มและเกณฑ์

ความตรงมักจะเป็นของเฉพาะกลุ่ม เนื่องจากสิ่งที่แบบสอบวัดนั้นมักได้รับอิทธิพลของตัวแปรต่าง ๆ เช่น อายุ เพศ ระดับความสามารถ ภูมิหลังของการศึกษา และวัฒนธรรมของผู้ตอบ ดังนั้นในการพิจารณาหาความตรงของแบบสอบที่เป็นมาตรฐาน ควรคำนึงถึงลักษณะเฉพาะของกลุ่มที่ใช้หาความตรงนอกจากนั้นต้องพิจารณาเกณฑ์ที่ใช้หาความตรงของแบบสอบด้วย

งานวิจัยที่เกี่ยวข้อง

ริชาร์ดสันและสตอลเนเกอร์ (Richardson and Stalnaker) วิเคราะห์การสอบผลสัมฤทธิ์ เขาสรุปความเห็นในเรื่องแบบสอบไว้ว่า รูปแบบของข้อ กระถงไม่อาจบอกสมรรถภาพหรือคุณภาพของแบบสอบได้ และที่มีผู้กล่าวว่าข้อมีส่วนช่วย ให้ผู้สอบได้คะแนนมากหรือน้อยนั้นไม่เป็นความจริง โดยทั่วไปแล้ว แบบสอบทั้งแบบอัตนัย แบบปรนัย และแบบแก้ปัญหานั้น ผู้สอบต่างก็ต้องใช้ระดับและชนิดของความสามารถอย่างเดียวกัน และถ้าสร้างอย่างใดแล้วจะให้ค่าความเที่ยงเท่ากัน แบบสอบแบบอัตนัยและแบบปรนัยที่ดี สามารถจำแนกหรือจัดอันดับนักเรียนได้ก็เท่า ๆ กัน แต่ทั้งนี้ไม่ได้หมายความว่าทั้ง 3 แบบ นี้ จะใช้แทนกันได้เสมอไป¹

รอสส์ (Ross) สรุปผลการทดลองเกี่ยวกับการเปรียบเทียบแบบของแบบสอบ ชนิดต่าง ๆ ว่า การเปรียบเทียบแบบของแบบสอบควรเปรียบเทียบในเนื้อหาเดียวกัน จุดมุ่งหมายเดียวกัน แบบสอบนั้นต้องสร้างให้ดีเท่ากัน และใช้เวลาในการสอบเท่ากัน เรา จัดระดับของความรู้ไว้เป็น 4 ระดับ เรียงจากต่ำไปสูงดังนี้

1. ระดับความจำ (Recognition)
2. ระดับย้อนระลึก (Recall)
3. ระดับตีความและประเมินค่า (Interpret and Evaluation)
4. ระดับการนำไปใช้ (Application)

¹M.W. Richardson and J.M. Stalnaker, "Comments on Achievement Examinations" Journal of Educational Research 28(1935) quoted in Robert L. Ebel, Measuring Education Achievement, p. 59.

การสอบวัดความรู้ระดับใดควรเลือกใช้ชนิดของแบบสอบให้เหมาะสมกับระดับความรู้ที่ต้องการวัด แบบสอบแบบเลือกตอบกับแบบจับคู่เหมาะกับระดับแรก คือ ระดับความจำ ส่วนแบบเติมคำเหมาะกับระดับที่สูงขึ้นไปทั้ง 3 ระดับ แต่ทั้งนี้ก็ขึ้นอยู่กับความสามารถของผู้สร้างแบบสอบด้วย แบบสอบแบบถูกผิด แบบเลือกตอบ และแบบจับคู่ ถ้าสร้างให้ดีแล้วสามารถวัดความเข้าใจ การประเมินค่าการนำไปใช้และความสามารถอื่นได้ ดังนั้นถ้าจะเปรียบเทียบคุณภาพของแบบสอบแต่ละแบบ จำต้องอาศัยความสามารถและทักษะในการสร้างแบบสอบแต่ละแบบด้วย¹

ใน ค.ศ. 1929 ริช (Ruch) รวบรวมผลการทดลองเกี่ยวกับการเปรียบเทียบแบบสอบชนิดต่าง ๆ แล้วสรุปว่า แบบสอบแบบปรนัยชนิดต่าง ๆ อย่างน้อยก็มีความตรงเท่ากับแบบสอบแบบอัตนัย และเขาได้สรุปต่อไปอีกว่า ถ้าให้เวลาในการสอบเท่ากันแบบสอบแบบเติมคำให้สมบูรณ์ แบบถูกผิด แบบเลือกตอบ และแบบจับคู่ จะมีค่าความตรงเท่ากัน²

เมเยอร์ (Mayer) ศึกษาผลของการใช้แบบสอบแบบอัตนัย และแบบปรนัยแบบถูกผิด แบบเลือกตอบ และแบบเติมคำ โดยศึกษากับเด็กนักเรียน 124 คน แบ่งออกเป็น 4 กลุ่ม แล้วให้สอบแบบสอบดังนี้

กลุ่มที่ 1 สอบแบบสอบแบบถูกผิด

กลุ่มที่ 2 สอบแบบสอบแบบเลือกตอบ

¹C.C. Ross, Measurement in Today's Schools, revised by Julian C. Stanley, 3d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1954), p. 167.

²G.M. Ruch, The Objective or New-Type Examination, (Chicago: Scott, Foresman & Co., 1929), quoted in Ross, C.C. Measurement in Today's School, p. 167.

- กลุ่มที่ 3 สอบแบบสอบแบบเติมคำ
 กลุ่มที่ 4 สอบแบบสอบแบบอัตนัย

ผลจากการศึกษา แบบสอบทั้ง 4 แบบให้แรงกระตุ้นในการเรียนแก่เด็กเท่า ๆ กันหลังจากสอบเสร็จแล้ว 5 สัปดาห์ ให้เด็กทำแบบสอบเติมคำอีก ปรากฏว่ากลุ่มที่สอบแบบสอบแบบอัตนัยและแบบเติมคำทำคะแนนได้สูงขึ้น ส่วนอีก 2 แบบคือ แบบเลือกตอบ และแบบถูกผิด คะแนนไม่สูงขึ้น แสดงว่าแบบสอบ 2 แบบหลังนี้ มีความเที่ยงมากกว่า 2 แบบแรก¹

ลินควิสและคณะ (Lindquist and Others) ศึกษาเปรียบเทียบความเที่ยงและความตรงของแบบสอบแบบต่าง ๆ เขาชี้ให้เห็นว่า การศึกษาเปรียบเทียบรูปแบบของแบบสอบ ผู้ศึกษามักลืมนึกถึงธรรมชาติของความตรง ไปเน้นที่ความเที่ยงมากเกินไป ทั้งผู้ศึกษาเองก็ไม่สามารถควบคุมตัวแปรเกี่ยวกับความชำนาญในการสร้างข้อกระทงแต่ละแบบได้ควย เขาเสนอว่าการเลือกใช้แบบสอบแบบต่าง ๆ ในจุดมุ่งหมายใด จุดมุ่งหมายหนึ่ง การคำนึงถึงวิธีการพิจารณาอย่างมีเหตุผล มีประโยชน์กว่าการคำนึงถึงผลการทดลองหรือหลักฐานที่ปรากฏชัดอื่น ๆ²

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.

²E.F. Lindquist and Others, The Construct and Use of Achievement Examination, (Boston : Houghton Mifflin Co., 1936) quoted in Ross, . Measurement in Today's Schools, p. 165.



รินส์แลนด์ (Rinsland) รวบรวมผลการทดลองที่เกี่ยวกับแบบสอบแบบต่าง ๆ เขาสรุปไว้ว่า แบบสอบแบบปรนัยทั้งหมดยกเว้นแบบถูกผิดมีค่าความตรงพอ ๆ กับแบบอัตนัย หรือน้อยกว่าเพียงเล็กน้อย ส่วนค่าความเที่ยงโดยทั่วไปแบบสอบแบบปรนัยทุกชนิดจะให้ค่าความเที่ยงเท่ากัน เมื่อใช้เวลาในการสอบเท่ากัน แต่ที่จะมีความเที่ยงต่างกันนั้นอาจจะ เป็นเพราะการใช้คำหรือภาษาที่แตกต่างกันมากกว่าจะเป็นเพราะรูปแบบของคำถามหรือข้อ กระทบ¹

เออร์วิน, ฮัลเปอร์น และ แลนด์แมน (Irvin, Halpern, and Landman) ได้ประเมินผลสัมฤทธิ์ที่เด็กฝึกปรักติด้วยแบบสอบมาตรฐานแบบถูกผิดและแบบเลือกตอบ เครื่องมือที่ใช้ได้แก่แบบสอบ 3 ฉบับ ฉบับที่ 1 เป็นแบบสอบมาตรฐานชุดสังคมและอาชีพ (SPIB) 3 ชุด ได้แก่ชุด การธนาคาร การซื้อขาย และทักษะในการทำงาน ซึ่งมี รูปแบบของข้อกระทบเป็นแบบถูกผิด ฉบับที่ 2 เป็นแบบเลือกตอบ 3 ตัวเลือกที่สร้างขึ้น เพื่อให้วัดในเนื้อหาเดียวกับฉบับที่ 1 ฉบับที่ 3 เป็นแบบเลือกตอบ 2 ตัวเลือกสร้างโดย วิธีตัดตัวลงที่มีค่าอำนาจจำแนกต่ำสุดของข้อกระทบในฉบับที่ 2 ออกข้อละ 1 ตัว กลุ่ม ตัวอย่างเป็นนักเรียนฝึกปรักติเกรด 9 - 12 จำนวน 425 คน ซึ่งแบ่งออกเป็น 3 กลุ่ม แต่ละกลุ่มให้ทำแบบสอบฉบับใดฉบับหนึ่งใน 3 ฉบับ ผลการวิจัยพบว่า ทั้งค่าความเที่ยงและ ความตรงของแบบสอบแบบถูกผิด กับแบบเลือกตอบมีค่าสูงและใกล้เคียงกัน²

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

² 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) : 51-58.

อีเบล (Ebel, 1971) ศึกษาเปรียบเทียบค่าความเที่ยงและความตรงของแบบสอบแบบถูกผิด และแบบเลือกตอบ โดยใช้อัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้ในเวลาเท่ากัน (ในที่นี้มีค่าอัตราส่วนเป็น 2:1) เป็นตัวปรับค่าความเที่ยงของแบบสอบแบบถูกผิดเพื่อให้เป็นค่าความเที่ยงของแบบสอบแบบถูกผิดที่ใช้เวลาในการสอบเท่ากับแบบสอบแบบเลือกตอบ ผลการศึกษาพบว่า

1. ค่าความเที่ยงของแบบสอบแบบเลือกตอบและแบบถูกผิดไม่แตกต่างกัน
2. แบบสอบแบบเลือกตอบและแบบถูกผิดสามารถวัดได้ในสิ่งเดียวกัน¹

กรีน (Green) ศึกษาเปรียบเทียบค่าความเที่ยงและความตรงของแบบสอบแบบเลือกตอบกับแบบถูกผิด โดยใช้แบบสอบสัมฤทธิ์ผลวิชาการทดสอบและการวัดผลที่ครูสร้างขึ้นสำหรับสอบกลางเทอมที่ไม่เคยผ่านการปรับปรุงมาก่อน 2 ฉบับ ฉบับที่ 1 เป็นแบบเลือกตอบ 3 ตัวเลือก 40 ข้อ แต่ละข้อ ข้อคำถามหรือตัวยืม (stem) อยู่ร่วมกับตัวเลือก (option) แล้วแบ่งข้อกระทง 40 ข้อ ออกเป็น 3 ส่วน ๆ แรกมี 14 ข้อ ส่วนที่สองมี 13 ข้อ และส่วนที่สามมี 13 ข้อ ฉบับที่ 2 เป็นแบบถูกผิดที่แปลงมาจากฉบับที่ 1 ด้วยวิธีแปลงข้อกระทงของฉบับที่ 1 แต่ละข้อให้เป็นแบบถูกผิด 3 ข้อ โดยเป็นข้อถูกตามตัวเลือกที่เป็นตัวถูก 1 ข้อ และเป็นข้อผิดตามตัวลวง 2 ข้อ ได้ข้อกระทงแบบถูกผิดทั้งหมด 120 ข้อ แบ่งข้อกระทงทั้งหมดออกเป็น 3 ส่วนเท่า ๆ กัน ด้วยวิธีนำข้อกระทงทุก ๆ 3 ข้อที่แปลงมาจากข้อกระทงแบบเลือกตอบแต่ละข้อมาแยกไปอยู่ข้อละส่วนโดยวิธีสุ่ม หลังจากนั้นจัดลำดับของข้อกระทงในแต่ละส่วนโดยวิธีสุ่มเช่นเดียวกัน กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือนิสิตระดับปริญญาตรี จำนวน 50 คน แบ่งออกเป็น 2 กลุ่ม ๆ ละ 25 คน ให้แต่ละกลุ่มทำแบบสอบ

¹Ebel, 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.

ฉบับใดฉบับหนึ่งจาก 2 ฉบับ โดยให้ทำที่ละส่วนเสร็จแล้วจึงให้ทำส่วนต่อไป และเมื่อหมดเวลา 12 นาทีแรกของการสอบ ให้ให้นักบันทึกจำนวนข้อที่ทำได้ไว้ด้วย ผลการศึกษาพบว่า

1. ความเที่ยงของแบบสอบแบบถูกผิดกับแบบเลือกตอบหลังจากปรับแล้ว (โดยเพิ่มความยาวของแบบเลือกตอบเป็น 1.25 เท่าของจำนวนข้อเดิมเพื่อให้ใช้เวลาในการสอบเท่ากับแบบสอบแบบถูกผิด) มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ
2. ค่าความตรงของแบบสอบแบบเลือกตอบและแบบถูกผิด ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ
3. ค่าอัตราส่วนของจำนวนข้อแบบถูกผิดต่อแบบเลือกตอบในกลุ่มตัวอย่างทำได้ในเวลาเท่ากัน เป็น 2.4:1¹

สุธรรม์ จันทรหอม ศึกษาเปรียบเทียบผลของการใช้แบบสอบแบบถูกผิด แบบเลือกตอบ และแบบเติมคำ วัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 7 กลุ่มตัวอย่างจำนวน 1350 คน เครื่องมือที่ใช้ศึกษาคือแบบสอบวิชาคณิตศาสตร์ ซึ่งแยกเป็น 3 ฉบับ คือ คณิตศาสตร์ทักษะ คณิตศาสตร์ปัญหา และคณิตศาสตร์ความเข้าใจ เก็บรวบรวมข้อมูลโดยให้นักเรียนสอบคนละ 3 ฉบับ แต่ละคนไม่ให้รูปแบบของข้อกระทงซ้ำกัน การศึกษาครั้งนี้ใช้จำนวนข้อของแบบสอบทั้ง 3 แบบเท่ากัน และวัดในเนื้อหาเดียวกัน ผลการศึกษาเกี่ยวกับค่าความเที่ยงและความตรงของแบบสอบพบว่า

1. ค่าความเที่ยงของแบบสอบแบบเลือกตอบ แบบถูกผิด และแบบเติมคำ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบเติมคำมีความเที่ยงสูงสุด และแบบถูกผิดมีค่าต่ำสุด

¹Kathy Green, Multiple Choice and True-False : Reliability and Validity Compared," Journal of Experimental Education 48(Fall 1979) : 42-44.

2. ค่าความตรงของแบบสอบแบบถูกผิด แบบเลือกตอบ และแบบเติมคำของวิชาคณิตศาสตร์ทักษะ และคณิตศาสตร์ปัญหาไม่แตกต่างกัน ส่วนวิชาคณิตศาสตร์ความเข้าใจแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01¹

ฟรีสบี (Frisbie, 1973) ศึกษาเปรียบเทียบค่าความเที่ยงและความตรงของแบบสอบแบบเลือกตอบ และแบบถูกผิด กลุ่มตัวอย่างเป็นนักเรียนนอกเขตนครหลวงจำนวน 1018 คน แบบสอบที่ใช้ศึกษาคือ แบบสอบสัมฤทธิ์ผลวิชาสังคมศึกษา (Social Study) และวิชาวิทยาศาสตร์ธรรมชาติ (Natural Science) ทั้ง 2 วิชา สร้างข้อกระทงเป็นแบบเลือกตอบก่อน แบบสอบวิชาสังคมศึกษาเรียกว่า ฟอรั่ม (form) SM วิชาวิทยาศาสตร์ธรรมชาติเรียกว่าฟอรั่ม NM หลังจากนั้นแปลงข้อกระทงแบบเลือกตอบให้เป็นแบบถูกผิด ซึ่งมีวิธีแปลง 2 วิธีดังนี้

1. โดยให้ครูผู้สอนวิชาละ 5 คน เป็นผู้ตัดสินคุณภาพของตัวดวงของข้อกระทงแต่ละข้อ ถ้าครูอย่างน้อย 4 คน ตัดสินว่าตัวดวงใดดีที่สุดตรงกันแล้ว ข้อนั้นจะแปลงเป็นข้อผิดตามตัวดวงนั้น ถ้าไม่เป็นไปตามเกณฑ์ดังกล่าว จะแปลงเป็นข้อถูกตามตัวเลือกที่เป็นตัวถูก เมื่อแปลงแล้ว แบบสอบวิชาสังคมศึกษาเรียกว่าฟอรั่ม SJ และวิชาวิทยาศาสตร์ธรรมชาติเรียกว่าฟอรั่ม NJ

2. โดยวิธีพิจารณาค่าอำนาจจำแนกของตัวดวงของข้อกระทงแต่ละข้อ ตัวดวงที่ดีที่สุดของข้อใดมีค่าอำนาจจำแนกตั้งแต่ $-.20$ ลงมา ข้อนั้นจะแปลงเป็นข้อผิดตามตัวดวงนั้น ถ้าไม่เป็นไปตามเกณฑ์นี้จะแปลงเป็นข้อถูกตามตัวเลือกที่เป็นตัวถูก เมื่อแปลงแล้วแบบสอบวิชาสังคมศึกษา เรียกว่าฟอรั่ม SD และวิชาวิทยาศาสตร์ธรรมชาติเรียกว่าฟอรั่ม ND

¹สุธรรม จันทร์หอม "การศึกษาเปรียบเทียบการใช้แบบสอบแบบถูกผิด แบบเลือกตอบ และแบบเติมคำ วัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 7" วิทยานิพนธ์ปริญญาโทบริหารศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2513.

หลังจากนั้นนำแบบสอบทั้ง 6 ฟอรม (SM, NM, SJ, NJ, SD และ ND) มาจัดใหม่ให้เป็น 8 ฟอรม ซึ่งแตกต่างกันตามเนื้อหาวิชาและวิธีแปลงดังนี้

ฟอรมที่ 1	SJA	ประกอบด้วยข้อ	1-35	เป็นของฟอรม SM	และข้อ	36-70	เป็นของฟอรม SJ
ฟอรมที่ 2	SJB	ประกอบด้วยข้อ	1-35	เป็นของฟอรม SJ	และข้อ	36-70	เป็นของฟอรม SM
ฟอรมที่ 3	SDA	ประกอบด้วยข้อ	1-35	เป็นของฟอรม SM	และข้อ	36-70	เป็นของฟอรม SD
ฟอรมที่ 4	SDB	ประกอบด้วยข้อ	1-35	เป็นของฟอรม SD	และข้อ	36-70	เป็นของฟอรม SM
ฟอรมที่ 5	NJA	ประกอบด้วยข้อ	1-35	เป็นของฟอรม NM	และข้อ	36-70	เป็นของฟอรม NJ
ฟอรมที่ 6	NJB	ประกอบด้วยข้อ	1-35	เป็นของฟอรม NJ	และข้อ	36-70	เป็นของฟอรม NM
ฟอรมที่ 7	NDA	ประกอบด้วยข้อ	1-35	เป็นของฟอรม NM	และข้อ	36-70	เป็นของฟอรม ND
ฟอรมที่ 8	NDB	ประกอบด้วยข้อ	1-35	เป็นของฟอรม ND	และข้อ	36-70	เป็นของฟอรม NM

เก็บรวบรวมข้อมูลโดยให้นักเรียนทำแบบสอบคนละ 1 ฟอรม โดยแบ่งเวลาเป็น 2 ช่วง ๆ แรกทำ 8 นาที แล้วให้นักเรียนบันทึกจำนวนข้อที่ทำได้ หลังจากนั้นให้เวลาทำต่ออีก 50 นาที

ผลการศึกษาสรุปได้ว่า

- อัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้ภายในเวลา 8 นาทีแรกของการสอบเป็น 1.5:1
- ค่าความเที่ยงของแบบสอบแบบเลือกตอบ กับ แบบถูกผิดหลังจากปรับแล้ว (โดยให้ความยาวของแบบสอบแบบถูกผิดเป็น 1.5 เท่าของจำนวนข้อเดิม เพื่อให้ใช้เวลาในการสอบเท่ากับแบบเลือกตอบ) แตกต่างกันอย่างมีนัยสำคัญทางสถิติ
- แบบสอบแบบเลือกตอบและแบบถูกผิดมีแนวโน้มที่จะวัดได้ในสิ่งเดียวกัน¹

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

ต่อมาปี ค.ศ. 1974 ฟรีสบี ศึกษาอิทธิพลของข้อกระทงแบบเลือกตอบและแบบ
 ถูกผิดที่มีต่อค่าความเที่ยงของแบบสอบอีกครั้งหนึ่ง โดยมีวัตถุประสงค์เพื่อเปรียบเทียบค่า
 ความเที่ยงของแบบสอบแบบเลือกตอบกับแบบถูกผิดที่ยาวเป็น 1.5 เท่าของแบบสอบแบบ
 เลือกตอบ เปรียบเทียบค่าความตรงของแบบสอบแบบเลือกตอบกับแบบถูกผิด รวมทั้งหาอัตรา
 ส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้ภายในเวลาที่เท่ากัน เพื่อ
 ยืนยันกับผลการวิจัยครั้งก่อน กลุ่มตัวอย่างเป็นนักเรียนนอกเขตนครหลวง จำนวน 529 คน
 แบบสอบที่ใช้ศึกษาคือแบบสอบสัมฤทธิ์ผลแบบเลือกตอบวิชาสังคมศึกษา และวิชาวิทยาศาสตร์
 ชรรมาชาติที่ใช้ในการศึกษาครั้งก่อน แต่แปลงให้เป็นแบบสอบแบบถูกผิดโดยวิธีพิจารณา
 อ่านาจำแนกของตัวลวงวิธีเดียวกันนั้น หลังจากนั้นจัดแบบสอบให้เป็น 4 ฟอรั่ม โดยให้
 แต่ละฟอรั่มประกอบด้วยข้อกระทงแบบถูกผิดและแบบเลือกตอบ และให้แบบถูกผิดมีความยาว
 เป็น 1.5 เท่าของแบบเลือกตอบ อัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบ
 แต่ละฟอรั่มยึดตามค่าอัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้
 ในเวลาที่กำหนด ตามผลการศึกษาก่อน (1.5:1) วิธีการเก็บรวบรวมข้อมูลเหมือน
 ครั้งแรกทุกประการ

ผลการศึกษาพบว่า แม้ว่าแบบสอบแบบถูกผิดจะยาวเป็น 1.5 เท่าของแบบเลือก
 ตอบ แต่ค่าความเที่ยงของแบบสอบแบบถูกผิดก็ยังต่ำกว่าแบบเลือกตอบ อย่างมีนัยสำคัญทาง
 สถิติ ค่าอัตราส่วนของจำนวนข้อของแบบถูกผิดต่อแบบเลือกตอบที่นักเรียนทำได้ในเวลา
 กำหนด วิชาสังคมศึกษาเท่ากับ 1.44:1 และวิชาวิทยาศาสตร์ชรรมาชาติมีค่าเท่ากับ 1.53:1
 และแบบสอบทั้งสองรูปแบบนี้วัดได้ในสิ่งเดียวกัน ผลการศึกษานี้สอดคล้องกับผลการศึกษา
 ครั้งก่อนทุกประการ¹

¹ David A. Frisbie, "The Effect of Item Format on Reliability and Validity : A Study of Multiple Choice and True-False Achievement Tests," Educational and Psychological Measurement 34(1974) : 885-892.

ปี ค.ศ. 1974 เช่นเดียวกัน อูสเตอร์ฮอฟและกลาสแนพ (Oosterhof and Glasnapp) ศึกษาเปรียบเทียบค่าความเที่ยงและความยากของแบบสอบแบบเลือกตอบ และแบบถูกผิด โดยใช้แบบสอบสัมฤทธิ์ผลวิชาการศึกษาเบื้องต้นที่ครูสร้างขึ้น 2 ฉบับ ฉบับที่ 1 เป็นแบบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ฉบับที่ 2 เป็นแบบถูกผิดที่แปลงมาจากฉบับที่ 1 ด้วยวิธีแปลงข้อกระทงของฉบับที่ 1 แต่ละข้อให้เป็นแบบถูกผิด 2 ข้อ คือ ข้อถูกและข้อผิด ข้อถูกได้มาจากการรวมข้อคำถามกับตัวเลือกที่เป็นตัวถูก ส่วนข้อผิดได้มาจากการรวมข้อคำถามกับตัวลวงที่มีค่าอำนาจจำแนกสูงสุด หลังจากแปลงแล้วได้ข้อกระทงแบบถูกผิด 80 ข้อ นำข้อกระทงทั้ง 80 ข้อมาแยกเป็น 2 ส่วน โดยแยกข้อกระทงแบบถูกผิดที่จะถูกไปอยู่ข้อละส่วนโดยวิธีสุ่ม แล้วจัดลำดับของข้อกระทงในแต่ละส่วนด้วยวิธีสุ่มเช่นเดียวกัน กลุ่มตัวอย่างของการวิจัยคือ นิสิตระดับปริญญาตรี จำนวน 101 คน แบ่งออกเป็น 2 กลุ่ม ๆ แรก 51 คน ให้ทำแบบสอบฉบับที่ 1 ก่อน แล้วจึงทำฉบับที่ 2 ส่วนกลุ่มที่ 2 มี 50 คน ให้ทำแบบสอบฉบับที่ 2 ก่อนแล้วจึงทำฉบับที่ 1 ในการทำแบบสอบให้บันทึกเวลาที่ใช้เมื่อทำเสร็จ 40 ข้อ, 80 ข้อ และ 120 ข้อ ผลการศึกษาได้ข้อสรุปที่น่าสนใจ คือ

1. อัตราส่วนของจำนวนข้อกระทงแบบถูกผิดต่อแบบเลือกตอบที่กลุ่มตัวอย่างทำได้ในเวลาเท่ากันเป็น 1.73:1
2. ค่าความเที่ยงของแบบสอบแบบถูกผิดหลังจากปรับแล้วมีค่าต่ำกว่าค่าความเที่ยงของแบบสอบแบบเลือกตอบ และถ้าจะให้แบบสอบแบบถูกผิดมีค่าความเที่ยงเท่ากับแบบสอบแบบเลือกตอบแล้ว แบบสอบแบบถูกผิดจะต้องยาวเป็น 3.75 เท่าของแบบสอบแบบเลือกตอบ
3. ค่าความเที่ยงของแบบสอบแบบถูกผิด ส่วนที่เป็นข้อผิดทั้งหมด สูงกว่าส่วนที่เป็นข้อถูกทั้งหมด¹

¹Albert C. Oosterhof, and Douglas R. Glasnapp, "Comparative Reliabilities and Difficulties of the Multiple-Choice and True-False Formats," Journal of Experimental Education 42(Spring 1974) : 62-64.