

บทที่ 2



เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสังเคราะห์งานวิจัยเกี่ยวกับคุณภาพของแบบสอบหลายตัวเลือก อันประกอบไปด้วยค่าความเที่ยง ค่าความตรง ค่าความยากง่าย และค่าอำนาจจำแนก ที่มีรูปแบบของแบบสอบ บริบทของแบบสอบ วิธีการตรวจให้คะแนนแบบสอบ และประเภทของแบบสอบต่างกัน ด้วยวิธีการวิเคราะห์ห่อภิมาณ

การศึกษาเอกสารที่เกี่ยวข้องกับการวิจัยครั้งนี้แบ่งออกเป็น 5 ตอน ตอนแรกเป็นการเสนอสาระเกี่ยวกับสังกับเบื้องต้นของแบบสอบ ตอนที่สองเป็นเรื่องเกี่ยวกับคุณภาพของแบบสอบ ตอนที่สามเป็นเรื่องเกี่ยวกับปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบ ตอนที่สี่เป็นเรื่องเกี่ยวกับการสังเคราะห์งานวิจัยด้วยการวิเคราะห์ห่อภิมาณ และตอนสุดท้ายเป็นงานวิจัยที่เกี่ยวข้องกับการสังเคราะห์งานวิจัยด้วยการวิเคราะห์ห่อภิมาณ โดยมีรายละเอียดของรายงานเอกสารในแต่ละตอน ดังต่อไปนี้

ตอนที่ 1 สังกับเบื้องต้นเกี่ยวกับแบบสอบ

ศิริชัย กาญจนวาสิ (2542) กล่าวว่า แบบสอบ (test) เป็นเครื่องมือที่นิยมใช้ในการทดสอบ ซึ่งเป็นวิธีการที่สำคัญอย่างหนึ่งสำหรับการวัดผลการเรียนรู้จากผู้เรียน โดยสามารถแบ่งประเภทของแบบสอบตามรูปแบบของการตอบแบบสอบได้เป็น 2 ประเภท คือ แบบสอบประเภทเสนอคำตอบ (supply type) และแบบสอบประเภทเลือกตอบ (selection type) ซึ่งแบบสอบแต่ละประเภทมีรายละเอียดดังนี้

แบบสอบประเภทเสนอคำตอบ (Supply Type)

แบบสอบประเภทเสนอคำตอบ เป็นแบบสอบที่ผู้ตอบจะต้องกำหนดหรือเขียนคำตอบด้วยตนเอง โดยสามารถแบ่งออกได้เป็น 2 รูปแบบ คือ (1) ข้อสอบแบบความเรียง (essay tests) และ (2) ข้อสอบแบบตอบสั้น (short answer) และข้อสอบแบบเติมคำ (completion)

1. ข้อสอบแบบความเรียง (Essay Tests)

Kubiszyn และ Borich (1984) กล่าวถึงข้อสอบแบบความเรียงว่าเป็นข้อสอบที่ให้นักเรียนเขียนคำตอบที่ถูกต้อง โดยต้องการที่จะให้นักเรียนเรียบเรียงหรืออธิบายคำตอบ โดยที่คำถามที่ถาม

จะไม่ได้มีคำตอบเพียงคำตอบเดียวหรือมีรูปแบบการตอบเพียงรูปแบบเดียว แต่ผู้ตอบสามารถขยายคำตอบได้ ความถูกต้องแม่นยำและคุณภาพของแต่ละคำตอบสามารถตัดสินได้โดยผู้ที่มีทักษะและมีความรู้ในวิชานั้นๆ

เขาวดี วิบูลย์ศรี (2539) กล่าวถึงข้อสอบแบบความเรียงว่าเป็นแบบสอบประเภทเขียนคำตอบ คือ เป็นแบบสอบที่ผู้สอบจะต้องเรียบเรียงแนวความคิดและความรู้ที่ได้เรียนมา ตลอดจนเรียบเรียงภาษาและผูกเป็นรูปประโยคให้เป็นข้อความที่ชัดเจน แล้วเขียนเป็นคำตอบให้เหมาะสมกับความต้องการของคำถาม

ศิริชัย กาญจนวาสี (2542) กล่าวถึงข้อสอบแบบความเรียงว่าเป็นข้อสอบที่ให้เสรีภาพแก่ผู้ตอบในการประมวล คัดเลือกความรู้ ความสามารถที่ตนมีอยู่ นำมาจัดระบบ เรียบเรียงและเขียนเป็นคำตอบ คำตอบที่ได้จากผู้สอบจึงมีความหลากหลายในระดับของคุณภาพและความถูกต้อง

จากคำกล่าวข้างต้นสรุปได้ว่า ข้อสอบแบบความเรียง เป็นแบบสอบที่ผู้ตอบมีอิสระในการตอบ ซึ่งจะต้องทำการกำหนด เรียบเรียงและเขียนคำตอบด้วยตนเอง

ข้อสอบแบบความเรียงสามารถแบ่งออกได้เป็น 2 ประเภทย่อยๆ คือ ข้อสอบความเรียงแบบจำกัดความยาวในการตอบ (restricted response) และข้อสอบความเรียงแบบไม่จำกัดความยาวในการตอบ (extended response) (Mehrens และ Lehmann, 1984)

(1) ข้อสอบความเรียงแบบจำกัดความยาวในการตอบ (Restricted Response)

ข้อสอบความเรียงแบบจำกัดความยาวในการตอบเป็นข้อสอบที่ผู้ตอบจะถูกจำกัดรูปแบบและขอบเขตของคำตอบ โดยที่คำถามจะเป็นลักษณะที่ให้ผู้ตอบตอบในประเด็นและขอบเขตที่กำหนดไว้

(2) ข้อสอบความเรียงแบบไม่จำกัดความยาวในการตอบ (Extended Response)

ข้อสอบความเรียงแบบไม่จำกัดความยาวในการตอบ เป็นข้อสอบที่ให้ผู้ตอบตัดสินใจเกี่ยวกับความยาวและความซับซ้อนในการตอบข้อความเอง โดยที่ข้อสอบความเรียงประเภทนี้จะใช้วัดระดับความรู้ขั้นการสังเคราะห์หรือขั้นการประเมินผลได้

ข้อดีและข้อจำกัดของข้อสอบความเรียง

เมื่อพิจารณาถึงข้อดีและข้อจำกัดของข้อสอบความเรียงแล้ว พบว่าข้อสอบความเรียงมีข้อดี 3 ประการคือ ประการแรก ข้อสอบความเรียงสร้างง่าย ประหยัดเวลาและค่าใช้จ่ายในการสร้าง ประการที่สอง ข้อสอบความเรียงสามารถวัดพฤติกรรมการเรียนรู้ได้ในระดับสูง ตั้งแต่ระดับความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์และการประเมินผล และประการที่สาม ผู้ตอบข้อสอบความเรียงไม่สามารถเดาคำตอบได้ โดยที่ผู้ตอบจะต้องมีความรู้ที่แท้จริง

ส่วนข้อจำกัดของข้อสอบความเรียงพบว่ามี 2 ประการคือ ประการแรก ข้อสอบความเรียงตรวจให้คะแนนได้ยาก เสียเวลาในการตรวจให้คะแนน และขาดความคงที่ในการให้คะแนน

ประการที่สอง ข้อสอบความเร็วไม่สามารถออกข้อสอบได้ครอบคลุมเนื้อหาวิชา เพราะข้อสอบในแต่ละข้อจะต้องใช้เวลาในการตอบมาก

2. ข้อสอบแบบตอบสั้น (Short Answer) และข้อสอบแบบเติมคำ (Completion)

Ebel (1965) กล่าวถึงข้อสอบแบบตอบสั้นว่าเป็นข้อสอบที่ให้นักเรียนเติมคำตอบ ซึ่งอาจจะเป็นคำ วลี ตัวเลขหรือสัญลักษณ์ต่างๆ ลงในช่องว่าง

Mehrens และ Lehmann (1984) กล่าวถึงข้อสอบแบบตอบสั้นว่าเป็นข้อสอบที่เกี่ยวกับความจำง่าย ๆ โดยให้ผู้สอบเป็นผู้เขียนคำตอบซึ่งอาจจะเป็นคำ วลี สัญลักษณ์ทางคณิตศาสตร์หรือสูตรลงในช่องว่างที่ปรากฏในข้อสอบ ซึ่งอาจจะมีมากกว่า 1 ช่องว่าง ซึ่งข้อสอบแบบตอบสั้นโดยทั่วไปมีความหลากหลายใน 3 ลักษณะ คือ ความหลากหลายของคำถาม ความหลากหลายของสิ่งที่เติม และความหลากหลายของความสัมพันธ์ ดังนี้

(1) ความหลากหลายของคำถาม (question variety) โดยที่ข้อสอบจะถามคำถามที่ตรงประเด็น

ตัวอย่างเช่น เมืองใดที่เป็นที่ตั้งของห้องทดลองทางจิตวิทยาแห่งแรก? (ไลป์ซิก)

(2) ความหลากหลายของสิ่งที่เติม (completion variety) โดยที่ลักษณะของข้อสอบจะเป็นข้อความที่ไม่สมบูรณ์

ตัวอย่างเช่น ห้องทดลองทางจิตวิทยาแห่งแรกตั้งอยู่ที่เมือง.....(ไลป์ซิก)

(3) ความหลากหลายของความสัมพันธ์ (association variety) โดยจะให้เติมคำตอบที่สัมพันธ์กับสิ่งที่กำหนดไว้ในข้อคำถาม

ตัวอย่างเช่น จากชื่อเมืองที่กำหนดมา ให้เติมชื่อรัฐที่เมืองเหล่านี้ตั้งอยู่

คีทรอย(มิชิแกน)

ชิคาโก(อิลลินอยส์)

บอสตัน(แมสซาชูเซต)

เยวคิ วินูลย์ศรี (2539) กล่าวถึงข้อสอบแบบตอบสั้นว่าเป็นข้อสอบที่มุ่งให้ผู้ตอบเขียนคำตอบขึ้นเองทั้งหมด แต่ให้ใช้ถ้อยคำหรือข้อความที่สั้นและกระชับ ซึ่งตามปกติจะไม่ให้ตอบเกินเนื้อที่หรือช่องว่างที่เว้นไว้ให้ และในทางปฏิบัติแล้วจะถือว่าข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำ (completion) เป็นข้อสอบประเภทเดียวกัน ซึ่งข้อสอบแบบเติมคำนั้นเป็นข้อสอบที่มุ่งให้ผู้ตอบเติมคำ วลี หรือประโยคที่ถูกต้องต่อจากข้อความที่ได้เขียนค้างไว้ เพื่อให้เป็นข้อความที่ครบถ้วนและตรงตามข้อเท็จจริง

ศิริชัย กาญจนวาสี (2542) กล่าวถึงข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำว่ามีลักษณะที่คล้ายคลึงกัน กล่าวคือ เป็นข้อสอบที่ผู้สอบต้องคิดคำตอบขึ้นมาเอง แต่เป็นคำตอบสั้นๆ

หรือการเติมคำตอบ จึงเหมาะสำหรับวัดความรู้ ความจำเกี่ยวกับคำศัพท์ ข้อเท็จจริง หลักการและกฎเกณฑ์ต่างๆ

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำมีความคล้ายคลึงกัน โดยเป็นข้อสอบที่ให้ผู้ตอบเติมคำหรือวลีหรือประโยคในช่องว่างหรือขอบเขตที่กำหนด เพื่อให้ข้อความนั้นๆ ถูกต้องและครบถ้วนสมบูรณ์

ข้อดีและข้อจำกัดของข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำ

เมื่อพิจารณาถึงข้อดีและข้อจำกัดของข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำแล้ว พบว่าข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำมีข้อดี 3 ประการคือ ประการแรก ข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำสร้างง่าย และประหยัดเวลาในการสร้าง ประการที่สอง ข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำสามารถสร้างได้หลายข้อ จึงทำให้ครอบคลุมเนื้อหาวิชา และประการที่สามผู้ตอบข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำไม่สามารถเดาคำตอบได้ โดยจะต้องมีความรู้ในเนื้อหาวิชานั้นๆ

ส่วนข้อจำกัดของข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำพบว่ามี 3 ประการคือ ประการแรก หากทำการสร้างข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำได้ไม่ดี จะเป็นการวัดพฤติกรรมการเรียนรู้ระดับความรู้ความจำ ไม่สามารถวัดพฤติกรรมการเรียนรู้ขั้นสูงได้ ประการที่สอง หากทำการสร้างข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำได้ไม่ดีหรือไม่มีการกำหนดขอบเขตของคำตอบ จะทำให้มีคำตอบที่ถูกต้องมากกว่าหนึ่งคำตอบ ประการที่สาม ข้อสอบแบบตอบสั้นและข้อสอบแบบเติมคำตรวจให้คะแนนยาก และเสียเวลาในการตรวจให้คะแนน

แบบสอบประเภทเลือกคำตอบ (Selection Type)

แบบสอบประเภทเลือกคำตอบ เป็นแบบสอบที่มีการกำหนดคำตอบไว้ในข้อสอบ แล้วให้ผู้ตอบเลือกคำตอบที่ถูกต้อง แบบสอบประเภทนี้สามารถแบ่งออกเป็น 3 รูปแบบ คือ (1) ข้อสอบแบบถูก-ผิด (true-false) (2) ข้อสอบแบบจับคู่ (matching) และ (3) ข้อสอบแบบหลายตัวเลือก (multiple-choice)

1. ข้อสอบแบบถูก-ผิด (true-false)

Thorndike และ Hagen (1969) กล่าวถึงข้อสอบแบบถูก-ผิดว่าเป็นข้อสอบหรือข้อความที่มี 2 ตัวเลือกให้นักเรียนเลือกคำตอบที่ถูกต้อง

Mchrens และ Lehmann (1984) กล่าวถึงข้อสอบแบบถูก-ผิดว่าเป็นข้อสอบที่มี 2 ตัวเลือกซึ่งจะมีเพียงหนึ่งตัวเลือกเท่านั้นที่เป็นคำตอบ โดยที่ข้อสอบประเภทนี้จะให้นักเรียนเป็นผู้ตัดสินใจ

ประโยคที่กำหนดให้เป็นจริงหรือเท็จ ความหลากหลายของข้อสอบแบบถูก-ผิดแบ่งออกได้เป็น 4 ลักษณะ คือ

- (1) แบบที่ตอบว่าใช่หรือไม่ใช่ (yes-no)
- (2) แบบที่ตอบว่าถูกหรือผิด (right-wrong)
- (3) แบบความหลากหลายในกลุ่มคำถาม (cluster variety)

ตัวอย่างเช่น คำเฉลยเลขคณิตคือ

ถูก ผิด (ก) การวัดแนวโน้มเข้าสู่ส่วนกลาง

ถูก ผิด (ข) ได้รับผลกระทบจากคะแนนสูงสุดน้อยกว่าค่ามัธยฐาน

ถูก ผิด (ค) ใช้แสดงความหมายของสหสัมพันธ์ระหว่าง 2 ตัวแปร

- (4) แบบความหลากหลายในการแก้ไขคำตอบให้ถูกต้อง (correction variety)

ตัวอย่างเช่น ถูก ผิด รากที่ 2 ของ 64 คือ 9 คำตอบที่ถูกต้องคือ 8

อุทุมพร (ทองอุไทย) จามรมาน (2535) กล่าวถึงข้อสอบแบบถูก-ผิดว่าประกอบด้วยข้อสอบหรือข้อความหลายๆ ข้อ บางข้อเป็นข้อที่ถูกและบางข้อเป็นข้อที่ผิด ผู้ตอบทำเครื่องหมาย ✓ หน้าข้อที่คิดว่าถูก และ X หน้าข้อที่คิดว่าผิด ข้อความแต่ละข้อเกี่ยวข้องกับเนื้อหาแต่ละประเด็น ข้อสอบแบบถูก-ผิดจึงทำการวัดเนื้อหาได้มาก

เยาวดี วิบูลย์ศรี (2539) กล่าวถึงข้อสอบแบบถูก-ผิดว่าเหมาะที่จะใช้วัดความจำที่เกี่ยวข้องกับข้อเท็จจริง เช่น ชื่อบุคคล สถานที่ ปีที่มีความสำคัญสำหรับเนื้อหานั้นๆ ผู้ตอบจะต้องมีความสามารถในการจำแนกข้อความว่า “ถูก” หรือ “ผิด” ออกมาให้ได้

ศิริชัย กาญจนวาสิ (2542) กล่าวถึงข้อสอบแบบถูก-ผิดว่าเป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบที่เป็นไปได้ 2 อย่าง เช่น ข้อความที่กำหนดให้ นั้น ถูกหรือผิด ใช่หรือไม่ใช่ จริงหรือเท็จ เป็นต้น ข้อสอบแบบถูก-ผิดนี้สามารถใช้วัดความรู้ความจำ ความเข้าใจในหลักการและการนำไปใช้

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าข้อสอบแบบถูก-ผิดเป็นข้อสอบที่ให้ผู้สอบตัดสินใจเลือกตอบว่าถูกหรือผิด จริงหรือเท็จ ใช่หรือไม่ใช่ เปรียบเสมือนมีตัวเลือกให้ตอบ 2 ตัวเลือกที่มีความหมายตรงข้ามกัน

ข้อดีและข้อจำกัดของข้อสอบแบบถูก-ผิด

เมื่อพิจารณาถึงข้อดีและข้อจำกัดของข้อสอบแบบถูก-ผิดแล้ว พบว่าข้อสอบแบบถูก-ผิดมีข้อดี 3 ประการคือ ประการแรก ข้อสอบแบบถูก-ผิดสร้างง่ายและประหยัดเวลาในการสร้าง ประการที่สอง ข้อสอบแบบถูก-ผิดสามารถสร้างได้ครอบคลุมเนื้อหาวิชา ประการที่สาม ข้อสอบแบบถูก-ผิดตรวจให้คะแนนได้ง่ายและมีความรวดเร็ว

ส่วนข้อจำกัดของข้อสอบแบบถูก-ผิดพบว่ามี 2 ประการคือ ประการแรก ข้อสอบแบบถูก-ผิดเปิดโอกาสให้ผู้ตอบเดาถูกสูงถึงร้อยละ 50 ประการที่สอง ข้อสอบแบบถูก-ผิดวัดพฤติกรรมการเรียนรู้ได้ในระดับต่ำ คือ ระดับความรู้ความจำ ความเข้าใจ

2 ข้อสอบแบบจับคู่ (Matching)

Mehrens และ Lehmann (1984) กล่าวถึงข้อสอบแบบจับคู่ว่าเป็นข้อสอบที่ประกอบด้วยข้อความ 2 แถว แถวแรกจะประกอบด้วยคำถามหรือปัญหาที่ต้องการคำตอบ และแถวที่ 2 จะเป็นคำตอบ ซึ่งผู้ตอบจะต้องเลือกคำตอบที่มีความเกี่ยวข้องสัมพันธ์กับคำถามหรือปัญหา

อุทุมพร (ทองอุไทย) จามรรมาน (2535) กล่าวถึงข้อสอบแบบจับคู่ว่าประกอบด้วยข้อความ 2 ชุด ที่มีความสัมพันธ์กัน คำสั่งให้ผู้สอบทำก็คือให้จับคู่ข้อความทางซ้ายกับทางขวาที่ตนคิดว่ามีความสัมพันธ์กัน

บุญชม ศรีสะอาด (2535) กล่าวถึงข้อสอบแบบจับคู่ว่าเป็นข้อสอบที่มีข้อความหรือคำ 2 ชุด แยกไว้เป็น 2 คอลัมน์ทางซ้ายและขวา ให้ผู้สอบเลือกจับคู่ว่าแต่ละข้อความหรือคำในคอลัมน์หนึ่ง ควรจะเข้ากับข้อความใดหรือคำใดในอีกคอลัมน์หนึ่ง

ศิริชัย กาญจนวาสี (2542) กล่าวถึงข้อสอบแบบจับคู่ว่าเป็นข้อสอบที่ให้ผู้ตอบจับคู่ระหว่างคำหรือข้อความสองแถวที่มีความสอดคล้องหรือสัมพันธ์กัน โดยทั่วไปแถวทางซ้ายมือจะเป็นข้อความ ส่วนแถวทางขวามือจะเป็นคำตอบ ข้อสอบแบบจับคู่เหมาะสำหรับวัดความรู้ความจำเกี่ยวกับข้อเท็จจริง คำศัพท์ หลักการ ความสัมพันธ์และการตีความหมาย

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าข้อสอบแบบจับคู่เป็นข้อสอบที่ประกอบด้วยข้อความ 2 แถว โดยแถวแรกจะเป็นข้อความ และแถวที่สองจะเป็นคำตอบ ซึ่งทั้งสองแถวจะมีความสอดคล้องสัมพันธ์กัน ซึ่งผู้ตอบจะต้องเป็นผู้เลือกคำตอบที่มีความสอดคล้องกับข้อความ

ข้อดีและข้อจำกัดของข้อสอบแบบจับคู่

เมื่อพิจารณาถึงข้อดีและข้อจำกัดของข้อสอบแบบจับคู่แล้ว พบว่าข้อสอบแบบจับคู่มีข้อดี 3 ประการคือ ประการแรก ข้อสอบแบบจับคู่เหมาะสำหรับการใช้วัดการแสดงความสัมพันธ์หรือความสอดคล้องระหว่างสิ่งต่างๆ ประการที่สอง หากสร้างข้อสอบแบบจับคู่ให้มีจำนวนตัวเลือกที่มากจะสามารถลดการเดาได้ ประการที่สามข้อสอบแบบจับคู่ตรวจให้คะแนนได้ง่าย

ส่วนข้อจำกัดของข้อสอบแบบจับคู่พบว่ามี 2 ประการคือ ประการแรก ข้อสอบแบบจับคู่วัดพฤติกรรมการเรียนรู้ในระดับต่ำ คือ ระดับความรู้ความจำ ความเข้าใจ ประการที่สองข้อสอบแบบจับคู่มีระดับความยากของข้อสอบไม่เท่ากัน โดยข้อสอบข้อแรกๆ จะยากกว่าข้อสอบข้อหลังๆ เพราะจำนวนตัวเลือกที่ใช้จะลดลงเรื่อยๆ

3. ข้อสอบแบบหลายตัวเลือก (Multiple-Choice)

Thorndike และ Hagen (1969) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน คือ ส่วนที่เป็นคำถาม (stem) ที่เป็นปัญหาและส่วนที่เป็นคำตอบ ซึ่งรูปแบบมาตรฐานของข้อสอบแบบหลายตัวเลือกจะมีตัวเลือกหนึ่งเป็นคำตอบที่ถูกต้องหรือคำตอบที่ดีที่สุด และตัวเลือกอื่นๆ จะเป็นตัวลวง โดยที่ตัวคำถามอาจจะเป็นปัญหาอย่างใดอย่างหนึ่งหรือเป็นข้อความที่ไม่สมบูรณ์

Roid และ Haladyna (1982) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกทั้งหมดมีส่วนประกอบที่เหมือนกัน คือ (1) ตัวคำถามของข้อสอบข้อนั้นๆ จะเป็นข้อความเริ่มต้น เป็นตัวกระตุ้นผู้สอบ (2) ตัวเลือก (choices) ตัวเลือกตัวหนึ่งจะเป็นคำตอบที่ถูกต้อง และ (3) คำตอบผิด (foils) หรือตัวลวง (distractors) เป็นตัวเลือกที่ผิด ซึ่งถูกคาดว่าจะมีเหตุที่น่าดึงดูดให้นักเรียนที่ไม่รอบรู้ในข้อนั้นๆ ตอบ

Mehrens และ Lehmann (1984) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน คือ ส่วนแรกเป็นตัวคำถาม (stem) หรือปัญหา และส่วนที่สองเป็นคำตอบที่เป็นตัวถูก (key) และตัวลวง (distractors) โดยส่วนของคำถามอาจจะเป็นคำถามที่ถามโดยตรงหรืออาจจะเป็นข้อความที่ไม่สมบูรณ์

อุทุมพร (ทองอุไทย) จามรมาน (2535) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าประกอบด้วยข้อความที่เป็นคำถามและคำตอบให้เลือก ซึ่งอาจมีตั้งแต่ 3 ตัวเลือกขึ้นไป

บุญชม ศรีสะอาด (2535) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกประกอบด้วยส่วนสำคัญ 2 ส่วน คือ ส่วนที่เป็นตัวคำถามหรือตอนนำ (stem or problem) กับส่วนที่เป็นตัวเลือกหรือตอนเลือก (choices or alternatives) และในตอนเลือกนี้ประกอบด้วยตัวเลือกที่เป็นคำตอบที่ถูกต้อง (correct answer) หรือที่ดีที่สุด (best choice) กับตัวเลือกที่เป็นตัวลวง (distractors) ตัวลวงทุกตัวของข้อนั้นๆ อาจเป็นคำตอบผิดชนิดไม่มีส่วนถูกอยู่เลย หรือเป็นคำตอบถูกแต่มีส่วนถูกหรือมีความเหมาะสมน้อยกว่าตัวเลือกที่เป็นคำตอบที่ถูกต้อง

เยาวดี วิบูลย์ศรี (2539) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกเป็นที่นิยมใช้กันมากในปัจจุบัน เพราะเป็นข้อสอบที่สามารถจำแนกระดับความรู้ต่างๆ ได้ดีกว่าแบบสอบประเภทอื่นๆ โดยเฉพาะอย่างยิ่งถ้าสร้างให้มีมาตรฐาน โครงสร้างของข้อสอบแบบหลายตัวเลือกประกอบด้วยข้อกระทง 2 ส่วนใหญ่ๆ คือ ส่วนแรกเป็นคำถามนำ และส่วนที่สองเป็นตัวเลือก ข้อกระทงที่เป็นคำถามนำประกอบด้วยส่วนที่เป็นคำถามหลักของแต่ละข้อ ส่วนข้อกระทงที่เป็นตัวเลือกประกอบด้วยตัวคำตอบ 1 ตัว และที่เหลือเป็นตัวลวง ลักษณะของคำถามอาจจะเป็นคำถามโดยตรง หรืออาจจะเป็นข้อความที่ไม่สมบูรณ์

ศิริชัย กาญจนวาที (2542) กล่าวถึงข้อสอบแบบหลายตัวเลือกกว่าข้อสอบแบบหลายตัวเลือกเป็นที่นิยมใช้กันอย่างกว้างขวาง เพราะสามารถใช้วัดผลการเรียนรู้ ทั้งความรู้ ความเข้าใจ การนำไป

ใช้และผลการเรียนรู้ขั้นสูงได้ ข้อสอบแบบหลายตัวเลือกเป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบจากตัวเลือกที่กำหนดให้ ข้อสอบแบบหลายตัวเลือกประกอบด้วย 2 ส่วน ได้แก่ ตัวคำถาม (stem) และตัวเลือก (alternative หรือ options) ซึ่งนิยมใช้ 3-6 ตัวเลือก ในส่วนของตัวเลือกประกอบด้วยตัวเลือกที่เป็นคำตอบถูกเรียกว่าตัวคำตอบ (answer หรือ key) 1 ตัว ส่วนที่เหลือเป็นตัวเลือกที่ผิดเรียกว่าตัวลวง (distractors)

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าข้อสอบแบบหลายตัวเลือกเป็นข้อสอบที่นิยมใช้กันมากในปัจจุบัน โดยประกอบด้วย 2 ส่วนสำคัญ คือ ส่วนที่เป็นคำถามหรือปัญหา (stem หรือ problem) ซึ่งอาจจะเป็นการถามโดยตรงหรือเป็นข้อความที่ไม่สมบูรณ์ และส่วนที่เป็นคำตอบ (alternative หรือ options) จะประกอบด้วย ตัวเลือกที่เป็นคำตอบถูก 1 ตัวและตัวเลือกอื่นๆ นอกจากนั้นจะเป็นตัวลวง (distractors)

ข้อดีและข้อจำกัดของข้อสอบหลายตัวเลือก

เมื่อพิจารณาถึงข้อดีและข้อจำกัดของข้อสอบแบบหลายตัวเลือก สรุปได้ว่าข้อสอบแบบหลายตัวเลือกมีข้อดี 4 ประการ ประกอบด้วย ประการแรกข้อสอบแบบหลายตัวเลือกสามารถวัดได้ครอบคลุมพฤติกรรมการเรียนรู้ครบทุกระดับพฤติกรรม ตั้งแต่ความรู้ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินผล ประการที่สองข้อสอบแบบหลายตัวเลือกมีความเป็นปรนัยในการตรวจให้คะแนนสูง ตรวจง่ายและรวดเร็ว สามารถใช้คอมพิวเตอร์ในการตรวจให้คะแนนได้ ประการที่สาม ข้อสอบแบบหลายตัวเลือกทำให้ผลจากการเดาคำตอบลดลงกว่าข้อสอบแบบอื่นๆ เพราะมีหลายตัวเลือก ประการสุดท้ายข้อสอบแบบหลายตัวเลือกสามารถออกข้อสอบได้เป็นจำนวนมาก ทำให้สามารถวัดผลการเรียนรู้ได้ครอบคลุมเนื้อหาและวัดผลการเรียนรู้กับกลุ่มผู้สอบจำนวนมากๆ ได้

ข้อสอบแบบหลายตัวเลือกมีข้อจำกัด 3 ประการ ประกอบด้วย ประการแรกข้อสอบแบบหลายตัวเลือกสร้างได้ยาก เสียเวลาและค่าใช้จ่ายมาก เมื่อเปรียบเทียบกับการสร้างข้อสอบแบบอื่นๆ เนื่องจากต้องสร้างตัวลวงให้ใกล้เคียงกับตัวถูกให้มากที่สุด ประการที่สองข้อสอบแบบหลายตัวเลือกไม่เหมาะที่จะใช้วัดการแสดงความคิดเห็น ความสามารถในการเขียน เนื่องจากมีการกำหนดคำตอบให้ ไม่เปิดโอกาสให้ผู้เข้าสอบแสดงความคิดเห็นของคนในส่วนที่เป็นคำตอบที่หลากหลาย และประการสุดท้ายหากขาดความระมัดระวังในขั้นตอนการสร้างข้อสอบแบบหลายตัวเลือกแล้ว อาจจะมีคำตอบที่ถูกต้องมากกว่า 1 ข้อ

จากข้อดีและข้อจำกัดของข้อสอบแบบหลายตัวเลือกพบว่าข้อสอบแบบหลายตัวเลือกมีข้อดีที่มากกว่าข้อจำกัด จึงเป็นที่นิยมใช้ในการวัดผลการเรียนรู้ของผู้เรียนอย่างกว้างขวาง

ตอนที่ 2 คุณภาพของแบบสอบ

แบบสอบที่จะนำไปใช้ทดสอบวัดความรู้ผู้เรียน ควรที่จะมีลักษณะตามที่ ชวาล แพร์ตฤด (2516) กล่าวไว้เกี่ยวกับคุณลักษณะที่ดีของแบบสอบ 10 ประการ กล่าวคือ มีความตรง (validity) มีความเที่ยง (reliability) มีความเป็นปรนัย (objectivity) ต้องถามลึก (searching) มีความยุติธรรม (fairness) ต้องยั่วยุ-กระตุ้นผู้สอบ (exemplary) มีประสิทธิภาพ (efficiency) ต้องถามเฉพาะเจาะจง (definite) มีความยากง่าย (item difficulty) และมีอำนาจจำแนก (discrimination)

ซึ่งคุณลักษณะที่ดีของแบบสอบทั้ง 10 ประการข้างต้น สามารถสรุปได้ดังนี้

(1) แบบสอบที่ดีต้องมีความตรง (validity) ซึ่งถือเป็นหัวใจของแบบสอบ โดยแบบสอบที่ดีควรที่จะมีความตรงทั้ง 3 ประเภท คือ ความตรงตามเนื้อหา (content validity), ความตรงตามเกณฑ์สัมพัทธ์ (criterion-related validity) และความตรงตาม โครงสร้าง (construct validity)

(2) แบบสอบที่ดีต้องมีความเที่ยง (reliability) สูง ซึ่งควรที่จะมีค่ามากกว่า 0.80

(3) แบบสอบที่ดีต้องมีความเป็นปรนัยของแบบสอบ (objectivity) ซึ่งประกอบด้วย

- (ก) ต้องมีคำถามที่แจ่มชัด ผู้สอบทุกคนอ่านแล้วเข้าใจข้อคำถามได้ถูกต้องตรงกัน
- (ข) ต้องตรวจให้คะแนนได้ตรงกัน ไม่ว่าผู้ใดตรวจข้อสอบ ย่อมที่จะได้ผลออกมาเหมือนกัน
- (ค) ต้องแปลความหมายของคะแนนได้แจ่มชัด หรือแปลความหมายของคะแนนที่ได้ อย่างเดียวกันเพื่อที่จะสามารถนำไปใช้ในการเปรียบเทียบระหว่างผู้สอบได้

(4) แบบสอบที่ดีต้องถามลึก (searching) ซึ่งจะต้องวัดพฤติกรรมการเรียนรู้ได้ลึกซึ้งและครบทุกพฤติกรรม ตั้งแต่ความรู้ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินผล

(5) แบบสอบที่ดีต้องมีความยุติธรรม (fairness) โดยแบบสอบจะต้องมีความยุติธรรมกับผู้สอบทุกคน ไม่เปิดโอกาสให้เกิดการได้เปรียบเสียเปรียบกันในระหว่างกลุ่มผู้สอบ

(6) แบบสอบที่ดีต้องมีคำถามที่ยั่วยุหรือกระตุ้น (exemplary) โดยที่ลักษณะของคำถามจะต้องกระตุ้นให้ผู้สอบเกิดความพยายามที่จะเอาชนะใจปัญหา ซึ่งมีหลายวิธี เช่น ให้มีการเรียงลำดับข้อสอบข้อที่ง่ายไปสู่ข้อที่ยากทีละขั้น เป็นต้น

(7) แบบสอบที่ดีต้องมีประสิทธิภาพ (efficiency) จะต้องใช้สอบวัดความรู้ของผู้สอบได้อย่างถูกต้อง โดยใช้เวลาและจำนวนข้อที่น้อย แต่มีคุณภาพเหมือนกับแบบสอบที่ใช้เวลาและจำนวนข้อที่มากกว่า

(8) แบบสอบที่ดีต้องถามเฉพาะเจาะจง (definite) โดยจะต้องถามเรื่องใดเรื่องหนึ่งเฉพาะเจาะจงในหัวข้อนั้นๆ และ/หรือมีการกำหนดขอบเขตของคำตอบไว้ชัดเจน เป็นต้น

(9) แบบสอบที่ดีจะต้องมีค่าความยากง่าย (item difficulty) อยู่ในระดับที่เหมาะสม คือ อยู่ในระดับปานกลาง ระหว่าง 0.2 ถึง 0.8 ซึ่งเป็นระดับที่แสดงให้เห็นว่าแบบสอบที่ใช้ไม่ยากหรือง่ายจนเกินไป

(10) แบบสอบที่ดีต้องมีค่าอำนาจจำแนก (discrimination) ในระดับที่สูง คือ ต้องสูงกว่า 0.2 ขึ้นไป เพื่อที่จะสามารถใช้เป็นเครื่องมือในการจำแนกกลุ่มผู้สอบออกเป็นกลุ่มที่มีความสามารถสูงหรือกลุ่มเก่ง และกลุ่มที่มีความสามารถต่ำหรือกลุ่มอ่อนได้อย่างมีประสิทธิภาพ

จากคุณลักษณะที่ดีของแบบสอบทั้ง 10 ประการ พบว่ามี 4 คุณลักษณะของแบบสอบที่เป็นที่นิยมใช้ในการตรวจสอบคุณภาพของแบบสอบ คือ ความตรง ความเที่ยง ความยากง่าย และอำนาจจำแนก เนื่องจากสามารถที่จำคำนวณออกมาเป็นตัวเลขได้ ซึ่งคุณภาพของแบบสอบทั้ง 4 ประเด็นมีสาระเกี่ยวกับความหมาย ประเภทและการวัด ดังสรุปในหัวข้อต่อไปนี้

1. ความตรง (Validity)

Lindquist (1951) กล่าวถึงความหมายของความตรงว่า ความตรง หมายถึง ความสามารถในการวัดสิ่งที่ต้องการวัด ได้ถูกต้องแม่นยำ หรือหมายถึงความสามารถในการให้ความหมายในสิ่งที่วัดได้อย่างไม่ผิดพลาด

Ebel (1972) กล่าวถึงความหมายของความตรงว่า ความตรง หมายถึง ความถูกต้องแม่นยำของคะแนนที่ได้จากการวัด ไม่ว่าจะทำการวัดสิ่งใด

ศิริชัย กาญจนวาที (2542) กล่าวถึงความหมายของความตรงว่า ความตรง หมายถึง ความถูกต้องแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการวัด

จากความหมายของความตรงข้างต้น สรุปได้ว่าความตรง (validity) หมายถึง คุณสมบัติของเครื่องมือหรือแบบสอบที่วัดได้ในสิ่งที่ต้องการวัด สามารถวัดได้ตรงตามจุดประสงค์ที่จะวัด จึงมีเกณฑ์ (criterion) ที่จะใช้พิจารณาว่ามีความตรงตามจุดประสงค์หรือไม่

ประเภทของความตรง

การแบ่งประเภทของความตรงมีหลายวิธี ซึ่งจะขึ้นอยู่กับจุดมุ่งหมายของการวัด และจากคู่มือมาตรฐานสำหรับการทดสอบทางจิตวิทยาและการศึกษา (Standards for Educational and Psychological Tests and Manuals) ของสมาคมจิตวิทยาอเมริกัน (American Psychological Association) เมื่อปีคริสต์ศักราช 1974 (อ้างถึงใน Mehrens และ Lehmann, 1984) ได้แบ่งประเภทของความตรงออกเป็น 3 ประเภท คือ ความตรงตามเนื้อหา (content validity), ความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) และความตรงตามโครงสร้าง (construct validity) ซึ่งสอดคล้องกับ Thorndike และ Hagen (1969), Kubiszyn และ Borich (1984) และเขาวดี วิบูลย์ศรี (2539) ซึ่งรายละเอียดของความตรงแต่ละประเภทมีดังนี้

1.1 ความตรงตามเนื้อหา (Content Validity)

Ebel (1972) กล่าวถึงความตรงตามเนื้อหาว่า เป็นความเกี่ยวข้องกับความพอเพียงของการเป็นตัวแทนในขอบเขตของเนื้อหา

กล่าวโดยสรุปแล้ว ความตรงตามเนื้อหา หมายถึง การที่เครื่องมือหรือแบบสอบสามารถวัดเนื้อหาได้ตรงตามเรื่องที่ต้องการวัด โดยความตรงตามเนื้อหาจะดูตารางวิเคราะห์หลักสูตรเป็นเกณฑ์ว่าแบบสอบมีความตรงตามเนื้อหาหรือไม่ หรือเป็นตัวแทนของมวลเนื้อหาที่ต้องการวัดหรือไม่ โดยที่น้ำหนักที่ระบุในตารางวิเคราะห์หลักสูตรจะเป็นเกณฑ์ ดังที่ Kubiszyn และ Borich (1984) กล่าวว่าความตรงตามเนื้อหาสามารถตอบคำถามได้ว่า แบบสอบที่ใช้วัดได้ตรงตามวัตถุประสงค์การเรียนรู้หรือไม่

1.2 ความตรงตามเกณฑ์สัมพัทธ์ (Criterion-Related Validity)

ความตรงตามเกณฑ์สัมพัทธ์เป็นเรื่องเกี่ยวกับเทคนิคของการศึกษาความสัมพันธ์ระหว่างคะแนนของแบบสอบหรือการวัดอื่นๆ (ซึ่งเป็นตัวทำนาย) กับการวัดตามเกณฑ์ภายนอก (ซึ่งเป็นเกณฑ์) เช่น คะแนนจากการทดสอบความถนัดเชิงวิชาการ (SAT) กับเกรดเฉลี่ย เป็นต้น (Mehrens และ Lehmann, 1984)

ความตรงตามเกณฑ์สัมพัทธ์แบ่งออกได้เป็น 2 ประเภท คือ ความตรงเชิงทำนาย (predictive validity) และความตรงร่วมสมัย (concurrent validity) ดังนี้

1.2.1 ความตรงเชิงทำนาย (Predictive Validity)

ความตรงเชิงทำนาย เป็นตัวบ่งชี้ถึงระดับของความเหมือนกันระหว่างคะแนนสอบจากข้อคำถาม และผลลัพธ์ที่เกิดขึ้นในอนาคต ซึ่งถือว่ามีความเกี่ยวข้องกันของลักษณะที่วัดโดยแบบสอบ (Tuckman, 1975) นั่นคือ ความตรงเชิงทำนาย หมายถึง ลักษณะของข้อสอบที่สามารถวัดเพื่อทำนายอนาคตได้ถูกต้องโดยอาศัยข้อเท็จจริงในปัจจุบัน

1.2.2 ความตรงร่วมสมัย (Concurrent Validity)

ความตรงร่วมสมัยเป็นความเกี่ยวข้องกับความสัมพันธ์ของคะแนนที่จะยอมรับเกณฑ์ในปัจจุบันของการปฏิบัติบนตัวแปรที่แบบสอบมีแนวโน้มที่จะวัด (Ebel, 1972) นั่นคือ ความตรงร่วมสมัย หมายถึง ลักษณะที่แบบสอบสามารถวัดได้ตรงตามสภาพความเป็นจริงในปัจจุบันของสิ่งที่จะวัด โดยการหาความสัมพันธ์ระหว่างคะแนนของแบบสอบกับคะแนนเกณฑ์ที่กำหนดขึ้นในเวลาเดียวกันหรือใกล้เคียงกัน

Mehrens และ Lehmann (1984) กล่าวถึงความแตกต่างระหว่างความตรงเชิงทำนายและความตรงร่วมสมัยว่า ข้อแตกต่างระหว่างความตรงทั้ง 2 ประเภทนี้คือ ช่วงระยะเวลาของการเก็บข้อมูลที่เป็นเกณฑ์ ถ้าเก็บรวบรวมข้อมูลเป็นเกณฑ์กับคะแนนการสอบในเวลาเดียวกันหรือใกล้เคียงกัน

จะเป็นความจริงร่วมสมัย แต่ถ้าเก็บรวบรวมข้อมูลที่เป็นเกณฑ์ที่ล่าช้าออกไปจะเป็นความจริงทำนายน

1.3 ความตรงตามโครงสร้าง (Construct Validity)

Ebel (1972) กล่าวถึงความตรงตามโครงสร้างว่า ความตรงตามโครงสร้างเกี่ยวข้องกับคุณสมบัติทางจิตวิทยาที่แบบสอบจะวัดและประเมินคุณสมบัตินั้น โดยการแสดงให้เห็นถึงโครงสร้างที่แน่นอน

ศิริชัย กาญจนวาสี (2542) กล่าวถึงความตรงตามโครงสร้างว่า ความตรงตามโครงสร้างหมายถึงความสามารถในการวัดได้ตรงตามคุณลักษณะทางจิตวิทยาของการเรียนรู้

จากคำกล่าวข้างต้นสรุปได้ว่า ความตรงตามโครงสร้างหมายถึง การที่แบบสอบวัดได้ตรงตามพฤติกรรมต่างๆ ครอบคลุมที่ระบุในวัตถุประสงค์ของหลักสูตร

2. ความเที่ยง (Reliability)

Ebel (1972) กล่าวถึงความเที่ยงว่า ความเที่ยงหมายถึง ความคงที่ของกลุ่มคะแนนที่ได้จากการวัดของแบบสอบ ไม่ว่าจะทำการวัดเมื่อใดก็ตาม

Tuckman (1975) กล่าวถึงความเที่ยงว่า ความเที่ยงหมายถึง ระดับความคงที่ในการวัดไม่ว่าจะทำการวัดเมื่อใดก็ตาม

Mchrens และ Lehmann (1984) กล่าวถึงความเที่ยงว่า ความเที่ยงหมายถึง ระดับของความคงที่ระหว่างการวัดสิ่งๆ เหมือนกัน 2 ครั้ง

เยาวดี วิบูลย์ศรี (2539) กล่าวถึงความเที่ยงว่า ความเที่ยงหมายถึง ระดับความคงที่หรือความคงเส้นคงวาของคะแนนสอบจากการทดสอบเรื่องเดียวกันในเวลาใดก็ตาม

ศิริชัย กาญจนวาสี (2542) กล่าวถึงความเที่ยงว่า ความเที่ยงหมายถึง ความสม่ำเสมอหรือความคงเส้นคงวาของค่าที่วัดได้เมื่อทำการวัดซ้ำๆ ถ้าการวัดสิ่งเดียวกันหลายๆ ครั้ง ได้ค่าที่ค่อนข้างคงเส้นคงวาเพียงใด ก็ถือว่าการวัดมีความเที่ยงมากขึ้นเพียงนั้น

จากความหมายของความเที่ยงข้างต้น สรุปได้ว่าความเที่ยงหมายถึง ความคงที่ของคะแนนในการวัดแต่ละครั้ง สอบวัดผู้ตอบกลุ่มเดียวกันครั้งใดๆ ก็จะได้คะแนนในการสอบเหมือนเดิมหรือใกล้เคียงกันทุกครั้ง

ทฤษฎีของความเที่ยง (Theory of Reliability)

Mchrens และ Lehmann (1984) กล่าวว่า ทฤษฎีของความเที่ยงสามารถอธิบายได้โดยเริ่มจาก คะแนนที่สังเกตได้ (observed score) ซึ่งได้มาจากการใช้เครื่องมือในการวัด โดยจะประกอบด้วยคะแนน 2 ส่วน คือ คะแนนจริง (true score) และคะแนนความคลาดเคลื่อน (error score) โดยเขียนเป็นสมการได้ดังนี้

	X	=	$T + E$
เมื่อ	X	คือ	คะแนนที่สังเกตได้ (observed score)
	T	คือ	คะแนนจริง (true score)
	E	คือ	คะแนนความคลาดเคลื่อน (error score)

คะแนนจริง (true score) เป็นส่วนของคะแนนที่สังเกตได้ ซึ่งเป็นคะแนนที่ไม่ได้รับผลกระทบจากความคลาดเคลื่อนที่เกิดจากการสุ่ม (random error) (Mehrens และ Lehmann, 1984)

คะแนนความคลาดเคลื่อน (error score) เป็นค่าความผิดพลาดที่เกิดจากการวัด โดยเป็นความคลาดเคลื่อนที่เกิดจากการสุ่ม (random error) ซึ่งความคลาดเคลื่อนที่เกิดจากการสุ่มได้แก่ความคลาดเคลื่อนที่อาจเกิดเนื่องมาจากตัวผู้สอบ แบบสอบ การตรวจให้คะแนน การควบคุมการสอบ เป็นต้น

ถ้าความคลาดเคลื่อนเกิดขึ้นโดยการสุ่ม ตามทฤษฎีแล้วความคลาดเคลื่อนที่เกิดขึ้นจะเป็นไปได้ทั้งบวกและลบ ซึ่งจะหักล้างกันหมดไป จนค่าเฉลี่ยของความคลาดเคลื่อนเป็น 0 ถ้าความคลาดเคลื่อนเกิดขึ้นโดยการสุ่มจะไม่มีความสัมพันธ์กับคะแนนจริงหรือไม่มีความสัมพันธ์ซึ่งกันและกัน โดยสามารถเขียนสมการของความแปรปรวนของแบบสอบได้ ดังนี้

$$S_x^2 = S_t^2 + S_e^2 \quad \dots(1)$$

เมื่อ	S_x^2	คือ	ความแปรปรวนของคะแนนที่สังเกตได้
	S_t^2	คือ	ความแปรปรวนของคะแนนจริง
	S_e^2	คือ	ความแปรปรวนของคะแนนความคลาดเคลื่อน

โดยทฤษฎีแล้ว ความเที่ยง (r_{tt}) คือ อัตราส่วนของความแปรปรวนของคะแนนจริงและคะแนนที่สังเกตได้ สามารถเขียนสมการได้ ดังนี้

$$r_{tt} = \frac{S_t^2}{S_x^2} \quad \dots(2)$$

จากสมการที่ (1) และ (2) จะได้สมการพื้นฐานในการหาค่าสัมประสิทธิ์สหสัมพันธ์ของความเที่ยง คือ

$$r_{tt} = 1 - \frac{S_e^2}{S_x^2}$$

การประมาณค่าความเที่ยงของแบบสอบ

ในการประมาณค่าความเที่ยงของแบบสอบสามารถทำได้หลายวิธี และในแต่ละวิธีก็มีข้อจำกัดที่แตกต่างกัน การที่จะเลือกใช้วิธีใดจะขึ้นอยู่กับความยุ่งเหยิง ลักษณะของคะแนน และชนิดของแบบสอบ (Guilford, 1978) ซึ่งในที่นี้จะนำเสนอวิธีการประมาณค่าความเที่ยงตามประเภทของ

แบบสอบ คือ การประมาณค่าความเที่ยงของแบบสอบอิงกลุ่ม (norm-referenced test) และการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ (criterion-referenced test) โดยมีรายละเอียดดังนี้

การประมาณค่าความเที่ยงของแบบสอบอิงกลุ่ม (Norm-Referenced Test)

การประมาณค่าความเที่ยงของแบบสอบอิงกลุ่ม แบ่งออกเป็น 3 วิธีคือ การประมาณค่าความเที่ยงแบบความคงที่ (stability) การประมาณค่าความเที่ยงแบบใช้แบบสอบคู่ขนาน (equivalence) และการประมาณค่าความเที่ยงแบบความสอดคล้องภายใน (internal consistency) ซึ่งมีรายละเอียดดังนี้

1. การประมาณค่าความเที่ยงแบบความคงที่ (Stability)

ความเที่ยงแบบความคงที่ (stability) หรือความเที่ยงแบบสอบซ้ำ (test-retest) หมายถึง ความคงเส้นคงวาของคะแนนจากการวัดในช่วงเวลาต่างกัน (ศิริชัย กาญจนวาสี, 2542)

วิธีการประมาณค่าความเที่ยงแบบความคงที่หรือความเที่ยงแบบสอบซ้ำ เป็นวิธีที่ต้องนำแบบสอบเดียวกันไปทดสอบกับกลุ่มผู้สอบซึ่งเป็นกลุ่มเดียวกันจำนวน 2 ครั้ง โดยเว้นระยะห่างพอสมควร เพื่อป้องกันการจำคำตอบของแบบสอบได้ ซึ่งโดยปกติทั่วไปนิยมเว้นระยะห่างประมาณ 2 สัปดาห์ แล้วนำคะแนน 2 ชุด จากการทดสอบแต่ละครั้งมาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ โดยใช้สูตรของเพียร์สัน (Pearson product moment correlation coefficient) คือ

$$r_{tt} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

เมื่อ	r_{tt}	คือ ค่าความเที่ยงของแบบสอบ
	X	คือ คะแนนจากการสอบครั้งที่ 1
	Y	คือ คะแนนจากการสอบครั้งที่ 2
	N	คือ จำนวนผู้สอบ

2. การประมาณค่าความเที่ยงแบบใช้ข้อสอบคู่ขนาน (Equivalence)

การประมาณค่าความเที่ยงแบบใช้ข้อสอบคู่ขนาน (equivalence) เป็นวิธีที่ต่างจากการประมาณค่าความเที่ยงแบบสอบซ้ำ (test-retest) กล่าวคือ การประมาณค่าความเที่ยงแบบใช้ข้อสอบคู่ขนานจะต้องใช้แบบสอบ 2 ฉบับที่มีเนื้อหาเท่าๆกัน มีค่าเฉลี่ยและความแปรปรวนของแบบสอบเท่ากัน ทำการทดสอบนักเรียนในวันเดียวกัน แล้วนำผลการทดสอบทั้ง 2 ฉบับมาประมาณค่าความเที่ยง

การประมาณค่าความเที่ยงด้วยวิธีใช้แบบสอบคู่ขนานและวิธีสอบซ้ำจะได้ค่าความเที่ยงที่แตกต่างกัน การที่จะเลือกใช้วิธีการประมาณค่าความเที่ยงที่เหมาะสมจึงขึ้นอยู่กับความมุ่งหมายของ

การทดสอบ หากจะใช้การทดสอบสำหรับทำนายผลการเรียนในระยะยาว ควรที่จะใช้วิธีการสอบซ้ำ แต่หากต้องนำคะแนนไปใช้เพื่อพิจารณาความรู้ของนักเรียนในหัวข้อเนื้อหาวิชา จะใช้วิธีใช้ข้อสอบคู่ขนาน (Mehrens และ Lehmann, 1984)

3. การประมาณค่าความเที่ยงแบบความสอดคล้องภายใน (Internal Consistency)

ความเที่ยงแบบความสอดคล้องภายใน (internal consistency) หมายถึง ความสอดคล้องกันระหว่างคะแนนรายข้อหรือความเป็นเอกพันธ์ของเนื้อหาหรือข้ออันเป็นตัวแทนของคุณลักษณะเด่นอันเดียวกันที่ต้องการวัด (ศิริชัย กาญจนวาสี, 2542)

การประมาณค่าความเที่ยงแบบความสอดคล้องภายในเป็นวิธีที่ใช้แก้ปัญหของวิธีการสอบซ้ำ โดยการนำแบบสอบไปทดสอบกับกลุ่มตัวอย่างกลุ่มเดียว จำนวน 1 ครั้ง แล้วนำมาประมาณค่าความเที่ยง ซึ่งมี 4 วิธี คือ การประมาณค่าความเที่ยงแบบแบ่งครึ่งข้อสอบ (split-half), การประมาณค่าความเที่ยงด้วยวิธีคูเดอร์-ริชาร์ดสัน (Kuder-Richardson estimates), การประมาณค่าความเที่ยงด้วยวิธีหาค่าแอลฟาของครอนบาค (Cronbach's alpha) และการประมาณค่าความเที่ยงด้วยวิธีฮอยท์ (Hoyt's method) ซึ่งมีรายละเอียดดังนี้

(ก) การประมาณค่าความเที่ยงแบบแบ่งครึ่งข้อสอบ (split-half)

การประมาณค่าความเที่ยงแบบแบ่งครึ่งข้อสอบ จะทำการแบ่งคะแนนที่ได้จากการทดสอบออกเป็น 2 ชุดย่อย โดยแบ่งการตรวจเป็น 2 ครั้งๆ ละครึ่งฉบับ ซึ่งการแบ่งครึ่งฉบับของแบบสอบสามารถทำได้หลายวิธี เช่น การแบ่งครึ่งข้อสอบเป็นข้อคู่และข้อคี่ การแบ่งครึ่งข้อสอบเป็นครึ่งฉบับแรกและครึ่งฉบับหลัง โดยวิธีนี้จะมีข้อตกลงเบื้องต้นว่า แบบสอบที่ทำการแบ่งครึ่งออกเป็น 2 ฉบับนั้นจะต้องเป็นแบบสอบคู่ขนานหรือทดสอบเทียบกัน

การประมาณค่าความเที่ยง สามารถทำได้โดยการนำคะแนน 2 ชุด ไปหาค่าสหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation coefficient) จากนั้นทำการปรับค่าความเที่ยงครึ่งฉบับให้เป็นค่าความเที่ยงแบบเต็มฉบับ โดยใช้สูตรของสเปียร์แมน-บราวน์ (Spearman-Brown) คือ

$$r_{tt} = \frac{2 r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}}$$

เมื่อ r_{tt} คือ ค่าความเที่ยงของแบบสอบเต็มฉบับ
 $r_{\frac{1}{2}\frac{1}{2}}$ คือ ค่าความเที่ยงของแบบสอบครึ่งฉบับ

(ข) การประมาณค่าความเที่ยงด้วยวิธีคูเดอร์-ริชาร์ดสัน (Kuder-Richardson estimates)

การประมาณค่าความเที่ยงด้วยวิธีคูเดอร์-ริชาร์ดสัน เป็นวิธีการคำนวณค่าความเที่ยงโดยไม่ต้องแบ่งครึ่งข้อสอบ ซึ่งจะใช้สูตรการประมาณค่าความเที่ยงของ Kuder-Richardson (KR) ซึ่งมี 2 สูตร คือ KR-20 และ KR-21

ข้อตกลงเบื้องต้นของการใช้สูตร KR-20 และ KR-21 ในการประมาณค่าความเที่ยง คือ (เขาวดี วิบูลย์ศรี, 2539)

- (1) ข้อสอบในแบบสอบสามารถที่จะให้คะแนนในกรณีที่ตอบถูกคือ จะได้คะแนนเท่ากับ 1 คะแนน และตอบผิดจะได้คะแนนเท่ากับ 0 คะแนน นั่นคือ คะแนนรวมของแบบสอบได้แก่คะแนนรวมของข้อสอบที่ตอบถูก
 - (2) ข้อสอบในแบบสอบจะต้องมีลักษณะที่เป็นเอกพันธ์คือ วัดคุณลักษณะเดียวกัน
- สูตรที่ใช้ในการประมาณค่าความเที่ยงของคูเดอร์-ริชาร์ดสัน

สูตร KR-20

$$r_{tt} = \frac{k}{k-1} \left[1 - \frac{\sum pq}{S^2} \right]$$

สูตร KR-21

$$r_{tt} = \frac{k}{k-1} \left[1 - \frac{X(k-X)}{k S^2} \right]$$

เมื่อ	r_{tt}	คือ ค่าความเที่ยงของแบบสอบ
	k	คือ จำนวนข้อสอบทั้งฉบับ
	p	คือ สัดส่วนของผู้ตอบถูกในแต่ละข้อ
	q	คือ สัดส่วนของผู้ตอบผิดในแต่ละข้อ ($q = 1 - p$)
	S^2	คือ ค่าความแปรปรวนของคะแนนรวมของผู้สอบทั้งหมด
	X	คือ ค่าเฉลี่ยของคะแนนจากแบบสอบ

(ค) การประมาณค่าความเที่ยงด้วยวิธีหาค่าแอลฟาของครอนบาค (Cronbach's alpha)

การประมาณค่าความเที่ยงด้วยวิธีหาค่าแอลฟาของครอนบาค เป็นวิธีที่คล้ายกับวิธีการแบ่งครึ่งข้อสอบ โดยวิธีนี้เป็นวิธีที่นิยมใช้เพราะสามารถนำไปใช้กับแบบสอบประเภทต่างๆ ได้ แม้จะเป็นแบบสอบที่มีค่าคะแนนเฉพาะในแต่ละข้อที่ต่างกัน (เขาวดี วิบูลย์ศรี, 2539) สูตรที่ใช้ในการประมาณค่าความเที่ยง คือ

$$\alpha = \frac{[k]}{k-1} \left[1 - \frac{\sum S^2}{S^2} \right]$$

- เมื่อ k คือ จำนวนข้อสอบในแบบสอบ
- ΣS^2 คือ ผลรวมของความแปรปรวนของการกระจายของคะแนนในข้อสอบข้อต่าง ๆ
- S^2 คือ ความแปรปรวนของการกระจายของคะแนนสอบทั้งหมด

(ง) การประมาณค่าความเที่ยงด้วยวิธีฮอยท์ (Hoyt's Method)

การประมาณค่าความเที่ยงด้วยวิธีฮอยท์เป็นวิธีที่จะให้ค่าความเที่ยงเหมือนกับวิธีคูเดอร์-ริชาร์ดสัน 20 (Kuder-Richardson : KR-20) หรือวิธีหาค่าแอลฟาของครอนบาค (Cronbach's alpha) (Mehrens และ Lehmann, 1984) ซึ่งทำการประมาณค่าความเที่ยงโดยใช้สูตร

$$r_{tt} = \frac{1 - MS_{error}}{MS_{person}}$$

- เมื่อ MS_{error} คือ คะแนนความแปรปรวนของความคลาดเคลื่อน (error)
- MS_{person} คือ คะแนนความแปรปรวนระหว่างคน (between person)

การประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ (Criterion-Referenced Test)

การประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ แบ่งออกเป็น 3 วิธีคือ การประมาณค่าความเที่ยงแบบความคงที่ของผลการตัดสินใจความรอบรู้ (decision stability reliability) การประมาณค่าความเที่ยงแบบความสอดคล้องของผลการตัดสินใจความรอบรู้ (decision consistency reliability) และการประมาณค่าความเที่ยงจากการสอบครั้งเดียว ดังรายละเอียดต่อไปนี้

1. การประมาณค่าความเที่ยงแบบความคงที่ของผลการตัดสินใจความรอบรู้ (Decision Stability Reliability)

การประมาณค่าความเที่ยงแบบความคงที่ของผลการตัดสินใจความรอบรู้ สามารถทำได้โดยการนำแบบสอบฉบับเดียวกัน ไปทดสอบซ้ำ 2 ครั้ง จากนั้นนำคะแนนที่ได้มาหาค่าความคงที่ของการรอบรู้และไม่รอบรู้ โดยใช้คะแนนจุดตัด (cut-off score) เป็นจุดตัดสินใจ

		การทดสอบครั้งที่ 1		
		ไม่รอบรู้	รอบรู้	รวม
การทดสอบครั้งที่ 2	ไม่รอบรู้	A	B	A+B
	รอบรู้	C	D	C+D
รวม		A+C	B+D	N

สูตรที่ใช้ในการประมาณค่าความเที่ยง คือ

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

2. การประมาณค่าความเที่ยงแบบความสอดคล้องของผลการตัดสินความรอบรู้ (Decision Consistency Reliability)

การประมาณค่าความเที่ยงแบบความสอดคล้องของผลการตัดสินความรอบรู้ สามารถทำได้โดยการนำแบบสอบเดียวกันหรือแบบสอบคู่ขนานไปทดสอบซ้ำ 2 ครั้ง จากนั้นนำคะแนนที่ได้มาหาค่าความสอดคล้องของการรอบรู้และไม่รอบรู้ โดยใช้คะแนนจุดตัด (cut-off score) เป็นจุดตัดสิน ซึ่งสูตรที่ใช้ในการประมาณค่าความเที่ยงมีหลายสูตร คือ

(ก) วิธีของคาร์เวอร์ (Carver, 1972)

$$r_{cc} = \frac{A + D}{N}$$

(ข) วิธีของแฮมเบิลตันและโนวิก (Hambleton and Novick, 1973)

$$P = P_A + P_D$$

เมื่อ P_A คือ สัดส่วนของผู้ไม่รอบรู้ทั้ง 2 ครั้ง
 P_D คือ สัดส่วนของผู้รอบรู้ทั้ง 2 ครั้ง

(ค) วิธีโคเฮน แคปปา (Cohen's Kappa)

$$K = \frac{P - P_C}{1 - P_C}$$

เมื่อ P คือ สัดส่วนของความสอดคล้องในการจำแนกว่าเป็นผู้รอบรู้และผู้ไม่รอบรู้ เท่ากับ $P_A + P_D$
 P_C คือ สัดส่วนของความสอดคล้องที่คาดหวังไว้ เท่ากับ

$$\frac{[(A+B)(A+C) + (B+D)(C+D)]}{N \quad N \quad N \quad N}$$

3. การประมาณค่าความเที่ยงจากการสอบครั้งเดียว

การประมาณค่าความเที่ยงจากการสอบครั้งเดียว สามารถหาได้จากสูตรของลิวิงสตัน (Livingston, 1972) (ศิริชัย กาญจนวาสี, 2542)

$$K^2 = \frac{\sigma^2 (KR-20) + (M_x - KC)^2}{\sigma^2 + (M_x - KC)^2}$$

เมื่อ K^2 คือ ค่าความเที่ยงของแบบสอบอิงเกณฑ์
 σ^2 คือ ความแปรปรวนของคะแนน
 KR-20 คือ ความเที่ยงของแบบสอบตามสูตร KR-20
 M_x คือ คะแนนเฉลี่ยของแบบสอบอิงเกณฑ์

- K คือ จำนวนข้อสอบ
C คือ สัดส่วนของคะแนนเกณฑ์

3. ความยากง่าย (Item Difficulty)

Ebel (1972) กล่าวถึงความยากง่ายของข้อสอบว่า ความยากง่ายของข้อสอบ คือ สัดส่วนของกลุ่มผู้สอบที่ตอบข้อสอบถูก

Kubiszyn และ Borich (1984) กล่าวถึงความยากง่ายของข้อสอบว่า ความยากง่ายของข้อสอบ (ใช้ตัวย่อว่า p) คือ สัดส่วนของนักเรียนที่ตอบข้อสอบถูก

Mehrens และ Lehmann (1984) กล่าวถึงความยากง่ายของข้อสอบว่า ความยากง่ายของข้อสอบ คือ อัตราส่วนระหว่างคะแนนเฉลี่ยและคะแนนรวมของแบบสอบ ตัวอย่างเช่น ถ้าแบบสอบมีคะแนนสูงสุด 100 คะแนน คะแนนเฉลี่ยเท่ากับ 70 คะแนน แบบสอบนี้จะมีค่าความยากง่ายเป็น .70

จากคำกล่าวข้างต้น สามารถสรุปได้ว่า ความยากง่ายของข้อสอบ คือ สัดส่วนของจำนวนผู้สอบที่ตอบข้อสอบข้อนั้นๆ ได้ถูกต้อง โดยระดับความยากง่ายของข้อสอบจะมีพิสัยอยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อใดมีค่าความยากง่ายเข้าใกล้ 0 แสดงว่าข้อสอบข้อนั้นยาก และข้อสอบข้อใดที่มีค่าความยากง่ายเข้าใกล้ 1 แสดงว่าข้อสอบข้อนั้นง่าย

ข้อสอบที่ดีควรที่จะมีค่าความยากง่ายอยู่ในระดับที่เหมาะสม ไม่ยากหรือง่ายจนเกินไป กล่าวคือ ในแบบสอบฉบับหนึ่งๆ ควรที่จะมีข้อสอบที่มีค่าความยากง่ายในระดับปานกลางจำนวนมากข้อ และมีข้อสอบที่มีความยากหรือค่อนข้างยากและข้อสอบที่มีความง่ายหรือค่อนข้างง่ายจำนวนน้อยข้อ

วิธีการหาค่าความยากง่ายของแบบสอบ

วิธีการหาค่าความยากง่ายของแบบสอบในที่นี้จะนำเสนอโดยแบ่งออกเป็น 2 วิธีตามประเภทของแบบสอบ คือ การหาค่าความยากง่ายของแบบสอบอิงกลุ่ม (norm-referenced test) และการหาค่าความยากง่ายของแบบสอบอิงเกณฑ์ (criterion-referenced test) ดังต่อไปนี้

(ก) การหาค่าความยากง่ายของแบบสอบอิงกลุ่ม (Norm-Referenced Test)

การหาค่าความยากง่ายของแบบสอบอิงกลุ่ม สามารถทำได้โดยการแบ่งผู้สอบออกเป็นกลุ่มสูงและกลุ่มต่ำ กลุ่มละเท่าๆ กันเมื่อกลุ่มผู้สอบมีขนาดเล็ก (มีจำนวน 40 คนหรือน้อยกว่า) (Mehrens และ Lehmann, 1984) หากกลุ่มผู้สอบมีขนาดใหญ่และการแจกแจงของคะแนนสอบเป็นการแจกแจงแบบปกติ (normal distribution) ควรใช้กลุ่มสูงและกลุ่มต่ำเพียงกลุ่มละ 27% แต่ถ้าการแจกแจงของคะแนนสอบไม่เป็นปกติ จะต้องใช้กลุ่มสูงและกลุ่มต่ำกลุ่มละ 33% (ศิริชัย กาญจนวาสี, 2542)

สูตรที่ใช้ในการคำนวณค่าความยากง่าย คือ

$$p = \frac{p_H + p_L}{2}$$

เมื่อ p คือ ค่าความยากง่ายของข้อสอบ
 p_H คือ สัดส่วนของผู้ตอบถูกในกลุ่มสูง
 p_L คือ สัดส่วนของผู้ตอบถูกในกลุ่มต่ำ

โดยมีเกณฑ์ที่ใช้ตัดสินระดับความยากง่ายของข้อสอบ ดังนี้ (ศิริชัย กาญจนวาสี, 2542)

ตารางที่ 2.1 เกณฑ์ในการตัดสินความยากง่ายของแบบสอบ

ความยากง่ายของข้อสอบ (p)	ความหมาย
0.80 ถึง 1.00	ง่ายมาก
0.60 ถึง 0.79	ค่อนข้างง่าย
0.40 ถึง 0.59	ปานกลาง
0.20 ถึง 0.39	ค่อนข้างยาก
0 ถึง 0.19	ยากมาก

ข้อสอบที่ดีควรที่จะมีค่าความยากง่ายระหว่าง 0.20 ถึง 0.80

(ข) การหาค่าความยากง่ายของแบบสอบอิงเกณฑ์ (Criterion- Referenced Test)

การหาค่าความยากง่ายของแบบสอบอิงเกณฑ์ สามารถทำได้โดยการแบ่งกลุ่มผู้สอบตามกลุ่มเป้าหมาย (criterion group) ประกอบด้วย กลุ่มที่เคยเรียนเนื้อหาแล้วหรือกลุ่มที่รอบรู้แล้ว และกลุ่มที่ยังไม่เคยเรียนหรือกลุ่มที่ยังไม่รอบรู้ (ศิริชัย กาญจนวาสี, 2542)

สูตรที่ใช้ในการคำนวณค่าความยากง่าย คือ

$$P = \frac{C}{N} \times 100$$

เมื่อ P คือ ค่าความยากง่ายของข้อสอบ
 C คือ จำนวนผู้ตอบถูกในข้อนั้น
 N คือ จำนวนผู้ตอบทั้งหมด

โดยมีเกณฑ์ที่ใช้ตัดสินระดับความยากง่ายของข้อสอบ คือ

- (1) กลุ่มที่เรียนเนื้อหาหรือกลุ่มที่รอบรู้แล้ว ควรมีค่าความยากง่ายอยู่ในระดับ 70 ถึง 100%
- (2) กลุ่มที่ยังไม่เคยเรียนหรือกลุ่มที่ยังไม่รอบรู้ ควรมีค่าความยากง่ายอยู่ในระดับ 0 ถึง 40%

4. ค่าอำนาจจำแนก (Discrimination)

Ebel (1972) กล่าวถึงค่าอำนาจจำแนกของข้อสอบว่า ค่าอำนาจจำแนกของข้อสอบเป็นดัชนีชี้ถึงความแตกต่างระหว่างนักเรียนเก่งและนักเรียนอ่อนในสัดส่วนของการตอบข้อสอบได้ถูกต้อง

เยาวดี วิบูลย์ศรี (2539) กล่าวถึงค่าอำนาจจำแนกของข้อสอบว่า ค่าอำนาจจำแนกของข้อสอบหมายถึง ความสามารถของข้อสอบที่จะจำแนกหรือแยกผู้สอบออกได้ตามระดับความสามารถ เช่น สามารถจำแนกคนเก่งออกจากคนอ่อน หรือคนที่มีความถนัดออกจากคนที่ไม่มีความถนัด เป็นต้น

ศิริชัย กาญจนวาสี (2542) กล่าวถึงค่าอำนาจจำแนกของข้อสอบว่า ค่าอำนาจจำแนกของข้อสอบหมายถึง ความสามารถของข้อสอบที่จะจำแนกหรือแยกให้เห็นความแตกต่างระหว่างคนที่มีผลสัมฤทธิ์ต่างกัน

จากคำกล่าวข้างต้น สามารถสรุปได้ว่าค่าอำนาจจำแนกของข้อสอบ หมายถึง ระดับความสามารถของข้อสอบในการจำแนกผู้สอบที่มีความสามารถแตกต่างกันออกจากกัน

ค่าอำนาจจำแนกของข้อสอบจะมีพิสัยอยู่ระหว่าง -1 ถึง $+1$ ถ้าค่าอำนาจจำแนกของข้อสอบที่ได้มีค่าเป็นบวก แสดงว่าข้อสอบนั้นสามารถจำแนกกลุ่มผู้สอบได้ดี กล่าวคือ สัดส่วนของผู้ที่ตอบถูกในกลุ่มเก่งจะมากกว่ากลุ่มอ่อน ถ้าค่าอำนาจจำแนกของข้อสอบที่ได้มีค่าเป็นศูนย์ แสดงว่าข้อสอบนั้นไม่สามารถจำแนกกลุ่มผู้สอบได้ และถ้าค่าอำนาจจำแนกของข้อสอบที่ได้มีค่าเป็นลบ แสดงว่าข้อสอบมีอำนาจจำแนกกลับกัน กล่าวคือ สัดส่วนของผู้ตอบถูกในกลุ่มอ่อนจะมากกว่ากลุ่มเก่ง

วิธีการหาค่าอำนาจจำแนกของแบบสอบ

วิธีการหาค่าอำนาจจำแนกของแบบสอบในที่นี้จะนำเสนอโดยแบ่งออกเป็น 2 วิธีตามประเภทของแบบสอบ คือ การหาค่าอำนาจจำแนกของแบบสอบอิงกลุ่ม (norm-referenced test) และการหาค่าอำนาจจำแนกของแบบสอบอิงเกณฑ์ (criterion-referenced test) ดังต่อไปนี้

(ก) การหาค่าอำนาจจำแนกของแบบสอบอิงกลุ่ม (Norm-Referenced Test)

การหาค่าอำนาจจำแนกของแบบสอบอิงกลุ่ม สามารถทำได้โดยการแบ่งกลุ่มผู้สอบออกเป็นกลุ่มสูงและกลุ่มต่ำ และคำนวณค่าอำนาจจำแนกของข้อสอบ

สูตรที่ใช้ในการหาค่าอำนาจจำแนก คือ

	p	=	$p_H - p_L$
เมื่อ	p	คือ	ค่าอำนาจจำแนกของข้อสอบ
	p_H	คือ	สัดส่วนของผู้ตอบถูกในกลุ่มสูง
	p_L	คือ	สัดส่วนของผู้ตอบถูกในกลุ่มต่ำ

โดยมีเกณฑ์การแปลความหมายของค่าอำนาจจำแนกของข้อสอบ ดังนี้ (Ebel, 1972)

ตารางที่ 2.2 เกณฑ์ในการตัดสินค่าอำนาจจำแนกของแบบสอบ

ค่าอำนาจจำแนกของข้อสอบ (r)	ความหมาย
0.40 หรือมากกว่า	เป็นข้อสอบที่ดีมาก
0.30 ถึง 0.39	เป็นข้อสอบที่ดีพอสมควร อาจจะต้องปรับปรุง
0.20 ถึง 0.29	เป็นข้อสอบที่พอใช้ได้ แต่ต้องปรับปรุง
0.19 หรือต่ำกว่า	เป็นข้อสอบที่ไม่ดี ควรตัดทิ้งหรือแก้ไข

ข้อสอบที่ดี ควรมีค่าอำนาจจำแนกมากกว่าหรือเท่ากับ 0.20

(ข) การหาค่าอำนาจจำแนกของแบบสอบอิงเกณฑ์ (Criterion- Referenced Test)

การหาค่าอำนาจจำแนกของแบบสอบอิงเกณฑ์ สามารถทำได้โดยการพิจารณาความแตกต่างระหว่างสัดส่วนของผู้ที่ตอบถูกในกลุ่มที่ได้รับการเรียน (instructed group) กับกลุ่มก่อนเรียน (uninstructed group) (ศิริชัย กาญจนวาสี, 2542)

สูตรที่ใช้คำนวณค่าอำนาจจำแนก คือ

$$S_i = P_{\text{post}} - P_{\text{pre}}$$

เมื่อ S_i คือ ค่าความไวในการจำแนกกลุ่มของผู้สอบข้อที่ i

P_{post} คือ สัดส่วนการตอบถูกหลังเรียน

P_{pre} คือ สัดส่วนการตอบถูกก่อนเรียน

โดยมีเกณฑ์การตัดสิน คือ ค่าความไวในการจำแนกกลุ่มของผู้สอบต้องมากกว่า 0

ตอนที่ 3 ปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบ

3.1 งานวิจัยในประเทศที่เกี่ยวกับคุณภาพของแบบสอบ

จากการศึกษาวิทยานิพนธ์และรายงานการวิจัยเกี่ยวกับแบบสอบเลือกตอบภายในประเทศ จากจุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ มหาวิทยาลัยเกษตรศาสตร์ และรายงานการวิจัย จำนวน 96 เรื่อง ซึ่งพิมพ์เผยแพร่ระหว่างปี 2511 ถึง 2542 พบว่า

งานวิจัยที่พบจะทำการศึกษาเกี่ยวกับคุณภาพของแบบสอบเลือกตอบ โดยมีวัตถุประสงค์เพื่อทำการตรวจสอบคุณภาพของแบบสอบทั้งค่าความเที่ยง (reliability) ความตรง (validity) ความยากง่าย (item difficulty) และค่าอำนาจจำแนก (discrimination) เมื่อแบบสอบมีลักษณะแตกต่างกัน ซึ่งสามารถแบ่งออกได้เป็น 4 กลุ่ม คือ กลุ่มแรกคือ ด้านรูปแบบของแบบสอบ ทำการศึกษาในเรื่องจำนวนตัวเลือก ลักษณะของตัวเลือก รูปแบบตัวเลือก และการเรียงตัวเลือก กลุ่มที่ 2 คือ ด้านบริบท

ของแบบสอบ ทำการศึกษาในเรื่องคำชี้แจงในการตอบ และการเรียงลำดับข้อสอบ กลุ่มที่ 3 คือ ด้านวิธีการตรวจให้คะแนนแบบสอบ และกลุ่มสุดท้าย คือ ด้านเกี่ยวกับประเภทของแบบสอบ ซึ่งมีรายละเอียดดังตารางที่ 2.3

ตารางที่ 2.3 งานวิจัยที่ศึกษาเรื่องปัจจัยที่มีผลต่อคุณภาพของแบบสอบ

ตัวแปรต้น	ผู้วิจัย/ปีที่ทำ	ตัวแปรตาม				กลุ่มตัวอย่าง	แบบสอบวิชา
		ความเที่ยง	ความตรง	ความยาก	อำนาจจำแนก		
จำนวนตัวเลือก	วารุณี ปิณฑวิชัย (2514)	✓		✓	✓	นร. ป.7	ภาษา
	มาลี ชิดสวน (2515)	✓	✓	✓	✓	นร. ม.3	วิทยาศาสตร์
	ครุณ หาญตระกูล (2518)	✓	✓		✓	นร. ป.7	คณิตศาสตร์
	ชนะ ทานะวงศ์ (2520)	✓	✓			นร. ป.4	วิทยาศาสตร์
	วิภาดา วัฒนกุลกิตติ (2528)	✓		✓		นร. ม.2	วิทยาศาสตร์
	ธีระศักดิ์ อูร์จนาพันธ์(2530)	✓				นร. ม.1	ภาษา
	วิชัย สันวิวัฒนาพานิช(2531)	✓		✓	✓	นร. ม.4	ฟิสิกส์
	นิสากร ล้อถาวร (2534)	✓		✓	✓	นร. ม.2	อนุกรมมิติ
	ปณยา แพร่เจริญวัฒนา (2540)	✓				นร.ม.1	-
ลักษณะตัวเลือก	นवलน้อย แต่บรรพกุล(2519)	✓	✓	✓	✓	นร. ป.7	ภาษาไทย
	นิรมล บุญตระกูล (2524)	✓	✓	✓	✓	นร. ม.1	คณิตศาสตร์
	จินดา โตอนันต์ (2527)	✓		✓	✓	นร. ม.1	วิทยาศาสตร์
	ปราณี ร่วมทอง (2527)	✓	✓	✓	✓	นร. ม.3	วิทยาศาสตร์
	วรัญญา ปราบปัจจะ(2528)	✓		✓		นร. ม.2	วิทยาศาสตร์
	สุพัทธรา เสวตะกุล (2528)	✓	✓	✓	✓	นร. ม.1	คณิตศาสตร์
	กมลทิพย์ อธิการยานันท์(2529)	✓	✓			นร. ม.2	วิทยาศาสตร์
	อำนวยการ ทองคัง (2530)	✓		✓	✓	นร. ป.6	คณิตศาสตร์
	พรทิพย์ แดงน้อย (2531)	✓	✓	✓	✓	นร. ม.3	คณิตศาสตร์
	วรรณพิศ นามเย็น (2531)	✓		✓	✓	นร. ม.1	คณิตศาสตร์
	จิราพร เนียมสุวรรณ(2532)	✓				นร. ม.2	คณิตศาสตร์
	เสรี สุกันตะ (2534)	✓		✓	✓	นร. ป.6	คณิตศาสตร์
	สมบูรณ์ เรืองแก้ว (2534)	✓		✓	✓	นร. ม.3	คณิตศาสตร์
	ยศลินี วิพชานานนท์ (2537)	✓		✓	✓	นร. ป.6	ภาษาไทย
	ฉัตรศิริ ปิยะพิมลสิทธิ์ (2539)	✓	✓			นร. ม.2	ภาษาไทย

ตารางที่ 2.3 (ต่อ)

ตัวแปรต้น	ผู้วิจัย / ปีที่ทำ	ตัวแปรตาม				กลุ่มตัวอย่าง	แบบสอบวิชา
		ความเที่ยง	ความตรง	ความยาก	อำนาจจำแนก		
รูปแบบตัว लग	สนั่น สิทธิวง (2511)	✓	✓	✓	✓	นร. ป.5	คณิตศาสตร์
	สมบูรณ์ วัฒนา (2518)	✓	✓	✓	✓	นักศึกษา	วิทยาศาสตร์
	บุญส่ง หาญพานิช (2521)	✓	✓	✓	✓	นักศึกษา	คณิตศาสตร์
	โกมล สรรคเสรมณี (2528)	✓	✓		✓	นร. ม.4	คณิตศาสตร์
	รณชัย ศรีสุธัญญาวงศ์ (2532)	✓		✓	✓	นร. ม.6	คณิตศาสตร์
การเรียงตัว เลือก	ชุมพล เขมาปทุมศักดิ์ (2525)	✓				นร. ม.2	คณิตศาสตร์
การชี้แจง ในการตอบ	ทัศนีย์ สุขสมบูรณ์ (2518)	✓				นร. ม. 3	คณิตศาสตร์
	ปราณีต เลิศไกร (2520)	✓				นร. ป.7	คณิตศาสตร์
	กิ่งแก้ว บุญฤทธิกุล (2525)	✓				นร. ม.4	เคมี
การเรียง ลำดับข้อ สอบ	เดือนใจ เศรษฐศักดิ์โก (2511)	✓	✓			นร. ป.7	ภาษาไทย, คณิตศาสตร์, วิทยาศาสตร์, อังกฤษ
	ศศิธร สุวรรณสุข (2518)	✓	✓			นร. ป.7	คณิตศาสตร์
	นิโลบล นิมกักรัตน์ (2519)	✓	✓	✓	✓	นร. ม.3	ภาษาไทยและ คณิตศาสตร์
	แนนน้อย เพชรรัตน์ (2520)	✓	✓	✓	✓	นร. ม.3	วิทยาศาสตร์
	ศิริพร พุ่มแก้ว (2521)	✓	✓			นร. ม.3	ภาษาไทยและ คณิตศาสตร์
	ปราณีต เลิศไกร (2521)	✓				นร. ป.7	คณิตศาสตร์
	โสภณา เหลืองวิลาวงษ์ (2523)	✓	✓	✓		นร. ม.3	คณิตศาสตร์
	เกื้อกูล อุดตมากร (2524)	✓	✓		✓	นร. ม.1	คณิตศาสตร์
	กิ่งแก้ว บุญฤทธิกุล (2525)	✓				นร. ม.4	เคมี
	สุทัศน์ ลักษณะวิลาส (2526)	✓	✓	✓	✓	นร. ม.2	คณิตศาสตร์
	จินตนา ต่อทรัพย์เงินชัย (2532)	✓	✓			นร. ม.3	คณิตศาสตร์
	มานะ คิวเจริญ (2538)	✓				นร. ม.2	คณิตศาสตร์
	อุดม ชูดีวรรณ (2538)	✓				นร. ม.1	คณิตศาสตร์
	ศิริวิมล สุขสมบูรณ์ (2540)	✓				นักศึกษา	-
	ดวงค์ ร่มเงิน (2540)	✓				นร. ม.2	คณิตศาสตร์
รัชณี ถิ่นภัย (2542)	✓	✓			นร. ม.3	คณิตศาสตร์	

ตารางที่ 2.3 (ต่อ)

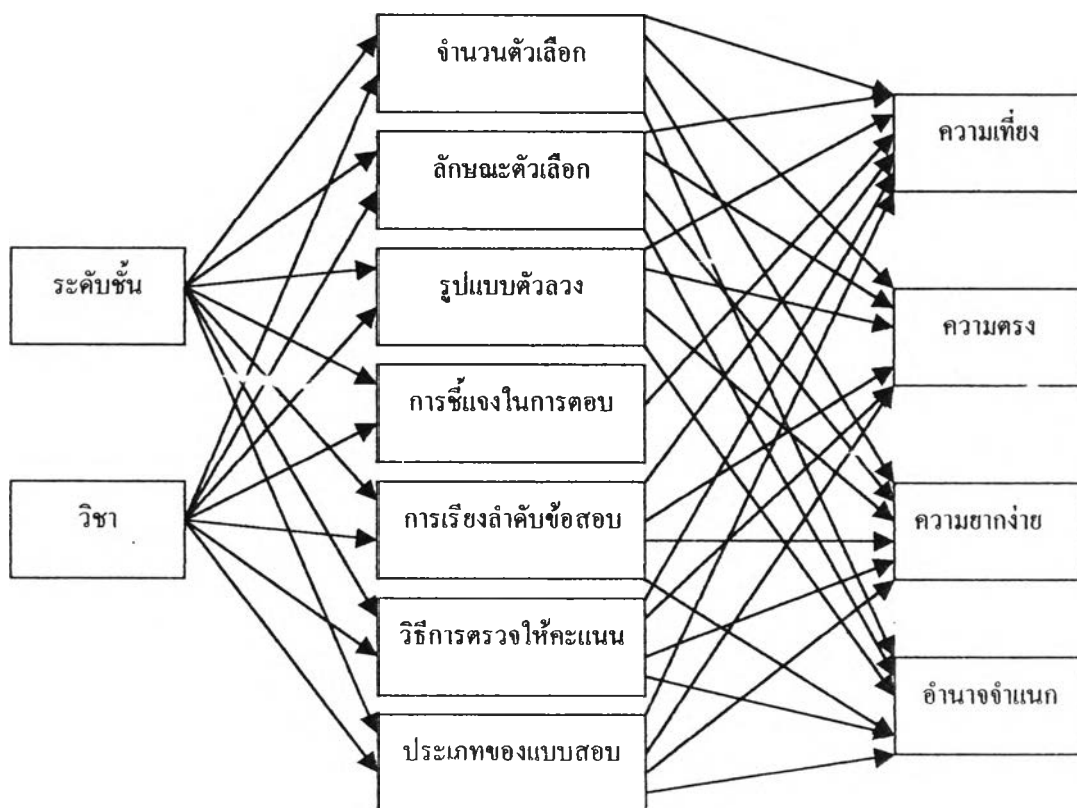
ตัวแปรตาม	ผู้วิจัย / ปีที่ทำ	ตัวแปรตาม				กลุ่มตัวอย่าง	แบบสอบวิชา
		ความ เที่ยง	ความ ตรง	ความ ยาก	อำนาจ จำแนก		
การตรวจ ให้คะแนน ด้วยวิธีต่าง กัน	อรรธรรม ดันท์เจริญรัตน์ (2516)	✓				นักศึกษา	ภาษาไทย
	กาญจนา ศรีวัฒนพงษ์ (2519)	✓	✓		✓	นร. ม.3	วิทยาศาสตร์
	ธีรศักดิ์ อินทรมาตย์ (2519)	✓	✓			นร. ป.7	วิทยาศาสตร์
	ชลัยทิพย์ เลิศกวีพร (2520)	✓		✓	✓	นร. ม.3	คณิตศาสตร์
	เพ็ญศรี สว่างเนตร (2520)	✓				นักศึกษา	อังกฤษ
	เกริกชัย ฮวบเจริญ (2524)	✓				นร. ป.6	วิทยาศาสตร์
	ตำรวจ มีแจ้ง (2525)	✓	✓		✓	นร. ม.4	คณิตศาสตร์
	ทวี ทองคำ (2526)	✓	✓		✓	ปวช.ป.1	คณิตศาสตร์
	นภา กาญจนกิจ โสภณ (2528)	✓				นร. ม.6	คณิตศาสตร์, ฟิสิกส์, วิทยาศาสตร์
	สุพรรณ สุกมลสันต์ (2528)	✓	✓			นิสิต	อังกฤษ
	พนิชา สวัสดิ์มงคล (2530)	✓	✓			นักศึกษา	การพยาบาล
	เพลินพิศ คุณคำ (2531)	✓	✓			นร. ป.2	วิทยาศาสตร์
	จินตนา ต่อทวีชัยสินชัย (2532)	✓	✓			นร. ม.3	คณิตศาสตร์
	พินิจ อุไรรักษ์ (2533)	✓	✓			นร. ม.1	วิทยาศาสตร์
	อนันต์ ศรีโสภณ (2533)	✓	✓			นักศึกษา	วิจัยเบื้องต้น
	เขवालักษณ์ น้อยนาแสง (2534)	✓	✓		✓	นร. ม.3	ความถนัด
	พัชนี ประภาสวัต (2534)	✓	✓			นร. ม.3	วิทยาศาสตร์
	วัชรภรณ์ จิตรมาศ (2537)	✓		✓	✓	นร. ป.5	สປข.
	มานะ คิวเจริญ (2538)	✓				นร. ม.2	คณิตศาสตร์
	ฉัตรศิริ ปิยะพิมลสิทธิ์ (2539)	✓	✓			นร. ม.2	ภาษาไทย
สมพงษ์ อิศวสุภฤกษ์ (2541)	✓				นักศึกษา	สถิติธุรกิจ	
วนิดา ปานโต (2542)	✓				ม.ปลาย	-	

ตารางที่ 2.3 (ต่อ)

ตัวแปรตาม	ผู้วิจัย / ปีที่ทำ	ตัวแปรตาม				กลุ่มตัวอย่าง	แบบสอบวิชา
		ความเที่ยง	ความตรง	ความยาก	อำนาจจำแนก		
ประเภทของแบบสอบต่างกัน	สุธรรม์ จันทร์หอม (2512)	✓	✓	✓	✓	นร. ป.7	คณิตศาสตร์
	รุจิรี ภู่อาระ (2513)	✓	✓	✓	✓	นร. ป.7	ภาษาไทย
	ไพบุลย์ จิตรโคต (2513)	✓	✓	✓	✓	นร. ป.4	คณิตศาสตร์
	นภาพร อมรเลิศสินไทย (2513)	✓	✓	✓	✓	นร. ป.7	สังคม
	ศุภชัย เอื้ออิสระวิมล (2518)	✓	✓	✓	✓	นร. ป.7	วิทยาศาสตร์
	เกื้อกุล ไทยน้อย (2521)	✓	✓	✓	✓	นักศึกษา	คณิตศาสตร์
	จิตรา หมั่นคำ (2521)	✓	✓	✓	✓	นร. ม.4	ฟิสิกส์
	มยุรี ยาพิมาย (2521)	✓	✓	✓	✓	นักศึกษา	คณิตศาสตร์
	รัชนี ชาญจักร (2521)	✓	✓	✓	✓	นร. ม.4	คณิตศาสตร์
	สุณีย์ ต่องประเสริฐ (2521)	✓	✓	✓	✓	นักศึกษา	คณิตศาสตร์
	ไพบุลย์ วิริยะวัฒนะ (2522)	✓	✓	✓	✓	นักศึกษา	ฟิสิกส์
	อรุณ มณีคำ (2523)	✓	✓			นร. ม.3	วิทยาศาสตร์
	สมหวัง พิธิยานุวัฒน์ และ เพ็ญศิริ คำนชนะ (2523)	✓	✓	✓		นักศึกษา	-
	ครองทอง เทพศิริอำนวย (2524)	✓	✓	✓	✓	นร. ม.5	คณิตศาสตร์
	นิภาพร จุลกมนตรี (2524)	✓		✓		นร. ม.4	คณิตศาสตร์
	ปนัดดา หารษาศิริพงษ์ (2528)	✓		✓		นร. ม.1	วิทยาศาสตร์
	สุรินทร์ แห่งจันทัก (2528)	✓				นร. ม.3	คณิตศาสตร์
	ประนอม พันธุ์สาว (2529)	✓				นร. ป.5	วิทยาศาสตร์
	สมศรี อุปริพพุทธิกุล (2529)	✓	✓	✓	✓	นร. ม.3	คณิตศาสตร์
	อมรรัตน์ แกมทอง (2529)	✓	✓			นร. ม.3	อังกฤษ
อัญชลี ปิ่นทองคำ (2531)	✓	✓	✓	✓	นร. ม.3	สังคม	
โกวิทย์ ทองอยู่ (2532)	✓	✓			นร. ป.6	คณิตศาสตร์	
กฤติกา เศษศรี (2535)	✓		✓	✓	นร. ป.3	ภาษาไทย	
จุฬารัตน์ ตันเจริญ (2537)	✓		✓	✓	นร. ม.3	คณิตศาสตร์	
ภูมิ หลอดกระโทก (2539)	✓	✓			นร. ป.5	ภาษาไทย	
บังอร กมลวัฒนา (2541)	✓				นร. ป.6	วิทยาศาสตร์	

จากตารางที่ 2.3 งานวิจัยภายในประเทศที่เกี่ยวกับแบบสอบหลายตัวเลือก เป็นการตรวจสอบคุณภาพของแบบสอบ โดยมีตัวแปรต้นเป็นปัจจัยเกี่ยวกับแบบสอบหลายตัวเลือก ได้แก่ จำนวนตัวเลือก ลักษณะตัวเลือก รูปแบบตัวลง การเรียงตัวเลือก การชี้แจงในการตอบ การเรียง

ลำดับข้อสอบ วิธีการตรวจให้คะแนน และประเภทของแบบสอบ ตัวแปรตามเป็นค่าคุณภาพของแบบสอบ ได้แก่ ความเที่ยง ความตรง ความยากง่าย และอำนาจจำแนก กลุ่มตัวอย่างที่ใช้ในงานวิจัยเป็นนักเรียนและนักศึกษา ระดับชั้นประถมศึกษา มัธยมศึกษา อาชีวศึกษาและอุดมศึกษา เครื่องมือที่ใช้เป็นแบบสอบหลายตัวเลือก และแบบสอบอื่นๆ เช่น แบบเติมคำ แบบโคลง เป็นต้น โดยส่วนใหญ่จะเป็นแบบสอบวิชาคณิตศาสตร์และวิทยาศาสตร์ ซึ่งสามารถสรุปเป็นแผนภาพปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบหลายตัวเลือก ดังแผนภาพที่ 2.1



แผนภาพที่ 2.1 ปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบหลายตัวเลือก

3.2 งานวิจัยต่างประเทศที่เกี่ยวกับคุณภาพของแบบสอบ

Marcus (1963) ได้ทำการศึกษาเรื่องผลจากตำแหน่งของคำตอบถูกต้องระดับความยากของคำถามของแบบสอบเลือกตอบ โดยมีวัตถุประสงค์เพื่อศึกษาอิทธิพลของการกำหนดตำแหน่งของตัวเลือกของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนแบบเลือกตอบ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนจำนวน 434 คน แบ่งเป็น 4 กลุ่มสำหรับแบบสอบแต่ละรูปแบบ (รูปแบบละ 100 ข้อ) การจัดเตรียมตัวเลือกถูกและตัวลวงใช้การสุ่มอย่างมีแบบแผนโดยให้มีจำนวนเท่าๆ กันของตัวเลือก โดยตัวเลือกถูกในแต่ละข้อจะอยู่ในตำแหน่งที่แตกต่างกัน ผลการศึกษาพบว่า ในแบบสอบเลือกตอบแบบหลาย

ตัวเลือกจะค่อนข้างมีอิสระในการวางตำแหน่งของตัวเลือก ซึ่งแบบสอบเลือกตอบนี้แสดงให้เห็นว่าตำแหน่งของตัวเลือกสามารถมองข้ามไปได้

Costin (1970) ได้ทำการศึกษาถึงจำนวนตัวเลือกที่ดีที่สุดในระบบสอบวัดผลสัมฤทธิ์ทางการเรียนประเภทหลายตัวเลือก กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาที่เรียนวิชาจิตวิทยาเบื้องต้น มหาวิทยาลัยอิลลินอยส์ จำนวน 200 คน เครื่องมือที่ใช้ เป็นแบบสอบวิชาจิตวิทยาเบื้องต้น ประกอบด้วยเนื้อหาจำนวน 4 เรื่อง คือ การรับรู้ (50 ข้อ) แรงงูใจ (60 ข้อ) การเรียนรู้ (60 ข้อ) และเขาวนัปัญหา (50 ข้อ) โดยในแต่ละเรื่องจะแบ่งแบบสอบออกเป็น 2 ฉบับ ฉบับที่ 1 มี 4 ตัวเลือก และฉบับที่ 2 มี 3 ตัวเลือก โดยการสุ่มตัดตัวลวงออก 1 ตัวเลือก ผลการวิจัยพบว่า

ตารางที่ 2.4 ค่าอำนาจจำแนก ความยากง่าย และค่าความเที่ยงจากการวิจัยของ Costin (1970)

เนื้อหา (เรื่อง)	จำนวนตัวเลือก	อำนาจจำแนก	ความยากง่าย	ความเที่ยง
การรับรู้	3	.44	63.5	.56
	4	.38	61.8	.50
แรงงูใจ	3	.47	64.5	.62
	4	.40	61.3	.52
การเรียนรู้	3	.41	65.5	.54
	4	.39	63.8	.52
เขาวนัปัญหา	3	.45	64.7	.57
	4	.37	61.1	.50

(1) ค่าอำนาจจำแนก พบว่า แบบสอบเลือกตอบแบบ 3 ตัวเลือก มีค่าอำนาจจำแนกสูงกว่าแบบสอบ 4 ตัวเลือก

(2) ค่าความยากง่าย พบว่า แบบสอบเลือกตอบแบบ 3 ตัวเลือก มีค่าความยากง่ายสูงกว่าแบบสอบ 4 ตัวเลือก

(3) ค่าความเที่ยง พบว่า แบบสอบเลือกตอบแบบ 3 ตัวเลือก มีค่าความเที่ยงสูงกว่าแบบสอบ 4 ตัวเลือก

Costin (1972) ได้ทำการศึกษาค่าความเที่ยงและความตรงของแบบสอบวัดผลสัมฤทธิ์เมื่อมีจำนวนตัวเลือกต่างกัน คือ 3 และ 4 ตัวเลือก กลุ่มตัวอย่างคือ นักศึกษาที่เรียนวิชาจิตวิทยาเบื้องต้น มหาวิทยาลัยอิลลินอยส์ จำนวน 1566 คน เครื่องมือที่ใช้เป็นแบบสอบเลือกตอบวิชาจิตวิทยาเบื้องต้น 2 ฉบับ ที่มีลักษณะเหมือนกัน แต่มีตัวเลือกต่างกันคือ แบบสอบฉบับที่ 1 มี 4 ตัวเลือก และแบบสอบฉบับที่ 2 มี 3 ตัวเลือก ซึ่งเกิดจากการสุ่มตัดตัวลวงออก 1 ตัวเลือก จำนวนฉบับละ 50 ข้อ ผลการวิจัยพบว่า ค่าความเที่ยงที่ได้ไม่มีความแตกต่างกัน (โดยแบบสอบ 4 ตัวเลือกมีค่าความเที่ยงเท่ากับ .78 และแบบสอบ 3 ตัวเลือกมีค่าความเที่ยงเท่ากับ .75)

Dudycha และ Carpenter (1973) ได้ทำการศึกษาเรื่องผลของรูปแบบของข้อสอบที่มีต่อค่าอำนาจจำแนกและค่าความยากของข้อสอบ โดยมีวัตถุประสงค์เพื่อรวบรวมข้อมูลเชิงประจักษ์เกี่ยวกับการเลือกกฎสำหรับโครงสร้างของข้อสอบเลือกตอบ ซึ่งได้กำหนดตัวแปรที่ใช้ในการวิจัย คือ ตัวแปรอิสระ ได้แก่ (1) รูปแบบของข้อสอบที่มีลักษณะคำถามเป็นแบบเปิดและแบบปิด (2) ข้อคำถามของข้อสอบเป็นแบบเชิงบวกและแบบเชิงลบ (3) ข้อสอบที่มีตัวเลือกถูกเพียงตัวเดียวและข้อสอบที่มีตัวเลือกปลายเปิด/ปลายปิด ตัวแปรตามคือ ค่าความยากและค่าอำนาจจำแนก ผลการวิจัยพบว่า (1) ค่าความยาก พบว่า ข้อสอบที่มีลักษณะคำถามเป็นแบบเปิด ($p = .53$) มีความยากกว่าข้อสอบที่มีลักษณะคำถามเป็นแบบปิด ($p = .54$) ข้อคำถามของข้อสอบที่เป็นเชิงลบ ($p = .50$) มีความยากกว่าข้อคำถามของข้อสอบที่เป็นเชิงบวก ($p = .58$) และข้อสอบที่มีตัวเลือกปลายเปิด/ปลายปิด ($p = .50$) มีความยากกว่าข้อสอบที่มีตัวเลือกถูกเพียงตัวเดียว ($p = .58$) (2) ค่าอำนาจจำแนก พบว่า เมื่อใช้ตัวเลือกปลายเปิด/ปลายปิด จำทำให้ค่าอำนาจจำแนกลดลง ข้อคำถามที่เป็นแบบเปิดเชิงบวกและข้อคำถามที่เป็นแบบปิดเชิงลบ จะให้ค่าอำนาจจำแนกลดลงกว่าข้อคำถามแบบอื่นๆ จึงอาจสรุปได้ว่ารูปแบบของข้อสอบจะมีผลต่อค่าคุณภาพของแบบสอบไม่ว่าจะเป็นค่าความยากและค่าอำนาจจำแนก

Pugh และ Brunza (1975) ได้ทำการศึกษาค่าความเที่ยงของแบบสอบเมื่อมีวิธีการตอบและตรวจให้คะแนนต่างกัน (วิธีการตอบแบบธรรมดาและวิธีตอบแบบให้ระบุความมั่นใจในการตอบ) เครื่องมือที่ใช้ในการวิจัยคือแบบสอบหลายตัวเลือกเกี่ยวกับคำศัพท์ ในแบบสอบวัดด้านสติปัญญา (I.E.R. Intelligence Scale, 1946) จำนวน 48 ข้อ โดยแบ่งออกเป็น 2 ส่วนๆ ละ 24 ข้อ กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาระดับบัณฑิตศึกษาที่ลงทะเบียนเรียนวิชาการวัดผลการศึกษา มหาวิทยาลัยอินเดียน่า จำนวน 84 คน (เป็นชาย 29 คนและหญิง 55 คน) แบ่งออกเป็น 3 กลุ่มๆ ละ 28 คน เพื่อทำแบบสอบ 3 ชุด คือ กลุ่มแรกทำแบบสอบชุด A เป็นการตอบแบบธรรมดาทั้งสองส่วน กลุ่มที่ 2 ทำแบบสอบชุด B เป็นการตอบแบบธรรมดาในส่วนที่หนึ่งและตอบแบบระบุความมั่นใจในส่วนที่สอง และกลุ่มที่ 3 ทำแบบสอบชุด C เป็นการตอบแบบระบุความมั่นใจทั้งสองส่วน ผลการศึกษาพบว่า

ตารางที่ 2.5 ค่าความเที่ยงจากการวิจัยของ Pugh และ Brunza (1975)

แบบสอบ	ชุด A			ชุด B		ชุด C		
	ส่วน 1	ส่วน 2	รวม 2 ส่วน	ส่วน 1	ส่วน 2	ส่วน 1	ส่วน 2	รวม 2 ส่วน
ค่าความเที่ยง	.48	.42	.57	.62	.48	.76	.78	.85

การตอบแบบสอบโดยระบุความมั่นใจในการตอบให้ค่าความเที่ยงสูงกว่าการตอบแบบธรรมดา

Abu-Sayf และ Diamond (1976) ได้ทำการศึกษาผลของการตอบแบบสอบถามโดยระบุความมั่นใจในแบบสอบถามหลายตัวเลือกที่มีต่อค่าความเที่ยงและค่าความตรง กลุ่มตัวอย่างคือ นักเรียนระดับชั้นมัธยมศึกษาจำนวน 152 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบถามวิชาคำศัพท์ภาษาอังกฤษ 4 ตัวเลือก จำนวน 60 ข้อ โดยวิธีการในการตอบแบบสอบถามจะให้ผู้ตอบเลือกคำตอบที่ตนคิดว่าถูกและให้ระบุระดับความมั่นใจในคำตอบนั้นๆ ด้วย ซึ่งมี 3 ระดับ คือ มั่นใจมาก มั่นใจน้อย และเดา วิธีการตรวจให้คะแนนจะทำการตรวจให้คะแนน 3 วิธี คือ วิธีแรกให้คะแนนเฉพาะข้อที่ถูกและตอบอย่างมั่นใจ วิธีที่สองให้คะแนนข้อที่ถูกโดยไม่ได้เดา และวิธีที่สามให้คะแนนข้อที่ถูกโดยไม่คำนึงถึงระดับความมั่นใจ แล้วทำการหาค่าความเที่ยงโดยใช้สูตรครอนบาคแอลฟา (Cronbach's alpha) และค่าสัมประสิทธิ์ความตรง ผลการวิจัยพบว่า

ตารางที่ 2.6 ค่าความเที่ยงและค่าสัมประสิทธิ์ความตรงจากการวิจัยของ Abu-Sayf และ Diamond (1976)

วิธีการตอบและตรวจให้คะแนน	จำนวนข้อที่ตอบถูก	จำนวนข้อที่ตอบผิด	จำนวนข้อที่เว้น	ค่าความเที่ยง	ค่าความตรง
ข้อที่ถูกและตอบอย่างมั่นใจ	2645	846	5629	.898	.219
ข้อที่ถูกโดยไม่ได้เดา	3610	2529	2981	.871	.260
ข้อที่ถูกโดยไม่คำนึงถึงระดับความมั่นใจ	4347	4574	199	.774	.236

(1) ค่าความเที่ยง วิธีการตรวจให้คะแนนเฉพาะข้อที่ถูกและระบุความมั่นใจให้ค่าความเที่ยงสูงที่สุด รองลงมาเป็นวิธีการตรวจให้คะแนนข้อที่ถูกโดยไม่ได้เดา และวิธีการตรวจให้คะแนนข้อที่ตอบถูกโดยไม่คำนึงถึงระดับความมั่นใจ ตามลำดับ

(2) ค่าความตรง วิธีการตรวจให้คะแนนข้อที่ถูกโดยไม่ได้เดาให้ค่าความตรงสูงที่สุด รองลงมาเป็นวิธีการตรวจให้คะแนนข้อที่ตอบถูกโดยไม่คำนึงถึงระดับความมั่นใจ และวิธีการตรวจให้คะแนนเฉพาะข้อที่ถูกและระบุความมั่นใจ ตามลำดับ

Straton และ Catts (1980) ทำการศึกษาเปรียบเทียบแบบสอบถาม 2, 3 และ 4 ตัวเลือก เมื่อมีการกำหนดจำนวนรวมของตัวเลือกทั้งหมดของแบบสอบถามเงื่อนไขของ Tversky กลุ่มตัวอย่างคือ นักเรียนเกรด 12 ที่เรียนวิชาเศรษฐศาสตร์ จำนวน 260 คน เครื่องมือที่ใช้เป็นแบบสอบถามเลือกตอบที่มีการกำหนดให้มีจำนวนรวมของตัวเลือกทั้งหมด 120 ตัวเลือกตามเงื่อนไขของ Tversky โดยแบ่งแบบสอบถามออกเป็น 4 รูปแบบ คือ (1) รูปแบบ A จำนวน 60 ข้อ 2 ตัวเลือก (2) รูปแบบ B จำนวน 40 ข้อ 3 ตัวเลือก (3) รูปแบบ C จำนวน 40 ข้อ 3 ตัวเลือก และ (4) รูปแบบ D จำนวน 30 ข้อ 4 ตัวเลือก โดยที่แบบสอบถามรูปแบบที่ 2 และ 3 แตกต่างกันที่ แบบสอบถามรูปแบบที่ 2 ตัดตัวลวงโดยการสุ่ม และแบบสอบถามรูปแบบที่ 3 ทำการตัดตัวลวงโดยผู้เชี่ยวชาญ ผลการวิจัยพบว่า

ตารางที่ 2.7 ค่าความยากง่ายและสัมประสิทธิ์ความเที่ยงจากการวิจัยของ Stratton and Catts (1980)

แบบสอบ	ความยากง่าย	สปส.ความเที่ยง
รูปแบบ A จำนวน 60 ข้อ 2 ตัวเลือก	.75	.47
รูปแบบ B จำนวน 40 ข้อ 3 ตัวเลือก	.67	.73
รูปแบบ C จำนวน 40 ข้อ 3 ตัวเลือก	.65	.66
รูปแบบ D จำนวน 30 ข้อ 4 ตัวเลือก	.59	.66

(1) ความยากง่าย พบว่า แบบสอบรูปแบบ D (30 ข้อ 4 ตัวเลือก) มีค่าความยากง่ายสูงกว่าแบบสอบรูปแบบอื่น

(2) สัมประสิทธิ์ความเที่ยง พบว่า แบบสอบรูปแบบ B มีค่าสัมประสิทธิ์ความเที่ยงสูงที่สุด โดยสูงกว่าแบบสอบรูปแบบ D และรูปแบบ C เล็กน้อย ซึ่งไม่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบสอบรูปแบบ A มีค่าสัมประสิทธิ์ความเที่ยงต่ำที่สุด

Tollefson (1987) ได้ทำการศึกษาเรื่องการเปรียบเทียบค่าความยากและค่าอำนาจจำแนกของแบบสอบเลือกตอบ โดยใช้ตัวเลือก “ไม่มีข้อถูก” และมีคำตอบถูกหนึ่งตัวเลือก โดยมีวัตถุประสงค์เพื่อเปรียบเทียบค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยงของแบบสอบ จาก 3 รูปแบบของข้อสอบเลือกตอบแบบหลายตัวเลือก คือ (1) มีคำตอบถูกหนึ่งตัวเลือก (2) มีตัวเลือก “ไม่มีข้อถูก” เป็นตัวลวง และ (3) มีตัวเลือก “ไม่มีข้อถูก” เป็นตัวถูก เครื่องมือที่ใช้เป็นแบบสอบวิชาสถิติ จำนวน 12 ข้อซึ่งคัดเลือกมาจากข้อสอบทั้งหมด 73 ข้อ ซึ่งมีค่าถามเหมือนกันทั้ง 3 รูปแบบ โดยทดสอบกับกลุ่มตัวอย่างที่เป็นนักเรียนจำนวน 81 คน ผลการวิจัยมีดังนี้

ตารางที่ 2.8 ค่าความยากง่าย อำนาจจำแนก และค่าความเที่ยงจากการวิจัยของ Tollefson (1987)

รูปแบบตัวเลือก	ความยากง่าย	อำนาจจำแนก	ความเที่ยง
มีคำตอบถูกหนึ่งตัวเลือก	.83	.60	.74
ตัวเลือก “ไม่มีข้อถูก” เป็นตัวลวง	.87	.42	.51
ตัวเลือก “ไม่มีข้อถูก” เป็นตัวถูก	.63	.46	.56

(1) ค่าความยากง่ายพบว่า ข้อสอบที่ใช้ตัวเลือก “ไม่มีข้อถูก” เป็นตัวถูก มีค่าความยากสูงสุดจากทั้งสามรูปแบบ

(2) ค่าอำนาจจำแนกพบว่า ทั้งสามรูปแบบให้ค่าอำนาจจำแนกที่ไม่แตกต่างกัน

(3) ค่าความเที่ยงพบว่า ข้อสอบที่มีคำตอบถูกเป็นหนึ่งในตัวเลือกมีค่าความเที่ยงสูงกว่าแบบสอบที่ใช้ตัวเลือก “ไม่มีข้อถูก” ทั้งสองรูปแบบ

Crehan, Haladyna และ Brewer (1993) ได้ทำการศึกษาเรื่องลักษณะของตัวเลือกแบบ “ไม่มีข้อถูก” และจำนวนตัวเลือก “3 และ 4 ตัวเลือก” ที่มีต่อคุณภาพของแบบสอบ กลุ่มตัวอย่างที่ใช้

เป็นนักศึกษาที่เรียนวิชาจิตวิทยาเบื้องต้น ในมหาวิทยาลัยเซาท์-เวสต์เทิร์น จำนวน 220 คน แบ่งออกเป็น 4 กลุ่มๆ ละ 55 คน เพื่อตอบแบบสอบถามวิชาจิตวิทยาของ Myer (1986) จำนวน 48 ข้อ ที่แบ่งออกเป็น 4 ลักษณะที่แตกต่างกัน คือ แบบแรกเป็นแบบสอบ 4 ตัวเลือกแบบธรรมดา แบบที่สองเป็นแบบสอบ 4 ตัวเลือกแบบปลายเปิด “ไม่มีข้อถูก” และแบบที่สามเป็นแบบสอบ 3 ตัวเลือกแบบธรรมดา และแบบที่สี่เป็นแบบสอบ 3 ตัวเลือกแบบปลายเปิด “ไม่มีข้อถูก” และทำการหาค่าความยากง่าย ค่าอำนาจจำแนกและค่าความเที่ยง ผลการศึกษาพบว่า

ตารางที่ 2.9 ค่าความยากง่าย ค่าอำนาจจำแนกและค่าความเที่ยงจากผลการวิจัยของ Crehan, Haladyna และ Brewer (1993)

ค่าคุณภาพ	4 ตัวเลือกธรรมดา	4 ตัวเลือกปลายเปิด “ไม่มีข้อถูก”	3 ตัวเลือกธรรมดา	3 ตัวเลือกปลายเปิด “ไม่มีข้อถูก”
ความยากง่าย	.766	.716	.803	.741
อำนาจจำแนก	.357	.368	.352	.326
ความเที่ยง	.610	.600	.580	.520

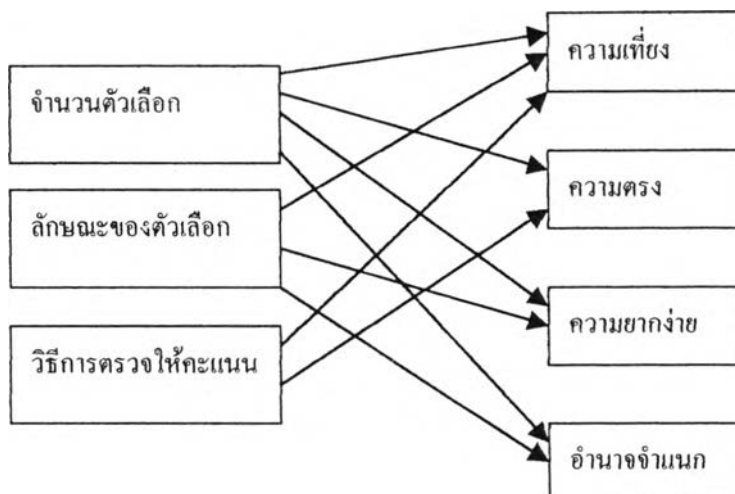
(1) ค่าความยากง่ายของแบบสอบที่มี 3 ตัวเลือกแตกต่างจากแบบสอบที่มี 4 ตัวเลือกอย่างมีนัยสำคัญทางสถิติที่ระดับ .024

(2) ค่าความยากง่ายของแบบสอบตัวเลือกธรรมดาแตกต่างจากแบบสอบตัวเลือกปลายเปิด “ไม่มีข้อถูก” อย่างมีนัยสำคัญทางสถิติที่ระดับ .001

(3) ค่าอำนาจจำแนกของแบบสอบที่มี 3 ตัวเลือกแตกต่างจากแบบสอบที่มี 4 ตัวเลือกอย่างไม่มีนัยสำคัญทางสถิติ

(4) ค่าอำนาจจำแนกของแบบสอบตัวเลือกธรรมดาแตกต่างจากแบบสอบตัวเลือกปลายเปิด “ไม่มีข้อถูก” อย่างไม่มีนัยสำคัญทางสถิติ

จากงานวิจัยต่างประเทศที่เกี่ยวกับแบบสอบหลายตัวเลือก เป็นการตรวจสอบคุณภาพของแบบสอบหลายตัวเลือก โดยมีตัวแปรต้นเป็นปัจจัยที่เกี่ยวกับแบบสอบหลายตัวเลือก ได้แก่ จำนวนตัวเลือก ลักษณะของตัวเลือก วิธีการตรวจให้คะแนน ตัวแปรตามเป็นคุณภาพของแบบสอบ ได้แก่ ความเที่ยง ความตรง ความยากง่าย และอำนาจจำแนก ซึ่งสามารถสรุปเป็นแผนภาพปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบหลายตัวเลือก ดังแผนภาพที่ 2.2



แผนภาพที่ 2.2 ปัจจัยที่มีอิทธิพลต่อคุณภาพของแบบสอบหลายตัวเลือก

ตอนที่ 4 การวิเคราะห์ห่อภิมาณ

การสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณ (meta-analysis) เริ่มเป็นที่รู้จักในหมู่นักวิจัยในปีคริสต์ศักราช 1976 โดย Glass (1976) เป็นผู้บัญญัติศัพท์คำว่า meta-analysis เป็นคนแรกในที่ประชุมสมาคมวิจัยการศึกษาอเมริกัน (The American Educational Research Association) โดยหมายถึง การวิเคราะห์ผลของการวิเคราะห์ทางสถิติเพื่อการสรุปรวม (Glass, 1976 อ้างถึงใน Hedges, 1992 ; Cooper และ Hedges, 1994 ; Cooper, 1998)

การสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณ เป็นการสังเคราะห์งานวิจัยเชิงปริมาณที่เกิดจากการนำงานวิจัยหลายๆ เรื่องที่ทำการศึกษาปัญหาเดียวกันมาทำการวิเคราะห์ผลการวิจัยเพื่อให้ได้ค่าดัชนีมาตรฐาน (standard index) ซึ่งมีชื่อเรียกต่างกันไป คือ ค่าดัชนีมาตรฐานที่ได้จากงานวิจัยเชิงทดลองจะเรียกว่า ค่าขนาดอิทธิพล (effect size : d) และค่าดัชนีมาตรฐานที่ได้จากงานวิจัยเชิงสหสัมพันธ์จะเรียกว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient : r) แล้วทำการศึกษาคงความแปรปรวนของค่าดัชนีมาตรฐานของงานวิจัยในแต่ละเรื่องที่น่ามาสังเคราะห์ว่ามีความแตกต่างกันหรือไม่ หากพบว่ามีความแตกต่างระหว่างค่าดัชนีมาตรฐานเกิดขึ้นก็จะทำการศึกษาวเคราะห์ต่อไปว่าความแตกต่างของค่าดัชนีมาตรฐานที่เกิดขึ้นเนื่องมาจากตัวแปรใด ซึ่งเรียกตัวแปรนี้ว่าตัวแปรปรับ (moderator variable)

การประมาณค่าดัชนีมาตรฐานในการวิเคราะห์ห่อภิมาณ

ในการสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณ นักสังเคราะห์งานวิจัยจะต้องทำการประมาณค่าดัชนีมาตรฐาน (standard index) ซึ่งจะได้มาจากผลการวิจัยของงานวิจัยในแต่ละเรื่องก่อน ซึ่งดัชนีมาตรฐานนี้สามารถแบ่งออกได้เป็น 2 ประเภทตามลักษณะของงานวิจัย คือ ค่าขนาดอิทธิพล (effect size : d) สำหรับการวิจัยเชิงทดลอง และค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient : r) สำหรับการวิจัยเชิงสหสัมพันธ์ โดยมีสูตรในการประมาณค่าดัชนีมาตรฐานทั้งค่าขนาดอิทธิพลและค่าสัมประสิทธิ์สหสัมพันธ์จากงานวิจัยได้ 2 วิธี คือ วิธีแรกเป็นการประมาณค่าดัชนีมาตรฐานจากการคำนวณโดยตรงจากค่าสถิติที่ได้จากกลุ่มตัวอย่าง และวิธีที่สองเป็นการประมาณค่าดัชนีมาตรฐานจากการคำนวณจากสถิติการทดสอบความมีนัยสำคัญ ซึ่ง นงลักษณ์ วิรัชชัย (2542) ได้สรุปวิธีการประมาณค่าดัชนีมาตรฐานไว้ดังนี้

1. วิธีการประมาณค่าดัชนีมาตรฐานจากการคำนวณโดยตรง

การประมาณค่าขนาดอิทธิพลและค่าสัมประสิทธิ์สหสัมพันธ์จากงานวิจัยที่มีแผนแบบการวิจัยเชิงทดลอง ซึ่งมีตัวแปรตาม Y เป็นตัวแปรต่อเนื่องและตัวแปรไม่ต่อเนื่อง และการประมาณค่าจากงานวิจัยที่มีแผนแบบการวิจัยเชิงสหสัมพันธ์ โดยการคำนวณโดยตรงจากค่าเฉลี่ยของกลุ่มตัวอย่างทำได้โดยใช้สูตรซึ่งมีสัญลักษณ์ดังต่อไปนี้

\bar{y}_E, \bar{y}_C = ค่าเฉลี่ยกลุ่มทดลอง และกลุ่มควบคุม

S = ส่วนเบี่ยงเบนมาตรฐาน

G = คะแนนการเปลี่ยนแปลง (change score)

g = คะแนนการเปลี่ยนแปลงเศษเหลือ (residual gain score)

ay = คะแนนปรับแก้ตัวแปรร่วม (score adjusted for covariate)

SS = ผลรวมกำลังสองของคะแนนเบี่ยงเบน (sum of square)

MS = ค่าเฉลี่ยผลรวมกำลังสองของคะแนนเบี่ยงเบน (mean square)

df = องศาอิสระ (degree of freedom)

A, B = ตัวแปรต้น หรือตัวแปรจัดกระทำ

d = ขนาดอิทธิพล (effect size)

r = สัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson's correlation coefficient)

r_{pb} = สัมประสิทธิ์สหสัมพันธ์พอยท์ไบซีเรียล (point biserial correlation coefficient)

p = สัดส่วน (proportion)

n = ขนาดของกลุ่มตัวอย่าง

1.1 แผนแบบการวิจัยแบบสองกลุ่มวัดครั้งหลัง (Two-group Posttest Only Design)

การประมาณค่าขนาดอิทธิพลกรณีที่มีตัวแปรตามเป็นตัวแปรต่อเนื่อง ใช้สูตรการประมาณค่าขนาดอิทธิพลจากค่าเฉลี่ย ดังนี้

$$d = \frac{[\bar{y}_E - \bar{y}_C]}{S_Y}$$

การประมาณค่าขนาดอิทธิพลกรณีที่มีตัวแปรตามเป็นตัวแปรไม่ต่อเนื่อง มีการหาสัดส่วน (proportion : p) ในกลุ่มตัวอย่างแต่ละกลุ่ม ใช้สูตรการประมาณค่าจากสัดส่วนดังนี้

$$d = \frac{p_E - p_C}{\sqrt{p_C(1 - p_E)}}$$

$$d = p_1 - p_2$$

เมื่อทราบค่าขนาดอิทธิพล จะประมาณค่าสัมประสิทธิ์สหสัมพันธ์จากขนาดอิทธิพลได้ ดังสูตรต่อไปนี้

$$r = \frac{d}{\sqrt{\frac{n}{nd^2 + 4n - 8}}} \quad ; n = n_E + n_C$$

1.2 แผนแบบการวิจัยแบบสองกลุ่มวัดก่อนและหลัง (Two-group, Pretest Posttest Design)

การประมาณค่าขนาดอิทธิพลกรณีที่มีตัวแปรตามเป็นตัวแปรต่อเนื่อง และใช้คะแนนการเปลี่ยนแปลง (G) คะแนนการเปลี่ยนแปลงเศษเหลือ (g) และคะแนนเมื่อปรับแก้ตัวแปรร่วม (ay) มีสูตรการประมาณค่าขนาดอิทธิพล ดังนี้

$$d = \frac{[\bar{G}_E - \bar{G}_C]}{S_Y} \quad ; \quad S_Y = \frac{S_G}{\sqrt{2(1 - r_{EC})}}$$

$$d = \frac{[\bar{g}_E - \bar{g}_C]}{S_Y} \quad ; \quad S_Y = \frac{S_g}{\sqrt{1 - r_{EC}^2}}$$

$$d = \frac{[\bar{ay}_E - \bar{ay}_C]}{S_Y} \quad ; \quad S_Y = \frac{S_{ay}}{\sqrt{1 - r_{EC}^2}}$$

1.3 แผนแบบการวิจัยแบบแฟคตอเรียล (Factorial Design)

การประมาณค่าขนาดอิทธิพลในแผนแบบการวิจัยแบบแฟคตอเรียลกรณีที่มีตัวแปรจัดกระทำสองตัวแปร คือ A และ B ใช้สูตรการประมาณค่าขนาดอิทธิพลสำหรับตัวแปรจัดกระทำ A ดังสูตรต่อไปนี้

$$d = \frac{[\bar{y}_{E1} - \bar{y}_{C1}]}{S_Y} \quad ; \quad S_Y = \sqrt{\frac{SS_B + SS_{AB} + SS_W}{df_B + df_{AB} + df_W}}$$

กรณีที่มีการประมาณค่าขนาดอิทธิพลสำหรับตัวแปรจัดกระทำในแต่ละระดับมีสูตรดังนี้

$$d = \frac{[\bar{y}_{E1} - \bar{y}_{C1}]}{S_Y} \quad ; \quad S_Y = \sqrt{MS_W}$$

1.4 แผนแบบการวิจัยเชิงสหสัมพันธ์ (Correlational Design)

การประมาณค่าขนาดอิทธิพลในแผนแบบการวิจัยเชิงสหสัมพันธ์ เป็นการประมาณค่าขนาดอิทธิพลจากค่าสัมประสิทธิ์สหสัมพันธ์โดยใช้สูตรดังนี้

$$d = \sqrt{\frac{n-2}{n}} \left[\frac{2r}{\sqrt{1-r^2}} \right]$$

2. วิธีการประมาณค่าดัชนีมาตรฐานจากค่าสถิติ

ค่าสถิติที่ใช้ในการประมาณค่าขนาดอิทธิพล และค่าสัมประสิทธิ์สหสัมพันธ์ หมายถึงค่าสถิติที่ได้จากการทดสอบสมมติฐานทางสถิติ ได้แก่ค่าสถิติ t , F , χ^2 มีสูตรในการประมาณค่าดังนี้

2.1 การประมาณค่าจากค่าสถิติ t ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ย

ค่าสถิติ t ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยมีสองกรณี คือ กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน และกรณีกลุ่มตัวอย่างสัมพันธ์กัน มีสูตรในการประมาณค่าขนาดอิทธิพลดังนี้

$$d = t \sqrt{1/n_E + 1/n_C}$$

$$d = t \sqrt{\frac{2}{n(1-r_{EC}^2)}}$$

2.2 การประมาณค่าจากสถิติ t ในการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์

เมื่อทราบค่าสถิติ t ในการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน จะประมาณค่าสัมประสิทธิ์สหสัมพันธ์และประมาณค่าขนาดอิทธิพลจากสัมประสิทธิ์สหสัมพันธ์ได้ดังสูตร

$$r = \frac{t}{\sqrt{t^2 + n - 2}}$$

$$d = \sqrt{\frac{n-2}{n}} \left[\frac{2r}{\sqrt{1-r^2}} \right]$$

2.3 การประมาณค่าจากสถิติไค-สแควร์

เมื่อทราบค่าสถิติไค-สแควร์ในการทดสอบความสัมพันธ์ระหว่างตัวแปรจะประมาณค่าสัมประสิทธิ์สหสัมพันธ์ได้ดังนี้

$$r = \frac{\chi^2}{\chi^2 + n}$$

2.4 การประมาณค่าจากค่าสถิติ F

เมื่อมีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยด้วยการวิเคราะห์ความแปรปรวน และทราบค่าสถิติ F จะประมาณค่าสัมประสิทธิ์สหสัมพันธ์และค่าขนาดอิทธิพลได้จากสูตรต่อไปนี้

$$r^2 = \frac{SS_B}{SS_T} = \frac{SS_B}{SS_W + SS_B} = \sqrt{\frac{F(df_B)}{[F(df_W) + (df_B)]}}$$

$$d = \frac{2 \sqrt{\frac{F(1 - r_{EC}^2)(df_W - 1)}{(n_E + n_C)(df_W - 2)}}}{\text{เมื่อการทดลองมีกลุ่มตัวอย่าง 2 กลุ่ม}}$$

ในกรณีที่ทำการวิเคราะห์ความแปรปรวนเมื่อมีกลุ่มตัวอย่าง k กลุ่ม นงลักษณ์ วิรัชชัย (2542) สรุปจากรายงานของ West และ Wiratchai (1984) ได้สูตรการประมาณค่าขนาดอิทธิพลดังนี้

$$d = 2f \sqrt{\frac{3(k-1)}{k+1}}$$

เมื่อ

$$f = \frac{\eta^2}{1 - \eta^2} \quad \text{และ} \quad \eta^2 = \frac{F(k-1)}{F(k-1) + (n-k)}$$

2.5 การประมาณค่าจากค่าสถิติ Mann-Whitney U

เมื่อมีการทดสอบความสัมพันธ์ระหว่างตัวแปรด้วยสถิติ Mann-Whitney U จะประมาณค่าสัมประสิทธิ์สหสัมพันธ์พอยท์ไบซีเรียล (r_{pb}) จากค่าสถิติ U ก่อน แล้วจึงประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน จากสูตรต่อไปนี้

$$r = r_{pb} \sqrt{\frac{n_1 n_2}{un}} \quad ; \quad u = \text{ordinate of unit normal distribution}$$

$$r_{pb} = 1 - \frac{2U}{n_1 n_2} \quad ; \quad U = \text{ค่าสถิติ Mann-Whitney}$$

ขั้นตอนในการวิเคราะห์ห่อภิมาณ

ขั้นตอนในการวิเคราะห์ห่อภิมาณที่นักวิจัยหลายท่านได้เสนอไว้ อาทิ อุทุมพร จามรมาน (2531) สุวัฒนา สุวรรณเขตนิกม (2529) ศิริยุภา พูลสุวรรณ (2530) นงลักษณ์ วิรัชชัย และสุวิมล ว่องวานิช (2541) และนงลักษณ์ วิรัชชัย (2542) มีขั้นตอนที่คล้ายคลึงกัน ผู้วิจัยจึงขอนำเสนอขั้นตอนการวิเคราะห์ห่อภิมาณตามที่ นงลักษณ์ วิรัชชัย (2542) เสนอไว้ 5 ขั้นตอน ดังนี้

1. การกำหนดปัญหา เป็นการตัดสินใจว่า การวิเคราะห์ห่อภิมาณมุ่งหมายจะตอบคำถามวิจัยในประเด็นใด การดำเนินงานในขั้นตอนนี้เป็นแบบเดียวกันกับขั้นตอนการกำหนดปัญหาวิจัยใน

ต้นฉบับ หน้าขาดหาย

ได้ การคำนวณค่าขนาดอิทธิพลวิธีนี้ต่างจากวิธีอื่นตรงที่ใช้ค่าส่วนเบี่ยงเบนมาตรฐานของกลุ่มควบคุมในการคำนวณ จุดเด่นในการวิเคราะห์ห่อภิมาณตามวิธีนี้คือ การมีสูตรในการประมาณค่าขนาดอิทธิพลจากงานวิจัยที่มีรูปแบบการทดลองแตกต่างกันทุกแผนแบบการวิจัย และมีสูตรในการปรับเปลี่ยนสัมประสิทธิ์สหสัมพันธ์แบบอื่นๆ มาเป็นสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สูตรการประมาณค่าดัชนีมาตรฐานทั้งสองแบบดังกล่าวนี้มีทั้งสูตรจากการประมาณค่าสถิติโดยตรง และสูตรการประมาณค่าจากผลการทดสอบสมมติฐานทางสถิติ

3. การวิเคราะห์ห่อภิมาณตามวิธีของ Hedges ให้ความสำคัญกับคุณสมบัติทางสถิติของดัชนีมาตรฐาน โดยเฉพาะค่าขนาดอิทธิพลเป็นอย่างมาก การคำนวณค่าขนาดอิทธิพลใช้ค่าส่วนเบี่ยงเบนมาตรฐานรวมตามแนวคิดของ Cohen (1969) และมีการปรับแก้ให้ค่าประมาณไม่คลาดเคลื่อน นอกจากนี้ยังมีสูตรสำหรับคำนวณค่าขนาดอิทธิพลจากค่าที่ได้ตามวิธีของ Glass และวิธีของ Hunter อีกด้วย Hedges และ Olkin ได้เสนอแนะให้มีการปรับแก้ความคลาดเคลื่อนเนื่องจากการวัดและความจำกัดของพิสัยตามแบบของ Hunter แต่ไม่มีการปรับแก้ความคลาดเคลื่อนจากการสุ่มตัวอย่าง สำหรับการสังเคราะห์ Hedges และ Olkin เสนอสูตรการคำนวณค่าเฉลี่ยโดยการถ่วงน้ำหนัก และเสนอว่าการแจกแจงของค่าขนาดอิทธิพลเป็นการแจกแจงแบบไค-สแควร์ รวมทั้งให้ค่าสถิติ Q ในการทดสอบว่าพารามิเตอร์ขนาดอิทธิพลมีขนาดเท่ากันสำหรับทุกกลุ่มประชากรก่อนที่จะทำการสังเคราะห์

4. การวิเคราะห์ห่อภิมาณตามวิธีของ Slavin ให้ความสำคัญกับการคัดเลือกงานวิจัยที่นำมาสังเคราะห์ เนื่องจาก Slavin เชื่อว่าถ้างานวิจัยเดิม ไม่มีคุณภาพ ผลการสังเคราะห์ย่อมที่จะไม่มีคุณภาพด้วย ดังนั้นจึงต้องมีการประเมินคุณภาพงานวิจัยและคัดเลือกเฉพาะงานวิจัยที่มีคุณภาพตามเกณฑ์มาสังเคราะห์ วิธีการของ Slavin มีข้อจำกัดในการปฏิบัติจริงเพราะงานวิจัยที่นำมาสังเคราะห์มีจำนวนไม่มาก เมื่อคัดเฉพาะงานที่มีคุณภาพดีทำให้ขนาดกลุ่มตัวอย่างสำหรับการวิเคราะห์ห่อภิมาณมีจำนวนน้อยลงไปอีก การสรุปอ้างอิงผลการวิจัยจึงมีข้อจำกัด

5. การวิเคราะห์ห่อภิมาณตามวิธีของ Mullen เป็นวิธีการที่ได้รับการพัฒนาน่าสุด วิธีการนี้ใช้การประมาณค่าขนาดอิทธิพลตามแนวคิดของ Cohen (1969) และใช้การประมาณค่าความสัมพันธ์ระหว่างตัวแปรโดยใช้คะแนนพิชเซอร์-สกีในการวิเคราะห์การประมาณค่าเฉลี่ยของดัชนีมาตรฐานใช้การหาค่าเฉลี่ยแบบถ่วงน้ำหนัก จุดเด่นของวิธีนี้คือ การพัฒนาโปรแกรมคอมพิวเตอร์เพื่อใช้สำหรับการวิเคราะห์ห่อภิมาณโดยตรง โปรแกรมสามารถทดสอบความเป็นเอกพันธ์ของดัชนีมาตรฐาน และตรวจสอบว่าตัวแปรปรับตัวใดสามารถอธิบายความแปรปรวนในดัชนีมาตรฐานได้

การวิเคราะห์ห่อภิมาณทั้ง 5 วิธีนี้ เหมาะสำหรับการสังเคราะห์งานวิจัยต่างๆ ไป แต่ในการสังเคราะห์งานวิจัยเกี่ยวกับคุณภาพของแบบสอบ Hunter และ Schmidt (1990) และ Knowles และ Welch (1992) เสนอแนะว่า ควรใช้วิธีของ Hunter และ Schmidt (1990) ดังมีรายละเอียดของวิธีการดังต่อไปนี้

6. การวิเคราะห์ห่อภิมาณตามวิธีของ Hunter เป็นการวิเคราะห์ห่อภิมาณงานวิจัยเชิงสหสัมพันธ์ โดยมีหลักการวิเคราะห์ข้อมูลเป็นการปรับแก้ลดความคลาดเคลื่อนของดัชนีมาตรฐานก่อนที่จะนำมาสังเคราะห์สรุปผล

Hunter และ Schmidt (1990) ได้กล่าวถึงค่าดัชนีมาตรฐานที่มีความคลาดเคลื่อนจาก 7 แหล่ง คือ (1) ความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง (sampling error) (2) ความคลาดเคลื่อนในการวัด (measurement error) (3) ความคลาดเคลื่อนเนื่องจากการจำกัดของพิสัย (error due to range restriction) (4) ความคลาดเคลื่อนจากการลดระดับการวัดของตัวแปร (dichotomization of continuous variable) (5) ความเบี่ยงเบนจากความตรงตามโครงสร้างแบบสมบูรณ์ของตัวแปร (deviation from perfect construct validity) (6) ความคลาดเคลื่อนจากการรายงาน (reporting or transcriptional error) และ (7) ความแปรปรวนเนื่องจากปัจจัยของตัวแปรแทรกซ้อน (variance due to extraneous factors) ซึ่ง Hunter และ Schmidt เสนอแนะว่าความคลาดเคลื่อนเหล่านี้สามารถที่จะปรับแก้ด้วยวิธีการทางสถิติได้ 3 วิธี คือ ความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง (sampling error) ความคลาดเคลื่อนในการวัด (measurement error) และความคลาดเคลื่อนเนื่องจากการจำกัดของพิสัย (error due to range restriction) ซึ่งการปรับแก้ความคลาดเคลื่อนของดัชนีมาตรฐานก่อนที่จะนำมาสังเคราะห์ผลนี้ถือว่าเป็นหัวใจของการวิเคราะห์ข้อมูลตามวิธีของ Hunter (นงลักษณ์ วิรัชชัย, 2542)

Hunter และ Schmidt ได้เสนอวิธีในการปรับแก้ความคลาดเคลื่อนในการวัด (measurement error) ความคลาดเคลื่อนเนื่องจากการจำกัดของพิสัย (error due to range restriction) และความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง (sampling error) ด้วยวิธีการทางสถิติ ดังนี้

6.1 การปรับแก้ความคลาดเคลื่อนจากการวัด

ทำการปรับแก้ความคลาดเคลื่อนจากการวัดโดยใช้สูตร

$$\rho_{TU} = \frac{\rho_{XY}}{\sqrt{r_x} \sqrt{r_y}}$$

6.2 การปรับแก้ความคลาดเคลื่อนจากความจำกัดของพิสัย

ทำการปรับแก้ความคลาดเคลื่อนจากความจำกัดของพิสัย โดยใช้สูตร

$$\rho_2 = \frac{U\rho_1}{\sqrt{(U^2 - 1)\rho_1^2 + 1}}$$

เมื่อ

$$U = \frac{\sigma_{r_2}}{\sigma_{r_1}}$$

โดยที่ ρ_1 คือ ค่าสหสัมพันธ์ของประชากรที่มีอยู่จริง ρ_2 คือ ค่าสหสัมพันธ์ของประชากรที่ศึกษา

U คือ อัตราส่วนของส่วนเบี่ยงเบนมาตรฐานของ 2 กลุ่ม

6.3 การปรับแก้ความคลาดเคลื่อนจากการสุ่มตัวอย่าง

การปรับแก้ความคลาดเคลื่อนจากการสุ่มตัวอย่าง มีขั้นตอนดังนี้

6.3.1 กำหนดค่าเฉลี่ยถ่วงน้ำหนัก (\bar{r}) จากสูตร

$$\bar{r} = \frac{\sum [N_i r_i]}{\sum N_i}$$

เมื่อ r_i คือ ค่าสหสัมพันธ์ของสิ่งที่ศึกษาครั้งที่ i N_i คือ จำนวนคนที่ใช้ในการศึกษาครั้งที่ i

6.3.2 กำหนดค่าความแปรปรวนของค่าเฉลี่ยถ่วงน้ำหนัก จากสูตร

$$\sigma_r^2 = \frac{\sum [N_i (r_i - \bar{r})^2]}{\sum N_i}$$

6.3.3 กำหนดค่าเฉลี่ยขนาดกลุ่มตัวอย่าง (\bar{N}) จากสูตร

$$\bar{N} = T / K$$

เมื่อ T คือ ผลรวมของขนาดกลุ่มตัวอย่าง

K คือ ผลรวมของจำนวนกลุ่มตัวอย่าง

6.3.4 กำหนดค่าความแปรปรวนของความคลาดเคลื่อนจากการสุ่มตัวอย่าง จาก

สูตร

$$\sigma_p^2 = \frac{(1 - \bar{r}^2)^2}{(\bar{N} - 1)}$$

6.3.5 การประมาณค่าความแปรปรวนของค่าสหสัมพันธ์ของประชากร จากสูตร

$$\sigma_p^2 = \sigma_r^2 - \sigma_p^2$$

ตามหลักการวิเคราะห์ห่อภิมาณคามวิธีของHunter นี้ เริ่มต้นจากการปรับแก้ความคลาดเคลื่อนทั้ง 3 แบบ และเมื่อได้ σ_p^2 แล้วจะพิจารณาตรวจสอบว่ามีนัยสำคัญหรือไม่ โดยใช้ค่า χ^2 ตรวจสอบ ตามสูตร

$$\chi^2_{k-1} = \frac{\sum (N_i - 1)(r_i - \bar{r})^2}{(1 - \bar{r}^2)^2}$$

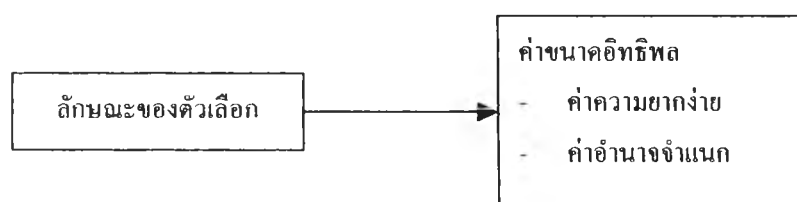
ถ้าผลการทดสอบพบว่ามีนัยสำคัญทางสถิติ Hunter เสนอให้แบ่งงานวิจัยออกเป็นกลุ่มตามลักษณะตัวแปรตาม ลักษณะงานวิจัยนี้เรียกว่าตัวแปรปรับ (moderator variable) แล้วทำการวิเคราะห์ทั้ง 3 ชั้นคอนในแต่ละกลุ่ม เพื่อให้ได้ข้อสรุปสุดท้าย

จากการสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณในครั้งนี้ ผู้วิจัยเลือกใช้วิธีการวิเคราะห์ห่อภิมาณของ Hunter ตามที่ได้เสนอไว้ข้างต้น

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

การสังเคราะห์งานวิจัยด้วยวิธีวิเคราะห์ห่อภิมาณ

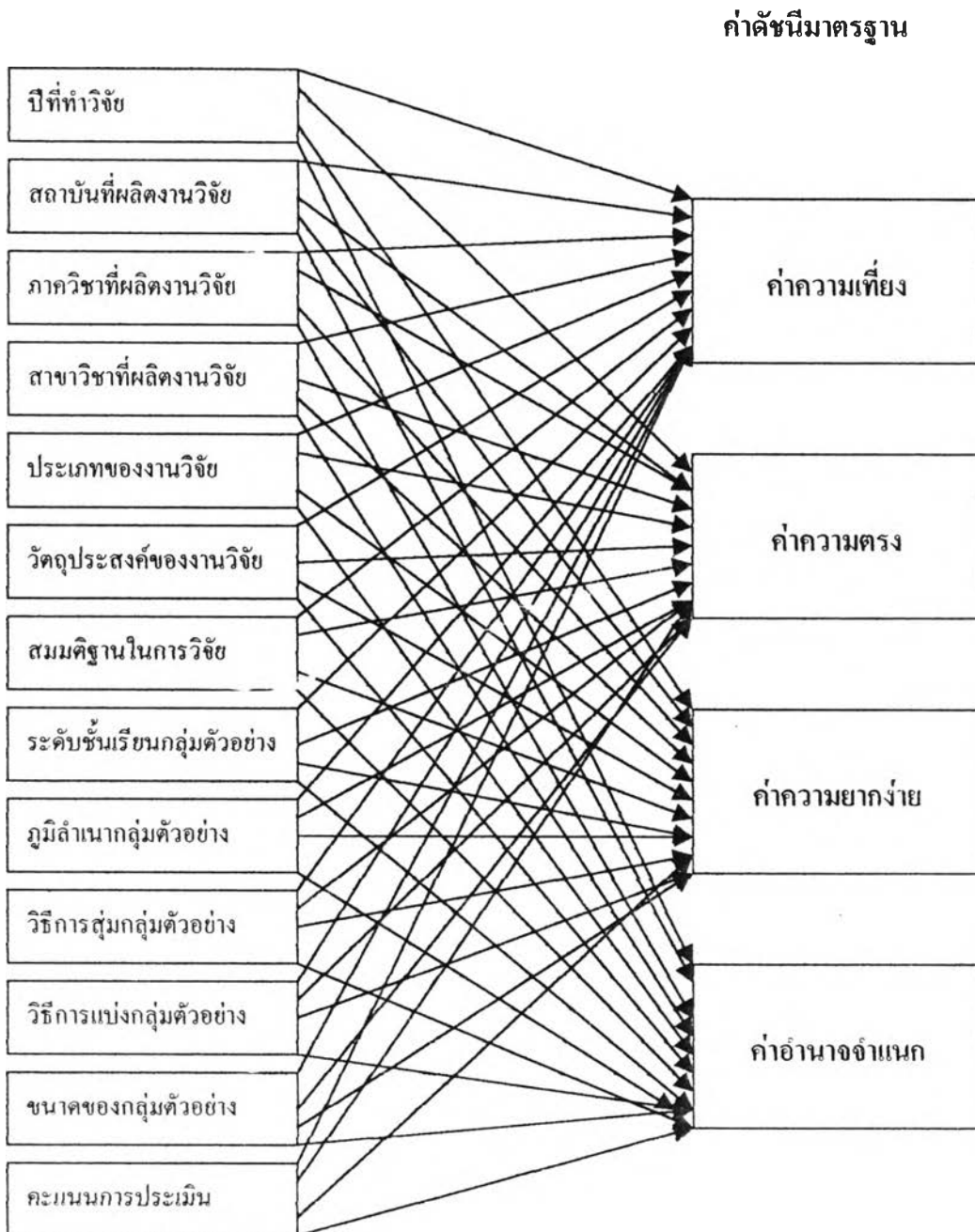
Knowles and Welch (1992) ได้ทำการสังเคราะห์งานวิจัยด้วยการวิเคราะห์ห่อภิมาณเกี่ยวกับค่าความยากง่ายและค่าอำนาจจำแนกของแบบสอบเลือกตอบประเภทหลายตัวเลือกที่ใช้ตัวเลือก “ไม่มีข้อถูก” โดยนำงานวิจัยที่เกี่ยวกับค่าความยากมาทำการสังเคราะห์ จำนวน 12 เล่ม และงานวิจัยที่เกี่ยวกับค่าอำนาจจำแนก จำนวน 7 เล่ม ได้ค่าขนาดอิทธิพลจำนวน 12 และ 11 ค่า ตามลำดับ วิธีการวิเคราะห์ข้อมูล ทำการคำนวณค่าขนาดอิทธิพลตามวิธีการสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณตามแนวคิดของ Hunter and Schmidt (1990) และได้ทำการคำนวณค่าเฉลี่ยของขนาดอิทธิพล (d) ความแปรปรวนของค่าขนาดอิทธิพลที่สังเกตได้ ความแปรปรวนเนื่องมาจากความคลาดเคลื่อนในการสุ่มตัวอย่าง และคำนวณช่วงระดับความเชื่อมั่น 90% ผลการสังเคราะห์พบว่า การใช้ตัวเลือกปลายเปิด “ไม่มีข้อถูก” ไม่ได้ทำให้ค่าความยากง่ายเพิ่มขึ้นและค่าอำนาจจำแนกลดลงกว่าการใช้แบบสอบที่มีตัวเลือกถูกเพียงตัวเดียว ซึ่งจากผลการวิจัยในประเทศและการสังเคราะห์งานวิจัยในต่างประเทศข้างต้นก็ยังไม่สามารถสรุปได้ชัดเจนว่าควรจะมีการนำตัวเลือกประเภทนี้มาใช้หรือไม่ โดยสามารถเขียนแผนภาพสรุปปัจจัยที่ส่งผลต่อค่าขนาดอิทธิพลค่าคุณภาพของแบบสอบหลายตัวเลือกที่นำมาวิเคราะห์ห่อภิมาณได้ตามแผนภาพที่ 2.3



แผนภาพที่ 2.3 ปัจจัยที่ส่งผลต่อค่าขนาดอิทธิพลในการวิเคราะห์ห่อภิมาณคุณภาพของแบบสอบ

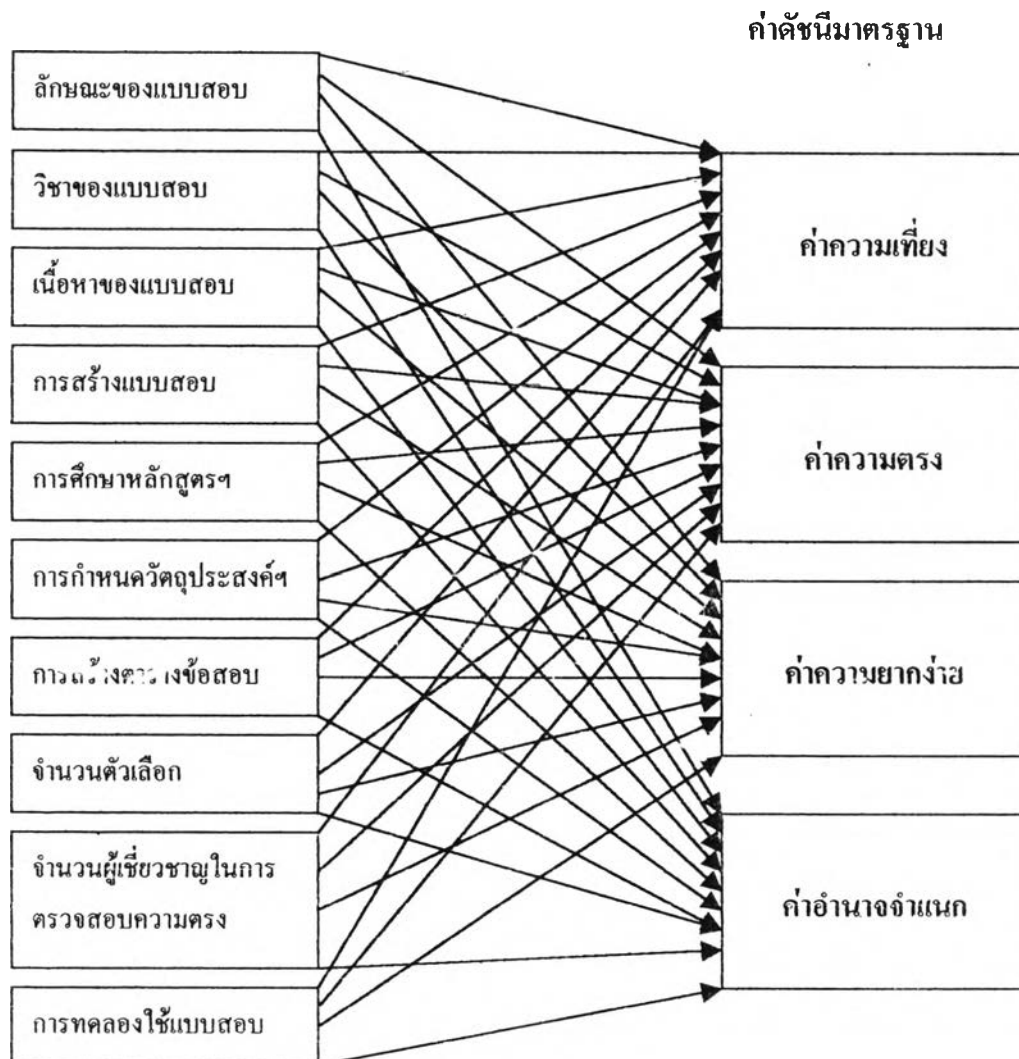
ต้นฉบับ หน้าขาดหาย

1. กรอบแนวคิดในการวิจัยของข้อมูลระดับเล่มวิทยานิพนธ์



แผนภาพที่ 2.5 กรอบแนวคิดในการวิจัยของข้อมูลระดับเล่มวิทยานิพนธ์

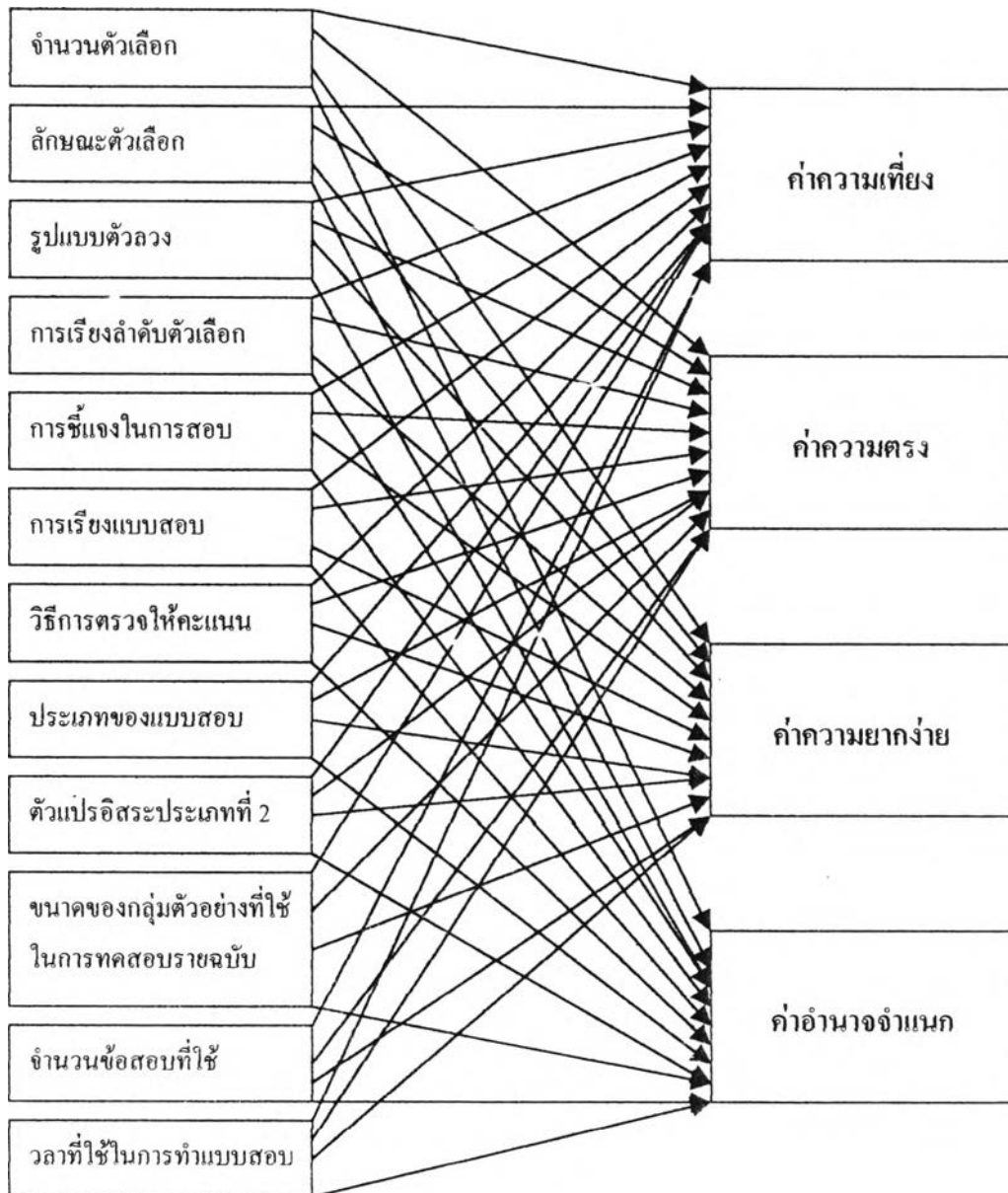
2. กรอบแนวคิดในการวิจัยของข้อมูลระดับแบบสอบที่สร้างขึ้นก่อนนำไปใช้ในการวิจัย



แผนภาพที่ 2.6 กรอบแนวคิดในการวิจัยของข้อมูลระดับแบบสอบที่สร้างขึ้นก่อนนำไปใช้ในการวิจัย

3. กรอบแนวคิดในการวิจัยของข้อมูลระดับค่าคุณภาพของแบบสอบ

ค่าดัชนีมาตรฐาน



แผนภาพที่ 2.7 กรอบแนวคิดในการวิจัยของข้อมูลระดับค่าคุณภาพของแบบสอบ