

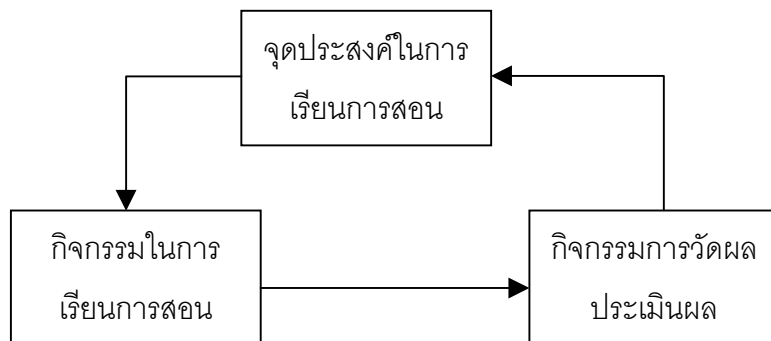


การวิเคราะห์แบบทดสอบ

บทนำ

การจัด กระบวนการเรียนรู้ หรือ กระบวนการเรียนการสอน จะประกอบด้วยองค์ประกอบหลัก 3 ประการเสมอ ได้แก่

- 1.1 จุดประสงค์ในการเรียนการสอน ได้แก่พฤติกรรมที่ต้องการ หรือพฤติกรรมที่คาดหวังให้เกิดการเปลี่ยนแปลงขึ้นที่ผู้เรียน
- 1.2 กิจกรรมการเรียนการสอนหรือกิจกรรมการเรียนรู้ ได้แก่กิจกรรมต่าง ๆ ที่จัดขึ้น เพื่อให้ผู้เรียนเกิดการเปลี่ยนแปลงพฤติกรรมให้เป็นไปตามจุดประสงค์ที่กำหนดไว้
- 1.3 การวัดผลประเมินผล ได้แก่กิจกรรมต่าง ๆ ที่จัดขึ้น เพื่อตรวจสอบผู้เรียนว่าเกิดพฤติกรรมตามที่กำหนดไว้มากน้อยเพียงใด



กิจกรรมการวัดผลประเมินผล เป็นกิจกรรมที่สำคัญไม่น้อยและเป็นกิจกรรมหลักที่จะส่งผลต่อคุณภาพการศึกษาโดยตรง เพราะว่าการวัดผลประเมินผลเป็นกระบวนการรวบรวมข้อมูลเกี่ยวกับผู้เรียนด้านต่าง ๆ ตามจุดประสงค์ที่กำหนดไว้ในหลักสูตร วิชา และรายวิชาต่าง ๆ แล้วนำข้อมูลเหล่านั้นมาวิเคราะห์ ตัดสินใจ เช่น การปรับปรุงการเรียนการสอน การสอนซ่อมเสริม การให้เรียนซ้ำ การจบหลักสูตร หรือการตัดสินใจเพื่อสิทธิต่าง ๆ ของผู้เรียนหรือผู้เข้ารับการวัดผล ประเมินผล

การวัดผล มีเครื่องมือที่สำคัญ คือ แบบทดสอบ หรือ แบบสอบวัด หรือ เรียกกันทั่วไปว่า ข้อสอบ ซึ่งมีหลายประเภท แล้วแต่ผู้สอบวัดจะเลือกใช้ เพื่อให้เหมาะสม

สอดคล้องกับจุดประสงค์ที่ต้องการสอบวัดและเครื่องมือสอบวัดทุกชนิดจะเก็บข้อมูลออกมาเป็น ตัวเลขหรือ คะแนน ทั้งสิ้น แล้วนำข้อมูลมาแปรผล ประเมิน และตัดสินใจ โดยทั่วไปข้อสอบ หรือแบบทดสอบประเภทต่าง ๆ มีลักษณะดังต่อไปนี้

1. แบบทดสอบ สำหรับสอบวัดด้านความรู้ ความคิด (Cognitive Domain)
ส่วนมากมักจะมีรูปแบบ หรือลักษณะดังต่อไปนี้

1.1 แบบอัตนัย (Essay Item)

1.1.1 แบบตอบสั้น

1.1.2 แบบตอบยาว

1.2 แบบปรนัย

1.2.1 แบบเลือกตอบ (Multiple Choice Item)

1.2.2 แบบถูก ผิด (True-False Item)

1.2.3 แบบจับคู่ (Matching Item)

1.2.3 แบบเติมคำ (Completion Item)

2. แบบทดสอบ วัดด้านความรู้สึก (Affective Domain)

2.1 แบบสังเกต

2.2 การบรรยายความรู้สึกเกี่ยวกับตนเอง

2.3 การตั้งคำถาม

2.4 การจัดอันดับคุณภาพ (Rating scale)

3. แบบทดสอบ วัดด้านการปฏิบัติ (Psychomotor Domain)

3.1 แบบสำรวจรายการ (Checklist)

3.2 การจัดอันดับคุณภาพ (Rating Scale)

จากประสบการณ์ของผู้เขียนซึ่งเกี่ยวข้องอยู่กับการนิเทศ อบรม และพัฒนา งานวัดผลประเมินผลในโรงเรียนมัธยมศึกษามาตั้งแต่ปี พ.ศ.2524-2540 พบว่า เครื่องมือวัดผลประเมินผลในโรงเรียนมัธยมศึกษา ร้อยละ 97.00 เป็นข้อสอบปรนัย แบบเลือกตอบ(Multiple Choice Item) ที่เหลือ เป็นข้อสอบแบบอื่น ๆ

หลังจากปี พ.ศ.2540 เป็นต้นมา มีการเปลี่ยนแปลง พัฒนาการวัดผล ประเมินผล โดยการเน้น การประเมินผลตามสภาพจริง มากขึ้น ทำให้มีการพัฒนา เครื่องมือวัดที่หลากหลายขึ้น เช่น มีการสอบวัดโดยข้อสอบแบบอัตนัย แบบสังเกต ภาควิปฏิบัติ และแบบสอบถามแบบมาตราประมาณค่า ซึ่งมีสัดส่วนมากขึ้นในปัจจุบัน

การกำหนดให้คะแนนแบบทดสอบหรือแบบสอบวัดแต่ละข้อ อาจจะจำแนกได้เป็น 3 ประเภท คือ

1. แบบ 0-1 (Zero-One Method) แบบสอบวัดประเภทนี้ แต่ละข้อจะกำหนดคะแนนเป็น 0 และ 1 ถ้าตอบผิด ได้ 0 คะแนน และ ถ้าตอบถูก ได้ 1 คะแนน ได้แก่ข้อสอบแบบปรนัย โดยเฉพาะอย่างยิ่งแบบเลือกตอบ (Multiple Choice Item) แบบถูก-ผิด เป็นต้น

2. แบบไม่ใช่ 0-1 (Non Zero-One Method) ซึ่ง คะแนนแต่ละข้อ ไม่ใช่ 0-1 ซึ่งแต่ละข้อจะกำหนดคะแนนเต็มไม่เท่ากัน เช่น 5 คะแนน 8 คะแนน 10 คะแนน เป็นต้น แบบสอบวัดประเภทนี้ได้แก่ข้อสอบแบบอัตนัย (Essay Item) เป็นต้น

3. แบบมาตราประมาณค่า (Rating Scale) แบบสอบวัดแบบนี้ มีการกำหนดให้คะแนนแต่ละข้อเป็นมาตราประมาณค่าคงที่เท่ากัน เช่น คะแนนแต่ละข้อกำหนดไว้เป็น 5 ระดับ คือ 1 2 3 4 และ 5 เป็นต้น การให้คะแนน จะขึ้นอยู่กับตามสิ่งที่สังเกตได้ หรือ ความรู้สึกที่มีอยู่ ถ้ามีมากที่สุด จะได้ 5 คะแนน และถ้ามีน้อยที่สุด ได้ 1 คะแนน เป็นต้น แบบสอบวัดประเภทนี้ได้แก่ แบบสอบถาม แบบสังเกต แบบวัดเจตคติ เป็นต้น

กระบวนการวัดผลประเมินผล มีความสำคัญไม่น้อยไปกว่ากระบวนการจัดการเรียนรู้หรือกระบวนการเรียนการสอน หรือบางทีอาจจะมีความสำคัญมากกว่า เนื่องจากกระบวนการวัดผลประเมินผลมีการตัดสินใจ กำหนดค่า ให้แก่ผู้ถูกวัดและประเมิน ซึ่งส่งผลต่อการตัดสินใจอนาคตของผู้ถูกประเมิน ถ้าการสอบวัดได้ข้อมูลที่มีความถูกต้องย่อมเป็นผลดีต่อผู้ถูกประเมิน แต่ถ้าข้อมูลไม่ถูกต้อง ย่อมเป็นผลเสียอย่างร้ายแรงอย่างยิ่งต่อผู้ถูกประเมิน

เครื่องมือสำคัญของการวัดผลคือ ข้อสอบ หรือ แบบสอบวัดประเภทต่าง ๆ ที่กล่าวมาแล้ว จากการศึกษาของหน่วยศึกษานิเทศก์ กรมสามัญศึกษา (หน่วยศึกษา นิเทศก์ กรมสามัญศึกษา, 2525 : 174) ซึ่งได้ศึกษาวิเคราะห์จากกลุ่มตัวอย่างโรงเรียนมัธยมศึกษาทั้งส่วนกลางและส่วนภูมิภาค จำนวน 603 โรงเรียน พบว่า สัดส่วนของคะแนนจากข้อสอบแบบอัตนัยต่อข้อสอบแบบปรนัย เป็น 1 ต่อ 6.09 หรือคะแนนร้อยละ 83.58 เป็นคะแนนจากข้อสอบแบบปรนัยชนิดต่าง ๆ และจากการวิเคราะห์ต่อไปยังพบว่าร้อยละ 91.00 ของข้อสอบแบบปรนัยทั้งหมด เป็นข้อสอบ ปรนัยแบบเลือกตอบ (Multiple Choice) ซึ่งในเรื่องเดียวกันนี้ ผู้เขียนได้ศึกษาประเภทและชนิดของข้อสอบวิชาสามัญประจำภาคเรียนต่าง ๆ ที่ใช้ในโรงเรียนมัธยมศึกษาปีการศึกษา 2536 จำนวน 5 โรงเรียน ทั้งโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก พบว่าคะแนน

สอบเป็นคะแนนจากแบบทดสอบแบบปรนัย เฉลี่ยร้อยละ 95.50 นอกนั้นเป็นข้อสอบแบบอัตนัย และข้อสอบแบบปรนัย ร้อยละ 97.00 เป็นข้อสอบแบบเลือกตอบ ซึ่งอาจกล่าวได้ว่า ครู โดยทั่ว ๆ ไปนิยมใช้ข้อสอบแบบปรนัยมากกว่าแบบอัตนัย โดยเฉพาะอย่างยิ่งนิยมข้อสอบปรนัยแบบเลือกตอบมากที่สุด

สิ่งที่สำคัญที่สุดจากการศึกษาครั้งนี้ พบว่าข้อสอบปรนัยแบบเลือกตอบนั้น มีความบกพร่องถึงร้อยละ 21.88 หรือประมาณ 1 ใน 5 ของข้อสอบแบบปรนัยแบบเลือกตอบทั้งหมด ซึ่งข้อบกพร่องต่าง ๆ ได้แก่ข้อคำถามหรือความนำบกพร่องร้อยละ 10.63 ตัวเลือกบกพร่องร้อยละ 8.35 และบกพร่องเพราะสาเหตุอื่น ๆ อีกร้อยละ 2.90 นอกจากนี้ ข้อสอบที่ครูใช้อยู่ยังไม่เป็นมาตรฐานเดียวกัน และเกณฑ์การผ่านก็แตกต่างกัน

ด้วยเหตุนี้ ผู้เขียนและคณะศึกษานิเทศก์ ในหน่วยศึกษานิเทศก์ กรมสามัญศึกษา เขตการศึกษา 8 จึงได้ดำเนินการทุกวิถีทางที่ให้ครู อาจารย์ในโรงเรียนมัธยมศึกษา กรมสามัญศึกษา ได้มีการพัฒนาเครื่องมือวัดผลให้มีคุณภาพ เพื่อผลการวัดที่ถูกต้อง

ผู้เขียนได้พัฒนาโปรแกรมตรวจและวิเคราะห์ข้อสอบแบบอิงเกณฑ์โดยวิธีหาค่าดัชนีจำแนก B (B-Index) ออกเผยแพร่อย่างเป็นทางการ ตั้งแต่ปี พ.ศ.2536 เป็นต้นมา ปรากฏว่าได้รับความนิยมมาก มีการนำไปใช้ทั่วประเทศ และผู้เขียนได้มีการพัฒนาปรับปรุงโดยตลอด ปัจจุบันเกือบจะเรียกได้ว่า เป็นความต้องการจำเป็นในกระบวนการพัฒนาการเรียนการสอน การพัฒนางานวิชาการทั่วไปเลยทีเดียว

เนื่องจากเทคโนโลยีคอมพิวเตอร์เปลี่ยนแปลงไปมากประกอบกับแนวคิดเรื่องการวัดผลเปลี่ยนแปลงไป ครู อาจารย์ มีการใช้เครื่องมือวัดผลที่หลากหลายขึ้น ผู้เขียนจึงได้มีการพัฒนาโปรแกรมรุ่นใหม่ให้สอดคล้องกับเทคโนโลยีและความต้องการของครู อาจารย์ ครอบคลุมกระบวนการวัดผลในโรงเรียนมากขึ้น ผู้เขียนได้พัฒนาขีดความสามารถของโปรแกรมให้สามารถวิเคราะห์และพัฒนาเครื่องมือวัดผล 3 ประเภท ในโปรแกรมเดียวกัน ตามชนิดของเครื่องมือวัดที่ใช้กันมากที่สุด ดังนี้

1. ตรวจและวิเคราะห์ข้อสอบแบบอิงเกณฑ์โดยวิธีหาค่าดัชนีจำแนก B (B-Index) ซึ่งเป็นข้อสอบแบบเลือกตอบ (Multiple Choice Item)
2. วิเคราะห์ข้อสอบแบบอัตนัย (Essay Item) ซึ่งมีการให้คะแนนไม่ใช่แบบ 0-1 (Non Zero-One Method)
3. วิเคราะห์แบบสอบถามแบบ Rating Scale

แนวคิด หลักการ การวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบ เป็นเทคนิคในการปรับปรุงคุณภาพของข้อสอบให้ดีขึ้น (อนันต์ ศรีโสภา, 2525 : 205) โดยทั่วไปมีจุดประสงค์ในการปรับปรุงเกี่ยวกับ

1. ความกำกวมของภาษาในข้อสอบ
2. การเฉลยผิด
3. ความยากและง่ายเกินไปของข้อสอบ
4. ข้อที่ไม่สามารถจำแนกความสามารถของนักเรียนให้เห็น ได้ชัดเจน

นอกจากนี้ การวิเคราะห์ข้อสอบ ยังมีจุดมุ่งหมายในการ ช่วยปรับปรุง เทคนิคการสอนของครู ช่วยในการค้นพบข้อบกพร่องต่าง ๆ ในการเรียนของ นักเรียนโดยพิจารณาว่านักเรียนยังอ่อนในเนื้อหาวิชาใด จุดประสงค์ใด ครูควร จะสอนซ่อมเสริมในจุดประสงค์ใดบ้าง ดังนั้น การวิเคราะห์ข้อสอบ จึงเป็น กระบวนการที่ ส่งผลถึงคุณภาพการศึกษาโดยตรง

การวิเคราะห์ข้อสอบช่วยให้ครูมีความเข้าใจเกี่ยวกับจุดมุ่งหมายของการสอน และพฤติกรรมการเรียนรู้ของนักเรียนและการวิเคราะห์ข้อสอบยังช่วยประหยัดเวลา ในการสร้างข้อสอบที่ดีอีกด้วย จากการศึกษาของ Lange et al (อ้างอิงใน อนันต์ ศรี โสภา, 2525 : 206) พบว่าต้องใช้เวลาถึง 5 เท่าในการสร้างข้อสอบแบบเลือกตอบ (Multiple Choice Item) ใหม่ แทนที่จะนำข้อสอบเหล่านั้นไปทำการวิเคราะห์เพื่อ ปรับปรุงแก้ไขเพียงบางส่วนซึ่งเสียเวลาน้อยกว่ามาก

แนวคิดในการการวิเคราะห์ข้อสอบ มี 2 แบบ คือ การวิเคราะห์ข้อสอบแบบ อิงกลุ่ม (Norm Referenced Test) และการวิเคราะห์ แบบอิงเกณฑ์ (Criterion Referenced Test)

การวิเคราะห์ข้อสอบแบบอิงเกณฑ์ (Criterion Referenced Test) เป็นการแปล ความหมายของคะแนนของผู้เข้าสอบแต่ละคนเทียบกับคะแนนเกณฑ์มาตรฐานที่ กำหนดไว้ (โกวิท ประวาลพุกษ์, 2527 :10) ถ้าผู้เข้าสอบได้คะแนนถึงเกณฑ์ที่ กำหนดไว้หรือสูงกว่า แสดงว่าผู้เรียนมีความสามารถในเรื่องนั้น ๆ หรือจุดประสงค์ นั้น ๆ แล้ว และถ้าได้คะแนนต่ำกว่าเกณฑ์ที่กำหนดไว้ แสดงว่าผู้เรียนยังไม่มี ความสามารถในเรื่องนั้น ๆ หรือจุดประสงค์นั้น ๆ

ดังนั้นการนิยามความหมายของผู้ที่ผ่านเกณฑ์ (กลุ่มสูง) และ ไม่ผ่านเกณฑ์ (กลุ่มต่ำ) จึงใช้เกณฑ์ ที่กำหนดไว้เป็นจุดตัด

การวิเคราะห์แบบอิงกลุ่ม (Norm Referenced Test) การนิยามความหมายของผู้ที่ผ่าน(กลุ่มสูง) และ ไม่ผ่าน(กลุ่มต่ำ) จะใช้คะแนนเฉลี่ยของกลุ่มเป็นเกณฑ์ตัดสิน กลุ่มที่ได้คะแนนสูงกว่าคะแนนเฉลี่ยเรียกว่า กลุ่มสูง กลุ่มที่ได้คะแนนต่ำกว่าคะแนนเฉลี่ยเรียกว่า กลุ่มต่ำ

ในการวิเคราะห์แบบทดสอบทั่วไปโดยทั่วไป จะมีการเรียงคะแนนจากสูงสุดไปถึงต่ำสุด คัดเลือกมาเพียง 27% จากคะแนนสูงสุดลงมา กลุ่มนี้เรียกว่ากลุ่มสูงและคัดเลือกเพียง 27% จากคะแนนต่ำสุดขึ้นไป กลุ่มนี้เรียกว่ากลุ่มต่ำ เพราะส่วนที่เหลือถึงแม้นำมาวิเคราะห์ ผลที่ได้จะไม่มี ความแตกต่างกัน ทำให้ประหยัดเวลาในการวิเคราะห์

การวิเคราะห์ข้อสอบทั้งแบบอิงกลุ่มและแบบอิงเกณฑ์เป็นการวิเคราะห์ เพื่อหาคุณภาพ 2 ประการ คือ

1. ความเที่ยงตรง (Validity) ได้แก่

1.1 ความเที่ยงตรงเชิงเนื้อหา (Content Validity)

1.2 ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ซึ่งได้แก่คุณภาพของข้อสอบ เป็นรายชื่อ

1.2.1 ค่าความยากของข้อสอบ (Item Difficulty)

1.2.2 ค่าอำนาจจำแนกของข้อสอบ (Discriminating Power)

2. ความเชื่อมั่น (Reliability) ซึ่งเป็นคุณภาพของข้อสอบทั้งฉบับ

การวิเคราะห์ข้อสอบที่ผู้เขียนเลือกใช้ คือ การวิเคราะห์ข้อสอบแบบเลือกตอบ การวิเคราะห์ข้อสอบแบบอัตนัย และ การวิเคราะห์แบบสอบถามแบบ Rating Scale จะมีการวิเคราะห์หาคุณภาพดังที่กล่าวมา

สำหรับการวิเคราะห์ความเที่ยงตรงตามเนื้อหาไม่ได้นำมาผนวกไว้ในโปรแกรมวิเคราะห์ที่ได้พัฒนาขึ้นเนื่องจากไม่มีความจำเป็นต้องใช้คอมพิวเตอร์ช่วยในการวิเคราะห์ เพียงแต่ให้ผู้เชี่ยวชาญด้านเนื้อหาและผู้เชี่ยวชาญด้านการวัดผล ประเมินผลร่วมกันวิเคราะห์เท่านั้น แต่ผู้เขียนได้นำวิธีการที่ง่ายและนิยมใช้กันมานานำเสนอเพื่อเป็นแนวทางปฏิบัติก่อนที่จะนำแบบสอบวัดหรือข้อสอบที่สร้างขึ้นและ นำไปพัฒนาต่อไป

การวิเคราะห์ความเที่ยงตรงตามเนื้อหา (Content Validity)

การวิเคราะห์ความเที่ยงตรงตามเนื้อหา (Content Validity) เป็นขั้นตอนสำคัญที่การพัฒนาข้อสอบทุกประเภทจะต้องดำเนินการก่อนการนำข้อสอบไปทดลองสอบ เพื่อการวิเคราะห์หาคุณภาพอื่น ๆ วิธีการวิเคราะห์ความเที่ยงตรงตามเนื้อหาส่วนมากนิยมดำเนินการตามหลักการและวิธีการต่อไปนี้

1. อาศัยดุลพินิจของผู้เชี่ยวชาญเนื้อหาวิชา
2. อาศัยเทคนิคการตรวจสอบจากการทดลองหรือ เทคนิคเชิงประจักษ์

(Empirical Techniques)

การวิเคราะห์ความเที่ยงตรงเชิงเนื้อหาโดยอาศัยดุลพินิจของผู้เชี่ยวชาญเนื้อหาวิชา เป็นผู้พิจารณาว่า ข้อสอบกับจุดประสงค์เชิงพฤติกรรม หรือ ลักษณะเฉพาะของมวลความรู้ที่ต้องการวัดมีความสอดคล้องกันหรือไม่ การวิเคราะห์ความเที่ยงตรงโดยอาศัยดุลพินิจของผู้เชี่ยวชาญเนื้อหาวิชาตัดสินนี้ โรวินेलลี และ แฮมเบิลตัน

(Hambleton et al. 1978 : 34 citing Rovinelli and Hambleton. 1977)

(อ้างอิงใน บุญเชิด ภิญโญนนตพงษ์, 2527 : 68) ได้เสนอไว้ 3 วิธี ดังนี้

1. ใช้ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์
2. ใช้ดัชนีความเหมาะสมระหว่างข้อสอบกับจุดประสงค์
3. ใช้ดัชนีการจับคู่ระหว่างข้อสอบกับจุดประสงค์

1. การวิเคราะห์หาค่าดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์

คำนวณได้จากสูตร

$$IOC = \frac{\sum R}{N}$$

IOC คือ ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์

R คือ ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชาทั้งหมด

N คือ จำนวนผู้เชี่ยวชาญเนื้อหาวิชา

วิธีดำเนินการวิเคราะห์ มีดังนี้

1. นำจุดประสงค์เชิงพฤติกรรม และข้อสอบที่สอบวัดจุดประสงค์นั้น ๆ ไปให้ผู้เชี่ยวชาญเนื้อหาวิชา อย่างน้อย 3 คน แต่ละคนพิจารณาถึงความเห็นว่าเป็นข้อสอบ แต่ละข้อวัดวัตถุประสงค์เชิงพฤติกรรมที่ต้องการวัดหรือไม่ โดยกำหนดคะแนน ความคิดเห็นดังนี้

- +1 หมายถึง แนใจว่าข้อสอบวัดจุดประสงค์ข้อนั้น
- 0 หมายถึง ไม่แนใจว่าข้อสอบวัดจุดประสงค์ข้อนั้นหรือไม่
- 1 หมายถึง แนใจว่าข้อสอบไม่วัดจุดประสงค์ข้อนั้น

2. บันทึกผลการพิจารณาของผู้เชี่ยวชาญเนื้อหาวิชาแต่ละคนในแต่ละข้อ หากคะแนนผลรวมรายข้อแล้วแทนค่าในสูตร

$$IOC = \frac{\sum R}{N}$$

- 3. กำหนดคะแนนจุดตัด ซึ่งคะแนนจุดตัดของทุกข้อจะเท่ากับ 0.5
- 4. แปลความหมาย ดัชนีความสอดคล้องระหว่างข้อสอบ กับ จุดประสงค์ ถ้าค่าดัชนีความสอดคล้อง (IOC) ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับ 0.5 แสดงว่า ข้อสอบข้อนั้น เป็นตัวแทนของจุดประสงค์เชิงพฤติกรรมข้อนั้น ถ้าดัชนีความสอดคล้อง ที่คำนวณได้ต่ำกว่า 0.5 แสดงว่าข้อสอบไม่วัด หรือไม่ตัวแทนของจุดประสงค์เชิงพฤติกรรมข้อนั้น
- 5. คัดเลือกข้อสอบที่มีค่าดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ ตามเกณฑ์ที่กำหนดไว้ใช้ต่อไป

2. การวิเคราะห์หาค่าดัชนีความเหมาะสมระหว่างข้อสอบกับจุดประสงค์
คำนวณได้จากสูตรค่าเฉลี่ยของคะแนนและค่าความแปรปรวนของคะแนน ดังนี้

$$1. \quad \bar{X} = \frac{\sum X}{N}$$

$$2. \quad s = \sqrt{\frac{[N \sum X^2 - (\sum X)^2]}{N(N-1)}}$$

\bar{X} คือ ค่าเฉลี่ยคะแนนความคิดเห็น

s คือ ค่าความเบี่ยงเบนมาตรฐานของคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชา

N คือ จำนวนผู้เชี่ยวชาญเนื้อหาวิชา

ค่าเฉลี่ยคะแนนความคิดเห็นเป็นดัชนีบ่งชี้ความเหมาะสมและค่าความแปรปรวนของคะแนนความคิดเห็นจะเป็นดัชนีบ่งชี้มติความคิดเห็นว่าเป็นเอกฉันท์หรือมีความขัดแย้งกัน

วิธีดำเนินการวิเคราะห์ มีดังนี้

1. นำจุดประสงค์เชิงพฤติกรรมและข้อสอบที่วัดจุดประสงค์ข้อนั้น ๆ ไปให้ผู้เชี่ยวชาญเนื้อหาวิชาแต่ละวิชาพิจารณาว่า ข้อสอบเหมาะสมกับจุดประสงค์เชิงพฤติกรรมมากน้อยเพียงใด โดยอาศัยมาตราส่วนประเมินดังนี้

4 หมายถึง เหมาะสมมากที่สุด

3 หมายถึง เหมาะสมมาก

2 หมายถึง เหมาะสมปานกลาง

1 หมายถึง เหมาะสมน้อย

0 หมายถึง เหมาะสมน้อยที่สุดหรือไม่เหมาะสมเลย

2. บันทึกผลการลงความเห็นของผู้เชี่ยวชาญเนื้อหาวิชาแต่ละคนเป็นรายข้อ แล้วคำนวณค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของคะแนนความคิดเห็นตามสูตรในสมการ 1 และ 2 ตามลำดับ

3. กำหนดความหมายค่าเฉลี่ยของคะแนนความคิดเห็นดังนี้

ค่าเฉลี่ย	ข้อสอบมีความเหมาะสมกับจุดประสงค์
3.5 - 4.0	มากที่สุด
2.5 - 3.4	มาก
1.5 - 2.4	ปานกลาง
0.5 - 1.4	น้อย
0.0 - 0.4	น้อยที่สุด

สำหรับค่าเบี่ยงเบนมาตรฐาน ถ้ามีค่าเป็นศูนย์ แสดงว่า การลงความคิดเห็น มีมติเป็นเอกฉันท์ แต่ถ้ามีค่ามากขึ้นเท่าไรแสดงว่า การลงความคิดเห็นมีความขัดแย้งกัน ไม่ตรงกันมากขึ้นเท่านั้น ดังนั้น จากมาตราส่วนประมาณค่า 5 ระดับนี้ ค่าส่วนเบี่ยงเบนมาตรฐานไม่เกิน 1.00 แสดงว่าการลงความคิดเห็นมีมติเป็นเอกฉันท์

4. คัดเลือกข้อสอบที่มีค่าเฉลี่ย ตั้งแต่ 2.5-4.0 และต้องมีค่าส่วนเบี่ยงเบนมาตรฐานไม่เกิน 1.00 ไว้ ส่วนข้อสอบข้อที่ไม่เป็นไปตามเกณฑ์ให้คัดออกหรือปรับปรุงแก้ไขใหม่

ตัวอย่างผลการวิเคราะห์ข้อสอบจำนวน 5 ข้อ ซึ่งวัดจุดประสงค์เชิงพฤติกรรม ข้อ 1 ไปให้ครูผู้เชี่ยวชาญเนื้อหาวิชา จำนวน 5 คน พิจารณาให้ความ

คิดเห็น ว่า ข้อสอบแต่ละข้อมีความเหมาะสมกับจุดประสงค์เชิงพฤติกรรมที่ต้องการวัดข้อนั้น มากน้อยเพียงใด ผลการวิเคราะห์แสดงดังตารางต่อไปนี้

จุดประสงค์ข้อที่	ข้อสอบข้อที่	คะแนนความคิดเห็น					\bar{X}	s
		คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5		
1	1	0	1	2	3	4	2.0	1.58
	2	4	4	4	3	3	3.6	0.55
	3	4	4	4	4	4	4.0	0.00
	4	0	0	1	2	2	1.0	1.00
	5	1	1	1	1	1	1.0	0.00

จากตาราง ปรากฏว่า ข้อ 2 และข้อ 3 มีค่าเฉลี่ยคะแนน เป็น 3.6 และ 4.0 แสดงว่าข้อสอบทั้ง 2 ข้อ มีความเหมาะสมมากกับจุดประสงค์ข้อ 1 และมีส่วนค่าเบี่ยงเบนมาตรฐานของคะแนนเป็น 0.55 และ 0.00 แสดงว่าการลงความคิดเห็นมีมติเป็นเอกฉันท์ ดังนั้น จึงเลือกข้อสอบทั้ง 2 ข้อไว้ใช้ ส่วนข้อ 1 ข้อ 4 และ ข้อ 5 มีความเหมาะสมต่ำกว่าเกณฑ์ที่ต้องการ

3. การวิเคราะห์โดยใช้ดัชนีการจับคู่ระหว่างข้อสอบกับจุดประสงค์ จำนวนได้จากสูตร ดังนี้

$$M = [f / N] \times 100$$

M คือ เปอร์เซนต์ความถี่ของการจับคู่แทนดัชนีการจับคู่

f คือ จำนวนความถี่ของผู้เชี่ยวชาญเนื้อหาวิชาที่ลง

ความคิดเห็นว่าข้อสอบจับคู่กับจุดประสงค์

N คือ จำนวนผู้เชี่ยวชาญเนื้อหาวิชา

วิธีดำเนินการวิเคราะห์ดำเนินการตามขั้นตอนต่อไปนี้

1. นำจุดประสงค์เชิงพฤติกรรมชุดหนึ่ง(มากกว่า 1 จุดประสงค์) และข้อสอบชุดหนึ่งซึ่งวัดจุดประสงค์เชิงพฤติกรรมข้อใดข้อหนึ่งไปให้ผู้เชี่ยวชาญเนื้อหาวิชาพิจารณาลงความคิดเห็นว่าข้อสอบข้อใดวัดจุดประสงค์เชิงพฤติกรรมข้อใดให้จับคู่กัน

2. นำผลการจับคู่ข้อสอบกับจุดประสงค์เชิงพฤติกรรมของผู้เชี่ยวชาญเนื้อหาวิชาแต่ละคนมาแจกแจงความถี่ลงในตารางที่จำแนกตามข้อสอบและจุดประสงค์

3. นำความถี่ของผู้เชี่ยวชาญเนื้อหาวิชาที่จับคู่ข้อสอบกับจุดประสงค์เชิงพฤติกรรมมาคำนวณค่าไคกำลังสอง (χ^2) เพื่อทดสอบความเป็นอิสระแก่กันของข้อมูล อาจใช้วิธีที่ง่ายและสะดวกคือสังเกตจากร้อยละของความถี่ที่สามารถบ่งชี้ได้

4. กำหนดเกณฑ์ของเปอร์เซ็นต์ความถี่ เพื่อจะยอมรับว่าข้อสอบวัดพฤติกรรมตรงตามที่ระบุไว้ในจุดประสงค์ ในที่นี้กำหนดเกณฑ์เท่ากับ 80%

5. คัดเลือกข้อที่มีเปอร์เซ็นต์ความถี่ตั้งแต่ 80% ขึ้นไป แสดงว่าข้อสอบข้อนั้นวัดพฤติกรรมตรงตามที่ระบุไว้ในจุดประสงค์ข้อที่จับคู่จริง ส่วนข้อที่มีเปอร์เซ็นต์ความถี่น้อยกว่า 80% จะถูกกำจัดออกไปหรือนำไปปรับปรุงแก้ไขต่อไป

ตัวอย่างการคำนวณ

นำข้อสอบ 4 ข้อ และจุดประสงค์เชิงพฤติกรรม 2 ข้อ ไปให้ผู้เชี่ยวชาญเนื้อหาวิชา จำนวน 10 คน พิจารณาลงความคิดเห็นว่า ข้อสอบข้อใดวัดพฤติกรรมตามจุดประสงค์ข้อใดบ้าง ผลการพิจารณาปรากฏตามที่ยบันทึกในตารางคำนวณ ต่อไปนี้

ข้อสอบ	จุดประสงค์			
	ข้อ 1		ข้อ 2	
	1	2	1	2
1	10	100	0	0
2	8	80	0	0
3	0	0	10	100
4	4	40	6	60

จากตาราง จะเห็นว่า

ข้อสอบข้อ 1 กับจุดประสงค์ข้อ 1 มีดัชนีการจับคู่ 100% แสดงว่า ผู้เชี่ยวชาญเนื้อหาวิชา มีมติเป็นเอกฉันท์ว่า ข้อสอบข้อ 1 วัดพฤติกรรมตรงตามที่ระบุไว้ในจุดประสงค์ ข้อ 1

ข้อสอบข้อ 2 กับจุดประสงค์ข้อ 1 มีดัชนีการจับคู่ 80% แสดงว่า ผู้เชี่ยวชาญเนื้อหาวิชาส่วนมากมีความคิดเห็นตรงกันว่า ข้อสอบข้อ 2 วัดพฤติกรรมตรงตามที่ระบุไว้ใน จุดประสงค์ ข้อ 1

ข้อสอบข้อ 3 กับจุดประสงค์ข้อ 2 มีดัชนีการจับคู่ 100% แสดงว่าผู้เชี่ยวชาญ เนื้อหาวิชา มีมติเป็นเอกฉันท์ว่า ข้อสอบข้อ 3 วัดพฤติกรรมตามที่ระบุไว้ใน จุดประสงค์ ข้อ 2

ข้อสอบ ข้อ 4 กับจุดประสงค์ทั้ง 2 ข้อ มีดัชนีการจับคู่ต่ำกว่า 80% แสดงว่า ข้อสอบข้อ 4 ไม่วัดพฤติกรรมที่ระบุไว้ในจุดประสงค์ทั้งข้อ 1 และ ข้อ 2

ดังนั้น จึงเลือกข้อสอบข้อ 1 ข้อ 2 และข้อ 3 ไว้ ส่วน ข้อ 4 กำจัดออกไปหรือ ปรับปรุงแก้ไขต่อไป

การวิเคราะห์ความเที่ยงตรงตามโครงสร้าง (Construct Validity)

ความเที่ยงตรงเชิงโครงสร้าง เป็นความเที่ยงตรงของแบบทดสอบหรือแบบ สอบวัด ที่สามารถวัดได้ตรงตาม เจตนา ของการวัด หรือสะท้อนสิ่งที่ต้องการวัดออกมาได้อย่างถูกต้อง เช่น ต้องการวัดผู้ที่มีความสามารถในการคิด และ ผู้ที่ยังไม่มีความ สามารถในการคิด แบบสอบวัดสามารถจำแนกออกมาได้ หรือ ต้องการสอบวัด เจตคติต่อเรื่องใดเรื่องหนึ่ง แบบสอบวัดสามารถแยกผู้ที่ถูกสอบวัดออกมาได้อย่าง ถูกต้อง เป็นต้น

การวิเคราะห์ความเที่ยงตรงตามโครงสร้าง จะเป็นการวิเคราะห์หาค่า อำนาจ จำแนก (Discriminating Power) และ ค่าความยาก (Item Difficulty) ของข้อสอบ

1. การวิเคราะห์ค่าอำนาจจำแนก (Discriminating Power)

การวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ มีวิธีวิเคราะห์หลายวิธีแต่ในที่นี้ จะขอเสนอวิธีการวิเคราะห์โดยวิธีหาค่าดัชนีอำนาจ B (B-Index) สำหรับวิเคราะห์ ข้อสอบแบบเลือกตอบ (Multiple Choice Item)

วิธีการวิเคราะห์วิธีนี้ โดยทั่ว ๆ ไปเรียกว่า วิธีหาค่าดัชนีอำนาจ B (B-Index) ซึ่งเป็นวิธีหนึ่งในการวิเคราะห์ความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ

ค่าดัชนีอำนาจ B (B-Index) คือตัวเลขที่บ่งชี้ว่าแบบทดสอบมีความสามารถ จำแนกผู้รอบรู้หรือผู้ที่ผ่านเกณฑ์และผู้ไม่รอบรู้หรือผู้ที่ไม่ผ่านเกณฑ์ของแต่ละ จุด ประสงค์ได้มากน้อยเพียงใด

วิธีนี้ Brennan (1972) ได้พัฒนาสูตรขึ้นโดยอาศัยแนวคิดของ Johnson (1951) ซึ่งเรียกว่า Discrimination Index B (อ้างอิงใน โกวิท ประมวลพฤษ์ และ สมศักดิ์ สันทรเวชญ์, 2523 : 286) ซึ่งมีรายละเอียดดังนี้

$$B = \frac{U(n2 - L) - L(n1 - U)}{n1n2}$$

$$B = \frac{Un2 - Ln1}{n1n2}$$

$$B = \frac{U}{n1} - \frac{L}{n2}$$

B = the index under consideration.

U = the number of students in the upper group who get the item correct.

L = the number of students in the lower group who get the item correct.

n1 = the total number of students in upper group.

n2 = the total number of students in lower group.

n2-L) = the number of correct discriminations.

L(n1-U) = the number of incorrect discriminations.

n1n2 = the total number of possible discriminations.

ตัวอย่างขั้นตอนการวิเคราะห์ความเที่ยงตรงเชิงโครงสร้าง โดยวิธีหาค่าดัชนีจำแนก B (B-Index) (บุญเชิด ภิญโญนนตพงษ์, 2527 : 84)

1. เลือกกลุ่มตัวอย่างนักเรียนจำนวนหนึ่ง ซึ่งผ่านการเรียนรู้ตามจุดประสงค์ของข้อสอบที่ต้องการจะวัด
2. นำข้อสอบในจุดประสงค์เดียวกันไปสอบกับกลุ่มตัวอย่าง
3. กำหนดคะแนนจุดตัดจากจำนวนข้อสอบและตัวเลือกของข้อสอบ
4. แบ่งนักเรียนออกเป็นกลุ่มรอบรู้ (ผู้ที่ได้คะแนนสูงกว่าคะแนนจุดตัด) และกลุ่มไม่รอบรู้ (ผู้ที่ได้คะแนนต่ำกว่าจุดตัด)
5. แจกแจงความถี่ของกลุ่ม รอบรู้-ไม่รอบรู้ ตอบถูก
6. หาค่าดัชนีอำนาจจำแนก โดยการแทนค่าตามสูตร
7. คัดเลือกข้อสอบ

สมมติว่านำข้อสอบแบบเลือกตอบ 4 ตัวเลือก ซึ่งวัตถุประสงค์เดียวกัน จำนวน 5 ข้อ ไปสอบกับนักเรียนจำนวน 40 คน ผลการสอบแบ่งจำนวนกลุ่มผู้รอบรู้ จำนวน 15 คน กลุ่มไม่รอบรู้ จำนวน 25 คน ดังนั้น $n_1 = 15$ $n_2 = 25$

เมื่อแจกแจงความถี่ของการตอบถูกของข้อสอบแต่ละข้อ และแต่ละกลุ่ม ได้ค่า U และ L ตามตาราง

ข้อสอบข้อที่	U	L	U/n ₁	L/n ₂	B-Index
1	15	5	1.00	0.20	0.80
2	3	20	0.20	0.80	-0.60
3	10	20	0.67	0.80	-0.13
4	10	10	0.67	0.40	0.27
5	15	25	1.00	1.00	0.00

จากตารางค่า B-Index คือ ค่าดัชนีจำแนกของข้อสอบ แต่ละข้อ ซึ่งแสดงให้เห็นว่าข้อสอบข้อ 1 สามารถจำแนกผู้รอบรู้ ไม่รอบรู้ได้ดีมาก ข้อ 4 สามารถจำแนกผู้รอบรู้ ไม่รอบรู้ได้บ้าง ส่วนข้อ 2 ข้อ 3 และ ข้อ 5 จำแนกไม่ได้ควรตัดทิ้งไปหรือแก้ไขปรับแต่งข้อสอบใหม่

2. การวิเคราะห์ค่าความยากของข้อสอบ (Item Difficulty)

ค่าความยากของข้อสอบ ในที่นี้เลือกใช้นิยาม ค่าความยากของข้อสอบ หมายถึงจำนวนเปอร์เซ็นต์ (Percentage) หรือ สัดส่วน (Proportion) ของนักเรียนที่เลือกตอบตัวเลือกนั้น เมื่อพิจารณาเปรียบเทียบกับนักเรียนทั้งหมด โดยมีสูตรการคำนวณดังนี้

$$P = \frac{RH + RL}{nH + nL}$$

P หมายถึง ค่าความยากของข้อสอบ

$RH+RL$ หมายถึงจำนวนนักเรียนที่ตอบข้อสอบข้อนั้นถูก ทั้งกลุ่มผ่านและไม่ผ่านเกณฑ์

$nH+nL$ หมายถึงจำนวนนักเรียนทั้งสิ้นที่ตอบข้อสอบ

การวิเคราะห์ความเชื่อมั่นของแบบทดสอบ (Reliability)

จุดมุ่งหมายของ การทดสอบแบบอิงกลุ่ม เพื่อที่จะเปรียบเทียบความสามารถของนักเรียนที่เข้าสอบด้วยกัน ส่วน การทดสอบแบบอิงเกณฑ์ เป็นการทดสอบที่ต้องการเปรียบเทียบความสามารถของนักเรียน กับเกณฑ์มาตรฐานของเนื้อหาวิชาที่กำหนดให้

ความเชื่อมั่นของแบบทดสอบแบบอิงกลุ่ม หมายถึง ความคงที่แน่นอนซึ่งได้จากการสอบซ้ำกับกลุ่มตัวอย่างกลุ่มเดียวกันด้วยแบบทดสอบฟอร์มเดียวกันหรือแบบทดสอบคู่ขนานสองฟอร์มซึ่งมีเซตของข้อสอบที่เทียบเท่ากัน (Anastasi, 1968: 71) (อ้างอิงใน บุญเชิด ภิญโญอนันตพงษ์, 2527 : 149)

ความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ หมายถึง คะแนนความ สอดคล้องในการตัดสินจำแนก ผู้รอบรู้-ไม่รอบรู้ จากการสอบซ้ำด้วยแบบทดสอบฉบับเดียวหรือแบบทดสอบคู่ขนานหรือเทียบเท่ากันสองฟอร์ม หรือ หมายถึง ความสอดคล้องของคะแนนแต่ละคนที่แปรปรวนจากคะแนนจุดตัด จากการสอบด้วยแบบทดสอบ คู่ขนานสองฟอร์ม

จากนิยามของความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์จะเห็นได้ว่ามีคะแนน เกณฑ์ หรือคะแนนจุดตัดมาเกี่ยวข้องด้วย การหาค่าความเชื่อมั่นแบบอิงกลุ่มวิธีทั่วไปจึงไม่เหมาะสมที่จะนำมาใช้กับ การทดสอบแบบอิงเกณฑ์ (Criterion-Referenced Test) ซึ่งเป็นแบบ Mastery Test ส่วนมากการวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบ เป็นการวิเคราะห์หาสัมประสิทธิ์ความสอดคล้องในการตัดสินจำแนกผู้รอบรู้และผู้ไม่รอบรู้ของการสอบสองครั้งด้วยข้อสอบฉบับเดียวหรือข้อสอบคู่ขนาน

การหาความเชื่อมั่นของแบบทดสอบแบบอิงกลุ่ม (Norm Referenced Test) โดยทั่วไป มี 4 วิธี คือ (สำนักทดสอบทางการศึกษา, 2531 : 200-207)

1. Coefficient of Stability ความเชื่อมั่นชนิดนี้จะต้องมีการสอบ 2 ครั้ง ด้วยแบบทดสอบฉบับเดียวกันและผู้สอบกลุ่มเดียวกัน (Test-Retest) โดยเว้นระยะเวลาระหว่างการสอบทั้ง 2 ครั้งนั้นให้เพียงพอที่ผู้เข้าสอบจะระลึกข้อสอบไม่ได้ แล้วนำคะแนนจากการสอบทั้ง 2 ครั้งนั้นมาหาค่าสหสัมพันธ์ ดังนี้

1.1 ถ้าคะแนนจากการสอบอยู่ในระบบ Interval Scale ใช้ Pearson Product Moment Correlation ของค่าความเชื่อมั่น

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

- เมื่อ r_{XY} หมายถึง สัมประสิทธิ์ของความเชื่อมั่น
 N หมายถึง จำนวนผู้เข้าสอบ
 X หมายถึง คะแนนจากการสอบครั้งที่ 1
 Y หมายถึง คะแนนจากการสอบครั้งที่ 2

1.2 ถ้าคะแนนจากการสอบอยู่ในระดับ Ordinal Scale ใช้ Spearman Rank Correlation

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

- เมื่อ ρ หมายถึง สหสัมพันธ์ระหว่างการสอบวัด 2 ชุดที่อยู่ในรูปเรียงอันดับ
 N หมายถึง จำนวนคู่ของการเรียงลำดับ
 D หมายถึง ผลต่างของคะแนนอันดับแต่ละคู่

2. Coefficient of Equivalent ซึ่งเป็นวิธีการใช้แบบทดสอบคู่ขนานสองฉบับ สอบติดต่อกันเพื่อขจัดปัญหาอันเกิดจากการเว้นช่วงระยะเวลาการสอบ แล้วนำคะแนนจากการสอบทั้ง 2 ฉบับมาหาค่าสัมประสิทธิ์สหสัมพันธ์เช่นเดียวกับวิธีที่ 1

3. Coefficient of Stability and Equivalent วิธีนี้เป็นวิธีผสมผสานระหว่างวิธีที่ 1 และวิธีที่ 2 โดยใช้แบบทดสอบคู่ขนาน แต่เว้นช่วงระยะเวลาของการสอบครั้งที่ 1 และ ครั้งที่ 2

4. Coefficient of Internal Consistency การหาความเชื่อมั่นวิธีนี้ เป็นการสอบเพียงครั้งเดียว แล้วแบ่งแบบทดสอบออกเป็น 2 ตอน นำคะแนนทั้ง 2 ชุด มาคำนวณดังนี้

4.1 วิธี Split-Half วิธีนี้มีข้อตกลงว่าแบบทดสอบที่แบ่งเป็น 2 ตอน ต้องมีความเท่าเทียมกัน ความแปรปรวนของข้อสอบทั้ง 2 ฉบับต้องเท่ากัน

4.2 วิธี Kuder-Richardson ซึ่งมีข้อตกลงพื้นฐานว่าเนื้อหาของแบบทดสอบจะต้องเป็นเอกพันธ์ คือ จะต้องวัดความสามารถ ความรู้ หรือทักษะเดียวกัน และข้อสอบมีระบบการให้คะแนนแบบ 0-1 (Zero- one Method) โดยมีสูตรในการคำนวณดังนี้

$$[KR_{20}] \quad r_{tt} = \left(\frac{n}{n-1} \right) \left(\frac{s_t^2 - \sum pq}{s_t^2} \right)$$

r_u หมายถึง ความเชื่อมั่นของแบบทดสอบ

n หมายถึง จำนวนของข้อสอบ

s_t^2 หมายถึง ค่าความแปรปรวนของข้อสอบทั้งฉบับ

p หมายถึง สัดส่วนของคนทำถูกในแต่ละข้อ

$$q = 1 - p$$

อีกสูตรหนึ่ง คือ KR_{21} สูตรนี้มีข้อตกลงเพิ่มขึ้น คือ นอกจากข้อสอบเป็นลักษณะเดียวกันแล้ว แต่ละข้อจะต้องมีระดับความยากเท่ากันหรือไม่แตกต่างกันมาก

$$[KR_{21}] r_u = \left(\frac{n}{n-1} \right) \left(1 - \frac{\bar{x}(n-\bar{x})}{ns_t^2} \right)$$

r_u หมายถึง ความเชื่อมั่นของแบบทดสอบ

n หมายถึง จำนวนข้อสอบ

\bar{x} หมายถึง ค่าเฉลี่ยของแบบทดสอบ

s_t^2 หมายถึง ค่าความแปรปรวนของแบบทดสอบ

การหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์

การวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบแบบอิงกลุ่มวิธีต่าง ๆ ที่กล่าวมาเป็นการคำนวณจากอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงและความแปรปรวนของคะแนนผลการสอบ

$$r_u = 1 - \frac{\delta_t^2}{\delta_0^2}, \delta_t^2 = \text{ความแปรปรวนของ True Score}$$

$$\text{เนื่องจาก } \delta_t^2 = \delta_0^2 - \delta_e^2, \delta_e^2 = \text{ความแปรปรวนของ Error Score}$$

$$\text{ดังนั้น } r_u = 1 - \frac{\delta_e^2}{\delta_0^2}, \delta_0^2 = \text{ความแปรปรวนของ Observed Score}$$

เนื่องจากการวิเคราะห์ดังกล่าวเกี่ยวข้องกับความแปรปรวนของคะแนน ซึ่งถ้าความแปรปรวนของคะแนนมาก ค่าความเชื่อมั่นจะยิ่งสูง แต่การทดสอบแบบอิงเกณฑ์นั้น ความแปรปรวนจะน้อย ค่าความเชื่อมั่นเข้าใกล้ศูนย์ (Pophan and Husek, 1969, Stanley, 1971) (อ้างอิงใน โกวิท ประมวลพฤษ์ และสมศักดิ์ สินธุระเวชย์, 2527 : 256) ฉะนั้นวิธีหาค่าความเชื่อมั่นแบบอิงกลุ่มจึงไม่เหมาะสมที่จะนำมาใช้กับระบบอิงเกณฑ์

การวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ส่วนมาก เป็นการวิเคราะห์หาค่าสัมประสิทธิ์ความสอดคล้องในการตัดสินจำแนก ผู้รอบรู้ และผู้ไม่รอบรู้ ของการสอบสองครั้งด้วยข้อสอบฉบับเดียวหรือข้อสอบคู่ขนาน

วิธีการวิเคราะห์ความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์มีหลาย ๆ วิธี เช่น

1. วิธีของคาร์เวอร์ (1970)
2. วิธีของแฮมเบิลตัน และ โนวิก (Hambleton and Novick, 1973)
3. วิธีของบุญเชิด ภิญโญอนันตพงษ์
4. วิธีของสวามินาธาน,แฮมเบิลตัน และ อัลจินา(Swaminathan, Hambleton and Algina : 1974)
5. วิธีของสับโคเวียก (Subkoviak, 1976)
6. วิธีของ ฮวิน(Huynh,1976)

ตัวอย่างวิธีการวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์

1. วิธีของสวามินาธาน, แฮมเบิลตัน และ อัลจินา

เป็นวิธีการใช้หลักการสอบ 2 ครั้ง และถ้าข้อสอบฉบับนั้นวัดหลาย ๆ จุดประสงค์หรือหลาย ๆ พฤติกรรม การพิจารณาความเชื่อมั่นจะแยกเป็น Subtest ของแต่ละจุดประสงค์นั้น เมื่อได้ผลการสอบ 2 ครั้งแล้ว นำไปแจกแจงตาราง (โกวิท ประมวลพฤษย์ และ สมศักดิ์ สิ้นธุระเวชญ์, 2527 : 257)

ผลการสอบครั้งที่ 1

		ผ่าน	ไม่ผ่าน	รวม
ผลการสอบ ครั้งที่ 2	ผ่าน	a	b	a+b
	ไม่ผ่าน	c	d	c+d
	รวม	a+c	b+d	a+b+c+d

a, b, c, d คือสัดส่วนของจำนวนนักเรียนที่เป็นไปตามเงื่อนไข เช่น a คือ สัดส่วนจำนวนนักเรียนที่สอบผ่านเกณฑ์ทั้งสองครั้ง b คือ สัดส่วนจำนวนนักเรียนที่สอบไม่ผ่านเกณฑ์ครั้งที่ 1 แต่ผ่านเกณฑ์ครั้งที่ 2 เป็นต้น

เมื่อคิดสัดส่วนในตาราง ทุกช่องแล้ว นำไปคำนวณจากสูตร

$$K = \frac{Po - Pc}{1 - Pc}$$

K = ความเชื่อมั่นของแบบทดสอบ

Po = สัดส่วนความสอดคล้องที่สังเกตได้จากการสอบซ้ำ

Pc = สัดส่วนความสอดคล้องที่คาดหวังได้จากความบังเอิญ

โดย $Po = a+d$

$$Pc = (a+b)(a+c) + (c+d)(b+d)$$

K จะมีค่าอยู่ระหว่าง -1.00 ถึง +1.00 ถ้า $K = 1$ แสดงถึงความสอดคล้องอย่างสมบูรณ์

2. วิธีของ Carver

วิธีนี้ ใช้แบบทดสอบคู่ขนาน สอบนักเรียนกลุ่มเดียวกัน แล้วนำผลการสอบไปแจกแจงลงตาราง

		แบบทดสอบ B	
		ผ่าน	ไม่ผ่าน
แบบทดสอบ A	ผ่าน	b	a
	ไม่ผ่าน	c	d

$$Po = \frac{(a+c)}{(a+b+c+d)}$$

Po คือ สัดส่วนของความสอดคล้องในการตัดสินใจรอบรู้

3. วิธีของ Livingston

การวิเคราะห์หาค่าความเชื่อมั่นแบบอิงเกณฑ์อีกวิธีหนึ่งเป็นการวิเคราะห์หาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนแต่ละข้อกับคะแนนรวมเป็นการวิเคราะห์แบบ Internal Consistency วิเคราะห์จากแบบทดสอบฉบับเดียว ซึ่ง ลิฟวิงสตัน (Livingston, 1970) (อ้างอิงในอนันต์ ศรีโสภณ, 2525 : 85) ได้สร้างสูตรสำหรับหาค่าความเชื่อมั่นแบบอิงเกณฑ์ (Criterion Referenced Test) ดังนี้

$$r_{cc} = \frac{r_{tt}\delta_0^2 + (\bar{x} - c)^2}{\delta_0^2 + (\bar{x} - c)^2}$$

- เมื่อ r_{cc} หมายถึง ความเชื่อมั่นของ Criterion Referenced Test
 r_{tt} หมายถึง ความเชื่อมั่นของแบบทดสอบ
 δ_0^2 หมายถึง ความแปรปรวนของคะแนนการสอบ
 \bar{x} หมายถึง ค่าเฉลี่ยของคะแนน
 C หมายถึง คะแนนเกณฑ์ (Criterion Score)

ถ้าคะแนนเกณฑ์ (C) เท่ากับค่าคะแนนเฉลี่ย (\bar{x}) r_{cc} จะเท่ากับ r_{tt} แต่โดยทั่วไปแล้วคะแนนเกณฑ์ (C) จะเบี่ยงเบนออกไปจากคะแนนเฉลี่ย (\bar{x}) เสมอ จะทำให้ความเชื่อมั่นของ Criterion Referenced Test สูงขึ้น

บรรณานุกรม

กองวิจัยทางการศึกษา, วิชาการ, กรม. การวิจัยเพื่อพัฒนาการเรียนรู้ ตามหลักสูตรการ
ศึกษา **ขั้นพื้นฐาน**, กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว, 2545.

โกวิท ประวาลพุกษ์ และ สมศักดิ์ สิ้นธุระเวชญ์. การประเมินผลในชั้นเรียน. กรุงเทพฯ :
พิมพ์ครั้งที่ 2, วัฒนาพานิช, 2527.

บุญเชิด ภิญโญอนันตพงษ์. การทดสอบแบบอิงเกณฑ์ : แนวคิดและวิธีการ. กรุงเทพฯ :
O.S.PRINTING HOUSE, 2527.

วิชาการ, กรม. **คู่มือครูการประเมินผลการเรียนระดับมัธยมศึกษา** ตามหลักสูตรฉบับปรับ
ปรุง พ.ศ.2533. กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว, 2535.

_____ **คู่มือครูการประเมินผลการเรียนระดับมัธยมศึกษาตอนต้น พุทธศักราช**
2521 (ฉบับปรับปรุง พ.ศ.2533. กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว, 2535.

สำนักงานทดสอบทางการศึกษา, วิชาการ, กรม. **การประเมินผลการเรียนระดับมัธยมศึกษา**.
กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว, 2531.

_____ **การประเมินผลการเรียนตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช**
2544. กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว, 2545.

สุชีรา ภัทรายุตวรรัตน์. **คู่มือการวัดทางจิตวิทยา**. กรุงเทพฯ : สำนักพิมพ์ เมติคัล มีเดีย , 2545.

สุพัฒน์ สุขมลสันต์. **การวิเคราะห์ข้อสอบแนวใหม่ด้วยคอมพิวเตอร์**. กรุงเทพฯ : นำอักษร
การพิมพ์, 2538.

หน่วยศึกษานิเทศก์ กรมสามัญศึกษา. **ชุดฝึกอบรมเรื่องการวิจัยในชั้นเรียน**. กรุงเทพฯ : พิมพ์
ครั้งที่ 2, โรงพิมพ์การศาสนา, 2544.