



บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ดั่งนี้
ในบทนี้ผู้วิจัยได้นำเสนอวรรณคดีที่เกี่ยวข้องในการวิจัย โดยแบ่งออกเป็น 3 ตอน

- ตอนที่ 1 มโนทัศน์เกี่ยวกับการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบ
- ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการสอบ
 - 2.1 ทฤษฎีคลาสสิกคอลล
 - 2.2 ทฤษฎีการตอบสนองข้อสอบ
 - 2.3 ความแตกต่างระหว่างทฤษฎีคลาสสิกคอลลกับทฤษฎีการตอบสนองข้อสอบ
- ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง
 - 3.1 ตามแนวของทฤษฎีคลาสสิกคอลล
 - 3.2 ตามแนวของทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 1 มโนทัศน์เกี่ยวกับการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบ

ในการวัดผลสัมฤทธิ์ทางการเรียนนั้น เครื่องมือที่ใช้ในการวัดคือ แบบสอบ ซึ่งแบบสอบที่ได้รับการนิยมนำใช้มากที่สุดคือ แบบสอบชนิดเลือกตอบ ในการสร้างแบบสอบชนิดเลือกตอบ ถือว่าการสร้างตัวเลือกเป็นสิ่งสำคัญที่สุดอย่างหนึ่ง ทั้งนี้เนื่องจากคุณภาพของข้อสอบชนิดนี้จะมีมากน้อยเพียงใด ย่อมขึ้นอยู่กับ การเขียนตัวเลือกเป็นสำคัญ เพราะตัวลวงที่ดีนั้นช่วยให้ข้อสอบมีคุณภาพดี (Georgia S.Adam 1964:340) ถ้าตัวลวงนั้นไม่สามารถลวงผู้สอบได้แล้ว นอกจากจะทำให้เสียเวลาในการออกข้อสอบ ยังทำให้ค่าความเที่ยงและความตรงของแบบสอบลดลงอีกด้วย (Jon C.Marshall Loyde 1972:62) ฉะนั้นในการสร้างข้อสอบชนิดนี้ จำเป็นต้องใช้ตัวเลือกที่ใกล้เคียงกับคำตอบถูกมากที่สุด เพื่อให้นักเรียนที่ไม่มีความรู้เลือกตอบตัวลวงเหล่านั้นกระจายออกไปทุกตัว ถ้าตัวลวงใดตัวหนึ่งไม่ใกล้เคียงกับคำตอบที่เป็นคำตอบถูก นักเรียนก็จะไม่เลือกตัวลวงเหล่านั้นเลย หากเป็นข้อสอบชนิด 5 ตัวเลือก ข้อสอบนั้นก็จะมีค่าเท่ากับข้อสอบชนิดเลือกตอบ 4 ตัวเลือก ยิ่งถ้าข้อสอบใดมีตัวลวงที่ไม่มีประสิทธิภาพหลาย ๆ ตัวก็จะทำให้ข้อสอบนั้น มีคุณภาพลดลงเรื่อย ๆ

ในการสร้างแบบสอบชนิดเลือกตอบโดยเฉพาะในส่วนที่เกี่ยวกับการสร้างตัวเลือกนั้น ได้มีผู้สรุปหลักเกณฑ์และให้ข้อเสนอแนะไว้ดังนี้

สมหวัง พิธิยานุวัฒน์ (2523 : 102-103) ได้กล่าวถึงหลักการและข้อเสนอในการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบไว้ คือ

1. ตัวลวงควรสร้างจากความเข้าใจของผู้สอบ ซึ่งมักจะได้จากคำตอบในการตอบข้อสอบแบบเรียงความ ข้อสอบแบบตอบสั้น ๆ หรือข้อสอบแบบเติมคำ
2. ควรหลีกเลี่ยงคำซ้ำในตัวเลือก พยายามตัดตอนให้สั้นกระชับรัดกุมที่สุด
3. ควรให้ตำแหน่งของคำตอบที่ถูกต้องอยู่แบบสุ่ม หรืออยู่อย่างกระจุกกระจาย
4. ไม่ควรทำให้คำตอบที่ถูกต้องยาวหรือสั้นกว่าตัวลวงจนสังเกตเป็นรูปแบบได้
5. หลีกเลี่ยงคำหรือข้อความที่ชี้แนะถึงคำตอบที่ถูกต้อง
6. ควรหลีกเลี่ยงตัวเลือกที่ว่า " ทุกข้อที่กล่าวมาแล้ว "
7. ใช้ตัวเลือกอย่างน้อย 4 ตัว เพื่อลดโอกาสเดาและควรให้ข้อสอบแบบเลือกตอบทุกข้อมีจำนวนตัวเลือกเท่ากัน
8. ควรหาตัวลวงซึ่งคล้ายคลึงกับตัวเลือกที่เป็นคำตอบถูกต้องให้มากที่สุด อีกทั้งเป็นตัวลวงที่ดึงดูดผู้ตอบที่ขาดความรู้ความสามารถที่มุ่งวัด
9. ตัวลวงแต่ละตัวควรเป็นที่สนใจแก่ผู้ตอบที่ไม่ทราบคำตอบที่ถูกต้องพอ ๆ กัน ถ้าไม่เช่นนั้นแล้วข้อสอบชนิดเลือกตอบ 4 ตัวเลือก อาจกลายเป็นข้อสอบแบบถูกผิดได้ ถ้าตัวเลือก 2 ตัวไม่ดึงดูดผู้ตอบที่ไม่ทราบคำตอบเลย ถ้าไม่สามารถหาตัวลวงที่ดีได้ควรเลือกใช้ข้อสอบแบบอื่นมากกว่าข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก

ยัง พิทยานิยม (2523 : อ้างถึงใน นิรมล บุญตะรัตน์ 2525:22) ได้แนะนำถึงการสร้างตัวเลือกของแบบสอบชนิดเลือกตอบไว้ดังนี้

1. เขียนคำตอบที่ถูกลงไปก่อน และสร้างตัวลวงทีหลัง
2. เขียนตัวเลือกให้เป็นพวกเดียวกัน
3. ใช้ตัวเลือกที่เป็นอิสระขาดจากกัน ไม่ก้าวร้าวหรือมีความหมายซ้ำซ้อนกัน
4. ตัวเลือกทุกตัวมีความหมายยาวเท่ากัน เพื่อไม่เป็นการแนะนำคำตอบ
5. ตัวเลือกทุกตัวต้องใช้ประโยชน์ได้
6. คำถามข้อเดียวให้มีตัวเลือกที่ถูกเพียงข้อเดียว
7. ให้ตัวเลือกที่เป็นคำตอบกระจายแบบสุ่ม
8. อย่าให้คำถามหรือตัวเลือกข้อต้น ๆ ไปมีอิทธิพลกับคำตอบข้อถัดไป

9. ใช้ตัวเลือกแบบปลายเปิดและปลายปิดให้เหมาะสม
10. อย่าให้คำตอบถูกและตัวลวง ถูกหรือผิดเด่นชัดเกินไป
11. ให้ตัวลวงมีส่วนถูกเจียด ๆ กับคำตอบถูก ตัวลวงที่จัดว่ามีประสิทธิภาพจะต้องสามารถทำให้นักเรียนที่ไม่รู้จริง เลือกตอบตัวลวงเท่านั้น
12. ตัวถูกหรือตัวผิดต้องถูกหรือผิดตามหลักการ หรือหลักวิชา และเป็นที่ยอมรับของบุคคลทั่วไป
13. กำหนดจำนวนตัวเลือกให้เหมาะสมกับวัยของเด็ก เช่น ระดับประถมศึกษาอาจใช้ตัวเลือกเพียง 3 ตัว ระดับมัธยมศึกษาตอนต้นควรรู้ใช้ 4 ตัวเลือก ส่วนในระดับสูงกว่านี้ควรรู้ใช้ 5 ตัวเลือก แต่ทั้งนี้ขึ้นอยู่กับภาระเขียนตัวเลือกด้วย
14. จัดเรียงตัวเลือกให้เป็นระเบียบ ถ้าไม่สามารถเขียนตัวเลือกให้ยาวเท่ากันได้ ก็พยายามเรียงตัวเลือกให้ตัวที่สั้นที่สุดอยู่ข้างบน ตัวที่ยาวกว่าอยู่รองลงมาจนถึงตัวสุดท้ายเป็นตัวเลือกที่ยาวที่สุดของข้อนั้น

วิรัช วรรณรัตน์ (2524 : อ้างถึงใน กัลญา เอื้อจิตติโกศา 2531:22-23) ได้กล่าวถึงกระบวนการสร้างข้อคำถามและตัวเลือกนั้น รวมทั้งเทคนิคการเขียนข้อสอบเลือกตอบไว้ดังนี้

1. การเขียนตัวคำถาม มีสาระและข้อควรคำนึงถึงดังนี้ คือ
 - 1.1 ควรทำเป็นประโยคคำถาม เพื่อให้ผู้สอบเข้าใจตัวคำถามอย่างชัดเจน
 - 1.2 เน้นจุดจะถามให้ชัดเจน
 - 1.3 ในคำถามหนึ่งข้อ ควรถามเพียงเรื่องเดียว
 - 1.4 ตัดคำฟุ่มเฟือยในตัวคำถามออก
 - 1.5 หลีกเลี่ยงการใช้คำปฏิเสธ ถ้าจำเป็นให้ขีดเส้นใต้
 - 1.6 ภาษาที่ใช้ควรเหมาะสมกับผู้สอบ ไม่ใช่ภาษาที่ซับซ้อน
 - 1.7 การใช้ภาพหรือข้อความประกอบคำถาม จะช่วยผู้สอบได้มาก
 - 1.8 ระวังคำถามต่างข้อจะแนะคำตอบซึ่งกันและกัน
 - 1.9 บางครั้งคำถามต่างทิศทางหรือต่างแนวกันอาจวัดในสิ่งเดียวกัน
 - 1.10 หลบเลี่ยงคำถามที่คล่องปาก
2. การเขียนตัวเลือก ควรคำนึงถึงสิ่งต่อไปนี้
 - 2.1 ถ้าตัวเลือกเป็นตัวเลขควรจัดลำดับกัน
 - 2.2 ควรระมัดระวังการใช้ตัวเลือกแบบปลายเปิด เช่น ไม่มีข้อถูกแบบใดก็ได้ ยังสรุปแน่นอนไม่ได้ และแบบปลายเปิด เช่น ถูกทุกข้อ ข้อ ก และ ข้อ ข ถูก เป็นต้น

- 2.3 ตัดคำฟุ่มเฟือยในตัวเลือกออก โดยนำคำที่ซ้ำ ๆ กันไปรวมไว้ในคำถาม
- 2.4 ควรมีตัวถูกเพียงตัวเดียว ทั้งตัวถูกและตัวผิดควรเป็นไปตามหลักการวิชาการ
- 2.5 หลีกเลี่ยงการใช้ภาษาแปลก ๆ และสะกดผิดในตัวถูก
- 2.6 ตัวเลือกควรเป็นอิสระจากกัน
- 2.7 อย่าให้ตัวถูก ตัวลวง เด่นชัดจนเกินไป
- 2.8 ข้อความในตัวถูกควรมีความสั้นยาวพอ ๆ กับตัวเลือกอื่น ๆ
- 2.9 หลบเลี่ยงคำขยายที่แนะนำคำตอบ

นอกจากนี้ ในการเขียนข้อคำถามและตัวเลือก ควรพิจารณาถึงสิ่งต่อไปนี้ประกอบด้วย

คือ

1. ตัวคำถามและตัวเลือกควรมีความเป็นเอกพันธ์กัน
2. ตัวถูกไม่ควรใช้คำหรือเสียงซ้ำกับคำถาม
3. ตัวคำถามและตัวเลือกควรเป็นเชิงอย่างหรือจงใจในทางที่ดี
4. ถ้าแต่งตัวเลือกไม่ได้ ควรลองนำตัวเลือกมาเป็นตัวคำถาม
5. ตำแหน่งของคำตอบถูกในแบบสอบชุดหนึ่ง ๆ ควรอยู่อย่างกระจาย ไม่ควร

เป็นระบบ

ธอร์นไดค์ (Thorndike 1971 : 113-119) ได้กล่าวถึงตัวเลือกของแบบสอบชนิดเลือกตอบไว้ดังนี้

1. ตัวเลือกที่ใช้ควรเหมาะสม สมเหตุสมผลและสอดคล้องกัน
2. ตัวเลือกทุกตัวต้องสอดคล้องกับคำถาม
3. ให้ตัวลวงทุกตัวดึงดูดความสนใจของผู้ตอบที่ไม่รู้จริง
4. หลีกเลี่ยงตัวเลือกที่หลอมน้ำกัน
5. ไม่ควรใช้คำว่า " ข้างบนทั้งหมด " และ " ไม่ใช่ข้างบนทั้งหมด "
6. เรียงลำดับตัวเลือกตามลำดับตรรกวิทยา (logical Order)
7. ไม่เขียนตัวเลือกเป็นลักษณะคล้ายแบบถูกผิด

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการทดสอบ

2.1 ทฤษฎีคลาสสิกอล (Classical Theory)

ในปัจจุบันทฤษฎีการวัดและประเมินผลการศึกษาที่ได้พัฒนาและนิยมใช้กันอย่างแพร่หลาย ซึ่ง กุลลิคเซิน (Gulliksen) 1950 ได้เป็นผู้เสนอทฤษฎีนี้ขึ้น (Warm 1978 :15) คือ ทฤษฎีคลาสสิกอล ซึ่งทฤษฎีนี้เป็นทฤษฎีดั้งเดิมที่นำมาใช้ในการวัดผลการศึกษาโดยเฉพาะการทดสอบ โดยที่นักวัดผลการศึกษาได้ใช้แบบสอบเป็นเครื่องมือในการกระตุ้นผู้เข้าสอบให้แสดงพฤติกรรมของความสามารถที่แท้จริงภายในตัว ซึ่งไม่สามารถวัดได้โดยตรง จากนั้นแล้วนำเอาพฤติกรรมที่ได้จากจำนวนข้อที่ตอบถูกซึ่งกำหนดในรูปของคะแนน โดยถือเป็นตัวแทนความสามารถที่แท้จริงของผู้สอบ และคะแนนจากการสอบที่นักเรียนตอบถูก (observed score) นั้นจะประกอบด้วย ส่วนที่เป็นคะแนนจริง (true score) และส่วนของคะแนนจากความคลาดเคลื่อน (error score) ซึ่งทฤษฎีนี้มีข้อตกลงเบื้องต้นต่อคะแนนที่วัดได้ว่า จะมีค่าเท่ากับ ผลรวมของคะแนนจริงกับคะแนนความคลาดเคลื่อน ซึ่งค่าความคาดหวังของคะแนนที่วัดได้ เท่ากับ คะแนนจริง และคะแนนความคลาดเคลื่อนกับคะแนนจริงไม่มีสหสัมพันธ์กัน

การตรวจสอบคุณภาพของแบบสอบ

กระบวนการตรวจสอบคุณภาพของแบบสอบตามแนวทฤษฎีคลาสสิกอลแบ่งออกเป็น 2 กระบวนการ คือ

1. การวิเคราะห์ข้อสอบรายข้อ โดยพิจารณาคุณลักษณะที่สำคัญของข้อสอบ 2 ประการ คือ

1.1 ค่าความยากของข้อสอบ

ค่าความยากของข้อสอบ หมายถึง สัดส่วนของนักเรียนที่ตอบข้อสอบข้อนั้นถูก จากคำนิยามความหมายของค่าความยากของข้อสอบนั้นจะเห็นได้ว่า ค่าความยากสามารถคำนวณได้จากจำนวนนักเรียนที่ตอบข้อสอบนั้นถูกหารด้วยจำนวนนักเรียนที่ตอบข้อสอบข้อนั้นทั้งหมด ดังนั้นค่าความยากของข้อสอบจึงมีค่าตั้งแต่ 0 ถึง 1 ถ้าค่าความยากของข้อสอบสูงหรือมีค่าใกล้ 1 แสดงว่า ข้อสอบนั้นง่ายมีนักเรียนจำนวนมากที่สามารถตอบถูก แต่ถ้าระดับค่าความยากต่ำหรือค่าใกล้ 0 แสดงว่า ข้อสอบนั้นยาก มีนักเรียนน้อยคนที่สามารถตอบถูก การพิจารณาระดับความยากที่พอเหมาะนั้นจะต้องทราบจุดมุ่งหมายของการวัด ในการวัดผลแบบอิงกลุ่ม (Norm Referenced) นั้นค่าความยากของข้อสอบที่เหมาะสมคือ .50 ทั้งนี้เพราะทำให้การกระจายของคะแนนมากที่สุด แต่การจะสร้างข้อสอบในแบบสอบให้มีความยากพอเหมาะทุก

ข้อนี้ทำได้ยาก ดังนั้นจึงแนะนำให้ใช้ข้อสอบที่มีความยากอยู่ในช่วง .20 - .80 ซึ่งก็นับว่าใช้ได้ (ชวาล แพร์ตกุล 2509 : 316) ส่วนข้อสอบที่มีค่าความยากต่ำกว่า .20 หรือสูงกว่า .80 เป็นข้อสอบที่ยากหรือง่ายเกินไป ข้อสอบเหล่านี้จำแนกความสามารถของนักเรียนไม่ได้หรือจำแนกได้น้อย สำหรับการวัดผลแบบอิงเกณฑ์ (Criterion-Referenced) ส่วนใหญ่จะเป็นข้อสอบที่มีค่าความยากสูง กล่าวคือ ข้อสอบง่าย ทำให้การกระจายของคะแนนที่ลักษณะเบ้ซ้าย กล่าวคือ คะแนนของนักเรียนส่วนใหญ่จะสูง ซึ่งจะช่วยให้ทราบว่านักเรียนคนใดบ้างที่มีความรู้ และทักษะไม่ถึงเกณฑ์ที่เราคาดหวัง (อนันต์ ศรีโสภา 2519 : 257)

องค์ประกอบที่มีอิทธิพลต่อค่าความยากของแบบสอบ

องค์ประกอบที่มีอิทธิพลต่อค่าความยากของแบบสอบมี 2 ประเภท (Campbell 1961: 899-913) คือ

1. องค์ประกอบภายใน (Intrinsic Factor) ประกอบด้วย

ก. เนื้อหาของแบบสอบแต่ละข้อ (Item Content) ในด้านความซับซ้อนความเป็นนามธรรมและความแปลกใหม่

ข. โครงสร้างของข้อสอบ (Item Structure) หมายถึง วิธีในการแสดงออกซึ่งเนื้อหา

2. องค์ประกอบภายนอก (Extrinsic Factor) ประกอบด้วย

ก. ความไม่คุ้นเคยต่อเนื้อหา (Unfamiliarity) คือ อยู่นอกเหนือประสบการณ์ของผู้สอบ

ข. สิ่งที่มีสัมพันธ์กับข้อสอบ (Item Context) เช่น ข้อสอบที่อยู่ใกล้เคียงกัน

ค. ตัวแปรด้านบุคลิกภาพ ได้แก่ สภาพร่างกาย ลักษณะนิสัย และความตั้งใจของผู้สอบ

1.2 อำนาจจำแนกของข้อสอบ

อำนาจจำแนก หมายถึง ประสิทธิภาพของข้อสอบที่จะสามารถแยกความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ ซึ่งมีค่าตั้งแต่ -1.00 ถึง 1.00 ข้อสอบที่มีค่าอำนาจจำแนกเข้าใกล้ -1.00 หรือ 1.00 แสดงว่ามีอำนาจจำแนกสูง แต่ข้อสอบที่มีคุณภาพตรงตามวัตถุประสงค์ของการสอบ คือ ข้อสอบที่ค่าอำนาจจำแนกเป็นบวก ซึ่งหมายความว่า นักเรียนในกลุ่มสูงตอบได้มากกว่านักเรียนในกลุ่มต่ำ และแสดงว่าข้อสอบสามารถจำแนกไปในทางเดียวกับคะแนนที่ได้จากแบบสอบ ส่วนข้อสอบที่มีค่าอำนาจจำแนกเป็นทางลบ ถือว่าจำแนกผิดทิศทาง

ไม่ตรงตามวัตถุประสงค์ที่ต้องการ ถ้าข้อสอบที่มีค่าอำนาจจำแนกเป็น ๑ แสดงถึงกลุ่มที่ได้คะแนนสูงตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากับกลุ่มที่ได้คะแนนต่ำซึ่งไม่สามารถจำแนกความแตกต่างระหว่างคน 2 กลุ่มนี้ได้เลย

อีเบล (Ebel 1965 :365) ได้เสนอเกณฑ์สำหรับตัดสินค่าอำนาจจำแนกของข้อสอบไว้ดังนี้

ตารางที่ 1 เกณฑ์การตัดสินค่าอำนาจจำแนก

ค่าอำนาจจำแนก	คุณภาพของข้อสอบ
.40 ขึ้นไป	ดีมาก
.30 ถึง .39	ดี
.20 ถึง .29	พอใช้ได้ แต่ต้องปรับปรุง
ต่ำกว่า .19	ใช้ไม่ได้ ต้องพิจารณาปรับปรุงใหม่

สิ่งสำคัญประการหนึ่งที่ควรคำนึงถึงในการประเมินคุณภาพของข้อสอบควบคู่กับค่าความยากของข้อสอบ เนื่องจากข้อสอบที่มีค่าความยากเท่า ๆ กันนั้นอาจจะต่างกันในด้านอำนาจจำแนก ดังได้กล่าวมาแล้วว่าข้อสอบที่มีค่าความยาก .50 อาจมีค่าอำนาจจำแนกสูงสุดหรือไม่มีเลยก็ได้ ฉะนั้นข้อสอบที่มีค่าความยาก .50 มีโอกาสที่จะทำให้ค่าความเที่ยงของแบบสอบต่ำก็ได้ เมื่อค่าอำนาจจำแนกเป็น ๑ แต่ข้อสอบที่มีค่าอำนาจจำแนกสูง จะมีผลทำให้ค่าความเที่ยงของแบบสอบสูงด้วยอย่างแน่นอน เนื่องจากค่าอำนาจจำแนกของข้อสอบ เกี่ยวข้องกับส่วนเบี่ยงเบนมาตรฐานของคะแนน ซึ่งเป็นองค์ประกอบสำคัญที่มีอิทธิพลต่อค่าความเที่ยงของแบบสอบ

(Ebel 1965 :302)

2. การวิเคราะห์แบบสอบทั้งฉบับ โดยพิจารณาคูณลักษณะที่สำคัญ 2 ประการคือ

2.1 ความตรงของแบบสอบ

นิยามของคำว่า " ความตรง " (Validity) ของแบบสอบ ได้มีผู้ให้ความหมายไว้หลายท่าน เช่น

ความตรง หมายถึง ความถูกต้องแน่นอนในการวัดสิ่งที่ต้องการวัด

(Linguist 1942 :213)

ความตรง หมายถึง คุณสมบัติของแบบสอบที่สามารถวัดในสิ่งที่ต้องการจะวัด (Ebel 1965 :273)

ความตรง หมายถึง ความถูกต้องแม่นยำของคะแนนในการพยากรณ์ที่จำเพาะเจาะจง (Stanley 1978 :101)

ความตรงของแบบสอบ หมายถึง ขอบเขตของผลการประเมินที่สอดคล้องกับจุดประสงค์นั้น (Gronlund 1976 :79)

ความตรงของแบบสอบ หมายถึง ค่าสหสัมพันธ์ระหว่างแบบสอบกับเกณฑ์ (Gulliksen 1950 :88)

จากความหมายของความตรงดังกล่าว สามารถสรุปความหมายของความตรงได้จากแบบสอบกับคะแนนเกณฑ์ 2 ประการ คือ

1. ความตรงของแบบสอบ หมายถึง ความแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด หรือสิ่งที่เครื่องมือควรจะวัด และคะแนนที่ได้จากแบบสอบที่ความตรงสูง สามารถบอกถึงสภาพที่แท้จริง และพยากรณ์ได้ถูกต้องแม่นยำ

2. ความตรงของแบบสอบ หมายถึง ค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากแบบสอบกับคะแนนเกณฑ์

ประเภทของความตรง

ความตรงของแบบสอบจำแนกออกได้หลายวิธี ทั้งนี้ขึ้นอยู่กับจุดมุ่งหมายของการวัดโดยทั่วไป การวัดทางการศึกษาและจิตวิทยาแบ่งความตรงออกเป็น 3 ประเภท (Mehrens and Lehmann 1973 :109)

1. ความตรงตามเนื้อหา (Content Validity) เป็นความตรงที่เกี่ยวกับการวิเคราะห์ตรวจสอบอย่างมีระบบในเนื้อหาของแบบสอบ ซึ่งความตรงนี้การพิจารณาขึ้นอยู่กับตัดสินของแต่ละบุคคลและการตัดสินแบบปรนัย ความตรงตามเนื้อหานี้แบ่งออกเป็น 2 ชนิด คือ

1.1 ความตรงตามปรากฏ เป็นความตรงที่มีผู้ตรวจแบบสอบเป็นผู้เชี่ยวชาญได้สรุปว่า แบบสอบนั้นวัดได้ตรงตามความต้องการที่จะวัด

1.2 ความตรงตามตรรกหรือการสุ่ม เป็นความตรงที่เกี่ยวข้องกับการสุ่มเนื้อหา โดยมีการนิยามโดเมนของพฤติกรรมที่จะวัดในการสอบ แล้วสุ่มเนื้อหาในการสร้างข้อสอบโดยให้ครอบคลุมเนื้อหาที่สำคัญของโดเมนที่นิยามไว้ ซึ่งมีประโยชน์มากในการพัฒนาแบบสอบวัดผลสัมฤทธิ์ โดยทั่วไปมักจะพิจารณาข้อสอบในแบบสอบโดยเทียบกับตารางวิเคราะห์หลักสูตร (Table of Specification) แล้วประมาณออกมาว่ามีความหมายตามเนื้อหามากน้อยเพียงใด ผู้ที่จะ

ตัดสินความตรงตามเนื้อหาได้ดีคือผู้เชี่ยวชาญสาขาอื่น ๆ (Mehrena and Lehmann 1978 :111)

2. ความตรงตามเกณฑ์สัมพัทธ์ (Criterion-Related Validity) เป็นความสามารถของแบบสอบในการทำนายพฤติกรรมของบุคคลในสถานการณ์เฉพาะ หรือแสดงถึงความสัมพันธ์ระหว่างคะแนนที่ผู้สอบกระทำได้จากการสอบกับเกณฑ์ โดยที่เกณฑ์เป็นพฤติกรรมที่ได้จากคะแนนการสอบที่ใช้ทำนาย ความตรงตามเกณฑ์สัมพัทธ์แสดงในรูปของสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบหรือตัวทำนายกับคะแนนเกณฑ์จำแนกได้ 2 ชนิดคือ

2.1 ความตรงตามสภาพ หรือความตรงร่วมสมัย (Concurrent Validity) หมายถึง ความตรงของแบบสอบที่มีความสามารถบอกให้เรารับว่าสิ่งที่วัดถูกต้องตามสภาพที่เป็นจริงในปัจจุบัน โดยการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบกับคะแนนเกณฑ์ ซึ่งคะแนนทั้งสองนี้วัดได้ในระยะเวลาเดียวกัน ความตรงร่วมสมัยนี้เป็นการใช้คะแนนการสอบประมาณเกณฑ์ร่วมสมัยมากกว่าใช้ทำนายเกณฑ์ในอนาคต

2.2 ความตรงเชิงทำนาย (Predictive Validity) หมายถึง ความสามารถของแบบสอบที่จะบ่งบอกผลที่วัดในขณะนั้น ได้ถูกต้องตามสภาพที่แท้จริงในอนาคต โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบกับคะแนนเกณฑ์ ซึ่งคะแนนเกณฑ์เป็นคะแนนที่รวบรวมภายหลัง ความตรงเชิงทำนายนี้เป็นการใช้คะแนนการสอบทำนายพฤติกรรมในอนาคต ความตรงเชิงทำนายเหมาะที่จะใช้กับแบบสอบความถนัดชนิดต่าง ๆ

3. ความตรงตามโครงสร้าง (Construct Validity) ข้อสอบที่สร้างขึ้นมานั้นถ้าสามารถวัดได้ในสิ่งนั้น ๆ จริง ย่อมถือว่าข้อสอบมีความตรงเชิงโครงสร้าง เนื่องจากมีคุณลักษณะ (trait) ที่ต้องการวัดนั้นอยู่หลายลักษณะ วิธีการหาความตรงตามโครงสร้างนี้มีหลายวิธี เช่น การหาสหสัมพันธ์ การหาความสอดคล้องภายใน การวิเคราะห์ตัวประกอบ เป็นต้น

องค์ประกอบที่มีอิทธิพลต่อค่าความตรงของแบบสอบ

ข้อสอบที่สร้างขึ้นมานั้นจะมีความตรงสูงหรือต่ำ ย่อมขึ้นอยู่กับองค์ประกอบหลาย ๆ ประการดังนี้ (Gronlund 1976 :98-102) คือ

1. องค์ประกอบที่เกิดจากตัวแบบสอบ มีดังนี้

1.1 ความไม่ชัดเจนของคำชี้แจง คำชี้แจงที่ไม่ชัดเจนเกี่ยวกับวิธีตอบข้อสอบ การบันทึกคำตอบ เป็นต้น จะทำให้ค่าความตรงลดลง

1.2 ใช้โครงสร้างของคำและประโยคยากเกินไป ทำให้นักเรียนสับสน แทนที่จะวัดได้ตรงกับวัตถุประสงค์ที่ต้องการ ก็จะกลายเป็นการวัดความสามารถทางภาษา

- 1.3 ข้อสอบที่สร้างไม่ดีบางข้อเป็นการแนะนำคำตอบให้นักเรียน
 - 1.4 ความกำกวมของภาษาที่ใช้ในแบบสอบที่จะวัดสิ่งที่ต้องการวัด
 - 1.5 ความไม่เหมาะสมของแบบสอบที่จะวัดสิ่งที่ต้องการวัด
 - 1.6 แบบสอบสั้นเกินไป ไม่สามารถวัดได้ครอบคลุมพฤติกรรมที่ต้องการจะวัด
 - 1.7 การจัดเรียงข้อสอบไม่เหมาะสม ข้อสอบที่เรียงจากยากไปหาง่าย ทำให้นักเรียนใช้เวลาคิดหาคำตอบนาน ทำให้ไม่มีเวลาทำข้อสอบที่ง่าย ๆ ข้อหลัง ๆ การเรียงข้อสอบแบบนี้มีผลต่อค่าความตรงโดยมีผลต่อแรงจูงใจในการทำแบบสอบของนักเรียน
 - 1.8 การเรียงคำตอบอย่างเป็นระบบ ช่วยให้นักเรียนเดาคำตอบได้ง่าย
 - 1.9 ชนิดของแบบสอบที่ใช้ไม่เหมาะสม แบบสอบที่ไม่เหมาะสมอาจวัดสิ่งที่ต้องการวัดได้ไม่ดีเพียงพอ ทำให้ไม่อาจประเมินสิ่งที่ต้องการวัดได้อย่างถูกต้อง
2. องค์ประกอบที่เกี่ยวข้องกับเนื้อหาวิชาและกระบวนการเรียนการสอน ข้อสอบที่สร้างขึ้นมาแล้ว ออกคำถามวัดในเนื้อหาวิชาที่นักเรียนยังไม่ได้เรียน หรือออกข้อคำถามวัดการคิดแก้ปัญหาแต่ไม่สอนและเฉลย เพื่อให้นักเรียนได้จดจำไปตอบ เช่นนี้ ย่อมทำให้ค่าความตรงของแบบสอบมีค่าต่ำลง จึงกล่าวได้ว่า ขบวนการเรียนการสอนในชั้นเรียนย่อมมีอิทธิพลต่อความตรงของแบบสอบที่สร้างขึ้นมาโดยตรง
3. องค์ประกอบของการตรวจคะแนนและการบริหารแบบสอบ เช่น เวลาในการสอบน้อยเกินไป การชี้แนะนักเรียนบางคนในการตอบ การให้คะแนนที่ไม่คงที่ เหล่านี้ทำให้ค่าความตรงลดลงได้
4. องค์ประกอบในการตอบแบบสอบของนักเรียนเอง รูปแบบการตอบแบบสอบของนักเรียนแต่ละคนแตกต่างกันไป ซึ่งทำให้ผลการสอบไม่ตรงกับความสามารถที่แท้จริง ตลอดถึงความวิตกกังวล ความเปลี่ยนแปลงทางอารมณ์ ทำให้ค่าความตรงลดลง
5. องค์ประกอบที่เกี่ยวข้องกับลักษณะของคนสอบในกลุ่มตัวอย่าง ตามปกติความตรงของแบบสอบจะมีค่าสูงเฉพาะในบางกลุ่ม เนื่องจากในการสอบวัดนั้น มีอิทธิพลจากองค์ประกอบอื่น ๆ เช่น อายุ เพศ ระดับความสามารถ เป็นต้น ดังนั้นในการใช้แบบสอบควรคำนึงถึงลักษณะเฉพาะของกลุ่มที่ใช้หาความตรง นอกจากนั้นต้องคำนึงถึงเกณฑ์ที่ใช้หาความตรงของแบบสอบให้เหมาะสมด้วย

2.2 ความเที่ยงของแบบสอบ

ได้มีผู้ให้ความหมายของความเที่ยงของแบบสอบไว้หลายท่าน เช่น

ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการสอบซ้ำกับกลุ่มตัวอย่างนักเรียนกลุ่มเดียวกันด้วยแบบทดสอบฟอร์มเดียว หรือแบบสอบคู่ขนาน (equivalent form) 2 ฟอร์ม ซึ่งมีเซตของข้อสอบที่เทียบเท่ากันหรือขนานกัน (อนาสตาซี 2519: 73)

ความเที่ยง หมายถึง ความคงที่ของผลที่ได้รับจากการใช้เครื่องมือชนิดเดียวกัน สอบมากกว่า 1 ครั้ง (Gronlund 1968 :163)

ความเที่ยง หมายถึง ความคงเส้นคงวาของชุดคะแนนจากแบบสอบที่วัดในสิ่งที่ต้องการจะวัด

จากที่กล่าวมาแล้วถึงความหมายของความเที่ยงของแบบสอบ พอสรุปได้ว่า ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการวัดด้วยเครื่องมือชุดเดียวกันในเวลาต่างกัน

นอกจากความหมายของความเที่ยงในลักษณะข้างต้น ยังมีนักการศึกษาบางท่านกล่าวถึงในเชิงของนิยามเชิงปฏิบัติการ (Operational definition) ว่า ค่าความเที่ยงของแบบสอบ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบคู่ขนาน 2 ชุด ซึ่งสอบนักเรียนกลุ่มเดียวกัน ด้วยวิธีดำเนินการสอบที่เป็นอิสระต่อกัน (Ebel 1965: 311 ; Allen and Yen 1979: 72)

ทฤษฎีของความเที่ยง (Theory of Reliability)

ในการวัดทางจิตมิติ จะเป็นลักษณะการวัดโดยอ้อม และสิ่งที่วัดจะไม่ค่อยคงที่แน่นอน เครื่องมือที่ใช้วัดก็ไม่ครอบคลุมทั้งหมด และจุดประสงค์หลักในการวัดแต่ละครั้งเพื่อต้องการทราบลักษณะพฤติกรรม หรือความสามารถที่แท้จริงของบุคคลมากกว่า แต่ความเป็นไปได้มีน้อยมาก เพราะแหล่งความคลาดเคลื่อนในการวัดไม่ได้มาจากตัวเครื่องมือเพียงลำพัง แต่มาจากแหล่งอื่น ๆ มากมาย เช่น โอกาส สถานที่ และสิ่งแวดล้อมต่าง ๆ ในขณะที่ดำเนินการสอบ ดังนั้นจึงทำให้มีการศึกษาในเรื่องความเที่ยงขึ้น เพื่อประมาณขนาดของผลจากความคลาดเคลื่อนในการวัดที่เกิดจากแหล่งต่าง ๆ หรืออาจกล่าวได้ว่าเป็นการศึกษาเพื่อประมาณระดับความสอดคล้องหรือความแม่นยำของเครื่องมือตนเอง คะแนนจากการวัดแต่ละครั้ง ประกอบด้วย คะแนนจริง และคะแนนความคลาดเคลื่อนรวมกัน ตามสมการซึ่งถือว่าเป็นสมการพื้นฐานของทฤษฎีการวัดแบบดั้งเดิม

(Classical Test Theory) ดังนี้คือ (Mehrens and Lehman 1984: 269)

$$X = T + E$$

เมื่อ	X	แทน คะแนนที่วัดได้ (observed score)
	T	แทน คะแนนจริง (True score)
	E	แทน คะแนนความคลาดเคลื่อน (Error score)

คะแนนที่วัดได้ (observed score) หมายถึง คะแนนที่ได้จากการวัดด้วยเครื่องมือ
 คะแนนจริง (True score) หมายถึง คะแนนที่ผู้สอบทำได้จากการวัดด้วยเครื่องมือ
 ที่มีคุณลักษณะสมบูรณ์ปราศจากความคลาดเคลื่อน หรือหมายถึง คะแนนเฉลี่ยของผู้สอบที่ได้จาก
 การทำแบบสอบเดิม หรือแบบสอบคู่ขนานหลาย ๆ ครั้ง โดยมีข้อตกลงว่า ไม่มีอิทธิพลจากฝึกฝน
 ความเมื่อยล้า และการเรียนรู้ในการทดสอบซ้ำ (Brown 1976: 50)

คะแนนความคลาดเคลื่อน (Error score) หมายถึง คะแนนที่เกิดจากความ
 คลาดเคลื่อนในการวัด ซึ่งเป็นความคลาดเคลื่อนที่เกิดขึ้นในลักษณะสุ่ม (Random error) ที่
 เกิดจากการวัด เช่น การตรวจให้คะแนน ตัวแบบสอบ การบริหารการสอบ เป็นต้น

ตามแนวคิดทฤษฎีของความเที่ยงกล่าวว่า ความคลาดเคลื่อนที่เกิดขึ้นจะมีค่าได้ทั้งทาง
 บวกและทางลบ ซึ่งจะหักล้างกันหมดไป นั่นคือ ค่าเฉลี่ยของความคลาดเคลื่อนจะเท่ากับ 0 ส่วน
 เบี่ยงเบนมาตรฐานเท่ากับ 1 สามารถเขียนเป็นสมการแสดงความแปรปรวนของคะแนนได้ดังนี้
 (Mehrens and Lehmann 1984: 269)

$$s^2_x = s^2_t + s^2_e$$

เมื่อ	s^2_x	คือ ความแปรปรวนของคะแนนที่วัดได้
	s^2_t	คือ ความแปรปรวนของคะแนนจริง
	s^2_e	คือ ความแปรปรวนของคะแนนความคลาดเคลื่อน

โดยทฤษฎีความเที่ยง หมายถึง อัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับ
 ความแปรปรวนของคะแนนที่วัดได้ (Brown 1976: 51) เขียนเป็นสมการได้ดังนี้

$$r_{tt} = \frac{s^2_t}{s^2_x}$$

เมื่อ	r_{tt}	แทน ค่าความเที่ยง
-------	----------	-------------------

การที่นิยมใช้สัญลักษณ์ r_{xx} แทนค่าความเที่ยงนั้นเป็นเพราะการประมาณค่าความเที่ยงส่วนใหญ่อาศัยพื้นฐานในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และใช้ subscript x 2 ตัว เพื่อบ่งชี้ว่าเป็นการวัดลักษณะนิสัย (trait) เดียวกัน จากสมการที่กล่าวมาข้างต้นเป็นที่มาของสมการที่บ่งชี้ค่าความเที่ยง คือ

$$r_{xx} = 1 - \frac{s^2_e}{s^2_x}$$

ซึ่งจะเห็นว่า ค่าความเที่ยงจะสูงขึ้นเมื่อความแปรปรวนคลาดเคลื่อนลดลง โดยกำหนดให้คะแนนผลการสอบมีค่าคงที่หรือเพิ่มขึ้น และค่าความเที่ยงจะมีค่าสูงสุดเท่ากับ 1

วิธีการประมาณค่าความเที่ยง

ในการประมาณค่าความเที่ยงสามารถทำได้หลายวิธี และในแต่ละวิธีก็มีข้อจำกัดที่แตกต่างกัน การที่จะเลือกใช้วิธีใดขึ้นอยู่กับ ความมุ่งหมาย ลักษณะของคะแนน และชนิดของแบบสอบ (Guilford 1978: 414) โดยทั่วไปวิธีการประมาณค่าความเที่ยงมี 4 รูปแบบ

(Mehrens and Lehmann 1978: 84) คือ

1. แบบสอบซ้ำ (Measures of stability)
2. แบบใช้ข้อสอบคล้ายกัน (Measures of equivalence)
3. แบบใช้ข้อสอบคล้ายกันและสอบซ้ำ (Measures of equivalence and stability)
4. แบบวัดความคงที่ภายใน (Measures of internal consistency) ซึ่งวิธีการนี้แบ่งออกเป็น

- 4.1 วิธีแบ่งครึ่งข้อสอบ (Split-half)
- 4.2 วิธีของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson estimates)
- 4.3 วิธีสัมประสิทธิ์แอลฟา (Coefficient alpha)
- 4.4 วิธีวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's analysis of variance procedure)

ในที่นี้จะกล่าวถึงเฉพาะวิธีการประมาณค่าความเที่ยงแบบคูเดอร์-ริชาร์ดสัน ซึ่งเป็นวิธีการประมาณค่าความเที่ยงในการวิจัยครั้งนี้

วิธีการประมาณค่าความเที่ยงแบบคูเดอร์-ริชาร์ดสัน

การประมาณค่าความเที่ยงแบบคูเดอร์-ริชาร์ดสัน เป็นวิธีการประมาณค่าความเที่ยงที่คูเดอร์-ริชาร์ดสัน (Kuder-Richardson) ได้เสนอไว้ในปี ค.ศ. 1934 เป็นวิธีที่รู้จักและนิยมแพร่หลายคือ สูตรคูเดอร์-ริชาร์ดสัน 20 (Kuder-Richardson 20) (Brown 1976:78)

$$r_{kk} = \frac{(K-1)(S^2_t - \sum P_i Q_i)}{S^2_t}$$

เมื่อ r_{kk}	แทน	ค่าความเที่ยงของแบบสอบ
K	แทน	จำนวนข้อในแบบสอบ
P_i	แทน	สัดส่วนของผู้ตอบถูกในแต่ละข้อ
Q_i	แทน	สัดส่วนของผู้ตอบผิดในแต่ละข้อ
S_t^2	แทน	ความแปรปรวนของคะแนนทั้งหมด

การประมาณค่าความเที่ยงแบบคูเดอร์-ริชาร์ดสัน เป็นการหาความสอดคล้องกันระหว่างข้อสอบซึ่งดำเนินการสอบเพียงครั้งเดียว และใช้แบบสอบชุดเดียว ความสอดคล้องกันระหว่างข้อนี้ได้รับอิทธิพลจากแหล่งความแปรปรวนคลาดเคลื่อน 2 แหล่ง (อนาสตาซี 2519: 84) คือ

1. เนื้อหาที่สับสน
2. ความเป็นวิวิธพันธ์ของพฤติกรรมที่สับสน ยิ่งข้อสอบมีความเป็นเอกพันธ์มาก ความสอดคล้องกันระหว่างข้อยิ่งสูง (Stanley 1978: 126)

มีข้อตกลงเบื้องต้นร่วมกันในการประมาณค่าความเที่ยงสูตรคูเดอร์-ริชาร์ดสัน คือ วิธีการใช้ตรวจให้คะแนนแบบสอบชุดนั้นจะต้องเป็นวิธีการ 0-1 (dichotomized) นั่นคือตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน ข้อสอบแต่ละข้อจะต้องเป็นเอกพันธ์กัน หรือวัดคุณลักษณะเดียวกัน นอกจากนี้แล้วสูตรคูเดอร์-ริชาร์ดสัน 20 นี้เหมาะที่จะใช้หาความเที่ยงของแบบสอบความสามารถ (Power test) เท่านั้น ไม่เหมาะที่จะใช้กับแบบสอบวัดความเร็ว (Speed test) ทั้งนี้เพราะค่า P และ ค่า Q ของแต่ละข้อต้องเป็นค่าที่ได้จากการที่ผู้สอบทุกคนมีโอกาสทำข้อสอบข้อนั้นเสร็จแล้ว นั่นหมายความว่า การใช้สูตร 20 นี้ข้อสอบจะต้องมีค่าความยากพอ ๆ กัน (ค่า p คงที่)

องค์ประกอบที่มีอิทธิพลต่อค่าความเที่ยง

1. องค์ประกอบภายในตัวแบบสอบ

1.1 ความยาวของแบบสอบ (Test Length) จะมีผลต่อค่าความเที่ยงของแบบสอบ เนื่องจากว่าถ้าแบบสอบมีมากข้อสอบขึ้น โอกาสที่ตนจะเดาถูกมีน้อย ค่าความเที่ยงของแบบสอบก็จะยาวขึ้น สูตรที่แสดงให้เห็นว่าเมื่อเพิ่มจำนวนข้อในแบบสอบทำให้ค่าความเที่ยงเพิ่มขึ้นคือสูตรของสเปียร์แมน-บราวน์ (Spearman-Brown) และครอนบาค (Cronbach) ได้เสนอตารางสำเร็จรูปเปลี่ยนค่าความเที่ยงของแบบสอบเป็น อัตราส่วนสัญญาณเสียง (Signal noise ratio หรือใช้สัญลักษณ์ว่า S/N) ซึ่งเป็นการหาค่าสัมประสิทธิ์ค่าความเที่ยงของแบบสอบหลังจากการเพิ่มความยาวของแบบสอบในขนาดต่าง ๆ ซึ่งคำนวณไม่ยุ่งยาก (Cronbach 1970: 170-171)

1.2 ความเป็นปรนัยของการวัด (Objectivity) แบบสอบที่มีความเป็นปรนัยสูง เช่น ปราศจากอคติของผู้ตรวจให้คะแนน ความชัดเจนของคำชี้แจง ถ้าการสอบในครั้งนี้มีความเป็นปรนัยมากก็จะมีค่าความเที่ยงสูง

1.3 ความยากของข้อสอบ (Item Difficulty) เนื่องจากค่าความเที่ยงของแบบสอบนั้นประมาณจากความแปรปรวนของคะแนนผลการสอบ ถ้าข้อสอบยากเกินไปทุกคนทำผิดหมด หรือถ้าข้อสอบง่ายเกินไปจะทำให้ทุกคนทำถูกหมด จะทำให้ผลของคะแนนลดลง ความแปรปรวนของคะแนนจะน้อย มีผลทำให้ความเที่ยงลดลงด้วย ถ้าข้อสอบมีค่าความยากเท่ากับ .50 จะทำให้ค่าความเที่ยงสูงกว่าข้อสอบที่มีระดับความยากปานกลาง (ตั้งแต่ .25-.75) โอกาสการเดาถูกก็เช่นกัน ถ้าเดาถูกได้ง่าย ค่าความเที่ยงจะต่ำลง

2. องค์ประกอบภายนอกแบบสอบ

2.1 ความเหมาะสมของแบบสอบกับลักษณะของผู้สอบ คือ แบบสอบไม่มีความลำเอียงกับผู้สอบกลุ่มใดกลุ่มหนึ่ง เช่น เพศ วัย วัฒนธรรม ระดับความสามารถของผู้สอบ ฯลฯ ถ้าเหมาะสมมากค่าความเที่ยงจะสูง

2.2 ลักษณะของผู้สอบ เช่น ผู้สอบที่มีลักษณะวิวิริพันธ์ (Heterogeneity) จะทำให้ค่าความเที่ยงสูง ถ้าผู้สอบไม่พึงพอใจกับสภาพการสอบวัด ค่าความเที่ยงจะต่ำลง

2.3 ความผิดปกติในตัวผู้สอบและสภาพแวดล้อมต่าง ๆ เช่น ความเจ็บป่วยในตัวบุคคล แสงสว่างและการระบายอากาศ การทุจริตในการสอบ เป็นต้น จะมีผลทำให้ค่าความเที่ยงลดลง

2.4 การเลือกวิธีการประมาณค่าความเที่ยง เพราะแต่ละวิธีจะมีข้อดีและข้อจำกัดต่างกัน ดังนั้นควรพิจารณาให้ดีก่อนตัดสินใจในการเลือกใช้ ถ้าเป็นแบบสอบประเภทความเร็ว (Speed test) ควรใช้วิธีสอบซ้ำ หรือวิธีสลับฟอร์ม ส่วนวิธีคูเดอร์ ริชาร์ดสัน และวิธีแอลฟาของครอนบาค ไม่เหมาะกับแบบสอบที่วัดลักษณะนิสัยหลาย ๆ ประเภท เพราะมีข้อตกลงพื้นฐานว่า ข้อสอบต้องเป็นเอกพันธ์ สูตรสเปียร์แมน บราวน์ (Spearman Brown) จะให้ค่าสัมประสิทธิ์ความเที่ยงที่อาจจะสูงหรือต่ำเกินความเป็นจริง ถ้าเนื้อหาของแบบสอบที่แบ่งครั้งออกจากกันนั้นไม่เทียบเคียงกัน

อนึ่งแบบสอบชุดเดียวกัน อาจจะให้ค่าความเที่ยงจากวิธีการประมาณค่าวิธีเดียวกันได้หลายค่า ถ้านำแบบสอบไปสอบกับผู้สอบหลายกลุ่ม และภายใต้เงื่อนไขในการดำเนินการสอบที่ต่างกัน (Allen and Yen 1979: 88)

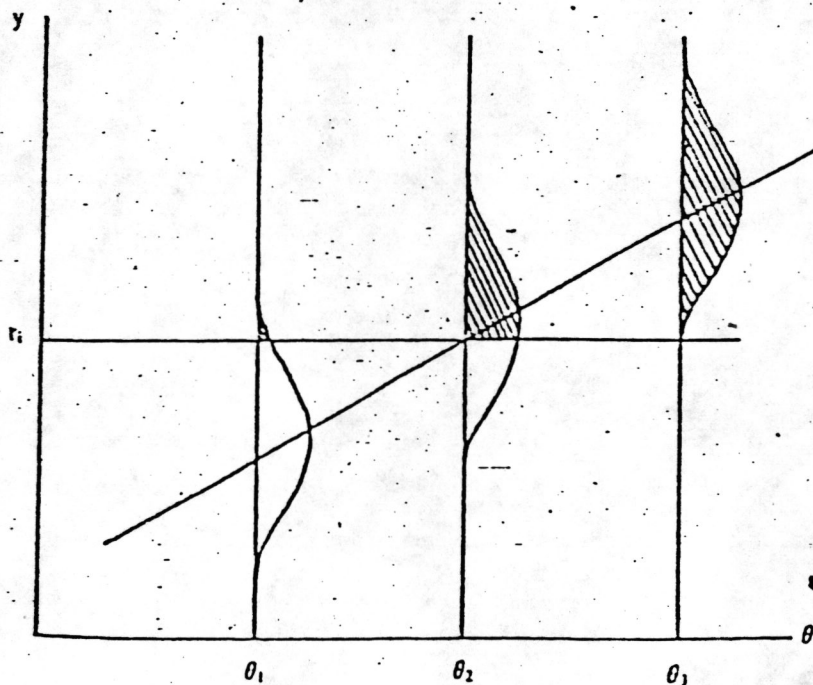
2.5 สภาพการณ์ของการดำเนินการสอบ เช่น แบบสอบความเร็ว (Speed test) หรือแบบสอบมาตรฐานที่จำกัดเวลา แต่ให้เวลาสอบเกินกว่าที่กำหนดให้ จะทำให้ความเที่ยงลดลง

2.2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการตอบสนองข้อสอบ (Item response theory) หรือทฤษฎีความสามารถแฝง (Latent trait theory) หรือทฤษฎีรูปโค้งลักษณะของข้อสอบ (Item characteristic curve theory) ได้เริ่มเข้ามามีบทบาทในวงการวัดและประเมินผลการศึกษาตั้งแต่ปี ค.ศ. 1936 โดยในปี 1942 และ 1943 เฟอร์กูสัน และ ลอเลย์ (Ferguson and Lawley) ได้เสนอแนวคิดและหลักการเบื้องต้นของทฤษฎีนี้ ซึ่งเป็นโมเดลทางคณิตศาสตร์ที่มีค่าพารามิเตอร์ 3 ตัว เรียกว่า โมเดลโลจิสติก 3 พารามิเตอร์ (Warm 1979 :19) ต่อมาในปี 1952 ลอร์ด (Lord) ได้เสนอทฤษฎีในรูปโค้งลักษณะของข้อสอบ (ICC) ของแต่ละข้อ โดยกล่าวว่าพฤติกรรมการตอบสนองข้อสอบแต่ละข้อของผู้สอบสามารถอธิบายได้ด้วยโค้งลักษณะซึ่งมีลักษณะเป็นโค้งปกติสะสม หรือที่เรียกว่า โมเดลโค้งปกติสะสม (Normal Ogive Model) ซึ่งโมเดลนี้จะกล่าวถึงพารามิเตอร์ 2 ตัว คือ ค่าความยาก ค่าอำนาจจำแนก แต่มีปัญหาในการนำแนวคิดของทฤษฎีนี้ไปใช้ เนื่องจากมีการคำนวณยุ่งยากและขาดแคลนคอมพิวเตอร์ที่จำเป็นต้องใช้วิเคราะห์ข้อมูลตามทฤษฎี ลอร์ดจึงหยุดความสนใจทฤษฎีนี้ไประยะหนึ่ง ต่อมาในปี 1960 ราสช์ (Rasch) ได้เสนอแนวคิดหนึ่งของทฤษฎีการตอบสนองข้อสอบ (IRT) ในรูปที่มีพารามิเตอร์ 1 ตัว คือ ค่าความยาก ซึ่งรูปแบบนี้เรียกกันทั่วไปว่า ราสช์โมเดล (Rasch Model) แนวคิดนี้ทำให้มีคนสนใจ IRT มากขึ้น จนในปี 1968 เบิร์นบอม (Birnbaum) ได้เสนอแนวคิดเกี่ยวกับโมเดลโลจิสติก (Logistic Model) ที่ใช้พารามิเตอร์ 2 ตัว คือ ค่าความยาก และค่าอำนาจจำแนก ซึ่งโมเดลนี้ง่ายกว่าของ ลอร์ด จึงทำให้โมเดลโลจิสติกเป็นที่นิยมแพร่หลาย

และมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งสามารถใช้กับพารามิเตอร์ 1 ตัว ,3 ตัว และ 4 ตัว

สำหรับแนวคิดของโมเดลโลจิสติกนั้น มาจากแนวคิดเดียวกับโมเดลโค้งปกติสะสม ซึ่งมุ่งที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (Y) แทนพฤติกรรมการตอบสนองต่อข้อสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคล และ r_i เป็นเกณฑ์ในการทำข้อสอบนั้นว่าถูกหรือผิด ดังนั้นถ้า $Y > r_i$ แสดงว่าผู้สอบทำข้อสอบข้อ i ได้ถูก และ $Y < r_i$ แสดงว่าผู้สอบทำข้อสอบข้อ i ผิด ความสัมพันธ์ระหว่างความสามารถ (θ) กับพฤติกรรมการตอบสนองข้อสอบ (Y) แสดงได้ดังภาพที่ 1



ภาพที่ 1 ความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบสนองต่อข้อสอบ

จากรูปภาพนี้แสดงว่า ถ้านำโอกาสการตอบถูก (พื้นที่ส่วนแรเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้โค้งลักษณะข้อสอบ (ICC) เป็นรูปต่าง ๆ ทั้งนี้ขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์ และจำนวนนารามิเตอร์ที่จะใช้อธิบายแต่ในการอธิบายโค้งลักษณะข้อสอบโดยใช้ฟังก์ชันโลจิสติก ซึ่งจะมีวิธีการคำนวณทางคณิตศาสตร์ได้ง่ายและสะดวกกว่าการใช้ฟังก์ชันโค้งปกติสะสม นอกจากนี้แล้วในสถานการณ์สอบจริงอาจจะมีผู้ที่มีความสามารถสูงตอบข้อสอบผิดด้วยความประมาท กรณีเช่นนี้ โมเดลโลจิสติกจะมีความแกร่งต่อข้อมูลแบบนี้มากกว่าโมเดลปกติสะสม จึงทำให้โมเดลโลจิสติกเป็นที่นิยมกันมากในทางปฏิบัติงานจริง (Lord 1980:14)

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

1. มิติของลาเท็นท์สเปซ (Dimensionality of the Latent Space)

คุณลักษณะภายในหรือความสามารถที่จะเป็นตัวกำหนดพฤติกรรมการตอบสนองต่อข้อสอบในข้อใดข้อหนึ่งอาจมีหลายลักษณะ และจำนวนคุณลักษณะทั้งหมดในลาเท็นท์สเปซ ก็คือ มิติของลาเท็นท์สเปซ ในทฤษฎีการตอบสนองข้อสอบโดยทั่วไปมักจะถือว่า ลักษณะหรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบสนองข้อสอบมีเพียงลักษณะเดียว ซึ่งยอมรับกันเป็นข้อตกลงเบื้องต้นที่สำคัญของแบบจำลองหลายแบบในทฤษฎีนี้ว่า ข้อสอบแต่ละข้อในแบบสอบวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือข้อสอบเหล่านี้มีลักษณะเป็นเอกพันธ์ (Homogeneous)

วิธีการตรวจสอบว่าข้อสอบชุดนี้วัดความสามารถหรือคุณลักษณะเดียวกันมีหลายวิธี ได้แก่การวิเคราะห์ตัวประกอบ (Factor Analysis) การเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง เนื่องจากข้อสอบที่มีค่าอำนาจจำแนกสูงจะมีสหสัมพันธ์กับคะแนนของแบบสอบทั้งหมดสูง ยูริ (Urry 1981) เสนอไว้ว่าควรเลือกข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ .80 ขึ้นไป (Green, et al. 1984:347)

2. ความเป็นอิสระในการตอบหรือสอบ (Local Independence) หมายถึง การตอบข้อสอบข้อต่าง ๆ ในแบบสอบของผู้สอบนั้นมีความเป็นอิสระ ซึ่งข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบสนองต่อข้อสอบนั้นมี 2 ลักษณะ คือ

2.1 ความเป็นอิสระทางสถิติ (Statistically Independence) กล่าวคือ ข้อสอบแต่ละข้อเป็นอิสระไม่เกี่ยวข้องกัน ดังนั้นคำตอบในแต่ละข้อของแต่ละคนเป็นอิสระต่อกัน แต่รวมกันแล้วจะวัดคุณลักษณะเดียวเท่านั้น

2.2 ความเป็นอิสระจากตำแหน่ง (Uncorrelated Independence) กล่าวคือ ข้อสอบแต่ละข้อจะปรากฏอยู่ในตำแหน่งใดของข้อสอบก็ได้ จะไม่มีผลต่อการตอบของข้อสอบ

ลอร์ด (Lord 1980: 19) กล่าวว่า เมื่อแบบสอบมีคุณสมบัติของการวัดเพียงมิติเดียวแล้วก็จะมีความเป็นอิสระในการตอบข้อสอบ ดังนั้นการตรวจสอบความเป็นอิสระในการตอบข้อสอบก็ใช้การวิเคราะห์ตัวประกอบได้เหมือนกัน

3. โค้งลักษณะข้อสอบ (Item characteristic curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะของข้อสอบมีหลายรูปแบบตามความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสที่จะตอบถูก เช่น Guttman Perfect Scale , Latent Distance Model , Linear Model , Logistic Model , Nominal Response Model และ Grade Response Model (จักรกฤษณ์ สาราणीใจ 2531:30-38) แต่รูปแบบที่นิยมมากที่สุดคือ โมเดลโลจิสติก (Logistic Model) รูปแบบนี้แบ่งออกเป็น 4 รูปแบบตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบดังนี้ (Hambleton and Swaminathan 1985 :37-49)

1. โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว (Three-Parameter Logistic Model)

เป็นโมเดลที่มีพารามิเตอร์อธิบายโค้งลักษณะข้อสอบ 3 ตัว ซึ่งมีรูปสมการดังนี้

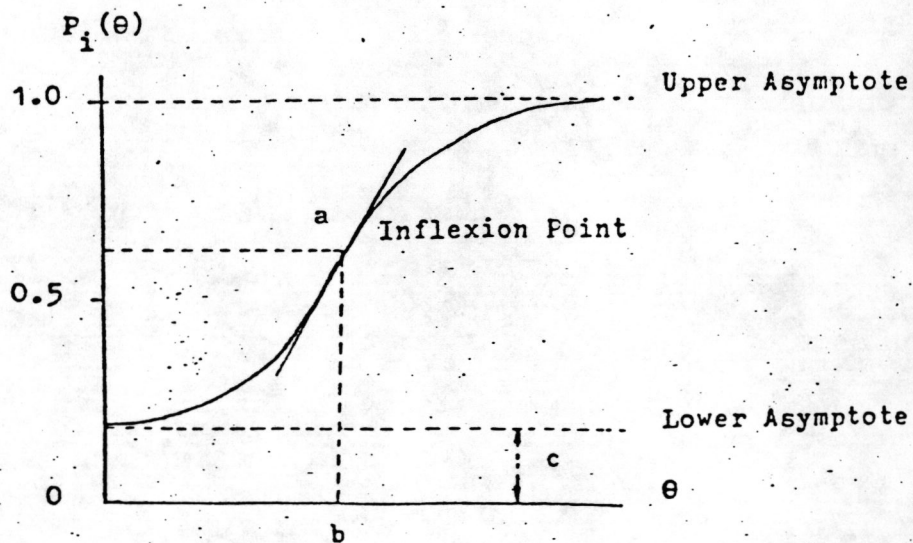
$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}}$$

$$\text{หรือ } P_i(\theta) = c_i + (1 - c_i) \left[1 + e^{-D a_i (\theta - b_i)} \right]^{-1}, \quad (i = 1, 2, \dots, n)$$

เมื่อ $P(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ ตอบข้อสอบข้อ i ถูกต้อง
 (θ) คือ ระดับความสามารถของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยมีพิสัยการกระจายอยู่ระหว่าง $-\infty$ ถึง $+\infty$

แต่ในทางปฏิบัติระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง $+3$ ซึ่งมีระดับความสามารถ -3 หมายถึง ผู้สอบมีระดับความสามารถต่ำมาก และระดับความสามารถ $+3$ หมายถึง ผู้สอบมีระดับความสามารถสูงมาก

- (ai) คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ θ . จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ โดยทั่วไปนิสัยของค่าอำนาจจำแนกของข้อสอบอยู่ระหว่าง 0 ถึง 2
- (bi) คือ ค่าความยากของข้อสอบข้อ i ที่แสดงถึงระดับความสามารถ ที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด โดยทั่วไปนิสัยขอบค่าความยากของข้อสอบระหว่าง -2 ถึง $+2$ ซึ่งค่าที่ใกล้ -2 หมายถึง ข้อสอบที่ง่ายมาก และค่าที่ใกล้ $+2$ หมายถึง ข้อสอบที่ยากมาก
- (ci) คือ ค่าการเดาของข้อสอบข้อ i ซึ่งเป็นค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมาก จะมีโอกาสตอบข้อสอบข้อ i ถูก หรือเป็นค่ากำกับโค้งที่ต่ำสุด (Lower asymptote) ของโค้งลักษณะข้อสอบมีค่า 0 ถึง 1
- D คือ a scaling factor มีค่าเท่ากับ 1.7
- e คือ ค่าคงที่ มีค่าเท่ากับ $2.71828\dots$



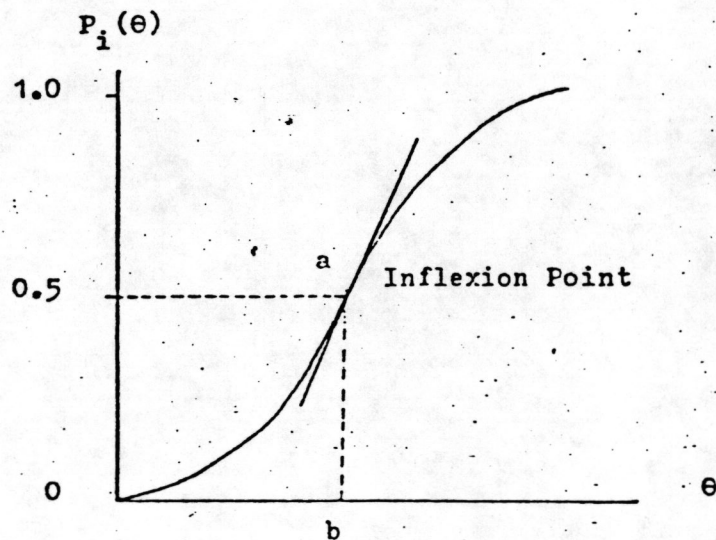
แผนภาพที่ 2 ความหมายของค่าพารามิเตอร์ของข้อสอบ 3 พารามิเตอร์

2. โมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว (Two-Parameter Logistic Model)

เป็นรูปแบบที่เบิร์นบอม (Birnbbaum 1957, 1958a, 1958b, 1968) เสนอไว้ แทนรูปแบบ Normal Ogive Mode เนื่องจาก รูปแบบโลจิสติกเป็นรูปแบบเชิงคณิตศาสตร์ ที่นำมาใช้ในทางปฏิบัติได้สะดวกกว่า รูปแบบนี้ถือว่าไม่มีการเดา ตั้งสมการ

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad \text{หรือ}$$

$$P_i(\theta) = \left[1 + e^{-Da_i(\theta-b_i)} \right]^{-1} ; (i = 1, 2, \dots, n)$$



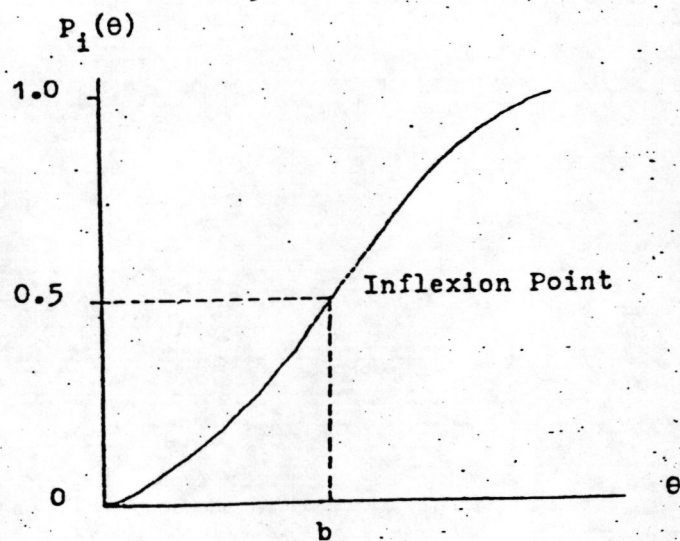
แผนภาพที่ 3

ความหมายของค่าพารามิเตอร์ของข้อสอบ 2 พารามิเตอร์

3. โมเดลโลจิสติกที่มีพารามิเตอร์ 1 ตัว (One-Parameter Logistic Model)
 เป็นโมเดลที่มีข้อตกลงเบื้องต้น คือ ค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเท่ากัน
 และค่าการเดาเท่ากับ 0 รูปสมการก็จะเหลือค่าพารามิเตอร์ความยากเพียง 1 ตัว มีที่มีค่า
 เปลี่ยนดังนี้

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \quad \text{หรือ}$$

$$P_i(\theta) = \left[1 + e^{-D(\theta - b_i)} \right]^{-1} \quad ; (i = 1, 2, \dots, n)$$



แผนภาพที่ 4 ความหมายของค่าพารามิเตอร์ของข้อสอบ 1 พารามิเตอร์

4. โมเดลโลจิสติกที่มีพารามิเตอร์ 4 ตัว (Four-Parameter Logistic Model)

รูปแบบนี้คำนึงถึงผู้สอบที่มีความสามารถสูง เนื่องจากผู้สอบที่มีความสามารถสูงบางครั้งอาจมีความสะเพร่าในการทำข้อสอบ จึงทำให้ตอบข้อสอบข้อนั้นผิดพลาดไปได้ แมคโดแนล (McDonald อ้างถึงใน Hambleton and Swaminathan 1985 :48) จึงเสนอรูปแบบนี้ขึ้น ดังสมการ

$$P_i(\theta) = c_i + (y_i - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad \text{หรือ}$$

$$P_i(\theta) = c_i + (y_i - c_i) \left[1 + e^{-Da_i(\theta - b_i)} \right]^{-1} \quad ; (i = 1, 2, \dots, n)$$

รูปแบบนี้ต่างจากรูปแบบ 3 พารามิเตอร์ ก็คือค่า y_i มีค่าต่ำกว่า 1 เล็กน้อย แต่อย่างไรก็ตามรูปแบบนี้ยังไม่สามารถนำมาใช้ในทางปฏิบัติได้

อนึ่งค่าพารามิเตอร์ต่าง ๆ ที่ได้จากการวิเคราะห์ข้อสอบรายข้อจากรูปแบบต่าง ๆ ข้างต้นนี้ เป็นค่าที่ไม่เปลี่ยนแปลงไปตามกลุ่มตัวอย่างและไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบคุณลักษณะดังกล่าวนี้คือ ความเป็นปรนัยของการวัดผล (objectivity of Measurement) ที่มีความสำคัญมากในทางวัดผล (Wright and Stones อ้างถึงใน สุนันต์ สุขมลสันต์ 2528: 49)

การตรวจสอบคุณภาพของแบบสอบตามแนวทฤษฎีการตอบสนองข้อสอบ

ในทฤษฎีการตอบสนองข้อสอบ จะใช้ผลการตอบแบบสอบประมาณค่าความสามารถของบุคคล การจะประเมินคุณภาพของแบบสอบอาจดูได้จากความถูกต้องแม่นยำในการประมาณค่าความสามารถโดยใช้คะแนนจากแบบสอบ ค่าสารสนเทศจากแบบสอบจะเป็นดัชนีชี้ถึงความถูกต้องแม่นยำในการประมาณค่า (Birnbaum 1968: 418) โดยที่ค่าฟังก์ชันสารสนเทศข้อสอบหรือแบบสอบเป็นดัชนีผสม (Composite index) ที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลายลักษณะ เช่น ค่าความยาก ค่าอำนาจจำแนก เป็นต้น รวมเป็นดัชนีเพียงตัวเดียว เพื่อชี้ถึงคุณภาพของข้อสอบหรือแบบสอบได้ และด้วยคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ ค่าฟังก์ชันสารสนเทศจึงเหมาะที่จะใช้เป็นดัชนีบอกคุณภาพของข้อสอบหรือแบบสอบตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) แทนการหาค่าความเที่ยงของคะแนนและความคลาดเคลื่อนมาตรฐานของการวัด ตามแนวคิดของทฤษฎีคลาสสิกอล (Hambleton 1977: 64)

ค่าฟังก์ชันสารสนเทศของแบบสอบ

ในสถานการณ์ทั่วไป ถ้ามั่นใจว่าเหตุการณ์อันหนึ่งจะเกิดขึ้นค่อนข้างมากแสดงว่ามีข่าวสารข้อมูลหรือสารสนเทศเกี่ยวกับเหตุการณ์นั้นมากพอสมควร ในทางตรงกันข้ามถ้าไม่มีข่าวสารเกี่ยวกับเหตุการณ์นั้นหรือมีน้อย ความมั่นใจก็จะน้อยตามไปด้วย ในกระบวนการอ้างอิงเชิงสถิติ ความแม่นยำของการประมาณค่าพารามิเตอร์ของกลุ่มประชากรอาจดูได้จากช่วงกว้างของค่าประมาณ ถ้าไม่มีสารสนเทศใด ๆ เกี่ยวกับประชากรเลย ก็อาจจะต้องประมาณค่าเป็นค่าใด ๆ ในช่วง $-\infty$ ถึง $+\infty$ ถ้ามีสารสนเทศเกี่ยวกับประชากรบ้าง ช่วงของการประมาณค่าพารามิเตอร์จะแคบเข้า นั้นหมายความว่า ความแม่นยำในการประมาณค่าเริ่มจะมีมากขึ้น ตามปกติความแม่นยำในการประมาณค่าพารามิเตอร์จะแสดงด้วยค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า คือ ถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีมาก ความแม่นยำของการประมาณก็จะน้อย เพราะช่วงของค่าประมาณจะกว้าง ในทางกลับกันถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่าน้อย ความแม่นยำของการประมาณค่าก็จะมาก ช่วงของค่าประมาณจะแคบ แสดงว่าค่าสารสนเทศมีความสัมพันธ์กับความคลาดเคลื่อนมาตรฐานของการประมาณค่า โดยมีความสัมพันธ์ในทิศทางกลับกัน คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าเท่ากับ $1/\sqrt{\text{สารสนเทศ}}$

โดยที่ค่าสารสนเทศคือ ปริมาณที่เป็นสัดส่วนกลับของกำลังสองของความกว้างของช่วงความเชื่อมั่นของค่าความสามารถด้วยคะแนนจากแบบสอบ (Lord 1980: 65) ถ้าให้ y เป็นคะแนนที่ได้จากการตอบแบบสอบ ค่าสารสนเทศสำหรับคะแนน y ด้วยสมการ

$$I(\theta, Y) = \frac{\left[d \mu_y | \theta \right]^2}{\text{Var}(Y | \theta)} \quad (\text{Lord 1980: 67})$$

จากสูตรจะเห็นว่าค่าสารสนเทศสำหรับคะแนน y ก็คือ กำลังสองของอัตราส่วนระหว่างความชันของเส้นถดถอยของ y บน θ กับความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถ θ นั้นเอง ดังนั้นค่าสารสนเทศของแบบสอบฉบับหนึ่งมีหลายค่าแตกต่างกันไปตามระดับของ θ และที่ระดับความสามารถค่าใดค่าหนึ่ง ค่าสารสนเทศจะมากหรือน้อยขึ้นอยู่กับ

ก. ค่าความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถนั้น ซึ่งถ้ามีความคลาดเคลื่อนน้อย ค่าสารสนเทศจาก y เกี่ยวกับ θ ก็จะมาก

ข. ความชันของเส้นถดถอยของ y บน θ ซึ่งถ้ามีความชันมาก ค่าสารสนเทศจาก y เกี่ยวกับ θ ก็จะมาก (Lord 1980: 68)

นอกจากนี้ค่าสารสนเทศของแบบสอบจะเพิ่มมากขึ้นเมื่อแบบสอบมีข้อสอบจำนวนมากขึ้น (Hambleton and Swaminathan 1985: 107)

สำหรับโมเดลโลจิสติก 3 พารามิเตอร์ จะสามารถหาค่าฟังก์ชันสารสนเทศข้อสอบในแต่ละระดับความสามารถ (θ) ได้จากสูตร (Lord 1980: 72-74)

$$I(\theta, u_i) = \frac{(1.7a_i)^2 (1-c_i)}{\left[\frac{1.7a_i (\theta - b_i)}{c_i + e} \right] \left[\frac{-1.7a_i (\theta - b_i)}{1 + e} \right]^2}$$

และสามารถกำหนดโค้งสารสนเทศข้อสอบได้จากสมการ

$$I(\theta, U_i) = (P_i')^2 / P_i Q_i \dots\dots\dots(1)$$

- $I(\theta, U_i)$ คือ ค่าฟังก์ชันสารสนเทศข้อสอบ
 P_i คือ ความชันของ ICC ที่ระดับความสามารถ θ
 P_i' คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง
 Q_i คือ $1 - P_i$

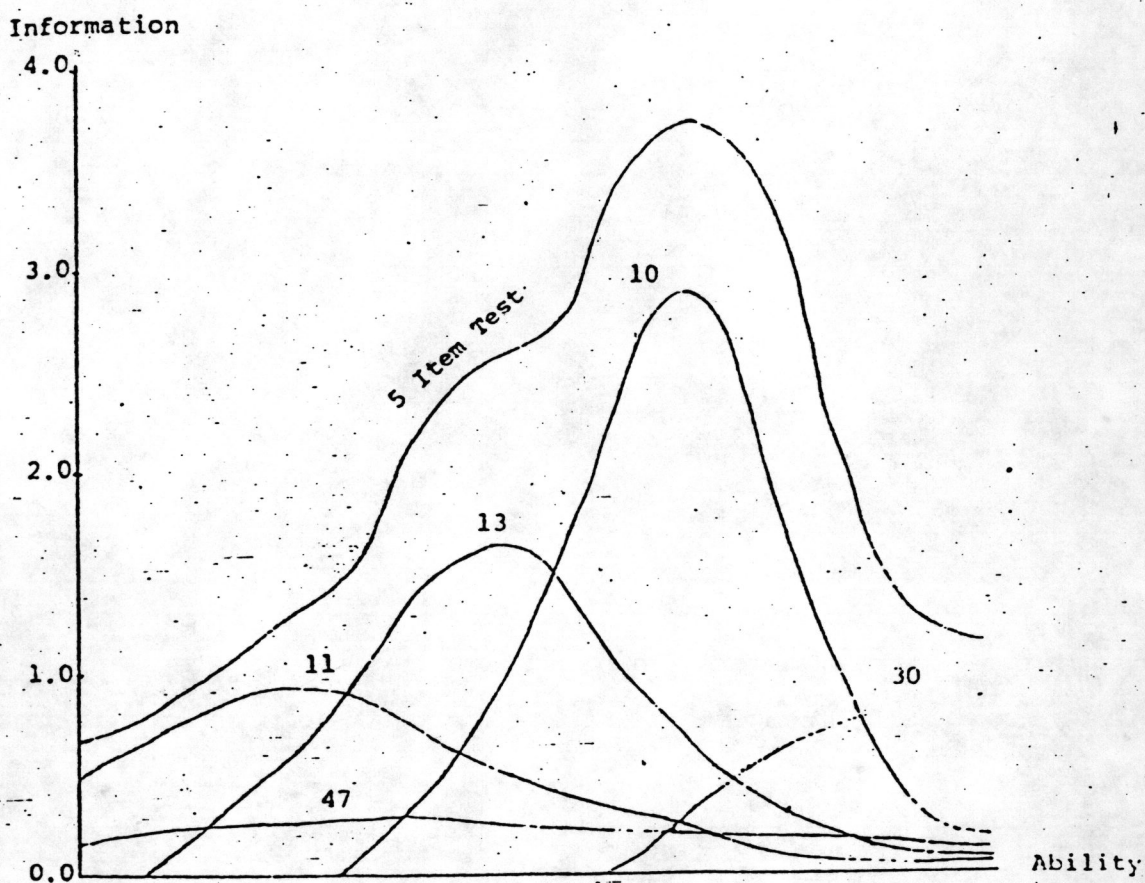
และสามารถหาค่าฟังก์ชันสารสนเทศแบบสอบได้จากผลรวมของฟังก์ชันสารสนเทศข้อสอบทั้งหมดในแบบสอบเขียนเป็นสมการได้ดังนี้

$$I(\theta) = \sum I(\theta, U_i) \dots\dots\dots(2)$$

เมื่อ $I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศแบบสอบ จากสมการ (1) จะเห็นได้ว่าค่าฟังก์ชันสารสนเทศแต่ละข้อขึ้นอยู่กับความชันของโค้งลักษณะข้อสอบ และความแปรปรวนของการตอบข้อสอบถูกต้องของแต่ละข้อในแต่ละระดับความสามารถ (θ) และยิ่งความชันของ ICC มีค่ามาก ประกอบกับค่าความแปรปรวนมีค่าน้อย ๆ โค้งสารสนเทศข้อสอบที่ระดับความสามารถนั้นจะยิ่งสูงขึ้น สำหรับโค้งสารสนเทศข้อสอบที่มีค่าสูงสุด ณ ระดับความสามารถใดก็จะจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton 1977: 66 อ้างถึงใน ประดิษฐ์ เรื่องตระกูล 2529: 23-24)

สำหรับค่าฟังก์ชันสารสนเทศของแบบสอบฉบับหนึ่งจะมีหลายค่าแตกต่างกันไปตามระดับของ θ และ ณ ระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศแบบสอบที่มีค่าสูงแสดงว่า แบบสอบมีคุณภาพดีมากในการประมาณค่าความสามารถที่ระดับนั้น แอมเบิลตัน และสวามิทาน (Hambleton and Swaminathan 1985: 107) กล่าวว่า เมื่อแบบสอบมีข้อสอบจำนวนมากขึ้น ค่าฟังก์ชันสารสนเทศแบบสอบก็จะเพิ่มมากขึ้น

ดังนั้นถ้าเรามีกลุ่มข้อสอบที่ทราบโค้งสารสนเทศ เราจะสามารถสร้างแบบสอบให้มีโค้งสารสนเทศแบบสอบ ๗ ระดับหนึ่งของความสามารถที่เราต้องการได้ เช่น สอบคัดเลือก ก็ต้องใช้ข้อสอบที่มีประสิทธิภาพสูงสุด ที่ระดับความสามารถสูง ๆ นั่นคือ ให้มีโค้งสารสนเทศแบบสอบสูง ๗ ระดับความสามารถสูง ๆ นั่นเอง



แผนภาพที่ 5 แสดงโค้งสารสนเทศข้อสอบและโค้งสารสนเทศแบบสอบของข้อสอบ 5 ข้อ (Lord 1980: 22)

ในการสอบทุกครั้ง จะมีความคลาดเคลื่อนในการวัดของแบบสอบ และสำหรับค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าสามารถ (The Standard Error of Estimate Ability : S.E.E) หมายถึง ส่วนเบี่ยงเบนมาตรฐานในการวัดของการประมาณค่าความสามารถ (θ) ในกลุ่มผู้สอบที่มีความสามารถแตกต่างกันไป ซึ่งประมาณค่าได้จากสูตร (Warm 1987: 77)

$$S.E.E = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (3)$$

ความคลาดเคลื่อนมาตรฐานในการประมาณค่านี้ขึ้นอยู่กับและผันแปรไปตามฟังก์ชันของข้อสอบและกลุ่มผู้สอบที่ใช้ ยิ่งความยากของข้อสอบเหมาะสมกับความสามารถของผู้สอบ และจำนวนข้อและจำนวนผู้สอบยิ่งมาก ความคลาดเคลื่อนมาตรฐานจะยิ่งต่ำลง ค่าฟังก์ชันสารสนเทศแบบสอบจะมีค่ามาก จะทำให้ค่า S.E.E มีค่าน้อย ซึ่งเป็นค่าที่เราต้องการมากในแบบสอบ

ประสิทธิภาพของแบบสอบและการเปรียบเทียบ

ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ

ดัชนีที่บ่งถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถตามแนวทฤษฎีการตอบสนองข้อสอบก็คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ ซึ่งจุดประสงค์ที่จัดให้มีการสอบขึ้นมานั้นก็คือ ต้องการให้ผลจากการสอบนำไปประมาณค่าความสามารถที่มีอยู่ภายในตัวบุคคล ซึ่งไม่สามารถสังเกตได้โดยตรง ดังนั้นการประเมินคุณภาพของแบบสอบจึงต้องพิจารณากันที่ความถูกต้องแม่นยำของการใช้คะแนนสอบประมาณค่าระดับความสามารถ ค่าฟังก์ชันสารสนเทศเป็นดัชนีที่ใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำมากที่สุด โดยกำหนดว่า ณ ระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศแบบสอบมีค่าสูง แสดงว่าแบบสอบมีคุณภาพที่ดีในการประมาณค่าความสามารถที่ระดับนั้น และนอกจากใช้พิจารณาคุณภาพของแบบสอบแล้ว ยังสามารถนำมาเปรียบเทียบคุณภาพของแบบสอบได้อีกด้วย

ในกรณีที่ต้องการเปรียบเทียบประสิทธิภาพของแบบสอบ 2 ฉบับที่ใช้วัดลักษณะหรือความสามารถอันเดียวกันก็สามารถหาค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของแบบสอบทั้งสองได้ โดยการนำเอาค่าฟังก์ชันสารสนเทศแบบสอบของแต่ละแบบสอบที่ต้องการเปรียบเทียบ 2 ฉบับ ณ ระดับความสามารถเดียวกัน มาหาค่าอัตราส่วนกัน ค่าอัตราส่วนของค่าฟังก์ชันสารสนเทศแบบสอบ 2 ฉบับที่นำมาเปรียบเทียบกัน ณ ระดับความสามารถเดียวกัน

ค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ y กับคะแนนแบบสอบ x หมายถึง อัตราส่วนของค่าฟังก์ชันสารสนเทศแบบสอบทั้ง 2 ฉบับ ดังสมการ

$$RE(y,x) = \frac{I(\theta,y)}{I(\theta,x)}$$

ค่าประสิทธิภาพสัมพัทธ์นี้ไม่ขึ้นอยู่กับมาตราที่ใช้วัดความสามารถ นั่นคือ ค่าดัชนีจะไม่แปรเปลี่ยนไปตามแปลงมาตรฐานของการวัดความสามารถ (Lord 1980: 89)

โดยที่คะแนน x และ y เป็นคะแนนจากแบบสอบที่แตกต่างกัน 2 ฉบับ ณ ระดับความสามารถเดียวกัน หรือ x และ y เป็นผลจากวิธีการให้คะแนนที่แตกต่างกัน 2 วิธีของแบบสอบฉบับเดียวกัน ตามนิยามของค่าประสิทธิภาพสัมพัทธ์ของแบบสอบนั้น ค่าความสามารถใน $I(\theta,y)$ ต้องเป็นค่าความสามารถเดียวกันใน $I(\theta,x)$ และค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ 2 ฉบับ มีค่าแปรเปลี่ยนไปตามระดับความสามารถด้วย (Lord 1980: 85)

การแปลความหมายของค่าประสิทธิภาพสัมพัทธ์

ณ ระดับความสามารถใด ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่ามากกว่า 1 แสดงว่า แบบสอบ y มีคุณภาพสูงกว่าแบบสอบ x ที่ระดับความสามารถนั้น ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าน้อยกว่า 1 แสดงว่า แบบสอบ y มีคุณภาพต่ำกว่าแบบสอบ x ที่ระดับความสามารถนั้น และถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าเท่ากับ 1 แสดงว่า แบบสอบ y และแบบสอบ x มีคุณภาพเท่ากัน ที่ระดับความสามารถนั้น (Warm 1978: 76)

ความแตกต่างระหว่างทฤษฎีคลาสสิกอลกับทฤษฎีการตอบสนองข้อสอบ

การเปรียบเทียบความแตกต่างระหว่างทฤษฎีคลาสสิกอลกับทฤษฎีการตอบสนองข้อสอบ ได้สรุปเป็นประเด็นสำคัญ ซึ่งแสดงด้วยตารางที่ 2 ดังนี้ (อวยพร วิบูลกาญจน์ 2526: 24)

ตารางที่ 2 ความแตกต่างระหว่างทฤษฎีคลาสสิกอลกับทฤษฎีการตอบสนองข้อสอบ

	ทฤษฎีคลาสสิกอล	ทฤษฎีการตอบสนองข้อสอบ
ลักษณะ	เป็นโมเดลทางคณิตศาสตร์ที่แสดงถึงความสามารถเคลื่อนของการวัดที่มีผลต่อคะแนนที่วัดได้หรือสังเกตได้ เป็นการเน้นการประมาณค่าคะแนนจริงหรือความสามารถที่แท้จริง โดยใช้คะแนนของกลุ่มผู้สอบเป็นสำคัญ	เป็นโมเดลทางคณิตศาสตร์ที่อธิบายถึงความสัมพันธ์ระหว่างการตอบข้อสอบของผู้สอบกับระดับความสามารถที่มีอยู่ภายในตัวผู้สอบ เป็นการเน้นการประมาณค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของความสามารถของผู้สอบ
ค่าความยาก	สัดส่วนของผู้สอบทั้งหมดที่ตอบข้อสอบแต่ละข้อถูกต้อง มีค่าตั้งแต่ 0 ถึง 1	ค่าที่แสดงถึงระดับความสามารถที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด โดยทั่วไปมีค่าอยู่ระหว่าง -2 ถึง 2
อำนาจจำแนก	ประสิทธิภาพของข้อสอบที่จะสามารถแยกความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ มีตั้งแต่ -1 ถึง 1	ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ โดยทั่วไปมีค่าอยู่ระหว่าง 0 ถึง 2

ตารางที่ 2 (ต่อ)

	ทฤษฎีคลาสสิกคอล	ทฤษฎีการตอบสนองข้อสอบ
ค่า การ เดา	อัตราส่วนของ 1 กับจำนวนตัวเลือก ในข้อสอบแต่ละข้อของแบบสอบ	ค่าความน่าจะเป็นที่ผู้สอบมีความ สามารถต่ำมาก มีโอกาสตอบข้อ สอบข้อนั้นถูก หรือเป็นค่าเท่ากับ โค้งลักษณะข้อสอบมีค่า 0 ถึง 1
ความ ตรง	ความสามารถในการวัดของสิ่งที่เรา ต้องการจะวัดได้ถูกต้อง มีค่าตั้งแต่ 0 ถึง 1	เป็นการใช้แบบสอบให้เป็นไปตาม วัตถุประสงค์ที่ต้องการของการสอบ โดยพิจารณาจากค่าฟังก์ชันสารสนเทศ แบบสอบ ซึ่งเป็นค่าฟังก์ชันของค่า พารามิเตอร์ของข้อสอบตามโมเดลที่ ใช้ มีค่าตั้งแต่ 0 ขึ้นไป
ความ เที่ยง	ความคงที่ของคะแนนที่ได้จากการ ทดสอบคนกลุ่มเดียวกันสองครั้งด้วย แบบสอบเดิมในเวลาที่แตกต่างกัน หรือ ทดสอบกลุ่มเดียวกันด้วยข้อสอบต่าง ชุดที่มีข้อสอบเทียบเท่ากัน หรือภาย ได้สถานการณ์สอบที่แตกต่างกัน มีค่า ตั้งแต่ 0 ถึง 1	ความถูกต้องแม่นยำของการประมาณ ค่าความสามารถที่แท้จริงของผู้สอบ โดยพิจารณาจากค่าฟังก์ชันสารสนเทศ แบบสอบ

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

3.1 การรอบทฤษฎีคลาสสิกอล

การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกที่แตกต่างกัน ตามกรอบทฤษฎีคลาสสิกอล ได้มีผลการศึกษาวิจัยของนักวัดและประเมินผลการศึกษาทั้งของต่างประเทศและไทย ซึ่งผลงานวิจัยดังกล่าวจะสรุปได้ดังนี้

งานวิจัยต่างประเทศ

เวสแมนต์ และเบนเนตต์ (Wesman and Benett 1943: 151-154) ได้ศึกษาผลการใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ในวิชาคณิตศาสตร์ และศัพท์ภาษาอังกฤษ ปรากฏว่าไม่พบความแตกต่างในด้านความยากของแบบสอบทั้ง 2 ชนิด แต่ในข้อสอบบางข้อพบว่า ข้อสอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" มีประสิทธิภาพมากและผู้วิจัยทั้งสองยังได้สรุปเพิ่มเติมอีกว่า การใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" จะมีประสิทธิภาพน้อยกว่าเพียงใดขึ้นอยู่กับคุณภาพของตัวเลือกตัวอื่น ๆ ด้วย

บอยนตัน (Boynton 1950: 565-570) ศึกษาถึงประสิทธิภาพของแบบสอบแบบเลือกตอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ในการสะกดคำ พบว่า ข้อสอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" มีระดับค่าความยาก ค่าอำนาจจำแนก สูงกว่าแบบสอบที่ไม่ได้ใช้ตัวเลือกแบบนี้

ริมแลนด์ (Rimland 1960: 533-539) ได้ทำการศึกษาเกี่ยวกับคุณภาพของแบบสอบเลือกตอบที่ใช้ตัวเลือกแบบธรรมดา และตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ในวิชาคณิตศาสตร์ โดยทำการศึกษากับกลุ่มตัวอย่าง 2 กลุ่ม ๆ ละ 3,600 คน โดยกลุ่มที่ 1 ทำแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา ส่วนกลุ่มที่ 2 ทำแบบสอบที่มีรูปแบบตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" โดยให้ 1 ใน 5 ของจำนวนข้อสอบในแบบสอบที่มีตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นคำตอบถูก ผลการศึกษาในครั้งนี้ ปรากฏว่า การใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ไม่ทำให้คุณภาพของแบบสอบในด้านค่าความยาก ค่าอำนาจจำแนก แตกต่างไปจากแบบสอบที่ใช้ตัวเลือกแบบธรรมดา

ฮิวส์ และทริมเบิล (Hughes and Trimble 1965: 117-125) ได้ทำการศึกษาวิจัยเกี่ยวกับคุณภาพของแบบสอบแบบเลือกตอบที่ใช้รูปแบบตัวเลือกที่แตกต่างกัน 3 รูปแบบ คือ รูปแบบตัวเลือกแบบธรรมดา แบบปลายเปิดและแบบปลายปิด และแบบผสม "ถูกทั้งข้อ 1 และข้อ 2" หรือ "ผิดทั้งข้อ 1 และข้อ 2" ในวิชาจิตวิทยาทั่วไป โดยศึกษากับนักศึกษาระดับปริญญาตรีปีที่ 1 จำนวน 63 คน แบ่งออกเป็น 3 กลุ่ม ให้ทำแบบสอบแต่ละฉบับ โดยกลุ่มที่ 1 ทำแบบสอบที่ใช้ตัวเลือกแบบธรรมดา กลุ่มที่ 2 ทำแบบสอบที่ใช้ตัวเลือกแบบปลายเปิดและแบบปลายปิด และกลุ่มที่ 3 ทำแบบสอบที่ใช้ตัวเลือกแบบผสมโดยที่แบบสอบแต่ละชุดจะประกอบด้วยข้อสอบจำนวน 55 ข้อ แบ่งออกเป็น 3 ส่วน คือ ส่วนแรก ใช้เป็นข้อสอบควบคุม ได้แก่ ข้อสอบที่มีตัวเลือกแบบธรรมดาเหมือนกันทั้ง 3 ชุด จำนวน 15 ข้อ ส่วนที่ 2 เป็นข้อสอบที่ใช้เปรียบเทียบจริง ได้แก่ ข้อสอบที่ใช้ตัวเลือกแบบธรรมดา หรือแบบปลายเปิดและแบบปลายปิด หรือแบบผสมต่างกันในแต่ละฉบับ จำนวน 35 ข้อ และส่วนที่ 3 เป็นข้อสอบที่ใช้ตัวเลือกแบบต่าง ๆ เป็นคำตอบถูกจำนวน 5 ข้อ ส่วนนี้ไม่ได้นำมาเปรียบเทียบ แต่เพื่อให้เห็นความสำคัญของตัวเลือกซับซ้อนบ้างไม่ใช่เป็นตัวเลือกอย่างเดียว ผลการวิจัยปรากฏว่า รูปแบบของตัวเลือกแบบต่าง ๆ มีผลต่อค่าความยาก แต่ไม่มีผลต่อค่าอำนาจจำแนกของแบบสอบ และเมื่อเปรียบเทียบคะแนนเฉลี่ยของนักเรียนปรากฏว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยคะแนนจากการสอบของนักเรียนที่ทำแบบสอบเลือกตอบแบบธรรมดาสูงที่สุด

มุลเลอร์ (Mueller 1975: 135-141) ได้ทำการศึกษาในทำนองเดียวกันกับฮิวส์ และทริมเบิล โดยใช้กลุ่มตัวอย่างที่เป็นนักศึกษาในมหาวิทยาลัยอินเดียนา จำนวน 2,000 คน ในวิชาเกี่ยวกับการประกอบอาชีพเป็นพนักงานขายของ ปรากฏว่า ผลการวิจัยที่ได้สอดคล้องกับการวิจัยของฮิวส์ และทริมเบิล คือ พบว่ารูปแบบตัวเลือกมีผลต่อค่าความยากของข้อสอบเช่นกัน โดยข้อสอบที่ใช้ตัวเลือกแบบผสมยากที่สุด รองลงมาคือ ข้อสอบที่ใช้ตัวเลือกแบบปลายเปิดและข้อสอบที่ใช้ตัวเลือกแบบธรรมดาง่ายที่สุด แต่ไม่มีผลต่อค่าอำนาจจำแนกของแบบสอบ นอกจากนี้ยังได้เปรียบเทียบตัวเลือกแบบต่าง ๆ เป็นคำตอบถูกด้วย พบว่า ข้อสอบที่ใช้ตัวเลือกแบบผสมยากที่สุด รองลงมาคือ ข้อสอบที่ใช้ตัวเลือกแบบปลายเปิด แต่ข้อสอบที่ง่ายที่สุดคือ ข้อสอบที่ใช้ตัวเลือกแบบปลายเปิดและตัวเลือกที่มีประสิทธิภาพน้อยที่สุด คือ ตัวลวงแบบธรรมดา

ฟอร์ซิท และสปรัทท์ (Forsyth and Spratt 1980: 31-43) ได้ศึกษาถึงผลของการเลือกข้อคำถามและคุณลักษณะของแบบสอบเลือกตอบที่วัดความสามารถในการแก้ปัญหาทางคณิตศาสตร์ โดยใช้แบบสอบโจทย์ปัญหาที่ปรับปรุงมาจากแบบสอบ ITBS (Iowa Test of Basic Skills) ของนักเรียนในระดับเกรด 7 และ เกรด 8 จำนวน 20 ข้อ ที่มีตัวเลือกแตกต่างกัน 4 รูปแบบคือ รูปแบบตัวเลือกแบบธรรมดา รูปแบบตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" รูปแบบตัวเลือกแบบกระบวนการคิด และรูปแบบตัวเลือกแบบกระบวนการคิดที่ระบุว่า "ไม่มีข้อใดถูก" โดยใช้กลุ่มตัวอย่างนักเรียนเกรด 7 และเกรด 8 จากโรงเรียนมัธยมศึกษา 3 โรงเรียน จำนวน 988 คน ซึ่งแบ่งออกเป็น 4 กลุ่ม โดยในแต่ละกลุ่มทำแบบเพียงฉบับเดียว ผลการวิจัยปรากฏว่า แบบสอบที่มีรูปแบบตัวเลือกแบบกระบวนการคิดยากกว่าแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา และตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ทำให้แบบสอบยากขึ้นแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดามีค่าอำนาจจำแนกสูงกว่าแบบสอบที่มีรูปแบบตัวเลือกเป็นกระบวนการคิด และแบบสอบที่มีตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ทำให้ความแตกต่างของค่าอำนาจจำแนกของแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา และที่เป็นกระบวนการคิดลดลง นอกจากนี้ยังพบว่า แบบสอบที่มีตัวเลือกแบบธรรมดามีความตรงเชิงโครงสร้างสูงกว่า แบบสอบที่มีตัวเลือกเป็นกระบวนการคิดเมื่อใช้แบบสอบ ITBS เป็นเกณฑ์

ออสเตอร์ฮอฟ และพามาเลา (Oosterhof and Pamela 1964: 287-295) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีตัวเลือกแตกต่างกัน 3 แบบ กับแบบสอบเติมคำในวิชาคณิตศาสตร์ที่เป็นโจทย์ปัญหา แบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกัน 3 รูปแบบคือ ตัวเลือกแบบธรรมดา ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" และตัวเลือกเป็นช่วง โดยแบบสอบแต่ละฉบับมีข้อสอบจำนวน 12 ข้อ โดยทำการศึกษากับกลุ่มตัวอย่างที่เป็นนักศึกษาธุรกิจ การศึกษาจำนวน 232 คน ผลการวิจัยพบว่า ค่าความยากของแบบสอบเลือกตอบทั้ง 3 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนแบบสอบเติมคำยากกว่าแบบสอบเลือกตอบทั้ง 3 ฉบับ อย่างมีนัยสำคัญทั้ง 3 ฉบับ และผลการศึกษาค้นคว้าพบว่า การใช้แบบสอบแบบปลายเปิด "ไม่มีข้อใดถูก" ทำให้แบบสอบยากกว่าแบบสอบที่มีตัวเลือกแบบธรรมดาและตัวเลือกแบบช่วง ถ้าใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นตัวถูกทำให้ข้อสอบยากขึ้น เพื่อพิจารณาถึงความเที่ยงของแบบสอบทั้ง 3 ฉบับ ที่คำนวณโดยใช้สูตร KR-20 พบว่า แบบสอบเติมคำมีความเที่ยงสูงสุด ส่วนแบบสอบเลือกตอบทั้ง 3 ฉบับ มีค่าความเที่ยงใกล้เคียงกัน โดยแบบสอบที่มีตัวเลือกแบบธรรมดามีค่าความเที่ยงสูงสุด รองลงมาคือ แบบสอบเลือกตอบแบบปลายเปิด "ไม่มีข้อใดถูก" และแบบสอบเลือกตอบที่มีตัวเลือกแบบช่วงตามลำดับ

แกรนดรู (Ghandour 1985: 515-A) ได้ทำการศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีลักษณะของตัวเลือกแตกต่างกัน 3 รูปแบบคือ แบบสอบที่มีตัวเลือกแบบธรรมดา แบบสอบที่มีตัวเลือกแบบปลายเปิด "ถูกทุกข้อ" และแบบสอบที่มีตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ในด้านค่าความยาก และค่าความเที่ยงของวิชาคณิตศาสตร์ ซึ่งแยกเนื้อหาออกเป็น แคลคูลัส แอลจีบรา (Algebra) และเรขาคณิต กลุ่มตัวอย่างเป็นนักศึกษาจำนวน 565 คน เป็นชาย 274 คน หญิง 291 คน โดยแยกการทดลองเป็นแบบ Factorial $2 \times 3 \times 3$ แบ่งออกเป็น 3 แฟคเตอร์ คือ เพศ เนื้อหา และรูปแบบของตัวเลือก แบบสอบแต่ละฉบับมีข้อสอบฉบับละ 18 ข้อ ในการตั้งสมมุติฐานเกี่ยวกับความเที่ยง 7 ข้อ เกี่ยวกับความยาก 7 ข้อ และสมมุติฐานเกี่ยวกับค่าอำนาจจำแนก 4 ข้อ ผลการวิจัยพบว่า ความเที่ยงของแบบสอบที่มีรูปแบบตัวเลือกแตกต่างกัน มีค่าความแตกต่างทั้งในลักษณะรวมเนื้อหาและแยกเนื้อหาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อพิจารณาค่าความยาก ปรากฏว่า มีความแตกต่างกันระหว่างเพศในกรณีแยกเนื้อหาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าความยากที่เกิดจากปฏิสัมพันธ์ (Interaction) ระหว่างรูปแบบของตัวเลือกกับเนื้อหาวิชา มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าอำนาจจำแนกมีความแตกต่างกันระหว่างเพศในวิชาเรขาคณิต และกับรูปแบบของแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดาเท่านั้น

โทเลฟสัน (Talefson 1987: 377-383) ได้ศึกษาและเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบของตัวเลือกแตกต่างกัน 3 รูปแบบคือ แบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา แบบสอบที่มีตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นตัวถูก และแบบสอบที่มีตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นตัวลวง แบบสอบทั้ง 3 ฉบับ มีข้อสอบฉบับละ 12 ข้อ ในเนื้อหาวิชาสถิติ กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาจำนวน 82 คน ผลการวิจัยพบว่า ค่าความยากของแบบสอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ยากกว่าแบบสอบที่ใช้ตัวเลือกแบบธรรมดาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ถ้าเปรียบเทียบระหว่างแบบสอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นตัวถูก กับแบบสอบที่ใช้ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" เป็นตัวลวง ปรากฏว่า ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" ที่เป็นตัวถูกยากกว่าตัวเลือกปลายเปิด "ไม่มีข้อใดถูก" ที่เป็นตัวลวง เมื่อพิจารณาค่าอำนาจจำแนกแล้วพบว่าแบบสอบที่มีตัวเลือกแบบธรรมดามีค่าอำนาจจำแนกสูงที่สุด แต่เมื่อทดสอบ ความแตกต่างทางสถิติ ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ ส่วนค่าความเที่ยงของแบบสอบทั้ง 3 ฉบับที่คำนวณโดยใช้สูตร KR-20 ปรากฏว่า แบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดามีค่าความเที่ยงสูงสุด

จากผลงานการวิจัยที่เกี่ยวกับคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันในต่างประเทศดังกล่าวข้างต้นจะเห็นได้ว่า ส่วนใหญ่เป็นการศึกษาคุณภาพของแบบสอบในวิชาคณิตศาสตร์ ส่วนในวิชาอื่น ๆ มีน้อยมาก และจากการศึกษาข้อค้นพบเหล่านี้พบว่าคุณภาพของแบบสอบในด้านต่าง ๆ นั้นยังไม่สอดคล้องกัน เช่น รีมแลนด์พบว่า การใช้ตัวเลือก "ไม่มีข้อใดถูก" ไม่ทำให้คุณภาพของแบบสอบในด้านค่าความยาก อำนาจจำแนกต่างจากแบบสอบที่มีรูปแบบบรรทัด แต่จากการศึกษาของฮิวส์และทริมเบิล (Hughes and Trimble) พบว่า แบบสอบที่มีรูปแบบตัวเลือกแบบบรรทัด เป็นต้น จึงไม่อาจทำให้ข้อสรุปที่แน่นอนได้

งานวิจัยภายในประเทศ

ไพบูลย์ จิตรโต (2514: 103-105) ได้เปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกต่างกัน 3 รูปแบบคือ ตัวเลือกแบบปลายเปิด "ไม่มีข้อใดถูก" แบบตัวลวงเจียด และแบบตัวลวงดับกับแบบสอบแบบเติมคำ ผลการวิจัยพบว่า ค่าความตรงต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในแบบสอบคณิตศาสตร์ทักษะ และคณิตศาสตร์ปัญหาชั้นประถมศึกษาปีที่ 7 มีความยากต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนแบบสอบคณิตศาสตร์ทักษะและคณิตศาสตร์ชั้นประถมศึกษาปีที่ 4 และชั้นประถมศึกษาปีที่ 7 มีค่าอำนาจจำแนกแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบสอบคณิตศาสตร์ปัญหาชั้นประถมศึกษาปีที่ 7 มีค่าความเที่ยงของแบบสอบทั้ง 3 แบบ ไม่แตกต่างกัน

นวลน้อย แต่บรรพกุล (2520: 62-65) ศึกษาแบบสอบเลือกตอบที่ใช้รูปแบบตัวเลือกต่าง ๆ ในวิชาภาษาไทยในระดับชั้นประถมศึกษาปีที่ 7 โดยแบ่งแบบสอบออกเป็น 6 ชนิดคือ แบบสอบที่ใช้ตัวเลือกแบบบรรทัด แบบปลายเปิด แบบปลายปิด และแบบผสม ผลการวิจัยพบว่า รูปแบบตัวเลือกไม่มีผลต่อค่าอำนาจจำแนก และค่าความเที่ยงของแบบสอบ แต่มีผลต่อความตรงของแบบสอบทุกฉบับ หากพิจารณาเฉพาะตัวเลือกแบบบรรทัดกับแบบปลายเปิด จะพบว่าไม่แตกต่างกันทั้งในด้านความยาก อำนาจจำแนก ความเที่ยง ความตรงและคะแนนผลการสอบ

นิรมล บุญตระกูล (2525: 59) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 โดยใช้รูปแบบตัวเลือกแตกต่างกัน 3 ฉบับคือ แบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดาทั้งฉบับ แบบสอบที่มีรูปแบบตัวเลือกแบบผสมระหว่างแบบธรรมดา กับแบบปลายเปิด "ไม่มีข้อใดถูก" จำนวนฉบับละ 50 ข้อ ในด้านความยาก อำนาจจำแนก ค่าความเที่ยง ความตรง ตลอดจนคะแนนเฉลี่ยจากการสอบของนักเรียนทั้ง 3 กลุ่ม กลุ่มตัวอย่างเป็นนักเรียนมัธยมศึกษาปีที่ 1 จังหวัดสงขลา จำนวน 200 คน โดยแต่ละคนทำแบบสอบเพียง 1 ฉบับ ผลการวิจัยพบว่า แบบสอบที่มีตัวเลือกแตกต่างกันทั้ง 3 ฉบับ ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้งในด้านความยาก อำนาจจำแนก ค่าความเที่ยง และความตรง ตลอดจนคะแนนเฉลี่ยจากการสอบของนักเรียน

จินดา ไทอนันต์ (2527: 61-62) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบทุกชนิดทุกตัวเลือก และแบบสอบเลือกตอบตัวเลือกเดียว ในด้านค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยงของแบบสอบวิชาวิทยาศาสตร์ เรื่อง สิ่งมีชีวิตและสิ่งแวดล้อม โดยการใช้แบบสอบ 2 ฉบับคือ แบบสอบแบบเลือกตอบทุกชนิดทุกตัวเลือก และแบบสอบเลือกตอบตัวเลือกเดียว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา กระทรวงศึกษาธิการ จำนวน 432 คน ผลการวิจัยปรากฏว่า ในด้านความยากแบบสอบเลือกตอบตัวเลือกเดียว มีค่าอำนาจจำแนกสูงกว่าแบบสอบเลือกตอบทุกชนิดทุกตัวเลือกอย่างมีนัยสำคัญทางสถิติ และค่าความเที่ยงของแบบสอบทั้ง 2 ฉบับไม่แตกต่างกัน

วรัญญา ปราบปัญจะ (2529: 44-46) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันในวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 โดยมีแบบสอบ 3 ฉบับ คือ แบบสอบเลือกตอบชนิดคำตอบเดียว แบบสอบเลือกตอบชนิดคำตอบซ้อนที่เงื่อนไขมีโอกาสได้รับเลือกเป็นตัวเลือกเท่ากัน และแบบสอบเลือกตอบชนิดคำตอบซ้อนที่เงื่อนไขมีโอกาสได้รับเลือกเป็นตัวเลือกไม่เท่ากันชุดละ 30 ข้อ โดยนำไปทดสอบกับกลุ่มตัวอย่างจำนวน 636 คน ผลการวิจัยปรากฏว่า แบบสอบชนิดคำตอบเดียวมีค่าความยากแตกต่างจากแบบสอบชนิดคำตอบซ้อนทั้ง 2 แบบ แต่ค่าความยากของแบบสอบชนิดคำตอบซ้อนทั้ง 2 แบบไม่แตกต่างกัน ส่วนค่าความเที่ยงของแบบสอบชนิดคำตอบซ้อนที่เงื่อนไขมีโอกาสรับเลือกเท่ากัน แต่แบบสอบชนิดคำตอบเดียว และแบบสอบชนิดคำตอบเดียวกับแบบสอบชนิดคำตอบซ้อนที่เงื่อนไขมีโอกาสรับเลือกเป็นตัวเลือกเท่ากัน มีค่าความเที่ยงไม่แตกต่างกัน

อำนาจ ทองด่าง (2531: 61-63) ได้ศึกษาผลของรูปแบบตัวเลือกของแบบสอบเลือกตอบที่มีต่อคุณภาพของแบบสอบเลือกตอบ โดยใช้แบบสอบเลือกตอบวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีรูปแบบตัวเลือกแตกต่างกัน 4 แบบคือ แบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแบบธรรมดา แบบสอบที่มีรูปแบบตัวเลือกแบบปลายเปิด แบบสอบที่มีรูปแบบตัวเลือกเป็นช่วง และแบบสอบที่มีรูปแบบตัวเลือกเป็นแบบใกล้เคียง โดยในแต่ละฉบับมีข้อสอบฉบับละ 30 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดบุรีรัมย์ จำนวน 824 คน ผลการวิจัยปรากฏว่า แบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันทั้ง 4 รูปแบบ มีค่าความยากแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งแบบสอบที่มีรูปแบบตัวเลือกเป็นแบบใกล้เคียงยากกว่าแบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา ในด้านอำนาจจำแนกของแบบสอบพบว่าอำนาจจำแนกของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกัน 4 รูปแบบ ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ในด้านค่าความเที่ยงของแบบสอบพบว่า ค่าความเที่ยงของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกัน 4 รูปแบบแตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยที่แบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดามีค่าความเที่ยงสูงสุด

พรทิพย์ แดงน้อย (2531: 77-85) ศึกษาเปรียบเทียบคุณภาพของแบบสอบเลือกตอบวิชาคณิตศาสตร์ที่มีรูปแบบตัวเลือกแตกต่างกัน 4 แบบคือ แบบสอบที่มีรูปแบบตัวเลือกแบบธรรมดา แบบสอบที่มีรูปแบบธรรมดาที่ระบุว่า "ไม่มีข้อใดถูก" แบบสอบที่มีรูปแบบตัวเลือกแบบกระบวนการคิด และแบบสอบที่มีรูปแบบตัวเลือกแบบกระบวนการคิดที่ระบุว่า "ไม่มีข้อใดถูก" กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคต้น จำนวน 492 คน แบ่งออกเป็น 4 กลุ่ม โดยแต่ละคนทำแบบสอบเพียงคนละ 1 ฉบับ ผลการวิจัยพบว่า แบบสอบที่มีรูปแบบตัวเลือกแตกต่างกันทั้ง 3 ฉบับ ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้งในด้านความยาก อำนาจจำแนก ความเที่ยงและความตรง

3.2 ตามกรอบทฤษฎีการตอบสนองข้อสอบ

การนำเอาแนวคิดเกี่ยวกับประสิทธิภาพสัมพัทธ์ไปใช้ในการเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบที่มีรูปแบบตัวเลือกต่างกันตามกรอบทฤษฎีการตอบสนองข้อสอบ ยังไม่เป็นที่แพร่หลายมากนัก จึงไม่ค่อยมีรายงานผลการวิจัยเกี่ยวกับเรื่องนี้มากนัก นอกจากผลงานของ Lord (1980: 96-100) ได้ใช้แนวคิดนี้ไปเปรียบเทียบคุณภาพของแบบสอบการอ่านเมโทรโพลิแทน (Metropolitan Reading Tests) ระดับกลาง (intermediate level) ฟอรัมเอฟ เฉพาะแบบสอบย่อยการวิเคราะห์คำ (Word Analysis subtest) [MAT] กับแบบสอบการ

อ่านอื่น ๆ อีก 6 ฉบับคือ

1. Sequential Test of Education Progress
Series II, level 4, Form A, Reading subtest. [STEP]
2. California Achievement Tests, level 4, Form A,
Reading Vocabulary. [CAT]
3. Iowa Test of Basic Skills, level 12, Form 5
Vocabulary. [ITBS]
4. Standard Reading Tests, Intermediate II, Form W
Word Meaning.[SRT]
5. Comprehensive Tests of Basic Skills, level 3, Form Q
Reading Vocabulary. [CTBS]
6. SRA Achievement Series (1971), Green Edition, Form E,
Vocabulary (SRA)

ผลการวิจัยพบว่า

STEP มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 20) และมีคุณภาพต่ำกว่าในระดับความสามารถอื่น ๆ

CAT มีคุณภาพต่ำกว่า MAT เกือบทุกระดับความสามารถ ยกเว้นที่ระดับความสามารถสูงมาก (สูงกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 95) CAT มีคุณภาพสูงกว่า MAT

ITBS มีคุณภาพต่ำกว่า MAT เกือบทุกระดับความสามารถ ยกเว้นที่ระดับความสามารถสูงกว่า (สูงกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 90) IBTS มีคุณภาพสูงกว่า MAT

SRT มีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 50) และมีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถสูง (สูงกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 50)

CTBS มีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 50) มีคุณภาพใกล้เคียงกันในกลุ่มนักเรียนที่มีความสามารถปานกลางและค่อนข้างสูง และมีคุณภาพสูงกว่าในกลุ่มที่มีความสามารถสูง (สูงกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 80)

SRA มีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 30) และมีคุณภาพสูงกว่าในระดับความสามารถอื่น ๆ

นอกจากเรื่องนี้แล้ว Lord ยังได้ใช้แนวคิดเกี่ยวกับการเปรียบเทียบคุณภาพของแบบสอบด้วยค่าสารสนเทศจากแบบสอบนี้ ไปศึกษาเปรียบเทียบคุณภาพของแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบที่มีจำนวนตัวเลือกต่างกัน พบว่า การลดจำนวนตัวเลือกต่อข้อลงในขณะที่เพิ่มจำนวนข้อสอบให้มากขึ้นมีผลทำให้คุณภาพของแบบสอบเพิ่มขึ้นในกลุ่มผู้ตอบที่มีความสามารถในระดับสูง และทำให้คุณภาพของแบบสอบลดลงในกลุ่มผู้ตอบที่มีความสามารถในระดับต่ำ และแบบสอบที่ประกอบด้วยข้อสอบที่มีจำนวนตัวเลือก 3 ตัว มีคุณภาพสูงที่สุด (Lord 1976: 36-39)

จากผลงานวิจัยที่เกี่ยวข้องกับคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันตามกรอบทฤษฎีคลาสสิกอลนั้น ทั้งในและนอกประเทศดังกล่าวข้างต้นจะเห็นได้ว่า ส่วนใหญ่แล้วเป็นการศึกษาคุณภาพของแบบสอบวิชาคณิตศาสตร์ ส่วนในวิชาอื่น ๆ นั้นมีน้อยมากเท่าที่พบก็มีวิชาวิทยาศาสตร์ และจากการค้นพบเหล่านี้พบว่า คุณภาพของแบบสอบในบ้านต่าง ๆ ยังไม่สอดคล้องกัน เช่น จากการศึกษาของ นิรมล บุญตระกูล พบว่า แบบสอบที่มีรูปแบบตัวเลือกแตกต่างกันไม่มีความแตกต่างกันในด้านความยาก อำนาจจำแนก ความเที่ยง ความตรง และคะแนนเฉลี่ย แต่ผลการศึกษาของ อำนาจ ทองด้วง พบว่า แบบสอบที่มีรูปแบบตัวเลือกต่างกัน ความยาก และความเที่ยงแตกต่างกัน เป็นต้น และจากผลงานวิจัยที่ขัดแย้งกัน ดังนั้นจึงยังไม่อาจสรุปได้อย่างชัดเจนว่าแบบสอบแบบเลือกตอบฉบับใดมีคุณภาพสูงกว่าเมื่อนำมาเปรียบเทียบกัน ทั้งนี้เป็นเพราะว่าการเปรียบเทียบคุณลักษณะที่สำคัญของการตรวจสอบคุณภาพของแบบสอบที่ละประเด็น

ส่วนการวิจัยที่เกี่ยวข้องกับคุณภาพของแบบสอบแบบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันตามกรอบทฤษฎีการตอบสนองข้อสอบในเรื่องนี้นั้นทั้งต่างประเทศและในประเทศยังไม่มีผู้ที่ทำการวิจัยมาก่อน ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับคุณภาพของแบบสอบเลือกตอบที่มีรูปแบบตัวเลือกแตกต่างกันโดยใช้ทฤษฎีการตอบสนองข้อสอบนั้น โดยการเปรียบเทียบคุณภาพของแบบสอบเลือกตอบตามกรอบทฤษฎีการตอบสนองข้อสอบน่าจะให้สารสนเทศเกี่ยวกับคุณภาพของแบบสอบได้ชัดเจนมากกว่าตามกรอบทฤษฎีคลาสสิกอล