



วรรณคดีที่เกี่ยวข้อง

การสร้างแบบสอบชนิดเลือกตอบให้มีคุณภาพดีนั้น นอกจากการสร้างคำถามที่ดีแล้ว การสร้างตัวเลือกให้มีคุณภาพก็เป็นสิ่งสำคัญอีกประการหนึ่งเช่นกัน และการเปรียบเทียบคุณภาพของแบบสอบ จำเป็นต้องศึกษาถึงคุณลักษณะที่ดีของแบบสอบว่าเป็นอย่างไร ก่อนที่จะนำคุณลักษณะเหล่านั้นมาเปรียบเทียบกัน ผู้วิจัยจึงได้ศึกษาวรรณคดีและงานวิจัยที่เกี่ยวข้องโดยแบ่งเป็น 3 ตอนดังนี้

- ตอนที่ 1 เกี่ยวกับการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบ
- ตอนที่ 2 เกี่ยวกับคุณลักษณะของแบบสอบที่ดี
- ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 เกี่ยวกับการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบ

ในการสร้างข้อสอบชนิดเลือกตอบ ถือว่าการสร้างตัวเลือกเป็นสิ่งสำคัญที่สุดอย่างหนึ่ง เนื่องจากคุณภาพของข้อสอบชนิดนี้จะมีมากน้อยเพียงใด ขึ้นอยู่กับ การเขียนตัวเลือกเป็นสำคัญ เพราะตัวลวงที่ดีนั้นช่วยให้ข้อสอบมีคุณภาพดี¹ ถ้าตัวลวงนั้นไม่สามารถลวงผู้สอบได้แล้ว นอกจากจะทำให้เสียเวลาในการออกข้อสอบ ยังทำให้ค่าความเที่ยงและความตรงของแบบสอบลดลงอีกด้วย² ฉะนั้นในการสร้างข้อสอบชนิดนี้ จำเป็นต้องใช้ตัวเลือกที่ใกล้เคียงกับคำตอบถูก

¹Georgia S. Adams, Measurement and Evaluation in Educational Psychology and Guidance (New York:Holt Rinehart and Winston Inc., 1964), p. 340.

²Jon C. Marshall and Loyde W. Hales, Essentials of Testing (Massachusetts:Addison-Wesley, 1972), p. 62

มากที่สุด เพื่อให้ให้นักเรียนที่ไม่มีความรู้จริง เลือกตอบตัวดวงเหล่านั้นกระจายกันออกไปทุกตัว¹ ถ้าตัวดวงตัวใดตัวหนึ่งไม่ใกล้เคียงกับตัวเลือกที่เป็นคำตอบถูก นักเรียนก็จะไม่เลือกตัวดวงเหล่านั้นเลย หากเป็นข้อสอบชนิด 5 ตัวเลือก ข้อสอบนั้นก็จะมีค่าเท่ากับข้อสอบชนิด 4 ตัวเลือก ยิ่งถ้าข้อสอบใดมีตัวดวงที่ไม่มีประสิทธิภาพหลายๆ ตัวก็จะทำให้ข้อสอบนั้นมีคุณภาพลดลงเรื่อยๆ

เกี่ยวกับเรื่องนี้ได้มีผู้สรุปหลักเกณฑ์ และให้ข้อเสนอแนะในการสร้างแบบสอบชนิดเลือกตอบโดยเฉพาะในส่วนที่เกี่ยวกับการสร้างตัวเลือกไว้ดังต่อไปนี้

สมหวัง พิธิยานุวัฒน์² ได้กล่าวถึงหลักการและข้อเสนอแนะในการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบไว้คือ

1. ตัวดวงควรสร้างจากความเข้าใจผิดของผู้สอบ ซึ่งมักจะได้จากคำตอบในการตอบข้อสอบแบบเรียงความ ข้อสอบแบบตอบสั้นๆ หรือข้อสอบแบบเติมคำ
2. ควรหาตัวดวงซึ่งคล้ายคลึงกับตัวเลือกที่เป็นคำตอบถูกให้มากที่สุด อีกหึ่งเป็นตัวดวงที่ดึงดูดผู้ตอบที่ขาดความรู้ความสามารถที่มุ่งวัด
3. ควรหลีกเลี่ยงคำซ้ำในตัวเลือก พยายามตัดทอนให้สั้นกระชับที่สุด
4. ใช้ตัวเลือกอย่างน้อย 4 ตัว เพื่อลดโอกาสเดาและควรให้ข้อสอบแบบเลือกตอบทุกข้อมีจำนวนตัวเลือกเท่ากัน
5. ตัวดวงแต่ละตัวควรเป็นที่สนใจแก่ผู้ตอบที่ไม่ทราบคำตอบที่ถูกต้องพอๆ กัน ถ้าไม่เช่นนั้นแล้วข้อสอบชนิดเลือกตอบ 4 ตัวเลือกอาจกลายเป็นข้อสอบแบบถูกผิดได้ ถ้าตัวเลือก 2 ตัวไม่ดึงดูดผู้ตอบที่ไม่ทราบคำตอบเลย ถ้าไม่สามารถหาตัวดวงที่ดีได้ควรเลือกใช้ข้อสอบแบบอื่นมากกว่าข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก

¹ Victor H. Noll, Introduction to Educational Measurement (Boston:Houghton Mifflin Company, 1965), p. 130.

² สมหวัง พิธิยานุวัฒน์, "หลักในการเขียนข้อสอบแบบเลือกตอบ," ใน การพัฒนาการเรียนการสอน (กรุงเทพมหานคร:คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์, 2523), หน้า 102-103. (อักษราเนา)

6. ควรให้ตำแหน่งของคำตอบที่ถูกต้องอยู่แบบสุ่ม หรืออยู่อย่างกระจุกกระจาย
7. ไม่ควรทำให้คำตอบที่ถูกต้องยาวหรือสั้นกว่าตัวดวงจนสังเกตเป็นรูปแบบได้
8. หลีกเลี่ยงคำหรือข้อความที่ชี้แนะถึงคำตอบที่ถูกต้อง
9. ควรหลีกเลี่ยงตัวเลือกที่ว่า "ทุกข้อที่กล่าวมาแล้ว"

ยัง พินยานิยม¹ ได้แนะนำถึงการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบไว้ดังนี้

1. เขียนคำตอบที่ถูกลงไปก่อน และสร้างตัวดวงที่หลัง
2. เขียนตัวเลือกให้เป็นพวกเดียวกัน
3. ใช้ตัวเลือกที่เป็นอิสระจากกัน ไม่ก้าวก่ายหรือมีความหมายซ้ำซ้อนกัน
4. ตัวเลือกทุกตัวต้องใช้ประโยชน์ได้
5. ให้ตัวเลือกทุกตัวมีความยาวเท่ากัน เพื่อไม่เป็นการแนะนำคำตอบ
6. ให้ตัวดวงมีส่วนถูกเฉียบๆ กับคำตอบถูก ตัวดวงที่จัดว่ามีประสิทธิภาพจะต้องสามารถทำให้นักเรียนที่ไม่รู้อันใดเลือกตอบตัวดวงเหล่านั้น
7. คำถามข้อเดียวให้มีตัวเลือกที่ถูกเพียงคำตอบเดียว
8. กำหนดจำนวนตัวเลือกให้เหมาะสมกับวัยของเด็ก เช่นระดับประถมศึกษาอาจใช้ตัวเลือกเพียง 3 ตัว ระดับมัธยมศึกษาตอนต้นควรใช้ 4 ตัวเลือก ส่วนในระดับสูงกว่านี้ควรใช้ 5 ตัวเลือก แต่ทั้งนี้ขึ้นอยู่กับ การเขียนตัวเลือกด้วย
9. ตัวถูกหรือตัวผิดต้องถูกหรือผิดตามหลักการ หรือหลักวิชา และเป็นที่ยอมรับของบุคคลทั่วไป
10. ให้ตัวเลือกที่เป็นคำตอบกระจายแบบสุ่ม
11. อย่าให้คำถามหรือตัวเลือกข้ออื่นๆ ไปมีอิทธิพลกับคำตอบข้อต่อไป
12. ใช้ตัวเลือกแบบปลายเปิดและปลายปิดให้เหมาะสม
13. อย่าให้คำตอบถูกและตัวดวง ถูกหรือผิดเด่นชัดเกินไป

¹ยัง พินยานิยม, การวัดผลการศึกษา (กรุงเทพมหานคร:หน่วยศึกษานิเทศกรรมอาชีวศึกษา กระทรวงศึกษาธิการ, 2523), หน้า 122-134.

14. จัดเรียงตัวเลือกให้เป็นระเบียบ ถ้าไม่สามารถเขียนตัวเลือกให้ยาวเท่ากันได้ ก็พยายามเรียงตัวเลือกให้ตัวที่สั้นที่สุดอยู่ข้างบน ตัวที่ยาวกว่าอยู่รองลงมาจนถึงตัวสุดท้ายเป็นตัวเลือกที่ยาวที่สุดของข้อนั้น

ธอร์นไดค์¹ (Thorndike) ได้กล่าวถึงตัวเลือกของแบบสอบถามชนิดเลือกตอบไว้ดังนี้

1. ตัวเลือกที่ใช้ควรเหมาะสม สมเหตุสมผล และสอดคล้องกัน
2. ตัวเลือกทุกตัวต้องสอดคล้องกับคำถาม
3. ให้ตัววางทุกตัวดึงดูดความสนใจของผู้ตอบที่ไม่รู้จริง
4. หลีกเลี่ยงตัวเลือกที่เหมือนกัน
5. ไม่ควรใช้คำว่า "ข้างบนทั้งหมด" และ "ไม่ใช่ข้างบนทั้งหมด"
6. เรียงลำดับตัวเลือกตามลำดับตรรกวิทยา (Logical Order)
7. ไม่เขียนตัวเลือกเป็นลักษณะคล้ายแบบถูก-ผิด

สำหรับตัวเลือกของแบบสอบถามชนิดศาสตร์ซึ่งโดยมากคำตอบมักเป็นจำนวนหรือตัวเลข และสัญลักษณ์ทางคณิตศาสตร์โดยเฉพาะ ควรสร้างตัววางให้เป็นจำนวนหรือตัวเลข และสัญลักษณ์ต่างๆ ที่ได้จากความเข้าใจผิดของนักเรียน² เรียงลำดับคำตอบตามเกณฑ์ เช่น จากมากไปหาน้อย หรือจากน้อยไปหามาก ถ้าต้องการคำตอบในหน่วยมาตราใด ก็ควรกำหนดไว้ให้แน่นอน และระวังอย่าให้ตัวเลือกเหมือนกัน หรือคำถามและคำตอบไปชี้แนะคำตอบให้กับข้อต่อไป นอกจากนี้ควรให้คำตอบถูกกระจายในลักษณะสุ่ม และแต่ละข้อควรมีคำตอบถูกประเภทถูกต้องเพียงคำตอบเดียว (Correct Answer)

¹Robert L. Thorndike, Educational Measurement (Washington D.C.:American Council on Education, 1971), pp. 113-119.

²Georgia S. Adams and Theodore Torgerson, Measurement and Evaluation for the Secondary School Teacher (New York:The Dryden Press, 1957), p. 242.

จากหลักเกณฑ์การสร้างตัวเลือกของแบบสอบชนิดเลือกตอบที่ไคกล่าวมา แสดงว่าการสร้างตัวเลือกที่มีประสิทธิภาพในการดึงดูดนักเรียนให้เลือกตอบนั้น เป็นสิ่งที่ทำได้ไม่มากนัก หากตัวเลือกไม่มีคุณภาพย่อมมีผลกระทบต่อคุณภาพของข้อสอบ คือทำให้ข้อสอบนั้นมีคุณภาพลดลง จึงจำเป็นอย่างยิ่งที่จะต้องสร้างตัวเลือกต่างๆ ให้เป็นไปตามหลักเกณฑ์ นอกจากนี้การสร้างตัวเลือกของข้อสอบชนิดเลือกตอบ ยังขึ้นอยู่กับทักษะของผู้สร้างข้อสอบด้วย ดังที่อีเบล¹ (Ebel) ไคกล่าวว่า ทักษะของผู้สร้างข้อสอบเป็นสิ่งจำเป็นมากในการสร้างข้อสอบชนิดเลือกตอบให้มีคุณภาพ ผู้ที่หัดเขียนข้อสอบใหม่ๆ มักไม่สามารถสร้างตัวเลือกที่ดีได้หมดทุกตัวและต้องใช้เวลานาน้อยในการคิดหาตัวเลือกที่ดีในข้อสอบแต่ละข้อ ดังนั้นการเขียนตัวเลือกของข้อสอบแต่ละข้อให้ได้ครบตามจำนวนที่ต้องการ บางครั้งก็เป็นปัญหาแก่ผู้สร้างข้อสอบเช่นกัน ผู้สร้างข้อสอบจึงแก้ปัญหาเหล่านี้โดย ใช้ตัวเลือกที่มีรูปแบบต่างๆ กันออกไป เช่นแบบปลายเปิด "ไม่มีข้อใดถูก" แบบปลายปิด "ถูกทุกข้อ" เป็นต้น แต่การใช้ตัวเลือกทั้ง 2 แบบนี้ไม่ควรใช้กับข้อสอบที่ให้เลือกคำตอบที่ถูกต้องที่สุด (Best Answer) จากตัวเลือกอื่นที่ถูกต้องน้อยกว่า แต่ควรใช้กับข้อสอบที่มีคำตอบถูกเพียงคำตอบเดียว (Correct Answer)² เนื่องจากอาจเกิดปัญหาคือ ผู้สอบอาจจะไม่เลือกตัวเลือกแบบปลายเปิดเลย เมื่อเห็นว่ามีตัวเลือกใดตัวหนึ่งถูกต้อง หรือผู้สอบอาจจะเลือกตัวเลือกแบบปลายปิด เมื่อพบว่าตัวเลือกทุกตัวถูกต้อง เพียงแต่ถูกมากหรือน้อยกว่ากันเท่านั้น ซึ่งสอดคล้องกับคำกล่าวของอีเบล³ (Ebel) ที่ว่า การใช้ตัวเลือกประเภท "ไม่มีข้อใดถูก" หรือ "ถูกทุกข้อ" ควรใช้กับข้อสอบที่มีคำตอบถูกต้องเพียงคำตอบเดียวจากตัวเลือกทั้งหมดที่ให้มา เช่นในแบบสอบสะกศคำ หรือแบบสอบคณิตศาสตร์ โจทย์ปัญหา เป็นต้น

¹Robert L. Ebel, Measuring Educational Achievement (New Jersey: Engle-Wood Cliffs, Prentice-Hall, Inc., 1965), p. 167.

²Robert L. Thorndike, Educational Measurement, p. 117.

³Robert L. Ebel, Measuring Educational Achievement, p. 159.

การใช้ตัวเลือกแบบปลายเปิด บางทีก็มีความจำเป็น และมีประสิทธิภาพได้เช่นกัน หากรู้จักใช้ให้เหมาะสม โดยเฉพาะอย่างยิ่งควรใช้ในแบบสอบคณิตศาสตร์ เพื่อไม่ให้ผู้สอบเดาคำตอบเพราะการใช้ตัวเลือกแบบนี้เท่ากับเป็นการบังคับให้ผู้สอบคิดคำนวณคำตอบเองทุกข้อ แต่ควรเปิดไว้หลายๆข้อ หรือทุกข้อเลยจะได้ไม่เป็นการแนะนำคำตอบ และควรให้ตัวเลือกประเภทนี้เป็นทั้งคำตอบถูกและเป็นตัวลงในบางข้อสลับกันไป¹ จึงจะสามารถดวงเด็กที่ไม่มี ความรู้จริงคิดว่าตัวเลือกนี้มีโอกาสเป็นข้อถูกได้เหมือนกัน นอกจากนั้นยังเป็นการป้องกันการ ผิดพลาดเช่น โจทย์ผิด² เป็นต้น ส่วนตัวเลือกแบบปลายปิด อาจใช้ในกรณีที่ตัวเลือกทั้งหมดไม่ สามารถแยกออกจากกันได้จริงๆ ว่าข้อไหนถูกมากกว่ากัน ไม่ใช่เป็นตัวเลือกแบบเฉียดๆ แต่ เป็นตัวเลือกที่มีน้ำหนักเท่ากันจริง จึงจำเป็นต้องใช้ "ถูกหมดทุกข้อ" เอาไว้และเป็นคำตอบ ถูกเสียด้วย³ ฉะนั้นในการใช้ตัวเลือกทั้ง 2 แบบนี้ หากใช้ให้เหมาะสมก็จะมีคุณค่ามาก แต่ การสร้างตัวเลือกแบบปลายปิดทำได้ยาก จึงไม่นิยมใช้เท่าตัวเลือกแบบปลายเปิด⁴ และการ ใช้ตัวเลือกทั้ง 2 แบบนี้ควรใช้เมื่อแน่ใจว่ามีความชำนาญในการเขียนข้อสอบพอสมควร หมายความว่าอยู่ในขั้นที่จะอธิบายวิธีเขียนข้อสอบให้ใครๆ ฟังได้ หรือพูดกับคนที่มีความชำนาญ ในการวัดผล⁵ เรื่อง⁵

ตอนที่ 2 เกี่ยวกับคุณลักษณะของแบบสอบที่ดี

เนื่องจากผลการสอบมีความสำคัญยิ่งต่อผู้สอบ จึงมีความจำเป็นที่ผู้สร้างแบบสอบ จะต้องทราบถึงคุณลักษณะของแบบสอบที่ดี เพื่อเป็นแนวทางในการสร้างแบบสอบ และใช้เป็น เกณฑ์ในการวินิจฉัยว่าแบบสอบที่สร้างขึ้นนั้นมีคุณภาพดี-เลว อย่างไร เพื่อจะได้ปรับปรุงแก้ไข

¹ชวาล แพทย์กุล, เทคนิคการวัดผล, หน้า 162-164.

²วิเชียร เกตุสิงห์, หลักการสร้างและวิเคราะห์ข้อสอบ (กรุงเทพมหานคร: มงคลการพิมพ์, 2515), หน้า 65-66.

³ยัง พิทยานิยม, การวัดผลการศึกษา, หน้า 133.

⁴Victor H. Noll, Introduction to Educational Measurement, p.157.

⁵ยัง พิทยานิยม, การวัดผลการศึกษา, หน้า 133.

ให้เป็นแบบสอบที่มีประสิทธิภาพยิ่งขึ้น

พวงแก้ว ปุณยกนก¹ ได้กล่าวถึงคุณลักษณะที่ดีของแบบสอบ ที่สามารถใช้เป็นเกณฑ์สำหรับวินิจฉัยคุณค่าของแบบสอบ และการวัดผลไว้ 10 ประการคือ แบบสอบที่ดีต้องมี

1. ความตรง (Validity) คือแบบสอบจะต้องมีความแม่นยำของคะแนนในการวัดสิ่งที่ต้องการวัด
2. ความเที่ยง (Reliability) คือต้องมีความคงเส้นคงวาของคะแนนสอบแบบสอบที่ดีนั้นใช้สอบซ้ำกับผู้สอบคนเดียวกันควรจะได้คะแนนเท่ากัน
3. ความเป็นปรนัย (Objectivity) คือแบบสอบควรจะไม่ปราศจากความคลุมเครือของคำถามหรือคำตอบ ปราศจากอคติในการให้คะแนน
4. ความสมดุลย์ (Balance) แบบสอบจะมีความตรงตามเนื้อหาได้จะต้องมีความสมดุลย์คือวัดพฤติกรรมที่เกี่ยวข้องกับเนื้อเรื่องที่สอบได้ครอบคลุมทั้งเนื้อหาและวัตถุประสงค์ สักส่วนของคำถามเป็นไปตามรายการตารางเฉพาะ
5. ความยุติธรรม (Fairness) คือผู้สอบทุกคนมีโอกาสที่จะตอบถูก ถ้ามีความรู้ในเรื่องที่ออกข้อสอบนั้นๆ
6. ประสิทธิภาพในการนำไปใช้ (Efficiency) คือแบบสอบจะต้องประหยัดเวลาในการสร้าง เวลาในการดำเนินการสอบ เวลาในการตรวจข้อสอบ ตลอดถึงอุปกรณ์ต่างๆ ในการสร้างข้อสอบ
7. ความยาก (Difficulty) ที่เหมาะสม ซึ่งความยากที่เหมาะสมของแบบสอบคือคะแนนตรงกลางระหว่างคะแนนสูงสุดที่ควรจะได้ กับคะแนนที่ได้โดยโอกาสการเดา

¹พวงแก้ว ปุณยกนก, "ลักษณะเครื่องมือวัดผลที่ดี," เอกสารประกอบการอบรมวิชาวัดผลและประเมินผลการศึกษา หลักสูตรระยะสั้นสำหรับครูอาจารย์โรงเรียนราษฎร์ อาชีวศึกษา กระทรวงศึกษาธิการ ระหว่าง 29 พ.ค.- 9 มิ.ย. 2521 ณ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (อัคราเนา)

8. อำนาจจำแนก (Discrimination Power) คือแบบสอบต้องจำแนกคนเก่งออกจากคนอ่อนได้

9. ความเฉพาะเจาะจง (Specificity) คือผู้ที่มีความรู้ความสามารถเฉพาะเรื่องนั้นๆ จึงจะตอบข้อสอบนั้นได้ ซึ่งวิธีหาความเฉพาะเจาะจงสามารถหาได้โดยให้ผู้มีความสามารถในเรื่องนั้นๆ กลุ่มหนึ่ง ทำแบบสอบนั้นเปรียบเทียบกับคนฉลาดที่ไม่มีความรู้ในเรื่องนั้น ถ้ากลุ่มแรกทำได้ดีกว่ากลุ่มหลัง ถือว่าแบบสอบนั้นมีความเฉพาะเจาะจง

10. ให้เวลาพอเหมาะ (Speededness) คือแบบสอบที่ไม่มีความจำกัดในเรื่องเวลาแม้ว่าจะต่อเวลาในการสอบนานเท่าใด คะแนนของผู้สอบก็คงไม่เพิ่มขึ้นอยู่นั่นเอง ดังนั้นแบบสอบวัดผลสัมฤทธิ์ควรให้เวลาพอเพียงที่เด็กเกินกว่า 90% ของจำนวนเด็กทั้งหมดทำแบบสอบนั้นเสร็จ

ชวาล แพร์ทกุล¹ ได้กล่าวถึงคุณลักษณะที่ดีของแบบสอบไว้ 10 ประการคือแบบสอบที่ดีคือ

1. มีความเที่ยง
2. ยุติธรรม
3. ถามลึก
4. ยั่วเย้า
5. จำเพาะเจาะจง
6. เป็นปรนัย
7. มีประสิทธิภาพ
8. ยากพอเหมาะ
9. มีอำนาจจำแนก
10. มีความตรง

¹ชวาล แพร์ทกุล, เทคนิคการวัดผล, หน้า 169-170.

ในการวิจัยครั้งนี้ผู้วิจัยได้เลือกคุณลักษณะที่คีของแบบสอบถามเพื่อพิจารณา และเปรียบเทียบระหว่างแบบสอบทั้ง 3 ชุด รวม 4 ด้านได้แก่ ความยาก อำนาจจำแนก ความเที่ยง และความตรงของแบบสอบ เนื่องจากเป็นตัวชี้บ่งคุณภาพเชิงการวัด ซึ่งจะขอล่าวถึงรายละเอียดดังต่อไปนี้

ความยากของแบบสอบ

ความยากของข้อสอบเป็นสัดส่วนหรือเปอร์เซ็นต์ของจำนวนผู้สอบที่ตอบข้อสอบนั้นๆ ถูกต้อง ถ้าผู้สอบตอบข้อสอบข้อใดถูกกันมาก แสดงว่าข้อสอบข้อนั้นง่าย ในทางตรงกันข้าม ถ้าผู้สอบตอบข้อสอบข้อนั้นถูกกันน้อยก็แสดงว่าข้อสอบนั้นยาก ฉะนั้นความยากของข้อสอบเป็นสิ่งที่ชี้ให้เห็นว่าข้อสอบมีความยากอยู่ในระดับใด เช่นยากมาก ปานกลาง หรือง่ายมากเท่านั้น ระดับความยากของข้อสอบมีค่าตั้งแต่ 0 ถึง 1.00 ระดับความยากเท่ากับ 0 หมายความว่า ข้อสอบยากมาก ไม่มีผู้ใดทำข้อสอบได้ถูกต้องเลย และระดับความยากเท่ากับ 1.00 หมายความว่าข้อสอบง่ายมาก ผู้สอบทุกคนทำข้อสอบได้ถูกต้อง แต่การกำหนดระดับความยากของข้อสอบขึ้นอยู่กับจุดมุ่งหมายของการสอบ เช่นต้องการวัดว่านักเรียนคนใดบ้างมีความรู้ความสามารถตามที่กำหนดไว้ (การสอบแบบอิงเกณฑ์) หรือต้องการจัดอันดับผลสัมฤทธิ์ทางการเรียนของนักเรียนในกลุ่ม (การสอบแบบอิงกลุ่ม)¹ ในการวัดความสามารถพื้นฐานแบบอิงเกณฑ์ (Criterion-Referenced Mastery Test) ความยากของข้อสอบขึ้นอยู่กับสิ่งที่เรียนรู้ หากเป็นการเรียนรู้ที่ง่าย ข้อสอบควรจะง่าย แต่ถาการเรียนรู้นั้นยากขึ้น ข้อสอบควรจะยากขึ้นด้วย ฉะนั้นการสอบในลักษณะนี้ ข้อสอบที่ง่ายมากหรือยากมาก ถือว่าเป็นข้อสอบที่คีไม่ควรแก้ไขหรือตัดทิ้งไป ส่วนในการสอบแบบอิงกลุ่มเพื่อคู้ว่าผู้ใดมีผลสัมฤทธิ์อยู่ในระดับใด ควรเลือกข้อสอบที่มีความยากประมาณ .50 คือมีจำนวนผู้ตอบถูกเป็นครึ่งหนึ่งของผู้สอบทั้งหมด

¹ Norman E. Gronlund, Measurement and Evaluation in Teaching,

สำหรับข้อสอบที่ยากมากผู้สอบทุกคนทำไม่ได้เลย และข้อสอบที่ง่ายมากผู้สอบทุกคนทำได้หมด จนไม่สามารถจำแนกความแตกต่างของผลสัมฤทธิ์ระหว่างบุคคลได้จึงไม่มีประโยชน์ คุณลักษณะในด้านความยากของข้อสอบเพื่อจำแนกระหว่างบุคคลจึงเป็นสิ่งสำคัญอย่างหนึ่งที่ต้องพิจารณา

เนื่องจากข้อสอบที่มีความยากประมาณ .50 จะทำให้พิสัยของคะแนนจากการสอบกว้างกว่าข้อสอบที่ยากหรือง่ายเกินไป ซึ่งพิสัยของคะแนนกว้างมากเท่าไรค่าความเที่ยงก็จะสูงมากเท่านั้น¹ แต่ข้อสอบที่มีค่าความยาก .50 อาจจะทำให้ค่าอำนาจจำแนกสูงสุดหรือค่าสุกก็ได้ หมายความว่าถ้าจำนวนคนในกลุ่มสูงทั้งหมดตอบถูก และจำนวนคนในกลุ่มต่ำทั้งหมดตอบผิด ค่าความยากของข้อสอบก็เท่ากับ .50 และค่าอำนาจจำแนกเท่ากับ 1.00 หรือจำนวนคนในกลุ่มสูงครึ่งหนึ่งตอบถูก และจำนวนคนในกลุ่มต่ำครึ่งหนึ่งตอบถูก ค่าความยากก็จะเท่ากับ .50 เช่นกัน แต่ค่าอำนาจจำแนกจะเท่ากับ 0 ฉะนั้นข้อสอบที่ดีควรจะมีค่าความยากที่เหมาะสมคือประมาณ .50 และมีค่าอำนาจจำแนกความแตกต่างระหว่างกลุ่มได้ด้วย

องค์ประกอบที่มีอิทธิพลต่อความยากของแบบสอบ

องค์ประกอบที่มีอิทธิพลต่อความยากของแบบสอบมี 2 ประเภท²คือ

1. องค์ประกอบภายใน (Intrinsic Factors) ประกอบด้วย
 - ก. เนื้อหาของข้อสอบแต่ละข้อ (Item Content) ในด้านความซับซ้อน ความเป็นนามธรรม และความแปลกใหม่
 - ข. ลักษณะโครงสร้างของข้อสอบ (Item Structure) หมายถึงวิธีในการแสดงออกซึ่งเนื้อหานั้น

¹Frank B. Womer, Basic Concepts in Testing (Boston:Houghton Mifflin Company, 1968), p. 76.

²Alison C. Campbell, "Some Determinants of the Difficulty of Non-Verbal Classification Item," Educational and Psychological Measurement 21(Winter 1961):899-913.

2. องค์ประกอบภายนอก (Extrinsic Factors) ประกอบด้วย

ก. ความไม่คุ้นเคยต่อเนื้อหา (Unfamiliarity) คืออยู่นอกเหนือประสบการณ์

ของผู้ตอบ

ข. สิ่งที่มีสัมพันธ์กับข้อสอบ (Item Context) เช่น ข้อสอบอื่นที่อยู่ในชุดเดียวกัน

ค. ตัวแปรด้านบุคลิกภาพ ได้แก่ สภาพร่างกาย ลักษณะนิสัย และความตั้งใจ

ของผู้สอบ

นอกจากนี้ธรรมชาติของเนื้อหา พฤติกรรมที่กองการวัด และตัวการที่สลับซับซ้อนอื่นๆ เช่นภาษาที่ใช้ในข้อสอบ รูปแบบของคำถาม และคำชี้แจงต่างๆ ยังเป็นองค์ประกอบที่อาจจะมีอิทธิพลต่อความยากของแบบสอบที่ควรคำนึงถึงด้วย

อำนาจจำแนกของแบบสอบ

อำนาจจำแนกของข้อสอบหมายถึงความสามารถของข้อสอบที่จะช่วยให้เห็นความแตกต่างระหว่างกลุ่มนักเรียนที่มีผลสัมฤทธิ์สูงกับกลุ่มนักเรียนที่มีผลสัมฤทธิ์ต่ำ อำนาจจำแนกของข้อสอบมีค่าตั้งแต่ -1.00 ถึง 1.00 ข้อสอบที่มีค่าอำนาจจำแนกเข้าใกล้ -1.00 หรือ 1.00 แสดงว่ามีอำนาจจำแนกสูง แต่ข้อสอบที่มีคุณค่าตรงตามวัตถุประสงค์ของการสอบคือข้อสอบที่มีค่าอำนาจจำแนกเป็นบวก ซึ่งหมายความว่านักเรียนในกลุ่มสูงตอบข้อสอบได้ถูกต้องมากกว่านักเรียนในกลุ่มต่ำ และแสดงว่าข้อสอบสามารถจำแนกไปในทางเดียวกับคะแนนรวมที่ได้จากการสอบ ส่วนข้อสอบที่มีค่าอำนาจจำแนกเป็นทางลบ ถือว่าจำแนกผิดทิศทาง ไม่ตรงตามวัตถุประสงค์ที่ต้องการ และข้อสอบที่มีค่าอำนาจจำแนกเป็น 0 ถือว่าจำแนกความแตกต่างระหว่างคน 2 กลุ่มไม่ได้เลย

ในการประเมินคุณภาพของข้อสอบโดยอาศัยค่าอำนาจจำแนก ควรถือเกณฑ์ดังนี้¹

<u>ค่าอำนาจจำแนก</u>	<u>คุณภาพของข้อสอบ</u>
.40 ขึ้นไป	ดีมาก
.30 ถึง .39	ดี
.20 ถึง .29	ใช้ได้
ต่ำกว่า .20	ไม่ดีควรตัดทิ้ง

อำนาจจำแนกของข้อสอบเป็นสิ่งสำคัญอีกอย่างหนึ่งที่เราควรคำนึงถึงในการประเมินคุณภาพของข้อสอบ ควบคู่กับความยากของข้อสอบ เนื่องจากข้อสอบที่มีความยากเท่าๆ กันยังต่างกันในเรื่องอำนาจจำแนกได้ด้วย ดังที่ได้กล่าวแล้วว่า ข้อสอบที่มีค่าความยาก .50 อาจจะมีค่าอำนาจจำแนกสูงสุดหรืออาจไม่มีเลยก็ได้ ฉะนั้นข้อสอบที่มีค่าความยาก .50 มีโอกาสที่จะทำให้ค่าความเที่ยงของแบบสอบต่ำได้ เมื่อข้อสอบนั้นมีค่าอำนาจจำแนกเป็น 0 แต่ข้อสอบที่มีค่าอำนาจจำแนกสูง จะมีผลให้ค่าความเที่ยงของแบบสอบสูงด้วยอย่างแน่นอน เนื่องจากค่าอำนาจจำแนกของข้อสอบ เกี่ยวข้องกับส่วนเบี่ยงมาตรฐานของคะแนนซึ่งเป็นองค์ประกอบสำคัญที่มีอิทธิพลต่อค่าความเที่ยงของแบบสอบ²

องค์ประกอบที่มีอิทธิพลต่ออำนาจจำแนกของแบบสอบ

องค์ประกอบที่สำคัญที่มีอิทธิพลต่ออำนาจจำแนกของแบบสอบได้แก่

1. คุณภาพของตัวलगและจำนวนตัวเลือก ข้อสอบที่สร้างตัวलगได้ก็มีความใกล้เคียงกับตัวเลือกที่เป็นคำตอบถูกหลายๆ ตัว จะทำให้จำนวนตัวเลือกเพิ่มขึ้น ซึ่งจะทำให้อำนาจจำแนกของข้อสอบสูงขึ้นด้วย³ เนื่องจากข้อสอบแบบนี้ทำให้ผู้สอบที่มีความรู้จริงเท่านั้น

¹ Robert L. Ebel, Measuring Educational Achievement, p. 364.

² Ibid., p. 302.

³ Ibid., p. 165.

สามารถทำข้อสอบได้ถูกต้อง แต่ผู้ที่ไม่มีความรู้จริงจะเลือกตอบตัวดวงที่มีความใกล้เคียงกับ คำตอบถูกกระจายกันออกไปทุกตัว

2. ความยากของข้อสอบ การจำแนกความแตกต่างของผลสัมฤทธิ์ระหว่างกลุ่ม ขึ้นอยู่กับความยากของข้อสอบ¹ คือข้อสอบที่มีความยาก .50 มีโอกาสทำให้ค่าอำนาจจำแนก สูงสุดได้ เมื่อนักเรียนในกลุ่มสูงทุกคนตอบข้อสอบได้ถูกต้อง ในขณะที่นักเรียนในกลุ่มต่ำทั้งหมด ตอบข้อสอบผิด และหากข้อสอบมีค่าความยากสูงหรือต่ำกว่านี้ ค่าอำนาจจำแนกของข้อสอบก็จะ ต่ำลง² ซึ่งหมายความว่าข้อสอบที่ง่ายหรือยากเกินไป จะเป็นข้อสอบที่มีค่าอำนาจจำแนกค่า นั้นเอง

ความเที่ยงของแบบสอบ

ความเที่ยง หมายถึงความคงที่ของคะแนนที่ได้จากการ สอบบุคคลกลุ่มเดียวกันด้วย แบบสอบชุดเดียวกัน ในเวลาที่ต่างกัน หรือสอบด้วยแบบสอบคนละชุดที่เทียบเท่ากัน หรือ ภายใต้สภาพการสอบที่ต่างกัน³

ความเที่ยง หมายถึงความคงที่ของผลที่ได้รับจากการใช้เครื่องมือชนิดเดียวกัน สอบมากกว่า 1 ครั้ง⁴

จากที่กล่าวมานี้ แสดงว่าความเที่ยงของแบบสอบ เป็นคุณสมบัติอย่างหนึ่งของ แบบสอบที่ทำให้คะแนนที่ได้จากการ สอบบุคคลกลุ่มใดกลุ่มหนึ่ง ในเวลาต่างกัน ด้วยแบบสอบ

¹Norman E. Gronlund, Measurement and Evaluation in Teaching, p. 154.

²Ibid., p. 270.

³แอน อนาสตาซี, การตรวจสอบเชิงจิตวิทยา, แปลโดย ประชุมสุข อาชวอำรุง และคณะ (กรุงเทพมหานคร:ไทยวัฒนาพานิช, 2519), หน้า 73.

⁴Norman E. Gronlund, Readings in Measurement and Evaluation (New York: The Macmillan Company, 1968), p. 163.

ชุดนั้นๆ มีความคงที่ไม่เปลี่ยนแปลง หมายความว่าผู้สอบที่ได้คะแนนสูงจากการสอบครั้งแรก จะได้คะแนนสูงในการสอบครั้งต่อๆ ไปด้วย และผู้สอบที่ได้คะแนนต่ำจากการสอบครั้งแรกก็จะ ได้คะแนนต่ำในการสอบครั้งต่อๆ ไปเช่นกัน แต่ในทางปฏิบัติจริง มักจะมีตัวแปรอื่นๆ มาเกี่ยว ข้อง เช่น การเรียนรู้จากการสอบครั้งแรก พัฒนาการในตัวผู้สอบ เป็นต้น ซึ่งตัวแปรต่างๆ นี้ ทำให้คะแนนในการสอบแต่ละครั้งมีการเปลี่ยนแปลงบ้าง อย่างไรก็ตามคะแนนสอบเหล่านั้น จะมีความสัมพันธ์กัน หากแบบสอบมีความเที่ยง

ทฤษฎีของความเที่ยง

ในการวัดทางค่านิจภาพ ถึงแม้เครื่องมือที่ใช้จะมีความเที่ยงสูง ก็ไม่มีเครื่องมือ ชนิดใดที่วัดได้ถูกต้องแม่นยำโดยสมบูรณ์ ฉะนั้นคะแนนจากการสอบที่นักเรียนทำได้ (Observed Scores) จะประกอบด้วยส่วนที่เป็นคะแนนจริง (True Scores) และส่วนของคะแนนจาก ความคลาดเคลื่อน (Error Scores) ซึ่งเขียนเป็นสมการ¹ ดังนี้

$$X_t = X_{\infty} + X_e$$

เมื่อ X_t แทนคะแนนที่ได้จากการวัด
 X_{∞} แทนคะแนนจริง
 X_e แทนคะแนนความคลาดเคลื่อน

คะแนนจริง คือคะแนนเฉลี่ยที่ผู้สอบได้รับจากการทำแบบสอบเดิมหลายๆ ครั้งโดย ปราศจากอิทธิพลของการเรียนรู้จากการสอบครั้งแรก แรงจูงใจ หรือสภาพทางอารมณ์ เป็นต้น

คะแนนความคลาดเคลื่อน คือคะแนนที่เกิดจากความผิดพลาดในการวัด ซึ่งแบ่งเป็น 2 ประเภท คือความคลาดเคลื่อนที่เกิดจากการสุ่ม (Random Error) กับความคลาดเคลื่อน

¹Joy Paul Guilford and Benjamin Fruchter, Fundamental Statistics in Psychology and Education, 6th ed. (Tokyo:McGraw-Hill Kogakusha, 1978), p. 408.

อย่างมีระบบ (Systematic Error) ซึ่งความคลาดเคลื่อนประเภทหลังนี้ไม่มีผลต่อค่าความเที่ยง แต่ความคลาดเคลื่อนที่เกิดจากการสุ่ม เช่น เกิดจาก ตัวข้อสอบ การดำเนินการสอบ การตรวจให้คะแนน และจากตัวผู้สอบเอง ซึ่งเป็นไปได้ทั้งทางบวกและทางลบ และจะหักล้างกันหมด หากคะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง หรือขนาดของคะแนนความคลาดเคลื่อนไม่เกี่ยวข้องกับขนาดของคะแนนจริงแล้ว ความแปรปรวนของคะแนนที่วัดได้ จะมีค่าเท่ากับผลบวกของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนความคลาดเคลื่อน¹ ซึ่งเขียนเป็นสมการ² ดังนี้

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

เมื่อ s_t^2 แทนความแปรปรวนของคะแนนที่วัดได้

s_∞^2 แทนความแปรปรวนของคะแนนจริง

s_e^2 แทนความแปรปรวนของคะแนนความคลาดเคลื่อน

ความเที่ยง หมายถึงอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่วัดได้ เขียนเป็นสมการ³ ได้ว่า

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

ฉะนั้นจากสมการ (1) และ (2) จะได้สมการพื้นฐานในการหาสัมประสิทธิ์สหสัมพันธ์ของความเที่ยง⁴ ดังนี้

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

¹ Robert L. Ebel, Measuring Educational Achievement, p. 331.

² Joy Paul Guilford and Benjamin Fruchter, Fundamental Statistics in Psychology and Education, p. 410.

³ Ibid.

⁴ Ibid.

แบบของความเที่ยง

ความเที่ยงแบ่งออกเป็น 5 แบบ¹คือ

1. ความเที่ยงแบบสอบซ้ำ (Test-Retest Reliability)
2. ความเที่ยงแบบสลับฟอร์ม (Alternate-Form Reliability)
3. ความเที่ยงชนิดแบ่งครึ่ง (Split-Half Reliability)
4. ความเที่ยงแบบคูเคอร์-ริชาร์ดสัน (Kuder-Richardson Reliability)
5. ความเที่ยงของผู้ให้คะแนน (Scorer Reliability)

ในที่นี้จะกล่าวถึงเฉพาะความเที่ยงแบบคูเคอร์-ริชาร์ดสัน ซึ่งเป็นแบบของความเที่ยงที่ใช้ในการวิจัยครั้งนี้

ความเที่ยงแบบคูเคอร์-ริชาร์ดสัน

ความเที่ยงแบบคูเคอร์-ริชาร์ดสัน เป็นการหาความสอดคล้องกันระหว่างข้อสอบ โดยดำเนินการสอบครั้งเดียวและใช้แบบสอบชุดเดียว ซึ่งความสอดคล้องกันระหว่างข้อนี้ได้รับอิทธิพลจากแหล่งความแปรปรวนคลาดเคลื่อน 2 ชนิด² คือ

1. เนื้อหาที่สุม
2. ความเป็นวิวิธพันธ์ของพฤติกรรมที่สุม ข้อสอบที่มีความเป็นเอกพันธ์มากขึ้น ความสอดคล้องกันระหว่างข้อก็ยิ่งสูงขึ้น

¹แอน อนาสตาซี, การตรวจสอบเชิงจิตวิทยา, หน้า 79-86.

²เรื่องเดียวกัน, หน้า 84.

วิธีหาความเที่ยงแบบคูเคอร์-ริชาร์คสัน

ในปี 1937 คูเคอร์และริชาร์คสัน (Kuder&Richardson) ได้เสนอสูตรสำหรับประมาณค่าความเที่ยงแบบนี้ คือสูตรคูเคอร์-ริชาร์คสัน 20 (Kuder-Richardson Formula 20)¹

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

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

ค่าความเที่ยงแบบคูเคอร์-ริชาร์คสัน ประมาณได้กับ

1. ค่าสหสัมพันธ์เฉลี่ยระหว่างข้อสอบ k ข้อในแบบสอบ
2. ค่าสัมประสิทธิ์ความเที่ยงของข้อสอบในแบบสอบ
3. ค่าเฉลี่ยของสัมประสิทธิ์ความเที่ยงที่ประมาณค่าจากสูตรสเปียร์แมน บราวน์ (Spearman-Brown Formula) ซึ่งแบ่งครึ่งข้อสอบด้วยวิธีต่างๆ²

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

²Julian C. Stanley and Kenneth D. Hopkins, Educational and Psychological Measurement and Evaluation, 5th ed. (Englewood Cliff:Prentice-Hall, 1972), p. 191.

ของผู้สอบในกลุ่มใกล้เคียงกันมาก (กลุ่มเอกพันธ์) การกระจายของคะแนนจะน้อย ทำให้ความเที่ยงมีค่าต่ำ

3. ความยากของแบบสอบ (Difficulty of Test) ค่าความเที่ยงของแบบสอบ เกี่ยวข้องกับความแปรปรวนของคะแนนสอบที่ได้ ข้อสอบที่ยากมากหรือข้อสอบที่ง่ายเกินไปจะทำให้ความแปรปรวนของคะแนนน้อย ความเที่ยงจะมีค่าต่ำ ฉะนั้นหากความยากของข้อสอบเหมาะสมที่จะทำให้ความแปรปรวนของคะแนนมากขึ้น ค่าความเที่ยงก็จะสูงขึ้นด้วย

4. ความเป็นปรนัย (Objectivity) แบบสอบที่มีความเป็นปรนัยสูง เช่น ปราศจากอคติของผู้ตรวจให้คะแนน ภาษาที่ใช้ในแบบสอบถูกต้อง ชัดเจน ทำให้ผู้สอบเข้าใจคำถามและสิ่งต่างๆ ในแบบสอบถูกต้องตรงกัน เป็นต้น ทำให้ผลจากการสอบตรงตามความสามารถในสิ่งที่ต้องการวัดที่ผู้สอบมีอยู่จริง การวัดนั้นก็就会有ความคงที่ ทำให้ความเที่ยงของแบบสอบมีค่าสูง

นอกจากนี้ยังมีองค์ประกอบอื่นๆ ที่มีอิทธิพลต่อค่าความเที่ยงของแบบสอบ เช่น การบริหารแบบสอบ องค์ประกอบเกี่ยวกับตัวผู้สอบ และวิธีการประมาณค่าความเที่ยง โดยเฉพาะในแบบสอบความเร็ว (Speed Test) หากประมาณค่าความเที่ยงของแบบสอบประเภทนี้ด้วยวิธีความคงที่ภายใน จะทำให้ค่าความเที่ยงที่ได้นั้นสูงกว่าที่ควรจะเป็น

ความตรงของแบบสอบ

ความตรง หมายถึงความถูกต้องแน่นอนในการวัดสิ่งที่ต้องการจะวัด¹

ความตรง หมายถึงคุณสมบัติที่แบบสอบสามารถวัดในสิ่งที่ต้องการจะวัด²

¹ Everet F. Lindquist, A First Course in Statistics (Boston: Houghton Mifflin Company, 1942), p. 213.

² Robert L. Ebel, Measuring Educational Achievement, p. 468.

จากความหมายของความตรงดังกล่าว สรุปได้ว่า ความตรงหมายถึงความแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด หรือสิ่งที่เครื่องมือควรจะวัด เช่นแบบสอบคณิตศาสตร์ ต้องการวัดความสามารถทางคณิตศาสตร์ ถ้ามีการให้คะแนนความถูกต้องในการสะกดคำ คะแนนที่ได้ย่อมขาดความตรง คือคะแนนให้ความหมายไม่ตรงจุดมุ่งหมายของการวัดนั่นเอง

ชนิดของความตรง

ความตรงของแบบสอบจำแนกได้เป็น 3 ชนิดใหญ่ๆ¹ คือ

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

การวิจัยครั้งนี้เกี่ยวข้องกับความตรงตามเนื้อหา และความตรงร่วมสมัย ซึ่งเป็นแบบหนึ่งของความตรงตามเกณฑ์สัมพันธ์ ดังจะกล่าวถึงรายละเอียดดังนี้

ความตรงตามเนื้อหา (Content Validity)

ความตรงตามเนื้อหา หมายความว่าแบบสอบฉบับนั้นมีคำถามสอดคล้องตรงตามเนื้อหาวิชาตามที่ระบุไว้ในหลักสูตร (ตัวเกณฑ์) เพียงใด สามารถวัดความรู้ตรงตามความตั้งใจที่จะวัดหรือไม่ เช่นสอนเรื่องเศษส่วนแล้วออกโจทย์ถามเรื่องเศษส่วนจริงๆ หรือไม่ หรือตั้งใจจะวัดความรู้ในเชิงวิชาการ แต่เขียนคำถามกลายเป็นวัดความถนัดไปหรือไม่ และข้อสอบแต่ละวิชานั้นๆ ถามครอบคลุมทั่วเนื้อหาตามที่ได้สอน ได้กำหนดไว้หรือไม่ เป็นต้น²

¹ Norman E. Gronlund, Measurement and Evaluation in Teaching, p. 81.

²ทบวงมหาวิทยาลัยของรัฐ และสำนักงานคณะกรรมการการศึกษาแห่งชาติ, การทดสอบสัมฤทธิ์ผล (กรุงเทพมหานคร : สำนักงานคณะกรรมการการศึกษาแห่งชาติ สำนักนายกรัฐมนตรี, 2518), หน้า 9-10.

การตัดสินว่าแบบสอบมีความตรงตามเนื้อหาเล็กน้อยเพียงใด ไม่สามารถคำนวณเป็นค่าตัวเลขได้ โดยทั่วไปมักจะพิจารณาข้อสอบเทียบกับหัวข้อเนื้อหาวิชา (Subject Matter) และชนิดของพฤติกรรมที่ต้องการจะวัด ที่กำหนดไว้ในตารางวิเคราะห์หลักสูตรหรือผังการสร้างข้อสอบ (Table of Specification or Test Blueprint) ของแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ซึ่งเป็นตาราง 2 มิติ (Two Way Dimensions) ที่จะช่วยให้การสุ่มหัวข้อเนื้อหาวิชา และชนิดของพฤติกรรมที่จะนำมาสร้างข้อสอบ เป็นตัวแทนที่ดีของการถามเนื้อหาวิชาและพฤติกรรมที่มีอยู่ทั้งหมด¹ และผู้ที่ตัดสินความตรงตามเนื้อหาได้ดีที่สุดคือ ผู้สอนหรือวิทยากรผู้ให้การอบรม หรือผู้เชี่ยวชาญในสาขาวิชานั้นๆ²

ความตรงร่วมสมัย (Concurrent Validity)

ความตรงร่วมสมัยของแบบสอบ หมายความว่าแบบสอบสามารถช่วยในการคาดคะเนสภาพที่แท้จริงของผู้สอบในปัจจุบันได้ถูกต้องเพียงใด³ การหาความตรงร่วมสมัยของแบบสอบทำได้โดยการหาความสัมพันธ์ระหว่างคะแนนจากการสอบด้วยแบบสอบฉบับนั้น กับเกณฑ์ที่เกิดในเวลาเดียวกัน หรือใกล้เคียงกันว่าสอดคล้องกันเพียงใด หรือโดยการตรวจดูว่าแบบสอบฉบับนั้นสามารถให้คะแนนกระจายมากน้อยเพียงใด และคะแนนเหล่านั้นสอดคล้องกับความสามารถที่แท้จริงของผู้สอบที่ปรากฏในปัจจุบันหรือไม่ หากแบบสอบใดสามารถให้คะแนน

¹อนันต์ ศรีโสภาก, การวัดและการประเมินผลการศึกษา (กรุงเทพมหานคร: ไทยวัฒนาพานิช, 2520), หน้า 69-70.

²สมหวัง พิธิยานุวัฒน์, "เทคนิคการวิเคราะห์คุณภาพของข้อสอบ," ใน การพัฒนาการเรียนการสอน (กรุงเทพมหานคร: คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์, 2523), หน้า 146. (อักษรสำนวน)

³ทบวงมหาวิทยาลัยของรัฐ และ สำนักงานคณะกรรมการการศึกษาแห่งชาติ, การทดสอบสัมฤทธิ์ผล, หน้า 11.

กระจายมาก ถือว่าแบบสอบนั้นมีความตรงร่วมสมัยได้ นอกจากนี้ยังสามารถหาความตรงร่วมสมัยได้โดยใช้สภาพความจริงในปัจจุบันที่สังเกตเห็นเป็นตัวเกณฑ์ คือการนำคะแนนของผู้สอบแต่ละคนมาจัดอันดับ และเปรียบเทียบกับอันดับความสามารถที่แท้จริง ว่าสอดคล้องกันมากน้อยเพียงใด¹

องค์ประกอบที่มีอิทธิพลต่อความตรงของแบบสอบ

กรอนลันด์² (Gronlund) ได้กล่าวถึงองค์ประกอบที่มีอิทธิพลต่อความตรงของแบบสอบไว้ดังนี้

1. องค์ประกอบภายในแบบสอบ ได้แก่

- 1.1 ความไม่ชัดเจนของคำชี้แจง คำชี้แจงที่ไม่ชัดเจนเกี่ยวกับวิธีตอบข้อสอบ การบันทึกคำตอบ เป็นต้น จะทำให้ค่าความตรงลดลง
- 1.2 โครงสร้างของคำและประโยคยากเกินไป ทำให้นักเรียนสับสน แทนที่จะวัดได้ตรงกับวัตถุประสงค์ที่ต้องการ ก็จะกลายเป็นการวัดความสามารถทางภาษา
- 1.3 ระดับความยากที่ไม่เหมาะสมของข้อสอบ ทำให้ไม่สามารถจำแนกนักเรียนได้ จะทำให้ความตรงลดลง
- 1.4 ข้อสอบที่สร้างไม่คืบข้างข้อเป็นการแนะนำคำตอบให้นักเรียน
- 1.5 ความกำกวมของภาษาที่ใช้ในแบบสอบทำให้นักเรียนสับสน
- 1.6 ความไม่เหมาะสมของแบบสอบที่จะวัดสิ่งที่ต้องการวัด
- 1.7 แบบสอบสั้นเกินไป ไม่สามารถวัดได้ครอบคลุมพฤติกรรมที่ต้องการวัด
- 1.8 การเรียงข้อสอบที่ไม่เหมาะสม ข้อสอบที่เรียงจากยากไปหาง่าย ทำให้

¹ทบวงมหาวิทยาลัยของรัฐ และสำนักงานคณะกรรมการการศึกษาแห่งชาติ, การทดสอบสัมฤทธิ์ผล, หน้า 11.

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

นักเรียนใช้เวลาคิดหาคำตอบนาน อาจจะไม่มีความเข้าใจข้อสอบที่ง่าย ๆ ทนหลัง การเรียงข้อสอบแบบนี้มีผลต่อความตรงโดยมีผลต่อแรงจูงใจในการทำแบบสอบของนักเรียน

- 1.9 การเรียงคำตอบอย่างเป็นระบบ จะช่วยให้นักเรียนเดาคำตอบได้ง่าย
2. เนื้อหาและวิธีดำเนินการสอน การสอนในเนื้อหาที่ตรงกับเนื้อหาที่ข้องการวัดในข้อสอบ จะทำให้เครื่องมือวัดความตรงในการวัดขบวนการทางสมองที่ซับซ้อนกว่า
3. การบริหารแบบสอบและการให้คะแนน เช่นเวลาในการสอบน้อยเกินไป การชี้แนะนักเรียนบางคนในการตอบ การให้คะแนนที่ไม่คงที่ เหล่านี้ทำให้ค่าความตรงลดลงได้
4. องค์ประกอบในการตอบแบบสอบของนักเรียนเอง รูปแบบการตอบแบบสอบของนักเรียนแต่ละคนแตกต่างกันออกไป ซึ่งทำให้ผลการสอบไม่ตรงกับความสามารถที่แท้จริง ตลอดจนถึงความวิตกกังวล ความเปลี่ยนแปลงทางอารมณ์ เป็นต้น ทำให้ค่าความตรงลดลง
5. ชรรวมชาติของกลุ่มและเกณฑ์ที่ใช้ ตามปกติความตรงของแบบสอบจะมีค่าสูงเฉพาะในบางกลุ่ม เนื่องจากในการสอบวันนั้น มีอิทธิพลจากองค์ประกอบอื่น เช่น อายุ เพศ ระดับความสามารถ ภูมิหลังทางการศึกษา และภูมิหลังทางวัฒนธรรม เป็นต้น ดังนั้นในการหาความตรงควรคำนึงถึงกลุ่มและเกณฑ์ที่ใช้ให้เหมาะสมด้วย

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

การวิจัยเกี่ยวกับผลของการใช้รูปแบบตัวเลือกแบบต่างๆ ที่มีต่อคุณภาพของแบบสอบชนิดเลือกตอบนี้ ในต่างประเทศได้มีผู้ทำการวิจัยไว้เท่าที่ค้นพบดังต่อไปนี้

ริมแลนค¹(Rimland) ได้ศึกษาเกี่ยวกับการใช้ตัวเลือกแบบปลายเปิด ในวิชาคณิตศาสตร์ โดยศึกษากับกลุ่มตัวอย่าง 2 กลุ่ม ๆ ละ 3,600 คน กลุ่มที่ 1 ทำแบบสอบที่ใช้

¹Bernard Rimland, "The Effects of Varying Time Limits and of Using 'Right Answer Not Given' in Experimental Forms of The U.S. Navy Arithmetic Test," Educational and Psychological Measurement 20 (Autumn 1960):533-538.

ตัวเลือกแบบธรรมดา กลุ่มที่ 2 ทำแบบสอบที่ใช้ตัวเลือกแบบปลายเปิด โดยให้ 1 ใน 5 ของจำนวนข้อสอบในแบบสอบมีตัวเลือกแบบปลายเปิดเป็นคำตอบถูก ผลการศึกษาครั้งนี้พบว่า การใช้ตัวเลือกแบบปลายเปิดนี้ไม่ทำให้คุณภาพของแบบสอบในด้านความยาก และอำนาจจำแนก แตกต่างจากแบบสอบที่ใช้ตัวเลือกแบบธรรมดา

ฮิวส์และทริมเบิล¹ (Hughes and Trimple) ได้ทำการวิจัยเกี่ยวกับคุณภาพของแบบสอบเลือกตอบ ที่ใช้ตัวเลือกแบบธรรมดา, แบบปลายเปิดและแบบปลายปิด และแบบผสม "ถูกทั้งข้อ 1 และข้อ 2" หรือ "ผิดทั้งข้อ 1 และข้อ 2" ในวิชาจิตวิทยาทั่วไป โดยศึกษากับนักศึกษาปริญญาตรีปีที่ 1 จำนวน 63 คน โดยแบ่งเป็น 3 กลุ่ม กลุ่มที่ 1 ทำแบบสอบที่ใช้ตัวเลือกแบบธรรมดา กลุ่มที่ 2 ทำแบบสอบที่ใช้ตัวเลือกแบบปลายเปิดและแบบปลายปิด และกลุ่มที่ 3 ทำแบบสอบที่ใช้ตัวเลือกแบบผสม โดยแบบสอบแต่ละชุดประกอบด้วย ข้อสอบจำนวน 55 ข้อ แบ่งเป็น 3 ส่วน คือส่วนแรกเป็นข้อสอบควบคุม ได้แก่ข้อสอบที่มีตัวเลือกแบบธรรมดาเหมือนกันทั้ง 3 ชุด จำนวน 15 ข้อ ส่วนที่ 2 เป็นข้อสอบที่ใช้เปรียบเทียบจริง ได้แก่ข้อสอบที่ใช้ตัวเลือกสุดท้ายเป็นแบบธรรมดา หรือแบบปลายเปิดและปลายปิด หรือแบบผสม ต่างกันในแต่ละชุดจำนวน 35 ข้อ และส่วนที่ 3 เป็นข้อสอบที่ใช้ตัวเลือกแบบต่างๆ เป็นคำตอบถูกจำนวน 5 ข้อ ได้ผลการศึกษาว่า รูปแบบของตัวเลือกแบบต่างๆ มีผลต่อความยากของแบบสอบ แต่ไม่มีผลต่ออำนาจจำแนกของแบบสอบ และเมื่อเปรียบเทียบผลการสอบของนักเรียนทั้ง 3 กลุ่ม ปรากฏว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยคะแนนจากการสอบของนักเรียนที่ทำแบบสอบที่ใช้ตัวเลือกแบบธรรมดาสูงที่สุด

หน้า 32

¹ Herbert H. Hughes and W. Eugene Trimple, "The Use of Complex Alternatives in Multiple Choice Items," Educational and Psychological Measurement 25 (Spring 1965):117-125.

มุลเลอร์¹ (Mueller) ได้ศึกษาในลักษณะเดียวกันกับ ฮิวส์และทริมเบิล โดยใช้กลุ่มตัวอย่างที่เป็นนักศึกษาในมหาวิทยาลัยอินเดียนา จำนวน 2,000 คน ในวิชาเกี่ยวกับการประกอบอาชีพเป็นพนักงานขายของ ปรากฏว่าผลการวิจัยที่ได้สอดคล้องกับการวิจัยของ ฮิวส์และทริมเบิล คือพบว่ารูปแบบของตัวเลือกมีผลต่อค่าความยากของข้อสอบเช่นกัน โดยข้อสอบที่ใช้ตัวเลือกแบบผสมยากที่สุด รองลงมาคือข้อสอบที่ใช้ตัวเลือกแบบปลายเปิด และข้อสอบที่ใช้ตัวเลือกแบบธรรมดาที่ง่ายที่สุด แต่ไม่มีผลต่อค่าอำนาจจำแนกของแบบสอบ นอกจากนี้ยังได้เปรียบเทียบเมื่อตัวเลือกแบบต่างๆ เป็นคำตอบถูกด้วย พบว่าข้อสอบที่ใช้ตัวเลือกแบบผสมยากที่สุด รองลงมาคือข้อสอบที่ใช้ตัวเลือกแบบปลายเปิด แต่ข้อสอบที่ง่ายที่สุดคือข้อสอบที่ใช้ตัวเลือกแบบปลายปิด และผลการพิจารณาประสิทธิภาพของตัวลองปรากฏว่าตัวลองแบบผสมมีประสิทธิภาพมากที่สุด รองลงมาคือแบบปลายเปิดและตัวลองที่มีประสิทธิภาพน้อยที่สุดคือตัวลองแบบธรรมดา

สำหรับในประเทศไทย มีงานวิจัยเกี่ยวกับรูปแบบของตัวเลือกแบบต่างๆ นี้เพียงเรื่องเดียวคืองานวิจัยของ นวลน้อย แต่บรรพกุล² ที่ได้ศึกษาถึงสมรรถนะในการวัดของแบบสอบชนิดเลือกตอบที่ใช้รูปแบบของตัวเลือกแบบต่างๆ ในวิชาภาษาไทย ระดับประถมศึกษาปีที่ 7 โดยแบ่งแบบสอบออกเป็น 4 ชุด คือแบบสอบที่ใช้ตัวเลือกแบบธรรมดา แบบปลายเปิด แบบปลายปิด และแบบผสม ซึ่งได้แก่ "ถูกทั้งข้อ ก และข้อ ข" หรือ "ถูกทั้งข้อ ก และข้อ ค"

¹ Daniel J. Mueller, "An Assessment of the Effectiveness of Complex Alternatives in Multiple Choice Achievement Test Items," Educational and Psychological Measurement 35 (Spring 1975):135-141.

² นวลน้อย แต่บรรพกุล, "การศึกษาสมรรถนะในการวัดของแบบสอบชนิดเลือกตอบที่มีตัวเลือกซับซ้อนแตกต่างกันในวิชาภาษาไทย ระดับชั้นประถมศึกษาปีที่ 7 จังหวัดภูเก็ต" (ปริญา-นิพนธ์การศึกษามหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2520).

เป็นต้น ซึ่งแบบสอบแต่ละชุดยังแบ่งย่อยตามเนื้อหาเป็น 3 ฉบับคือ แบบสอบสะกดคำ หลักภาษา และความเข้าใจในการอ่าน และแบบสอบแต่ละชุดประกอบด้วยข้อสอบจำนวน 35 ข้อ แบ่งเป็นข้อสอบที่ใช้เปรียบเทียบจริงจำนวน 25 ข้อ ซึ่งใช้ตัวเลือกแบบต่างๆ เป็นตัวลวงตัวสุดท้าย และอีกส่วนหนึ่งเป็นข้อสอบควบคุมหมายถึงข้อสอบที่มีตัวเลือกธรรมดา 5 ตัวเลือก ซึ่งเป็นข้อสอบที่ง่ายๆ เพื่อไม่ให้มีผลต่อข้อสอบที่จะใช้เปรียบเทียบจริง ผลการวิจัยครั้งนี้ปรากฏว่ารูปแบบของตัวเลือกไม่มีผลต่ออ่านจจำแนก และความเที่ยงของแบบสอบ แต่มีผลต่อความตรงของแบบสอบทุกฉบับ ยกเว้นแบบสอบสะกดคำ สำหรับความยากของแบบสอบ แตกต่างกันเฉพาะในแบบสอบความเข้าใจในการอ่านเท่านั้น จากผลการวิจัยนี้ หากพิจารณาเฉพาะตัวเลือกแบบธรรมดากับแบบปลายเปิด จะพบว่าไม่แตกต่างกันทั้งในด้านความยาก อ่านจจำแนก ความเที่ยง ความตรง และคะแนนผลการสอบ

จากงานวิจัยที่เกี่ยวข้องดังกล่าวมาทั้งหมด จะเห็นว่าวัตถุประสงค์ของการวิจัยใกล้เคียงกันมาก แต่ผลที่ได้รับจากการวิจัยบางส่วนยังไม่สอดคล้องกัน ทั้งนี้อาจจะเนื่องจากความแตกต่างในด้านเนื้อหาวิชาที่ใช้ศึกษา ระดับการศึกษาของกลุ่มตัวอย่าง และสัดส่วนในการกำหนดให้ตัวเลือกแบบต่างๆ เป็นคำตอบถูกหรือเป็นตัวลวงในแบบสอบที่ใช้เป็นเครื่องมือในการวิจัย แต่มีข้อน่าสังเกตว่างานวิจัยที่กล่าวถึงนี้ ไม่มีงานวิจัยใดที่ศึกษาในเนื้อหาวิชาเดียวกัน และงานวิจัยที่มีลักษณะโครงสร้างของเครื่องมือในการวิจัยคล้ายคลึงกันมากซึ่งได้แก่ งานวิจัยของฮิวส์และทริมเบิล กับงานวิจัยของนวลน้อย แต่บรรพกุล ยังให้ผลการวิจัยบางส่วนที่ขัดแย้งกัน แต่งานวิจัยที่ศึกษาในระดับการศึกษาเดียวกันคือในระดับปริญญา ได้แก่งานวิจัยของฮิวส์และทริมเบิล กับงานวิจัยของมุลเลอร์ กลับให้ผลการวิจัยที่สอดคล้องกัน จึงทำให้ไม่สามารถสรุปผลที่แน่นอนได้ว่ารูปแบบของตัวเลือกของแบบสอบชนิดเลือกตอบจะมีผลต่อคุณภาพของแบบสอบหรือไม่ และมีตัวแปรใดบ้างที่จะส่งผลกระทบต่อคุณภาพของแบบสอบร่วมกับรูปแบบของตัวเลือกแบบต่างๆ ซึ่งควรจะได้ทำการศึกษาต่อไป