

บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาแนวคิด หลักการและงานวิจัยที่เกี่ยวข้องตามลำดับดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้องกับรูปแบบฟาเซท

2.1.1 ความหมาย พัฒนาการ และรูปแบบของเทคโนโลยีการเขียนข้อสอบ

2.1.2 ความเป็นมาและความหมายของรูปแบบฟาเซท

2.1.3 หลักการและข้อเสนอแนะในการใช้รูปแบบฟาเซทสร้างข้อสอบและตัวเลือก

2.1.4 ประโยชน์และข้อจำกัดของรูปแบบฟาเซท

2.2 แบบทดสอบคู่ขนาน

2.3 หลักการสร้างและการตรวจสอบคุณภาพของแบบทดสอบอิงเกณฑ์

2.3.1 หลักการสร้างแบบทดสอบอิงเกณฑ์

2.3.2 การตรวจสอบคุณภาพแบบทดสอบอิงเกณฑ์

2.3.3 การกำหนดเกณฑ์มาตรฐาน

2.3.4 ความตรงและความเที่ยงของแบบทดสอบอิงเกณฑ์

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 งานวิจัยเกี่ยวกับการสร้างแบบทดสอบโดยใช้เทคโนโลยีการเขียนข้อสอบ

2.4.2 งานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบคู่ขนาน

2.4.3 งานวิจัยเกี่ยวกับการสร้างแบบทดสอบ และการวิเคราะห์หาคุณภาพของแบบทดสอบที่เขียนจากเทคโนโลยีการเขียนข้อสอบตามรูปแบบฟาเซท

2.1 ทฤษฎีที่เกี่ยวข้องกับรูปแบบฟาเซท

2.1.1 ความหมาย พัฒนาการและรูปแบบของเทคโนโลยีการเขียนข้อสอบ

รอย และฮาလာไดนา (Roid and Haladyna, 1982) ได้กล่าวถึงความหมายของเทคโนโลยีการเขียนข้อสอบไว้ดังนี้

1) เป็นวิธีการเขียนข้อสอบที่สามารถบอกถึงความสัมพันธ์ของการเรียนการสอนและการสอบได้อย่างมีเหตุผลและมีความเที่ยงตรง

2) เป็นวิธีการเขียนข้อสอบที่ผู้เขียนต้องมีประสบการณ์ด้านเนื้อหา แม้ว่าจะไม่มีประสบการณ์ หรือทักษะในการเขียนข้อสอบมากก็สามารถเขียนข้อสอบได้

3) เป็นวิธีการเขียนข้อสอบที่มีการกำหนดลักษณะเฉพาะ ซึ่งทำให้ง่ายสำหรับผู้เขียนข้อสอบอื่น ๆ ที่จะสามารถเขียนข้อสอบตามลักษณะเฉพาะที่กำหนดไว้แล้วได้

บุญเชิด ภิญโญนันตพงษ์ (2527) กล่าวว่า เทคโนโลยีการเขียนข้อสอบ คือ เครื่องช่วยผลิตข้อสอบที่คู่ขนานกันจากประชากรข้อสอบได้จำนวนมาก โดยการกำหนดลักษณะเฉพาะของมวลความรู้ ซึ่งครอบคลุมทั้งพฤติกรรมและลักษณะเนื้อหาวิชาอย่างชัดเจน พร้อมทั้งสร้างกฎเกณฑ์ในการเขียนข้อสอบ ตามคำนิยามดังกล่าว จะเห็นว่า หากกำหนดเนื้อหาที่จะสอบไว้อย่างละเอียด ชัดเจนและเฉพาะเจาะจงตามจุดประสงค์การเรียนการสอนแล้ว ก็จะช่วยให้เขียนข้อสอบได้โดยตรงตามต้องการและผลิตข้อสอบได้จำนวนมาก

ราวปี ค.ศ. 1968 ทีมมหาวิทยาลัยมินเนโซตา ไฮฟลี และคณะ (Hively and Other , 1968 อ้างถึงในชูศักดิ์ ชัมภลิจิต , 2529) ได้พัฒนาแบบทดสอบคู่ขนานขึ้นหลายฉบับเพื่อใช้วิจัยควบคู่ไปกับการพัฒนาหลักสูตรวิชาคณิตศาสตร์และวิทยาศาสตร์ระดับประถมศึกษาในการนี้ ไฮฟลี และคณะได้สร้างรูปแบบที่จะช่วยในการเขียนข้อสอบชนิดหนึ่งคือ ฟอรัมข้อสอบ (Item Form) ซึ่งถือว่าเป็นเครื่องมือชิ้นแรกของการพัฒนาการของเทคโนโลยีการเขียนข้อสอบ ในช่วงเวลาที่ใกล้เคียงกัน ทีมมหาวิทยาลัยแคลิฟอร์เนีย แห้งลอสแอนเจลิส (UCLA) ภายใต้การนำของปอปแฮม (Popham , 1978) ได้ตั้งสถาบันผู้เชี่ยวชาญทางด้านการสร้างแบบทดสอบอิงเกณฑ์ ชื่อ Institute of Objective Exchange (IOX) สถาบัน IOX นี้ จัดตั้งขึ้นเพื่อเป็นศูนย์รวบรวมจุดประสงค์เชิงพฤติกรรมของกระบวนการเรียนการสอนหลาย ๆ หมวดวิชาให้ครอบคลุมมากที่สุด และได้จัดพิมพ์จุดประสงค์ของแต่ละวิชา และตัวอย่างข้อสอบไว้บริการครู หรือผู้สนใจทั่วไป แต่สถาบัน IOX ไม่สามารถบรรลุเป้าหมายอันนี้ได้ เพราะจุดประสงค์เชิงพฤติกรรมของแต่ละข้อเป็นข้อความที่กำหนดไว้อย่างหลวม ๆ และไม่ได้กำหนดแนวทางเฉพาะให้กับผู้เขียนข้อสอบอย่างชัดเจน สถาบัน IOX จึงดัดแปลงฟอรัมข้อสอบของไฮฟลีและคณะ โดยสร้างรูปแบบที่ช่วยในการเขียนข้อสอบขึ้นใหม่สองเทคนิควิธีคือ จุดประสงค์ขยายความ (Amplified Objectives) และลักษณะเฉพาะของข้อสอบ (Test Specifications)

บอร์มูธ (Bormuth , 1970) เป็นบุคคลแรกที่เสนอเทคนิคเก่าแก่ของการเขียนข้อสอบอีกวิธีหนึ่ง ซึ่งเรียกว่า แบบข้อสอบความเรียง (Item for Prose Learning) ซึ่งมีหลักและวิธีการแปลงประโยคสำคัญของเนื้อหาวิชาให้เป็นองค์ประกอบของมวลความรู้ (Domain) แล้วเขียนข้อสอบจากมวลความรู้ดังกล่าว ชูศักดิ์ ชัมภลิจิต (2529) กล่าวว่าในระยะเวลา 6 ปีที่แล้วมีเทคโนโลยีการเขียนข้อสอบใหม่ ๆ เกิดขึ้นอีกหลายแนวคิด เช่น การประยุกต์ทฤษฎีฟาสเซทของ

กัตแมน (Guttman) มาใช้เป็นเทคนิคการเขียนข้อสอบตามแนวอิงเกณฑ์ ซึ่งเรียกว่า "The Mapping Sentence Method"

รอยและฮาไลดัยนา (Roid and Haladyna , 1982) ได้กล่าวถึงรูปแบบของเทคโนโลยีการเขียนข้อสอบไว้ 6 วิธี คือ

- 1) การใช้ข้อความเรียง (Item for Prose Learning)
- 2) การใช้ฟอร์มข้อสอบ (Item Form)
- 3) การใช้รูปแบบฟาเซทหรือการสร้างประโยชน์จากการจับคู่ (Facet Design or The Mapping Sentence Method)
- 4) การใช้เทคนิคการวัดมโนทัศน์ (Test Item for Concept)
- 5) การใช้เทคนิคเชิงตรรก (Logical Operations for Generation Intended Questions)
- 6) การใช้เทคนิค IQI (The Instructional Quality Inventory)

2.1.2 ความเป็นมาและความหมายของรูปแบบฟาเซท

รูปแบบฟาเซทมีรากฐานการพัฒนาจากทฤษฎีบุคลิกภาพโดยการนิยามรูปแบบทางจิตวิทยาของ Jung's Psychological Types) (Foa, 1965) ในปี 1965 และ 1969 กัตแมน (Guttman , 1969 quoted in Roid and Haladyna, 1982) เป็นผู้เสนอใช้ประโยชน์ของรูปแบบฟาเซทกับการวัดผลสัมฤทธิ์ และได้เสนอวิธีการจัดอันดับขั้นของวิชาต่าง ๆ กับพฤติกรรมทางสมอง (Cognitive Behavior) ดังตารางที่ 1

ตารางที่ 1 แสดงการใช้เมตริกซ์ลำดับขั้นของกัตแมน : ในพฤติกรรมทางสมอง

พฤติกรรมด้านความรู้ ลำดับขั้นของเนื้อหา	B.1 รูปภาพ (Figural)	B.2 ตัวเลข (Numerical)	B.3 ภาษา (Verbal)
A.1 การอ้างอิงกฎ (Rule Inferring)	ปัญหาเกี่ยวกับเรขาคณิต (Geometric Problems)	อันดับตัวเลข (Numerical Progression)	การให้คำที่เกี่ยวข้องกัน 2 คำ เช่น car: Ford
A.2 การประยุกต์กฎ (Rule Applying)	จับคู่ลูกบาศก์ (Matching cubes)	การใช้สูตร การ คำนวณ การแก้ ปัญหา เลขคณิต	การกำหนดประเภท สิ่งของ เช่น โรบิน:นก
A.3 ผลสัมฤทธิ์ (Achievement)	การเรียนรู้เกี่ยวกับภาษา สัญลักษณ์ (Learning a Symbolic Language)	คณิตศาสตร์ เคมี , สถิติ ,ฟิสิกส์	ประวัติศาสตร์ วรรณคดี

จากตารางที่ 1 เป็นการประยุกต์วิธีการกัตแมน มาใช้วัดพฤติกรรมทางสมอง โดยพฤติกรรมทางสมองทั้งหมดสามารถจัดลงในมิติใดมิติหนึ่ง ซึ่งประกอบด้วยรูปภาพ (B.1) ตัวเลข (B.2) หรือภาษา (B.3) และจัดลงในมิติอีกมิติหนึ่งคือ มิติที่ประกอบด้วย การอ้างอิงกฎ (A.1) การประยุกต์กฎ (A.2) และผลสัมฤทธิ์ (A.3) เช่น วิชาคณิตศาสตร์และฟิสิกส์ เป็นวิชาที่เกี่ยวกับผลสัมฤทธิ์และตัวเลข (A.3 และ B.2) ประวัติศาสตร์เป็นวิชาที่เกี่ยวกับผลสัมฤทธิ์ด้านภาษา (A.3 และ B.3)

แบบทดสอบวัดความสามารถทางสมองหลายชนิด ที่สามารถใช้กฎการอ้างอิงหรือกฎการประยุกต์ โดยการแยกแบบทดสอบเป็นแบบทดสอบเกี่ยวกับรูปภาพ ตัวเลข และภาษาลักษณะดังกล่าวมีความคล้ายคลึงกับโมเดลโครงสร้างทางสมองของกิลฟอร์ด (Roid and Haladyna ; 1982 : 128 ; eiting, 1967) ดังนั้น เนื้อหาของรูปแบบฟาเซท ถือว่าเป็นการนำทฤษฎีการเรียนรู้ที่ต่อเนื่อง (Hierarchical Theory of Learning) มาใช้กับการทดสอบผลสัมฤทธิ์ทางการเรียน

กัตแมน (Guttam, 1969 quoted in Roid and Haladyna, 1982) ได้อธิบายเกี่ยวกับรูปแบบฟาเซท ว่าเป็นเทคโนโลยีการเขียนข้อสอบชนิดหนึ่งที่ได้พัฒนามาจากฟอร์มข้อสอบ (Item Form) ซึ่งลักษณะของฟอร์มข้อสอบ คือ การกำหนดกฎเกณฑ์ในการเขียนข้อสอบให้ได้หลายๆ ข้อ โดยที่กลุ่มข้อสอบที่สามารถวัดมวลความรู้เดียวกันได้อย่างครอบคลุมฟอร์มข้อสอบมีส่วนประกอบที่สำคัญ 3 ส่วน คือ

- 1) ส่วนคงที่ เป็นส่วนประกอบของโครงสร้างข้อความที่กำหนดไว้ตายตัวเป็นเสมือนตัวฟอร์มข้อสอบหรือเปลือกหุ้มฟอร์มข้อสอบ (Item Form Shell) ที่ใช้สำหรับผลิตข้อสอบ
- 2) ส่วนที่แปรเปลี่ยน เป็นส่วนประกอบของโครงสร้างข้อความที่เว้นว่างไว้ อย่างน้อยที่สุดหนึ่งแห่ง เพื่อนำเนื้อหาของคำหรือ ข้อความที่เป็นคำถามใส่แทนลงไป
- 3) ส่วนที่นำไปเติมลงในส่วนที่แปรเปลี่ยน เป็นเซตของประโยคข้อสอบที่ได้รับการนิยามไว้อย่างชัดเจน สำหรับใช้เติมลงในส่วนที่แปรเปลี่ยน

รอยด์ และฮาလာไดนา (Roid and Haladyna, 1982) กล่าวว่า สำคัญของรูปแบบฟาเซท เป็นวิธีการในการรวบรวมและนิยามเนื้อหา โดยการกำหนดลักษณะเฉพาะของขอบเขตและลำดับขั้นขององค์ประกอบ และส่วนย่อยอื่น ๆ รูปแบบฟาเซทประกอบด้วยฟาเซทหลายๆ ฟาเซทภายในสมาชิกของฟาเซทมีส่วนที่คงที่และส่วนที่แปรเปลี่ยน ซึ่งก็คือลักษณะของฟอร์มข้อสอบ (Item Form) และภายในสมาชิกจะมีการกำหนดลักษณะเฉพาะของส่วนที่เป็นคำถามและตัวเลือก ซึ่งก็คือการกำหนดลักษณะเฉพาะของข้อสอบ (Item Specification) จากนั้นในการสร้างข้อ

สอบจะมีการเลือกจับคู่จากสมาชิกของฟาเซตตามลักษณะเฉพาะที่กำหนด มาสร้างเป็นคำถามและตัวเลือกเรียกว่า ประโยคที่เกิดจากการจับคู่ ซึ่งประกอบด้วยส่วนสำคัญ 3 ส่วน คือ

1) ส่วนที่คงที่ (Fixed Part) จะมีลักษณะคล้ายกับ เปลือกหุ้มฟอร์มข้อสอบ (Item Form Shell) หมายถึง ส่วนที่เป็นฟอร์มข้อสอบ

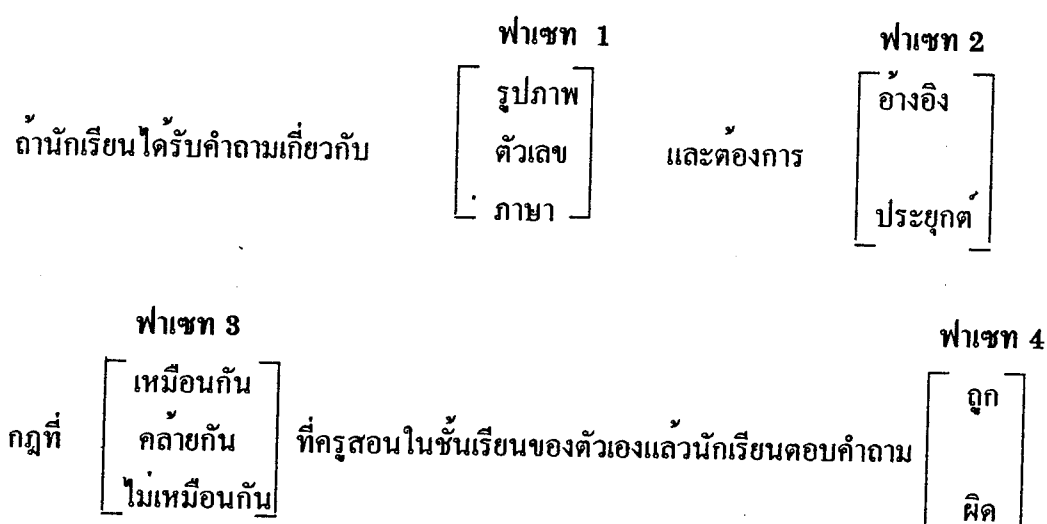
2) ส่วนที่เป็นฟาเซต (Facets) เป็นแง่มุมต่าง ๆ ของตัวแปร

3) ส่วนที่เป็นสมาชิกของฟาเซต (Facet Elements) หรือประเด็นปัญหา คือ ส่วนที่นำไปแทนที่ (Replacement) ของฟอร์มข้อสอบ

การสร้างประโยคจากการจับคู่จะใช้ส่วนที่เป็นฟาเซตและสมาชิกของฟาเซต ประกอบกันเป็นรูปแบบฟาเซต ซึ่งรูปแบบฟาเซตนี้จะเป็นการกำหนดลักษณะของมวลความรู้ที่จะนำไปเขียนข้อสอบ

กัตแมน (Guttman, 1969) ได้ยกตัวอย่างประกอบรูปแบบฟาเซต ซึ่งสามารถใช้อธิบายการวิเคราะห์พฤติกรรมทางสมอง ซึ่งแสดงการจับคู่ประโยคได้ดังแผนภาพที่ 1

แผนภาพที่ 1 ตัวอย่างการสร้างประโยคที่เกิดจากการจับคู่ระหว่างสมาชิกของฟาเซต



จากการแสดงการจับคู่สมาชิกของแต่ละฟาเซตข้างบน ซึ่งมีฟาเซต 4 ฟาเซต และมีจำนวนสมาชิกใน แต่ละฟาเซตเป็น 3,2,3 และ 2 ตามลำดับ สมาชิกของฟาเซตจะเป็นคำที่เปลี่ยนแปลงได้ ในประโยคตัวอย่างนี้ใช้สำหรับการอธิบายพฤติกรรมของการอ้างอิงกฎหรือการประยุกต์กฎเพื่อใช้ในการทำนายผลสัมฤทธิ์ ฟาเซตที่ 3 ในประโยคจับคู่ดังกล่าว เสนอแนะว่ากฎควรจะมีการจัดลำดับขั้นที่คล้ายกันในการบรรยายในการเรียนการสอน จุดสำคัญที่กัตแมน (Guttman)

ต้องการชี้ให้เห็นคือความสัมพันธ์ระหว่างชนิดของผลสัมฤทธิ์ทางการเรียนที่กำหนดไว้ใน การเรียน การสอน และความสามารถทางสมองที่จะนำไปปฏิบัติ เช่นนักเรียนคนหนึ่งมีระดับการอ้างอิงกฎ และการประยุกต์กฎอยู่ในระดับสูง ในการลำดับตัวเลขทำให้เขาสามารถมีผลสัมฤทธิ์ คือ ไป ปฏิบัติให้ถูกต้องได้

จากประโยคจับคู่ที่เกิดจากการจับคู่สมาชิกของแต่ละฟาเซท ที่แสดงให้เห็นในแผนภาพ ที่ 1 ทำให้เราสามารถสร้างประโยคได้ 36 ประโยคที่แตกต่างกัน ซึ่งมีทั้งถูกและผิด ดัง ตัวอย่าง 2 ประโยคต่อไปนี้

- 1) ถ้านักเรียนได้รับคำถามเกี่ยวกับภาษาและต้องการอ้างอิงกฎที่เหมือนกับที่ครูสอนใน ชั้นเรียนของตัวเองแล้ว นักเรียนควรจะตอบคำถามผิด
- 2) ถ้านักเรียนได้รับคำถามเกี่ยวกับตัวเลข และต้องการประยุกต์กฎที่คล้ายกันที่ครูสอน ในชั้นเรียนแล้ว นักเรียนควรจะตอบคำถามถูก

รูปแบบการสร้างข้อคำถาม จากการจับคู่ของสมาชิกของแต่ละฟาเซทให้เป็นประโยคนี้นอกจากจะใช้สร้างเป็นแบบทดสอบแบบถูกผิดแล้ว ยังสามารถสร้างเป็นแบบทดสอบแบบเลือกตอบ โดยการทำให้ฟาเซท 3 ฟาเซทหลังคงที่ จากนั้น สร้างประโยคใหม่แล้วให้ฟาเซทแรก (ภาษา ตัวเลข และรูปภาพ) ซึ่งมีคำตอบถูก 1 ตัว และผิด 2 ตัว และถ้าจะสร้างแบบทดสอบแบบเติมคำทำได้ โดยการดึงสมาชิกของฟาเซทใดฟาเซทหนึ่ง ออกจากประโยคทั้ง 36 ประโยค ในกรณีนี้เราสามารถสร้างแบบทดสอบเติมคำได้ 144 ข้อ เพราะว่าโดยอาศัยประโยค 36 ประโยคนี้นี้มีฟาเซทอยู่ 4 ฟาเซท โดยใน 1 ประโยคนี้สามารถสร้างข้อสอบเติมคำได้ 4 ข้อ ($36 \times 4 = 144$ ข้อ) ถ้าฟาเซทที่ 4 ในตารางนี้ขยายสมาชิกเป็น 4 ตัว ในกรณีนี้จึงเป็นการเพิ่มจำนวนประโยคหรือข้อคำถามที่สามารถสร้างได้เป็น 72 ประโยค เพื่อให้เห็นชัด เราจะพบว่ายังจำนวนสมาชิกมีมาก หรือจำนวนฟาเซทยิ่งมากในประโยคจับคู่หนึ่ง ๆ ยิ่งทำให้จำนวนข้อสอบเพิ่มมากขึ้น

2.1.3 หลักการและข้อเสนอแนะในการใช้ รูปแบบฟาเซทสร้างข้อสอบและตัวเลือก

2.1.3.1 หลักการและข้อเสนอแนะในการสร้างรูปแบบฟาเซท เพื่อสร้างข้อสอบ

เอ็นเจล และ มาร์ตูซา (Engel and Martuza, 1976 quoted in Roid and Haladyna, 1982) และเบิร์ก (Berk, 1978) ได้เสนอแนะเกี่ยวกับการสร้างแบบทดสอบวัดผลสัมฤทธิ์โดยใช้รูปแบบฟาเซท โดยให้ศึกษาจุดประสงค์การสอน เพื่อเป็นพื้นฐานในการสร้างข้อความสำหรับการสร้างประโยคจากการจับคู่ (Mapping Sentence) และได้กำหนดหลักการสร้างข้อสอบ โดยใช้รูปแบบฟาเซท 6 ขั้นตอนดังนี้

- 1) คัดเลือกจุดประสงค์การสอน (Selecting an Instructional Objective)

จุดประสงค์ที่จะใช้สร้างประโยชน์จากการจับคู่ จะต้องเป็น จุดประสงค์ที่สำคัญและให้ความคิดรวบยอด เพื่อเป็นแนวทางในการกำหนดรูปแบบฟาเซท

2) แสดงเนื้อหาที่จะใช้สร้างเครื่องมือ (List Instructional Material) รายละเอียดของเนื้อหาสาระ ความต้องการของผู้เขียนข้อสอบ เพื่อที่จะนำไปขยายให้เป็นจุดประสงค์ขยายความ (Amplifying Objective) โดยจะศึกษาได้จากหนังสือเรียน คู่มือครู ซึ่งมีรายละเอียดของเนื้อหาและจุดประสงค์

3) พัฒนาจุดประสงค์ขยายความ (Develop and Amplified Objective) ของเนื้อหาให้เฉพาะเจาะจงขึ้น ซึ่งประกอบด้วยตัวอย่างข้อคำถาม บรรยายสถานการณ์และลักษณะคำตอบของข้อคำถาม

4) การสร้างประโยคจับคู่ (Generate a Mapping Sentence) จุดประสงค์ขยายความจะใช้เป็นพื้นฐานในการสร้างประโยคจากการจับคู่ โดยสมาชิกในแต่ละฟาเซท จะต้องสร้างให้อยู่ในขอบเขตของความเป็นไปได้

5) การกำหนดโครงสร้างฟาเซทของข้อสอบ (Generate the Item Facet Structure)

คำนึงถึง 2 ลักษณะคือ ความสมบูรณ์ของความเป็นตัวแทนในมวลความรู้ และความเหมาะสมของสมาชิกฟาเซท

6) การเขียนข้อสอบ (Write the Item)

ในส่วนนี้สามารถใช้คอมพิวเตอร์ผลิตข้อสอบออกมาได้ หรือจะกระทำโดยการเลือกจากสมาชิกของแต่ละฟาเซท และการแปรเปลี่ยนไปอย่างเป็นระบบของตัวลง ทำได้เช่นเดียวกันกับคำตอบถูก

2.1.3.2 การพัฒนาการเขียนข้อสอบจากรูปแบบฟาเซท

รันเคิล และ แมคกราธ (Runkel and Macgraph, 1972 quoted in Roid and Haladyna, 1982) ได้เสนอกฎเกณฑ์สำหรับการพัฒนาการเขียนข้อสอบจากรูปแบบฟาเซทดังนี้

1) ฟาเซทต่าง ๆ ที่ถูกเลือกจะต้องสามารถอธิบายโน้ตหลักของจุดประสงค์นั้น ๆ ได้

2) ฟาเซทแต่ละฟาเซทจะประกอบด้วยสมาชิกที่เป็นไปได้ในแต่ละฟาเซท

3) สมาชิกของฟาเซทแต่ละฟาเซท ต้องเป็นสมาชิกที่แยกจากกัน (Mutually Exclusive) และถ้าจุดประสงค์ใดสามารถแบ่งออกเป็นฟาเซทได้มากกว่าหนึ่งฟาเซท แล้วฟาเซทที่แบ่งออกมานั้นจะต้องมีความสมบูรณ์ครบถ้วนในตัวเอง

4) ความสัมพันธ์ระหว่างฟาชะทควรจะมีเหตุผล และมีลักษณะเฉพาะ ต้องมีการจัดอันดับของฟาชะทที่แน่นอน และประโยชน์จับคู่สามารถเชื่อมโยงแง่มุมต่าง ๆ ให้เป็นเรื่องที่ต่อเนื่องกัน

5) สมาชิกของฟาชะทแต่ละฟาชะทควรมีเหตุผลที่สามารถอธิบายได้ว่า ทำไมจึงถูกเลือกเป็นสมาชิกของฟาชะทนั้น ๆ

6) จำนวนฟาชะทที่เป็นไปได้ทั้งหมด จะอยู่ภายใต้ขอบเขตเนื้อหาสาระที่สอน

2.1.3.3 การสร้างตัวเลือกโดยใช้รูปแบบฟาชะท

ตัวลวงเป็นส่วนหนึ่งของข้อสอบ และใช้ตัดสินระดับความยากของข้อสอบโดยเชื่อมกับส่วนคำถาม กัดแมน และชลีเซนเจอร์ (Guttman and Schlesenger, 1967) กล่าวว่า ส่วนของคำถามของข้อสอบจะไม่สามารถบอกค่าความยากของข้อสอบข้อนั้นได้เลย ถ้าตัวลวงของข้อสอบนั้นไม่ได้สร้างอย่างเป็นระบบ เพราะการเขียนตัวลวงอย่างเป็นระบบจะทำให้เกิดผลดี ดังนี้

1) ข้อสอบข้อนั้นสามารถพยากรณ์ความสัมพันธ์ของค่าความยากของตัวลวงแต่ละตัวได้

2) ครูสามารถแปลผลของแบบทดสอบได้อย่างชัดเจน เพราะได้มีการตัดตัวลวงที่ไม่เกี่ยวข้องออก

3) ความแตกต่างของคะแนนที่ได้จากการสอบของนักเรียนแต่ละคน ขึ้นอยู่กับรูปแบบของความสามารถในการตัดสินใจเลือกของตัวลวง

นอกจากนั้นการสร้างตัวลวงอย่างมีระบบยังทำให้สามารถสร้างแบบทดสอบที่มีความยาวนานกว่าปกติโดยไม่มีผลกระทบกับความเชื่อมั่นและค่าความตรงของแบบทดสอบ

การเขียนข้อสอบโดยใช้รูปแบบฟาชะทนั้น กัดแมนและชลีเซนเจอร์ (Guttman and Schlesenger, 1967) เสนอว่าใช้ได้ทั้งการสร้างข้อคำถาม (Stem) และใช้ในการสร้างตัวลวงที่เป็นระบบ ซึ่งสามารถใช้ในการวินิจฉัยได้ด้วย ซึ่งจะขอยกตัวอย่างของเอ็นเจล (Engel 1976 quoted in Roid and Haladyna, 1982) ดังแผนภาพที่ 2

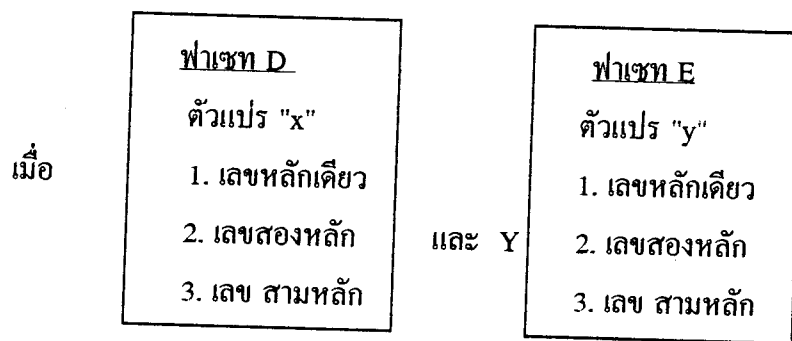
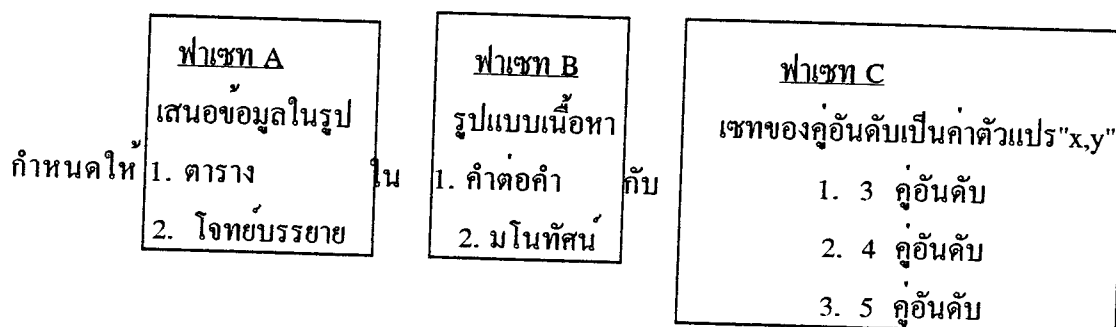
แผนภาพที่ 2 แสดงตัวอย่างฟาชะท ที่ใช้ในการเขียนข้อสอบโดยใช้รูปแบบฟาชะทของกัดแมน และชลีเซนเจอร์

จุดประสงค์

1. เมื่อกำหนดค่าของคู่อันดับ x และ y ให้ สามารถคำนวณหาค่า R_{xy} ได้ถูกต้อง
2. เมื่อกำหนดค่าสหสัมพันธ์แบบ Pearson r , Point Biserial, Phi และ Rank Order (ρ)

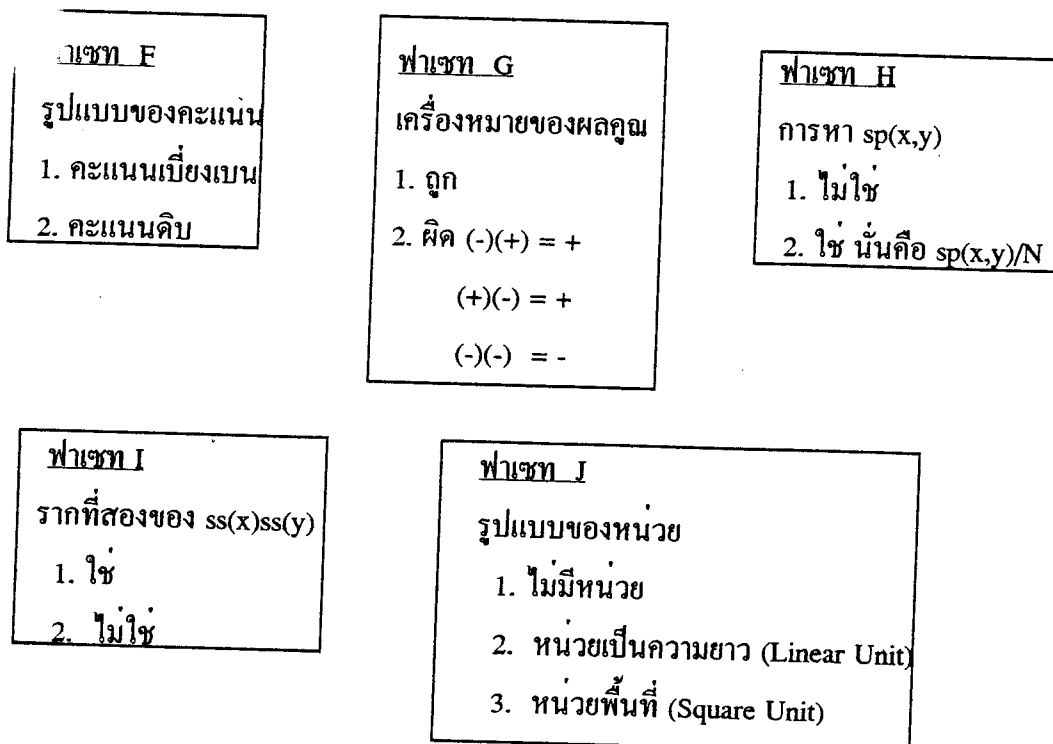
ผู้เรียนจะเปรียบเทียบความเหมือนและความต่างกันได้ถูกต้อง

ฟาเซทที่ใช้สร้างข้อคำถาม สำหรับจุดประสงค์ที่ 1 การสร้างประโยชน์จากการจับคู่



ฟาเซทที่ใช้สร้างตัวเลือก

นักเรียนจะเลือกคำตอบถูกของค่าสหสัมพันธ์ $r(x,y)$ จากเซตของตัวเลือกซึ่งสร้างจาก



จากประโยชน์จากการจับคู่เอ็นเจล (Engel) ได้สร้างข้อสอบไว้ดังนี้
 แสดงตัวอย่างการเขียนข้อสอบโดยใช้รูปแบบฟาชะทของกัตแมน และชลีเซนเจอร์ ที่สร้าง
 จากรูปแบบฟาชะท $A_1 B_1 C_1 D_1 E_1$ คือ
 (อ) ตารางข้างล่างนี้เป็นคะแนนทัศนคติก่อน-หลัง การสอบของนักเรียน 3 คน ในวิชา
 ออกแบบ (BU 307) ให้นักเรียนหาค่า Pearson r จากข้อมูลต่อไปนี้

แบบวัดทัศนคติ	
ก่อนสอบ	หลังสอบ
1	6
1	3
4	3

ก. -0.50 ข. 0.67 ค. 0.50 ง. -0.17

เอ็นเจล (Engel) กล่าวว่า ถ้านักเรียนคนใดเลือกตอบ (ตัวเลือก ก) โดยอาศัยรูปแบบ
 ฟาชะท ในการสร้างตัวลวงจะทำให้ทราบความสัมพันธ์ระหว่างตัวถูก (ตัวเลือก ก) และตัวลวง
 (ตัวเลือก ค) ดังนี้

ก. $F_1 G_1 H_1 I_1 J_1$

ค. $F_1 G_2 H_1 I_1 J_1$

ข้อสังเกต จะเห็นได้ว่าในระหว่างตัวเลือกถูก กับตัวลวงนั้นมีความแตกต่างกันที่ฟาชะท G โดยที่
 สมาชิกของฟาชะท G คือ G_2 นั้นทำให้ทราบว่านักเรียนมีความบกพร่องที่การใช้เครื่องหมาย
 ของผลคูณ อันจะเป็นแนวทางในการสอนซ่อมเสริม

2.1.4 ประโยชน์และข้อจำกัดของรูปแบบฟาชะท

เอ็นเจล และมาร์ทุซา (Engel and Martuza, 1976 quoted in Rioid and Haladyna, 1982)
 ได้กล่าวถึงประโยชน์ของประโยชน์จับคู่ไว้ดังนี้

1) คำถาม และตัวลวง สามารถเขียนเพิ่มขึ้นอย่างมีระบบ โดยที่ตัวลวงสามารถเลือกจาก
 สมาชิกของแต่ละฟาชะทได้ รอยด์และฮาလာไดนา (Roid and Haladyna, 1980) กล่าวว่า ในการ
 เลือกตัวลวงนั้นแต่เดิมนับได้ว่าเป็นปัญหาใหญ่ปัญหาหนึ่งของผู้เขียนข้อสอบแบบเลือกตอบ และ

ยังขึ้นกับความชำนาญของผู้เขียนข้อสอบแต่ละคน แต่ก็สามารถทำได้ง่ายขึ้นได้โดยการเขียนตัวลวงอย่างเป็นระบบของประโยชน์ที่เกิดจากการจับคู่สมาชิกของฟาเซท

2) ตัวลวงแต่ละตัวจะมีความเกี่ยวข้องกัน แต่จะมีบางตัวที่น่าจะเป็นตัวถูกมากกว่าตัวอื่น ดังนั้นการที่นักเรียนเลือกตอบตัวลวงตัวใดตัวหนึ่ง ครูย่อมวินิจฉัยได้ว่านักเรียนที่ตอบโดยเลือกตัวลวงที่มีความสมเหตุสมผลน้อยกว่า ควรจะได้รับการสอนซ่อมเสริมมากกว่านักเรียนที่เลือกตัวลวงที่มีเหตุผลมากกว่า

3) ในการเขียนข้อสอบจากประโยชน์จับคู่ นั้น เป็นวิธีการพื้นฐานของทั้งการเรียนการสอน และการสอบทำให้ข้อสอบที่เขียนออกมานั้นมีความสัมพันธ์เชิงตรรกศาสตร์ (Logical Relationship) กับเนื้อหาที่ใช้ในการสอน

4) การสร้างแบบทดสอบเลือกตอบหรือแบบทดสอบคู่ขนานนั้น สามารถทำได้ง่าย ข้อสอบที่เลือกมาเป็นแบบทดสอบแต่ละฉบับนั้น สามารถสุ่มได้มาโดยปราศจากความลำเอียง และมีความสัมพันธ์กันทั้งเนื้อหา ค่าความยาก และค่าอำนาจจำแนก

5) สามารถใช้เครื่องคอมพิวเตอร์ในการเขียนข้อสอบได้ (Millman, 1980)

6) ขั้นตอนแรกของรูปแบบฟาเซทนั้น เริ่มจากการใช้จุดประสงค์การเรียนการสอนนำมาเป็นพื้นฐานในการสร้างประโยชน์จับคู่ ถ้านักวางแผนหลักสูตรการศึกษาสามารถรวมจุดประสงค์การเรียนการสอน แล้วแปลงให้เป็นประโยชน์จับคู่ วิธีการนี้จะทำให้การกำหนดขอบเขตของรูปแบบฟาเซทได้ง่ายขึ้น

บุญเชิด ภิญโญอนันตพงษ์ (2527) ได้กล่าวถึงข้อดีและข้อจำกัดของรูปแบบฟาเซทไว้ดังนี้

ข้อดี

- 1) การวิเคราะห์เนื้อหาวิชาใช้การวิเคราะห์เชิงทฤษฎี หรือหลักเหตุผลพิจารณาไม่ต้องวิเคราะห์จากผลการปฏิบัติ
- 2) ใช้การจับคู่สมาชิกของแต่ละฟาเซท เป็นตัวช่วยนิยามมวลความรู้ในเนื้อหาวิชา และสมาชิกของข้อสอบที่เกี่ยวข้องที่ใช้วัดผลสัมฤทธิ์ของมวลความรู้นั้น
- 3) ง่ายและสะดวกกว่าการเขียนฟอร์มข้อสอบ
- 4) เหมาะสมสำหรับใช้เขียนข้อสอบถูกผิดหรือเติมคำมากที่สุดและสามารถดัดแปลงไปเป็นข้อสอบแบบเลือกตอบได้

ข้อจำกัด

- 1) ไม่มีกฎเกณฑ์ใด ๆ ที่จะช่วยในการสร้างประโยชน์จากการจับคู่สมาชิกของฟาชะท ทำให้ได้ประโยชน์ที่แตกต่างกันขึ้นกับผู้เขียน ซึ่งทำให้การเขียนข้อสอบแบบนี้ ยังคงต้องอาศัยประสบการณ์ส่วนตัว
- 2) เหมาะสำหรับบางเนื้อหาวิชา
- 3) เซทของประโยชน์ที่นำมาใช้จับคู่เพียงเซทเดียวอาจไม่เป็นตัวแทนของเนื้อหาที่ต้องการวัดทั้งหมด
- 4) ประโยชน์ที่ได้จากการจับคู่จะเป็นตัวกำหนดให้ได้ฟอร์มข้อสอบเพียงฟอร์มเดียว ไม่สามารถสร้างข้อสอบให้มีฟอร์มข้อสอบตามที่ต้องการได้

สำหรับรูปแบบฟาชะทที่ผู้วิจัยใช้ในการเขียนข้อสอบครั้งนี้ มีลักษณะดังต่อไปนี้

1. มีการนิยามขอบเขตของเนื้อหา
2. สร้างรูปแบบฟาชะท ซึ่งมีส่วนประกอบที่สำคัญ 3 ส่วน (รายละเอียดหน้า 4)
3. สร้างข้อสอบโดยการกำหนดลักษณะเฉพาะของการเลือกสมาชิกของแต่ละฟาชะท จากรูปแบบฟาชะทละ 1 ตัว ทั้งในส่วนที่เป็นคำถาม และส่วนของตัวเลือก
4. สร้างประโยชน์จากการเลือกจับคู่สมาชิกของแต่ละฟาชะทเชื่อมเข้ากับส่วนที่คงที่ เพื่อให้ได้ข้อสอบที่สามารถวัดจุดประสงค์นั้นได้

2.2 แบบทดสอบคู่ขนาน (parallel Test)

อัลเลน และเยน (Allen and Yen, 1976 : 51 - 71) ได้กำหนดข้อตกลงเบื้องต้นของทฤษฎีการทดสอบดั้งเดิม (The Assumptions of Classical True-Score Theory) ซึ่งเกี่ยวข้องกับความหมายของแบบทดสอบคู่ขนาน (Parallel Tests) ว่าจะต้องมีข้อตกลงเบื้องต้นดังนี้

ข้อตกลงเบื้องต้นที่ 1 กล่าวถึงคะแนนจากการวัด (X) ว่าเป็นผลรวมจากคะแนนจริง (T) และคะแนนซึ่งเกิดจากความคลาดเคลื่อน (E) ซึ่งเขียนเป็นสมการได้คือ

$$X = T + E$$

เมื่อ X แทน คะแนนจากการวัด (Observed Score)

T แทน คะแนนจริง (True Score)

E แทน คะแนนที่ได้จากความคลาดเคลื่อน (Errors Score)

ข้อตกลงเบื้องต้นที่ 2 กล่าวถึงค่าที่เป็นไปได้ตามทฤษฎีหรือค่าความคาดหวัง (Expected Value) ของคะแนนเฉลี่ยที่ได้จากประชากร จะมีค่าเท่ากับคะแนนจริง (T) ของการวัดในครั้งนั้น ๆ กล่าวคือ เป็นการแจกแจงคะแนนเฉลี่ยเชิงทฤษฎีของคะแนนจากการวัด $E(X)$ นอกจากนี้ยังหมายถึงคะแนนที่ได้จากการสอบกับนักเรียนเพียงคนเดียวแต่เป็นการทดสอบด้วยแบบทดสอบฉบับเดียวหลายหน ข้อตกลงเบื้องต้นนี้สามารถเขียนเป็นสมการได้ดังนี้

$$E(X) = T$$

เมื่อ $E(X)$ แทน คะแนนค่าความคาดหวังของ X (Expected Value)

X แทน คะแนนจากการสอบ (Observed Score)

T แทน คะแนนจริง (True Score)

ข้อตกลงเบื้องต้นที่ 3 กล่าวถึงคะแนนที่เกิดจากความคลาดเคลื่อนในการวัดและคะแนนจริง เมื่อนำไปทดสอบประชากร (จำนวนมาก) ด้วยแบบทดสอบชุดหนึ่ง ๆ ไม่มีความสัมพันธ์กัน ดังสมการต่อไปนี้

$$\rho_{ET} = 0$$

เมื่อ ρ_{ET} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความคลาดเคลื่อนในการวัดกับคะแนนจริง

E แทน ความคลาดเคลื่อนในการวัดด้วยแบบทดสอบ

T แทน คะแนนจริงที่ได้จากการวัดด้วยแบบทดสอบ

ข้อตกลงเบื้องต้นที่ 4 กล่าวถึง คะแนนที่เกิดจากความคลาดเคลื่อนจากแบบทดสอบสองฉบับ คือ E_1 และ E_2 เมื่อนำมาหาค่าสหสัมพันธ์กันแล้ว จะมีค่าเป็นศูนย์ หรือหมายถึงคะแนนที่เกิดจากความคลาดเคลื่อนของแบบทดสอบฉบับที่ 1 และฉบับที่ 2 ไม่มีความสัมพันธ์กัน ซึ่งเขียนเป็นสมการได้ดังนี้

$$\rho_{E_1E_2} = 0$$

เมื่อ $\rho_{E_1E_2}$ แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนความคลาดเคลื่อนของแบบทดสอบฉบับที่ 1 กับคะแนนจริงฉบับที่ 2

E_1 แทน คะแนนความคลาดเคลื่อนในการวัดด้วยแบบทดสอบฉบับที่ 1

E_2 แทน คะแนนความคลาดเคลื่อนในการวัดด้วยแบบทดสอบฉบับที่ 2

ข้อตกลงเบื้องต้นที่ 5 กล่าวถึงคะแนนจากความคลาดเคลื่อนจากแบบทดสอบฉบับที่ 1 (E_1) กับคะแนนจริงของแบบทดสอบฉบับที่ 2 (T_2) มีความสัมพันธ์กันดังสมการ

$$\rho_{E_1 T_2} = 0$$

เมื่อ $\rho_{E_1 T_2}$ แทน ค่าสหสัมพันธ์ระหว่างคะแนนความคลาดเคลื่อนของแบบทดสอบฉบับที่ 1 กับคะแนนจริงแบบทดสอบฉบับที่ 2

E_1 แทน คะแนนความคลาดเคลื่อนในการวัดด้วยแบบทดสอบฉบับที่ 2

T_2 แทน คะแนนจริงของการวัดด้วยแบบทดสอบฉบับที่ 2

ความคลาดเคลื่อนตามข้อตกลงเบื้องต้นที่ได้กล่าวมาแล้ว เกิดจากการวัดผลอย่างไม่เป็นระบบ (Unsystematic) หรือเป็นความคลาดเคลื่อนในลักษณะการสุ่ม (Random) ภายใต้สภาพการณ์ของคะแนนที่ได้จากการวัด ซึ่งเบี่ยงเบนไปในเชิงทฤษฎี

ข้อตกลงเบื้องต้นที่ 6 กล่าวถึง นิยามของแบบทดสอบคู่ขนานกล่าวคือ เมื่อมีแบบทดสอบอยู่ 2 ฉบับ แล้วนำไปทดสอบกับประชากรหรือกลุ่มผู้สอบกำหนดให้ X คือ คะแนนจากการสอบด้วยแบบทดสอบฉบับที่ 1 และ X' คือคะแนนที่ได้จากการทดสอบด้วยแบบทดสอบฉบับที่ 2 T เป็นคะแนนจริงของประชากรที่ได้รับการทดสอบฉบับที่ 1 T' เป็นคะแนนจริงของประชากรที่ได้รับการทดสอบด้วยแบบทดสอบฉบับที่ 2 และ σ^2_x เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนที่วัดด้วยแบบทดสอบฉบับที่ 1 $\sigma^2_{x'}$ เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนที่วัดด้วยแบบทดสอบฉบับที่ 2 ดังนั้นแบบทดสอบสองฉบับที่มีลักษณะคู่ขนานกันจะต้องมีความเท่าเทียมกันของค่าต่าง ๆ ดังนี้

$$1. T' = T$$

$$2. \sigma^2_{E_1} = \sigma^2_{E_2}$$

3. ค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) นอกจากนี้อัตราสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบทั้งสองฉบับจะต้องมีค่าสูงมาก ๆ อีกด้วย

ข้อตกลงเบื้องต้นที่ 7 เป็นข้อตกลงที่คล้ายกับข้อ 6 ซึ่งเป็นการให้นิยามของแบบทดสอบที่มีคะแนนจริงเทียบเท่า (T-Equivalent Test) กล่าวคือ แบบทดสอบที่มีลักษณะของคะแนนจริง

เทียบเท่ากันนั้น จะต้องมีแบบแผนของคะแนนเป็นแบบเดียวกันคือคะแนนจริงที่ได้จากแบบทดสอบฉบับหนึ่งฉบับใดบวกด้วยค่าคงที่ค่าหนึ่งเสมอ

ธอร์นไดค์ และฮาเกน (Thorndike and Hagen, 1968 : 181, 647) กล่าวว่า แบบทดสอบคู่ขนานหรือแบบทดสอบฟอร์มเทียบเท่า (Equivalent Form) จะสร้างขึ้นมาจากลักษณะเฉพาะของข้อสอบเดียวกัน เพื่อนำมาวัดคุณลักษณะด้านเดียวหรือหลายด้าน และต้องมีค่าสถิติแสดงคุณลักษณะแบบทดสอบเหมือนกัน

จริยา เสดบุตร (2522 : 136) กล่าวว่าแบบทดสอบคู่ขนานจะต้องมีเนื้อหาเดียวกันมีค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากัน และคะแนนจริงที่ได้รับจากแบบทดสอบเท่ากัน

ยาใจ พงษ์บริบูรณ์ (2522 : 98) กล่าวถึง การสร้างแบบทดสอบคู่ขนานว่าจะต้องวัดเนื้อหาและพฤติกรรมอย่างเดียวกัน มีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานใกล้เคียงกัน ระดับความยากของข้อสอบ และค่าอำนาจจำแนกใกล้เคียงกัน

พิตร ทองชั้น (2524 : 80 - 81) ได้กล่าวถึงการสร้างแบบทดสอบคู่ขนาน 2 วิธี คือ

- 1) โดยการสร้างแบบทดสอบขึ้นมา 2 ชุด โดยให้มีจำนวนมาก ๆ ข้อ แล้วนำไปทดลองกับเด็ก จากนั้นก็ทำการวิเคราะห์ข้อสอบหาค่าความยาก และค่าอำนาจจำแนกแล้ว นำข้อสอบที่มีค่าความยาก และค่าอำนาจจำแนกใกล้เคียงกันมาจัดเป็นคู่ ๆ เพื่อจัดให้อยู่ในแบบทดสอบทั้งสองฉบับ
- 2) โดยการสร้างตามโครงสร้างของเนื้อหาวิชาซึ่งได้แบบทดสอบที่มีเนื้อหาต่าง ๆ คล้ายคลึงกันจากนั้นก็นำไปทดลองกับเด็กพร้อมกันทั้งสองชุด เสร็จแล้วนำมาหาค่าสหสัมพันธ์ระหว่างแบบทดสอบทั้งสองฉบับ ถ้าหากมีความสัมพันธ์กันสูงก็ถือว่าใช้ทดสอบแทนกันได้

พวงรัตน์ ทวีรัตน์ (2531:129) กล่าวว่า แบบทดสอบคู่ขนานจะต้องมีคุณภาพเหมือนกันทุกประการ นั่นคือ เนื้อหา รูปแบบข้อคำถาม จำนวนข้อ ความยาก และมีค่าคะแนนเฉลี่ย และความแปรปรวนเท่ากันทั้งสองฉบับไปสอบวัดกับกลุ่มเดียวกัน เมื่อได้คะแนน 2 ชุดแล้ว นำมาหาค่าสัมประสิทธิ์ตามวิธีของเพียร์สัน ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้เรียกว่า สัมประสิทธิ์ของความเท่าเทียมกัน (Equivalent Coefficient)

บุญชม ศรีสะอาด (2532 : 94) กล่าวถึง แบบทดสอบคู่ขนาน คือ แบบทดสอบ 2 ฉบับ หรือมากกว่าที่ให้แทนกันได้ เพราะวัดในเนื้อหาและจุดประสงค์เดียวกัน มีระดับความยาก ค่าคะแนนเฉลี่ย และค่าความแปรปรวนเท่ากัน และมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบสูง

บุญเชิด ภิญโญนันตพงษ์ (2532 : 153) ได้กล่าวถึงรูปแบบของแบบทดสอบคู่ขนาน ซึ่งมีอยู่ 2 ลักษณะดังนี้

1) แบบทดสอบคู่ขนานที่สร้างตามทฤษฎีการทดสอบดั้งเดิม หมายถึง แบบทดสอบที่สร้างตามทฤษฎีการทดสอบดั้งเดิม ซึ่งแบบทดสอบคู่ขนานกันนั้นจะมีเนื้อหาสาระของข้อสอบเหมือนกัน ค่าเฉลี่ย ค่าความแปรปรวนของคะแนน และค่าสหสัมพันธ์ระหว่างข้อสอบเท่ากัน

2) แบบทดสอบคู่ขนานแบบลุ่ม หมายถึง แบบทดสอบที่สร้างมาจากเทคโนโลยีการเขียนข้อสอบอย่างใดอย่างหนึ่ง เช่น การเขียนข้อสอบจากจุดประสงค์ขยายความ ลักษณะเฉพาะของแบบทดสอบ ประโยคจากการจับคู่ (ทฤษฎีแ่งมุมปัญหา) การแปลงประโยคมาเป็นฟอรม์ข้อสอบ และการใช้คอมพิวเตอร์เขียนข้อสอบเป็นต้น การเขียนข้อสอบจากเทคโนโลยีการเขียนข้อสอบหรือกฎเกณฑ์ในการเขียนข้อสอบเดียวกัน จะมีลักษณะคู่ขนานกันหมด ดังนั้น เมื่อลุ่มข้อสอบมากก็จะได้ข้อสอบที่จัดเป็นแบบทดสอบคู่ขนานกันได้โดยไม่ต้องอาศัยคุณสมบัติทางสถิติเหมือนแบบแรก

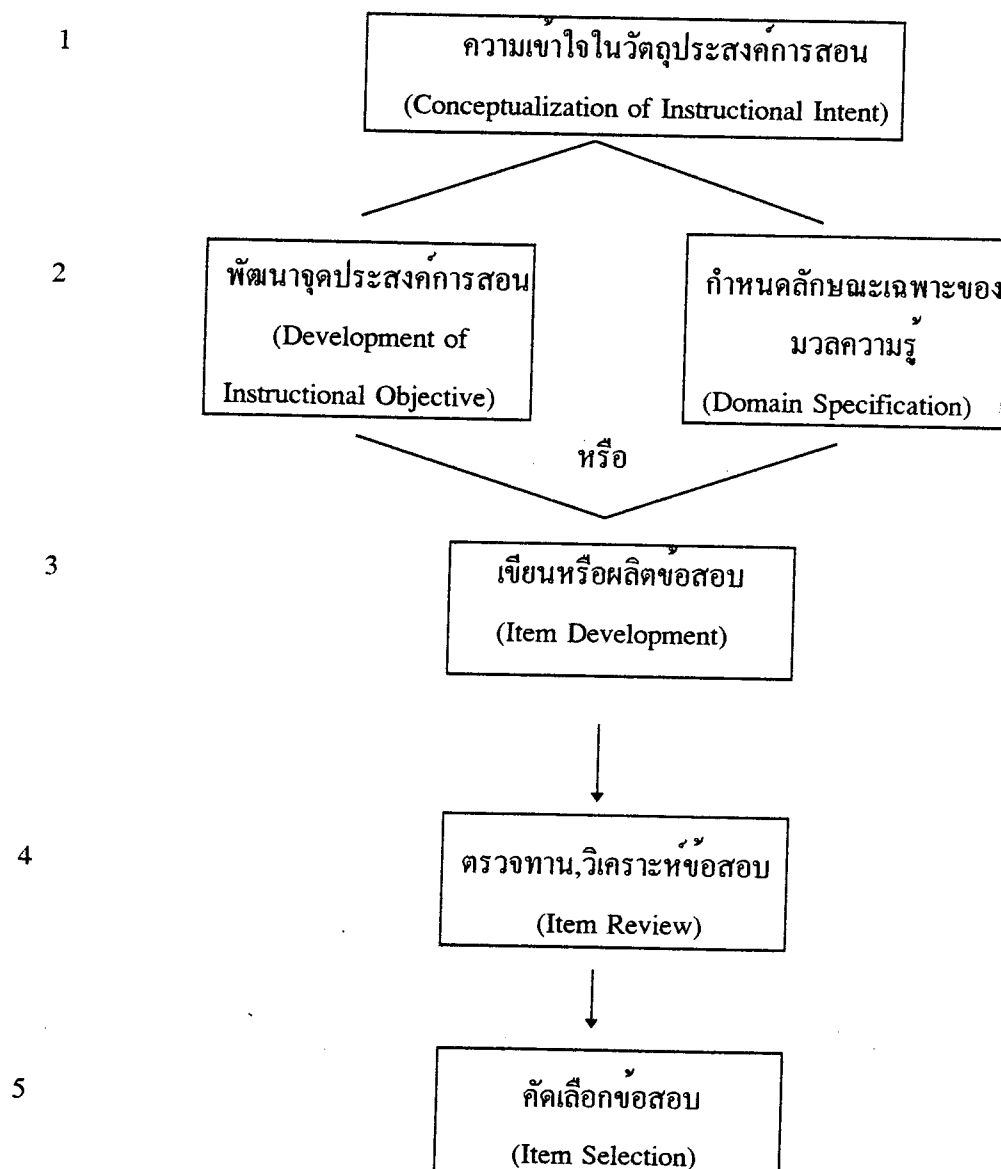
จากแนวคิดของนักวัดผลการศึกษาที่กล่าวมาข้างต้นสรุปได้ว่า แบบทดสอบคู่ขนาน หมายถึง แบบทดสอบที่วัดในเนื้อหาและจุดประสงค์เดียวกัน รูปแบบของข้อคำถามเหมือนกัน จำนวนข้อสอบและเวลาทำข้อสอบเท่ากันและค่าสถิติของแบบทดสอบ ซึ่งได้แก่ ค่าคะแนนเฉลี่ย คะแนนจุดตัด และคุณภาพของข้อสอบในด้านค่าความแปรปรวน ค่าความยาก ค่าอำนาจจำแนก และค่าความเชื่อมั่นไม่แตกต่างกัน นอกจากนี้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบคู่ขนานต้องมีค่าสูง ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการทดสอบความเป็นคู่ขนานของแบบทดสอบ 3 ฉบับโดยครู 6 คนสร้างในด้านค่าสถิติพื้นฐาน คือ คะแนนเฉลี่ย ค่าความแปรปรวน และคุณภาพของข้อสอบในด้านค่าความยาก เฉลี่ยและค่าความเชื่อมั่น เนื่องจากเป็นแบบทดสอบอิงเกณฑ์ เพื่อนำค่าคะแนนจุดตัดไปทดสอบ ผู้วิจัยจึงได้ทำการทดสอบค่าความเชื่อมั่นของแบบทดสอบ

2.3 หลักการสร้างและการตรวจสอบ คุณภาพของแบบทดสอบอิงเกณฑ์

2.3.1 หลักการสร้างแบบทดสอบอิงเกณฑ์

รอยด์และฮาလာไดนา (Roid and Haladyna. 1980) ได้เสนอหลักการเขียนข้อสอบ เพื่อพัฒนาแบบทดสอบอิงเกณฑ์ ซึ่งมีสาระสำคัญ 5 ขั้นตอนพอสรุปได้ดังแผนภาพที่ 3

แผนภาพที่ 3 ขั้นตอนการสร้างและพัฒนาแบบทดสอบอิงเกณฑ์
ขั้นตอน (Steps)



จากภาพที่ 1 จะกล่าวถึงรายละเอียดของแต่ละขั้นตอนดังนี้

ขั้นตอนที่ 1 ความเข้าใจในวัตถุประสงค์การสอน นับเป็นขั้นตอนแรกในการสร้างข้อสอบ ที่ผู้สร้างต้องทำความเข้าใจในจุดประสงค์ที่ต้องการวัดในหัวข้อนั้น ๆ การวิเคราะห์วัตถุประสงค์หลักนี้อาจต้องใช้วิธีการวิเคราะห์งาน (Task Analysis or Job Analysis)

ขั้นตอนที่ 2 การพัฒนาจุดประสงค์การสอน หรือการกำหนดลักษณะเฉพาะของมวลความรู้ขั้นตอนนี้จะเป็นการแปลงวัตถุประสงค์หลักจากขั้นตอนที่ 1 ให้อยู่ในรูปของวัตถุประสงค์

เชิงพฤติกรรม ซึ่งจะบอกพฤติกรรมของผู้เรียนที่ได้จากการสอน หรือเป็นการกำหนดลักษณะเฉพาะของมวลความรู้

ขั้นตอนที่ 3 การเขียนหรือผลิตข้อสอบ เป็นขั้นตอนที่ใช้เทคนิคการเขียนข้อสอบ โดยเขียนให้สอดคล้องกับจุดมุ่งหมายเชิงพฤติกรรม หรือมวลความรู้ที่กำหนดไว้ ขั้นตอนนี้อาจนำเทคโนโลยีการเขียนข้อสอบมาใช้ได้

ขั้นตอนที่ 4 ตรวจทานและวิเคราะห์ข้อสอบ เป็นขั้นตอนในการตรวจสอบข้อบกพร่องของข้อสอบ นำข้อสอบไปทดลองใช้และปรับปรุงข้อสอบก่อนนำไปใช้จริง

ขั้นตอนที่ 5 คัดเลือกข้อสอบ เป็นขั้นตอนสุดท้าย ในการพัฒนาแบบทดสอบที่จะใช้คัดเลือกข้อสอบที่มีคุณภาพไว้ต่อไป

ตามแนวคิดของแบบทดสอบอิงเกณฑ์ นอกจากจะต้องกำหนดวัตถุประสงค์ และขอบเขตของเนื้อหาที่ต้องการจะสรุปพาดพิงไปถึงให้ชัดเจนแล้ว ยังมีความจำเป็นที่จะต้องระบุถึงขอบเขตประชากรของข้อสอบให้เด่นชัดอีกด้วย แนวคิดพื้นฐานดั้งเดิมง่าย ๆ ที่จะทำให้ได้ข้อสอบเป็นจำนวนมากพอที่จะเชื่อถือได้ว่าครอบคลุมเนื้อหาทั้งหมดที่ต้องการศึกษา คือ การสร้างธนาคารข้อสอบ (Shoemaker, 1975) โดยอาจใช้วิธีระดมนักเขียนข้อสอบมาช่วยกันเขียนและเก็บรวบรวมไว้ อย่างไรก็ตามวิธีดังกล่าวจะมีปัญหาในประเด็นที่ว่าผู้สร้างข้อสอบต่างคนกัน แม้จะได้รับการมอบหมายให้สร้างข้อสอบที่วัตถุประสงค์เดียวกันแต่ข้อสอบที่เขียนออกมาก็อาจมีความแตกต่างกันอยู่มาก จึงได้มีการพยายามพัฒนาแนวคิดการสร้างข้อสอบที่มีความเป็นปรนัยมากขึ้น ซึ่งรอยด์และฮาไลดนา (อ้างถึงใน Roid and Haladyna, 1974) ได้สรุปแนวคิดการสร้างข้อสอบในแบบทดสอบอิงเกณฑ์ ที่มีความเป็นปรนัยไว้ ซึ่งเรียกว่า "เทคโนโลยีการเขียนข้อสอบ"

2.3.2 การวิเคราะห์คุณภาพแบบทดสอบอิงเกณฑ์จะแยกกล่าวในแต่ละประเด็นดังนี้

2.3.2.1 ค่าความยากของข้อสอบ

Cox และ Garagas (อ้างถึงใน โกวิท, 2523) ใช้การเปลี่ยนแปลงของค่าความยากง่ายของรายข้อกระทง เมื่อใช้กับกลุ่มนักเรียนก่อนสอนกับหลังการสอนเป็นตัวชี้แนะ ถ้าค่าความยากง่ายของข้อกระทงใดไม่มีการเปลี่ยนแปลงกับการสอนสองครั้งนี้ก็ถือว่าขาดความเที่ยงตรงเชิงโครงสร้าง

Mayo (อ้างถึงใน กมล ภ. 2518) มีความคิดเห็นว่าแบบทดสอบอิงเกณฑ์นั้นสร้างขึ้นตามจุดมุ่งหมายของการสอน ซึ่งตามแนวคิดของการเรียนให้บรรลุจุดมุ่งหมายถือว่าผู้เรียนเกือบทั้งหมดควรจะบรรลุได้ ดังนั้นผู้เรียนส่วนใหญ่ควรจะทำข้อสอบได้ค่าความยากของข้อคำถามอาจจะสูงถึงร้อยละ 85 หรือมากกว่านั้น

ส.วาสนา ประवालพฤษ (2517) กล่าวว่า แบบทดสอบอิงเกณฑ์ข้อสอบยากง่ายไม่สำคัญเขียนข้อสอบเอาตามเนื้อหา และจุดมุ่งหมายที่วางไว้ และเทคนิคการใช้ค่าความยากง่ายของข้อสอบแบบอิงเกณฑ์เป็นตัวชี้แนะความเที่ยงตรงตามโครงสร้าง หลักการที่สำคัญก็คือข้อกระทงต่าง ๆ ที่วัดจุดประสงค์เดียวกันควรมีค่าความยากง่ายใกล้เคียงกัน (โกวิท, 2523)

อังคณา สายยศ (2526) กล่าวว่าความยากง่ายของข้อสอบอิงเกณฑ์ควรจะค่อนข้างง่าย เพราะถ้านักเรียนได้เรียนเรื่องนั้นจนรอบรู้ตามเกณฑ์ขั้นต่ำ (Minimum Requirement) ก็ควรจะได้ ดังนั้นข้อสอบอิงเกณฑ์ควรจะมีค่าความยากง่ายตั้งแต่ .50 หรือร้อยละ 50 ของคนทำถูกขึ้นไป

จะเห็นว่าข้อสอบแบบอิงเกณฑ์นั้นจะไม่คำนึงถึงค่าความยากของข้อสอบ ทั้งนี้เพราะว่าข้อสอบแบบอิงเกณฑ์นั้นจะเป็นข้อสอบที่เขียนขึ้นมาตามวัตถุประสงค์เชิงพฤติกรรมของเนื้อหาที่ใช้สำหรับการเรียนการสอนเป็นตอน ๆ ไป

Berk (อ้างถึงใน อนันต์, ม.ป.ป.) ได้เสนอวิธีการหาค่าความยากของข้อสอบ โดยคิดจากเปอร์เซ็นต์ของผู้ที่ตอบถูกในข้อนั้น ซึ่งมีค่าตั้งแต่ 0 ถึง 100 เมื่อนักเรียนตอบถูกทั้งหมดจะมีค่าความยากเท่ากับ 100 แต่ถ้านักเรียนตอบผิดหมดจะมีค่าความยากเท่ากับ 0 อย่างไรก็ตาม ข้อสอบที่มีความยากเหมาะสมคือข้อสอบที่มีความยากในกลุ่มก่อนเรียน แต่ภายในกลุ่มหลังเรียน การหาความยากของข้อสอบแบบอิงเกณฑ์คำนวณได้จากสูตรดังนี้

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

เมื่อ

P = ค่าความยากของข้อสอบ

R = จำนวนนักเรียนที่ตอบถูกในข้อนั้น

N = จำนวนนักเรียนที่เข้าสอบทั้งหมด

2.3.2.2 ค่าอำนาจจำแนก

Hsu (อ้างถึงใน Haladyna, 1974) กล่าวว่าข้อสอบที่ดีตามแนวอิงเกณฑ์ควรมีอำนาจจำแนกดี กล่าวคือ ผู้เรียนที่รอบรู้แล้วควรทำข้อสอบได้ถูกต้อง ส่วนผู้ที่ยังไม่รอบรู้ควรทำผิด

Popham และ Husek (อ้างถึงใน มัทนี, 2526) กล่าวว่า ในการคัดเลือกข้อสอบอิงเกณฑ์ ข้อที่มีค่าอำนาจจำแนกเป็นบวกจะได้รับการคัดเลือกไว้ส่วนข้อที่มีค่าอำนาจจำแนกเป็นลบจะคัดทิ้งไป

เดียนใจ เกตุษา (2525) ได้เสนอว่า ในการคัดเลือกข้อสอบแบบอิงเกณฑ์ โดยพิจารณาผลการวิเคราะห์ข้อสอบจากการหาค่าอำนาจจำแนกของข้อสอบนั้น สามารถถือหลักตามแบบอิงกลุ่มได้ดังตารางที่ 2

ตารางที่ 2 แสดงความหมายของค่าอำนาจจำแนกในระดับต่าง ๆ

ค่าอำนาจจำแนก	ความหมาย
1.00	เป็นข้อสอบที่ดี เป็นไปตามทฤษฎี
.80 - .99	เป็นข้อสอบที่ดี หาได้ในเชิงปฏิบัติ
.20 - .79	เป็นข้อสอบดี พอใช้ได้
.01 - .19	เป็นข้อสอบที่ไม่ดี ควรคัดทิ้ง
-1.00 - .00	เป็นข้อสอบที่ไม่ดีอย่างยิ่ง ต้องคัดทิ้ง

สำหรับการคำนวณหาค่าอำนาจจำแนกของข้อสอบแบบอิงเกณฑ์ได้มีผู้เสนอไว้ดังนี้

Kosecoff และ Klein (อ้างถึงใน อนันต์, ม.ป.ป.) ได้เสนอวิธีการคำนวณค่าอำนาจการจำแนกของข้อสอบแบบอิงเกณฑ์ โดยคำนวณจากสัดส่วนผู้ตอบผิดในการสอบก่อนเรียนแต่ตอบถูกในการสอบหลังเรียนจำนวนผู้เข้าสอบทั้งหมดด้วยสัดส่วนของผู้ตอบผิดในการสอบทั้งสองครั้งกับจำนวนผู้เข้าสอบทั้งหมด ค่าอำนาจจำแนกนี้จะมีค่าอยู่ระหว่าง - 1.00 - 1.00 ซึ่งมีสูตรดังนี้

$$D = \frac{I_c}{N} - \frac{I_i}{N}$$

เมื่อ $D =$ ค่าอำนาจจำแนกของข้อสอบ

$I_c =$ จำนวนนักเรียนที่ตอบผิดในการสอบก่อนเรียนและตอบถูกในการสอบหลังเรียน

I_1 = จำนวนนักเรียนที่ตอบผิดทั้งสองครั้ง

N = จำนวนนักเรียนที่เข้าสอบ

Brennan (อ้างถึงใน อังคณา, 2526) ได้เสนอแนวทางในการหาค่าอำนาจจำแนกโดยใช้การสอบเพียงครั้งเดียว เรียกว่า ดัชนีบี B (B-Index) โดยคำนวณจากผลต่างของสัดส่วนผู้ตอบถูกในกลุ่มผ่านเกณฑ์ และกลุ่มไม่ผ่านเกณฑ์ ซึ่งมีสูตรดังนี้

$$B = \frac{U}{n_1} - \frac{L}{n_2}$$

เมื่อ B = ค่าดัชนีความไว หรือค่าอำนาจจำแนก

U = จำนวนนักเรียนที่ทำข้อสอบข้อนั้นถูก ในกลุ่มที่ผ่านเกณฑ์

L = จำนวนนักเรียนที่ทำข้อสอบข้อนั้นถูก ในกลุ่มที่ไม่ผ่านเกณฑ์

n_1 = จำนวนนักเรียนที่สอบผ่านเกณฑ์

n_2 = จำนวนนักเรียนที่สอบไม่ผ่านเกณฑ์

Kryspin และ Feldhusen (อ้างถึงใน สมศักดิ์, 2525) ได้เสนอวิธีวิเคราะห์ข้อสอบอิงเกณฑ์แต่ละข้อ ซึ่งเรียกว่า ดัชนี S (Sensitivity Index) โดยใช้หลักการหาความแตกต่างของการสอบก่อนสอนและหลังสอน โดยใช้แบบทดสอบชุดเดียวกัน ซึ่งมีสูตรดังนี้

$$S = \frac{R_{\text{pos}} - R_{\text{pre}}}{T}$$

เมื่อ S แทน ดัชนีความไวของการจำแนก (ค่าอำนาจจำแนกของข้อสอบอิงเกณฑ์)

R_{pos} แทน จำนวนที่นักเรียนตอบถูกหลังสอน

R_{pre} แทน จำนวนที่นักเรียนตอบถูกก่อนสอน

T แทน จำนวนนักเรียนที่เข้าสอบ

การวิจัยครั้งนี้ ผู้วิจัยหาค่าอำนาจจำแนกโดยใช้วิธีของ Brennan (อ้างถึงใน อังคณา, 2526) เพราะเหตุว่าทำการสอบเพียงครั้งเดียว โดยคำนวณจากผลต่างของสัดส่วนผู้ตอบถูกในกลุ่มผ่านเกณฑ์ และกลุ่มไม่ผ่านเกณฑ์

2.3.2.3 ความเชื่อมั่น (Reliability)

Berk (อ้างถึงใน บุญเชิด, 2526) ให้ความหมายของความเชื่อมั่นของ แบบทดสอบอิงเกณฑ์ว่า หมายถึงความสอดคล้องในการตัดสินจำแนกผู้รอบรู้ไม่รอบรู้ จากการสอบซ้ำด้วยแบบทดสอบฉบับเดียวกัน หรือแบบทดสอบคู่ขนาน

อังคณา สายยศ (2526) กล่าวว่า ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ เป็นการหาความคงที่ของกลุ่มที่สอบผ่านเกณฑ์ และไม่ผ่านเกณฑ์ในการสอบหลายครั้ง โดยใช้แบบทดสอบฉบับเดียวกัน ดังนั้นความเชื่อมั่นจะขึ้นอยู่กับคะแนนเกณฑ์ของแบบทดสอบฉบับนั้น และถ้าหากหาความเชื่อมั่นโดยวิธีของ Huynh ค่าความเชื่อมั่นจะมีค่าสูงสุดเมื่อคะแนนเกณฑ์อยู่ใกล้คะแนนเฉลี่ยและจะให้ค่าต่ำสุดเมื่อคะแนนเกณฑ์อยู่ใกล้ศูนย์หรือคะแนนเต็ม

ผจงจิต อินทสุวรรณ (2523) ให้ความหมายของความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ว่า เป็นความคงเส้นคงวาในการตัดสิน โดยใช้ข้อสอบคู่ขนานหรือสอบซ้ำ

การคำนวณหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์แบ่งได้เป็นสองลักษณะ ใหญ่ ๆ คือ

- 1) ใช้คะแนนจากผลการสอบสองครั้ง เช่น วิธีของ Swaminathan, Hambleton และ Algina วิธีของ Carver วิธีของ Popham วิธีของ Huynh
- 2) ใช้คะแนนจากผลการสอบครั้งเดียว เช่น วิธีของ Subkoviak วิธีของ Lovett วิธีของ Harris

ส่วนวิธีของ Livingston อาจจัดอยู่ในประเภทใดประเภทหนึ่งก็ได้ (ผจงจิต, 2523) ซึ่งวิธีของ Livingston ได้ปรับปรุงมาจากการหาความเชื่อมั่นแบบอิงกลุ่ม

Swaminathan, Hambleton และ Algina (อ้างถึงใน โกวิท และสมศักดิ์, 2523) ได้เสนอวิธีหาความเชื่อมั่นจากการสอบสองครั้ง และถ้าแบบทดสอบนั้นวัดหลายๆ จุดมุ่งหมายเชิงพฤติกรรมในการหาความเชื่อมั่นควรจะแยกเป็นแบบทดสอบย่อยของแต่ละจุดมุ่งหมาย ดังนั้นค่าความเชื่อมั่นในแบบทดสอบอิงเกณฑ์จึงเป็นค่าความเชื่อมั่นในแต่ละแบบทดสอบย่อย ซึ่งถ้าจัดการเรียนการสอนดี นักเรียนจะรอบรู้มากขึ้น ดังนั้นความเชื่อมั่นไม่ได้อยู่ที่ความยาวของแบบทดสอบเท่านั้น ยังขึ้นอยู่กับประสิทธิภาพการสอนด้วย ในที่นี้จะกล่าวถึงรายละเอียดเฉพาะหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่ใช้ในการสอบครั้งเดียว ได้แก่

Subkoviak (อ้างถึงใน สุโชติ, 2526) เสนอวิธีการหาความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่ใช้การสอบครั้งเดียวโดยใช้แบบทดสอบคู่ขนาน โดยนิยามสัมประสิทธิ์ของความสอดคล้องสำหรับบุคคล ว่าเป็นความน่าจะเป็นที่บุคคลจะถูกกำหนดลงในชั้นเดียวกันของการเรียนรู้ อันเนื่องมาจากผลการสอบโดยใช้แบบทดสอบคู่ขนาน เมื่อใช้คะแนนเกณฑ์สำหรับแบ่งระดับการเรียนรู้ออกเป็น 2 ระดับ คือ ผ่าน ไม่ผ่าน สูตรการหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ของ Subkoviak มีดังนี้ คือ

$$P_C = \frac{\sum_{i=1}^n P_C^{(i)}}{n}$$

$$\text{โดย } P_C^{(i)} = P(X_i \geq C, X'_i \geq C) + P(X_i < C, X'_i < C)$$

เมื่อ	X_i	แทน	คะแนนที่ได้จากการแบบทดสอบฉบับที่ 1
	X'_i	แทน	คะแนนที่ได้จากแบบทดสอบฉบับที่ 2
	C	แทน	คะแนนเกณฑ์
	n	แทน	จำนวนข้อในแบบทดสอบ

วิธีการของ Subkoviak มีข้อตกลง 2 ข้อ คือ

- 1) การกระจายของคะแนนจากแบบทดสอบคู่ขนาน ทั้งสองฉบับของผู้สอบแต่ละคนเป็นอิสระต่อกัน คือประสิทธิภาพในการทำแบบทดสอบฉบับที่ 1 ไม่มีผลต่อการทำแบบทดสอบฉบับที่ 2
- 2) การกระจายของคะแนนจากแบบทดสอบคู่ขนานทั้งสองฉบับของบุคคลหนึ่งเป็นแบบไบโนเมียลเหมือนกันตลอด คือมีการให้คะแนนแบบ 0-1 และความน่าจะเป็นของคำตอบถูกต้องที่ในข้อ

จากข้อตกลงทั้ง 2 ข้างต้น จะเห็นได้ว่า

$$\begin{aligned}
 P_C^{(i)} &= [P(X_i \geq C)]^2 + [P(X_i < C)]^2 \\
 &= [P(X_i \geq C)]^2 + [1 - P(X_i \geq C)]^2 \\
 P(X_i > C) &= \sum_{X_i=C}^n (X_i)^n X_i (1-P_i)^{n-x}
 \end{aligned}$$

ตัวพารามิเตอร์ P_i สามารถประมาณได้จากสมการ $P_i = \frac{\sum X_i}{n}$

เมื่อ X_i แทน จำนวนข้อของการตอบถูก
 n แทน จำนวนข้อในแบบทดสอบ

วิธีการของ Subkoviak นี้จะเห็นได้ว่า เมื่อคะแนนเกณฑ์เปลี่ยนแปลงจะมีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบเปลี่ยนแปลงไปด้วย

Lovett (อ้างถึงใน บุญชม, 2532) ได้เสนอวิธีการหาความเชื่อมั่นจากการทดสอบเพียงครั้งเดียว ซึ่ง Lovett เสนอไว้สองสูตร คือ

ก. สูตรโดยวิธี Binomial

$$r_{cc} = 1 - \frac{k \sum X_i - \sum X_i^2}{(k-1) \sum (X_i - C)^2}$$

เมื่อ r_{cc} แทน สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ
 K แทน จำนวนข้อสอบ
 X_i แทน คะแนนของข้อสอบแต่ละคน
 C แทน คะแนนจุดตัด

ข. สูตรโดยวิธี ANOVA

$$r_{cc} = \frac{(F - 1)}{F}$$

$$\text{เมื่อ } F = \frac{MS_p}{mMsq}$$

$$m = \frac{n(K - 1)}{n(K - 1) - 2}$$

$$MS_p = \frac{(K) \left(\sum \frac{X_i^2}{k} \right) - \frac{(C)^2}{n}}{k}$$

$$MS = \frac{\left(\sum \sum \frac{(X_{ij} - C)^2}{k} - n \sum \frac{(X_j - \bar{X})^2}{k} - k \sum \frac{X_i - \bar{C}}{k} \right)}{(K - 1)(n - 1)}$$

n คือ จำนวนคนเข้าสอบ

k คือ จำนวนข้อสอบ

X_i คือ คะแนนของแต่ละข้อ

X_{ij} คือ คะแนนของคนที่ i ข้อที่ j

Harris (อ้างถึงใน มัทนี, 2526) ได้เสนอแนะวิธีการหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ในรูปของการวิเคราะห์ความแปรปรวนจากผลการสอบเพียงครั้งเดียวซึ่งมีสูตรดังนี้

$$\mu_c^2 = \frac{SS_B}{SS_B + SS_W}$$

เมื่อ SS_B แทน ผลบวกกำลังสองระหว่างกลุ่ม
 SS_W แทน ผลบวกกำลังสองภายในกลุ่ม

Harris กล่าวว่าสำหรับการแจกแจงแบบสมการ ค่า μ_c^2 จะมากที่สุดเมื่อ $C = \mu$
 และ Subkoviak กล่าวว่าวิธีการนี้ค่า μ_c^2 จะมีความไวต่อสัดส่วนของกลุ่มผู้รอบรู้และไม่รอบรู้มาก
 Livingston (1972) ได้เสนอสูตรการหาค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
 โดยอาศัยแนวคิดแบบอิงกลุ่ม โดยมีแนวคิดว่าจะใช้ข้อสอบแบบนี้เมื่อต้องการรู้ว่าคะแนนของ
 นักเรียนคนหนึ่งเบนออกจากเกณฑ์มาตรฐานที่กำหนดไว้เพียงใด โดย Livingston นิยามความ
 เบี่ยงเบนของคะแนนว่าเป็นการเบี่ยงเบนไปจากคะแนนเกณฑ์ ไม่ใช่คะแนนเฉลี่ย สูตร การหาค่า
 ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ของ Livingston คือ

$$r_{cc} = \frac{r_{cc}^2 + (X - C)^2}{\sigma_0^2 + (X - C)^2}$$

เมื่อ r_{cc} แทน ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์
 r_u แทน ความเชื่อมั่นของแบบทดสอบอิงกลุ่ม
 σ_0^2 แทน ความแปรปรวนของคะแนนการสอบ
 X แทน คะแนนเฉลี่ย
 C แทน คะแนนเกณฑ์หรือคะแนนจุดตัด

สูตรของ Livingston นี้ใช้ได้ดีในกรณีที่การกระจายของคะแนนมีฐานนิยมนฐาน
 เดียว แต่ถ้าการกระจายของคะแนนมีสองฐานแล้ว ค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ที่
 คำนวณได้จะมีค่าสูง ซึ่ง Harris (1972) ชี้ให้เห็นว่าความเชื่อมั่นของแบบทดสอบนี้คำนวณโดยใช้
 สูตรของ Livingston จะสูงกว่าที่คำนวณโดยวิธีแบบเก่า แต่ทั้งคู่ก็ให้ค่าความคลาดเคลื่อนในการ
 วัดเท่ากัน ฉะนั้น Harris จึงสรุปว่าค่าความเชื่อมั่นที่สูงกว่าของ Livingston นี้ไม่ได้แสดงถึงความ
 น่าเชื่อถือของการกำหนดว่าคะแนนจริงจะอยู่ใต้หรือเหนือคะแนนจุดตัด (ผจงจิต อินทสุวรรณ,
 2523)

การวิจัยครั้งนี้ผู้วิจัยใช้วิธีการของ Lovett โดยวิธี Binomial หาค่าความเชื่อมั่น
 ของแบบทดสอบที่สร้างขึ้น เพราะเมื่อพิจารณาแล้ว เห็นว่าเหมาะสมกับสภาพการวิจัย คือ ทำ
 การทดสอบเพียงครั้งเดียว และเป็นวิธีการที่ไม่ยุ่งยากในทางปฏิบัติ

2.3.2.4 ความเที่ยงตรง (Validity)

Berk (อ้างถึงในสุโขตี, 2526) ให้ความหมายความเที่ยงตรงของแบบทดสอบอิงเกณฑ์ว่า คือความสอดคล้อง (Consistency) ระหว่างพฤติกรรมที่วัดได้จากข้อสอบแต่ละข้อในแบบทดสอบกับจุดมุ่งหมายที่แบบทดสอบกำหนดให้วัด

อนันต์ (ม.ป.ป.) กล่าวว่า แบบทดสอบอิงเกณฑ์จำเป็นต้องมีความเที่ยงตรงสองลักษณะ คือความเที่ยงตรงเชิงเนื้อหานั่น เป็นการตรวจสอบข้อสอบเหล่านั้นว่า เราพิจารณาลักษณะภายในตัวผู้สอบที่มีความรู้ความสามารถได้ครบถ้วนในจุดประสงค์ที่วัดโดยแบบทดสอบนี้ ซึ่งเรียกว่าผู้รอบรู้ หรือคุณลักษณะที่แสดงถึงการไม่มีคุณสมบัติครบถ้วนในจุดประสงค์ที่กำลังวัด ซึ่งเรียกว่าผู้ไม่รอบรู้ (ส.วาสนา, 2517) ให้ความเห็นว่า ความเที่ยงตรงเชิงเนื้อหาเป็นสิ่งที่สำคัญที่สุดสำหรับแบบทดสอบอิงเกณฑ์

สำหรับการตรวจสอบความเที่ยงตรงเชิงเนื้อหานี้มีผู้เสนอไว้ ดังนี้

Rovinelli และ Hambleton (อ้างถึงใน สงบ, 2523) เสนอวิธีการตรวจสอบความเที่ยงตรงเชิงเนื้อหา โดยกำหนดจุดมุ่งหมายเชิงพฤติกรรม 1 ข้อ และสร้างข้อสอบต่าง ๆ ที่วัดจุดมุ่งหมายเชิงพฤติกรรมนั้นแล้วนำเสนอผู้เชี่ยวชาญหลาย ๆ คน โดยให้ผู้เชี่ยวชาญแต่ละคนให้คะแนนข้อสอบในแต่ละข้อโดยให้คะแนนดังนี้

- +1 หมายถึง แน่ใจว่าข้อสอบนั้นวัดได้ตรงจุดประสงค์นั้น
- 0 หมายถึง ไม่แน่ใจว่าข้อสอบนั้นวัดได้ตรงจุดประสงค์นั้น
- 1 หมายถึง แน่ใจว่าข้อสอบนั้นไม่ได้วัดได้ตรงจุดประสงค์นั้น

แล้วนำคะแนนจากผู้เชี่ยวชาญทุกคนมารวมกันทีละข้อ แล้วหาคะแนนเฉลี่ยหรือค่าดัชนีความสอดคล้อง ถ้าดัชนีความสอดคล้องมีค่ามากกว่า .50 ถือว่า ข้อสอบข้อนั้นวัดจุดประสงค์ที่กำหนดไว้จริง

Cronbach (อ้างถึงในมัทนี, 2526) เสนอวิธีตรวจสอบความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบอิงเกณฑ์โดยการทดลอง โดยให้ผู้เชี่ยวชาญสองกลุ่มที่มีความสามารถทัดเทียมกันเขียนข้อสอบกลุ่มละ 1 ชุด วัดจุดประสงค์เดียวกัน แล้วนำไปสอบกับนักเรียนกลุ่มหนึ่งที่เรียนจุดประสงค์นั้นแล้ว ซึ่งถ้ากำหนดจุดประสงค์อย่างชัดเจนและการสุ่มข้อสอบจากโดเมน ทำได้อย่างเหมาะสมแล้ว ผลการทำข้อสอบทั้ง 2 ชุดของนักเรียนกลุ่มนี้จะต้องสอดคล้องกัน ส่วนข้อที่ไม่สอดคล้องกันจะถูกคัดทิ้งออกไปเนื่องจากขาดคุณสมบัติของความเที่ยงตรงเชิงเนื้อหา

สำหรับการตรวจสอบความเที่ยงตรงเชิงโครงสร้างมีผู้เสนอไว้ ดังนี้

Carver (ล่วน และ อังคณา , ม.ป.ป.) เสนอวิธีหาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบอิงเกณฑ์โดยการหาจากกลุ่มที่เรียนแล้ว และยังไม่ได้เรียนดังตารางที่ 3

ตารางที่ 3 แสดงกลุ่มนักเรียนที่เรียนแล้ว และยังไม่ได้เรียน

	กลุ่มที่ไม่ได้รับการสอน	กลุ่มที่ได้รับการสอน
ผ่าน	b	a
ไม่ผ่าน	c	d

ค่าสัมประสิทธิ์ความเที่ยงตรง = $(a + c)/N$

เมื่อ $N = a + b + c + d$

จะเห็นว่าค่าความเที่ยงตรงวิธีนี้จะมากหรือน้อยขึ้นอยู่กับจำนวนผู้ที่สอบผ่านหลังเรียนกับจำนวนผู้ที่สอบไม่ผ่านก่อนเรียน ดังนั้น ค่าความเที่ยงตรงนี้จะขึ้นอยู่กับประสิทธิภาพในการสอนเป็นสำคัญ

Millman (อ้างถึงใน อังคณา, 2526) เสนอวิธีหาค่าความเที่ยงตรงเชิงโครงสร้างโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบฟี (Phi-Correlation Coefficient) ซึ่งมีสูตรดังนี้

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

เมื่อ ϕ แทน ค่าความเที่ยงตรง
 a แทน จำนวนผู้สอบผ่านหลังเรียน
 b แทน จำนวนผู้สอบผ่านก่อนเรียน
 c แทน จำนวนผู้สอบไม่ผ่านหลังเรียน
 d แทน จำนวนผู้สอบไม่ผ่านก่อนเรียน

Brennan และ Stolurow (1971) ได้เสนอวิธีการตรวจสอบความเที่ยงตรงเชิงโครงสร้างโดยนำข้อสอบไปทดสอบกับเด็กกลุ่มที่ผ่านการเรียนจุดประสงค์นั้นมาแล้วจากนั้นก็คำนวณค่าความยากของข้อสอบแล้วใช้ Cochran's Q test ทดสอบว่าค่าความยากของข้อสอบแต่ละข้อมีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ ข้อใดที่มีค่าความยากเบี่ยงเบนไปจากข้อสอบส่วนใหญ่อย่างมีนัยสำคัญ ถือว่าไม่มีความเที่ยงตรงเชิงโครงสร้างก็จะถูกคัดทิ้งออกไป

การตรวจสอบความเที่ยงตรงเชิงโครงสร้างโดยใช้เทคนิคกลุ่ม ที่รู้คำตอบอยู่แล้ว (Known Group Technique) เป็นการตรวจสอบความเที่ยงตรงตามโครงสร้างวิธีหนึ่ง ซึ่งวิธีนี้จะ

ใช้กลุ่มตัวอย่างที่ทราบแล้วว่ามีรู้ความสามารถในเนื้อหาวิชานั้นมากน้อยเพียงไร เป็นหลัก ในการเปรียบเทียบโดยนำแบบทดสอบไปทดสอบกับกลุ่มนักเรียนที่ยังไม่รอบรู้ แล้วนำผลที่ได้ จากการทดสอบนักเรียนทั้งสองกลุ่มมาเปรียบเทียบกันโดยใช้วิธีการทางสถิติ ในการเปรียบเทียบ โดยใช้สูตร

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

t	แทน Distribution ของ t
$\overline{X}_1$	แทน คะแนนเฉลี่ยของกลุ่มนักเรียนที่รอบรู้
$\overline{X}_2$	แทน คะแนนเฉลี่ยของกลุ่มนักเรียนที่ไม่รอบรู้
S_1^2	แทน ความแปรปรวนของกลุ่มนักเรียนที่รอบรู้
S_2^2	แทน ความแปรปรวนของกลุ่มนักเรียนที่ไม่รอบรู้
n_1	แทน จำนวนนักเรียนที่รอบรู้
n_2	แทน จำนวนนักเรียนที่ไม่รอบรู้

คะแนนเฉลี่ยของกลุ่มนักเรียนที่รอบรู้จะต้องมีค่าสูงกว่าคะแนนเฉลี่ยของกลุ่มนักเรียนที่ยังไม่รอบรู้ อย่างมีนัยสำคัญทางสถิติ จึงจะถือว่าแบบทดสอบมีความเที่ยงตรงเชิงโครงสร้าง (สุรินทร์, ม.ป.ป.) การวิจัยครั้งนี้ผู้วิจัยหาค่าความเที่ยงตรงเชิงเนื้อหาโดยใช้วิธีของ Rovinelli และ Hambleton โดยให้ผู้เชี่ยวชาญในด้านเนื้อหาวิชาพิจารณาตรวจสอบแบบทดสอบแต่ละข้อที่สร้างขึ้นกับจุดประสงค์การเรียนรู้ เพราะเป็นวิธีที่สอดคล้องกับกระบวนการเรียนการสอน

2.3.2.5 การกำหนดคะแนนเกณฑ์หรือคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์

แนวคิดของแบบทดสอบอิงเกณฑ์นั้น เป็นแบบทดสอบที่มีการกำหนดเกณฑ์ หรือ จุดตัด (Cut - off Score) ขึ้นไว้ เพื่อใช้ในการจำแนกผู้สอบออกเป็นสองส่วน คือ ผู้สอบที่ได้คะแนนสูงกว่าเกณฑ์ จะถือว่าเป็นผู้รอบรู้ (Master) และผู้ที่ได้คะแนนต่ำกว่าเกณฑ์ จะถือว่าเป็นผู้ไม่รอบรู้ (Nonmaster)

การกำหนดเกณฑ์นับได้ว่าเป็นส่วนสำคัญของแบบทดสอบอิงเกณฑ์ ได้มี ผู้นำเสนอแนวคิด เทคนิค และวิธีการต่าง ๆ ในการกำหนดเกณฑ์ไว้จำนวนมาก พอสรุปได้ดังนี้

มิลแมน (Millman, 1975) ได้สรุปแนวคิดของการกำหนดเกณฑ์ไว้ว่าควรคำนึงถึงสิ่งต่าง ๆ ต่อไปนี้

1) ระดับความสามารถในการตอบแบบทดสอบของคนอื่น ๆ (Performance of Others) โดยเฉพาะผลการสอบของกลุ่มที่ผ่านการยอมรับแล้ว ยื่อนำมาพิจารณาในการกำหนดเกณฑ์ได้

2) เนื้อหาสาระของข้อสอบในแบบทดสอบ (Item Content) ควรพิจารณาตัดสินใจว่าปัญหาค้าง ๆ หรือจุดมุ่งหมายที่นำมาทำการทดสอบนั้น ผู้สอบที่มีความสามารถปานกลาง ควรทำได้ถูกต้องสักเท่าใด จึงจะอยู่ในระดับที่พอใจ

3) คำนึงถึงผลกระทบที่ตามมา (Educational Consequences) บางครั้งควรคำนึงว่าการกำหนดเกณฑ์สูงเกินไป อาจทำให้ผู้เรียนไม่ผ่านโดยไม่จำเป็นแต่ถ้าหากว่าต่ำเกินไป ก็จะทำให้มาตรฐานการศึกษาต่ำ หรือทำให้ผู้เรียนเกิดปัญหาในการเรียนต่อไป

4) ผลทางจิตวิทยาและการลงทุนทางการศึกษา (Psychological and Financial Cost) การกำหนดเกณฑ์ควรคำนึงถึงผลทางจิตวิทยาที่เกิดขึ้นกับผู้สอบ เช่น แรงจูงใจ ความเบื่อหน่าย การทำลายอัธมโนทัศน์ และรวมทั้งค่าใช้จ่ายในการสอนซ่อมเสริมผู้เรียนที่สอบไม่ผ่าน

5) ความคลาดเคลื่อนในการเดาและการสุ่มข้อสอบ (Error due to Chance and Item Sampling) การกำหนดเกณฑ์ควรคำนึงถึง โอกาสที่ผู้สอบจะเดาถูกรวมทั้งลักษณะการสุ่มข้อสอบจากประชากรข้อสอบในขอบเขตเนื้อหาที่กำหนดขึ้นด้วย

การกำหนดเกณฑ์ของแบบทดสอบอิงเกณฑ์ โดยมีผู้เสนอวิธีการต่าง ๆ ไว้อย่างกว้างขวาง กลาส (Glass, 1978) ได้รวบรวมวิธีการหาเกณฑ์มาตรฐานได้ 6 วิธี คือ

1) การพิจารณาความสามารถของผู้อื่น (Performance of Others as a Criterion) วิธีนี้ไม่แตกต่างไปจากแนวคิดของการประเมินแบบอิงเกณฑ์มาตรฐาน

2) การนับถอยหลัง (Counting Backward from 100%) วิธีนี้จะให้ผู้เชี่ยวชาญทางด้านเนื้อหาวิชา พิจารณาตามความสำคัญของจุดประสงค์และกำหนดค่าร้อยละตามลำดับความสำคัญ

3) การใช้เกณฑ์ภายนอกเข้ามาช่วยในการตัดสินใจ (Bootstrapping on Others Criterion Scores) วิธีนี้ใช้ได้ก็ต่อเมื่อ ตัวเกณฑ์ภายนอกกับคะแนนจากแบบทดสอบอิงเกณฑ์มีความสัมพันธ์กันอย่างสมบูรณ์ มิฉะนั้นจะเกิดความคลาดเคลื่อนได้

4) การกำหนดสมรรถภาพต่ำสุด (Judging Minimal Competence) วิธีนี้ใช้ผู้เชี่ยวชาญเป็นผู้ร่วมในการกำหนดเกณฑ์มาตรฐานต่ำสุดที่ยอมรับได้ เช่น วิธีการกำหนดเกณฑ์มาตรฐานของ อีเบล แองกอฟฟ์ และนี่เคลลี

5) การใช้ทฤษฎีการตัดสินใจประกอบ (Decision Theoretic Approach) วิธีนี้นับว่าเหมาะสมการทางสถิติเพื่อนำมาใช้เป็นวิธีการในการกำหนดเกณฑ์ เช่นการกำหนดเกณฑ์ที่

จะทำให้เกิดความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ต่ำสุด ได้แก่ วิธีการของกลาส (Glass, 1978) และวิธีการของเบิร์ก (Berk, 1976)

6) การใช้วิธีการวิจัยเชิงปฏิบัติการ (Operation Research Method) วิธีการนี้เป็นการนำเอาเกณฑ์ภายนอกที่เหมาะสมสัมพันธ์กับคะแนนที่ได้จากแบบทดสอบ เช่น การพิจารณาถึงทัศนคติต่อวิชาที่เรียน ซึ่งจะผันแปรไปตามเกณฑ์ของแบบทดสอบ

ในการกำหนดเกณฑ์หรือคะแนนจุดตัดในปัจจุบันยังไม่มีวิธีใดที่ดีที่สุด คงจะต้องทำเป็นเฉพาะกรณีไป โดยคำนึงถึงข้อมูลหลายด้าน และเมื่อได้เกณฑ์มาแล้ว ก็ควรจะใช้เกณฑ์นั้นอย่างมีความเป็นปรนัย การวิจัยครั้งนี้ ผู้วิจัยได้ใช้การหาคะแนนเกณฑ์หรือคะแนนจุดตัดโดยวิธีการของอีเบล (Ebel) ซึ่งเป็นการกำหนดเกณฑ์จากสมรรถภาพขั้นต่ำ เป็นการให้ผู้เชี่ยวชาญด้านเนื้อหาวิชาหรือครูประจำวิชา ศึกษาข้อสอบแล้วระบุว่า นักเรียนที่มีสมรรถภาพขั้นต่ำต้องสอบได้คะแนนเท่ากับเท่าใด และเป็นวิธีที่เหมาะสมเนื่องจากการพิจารณารายละเอียดรัดกุมกว่าวิธีการอื่น ๆ โดยอาศัยความคิดเห็นจากผู้เชี่ยวชาญหลาย ๆ คนมาเฉลี่ยทำให้ผลที่ได้น่าเชื่อถือยิ่งขึ้น Ebel (อ้างถึงใน บุญเชิด, 2527) หลักการของเทคนิคนี้อาศัยการวิเคราะห์ลักษณะข้อสอบ และระดับความยากของข้อสอบร่วมกัน โดยให้ครูประจำวิชาวิเคราะห์เทียบกับเปอร์เซ็นต์การสอบผ่านที่คาดหวัง จากลักษณะข้อสอบและความยากของข้อสอบ ซึ่งอีเบลได้ทำการจำแนกไว้แล้วคือ ลักษณะข้อสอบที่ตรงกับประเด็นปัญหาจำแนกเป็น 4 ระดับหรือกลุ่ม คือ

- ระดับที่ตรงกับจุดประสงค์และจำเป็นมาก
- ระดับที่ตรงกับจุดประสงค์และสำคัญ
- ระดับที่ตรงกับจุดประสงค์
- ระดับที่ไม่แน่ใจว่าตรงกับจุดประสงค์

ส่วนความยากของข้อสอบจำแนกเป็น 3 ระดับ คือ ระดับง่าย ระดับปานกลาง และระดับยากซึ่งข้อสอบที่มีลักษณะต่าง ๆ กันและความยากต่างกัน จะมีเปอร์เซ็นต์การสอบผ่านต่างกันไปดังตารางที่ 4

ตารางที่ 4 เปอร์เซนต์การสอบผ่านที่คาดหวังจากลักษณะข้อสอบ และระดับความยากของข้อสอบ เป็นดังนี้

ลักษณะข้อสอบ	ระดับความยาก		
	ง่าย	ปานกลาง	ยาก
ตรงกับจุดประสงค์และจำเป็นอย่างมาก	100%	-	-
ตรงกับจุดประสงค์และสำคัญ	90%	70%	-
ยอมรับว่าตรงกับจุดประสงค์	80%	60%	30%
ไม่แน่ใจว่าตรงกับจุดประสงค์	70%	50%	40%

วิธีการหาคะแนนจุดตัด มีลำดับขั้นดังนี้

- 1) นำข้อสอบที่ต้องการวิเคราะห์ไปให้ครูประจำวิชาพิจารณาว่า ข้อสอบมีลักษณะตรงกับจุดประสงค์มากน้อยเพียงไร มีระดับความยากง่ายเพียงไร
- 2) จำแนกข้อสอบตามลักษณะและความยาก โดยแจกแจงเป็นความถี่รวม
- 3) กำหนดเปอร์เซนต์การสอบผ่าน ตามลักษณะข้อสอบและความยากของข้อสอบ
- 4) คูณความถี่จำนวนข้อกับเปอร์เซนต์การสอบผ่านที่คาดหวัง และรวมผลคูณทั้งหมด
- 5) หาค่าเปอร์เซนต์การสอบผ่านที่คาดหวังโดยเฉลี่ย
- 6) กำหนดคะแนนจุดตัดจากค่าเปอร์เซนต์การสอบผ่านที่คาดหวังโดยเฉลี่ยจากขั้นที่ 5

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 งานวิจัยเกี่ยวกับการสร้างแบบทดสอบโดยใช้เทคโนโลยีการเขียนข้อสอบ

ในปี 1966 โฮฟลี และคณะ ซึ่งเป็นคณะอาจารย์ในมหาวิทยาลัยมินเนโซตา ได้เริ่มโครงการ MINNEMAST ซึ่งเป็นโครงการทดสอบอิงโดเมนในวิชาวิทยาศาสตร์ และ คณิตศาสตร์ ในระดับชั้นอนุบาลจนถึงชั้นประถมศึกษาปีที่ 6 ในมลรัฐมินเนโซตา ซึ่งจุดมุ่งหมายโครงการเป็นไปเพื่อการประเมินผลที่มีต่อผู้เรียนโดยเฉพาะ และพัฒนาแบบทดสอบอิงโดเมนเพื่อวัดผลสัมฤทธิ์ในการกำหนดโดเมนจะกำหนดเป็นโดเมนย่อย ๆ ที่ประกอบกันขึ้นเป็นโดเมนใหญ่โดยใช้วิธีอุปนัย (Induction) ซึ่งมีหลักการว่า พฤติกรรมที่แสดงออกถึงการเรียนรู้ตามที่นักวิชาการของสาขา

วิชาการไว้อย่างแน่นอนและชัดเจนแล้ว สามารถนำมาจัดเข้าเป็นกลุ่มหรือโดเมนของข้อสอบได้ และในการกำหนดโดเมนให้มีความเป็นตัวแทนที่ดีและวิธีการพัฒนาข้อสอบด้วยนั้น ไฮฟลิได้ใช้เทคนิครูปแบบฟอร์มข้อสอบ (Item Form) ซึ่งระบบทดสอบที่ใช้ อิงโดเมนเป็นรายบุคคล และในการสอบไม่มีการกำหนดเกณฑ์ผ่าน ทั้งนี้เพราะโปรแกรมอยู่ในระยะทดลองเพื่อการปรับปรุงให้ดีขึ้น

กาญจนา วัฒนสุนทร (2522) ได้ศึกษาเรื่องการสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ เรื่อง สมการ ชั้นมัธยมศึกษาปีที่ 1 โดยกำหนดฟอร์มข้อสอบ (Item Form) ทำการวิเคราะห์เนื้อเรื่องที่จะสร้างข้อสอบจากหลักสูตร นำมาสร้างโดเมนข้อสอบที่จะวัดลักษณะของข้อหัวเรื่อง 6 หัวข้อแล้วสร้างประชากรเนื้อเรื่องที่มุ่งวัดในแต่ละโดเมน โดเมนละ 40 ข้อ หากความเที่ยงตรงของเนื้อเรื่อง โดยใช้ผู้เชี่ยวชาญตัดสิน เขียนวัตถุประสงค์เชิงพฤติกรรม เพื่อนำมา สร้างตารางวิเคราะห์หลักสูตร เพื่อกำหนดจำนวนข้อสอบ แล้วสุ่มข้อสอบจากโดเมนของข้อสอบแต่ละโดเมนเพื่อนำมาใส่ในแบบทดสอบย่อย 4 ชุด คือ ความรู้เบื้องต้นเกี่ยวกับสมการแทนค่าสมการ การแก้ปัญหสมการโดยใช้คุณสมบัติการบวก ลบ คูณ และหาร ความรู้ในเรื่องการเปลี่ยนโจทย์สมการเป็นสมการ และความรู้เกี่ยวกับการแก้โจทย์สมการ กลุ่มตัวอย่างเป็นนักเรียนโรงเรียนสันติราษฎร์ จำนวน 156 คน วิเคราะห์หาคุณภาพรายข้อ โดยการหาค่าความยากและค่าดัชนีความไว หาค่าความเที่ยงตรงโดยวิธีของคาร์เวอร์ หาค่าความเชื่อมั่นโดยวิธีการของลิฟวิงตัน ผลการทดสอบค่าเฉลี่ยของคะแนนหลังการเรียนและก่อนการเรียนที่ได้จากการทดสอบด้วยแบบทดสอบย่อย 4 ชุด แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

พิศิษฐ์ ตันจาวนิช (2528) ได้ศึกษาเรื่อง การประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นเพื่อกำหนดความยาวของแบบทดสอบอิงเกณฑ์ โดยมุ่งตรวจสอบหาค่าความยากของแบบทดสอบที่เกิดความคลาดเคลื่อนในการจัดประเภทของบุคคลตามระดับการรอบรู้ไม่แตกต่างไปจากอัตราความคลาดเคลื่อนที่กำหนดที่ระดับ .05 ขอบเขตของเนื้อหาที่สร้างในหัวข้อ "การวิเคราะห์ข้อสอบ" ใช้วิธีการสร้างข้อสอบโดยการกำหนดภาระงานสอบเป็น "รูปแบบเฉพาะประจำข้อสอบ" จำนวน 14 รูปแบบ ให้ผู้มีประสบการณ์ในการสอน 18 คน ตรวจสอบความสอดคล้องระหว่างจุดประสงค์ ขอบเขตของเนื้อหาและรูปแบบเฉพาะประจำข้อสอบ จากนั้น ได้สร้างข้อสอบจากรูปแบบเฉพาะประจำข้อสอบ รูปแบบละ 5 ข้อ และสร้างข้อสอบจากรูปแบบเฉพาะประจำข้อที่ 6 และ 7 สุ่มแทรกเพิ่มเข้าไปอีกรูปแบบละ 1 ข้อ รวมเป็นแบบทดสอบเต็มฉบับ 72 ข้อ กลุ่มตัวอย่างเป็นนักศึกษาระดับประกาศนียบัตรวิชาการศึกษาระดับชั้นสูง จำนวน 282 คน แบบทดสอบที่พัฒนาขึ้นมีค่าดัชนีความไวของข้อสอบไม่ต่ำกว่า .20 เป็นแบบทดสอบที่มี

ความตรงเชิงจำแนก หากค่าความเชื่อมั่นด้วยวิธีการของสับโคเวียคมีค่าระหว่าง 0.90-0.97 การวิเคราะห์ข้อมูลมุ่งตรวจสอบอัตราความคลาดเคลื่อนที่ กำหนดโดยการทดสอบทวินาม เพื่อใช้คะแนนจากแบบทดสอบเต็มฉบับ เป็นตัวเกณฑ์กำหนดความถูกต้องในการจัดประเภทบุคคลตามระดับการรอบรู้ และคะแนนจากแบบทดสอบที่มีช่วงความยาวเป็น 10 ข้อ 14 ข้อ 20 ข้อ 25 ข้อ และ 30 ข้อเป็นเกณฑ์และหาค่าสหสัมพันธ์เคนคอลลี ทอ ระหว่างผลการจัดประเภทบุคคลตามระดับการรอบรู้ จากแบบทดสอบเต็มฉบับ และแบบทดสอบในช่วงความยาวต่าง ๆ กัน พบว่าจำนวนข้อสอบที่มีความเหมาะสมในการจัดสอบ เพื่อจำแนกบุคคลตามระดับการรอบรู้ คือ จำนวน 20 ข้อ และมีความสัมพันธ์กับแบบทดสอบเต็มฉบับอยู่ในช่วง 0.53 - 0.89

รุ่งนภา ตระกูลพัฑ (2528) ได้ศึกษาเปรียบเทียบคุณภาพของแบบทดสอบวิชาคณิตศาสตร์ ที่สร้างตามลักษณะเฉพาะของข้อสอบ โดยครูผู้สร้างข้อสอบมีสถานภาพต่างกัน กลุ่มตัวอย่างที่ใช้ในการศึกษามี 2 กลุ่ม คือ กลุ่มครูผู้สร้างข้อสอบและกลุ่มนักเรียน โดยกลุ่มครูผู้สร้างข้อสอบเป็นครูที่สอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 4 จำนวน 16 คน ส่วนกลุ่มตัวอย่างนักเรียนเป็นนักเรียนชั้นประถมศึกษาปีที่ 4 จำนวน 432 คน ผลการวิจัยพบว่าแบบทดสอบวิชาคณิตศาสตร์ที่สร้างตามลักษณะเฉพาะของข้อสอบ โดยกลุ่มครูผู้สร้างข้อสอบทั้ง 4 กลุ่ม (กลุ่มครูวิชาคณิตศาสตร์ที่ไม่ได้รับการอบรมกลุ่มครูวิชาเอกอื่น ๆ ที่ไม่ได้รับการอบรม กลุ่มครูวิชาเอก คณิตศาสตร์ที่ได้รับการอบรม กลุ่มวิชาเอกอื่น ๆ ที่ได้รับการอบรม) เมื่อเปรียบเทียบคุณภาพของแบบทดสอบทั้ง 4 ฉบับ พบว่า คุณภาพของแบบทดสอบไม่แตกต่างกัน ทั้งในด้านค่าความเชื่อมั่น ความเที่ยงตรง ความยากง่าย อำนาจจำแนกและประสิทธิภาพของตัวเลือก

เคือนเพ็ญ หวานณรงค์ (2528) ได้ศึกษาการสร้างแบบทดสอบวินิจัยทักษะพื้นฐานของความเข้าใจภาษาอังกฤษ สำหรับแบบทดสอบที่สร้างขึ้นนี้มีจำนวน 7 ฉบับ โดยใช้การกำหนดลักษณะของแบบทดสอบ (Test Specification) ของโปแฟม (Popham, 1978: 121-129) ผลจากการศึกษาพบว่า แบบทดสอบวินิจัยทั้ง 7 ฉบับ มีค่าความยากของ ข้อสอบตั้งแต่ 0.50 ถึง .95 ค่าอำนาจจำแนกของข้อสอบตั้งแต่ .90 ถึง 1.00 ค่าความเชื่อมั่นของแบบทดสอบมีค่าเรียงลำดับดังนี้ .9443, .5368, .8596, .7891, .7984, .8080 และ .7918 สำหรับค่าความเที่ยงตรงของแบบทดสอบหาโดยวิธีพิจารณาจากผู้เชี่ยวชาญทางด้านเนื้อหา เพื่อตัดสินความสอดคล้องระหว่าง .50 ถึง 1.00 แสดงว่าแบบทดสอบวินิจัยทั้ง 7 ฉบับ มีความเที่ยงตรงเชิงเนื้อหา

ดาวรุ่ง วีรกุล (2529) ได้ทำการศึกษาเปรียบเทียบคุณภาพของแบบทดสอบชั้นมัธยมศึกษาปีที่ 4 ที่สร้างตามลักษณะเฉพาะของข้อสอบ โดยผู้ออกข้อสอบมีสถานภาพ

ต่างกัน พบว่าลักษณะของข้อสอบที่กำหนดให้ระบุคำอธิบายชัดเจนดี และทำให้ครูสามารถออกข้อสอบได้มีคุณภาพเหมือนกับวัดพฤติกรรมเดียวกัน อย่างไรก็ตามผู้ออกข้อสอบบางคนรู้สึกว่าการอิสระในการออกข้อสอบ

2.4.2 งานวิจัยที่เกี่ยวข้องกับการสร้างแบบทดสอบคู่ขนาน

มนต์ชัย สังฆพันธ์ (2535) ได้ศึกษาเรื่องการใช้รูปแบบฟาเซทสร้างแบบทดสอบคู่ขนาน วิชาคณิตศาสตร์ เรื่อง ทศนิยม ชั้นประถมศึกษาปีที่ 6 ให้มีคู่ขนานกันสองฉบับ จำนวน 3 ชุด คือ ชุดที่ 1 วัดทักษะการคิดคำนวณ ชุดที่ 2 วัดความเข้าใจ และชุดที่ 3 วัดการแก้โจทย์ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2534 ของโรงเรียน ในสังกัดสำนักงานคณะกรรมการประถมศึกษา อำเภอเมืองขอนแก่น จังหวัดขอนแก่น จำนวน 377 คน ผลการวิจัยพบว่า

1) รูปแบบฟาเซทเป็นเทคโนโลยีการเขียนข้อสอบที่สามารถนำมาใช้สร้างแบบทดสอบให้มีลักษณะคู่ขนานได้ ทั้งในด้านเนื้อหาสาระของข้อสอบและคุณลักษณะทางสถิติตามเกณฑ์มาตรฐานของความเป็นแบบทดสอบคู่ขนาน กล่าวคือ แบบทดสอบคู่ขนาน ฉบับที่ 1 และ ฉบับที่ 2 ที่ใช้วัดความสามารถทางคณิตศาสตร์ ทั้ง 3 ชุด มีค่าความยากมาตรฐานเฉลี่ย ค่าอำนาจจำแนก ค่าคะแนนเฉลี่ยค่าความแปรปรวน และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบทดสอบคู่ขนานสองฉบับอยู่ในเกณฑ์สูงทั้ง 3 ชุด

2) แบบทดสอบคู่ขนานทั้ง 6 ฉบับ มีความเที่ยงตรงเชิงเนื้อหาและมีคุณภาพพอที่จะใช้ประเมินผลสัมฤทธิ์ทางการเรียนเรื่องทศนิยมได้

สุภาพรณ โลหะการก (2534) ได้ศึกษาเรื่อง การประยุกต์แนวคิดทฤษฎีรูปแบบฟาเซท ในการสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนในสังกัดสำนักงานการศึกษา จังหวัดอุทัยธานี ปีการศึกษา 2532 จำนวน 640 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบอิงเกณฑ์ ชั้นประถมศึกษาปีที่ 6 เรื่อง รูปทรงและปริมาตรและแบบทดสอบคู่ขนานอีก 4 ฉบับ ดำเนินการเก็บรวบรวมข้อมูลโดยทดสอบก่อนเรียนและหลังเรียนด้วยแบบทดสอบอิงเกณฑ์ เว้นระยะ 1 สัปดาห์ ทำการทดลองซ้ำด้วยแบบทดสอบคู่ขนาน นำผลการสอบมาวิเคราะห์หาคุณภาพของเครื่องมือ

ผลการวิจัยพบว่า แนวคิดทฤษฎีรูปแบบฟาเซทสามารถใช้สร้างแบบทดสอบอิงเกณฑ์ วิชาคณิตศาสตร์ได้อย่างมีคุณภาพ โดยมีความตรงเชิงเนื้อหา มีค่าความยากอยู่ระหว่าง .20 ถึง .89 มีค่าดัชนีความไวอยู่ระหว่าง .13 ถึง .43 ค่าสัมประสิทธิ์ความตรงในการตัดสินใจเท่ากับ .75 นอกจากนี้ยังพบว่า แนวคิดดังกล่าวสามารถใช้สร้างแบบทดสอบคู่ขนานได้หลายฉบับด้วยวิชา

การที่ง่าย ๆ นี้ ทั้งนี้แบบทดสอบคู่ขนาน 4 ฉบับ มีค่าความยากอยู่ระหว่าง .20 ถึง .93, .20 ถึง .96, .31 ถึง .94 และ .21 ถึง .93 ตามลำดับ มีค่าสัมประสิทธิ์ความเชื่อมั่นแบบความสอดคล้องภายใน เท่ากับ .91, .91, .91 และ .89 ตามลำดับ และมีค่าความตรงเชิงเกณฑ์สัมพันธ์กับแบบทดสอบอิง เกณฑ์ฉบับแรกเท่ากับ .75, .75, .76 และ .76 ตามลำดับ ในการประยุกต์ทฤษฎีรูปแบบฟาเซทในการเขียนข้อสอบนั้นพบว่ามีข้อจำกัดในขั้นตอนการสร้างรูปแบบฟาเซทเนื่องจากใช้เวลานาน แต่เมื่อสร้าง รูปแบบฟาเซทได้แล้วจะมีข้อดี คือสามารถสร้างข้อสอบที่มีคุณภาพได้จำนวนมาก และสร้างแบบทดสอบคู่ขนานได้ง่ายและรวดเร็ว

2.4.3 งานวิจัยเกี่ยวกับการสร้างแบบทดสอบ และการวิเคราะห์หาคุณภาพของแบบทดสอบที่เขียนจากเทคโนโลยีการเขียนข้อสอบที่เรียกว่ารูปแบบฟาเซท

รอยด์ และฮาလာไดนา (Roid and Haladyna, 1982 : 116) ได้ศึกษาเกี่ยวกับคุณภาพของแบบทดสอบที่กำหนดให้ครูจำนวน 3 คน ได้รับความรู้เกี่ยวกับการเขียนข้อสอบ โดยให้รูปแบบฟาเซทเป็นผู้สร้างแบบทดสอบ ซึ่งมีสมมติฐานในการศึกษาครั้งนี้ คือ

- 1) ผู้เขียนข้อสอบที่ได้รับความรู้เกี่ยวกับรูปแบบฟาเซท จะเขียนข้อสอบ ให้มีค่าความยากเทียบเท่ากัน
- 2) ผู้เขียนข้อสอบมีความรู้ความเข้าใจเกี่ยวกับรูปแบบฟาเซทเป็นอย่างดี จะสามารถสร้างตัวลองของแบบทดสอบแบบเลือกตอบ ให้สามารถวินิจฉัยความบกพร่องของผู้เรียน ได้ดีกว่าผู้ออกข้อสอบที่มีความรู้ความเข้าใจเกี่ยวกับรูปแบบฟาเซทเพียงเล็กน้อย ผลการศึกษาพบว่ากลุ่มผู้ที่ได้รับความรู้เกี่ยวกับรูปแบบฟาเซทเป็นอย่างดีเหมือนกันสร้างข้อสอบ แล้วพบว่าค่าความยากของ ข้อสอบในกลุ่มนี้มีค่าไม่แตกต่างกัน ซึ่งสอดคล้องกับสมมติฐานที่ตั้งขึ้นและเมื่อนำข้อสอบที่เขียนขึ้นจากสมาชิกของฟาเซทเดียวกันของผู้ออกข้อสอบ ที่มีความเข้าใจในรูปแบบฟาเซทเป็นอย่างดี และผู้ออกข้อสอบที่มีความรู้ความเข้าใจในรูปแบบฟาเซทเพียงเล็กน้อยมาเปรียบเทียบกันพบว่า มีค่าความยากแตกต่างกัน นอกจากนั้น ผลการศึกษายังชี้ให้เห็นว่าเมื่อ ผู้ออกข้อสอบมีความรู้ต่างกันหรือผู้ออกข้อสอบมีอิสระในการเลือกคำในการเขียนข้อสอบแล้วจะได้ข้อสอบที่มีค่าความยากแตกต่างกัน

สุนทรี ไกรกาบแก้ว (2532) ได้ศึกษาคุณภาพของแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชนิดโจทย์ปัญหา และโจทย์ทักษะการคิดคำนวณของครูผู้เขียนข้อสอบที่มีลักษณะต่างกันในด้านเพศ วุฒิ และประสบการณ์ ที่เขียนข้อสอบจากลักษณะเฉพาะของแบบทดสอบที่ใช้ในรูปแบบฟาเซทในการเขียนข้อคำถามและตัวเลือก กลุ่มตัวอย่างเป็นครู 16 คน และนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 312 คน ผลการวิจัยพบว่า ค่าความยากก่อนสอนและหลังสอน

ค่าอำนาจจำแนก และค่าความเชื่อมั่นของแบบทดสอบวิชาคณิตศาสตร์ชนิดโจทย์ทักษะ
 จำนวนที่เขียนข้อสอบโดยครูที่มีลักษณะด้านเพศ วุฒิ และประสบการณ์แตกต่างกันอย่างไม่มีนัย
 สำคัญทางสถิติที่ระดับ .05

จากเอกสารและงานวิจัยดังที่กล่าวมาแล้ว จะเห็นได้ว่างานวิจัยส่วนใหญ่ได้มีการใช้
 เทคโนโลยีการเขียนข้อสอบรูปแบบใดรูปแบบหนึ่งมากำหนดโครงสร้างหรือลักษณะของข้อสอบ
 เช่น ฟอรมข้อสอบ (Item Form) การกำหนดลักษณะเฉพาะของข้อสอบ (Test Specification) และ
 รูปแบบฟาเซท (Facet Design) ซึ่งแบบทดสอบที่ได้จะมีลักษณะเป็นคู่ขนานกัน และส่วนมากจะ
 เพื่อศึกษาเปรียบเทียบลักษณะของครูในการเขียนข้อสอบมิได้เน้นในประเด็นขั้นตอนการสร้างและ
 พัฒนาแบบทดสอบ ตลอดจนความเหมาะสมในการนำไปใช้ในสถานการณ์ต่าง ๆ เพื่อแก้ปัญหา
 ประโยชน์และข้อจำกัดที่พบจากการใช้เทคโนโลยี การสร้างข้อสอบเป็นประเด็นที่ผู้วิจัยสนใจ จึง
 ได้นำมาศึกษาต่อดังนี้

1. ใช้เทคโนโลยีการสร้างข้อสอบที่เรียกว่า รูปแบบฟาเซท เพื่อกำหนดโครงสร้างของ
 ข้อสอบสร้างแบบทดสอบคู่ขนาน
2. กำหนดลักษณะเฉพาะของข้อสอบจากรูปแบบฟาเซท
3. การออกข้อสอบจะยึดวัตถุประสงค์เป็นหลัก เพราะเหตุว่าผู้วิจัยมีความประสงค์ที่จะให้
 ได้ข้อสอบที่สามารถอ้างอิงได้ว่า ผู้สอบผ่านเกณฑ์หรือผ่านจุดประสงค์หรือไม่
4. ตรวจสอบค่าสถิติความเป็นคู่ขนานของแบบทดสอบ ดังนี้
 - 1) ทดสอบความแตกต่างของคะแนนเฉลี่ย
 - 2) ทดสอบความแตกต่างของค่าความแปรปรวน
 - 3) ทดสอบความแตกต่างของค่าความขาคเฉลี่ย
 - 4) ทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ความเชื่อมั่น