

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้องในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอเป็น 5 ประเด็นดังนี้

1. แนวคิดของการวัดผลแบบอิงเกณฑ์
2. ความหมายและลักษณะของแบบสอบอิงเกณฑ์
3. การสร้างและคุณภาพของแบบสอบอิงเกณฑ์
4. แบบสอบปรนัยแบบเลือกตอบและแบบตอบสั้น
5. เทคนิคการวิเคราะห์ตามลำดับขั้น และงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังนี้

1. แนวคิดของการวัดผลแบบอิงเกณฑ์

การให้การศึกษาเป็นการส่งเสริมความสามารถ ความคิดและสติปัญญาของผู้เรียน ให้พัฒนา งามองตามศักยภาพของแต่ละบุคคลโดยอาศัยวิธีการที่เหมาะสม การสอนจึงเป็นวิธีการที่ทำให้ผู้เรียนเกิดการเรียนรู้และการสอบวัดจะทำหน้าที่บอกระดับความสำเร็จในการเรียน (ไพศาล หวังพาณิชย์, 2526) การวัดผลที่ใช้กันแพร่หลายแต่ดั้งเดิม คือ การวัดผลแบบอิงกลุ่ม (Norm Reference Measurement) คะแนนที่ได้จากการวัดผลวิธีนี้จะนำมาจำแนกนักเรียนเป็นระดับต่างๆตามคะแนนที่ได้รับและจะตัดสินความสำเร็จของการเรียนรู้ โดยนำคะแนนที่ได้มาเปรียบเทียบกับคะแนนของนักเรียนอื่นๆ ในกลุ่ม (เบญจวรรณ ผ่องแผ้ว, 2529)

การวัดผลแบบอิงกลุ่มได้รับการวิพากษ์วิจารณ์กันมากกว่า เป็นการจัดการศึกษาที่มุ่งส่งเสริมเฉพาะนักเรียนที่เก่ง อีกทั้งยังไม่สามารถบอกจุดอ่อนหรือความสามารถ ในการทำข้อสอบของนักเรียนได้ ไม่เหมาะที่จะนำมาวัดการเรียนรู้ของแต่ละบุคคล การจัดการศึกษาควรจะให้ให้นักเรียนได้พัฒนาบรรลุเป้าหมายที่ต้องการทุกคน คือทั้งนักเรียนเก่งและอ่อน เมื่อเรียนแล้วก็ควรจะงอกงามก้าวหน้าได้เหมือนกัน

ในปัจจุบันได้เกิดแนวความคิดว่า นักเรียนส่วนมาก (อาจมากกว่า 90 %) ควรจะสามารถรอบรู้ได้ในสิ่งที่ครูสอน ผู้ที่เกี่ยวข้องทั้งปวงมีหน้าที่ที่จะแสวงหาวิธีการมาช่วยให้นักเรียนเรียนให้ "รอบรู้" ความที่ต้องการได้ การ "รอบรู้" ที่ต้องการหมายถึงการที่นักเรียนมีคุณสมบัติหรือความสามารถถึงระดับที่คาดหวังไว้ ฉะนั้นจุดมุ่งหมายในการเรียนการสอนจึงจะต้องชัดเจนและมี "เกณฑ์" ที่บอกให้ทราบว่าพฤติกรรมระดับใดจึงจะเรียกว่า "บรรลุ" ถึงระดับเรียกว่า "รอบรู้" นั้น (กมล ภูประเสริฐ, 2520) การเรียนในลักษณะนี้เหมาะกับการวัดผลแบบอิงเกณฑ์ (CRITERION-REFERENCED MEASUREMENT) ซึ่งสืบเนื่องมาจากทฤษฎีการเรียนรู้แบบ การเรียนเพื่อรอบรู้ (MASTERY LEARNING) ของบลูม (Bloom) ที่กล่าวว่า "การเรียนรู้ทั้งหลาย ควรจะเป็นการเรียนเพื่อรอบรู้ในเนื้อหาต่างๆ ครึ่งต้องการให้นักเรียนทุกคนบรรลุสู่ความเป็นผู้รอบรู้ หรือ ควรจะเป็น 80-90 % ของเนื้อหาที่เรียนรู้" (สมศักดิ์ สินธุระเวชช์, 2521) แนวความคิดนี้เชื่อว่า นักเรียนทุกคนสามารถเรียนบทเรียนใดๆ ก็ได้หากให้เวลาและโอกาสแก่เขามากพอ ดังนั้นการวัดผลแบบอิงเกณฑ์จึงสนใจความสามารถของนักเรียนแต่ละคน มากกว่าสนใจเปรียบเทียบกับคะแนนของนักเรียนผู้นั้นกับผู้อื่น

จากหลักการของการเรียนเพื่อรอบรู้ดังกล่าว ต่อมาได้พัฒนาเป็นการวัดผลแบบอิงเกณฑ์ โดยจะเห็นว่า การเรียนเพื่อรอบรู้สอดคล้องการวัดผลแบบอิงเกณฑ์มาก หรืออาจกล่าวได้ว่าการวัดผลแบบอิงเกณฑ์มีรากฐานมาจากการเรียนเพื่อรอบรู้ และการเรียนเพื่อรอบรู้ก็มีรากฐานมาจากหลักความแตกต่างระหว่างบุคคลนั่นเอง (เอนก เพียรอนุกุลบุตร 2524)

2. ความหมายและลักษณะของแบบสอบอิงเกณฑ์

แบบสอบอิงเกณฑ์ถูกนำมาใช้ในโรงเรียนมากขึ้น ได้มีการศึกษาค้นคว้า และให้นิยามของแบบสอบอิงเกณฑ์ไว้แตกต่างกัน ดังนี้

โพแฟม (Popham, 1975) ได้นิยามแบบสอบอิงเกณฑ์ไว้ว่า "แบบสอบอิงเกณฑ์เป็นแบบสอบที่ชี้ข้งถึงสถานภาพของผู้สอบเป็นรายบุคคล ซึ่งอ้างอิงถึงคะแนนโดเมน ทั้งนี้ต้องกำหนดพฤติกรรมในโดเมนที่ชัดเจน

เกลเซอร์และนิตโก (Glaser and Nitko in Thorndike, R.L., 1971) กล่าวว่า แบบสอบอิงเกณฑ์เป็นแบบสอบที่สร้างขึ้น โดยมีเจตนาให้เป็นการวัดที่สามารถแปลความหมายออกมาได้ตรง ๆ ในรูปมาตรฐานการกระทำที่เฉพาะ ซึ่งในที่นี้ข้งชี้



ได้โดยกำหนดโดเมนของงานที่นักเรียนหรือผู้สอบควรทำได้ จากนั้นนำตัวอย่างที่เป็นตัวแทนของงานจากโดเมนนี้มาจัดเป็นแบบสอบ ผลการวัดจะแสดงถึงพฤติกรรมของแต่ละคนโดยเทียบกับคะแนนโดเมนที่กำหนดไว้

ส่วนมิลแมน (Millman , 1974) ได้ให้ความหมายของแบบสอบอิงเกณฑ์ ได้ชัดเจนคือว่าแบบสอบนั้นต้องการวัดอะไร แบบสอบประกอบไปด้วยกลุ่มตัวอย่างข้อกระถงจากโดเมนที่อาจจะได้จากการสุ่มตัวอย่างแบบง่ายหรือสุ่มแบบแบ่งชั้น (Simple Random Sampling or Stratified Random Sampling)

สงข ลักษณ์ (2523) ได้กำหนดลักษณะของแบบสอบอิงเกณฑ์ออกเป็นกลุ่ม ๆ ดังนี้

1. ลินด์วอลล์ และ นิตโก (Lindvall and Nitko) และนักวัดผลอีกกลุ่มหนึ่ง มีแนวคิดว่ามี การสร้างแบบสอบ โดยอาศัยจุดประสงค์เชิงพฤติกรรมที่สร้างขึ้นโดยใช้หลัก สูตร และพฤติกรรมที่มุ่งหวังในแต่ละเนื้อหาวิชา ก็จัดได้ว่าเป็นแบบสอบอิงเกณฑ์

2. เกลเซอร์ เมเจอร์ และ โฟแฟม (Glasser, Mager and Popham) มีแนวคิดว่ามี แบบสอบอิงเกณฑ์จะต้องมีคะแนนจุดตัด (Cut off Score) เพื่อจำแนกผู้สอบ ออกเป็น 2 พวก คือ พวกเรียนรักรบถ้วน กับ พวกเรียนไม่ครบถ้วน

3. แฮร์ริส และ สจ๊วต (Harris and Stewart) และนักวัดผลอีกกลุ่ม หนึ่งเชื่อว่าแบบสอบอิงเกณฑ์ที่แท้จริงต้องมีการสุ่มตัวอย่างข้อกระถง เนื้อเป็นตัวแทนของข้อ กระถง ในจุดประสงค์นั้นทั้งหมด ซึ่งเรียกว่า โดเมน การคิดคะแนนทำโดยหาอัตราส่วน ที่ผู้สอบตอบข้อกระถงที่สุ่มมาได้ถูกต้อง

โกวิท ประวาลพฤษย์ (2523) ได้กล่าวถึง ลักษณะของแบบสอบอิงเกณฑ์มี 2 ประเภท คือ

1. การยึดจุดประสงค์เป็นหลัก หรือเรียกว่า แบบทดสอบอิงจุดประสงค์ (Objective Referenced Test) กล่าวคือ ในการเรียนการสอนมีการกำหนด จุดประสงค์เชิงพฤติกรรม และการสอบวัดก็ยึดจุดประสงค์เป็นเกณฑ์ เพื่อจะบ่งชี้ว่าใครผ่าน เกณฑ์ที่กำหนดไว้ในจุดประสงค์หรือยัง

2. การขยายจุดประสงค์ให้ครอบคลุมประชากรของพฤติกรรม ที่เรียกว่า โดเมน แบบสอบในลักษณะนี้เรียกว่าแบบทดสอบอิงโดเมน (Domain Referenced Test) ข้อสอบต่าง ๆ ได้จากกลุ่มตัวอย่างของพฤติกรรมที่กำหนดจากโดเมน เมื่อสอบวัดแล้วสามารถ อ้างอิงได้ว่า บุคคลมีความสามารถระดับใดตามคะแนนมาตรฐาน เนื้อหา หรือระดับความ รอบรู้แนวคิดขนาดใด

จากความหมายและลักษณะของแบบสอบอิงเกณฑ์ สรุปได้ว่า
แบบสอบอิงเกณฑ์ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้นตามจุดประสงค์
เชิงพฤติกรรมตามเนื้อหาวิชาหรือกลุ่มตัวอย่างของข้อสอบที่สุ่มมาจากโดเมนของพฤติกรรมต่างๆ
เพื่ออ้างอิงความสามารถของบุคคลตามมาตรฐานของเนื้อหาวิชาคะแนนที่ได้จากแบบสอบอิงเกณฑ์
จะแปลความหมายในลักษณะที่เป็นความสามารถเฉพาะบุคคลโดยเทียบกับเกณฑ์ที่กำหนดไว้

3. การสร้างและคุณภาพของแบบสอบอิงเกณฑ์

การสร้างแบบสอบอิงเกณฑ์

แบบสอบอิงเกณฑ์มีแนวความคิดที่ต้องการสรุปอ้างอิงไปยังเนื้อหา ที่ต้องการวัด
ได้ชัดเจนและเหมาะสม ดังนั้นการสร้างแบบสอบอิงเกณฑ์ จึงต้องมีการระบุนเขต ขอบเขต
ประชากรของข้อกระทงให้ชัดเจน เพื่อให้แบบสอบที่สร้างขึ้นมีความเป็นปรนัยในการแปล
ความหมาย โดยใช้คะแนนที่สอบได้เป็นค่าประมาณของระดับผลสัมฤทธิ์ทางการเรียนรูทั้ง
ขอบข่ายความรู้ และทักษะ หรือโดเมนที่กำหนด ซึ่งเป็นสิ่งสำคัญยิ่งต่อการสอนอย่างเป็น
ระบบ กรอนแลนด์ (Gronlund 1973 อ้างอิงใน เขาวดี วิบูลย์ศรี 2526) ได้
เสนอหลักการสร้างแบบสอบอิงเกณฑ์ไว้ให้พิจารณา 4 ประการ คือ

1. ตัดสินว่าการเรียนรู้ใดควรจะบรรจุในการเรียนรู้รอบ
2. จำกัดขอบข่ายของเนื้อหาและทักษะที่ต้องการทดสอบ
3. กำหนดผลลัพธ์ของการเรียนรู้ให้ชัดเจน
4. กำหนดเกณฑ์มาตรฐานในการปฏิบัติ

สมศักดิ์ สินธุระเวช (2522) ได้กล่าวถึง การสร้างแบบทดสอบอิงเกณฑ์
ว่านอกจากตัวคำถามจะต้องเขียนให้ชัดเจน ไม่คลุมเครือเพราะจะต้องแสดงให้เห็นว่าผู้สอบ
สามารถทำได้หรือไม่แล้ว แผนการต่าง ๆ ยังต้องเขียนให้ชัดเจนเช่นกันไม่ว่าจะเป็น
ตารางวิเคราะห์จุดมุ่งหมาย หรือรูปแบบข้อสอบ เพราะสิ่งต่าง ๆ เหล่านี้จะทำให้การสร้าง
ข้อสอบเป็นไปอย่างมีประสิทธิภาพ

ในศาล หวังพาณิชย์ (2526) ได้สรุปขั้นตอนในการสร้างข้อสอบแบบอิงเกณฑ์
ไว้ ดังต่อไปนี้

1. กำหนดจุดมุ่งหมายของการสอบโดยกำหนดในรูปจุดมุ่งหมายเชิงพฤติกรรม
2. เปรียบสร้างคำถามที่ใช้วัดคุณลักษณะ และเนื้อหาต่าง ๆ ที่สอดคล้องกับ
จุดมุ่งหมายเหล่านั้น

3. ใช้ข้อสอบทำการทดสอบเป็นระยะ
4. เมื่อจะมีการวัดเพื่อประเมินผลการเรียนขั้นสุดท้าย ก็ใช้วิธีเลือกจุดมุ่งหมายเชิงพฤติกรรมจากเนื้อหาต่าง ๆ เพื่อใช้เป็นตัวแทนคุณลักษณะที่จะวัด จากนั้นก็เขียนคำถามวัดพฤติกรรม และเนื้อหาต่าง ๆ ตามจุดมุ่งหมายเหล่านั้น สแตนเลย์ และ ฮอปกินส์ (Stanley and Hopkins, 1972 อ้างอิงใน บุญเชิด วิทยุณันต์พงษ์, 2527) ได้เสนอขั้นตอนในการสร้างแบบสอบอิงเกณฑ์ไว้ดังนี้
 - ขั้นที่ 1 ขั้นวางแผนสร้างแบบทดสอบ
 - ขั้นที่ 2 ขั้นเตรียมงานเขียนข้อสอบ และลงมือเขียนข้อสอบ
 - ขั้นที่ 3 ขั้นทดลอง
 - ขั้นที่ 4 ขั้นประเมินผลการสอบ

รอยด์ และ ฮาลาดินา (Roid and Haladyna, 1980) ได้เสนอวิธีการสร้างแบบสอบ ที่เรียกว่า เทคโนโลยีของการเขียนข้อสอบ โดยมีแนวคิดที่สำคัญ และที่นิยมใช้กัน คือ วิธีการกำหนดวัตถุประสงค์สำคัญ (Principal objectives) ซึ่งเทคนิคนี้สามารถแบ่งตามรายละเอียด ดังนี้

1. การกำหนดวัตถุประสงค์ขยาย (Amplified Objectives)

วิธีการนี้เป็นการพัฒนาจากวัตถุประสงค์เชิงพฤติกรรม ทั้งนี้ในตัวของวัตถุประสงค์ขยายจะมีการแจกแจงรายละเอียดเพิ่มขึ้น โดยบรรยายให้ทราบถึงสถานการณ์ของการสอบ วิธีการสร้างตัวเลือกและเกณฑ์ในการให้คะแนนกับวัตถุประสงค์เฉพาะนั้น ๆ จากวัตถุประสงค์ที่เด่นชัดดังกล่าว จะทำให้สามารถสร้างข้อสอบในแนวเดียวกันออกได้เป็นจำนวนมาก ทำให้รูปแบบการเขียนข้อสอบมีความเป็นปรนัยสูง
2. การกำหนด โครงสร้างรูปแบบเฉพาะประจำข้อสอบ (Item Form)

วิธีการสร้างข้อสอบจะเป็นการเขียนบรรยายคำ โครงสร้างหลักของคำถามที่จะใช้ เป็นหลักหรือแกน โดยที่บางส่วนของประโยคในโครงสร้างหลักจะถูกกำหนดให้มีลักษณะเป็นตัวแปร ลักษณะการสร้างข้อสอบนั้น ทุกๆ ครั้งที่มีการตั้งคำถาม โครงสร้างหลักหรือแกนของประโยคจะยังคงสภาพเดิม แต่ค่าของตัวแปรจะถูกเปลี่ยนแปลง โดยวิธีการนี้จะทำให้ได้ข้อสอบเป็นจำนวนมากที่วัดพฤติกรรมการเรียนรู้ได้เช่นกัน

นอกจากนี้ รอยด์ และ ฮาลาดินา (Roid and Haladyna, 1980) ยังได้ให้หลักการสร้างข้อสอบของแบบสอบผลสัมฤทธิ์ มี 5 ขั้น ดังนี้

- ขั้นที่ 1 มโนทัศน์ที่เป็นหลักการเหตุผล และแนวคิดในการเรียนการสอนผู้สอน ต้องกำหนดพฤติกรรมที่วัดหมาย ให้เกิดขึ้นในตัวผู้เรียนหลังจากได้รับ

- การเรียนการสอนแล้ว และสัมพันธ์กับโดเมนของเนื้อหาซึ่งอาจจะหา
ได้โดยการวิเคราะห์เนื้อหาวิชา หรือวิเคราะห์งาน
- ขั้นที่ 2 เป็นขั้นที่กำหนดเนื้อหาที่จะสอน แบ่งได้ ดังนี้
- ก. กำหนดจุดประสงค์การเรียนการสอน
 - ข. กำหนดลักษณะเฉพาะของโดเมนเนื้อหาให้สอดคล้องกับจุดมุ่งหมาย
ที่ทำให้เกิดแก่ผู้เรียน
- ขั้นที่ 3 ขั้นพัฒนาข้อกระทงซึ่งเป็นขั้นตอนการสร้างข้อกระทง โดยใช้เทคนิค
การเขียนข้อกระทง ซึ่งควรเขียนไปตามโดเมนที่อธิบายเนื้อหาให้เป็นตัว
แทนการเรียนการสอนตามเป้าหมาย
- ขั้นที่ 4 ปรับปรุงข้อกระทง ถ้ามีข้อบกพร่องในข้อคำถาม ต้องปรับปรุงแก้ไข
ให้เรียบร้อยก่อนนำไปใช้สอบ
- ขั้นที่ 5 คัดเลือกข้อกระทงมาใช้การลุ่มตัวอย่างข้อกระทง จากโดเมนข้อสอบ

เทคนิคการเขียนข้อสอบอิงเกณฑ์

การเขียนข้อสอบต้องเขียนให้สอดคล้องกับเนื้อหา และพฤติกรรมที่ต้องการวัด
ฉะนั้นจำเป็นต้องศึกษาว่าเนื้อหาที่ต้องการวัดนั้นมีลักษณะเป็นอย่างไร เพื่อนำมาแปลงให้
พฤติกรรมที่จะวัดได้เฉพาะ และชัดเจน ซึ่งจะทำได้โดยสามารถเขียนข้อสอบได้ตรงตามความ
ต้องการ โดยอาจเขียนเป็นจุดประสงค์เชิงพฤติกรรมหรือกำหนดลักษณะเฉพาะของมวลความรู้
หรือพฤติกรรมหรือเป็นการกำหนดประชากรข้อสอบ หรือมวลข้อสอบ ในลักษณะเช่นนี้จึงทำให้
เกิดเทคนิคการเขียนข้อสอบแบบอิงเกณฑ์ขึ้น โอนีลี (Hively, 1974) ได้เสนอ เทคนิค
การเขียนข้อสอบจากการกำหนดลักษณะเฉพาะของข้อสอบ ซึ่งจะบรรยายถึงรูปแบบ และ
เหตุการณ์ที่ข้อสอบกำหนด การกำหนดลักษณะเฉพาะของข้อสอบเรียกว่า " ฟอรัมข้อสอบ "
(Item Forms) กลุ่มข้อสอบที่เป็นตัวแทนของประชากรข้อสอบ ซึ่งสร้างขึ้นจากฟอรัม
ข้อสอบนี้ จะสามารถวัดโดเมนของเนื้อหาได้อย่างครอบคลุม ฟอรัมข้อสอบ รอดด์ และ
อาลาไดนา (Roid and Haladyna, 1982) ได้นิยามเทคนิคในโลยีการเขียนข้อสอบไว้ว่า
" เทคนิคในโลยีการเขียนข้อสอบ เป็นเทคนิคการกำหนดเนื้อหา ที่จะสอบวัดให้ครอบคลุมเข้า
ประสงค์ของการเรียนการสอนนั่นเอง " ตามคำนิยามดังกล่าว จะเห็นว่าหากกำหนดเนื้อหาที่
จะสอบไว้ได้ละเอียด และชัดเจนความเข้าประสงค์ของการเรียนการสอนแล้วจะช่วยให้การ
เขียนข้อสอบง่าย และถูก "Spec" ยิ่งขึ้น

มหาวิทยาลัยแคลิฟอร์เนีย เมืองลอสแอนเจลิส (UCLA) ภายใต้การนำของ โฟแฟม (Popham, 1978 อ้างอิงใน ชุคคักดิ์ นัมภลิต 2529) ได้ตั้งสถาบันผู้เชี่ยวชาญทางด้าน CRITERION REFERENCED TEST ชื่อ Institute of Objectives Exchange (IOX) เพื่อเป็นศูนย์รวบรวมจุดประสงค์เชิงพฤติกรรมของกระบวนการเรียน การสอนหลายวิชา แต่ไม่สามารถบรรลุเป้าหมายได้เพราะจุดประสงค์เชิงพฤติกรรมของแต่ละข้อเป็นข้อความที่กำหนดไว้อย่างหลวม ๆ และไม่ได้กำหนดแนวทางเฉพาะให้กับผู้เขียน ข้อสอบอย่างชัดเจน ทางสถาบัน IOX จึงดัดแปลง Item Form ของ Hively และ คณะ เป็นเทคนิค Amplified Objective และ Test Specification

สงข ลักษณ์ (2525) ได้ดัดแปลง Test Specification มาใช้เป็นแนว ในการกำหนด "ลักษณะเฉพาะของข้อสอบ" (Item Specification) ซึ่งมีส่วนประกอบ ที่สำคัญ 5 ส่วน ดังนี้

1. พฤติกรรมที่ต้องการวัด ส่วนนี้จะระบุพฤติกรรมหลักที่วิเคราะห์จากรายวิชา ที่จะสร้างข้อสอบ
2. พฤติกรรมย่อย ส่วนนี้จะระบุพฤติกรรมเฉพาะที่แตกมาจากพฤติกรรมหลัก
3. คำอธิบาย ส่วนนี้จะเขียนขยายพฤติกรรมย่อยโดยเขียนในรูปจุดประสงค์เชิง พฤติกรรมที่ครอบคลุมสถานการณ์เงื่อนไขอย่างละเอียด และมีพฤติกรรมคาดหวังที่สังเกตได้
4. ลักษณะคำถามและคำตอบ ส่วนนี้จะกำหนดฟอร์มข้อสอบทั้งส่วนที่เป็นคำถาม และคำตอบ ส่วนที่เป็นคำถามจะกำหนดสถานการณ์ เงื่อนไข และคำสั่งที่จะให้ผู้สอบทำ
5. ตัวอย่างข้อสอบ ส่วนนี้จะช่วยให้เข้าใจคำอธิบายชัดเจนยิ่งขึ้น

จุดเด่นของการกำหนดลักษณะเฉพาะของข้อสอบ คือจะช่วยขจัดความยุ่งยากของ การเขียนข้อสอบตามจุดประสงค์เชิงพฤติกรรมที่มีอยู่จำนวนมาก ต่อหนึ่งจุดประสงค์การสอน จุดประสงค์เหล่านี้ไม่สามารถนำมารวมกันและตีความหมายได้อย่างชัดเจนเพราะว่านักเรียน ทำข้อสอบในบางจุดประสงค์เชิงพฤติกรรมได้ แต่กลับทำข้อสอบไม่ได้ในบางจุดประสงค์เชิง พฤติกรรม นอกจากนั้นบางเนื้อหาสามารถเขียนจุดประสงค์เชิงพฤติกรรมได้มาก บางเนื้อหา ก็เขียนจุดประสงค์เชิงพฤติกรรมได้น้อย หรือเขียนไม่ได้เลย จุดประสงค์เชิงพฤติกรรม เหล่านี้จึงไม่สามารถบ่งชี้ถึงผลสัมฤทธิ์ในเนื้อหาวิชาในลักษณะรวมได้ จุดอ่อนนี้สามารถแก้ไข ได้ โดยใช้การกำหนดลักษณะเฉพาะของข้อสอบอันเป็นการจัดระเบียบ หรือสร้างกฎเกณฑ์ สำหรับการเขียนข้อสอบให้รัดกุมรอบคอบ เด่นชัด สมบูรณ์ ด้วยคุณภาพต่างๆ เช่น มีเหตุผล เป็นปรนัย วัดในสิ่งที่ต้องการวัด ผู้วิจัยจึงใช้ การกำหนดลักษณะเฉพาะของข้อสอบเป็น แนวทางในการเขียนแบบสอบอิงเกณฑ์ที่ใช้ในการวิจัยครั้งนี้

ความยาวของแบบสอบอิงเกณฑ์

ไคลน์ และ โคเซคอฟ (Klein and Kosecoff, 1972) ทำการสำรวจงานด้านวัดผลอิงเกณฑ์เกี่ยวกับการกำหนดจำนวนข้อสอบที่จะทำให้การวัดผล และการตัดสินผู้รอบรู้ และไม่รอบรู้ให้ถูกต้องแม่นยำ หรือมีความคลาดเคลื่อนน้อยนั้นควรใช้ข้อสอบจำนวน 3-5 ข้อ ในแต่ละจุดประสงค์

เบอร์ก (Berk, 1980) ได้ศึกษาวิจัยถึงความคลาดเคลื่อนประเภทที่ I (Type I Error) คือ ผู้รอบรู้แต่ถูกตัดสินให้เป็นผู้ไม่รอบรู้ (False non-mastery) และความคลาดเคลื่อนประเภทที่ II (Type II Error) คือ ผู้ไม่รอบรู้ แต่ถูกตัดสินให้เป็นผู้รอบรู้ (False mastery) โดยสรุปได้ว่า ถ้าใช้ข้อสอบ จำนวน 15 ข้อ ในแต่ละจุดประสงค์กับการส่งผลย้อนกลับทันที ที่สอบเสร็จ จะทำให้การตัดสินถูกต้องแม่นยำ แต่ถ้าใช้ 2-3 ข้อ ในแต่ละจุดประสงค์จะเกิดความคลาดเคลื่อนในการตัดสินมาก และ เบอร์ก (Berk) ได้เสนอแนะว่า การกำหนดจำนวนข้อสอบสำหรับการประเมินผลความก้าวหน้า (Formative Evaluation) หรือการประเมินผลเพื่อจัดตำแหน่ง (Placement) นั้น ควรใช้ข้อสอบจำนวน 5-10 ข้อ แต่ถ้าเป็นการประเมินผลรวมสรุป (Summative Evaluation) หรือการตัดสินแบบถาวร (Permanent Decision) จำเป็นต้องใช้ข้อสอบจำนวนมากขึ้นเป็นสองเท่า นั่นคือ จำนวน 10-20 ข้อ

โนวิก และ เลวิส (Novick and Lewis in Harris, 1974) กล่าวว่าแบบสอบอิงเกณฑ์ที่มีความยาวมากที่สุดไม่ควรเกิน 20 ข้อ สำหรับแบบสอบวัดความสามารถ และ เมื่อนิชิซึ ตันทวนิช (2528) ได้ตรวจสอบความคลาดเคลื่อนในการจำแนกระดับความรอบรู้ของแบบสอบที่ความยาว 10 ข้อ 14 ข้อ 20 ข้อ 25 ข้อ 28 ข้อ และ 30 ข้อ กับแบบสอบเต็มฉบับที่ความยาว 72 ข้อ ในขอบเขตเนื้อหา "การวิเคราะห์ข้อสอบ" กับนักศึกษาลัทธิสุทธาระดับชั้นมัธยมศึกษาชั้นสูง พบว่าแบบสอบที่มีความเหมาะสมที่สุดในการจัดประเภทบุคคลตามระดับความรอบรู้ คือ แบบสอบที่ประกอบด้วยจำนวนข้อ 20 ข้อ ซึ่งมีความคลาดเคลื่อนเพียงร้อยละ 9.76 ถึงแม้จะเพิ่มความยาวของแบบสอบเป็น 25 ข้อ 28 ข้อ และ 30 ข้อ ความคลาดเคลื่อนที่ปรากฏมีความคงที่ คือ เกิดขึ้นร้อยละ 9.15 9.84 และ 8.54 ตามลำดับ เมื่อนิจารณาแล้วความคลาดเคลื่อนดังกล่าวไม่แตกต่างไปจากแบบสอบซึ่งมีความยาว 20 ข้อ ซึ่งเป็นข้อค้นพบที่มีความสอดคล้องกับ โนวิก และ เลวิส

การคัดเลือกข้อกระทบง

ลักษณะของข้อกระทบงที่มีน้ำหนักมากที่สุดในการคัดเลือก คือ ความสอดคล้องระหว่างข้อกระทบงกับจุดประสงค์ ค่าความยาก และอำนาจจำแนก การพิจารณาอาจจะแยกจากกัน หรือรวมกันก็ได้ โดยทั่วไปข้อกระทบงที่มีความเหมาะสมตามเกณฑ์ที่กำหนดไว้ก็จะนำไปรวมเป็นแบบสอบ อย่างไรก็ตามเมื่อข้อกระทบงมีเกณฑ์ไม่เหมาะสม การพิจารณาจะยังคงไว้ ปรับปรุง หรือตัดทิ้ง มีรายละเอียดที่จะพิจารณาดังต่อไปนี้

1. ความสอดคล้องระหว่างข้อกระทบงกับจุดประสงค์ เป็นส่วนสำคัญของประสิทธิภาพของแบบสอบ และมีประโยชน์ในการแปลผล ข้อกระทบงที่ไม่สอดคล้องกับจุดประสงค์ไม่ควรนำมาไว้ในแบบสอบ

2. ความยากของข้อกระทบง ข้อกระทบงของแบบสอบอิงเกณฑ์ควรมีความยากก่อนการเรียนการสอนมากกว่าความยากหลังการเรียนการสอน มีนักวัดผลบางท่านไม่ถือเรื่องความยากของข้อกระทบงในแบบสอบอิงเกณฑ์เป็นสำคัญ เช่น บลุม (Bloom 1971 อ้างถึงใน สำเร็จ บุญเรืองรัตน์, 2527) กล่าวว่า ค่าความยากง่ายของแบบสอบอิงเกณฑ์มิได้ถือเป็นเรื่องสำคัญ สิ่งสำคัญ คือ เขียนคำถามให้สอดคล้องกับจุดประสงค์เชิงพฤติกรรม สมคักดี สินธุรเวชญ์ (2521) และ สงบ ลักษณะ (2523) ได้กล่าวในทำนองเดียวกันว่าความยากง่ายของข้อสอบอิงเกณฑ์นั้นขึ้นอยู่กับเนื้อหาวิชา และจุดมุ่งหมายเชิงพฤติกรรม ถ้าสิ่งที่ต้องการวัดง่าย ข้อสอบก็จะง่าย ถ้าสิ่งที่ต้องการวัดยากข้อสอบก็จะยากเช่นกัน

3. อำนาจจำแนก ข้อกระทบงควรมีความตรงต่อการจำแนกผู้รอบรู้ และ ไม่รอบรู้ หรือสามารถบ่งชี้ได้ว่าผู้สอบเป็นผู้รอบรู้ หรือไม่รอบรู้ ดัชนีค่าอำนาจจำแนกของแบบสอบอิงเกณฑ์นั้น โฟแฟม และฮูเซ็ค (Popham and Husek, 1969 อ้างอิงใน สำเร็จ บุญเรืองรัตน์ 2527) ได้แสดงความเห็นว่า ค่าอำนาจจำแนกที่เป็นลบแสดงว่าข้อสอบมีข้อบกพร่อง ส่วนค่าอำนาจจำแนกที่ค่าบวกสูง แสดงว่า การสอนยังไม่ได้ผลดี ถ้าค่าอำนาจจำแนกเป็นบวก และมีค่าต่ำใกล้ ๆ ศูนย์แล้วแสดงว่าการสอนได้ผลดี คริสปีน และ เฟรดอเชน เสนอดัชนี S เรียกว่า Sensitivity Index (ความไวในการเรียนการสอน) เพื่อวิเคราะห์ข้อสอบอิงเกณฑ์ ข้อสอบอิงเกณฑ์ที่ต้องการควรมีค่าดัชนีความไวอยู่ระหว่าง .00 - 1.00

จากที่กล่าวมานี้ ผู้วิจัยจึงใช้ลักษณะทั้งสามข้อนี้ ประกอบการคัดเลือกข้อกระทบง โดยเน้นความสอดคล้องระหว่างข้อกระทบงกับจุดประสงค์เป็นอันดับแรก

คุณภาพของแบบสอบอิงเกณฑ์

1. การวิเคราะห์ความตรงของข้อสอบอิงเกณฑ์

1.1 ความตรงเชิงเนื้อหา (Content Validity) หมายถึง ความสอดคล้องระหว่างพฤติกรรมที่ข้อสอบวัดได้ กับพฤติกรรมที่ระบุไว้ในจุดประสงค์ ซึ่งจะใช้ผู้เชี่ยวชาญในเนื้อหาตัดสินผลว่า ข้อกระทงแต่ละข้อสอดคล้องกับจุดประสงค์ที่กำหนดไว้หรือไม่ โดยกำหนดวิธีการให้คะแนนผลการตัดสินไว้ดังนี้

- +1 หมายถึง แน่ใจว่าข้อกระทงนั้นวัดจุดประสงค์นั้น
- 0 หมายถึง ไม่แน่ใจว่าข้อกระทงนั้นวัดจุดประสงค์นั้น
- 1 หมายถึง แน่ใจว่าข้อกระทงนั้น ไม่ได้ วัดจุดประสงค์นั้น

จากผลการตัดสิน และวิธีการให้คะแนนดังกล่าว นำมาคำนวณหาดัชนีความสอดคล้องของข้อกระทงแต่ละข้อกับจุดประสงค์ (Index of Item Objective Congruence) ซึ่งโรวินเนลลี และแฮมเบิลตัน (Rovinelli and Hambleton, 1976) เสนอไว้ดังนี้

$$I_{io} = \frac{(M - 1) S - S'}{2N(M - 1)}$$

เมื่อ	I_{io}	คือ	ดัชนีความสอดคล้องระหว่างข้อกระทง i กับจุดประสงค์ o
	M	คือ	จำนวนจุดประสงค์ที่วัด
	N	คือ	จำนวนผู้เชี่ยวชาญที่เป็นผู้ตัดสิน
	S_o	คือ	คะแนนรวมจากผู้ตัดสินทั้งหมดที่ให้กับจุดประสงค์ o
	S'_o	คือ	คะแนนรวมจากผู้ตัดสินทั้งหมดที่ให้กับจุดประสงค์ทั้งหมดยกเว้นจุดประสงค์ o

และอีกสูตรหนึ่งของ Rovinelli และ Hambleton ซึ่งเป็นสูตรดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ (IOC) เช่นกัน แต่ใช้คำนวณเป็นรายจุดประสงค์ โดยมีสูตรคำนวณดังนี้ (บุญเชิด วิทยุโณนันต์พงษ์ 2527 อ้างถึง Hambleton and others, 1978)

$$IOC = R / N$$

เมื่อ	IOC	คือ	ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์
	R	คือ	ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญ
	N	คือ	จำนวนผู้เชี่ยวชาญที่ใช้พิจารณาตัดสินทั้งหมด

การแปลความหมาย ดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ ถ้าดัชนีที่คำนวณได้มีมากกว่า หรือเท่ากับ .5 แสดงว่าข้อกระทงวัดหรือเป็นตัวแทนจุดประสงค์เชิงพฤติกรรมข้อนั้น ถ้าค่าดัชนีน้อยกว่า .5 แสดงว่าข้อกระทงข้อนั้นไม่วัดหรือไม่เป็นตัวแทนจุดประสงค์เชิงพฤติกรรมข้อนั้น คัดเลือกข้อกระทงที่มีความตรงเชิงเนื้อหาไว้ ข้อกระทงที่ขาดความตรงเชิงเนื้อหาจะถูกคัดออกไปหรือไม่ก็ปรับปรุงแก้ไขข้อกระทงใหม่

1.2 ความตรงเชิงจำแนก (Discriminant Validity) คือ

ความตรงของแบบสอบอิงเกณฑ์มีรากฐานอยู่บนแนวคิดเกี่ยวกับทฤษฎีของการวัดอิงเกณฑ์ที่ว่า เครื่องมือที่ใช้ควรเป็นเครื่องมือที่สามารถจะแยกผู้เรียนแล้ว และผู้ที่ยังไม่ได้เรียนออกจากกันได้ เช่น ถ้าเครื่องมือที่ใช้ คือแบบสอบ เมื่อนำแบบสอบไปทดสอบคนสองกลุ่ม คือกลุ่มผู้ที่ยังไม่เคยเรียนเนื้อหาในแบบสอบ และกลุ่มที่เคยเรียนมาแล้ว ในกลุ่มหลัง ควรจะทำข้อสอบได้มากกว่าผู้สอบในกลุ่มแรก จากความคิดนี้ ความตรงในความหมายของการวัดอิงเกณฑ์ ก็คือ ความตรงในการตัดสินผู้สอบนั่นเอง การหาความตรงในลักษณะนี้ เครเฮน (Crehan, 1974 อ้างถึงใน บุญเชิด วิทยุอนันต์พงษ์, 2527) ได้เสนอวิธีการนี้

$$C = (U/N_1) - (L/N_2)$$

เมื่อ	C	เป็นดัชนีความตรงของแบบสอบ
	U	เป็นจำนวนผู้ตอบข้อสอบถูกในกลุ่ม N_1
	L	เป็นจำนวนผู้ตอบข้อสอบถูกในกลุ่ม N_2
	N_1	เป็นจำนวนผู้สอบในกลุ่มที่ได้รับการสอน
	N_2	เป็นจำนวนผู้สอบที่ไม่ได้รับการสอน

ซึ่งได้ปรับมาจาก สูตรหาความเที่ยงของแบบสอบอิงเกณฑ์ของ Carver (1970) มาใช้ประมาณค่าความตรงของแบบสอบอิงเกณฑ์ Hambleton and others (1978) เรียกว่า วิธีนี้ว่า การหาดัชนีจำแนกของข้อสอบการประมาณค่าความตรงตามสูตรนี้ อาศัยเทคนิคการ

ทดลองเชิงประจักษ์ จากผลการสอบก่อนสอนและหลังสอนแล้วนำจำนวนนักเรียนที่มีคะแนน
 ลอผ่าน หรือตก (รอบรู้ หรือไม่รอบรู้) มาแจกแจงในตาราง 2 x 2 ดังนี้

	ก่อนสอน	หลังสอน
ผ่าน	b	a
ตก	c	d

$$\text{ประมาณค่าความตรง} = (a + c) / N$$

$$\text{เมื่อ } N = a + b + c + d$$

ถ้านำมาวิเคราะห์ความสามารถในการจำแนกผู้เรียน ระหว่างก่อนสอนและ
 หลังสอน โดยพิจารณาเป็นรายข้อ เมื่อ การตอบถูกหมายถึง รอบรู้หรือผ่านและ ตอบผิด
 หมายถึง ไม่รอบรู้หรือตก นำจำนวนนักเรียนตอบถูก-ผิด มาแจกแจงในตาราง 2x2 ได้
 ดังนี้

	ก่อนสอน	หลังสอน
ถูก	b	a
ผิด	c	d

$$\text{ประมาณค่าความตรง} = (a + c) / N$$

จากสูตรที่เสนอโดย Crehen นี้เป็นการแสดงสัดส่วนการจำแนกถูกกับจำนวน
 ผู้สอบทั้งหมด การจำแนกถูกนั้น หมายถึง ก่อนการเรียนการสอนผู้สอบตอบข้อนั้นผิด และ
 หลังการเรียนการสอนผู้สอบจะตอบข้อนั้นถูก นั่นคือ ถ้าข้อใดที่ผู้สอบตอบผิดมากในเหตุการณ์
 ก่อนการเรียนการสอน และหลังการเรียนการสอนแล้วตอบถูกมาก ข้อนั้นจะแสดงถึงความ
 สามารถในการจำแนกผู้เรียน หรือมีความตรง ค่าความตรงที่ได้มาจากการคำนวณตามสูตร
 นี้ ไม่ควรน้อยกว่า .50 จากเหตุผลที่ค่าความตรง .50 นั้น แสดงว่า ผลการสอบก่อน
 สอนและหลังสอนมีผู้ตอบถูกจำนวนเท่ากัน ซึ่งเป็นความสามารถในการจำแนกต่ำ แต่ใน



ลักษณะการวัดอิงเกณฑ์ก็ถือว่า มีความตรงใช้ได้

นอกจากนี้ยังสามารถคำนวณได้ด้วยแนวคิดอื่น ๆ ดังต่อไปนี้

คริสปิ่น และ เฟรดโฮเซน (Kryspin and Feldhusen ,1974 อ้างถึงใน โกวีท ประมวลพฤษ์ ,2523) ได้เสนอดัชนีความไว (Sensitive Index) ซึ่ง เป็นความแตกต่างความยากก่อนสอนกับหลังสอน เช่นกัน โดยมีสูตร

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

- เมื่อ S คือ ดัชนีความไว
- R_{pos} คือ จำนวนนักเรียนที่ตอบถูกหลังการเรียนการสอน
- R_{pre} คือ จำนวนนักเรียนที่ตอบถูกต้องก่อนการเรียนการสอน
- T คือ จำนวนนักเรียนทั้งหมด

ข้อกระทงที่ต้องการ คือ ข้อสอบที่มีค่าดัชนีความไวสูงและเป็นบวก (ระหว่าง 0.00 - 1.00) ซึ่งแสดงว่าข้อสอบนั้นสามารถจำแนกผู้เรียนแล้ว และผู้ที่ยังไม่ได้เรียนออกจากกัน

คุณภาพของแบบสอบอิงเกณฑ์ในด้านนี้มีผู้ศึกษาไม่มากนัก จึงไม่ยุ่งยากซับซ้อน เหตุผลส่วนหนึ่ง เนื่องมาจากดัชนีที่ใช้คัดเลือกข้อกระทงในการวิเคราะห์รายข้อ ก็เป็นตัวบ่งชี้ถึงความตรงโดยส่วนรวมของข้อกระทงเป็นข้อ ๆ อยู่แล้วก็อาจเป็นไปได้

แบบสอบอิงเกณฑ์ที่ผู้วิจัยได้พัฒนาขึ้น เพื่อเป็นเครื่องมือในการวิจัยครั้งนี้ ตรวจสอบความตรงเชิงจำแนกของแบบสอบทั้งฉบับ โดยยึดประเด็นที่ว่า แบบสอบอิงเกณฑ์มีความตรงเชิงจำแนก ก็ต่อเมื่อแบบสอบนั้นสามารถจำแนกผู้เรียนหรือออกจากผู้ไม่เรียนรู้ กล่าวคือ ผลสัมฤทธิ์ทางการ เรียนของผู้เรียนที่ได้จากการสอบหลังการเรียนเน้นขึ้น จากผลสัมฤทธิ์ทางการเรียนก่อนการสอน อย่างมีนัยสำคัญทางสถิติ ซึ่งสามารถคำนวณได้จากสูตร (Guilford, 1978)

$$t = \frac{\bar{X}_d}{\sqrt{\frac{\sum x^2_d}{N(N-1)}}$$

2. ความเที่ยง (Reliability) ของแบบสอบอิงเกณฑ์

ความเที่ยงของแบบสอบอิงเกณฑ์ หมายถึง คุณภาพของคะแนนสอบ (Observed Scores) ที่ได้จากแบบสอบควรจะเท่ากับ หรือใกล้เคียงกับคะแนนที่แท้จริง (True Scores) ของผู้สอบที่ประมาณได้ นั่นคือ การทดสอบที่ปราศจากความคลาดเคลื่อนของการวัด (Error of Measurement) (เขาวดี วิบลัยศรี , 2526)

เบอร์ก (Berk, R.A., 1980) ได้สรุปมโนทัศน์ของความเที่ยงของแบบสอบอิงเกณฑ์ ไว้ 3 มโนทัศน์ คือ

1. ความเที่ยง เป็นความคงที่ในการตัดสินผู้สอบเป็นผู้รอบรู้ และไม่รอบรู้ โดยการสอบซ้ำด้วยแบบสอบฉบับเดียวกัน หรือสอบด้วยแบบสอบคู่ขนาน
2. ความเที่ยง เป็นความคงที่ของความแตกต่างกำลังสองของคะแนนแต่ละคนที่เบี่ยงเบนไปจากคะแนนจุดตัด โดยการทดสอบด้วยแบบสอบคู่ขนาน
3. ความเที่ยง เป็นความคงที่ของคะแนนของนักเรียนแต่ละคน จากการทดสอบด้วยแบบสอบคู่ขนาน

จากข้อสรุปความเที่ยงตามมโนทัศน์ของ เบอร์ก (Berk) จะเห็นว่ามโนทัศน์ที่ 1 และมโนทัศน์ที่ 2 เป็นลักษณะความเที่ยงของแบบสอบอิงเกณฑ์ ที่มีความเกี่ยวข้องกับคะแนนจุดตัดของแบบสอบ สำหรับการวิจัยครั้งนี้ผู้วิจัยได้ประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ประเภทเลือกตอบและตอบสั้นตามแนวมโนทัศน์ที่ 2 ซึ่งเป็นความคงที่ของความแตกต่างกำลังสองของคะแนนแต่ละคนที่เบี่ยงเบนไปจากคะแนนจุดตัด วิธีการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ตามมโนทัศน์นี้ 2 วิธีคือ

1. วิธีของลิฟวิงตัน (Livington) ซึ่งเกี่ยวข้องกับความแปรปรวนที่เกิดจากปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ (Pearson x item interaction)
2. วิธีของเบรนนอนและเคน (Brennan and Kane) ซึ่งเกี่ยวข้องกับความแปรปรวนที่เกิดจาก ผลของข้อสอบรวมกับปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ (item plus Pearson x item interaction)

แต่ได้มีนักวัดผลบางท่านได้ดัดแปลงสูตรการประมาณค่าความเที่ยง โดยอาศัยหลักการวิเคราะห์ความแปรปรวนของคะแนน จึงทำให้เกิดวิธีประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ตามรูปแบบที่คล้ายกันอีก 3 วิธี คือ

1. วิธีของโคเฮน คอกซ์ และวอลเบสเซอร์ (Cohen, Cox and Walbesser)
2. วิธีของโลเวตต์ (Lovett, 1975) ใช้การแจกแจงแบบทวินาม
3. วิธีของโลเวตต์ (Lovett, 1978) ใช้การวิเคราะห์ความแปรปรวน ของ

คะแนนแบบ 0 - 1

การประมาณค่าความเที่ยงตามวิธีของโคเอน คอกซ์ และวอลเบลเซอร์ ไม่ได้นำคะแนนจุดตัดมาใช้ในการคำนวณ จึงไม่เหมาะกับแบบสอบอิงเกณฑ์ ที่กำหนดคะแนนจุดตัดไว้ (บุญเชิด วิทยุโณนันทพงษ์, 2527) ผู้วิจัยจึงใช้การประมาณความเที่ยงของ โลเวตต์ (Lovett, 1978) ซึ่งได้เสนอสูตรในการประมาณค่าความเที่ยงของคะแนนของแบบสอบอิงเกณฑ์โดยอาศัยส่วนประกอบของการวิเคราะห์ความแปรปรวนของคะแนน ซึ่งคะแนนนั้นมีความเบี่ยงเบนจากคะแนนจุดตัด สัมประสิทธิ์ของความเที่ยงจึงเป็นความเบี่ยงเบนของคะแนนผู้สอบแต่ละคนจากคะแนนเกณฑ์ ยิ่งความเบี่ยงเบนมีค่ามากเท่าไร ความถูกต้องของการตัดสินให้ผู้สอบได้หรือตกก็ยิ่งมากขึ้น

การประมาณค่าความเที่ยงจากการวิเคราะห์ความแปรปรวน (ANOVA) ตามวิธีของโลเวตต์ (Lovett, 1978) มีสูตรดังนี้

$$r_{cc} = 1 - \frac{MS_e}{MS_p}$$

เมื่อ

$$MS_p = \frac{k \sum_{t=1}^n (x_t/k - c)^2}{n}$$

$$MS_e = \frac{\sum_{t=1}^n \sum_{j=1}^k (x_{tj} - c)^2 - n \sum_{j=1}^k (x_j/n - x_{tj}) - k \sum_{t=1}^n (x_t/k - c)^2}{(n-1)(k-1)}$$

r_{cc}	คือ	ค่าประมาณความเที่ยงของแบบสอบซึ่งใช้การสอบเพียงครั้งเดียว
MS_p	คือ	ผลรวมกำลังสองเฉลี่ยของคะแนนเนื่องจากบุคคล
MS_e	คือ	ผลรวมกำลังสองเฉลี่ยของคะแนนเนื่องจากความคลาดเคลื่อน
\bar{x}_{tj}	คือ	ค่าเฉลี่ยรวม
c	คือ	คะแนนเกณฑ์
k	คือ	จำนวนข้อสอบ
n	คือ	จำนวนผู้สอบ

สูตรการประมาณค่าความเที่ยงดังกล่าวข้างต้นจะเป็นค่าประมาณที่อคติ โลเวตต์ได้เสนอสูตรปรับแก้ไว้ดังนี้

$$r_{cc} \text{ (ANOVA Corrected)} = 1 - \frac{n(k-1)}{n(k-1)-2} \cdot \frac{MS_e}{MS_p}$$

4. แบบสอบปรนัยชนิดเลือกตอบและตอบสั้น

แบบสอบปรนัย (Objective Tests) เป็นแบบสอบที่ประกอบด้วยคุณลักษณะ 3 ประการ คือ (อำนาจ เลิศชัยนัต , 2533)

1. คำถามที่ใช้ถาม เป็นคำถามที่ชัดเจน ถามตรงจุด อ่านแล้วรู้ว่าถามอะไร
2. เกณฑ์การตรวจให้คะแนนได้กำหนดไว้ชัดเจนใครๆตรวจก็ต้องได้คะแนนตรงกัน
3. การแปลผลทุกคนที่แปลผลได้ตรงกัน เช่น ใครทำข้อสอบได้ คือคนเก่ง
ใครทำข้อสอบไม่ได้ คือ คนอ่อน

ชนิดของข้อสอบปรนัยทั่วไปจะจัดแบ่งดังนี้

1. ชนิดเลือกตอบหลายตัวเลือก (Multiple Choices Type)
2. ชนิดเติมคำ (Completion Type)
3. ชนิดตอบสั้น (Short Response Type)
4. ชนิดจับคู่ (Match Type)
5. ชนิดถูกผิด (True - False Type)

นักการศึกษาหลายท่านได้ศึกษาเกี่ยวกับชนิดของแบบสอบปรนัยในลักษณะต่าง ๆ

ถึงนี้ มีเชลล์ (Michells, W.J., 1950) กล่าวถึง แบบสอบปรนัยชนิดเลือกตอบว่าเป็นแบบสอบที่มีค่ามากที่สุด ในบรรดาแบบสอบทั้งหลาย เพราะสามารถออกให้นักเรียนตีความวินิจฉัยเคราะห์ เลือกและนำสิ่งที่เรียนมาใช้ให้เป็นประโยชน์ และสามารถวัดความเข้าใจ การตัดสินใจ การวินิจฉัยเหตุผล และวัดความจำของนักเรียนได้อย่างกว้างขวางกว่าที่ใช้แบบสอบปรนัยชนิดเติมคำ ฮัตตี (Hatti, 1971) ได้ศึกษาเปรียบเทียบคะแนนความเที่ยงและความตรงของแบบสอบที่มีตัวเลือก และไม่มีตัวเลือก พบว่าคะแนน และความตรงของแบบสอบไม่พบความแตกต่างกันทางสถิติ แต่ค่าความเที่ยงของแบบสอบที่ไม่มีตัวเลือก มีค่าต่ำกว่าแบบสอบชนิดที่มีตัวเลือก อีเบล (Ebel, R.L, 1972) กล่าวถึง แบบสอบปรนัยชนิดเลือกตอบและตอบสั้น ว่าแบบสอบทั้งสองชนิดนี้มีความสัมพันธ์กันสูงมาก เมื่อทั้งสองแบบสอบวัดในสิ่งเดียวกัน และเป็นแบบสอบคู่ขนาน (Parallel Form) คือ ถ้านักเรียนสามารถเขียนคำตอบได้ถูกต้องก็น่าที่จะจำแนกตัวเลือกได้ถูกต้อง เช่นกัน แต่อย่างไรก็ตาม การจำแนกตัวเลือกน่าจะง่ายกว่าการสร้างคำตอบขึ้นมา เพราะโดยรูปแบบของแบบเลือก

ตอนมีการนำเสนอสิ่งเร้า และการตอบสนองให้เสรีจะง่ายกว่าการที่จะตอบถูก โดยรู้แต่สิ่งเร้า หรือคำถามเท่านั้น ซึ่งผู้สอบต้องใช้ความสามารถในการระลึกได้ (Recall) จึงทำให้ยากที่จะเดาคำตอบได้ถูกต้อง ถ้าจะกล่าวถึงประโยชน์ของแบบสอบทั้งสองแล้ว (Gronlund, N.E., 1976) แบบสอบแบบเลือกตอบสามารถวัดความรู้ที่มีประสิทธิภาพได้หลาย ๆ ชนิด และสามารถวัดผลการเรียนรู้ได้ละเอียดซับซ้อน ส่วนคำตอบสั้นนั้นสามารถตอบได้หลายแนวทาง แต่แบบเลือกตอบจะทำให้คำตอบของนักเรียนมีความเฉพาะเจาะจง สดแทนแล้ และ ฮอปกินส์ (Stanley and Hopkins, 1978) ได้เสนอว่าแบบสอบชนิดตอบสั้นเป็นแบบสอบที่มีคุณค่าและเหมาะสมเป็นพิเศษสำหรับวิชาคณิตศาสตร์ และฟิสิกส์ เพราะคำถามของวิชาเหล่านี้อยู่ในรูปของปัญหาที่ต้องการคำตอบ โดยการคำนวณ

สำหรับในประเทศไทย สมหวัง นิธิยานูวัฒน์ และเพ็ญศิริ ดำเนิน (2524) ศึกษาเปรียบเทียบคุณภาพของแบบสอบอิงกลุ่มแบบเลือกตอบกับแบบตอบสั้น พบว่า แบบสอบทั้งสองประเภทมีความตรงร่วมสมัยไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนค่าความเที่ยงของแบบสอบตอบสั้นสูงกว่าแบบสอบแบบเลือกตอบอย่างมีนัยสำคัญ ในกรณีที่กลุ่มตัวอย่างมีพื้นฐานความรู้คณิตศาสตร์ และแบบสอบแบบเลือกตอบง่ายกว่าแบบสอบตอบสั้น และได้เสนอแนะว่าถ้าสร้างแบบสอบเลือกตอบให้มีคุณภาพสูงแล้ว แบบสอบเลือกตอบก็สามารถที่จะทำหน้าที่แทนแบบสอบตอบสั้นได้ โดยเฉพาะกลุ่มตัวอย่างที่มีความสามารถในระดับปานกลาง หรือต่ำกว่า แต่ถ้ากลุ่มตัวอย่างมีความรู้ความสามารถในระดับสูงกว่าปานกลาง ถ้าจะใช้แบบสอบแบบเลือกตอบต้องเป็นแบบสอบชนิดที่มีความแหลมคมจริง ๆ ถ้ายังไม่เชี่ยวชาญในการสร้างแบบสอบแบบเลือกตอบควรใช้แบบสอบแบบตอบสั้น ซึ่งใช้ได้ผลดีในทุกกรณีแต่มีข้อเสียในเรื่องของการตรวจที่อาจต้องใช้เวลามากขึ้น กรองทอง เพนศิริอำนวย (2525) ศึกษาเปรียบเทียบคุณภาพของแบบสอบวิชาคณิตศาสตร์ชนิดเลือกตอบและเติมคำ พบว่า แบบสอบเลือกตอบง่ายกว่าแบบสอบเติมคำ และแบบสอบเลือกตอบมีค่าความเที่ยงน้อยกว่าแบบสอบชนิดเติมคำ สุรินทร์ แห่งจันทิก (2528) ศึกษาเปรียบเทียบคะแนนจุดตัดและความเที่ยงของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ระหว่างแบบสอบประเภทเลือกตอบและตอบสั้น จากการใช้ทฤษฎีการตัดสินใจของแกลส พบว่า แบบสอบเลือกตอบมีความเที่ยงน้อยกว่าแบบสอบตอบสั้น

จากรายงานวิจัยที่กล่าวมา จะเห็นว่า แบบสอบเลือกตอบและแบบสอบตอบสั้นนั้นแม้จะมีความตรงเชิงเนื้อหาเทียบเคียงกัน แต่แบบสอบทั้งสองประเภทก็ยังคงมีความแตกต่างในด้านอื่นที่น่าจะศึกษาคือไป

5. เทคนิคการวิเคราะห์ตามลำดับขั้น และงานวิจัยที่เกี่ยวข้อง

เทคนิคการวิเคราะห์ตามลำดับขั้นพัฒนา โดยนักคณิตศาสตร์ชื่อแอบราฮัม วอลด์ ในปี 1945 เทคนิคนี้มุ่งจะตรวจสอบเพื่อตัดสินว่าจะยอมรับสมมติฐานศูนย์ (Null Hypothesis) หรือ สมมติฐานอื่น (Alternative Hypothesis) ภายใต้โอกาสความคลาดเคลื่อนที่กำหนดไว้ล่วงหน้า โดยในการทดสอบจะมีการสุ่มตัวอย่างขึ้นมาทีละหน่วย แล้วพิจารณาว่าจากข้อมูลที่ได้ ควรจะตัดสินใจรับสมมติฐานใด ถ้ายังไม่สามารถจะตัดสินใจได้ ก็จะสุ่มตัวอย่างเพิ่มเข้าไปอีกทีละหน่วย เมื่อตัดสินใจยอมรับสมมติฐานตัวใดตัวหนึ่งได้ ก็จะยุติการทดสอบ จุดเด่นของหลักการนี้ คือ ต้องการให้กลุ่มตัวอย่างให้น้อยที่สุดในการตัดสินใจยอมรับสมมติฐานตัวใดตัวหนึ่ง โดยมีโอกาสเกิดความคลาดเคลื่อนในการตัดสินใจในระดับต่ำซึ่งจะกำหนดขึ้นไว้ล่วงหน้าก่อนที่จะดำเนินการทดสอบ

แนวคิดดังกล่าวนี้ได้พัฒนาขึ้นในโรงงานอุตสาหกรรม ในทางคณิตศาสตร์เป็นการมุ่งใช้เทคนิคการวิเคราะห์ตามลำดับขั้น เพื่อตรวจสอบค่ามัธยฐานเลขคณิตของตัวแปรที่มีลักษณะการแจกแจงเป็นแบบทวินาม

ในการสุ่มกลุ่มตัวอย่างมาทดสอบต้องกำหนดค่าพารามิเตอร์ (α , β , P_0 , P_1) ที่เหมาะสม โดยให้

	P_0	คือ	อัตราส่วนของข้อกระทงที่ยอมให้ผู้ทำผิดได้และจะถือเป็นผู้รอบรู้
	P_1	คือ	อัตราส่วนของข้อกระทงที่ถ้าผู้สอบทำผิดจะถือว่าเป็นผู้ไม่รอบรู้
จะยอมรับ	H_0	ก็ต่อเมื่อ	$P < P_0$ (สัดส่วนของข้อกระทงที่ทำผิดไม่เกินอัตรา P_0)
และยอมรับ	H_1	ก็ต่อเมื่อ	$P \geq P_1$ (สัดส่วนของข้อกระทงที่ทำผิดเกินอัตรา P_1)
ทั้งนี้กำหนดให้	α	เป็นโอกาสของความคลาดเคลื่อนในการปฏิเสธ	H_0 ที่เป็นจริง
	β	เป็นโอกาสของความคลาดเคลื่อนในการปฏิเสธ	H_1 ที่เป็นจริง
	$1 - \alpha$	เป็นโอกาสของความถูกต้องในการยอมรับ	H_0 ที่เป็นจริง
	$1 - \beta$	เป็นโอกาสของความถูกต้องในการยอมรับ	H_1 ที่เป็นจริง

สมมติฐานที่ทดสอบ

$$H_0 : P = P_0$$

$$H_1 : P = P_1$$

ในการทดสอบจะสุ่มข้อกระทงให้ x_i คือ ผลการทำข้อกระทงที่ i โดยให้ $x = 1$ ถ้าทำถูก และ $x = 0$ ถ้าทำผิด P คือ อัตราส่วนของข้อกระทงที่ทำผิดจากข้อกระทงที่ทดสอบ (x_1, x_2, \dots, x_m) โดยให้เป็น

$$P^{d_m} (1-P)^{m-d_m}$$

เมื่อ d_m คือจำนวนข้อกระทงที่ตอบผิด
 m คือจำนวนข้อกระทงที่ทดสอบ

ภายใต้สมมติฐาน $P = P_1$ ความน่าจะเป็นของอัตราส่วนที่ทำผิดจะเป็น

$$P_{1m} = P_1^{d_m} (1 - P_1)^{m-d_m}$$

ภายใต้สมมติฐาน $P = P_0$ ความน่าจะเป็นของอัตราส่วนที่ทำผิดจะเป็น

$$P_{0m} = P_0^{d_m} (1 - P_0)^{m-d_m}$$

การตรวจสอบแบบสอง m สามารถคำนวณได้ตามอัตราส่วนความน่าจะเป็น ดังนี้

$$\log \frac{P_{1m}}{P_{0m}} = d_m \log \frac{P_1}{P_0} + (m - d_m) \log \frac{1 - P_1}{1 - P_0}$$

จะทำการทดสอบต่อถ้า

$$\log \frac{\beta}{1 - \alpha} < \log \frac{P_{1m}}{P_{0m}} < \log \frac{1 - \beta}{\alpha}$$

จะสิ้นสุดการทดสอบ ถ้า

$$\log \frac{P_{1m}}{P_{0m}} \geq \log \frac{1 - \beta}{\alpha} \quad (\text{ปฏิเสธ } H_0)$$

$$\log \frac{P_{1m}}{P_{0m}} \leq \log \frac{\beta}{1 - \alpha} \quad (\text{ยอมรับ } H_0)$$

จากสมการข้างบน สามารถเขียนเป็นสมการในรูปที่ง่ายขึ้น

$$\begin{aligned}
 & \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + n \frac{\log \frac{1-P_0}{1-P_1}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} < d_n < \\
 & \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + n \frac{\log \frac{1-P_0}{1-P_1}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} \\
 d_n & \geq \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + n \frac{\log \frac{1-P_0}{1-P_1}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} \\
 \text{และ} & \\
 d_n & \leq \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + n \frac{\log \frac{1-P_0}{1-P_1}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}}
 \end{aligned}$$

การปฏิบัติอีกทางหนึ่ง ใช้การแปลความหมายจากเขตแดนที่ประกอบด้วยเส้นตรง 2 เส้นที่ขนานกัน คือ L_1 และ L_0 ทั้งนี้จะต้องกำหนดค่า P_0, P_1, α, β เพื่อนำไปคำนวณหา จุดตัดแกน Y และความชันของเส้น L_1 และ L_0 ซึ่งมีวิธีการดังนี้

$$S = \frac{\log \frac{1 - p_0}{1 - p_1}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}$$

$$h_0 = \frac{\log \frac{p}{1 - \alpha}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}$$

$$\text{และ } h_1 = \frac{\log \frac{1 - p}{\alpha}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}$$

จะได้ว่า $h_0 + ms < d_m < h_1 + ms$

$$L_1 = h_1 + ms \quad (\text{เส้นสมการที่แยกผู้ไม่รอบรู้ออกจากเขตทดสอบต่อ})$$

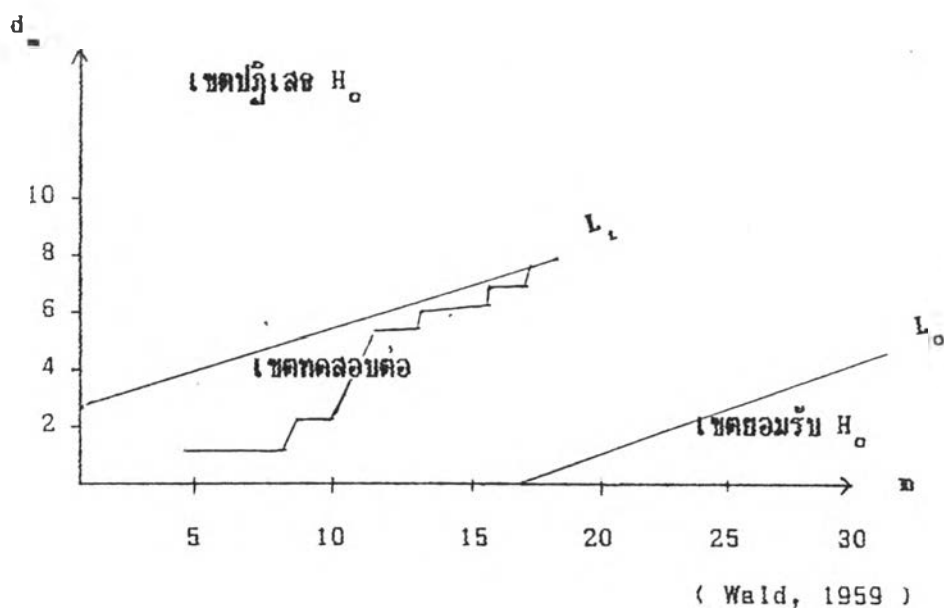
$$L_0 = h_0 + ms \quad (\text{เส้นสมการที่แยกผู้รอบรู้ออกจากเขตทดสอบต่อ})$$

โดยที่ถ้านำมาเขียนเป็นเส้นกราฟแล้ว h_1 และ h_0 จะเป็นจุดตัดกับแกนตั้ง และ s จะเป็นค่าความชันของเส้นตรงสองเส้น ซึ่งขนานกัน และการที่จะสร้างเส้นกราฟได้นั้น จะต้องกำหนดค่าต่าง ๆ ขึ้น 4 ค่า คือ ค่าพารามิเตอร์ เมื่อ H_0 เป็นจริง ค่าพารามิเตอร์เมื่อ H_1 เป็นจริง อัตราความคลาดเคลื่อนประเภทที่หนึ่ง และอัตราความคลาดเคลื่อนประเภทที่สอง

การเขียนเส้นกราฟ L_1 และ L_0 จะเริ่มต้นด้วยการเขียนค่าลำดับ (m, d_m) ซึ่งได้จากการแทนค่า m ลงในสมการข้างบน นำค่าลำดับที่ได้มาเขียนเส้นกราฟ

ในการทดสอบ ถ้ามีค่าลำดับใดอยู่บนเส้น L_0 หรือต่ำกว่า ก็จะยอมรับ H_0 แต่ถ้าอยู่บน L_1 หรือเหนือกว่าก็จะปฏิเสธ H_0 เส้นกราฟ L_1 และ L_0 สามารถเขียนได้ดังแผนภูมิที่ 1

แผนภูมิที่ 1 การสร้างเส้นกราฟของเทคนิคการวิเคราะห์ตามลำดับขั้นเพื่อพิจารณาตัดสินใจยอมรับสมมติฐานตัวใดตัวหนึ่ง



การประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นในการศึกษา

โคว์เดน (Cowden, 1946) ได้ทดลองใช้เทคนิคการวิเคราะห์ตามลำดับขั้นกับการสอบในชั้นเรียน โดยใช้นักศึกษา 10 คน ซึ่งมีเกรด A 1 คน เกรด B 1 คน เกรด C 5 คน เกรด D 1 คน และเกรด F 2 คน เรียนวิชาสถิติพื้นฐานที่มหาวิทยาลัยนอร์ธคาโรไลนา สอบแบบสอบถูก-ผิด ซึ่งประกอบด้วยคำถาม 200 ข้อ และใช้



เทคนิคการวิเคราะห์ตามลำดับชั้นมาตรวจสอบความยาวที่แบ่งออกเป็น 20 ข้อ 40 ข้อ 60 ข้อ และ 80 ข้อ ตามลำดับ โดยกำหนดเกณฑ์มาตรฐานให้ผ่าน (P_1) เป็น 0.30 และเกณฑ์มาตรฐานให้ตก (P_2) เป็น 0.40 อัตราความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตก (α) เป็น .20 และอัตราความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน (β) เป็น .10 ผลการทดลองปรากฏว่า นักเรียนที่มีเกรด A และเกรด F สามารถตัดสินผลได้เร็วกว่าคนที่มีความสามารถเรียนปานกลาง นอกจากนี้โค้วเตนยังตั้งคำถามเพื่อการศึกษาค้นคว้า 2 ประการ คือ

- (1) คำถามจำนวนกี่ข้อ จึงจะสามารถใช้ในการตัดสินความสามารถ
- (2) ผลที่ได้จากการตัดสินนั้น มีความถูกต้องเพียงใด

แพตเตอร์สัน (Patterson, 1962 Cited by Cleary, Linn and Rock, 1968) เปรียบเทียบรูปแบบการวิเคราะห์ตามลำดับชั้น กับการกระจายของประชากรในลักษณะต่าง ๆ เขาสรุปได้ว่า เทคนิคการวิเคราะห์ตามลำดับชั้นจะสามารถจำแนกนักเรียนที่มีความสามารถระดับสูงกับต่ำได้ถูกต้อง ด้วยข้อสอบจำนวนน้อยข้อ โดยไม่คำนึงถึงลักษณะการแจกแจงของข้อมูลในประชากร

เคลียร์ ลินน์ และรอก (Cleary, Linn and Rock, 1968) ศึกษาผลการจำแนกบุคคลตามระดับความรู้ กับนักเรียนเกรด 11 จำนวน 4,885 คน กลุ่มตัวอย่างจะถูกแบ่งเป็น 2 กลุ่ม กลุ่มแรกจะเป็นกลุ่มที่ใช้หาชนิดเกี่ยวกับพื้นฐาน ของกลุ่มตัวอย่างทั้งหมด (Original Sample) มีจำนวน 2,478 คน กลุ่มที่สองเป็นกลุ่มที่ไว้สำหรับตรวจสอบ (Cross Validation Sample) มีจำนวน 2,407 คน แยกระดับความสามารถออกเป็น 4 ระดับ ใช้วิธีการต่าง ๆ 4 วิธี คือ การทดสอบที่เป็นลักษณะสองขั้นตอน การจัดสอบจากข้อสอบที่มีค่าความยากแตกต่างกันมาก การใช้ผลความแตกต่างจากการสอบภายในกลุ่ม และการใช้เทคนิคการวิเคราะห์ตามลำดับชั้น จากการศึกษาพบว่า เทคนิคการวิเคราะห์ตามลำดับชั้น สามารถจำแนกบุคคลตามระดับความรู้ได้ถูกต้องสูงสุด

กรีน (Green 1971 Cited by Linn, Rock and Cleary 1972) พบว่า เทคนิคการวิเคราะห์ตามลำดับชั้น สามารถลดเวลาในการสอบลงไปได้ครึ่งหนึ่งจากเวลาที่ใช้ปกติ

ลินน์ , รอค และ เคลียร์ (Linn, Rock and Cleary , 1972) นำนาหลักการทางสถิติของเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้ในการศึกษาผลการตัดสินระดับความรอบรู้จากผลทดสอบของนักศึกษาระดับมหาวิทยาลัย จำนวน 4,840 คน ใน 3 รายวิชา คือ คณิตศาสตร์ จำนวน 75 ข้อ วิทยาศาสตร์ธรรมชาติและเรื่องความภาษาอังกฤษ วิชาละ 100 ข้อ นำผลการตัดสินระดับความรอบรู้จากเทคนิคการวิเคราะห์ตามลำดับขั้นเปรียบเทียบกับผลการสอบแบบปกติธรรมดา เมื่อกำหนดให้ผลการตัดสินมีความถูกต้องเท่ากัน

เทคนิคการวิเคราะห์ตามลำดับขั้นที่พัฒนาขึ้นนี้ได้นำค่าความยากที่ได้จากการสอบมารวมคำนวณ และตัดสินระดับความรอบรู้เพียง 2 ระดับ คือ รอบรู้ ไม่รอบรู้ ซึ่งมีวิธีการดังนี้

P_{n_i} คือ ค่าความยากของข้อสอบในกลุ่มสูง
 P_{l_i} คือ ค่าความยากของข้อสอบในกลุ่มต่ำ

ในการทดสอบ จะจัดเป็นผู้รอบรู้ และยุติการทดสอบเมื่อ

$$\Psi(n) = \sum_{i=1}^m \log R_i > A \quad (1)$$

และจะจัดเป็นผู้ไม่รอบรู้ และยุติการทดสอบเมื่อ

$$\Psi(n) = \sum_{j=1}^m \log R_j < -A \quad (2)$$

เมื่อ $R_i = \text{Phi} / P_{li}$ ถ้าผู้สอบทำข้อสอบข้อนั้น ๆ ถูก

$R_i = (1 - \text{Phi}) / P_{li}$ ถ้าผู้สอบทำข้อสอบข้อนั้น ๆ ผิด

A คือ ค่าคงที่ที่ยอมให้เกิดความคลาดเคลื่อน

n คือ จำนวนข้อสอบทั้งหมดที่ได้รับการทดสอบ

และถ้าผู้สอบทำข้อสอบ 50 ข้อ แล้วยังไม่สามารถตัดสินระดับความรอบรู้ได้ จะใช้เกณฑ์พิจารณา ดังนี้

$$\Psi (60) \geq 0 \quad \text{ถือว่าผู้สอบเป็นผู้รอบรู้}$$

$$\Psi (60) < 0 \quad \text{ถือว่าผู้สอบเป็นผู้ไม่รอบรู้}$$

ในการศึกษาของลินน์ , รอค และเคลียร์ ได้กำหนดค่า $A = 1.39, 2.30, 3.00$ และ 4.61 ผลการศึกษาปรากฏว่า การสอบแบบปกติธรรมดา จะต้องใช้จำนวนข้อสอบมากกว่าการสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น ประมาณ 2 เท่า ดังรายละเอียดในตารางที่ 2

ตารางที่ 2 เปรียบเทียบสัดส่วนจำนวนข้อสอบที่ต้องใช้ในการสอบปกติธรรมดากับการสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น เมื่อให้ผลความถูกต้องในการตัดสินระดับความรอบรู้มีจำนวนเท่ากัน

ค่าที่กำหนด	แบบสอบ		
	คณิตศาสตร์	วิทยาศาสตร์ธรรมชาติ	เรียงความภาษาอังกฤษ
1.39	2.17	2.67	2.08
2.30	2.35	1.97	1.95
3.00	2.69	1.79	1.84
4.61	2.33	1.65	1.71

(Linn, Rock and Cleary , 1972)

วิธีการที่ลินน์, รอค และเคลียร์ นั้นข้านั้นใช้ได้แม้ในกรณีที่ข้อสอบมีค่าความยากไม่เท่ากัน ทั้งนี้เพราะผลการสอบของข้อสอบแต่ละข้อไม่ว่าจะถูกหรือผิดจะส่งผลกระทบไปถึงโอกาสในการตัดสินระดับความรอบรู้ กล่าวคือ ถ้าตอบข้อสอบที่ยากได้ถูกต้อง ก็จะมีความเป็นไปได้สูงที่จะได้รับการตัดสินเป็นผู้รอบรู้ และถ้าตอบข้อง่ายผิด ก็จะมีความเป็นไปได้ที่จะได้รับการตัดสินเป็นผู้ไม่รอบรู้

ปี 1975 สเตคไลน์ และนิธยานุวัฒน์ (Stecklein and Pitiyenuwat) ได้ชี้ประเด็นความเหมาะสมในการนำเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้ทดสอบในระบบอิงเกณฑ์ สรุปได้ดังนี้

1. การทดสอบ โดยการใช้เทคนิคการวิเคราะห์ตามลำดับขั้น ในการทดสอบแบบอิงเกณฑ์ แม้จะใช้ข้อสอบเพียงจำนวนน้อยข้อ ข้อสอบเหล่านี้ก็มีความเป็นตัวแทนของเนื้อหาวิชาที่ต้องการทดสอบ ทั้งนี้เนื่องจากว่าแบบสอบอิงเกณฑ์จะมีการกำหนดขอบเขตของเนื้อหาให้มีความเป็นเอกพันธ์

2. การทดสอบ โดยการใช้เทคนิคการวิเคราะห์ตามลำดับขั้นจะสามารถลดเวลาในการสอบลงได้

3. ความยากของข้อสอบเป็นปัจจัยแรก ที่มีผลต่อการจำแนกบุคคลตามระดับความรอบรู้ได้อย่างถูกต้อง

นอกจากนี้ สเตคไลน์ และ นิธยานุวัฒน์ ได้กำหนดค่าพารามิเตอร์ที่เหมาะสมขึ้นทั้งหมด 7 แผน (ซึ่งได้เสนอไว้ในตารางที่ 1 ในบทที่ 1) จากแผนที่กำหนดขึ้น สเตคไลน์ และนิธยานุวัฒน์ ได้เลือกใช้แผน E ซึ่งมีค่าพารามิเตอร์ดังนี้ $P_0 = .25$, $P_1 = .45$, $\alpha = .05$ และ $\beta = .10$ นำไปทดลองใช้กับนักศึกษาจำนวน 24 คน แบบสอบที่ใช้มีจำนวน 30 ข้อ ผู้ที่สอบผ่านจะต้องได้คะแนน 24 คะแนนขึ้นไป เมื่อนำผลที่ได้มาเทียบกับการจำแนกระดับความรอบรู้โดยใช้เทคนิคการวิเคราะห์ตามลำดับขั้น ปรากฏผลดังนี้ มีผู้เรียนจำนวน 16 คน ที่เป็นผู้รอบรู้ เมื่อตัดสินโดยใช้เทคนิคการวิเคราะห์ตามลำดับขั้น ในจำนวนนี้มีอยู่ 13 คน ที่เป็นผู้รอบรู้ เมื่อใช้เกณฑ์จากคะแนนของแบบสอบเต็มฉบับ และมีผู้เข้าสอบ 1 คน มีผลการตัดสินจากเทคนิคการวิเคราะห์ตามลำดับขั้นเป็นผู้ยังตัดสินไม่ได้ แต่ผลการสอบจากแบบสอบเต็มฉบับเป็นผู้รอบรู้

สเตคไลน์, นิธยานุวัฒน์ และลิงก์ (Stecklein, Pitiyenuwat and Ling, 1980) นำเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้ในสถานการณ์การเรียนการสอนจริงกับนักศึกษา โดยกำหนดค่าพารามิเตอร์ดังนี้ $P_0 = .25$, $P_1 = .45$, $\alpha = .05$ และ $\beta = .10$ นักศึกษาแต่ละกลุ่มมีจำนวนตั้งแต่ 18 คน ถึง 152 คน ที่เรียนวิชาต่างๆ เช่น วิชาความสัมพันธ์ของมนุษย์ จิตวิทยาการศึกษา เป็นต้น เป็นต้น ผู้เข้าสอบจะทราบคำตอบของข้อสอบแต่ละข้อทันทีที่ตอบเสร็จจากกระดาษคำตอบที่ได้รับการออกแบบโดยเฉพาะ ผู้สอบจะได้รับการตัดสินว่าผ่าน ถ้าทำข้อสอบในช่วงความยาวต้น ๆ เป็นไปตามเกณฑ์ใดเกณฑ์หนึ่งดังต่อไปนี้

- ก. ทำข้อสอบ 8 ข้อแรกถูกต้อง
- ข. ทำข้อสอบ 11 ข้อแรกถูก 10 ข้อ
- ค. ทำข้อสอบ 14 ข้อแรกถูก 12 ข้อ
- ง. ทำข้อสอบ 17 ข้อแรกถูก 14 ข้อ
- จ. ทำข้อสอบ 20 ข้อแรกถูก 16 ข้อ

ถ้าผู้สอบทำข้อสอบได้ถูกต้องตามเกณฑ์ดังกล่าวแล้ว ก็จะได้รับรางวัลให้ทำแบบสอบต่อไปจนหมดทั้งฉบับ แต่ก็อาจยุติการสอบได้ถ้าผู้สอบต้องการ

จากการนำผลที่ได้มาศึกษาเทียบกับผลการตัดสินจากแบบสอบเต็มฉบับ พบว่าในรายวิชาแรกที่ใช้แบบสอบที่มีจำนวน 60 ข้อ สามารถตัดสินผู้สอบได้ถูกต้องที่มีความยาวต่างๆ ทั้ง 5 ช่วง เป็นจำนวน 68 คน จากคนทั้งหมด 152 คน และเกิดความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน 3 คน ในรายวิชาที่สองแบบสอบมี 60 ข้อ จำนวนผู้เข้าสอบ 53 คน เมื่อใช้เทคนิคการวิเคราะห์ตามลำดับขั้นเป็นเกณฑ์สามารถตัดสินให้ผู้สอบเป็นผู้รอบรู้ จำนวน 22 คน และเป็นผู้ไม่รอบรู้ 6 คน โดยไม่มีความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตกเกิดขึ้น และเกิดความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้ผ่านร้อยละ 3.6 ส่วนในรายวิชาที่สาม แบบสอบมี 40 ข้อ สามารถตัดสินใจให้ผู้สอบสอบผ่านได้ถึง 27 ใน 31 คน แต่อาจารย์ประจำวิชา มีความเห็นว่า แบบสอบในช่วงต้น ๆ ประกอบด้วยข้อสอบที่ง่ายจนเกินไป

นิคิษฐ์ ตัณฑวิช (2528) ได้ประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้น เพื่อกำหนดความยาวของแบบสอบอิงเกณฑ์ โดยมุ่งตรวจสอบหาช่วงความยาวของแบบสอบที่เกิดความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ ไม่แตกต่างไปจากอัตราความคลาดเคลื่อนที่กำหนดร้อยละ 5 พลวิชัยที่ใช้ในการศึกษา เป็นนักศึกษาวิทยาลัยศรีสุนทร จำนวน 282 คน แบบสอบที่ใช้ในการศึกษาเป็นแบบสอบอิงเกณฑ์ เรื่อง "การวิเคราะห์ข้อสอบ" มีจำนวน 72 ข้อ นำเทคนิคการวิเคราะห์ตามลำดับขั้นมากำหนดความยาวของแบบสอบออกเป็น 6 ช่วง คือ 10 ข้อ 14 ข้อ 20 ข้อ 25 ข้อ 28 ข้อ และ 30 ข้อ การวิเคราะห์ข้อมูลมุ่งตรวจสอบอัตราความคลาดเคลื่อนที่ปรากฏ กับอัตราความคลาดเคลื่อนที่กำหนดโดยการทดสอบทวินาม และใช้คะแนนจากแบบสอบเต็มฉบับเป็นตัวเกณฑ์กำหนดความถูกต้องในการจัดประเภทบุคคลตามระดับการรอบรู้ และ คะแนนจากแบบสอบในช่วงความยาวต่าง ๆ เป็นตัวอิงเกณฑ์ และ หาค่าสหสัมพันธ์เคนดอลล์ ทอ ระหว่างผลการจัดประเภทบุคคลตามระดับการรอบรู้จากแบบสอบเต็มฉบับ และแบบสอบช่วงความยาวต่าง ๆ

จากการศึกษา พบว่า การนำเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้กับแบบสอบ โดยกำหนดโอกาสความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตก และผู้ไม่รอบรู้สอบผ่านที่ระดับ 0.05 ทั้งสองกรณี กำหนดเกณฑ์มาตรฐานในการให้ผ่านที่ระดับ 0.20 และ กำหนดเกณฑ์มาตรฐานในการให้ตกที่ระดับ 0.40 จำนวนข้อสอบที่เหมาะสมในการจัดสอบ เพื่อจำแนกบุคคลตามระดับการรอบรู้ คือ จำนวน 20 ข้อ โดยความคลาดเคลื่อนที่ปรากฏสูงสุดไม่เกินร้อยละ 13.16 นอกจากข้อค้นพบนี้แล้ว ข้อมูลที่ได้จากการวิจัยยังพบว่า การประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้เป็นหลักในการจัดประเภทบุคคลนั้นยังสามารถช่วยลดความยาวของแบบสอบลงไปได้มากเช่นนี้ และยังมีประสิทธิภาพในการจัดประเภทบุคคลตามระดับการรอบรู้ไม่ต่างไปจากเมื่อใช้แบบสอบเต็มฉบับ ย่อมมีข้อดีในประเด็นของการลดค่าใช้จ่ายวัสดุอุปกรณ์ต่าง ๆ ที่จะต้องใช้ในการดำเนินการสอบและทั้งลดเวลาในการสอบลงได้

จากงานวิจัยที่นำเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้กับแบบสอบ จะพบว่า เทคนิคนี้สามารถลดจำนวนข้อสอบได้จริง แต่ก็ยังไม่สามารถระบุได้แน่ชัดว่า จำนวนข้อสอบเท่าใดจึงจะสามารถจำแนกระดับความรอบรู้ของบุคคลได้ถูกต้อง ปัจจัยอย่างหนึ่งที่มีผลต่อการจำแนกระดับความรอบรู้ด้วยเทคนิคการวิเคราะห์ตามลำดับขั้น คือ ความยากของแบบสอบ ซึ่งถ้ามาพิจารณาประกอบกับแบบสอบแต่ละชนิด จะพบว่า แบบสอบที่ต่างชนิดกันแม้จะมีความตรงเทียบเคียงกันแต่ความยากจะแตกต่างกัน จากลักษณะเช่นนี้ผู้วิจัยเห็นว่าชนิดของแบบสอบน่าจะส่งผลต่อการจำแนกระดับความรอบรู้ของบุคคล ดังนั้นในการศึกษาครั้งนี้จึงเปรียบเทียบความคลาดเคลื่อนของแบบสอบอิงเกณฑ์ชนิดเลือกตอบ และตอบสั้น ในการจำแนกระดับความรอบรู้ ด้วยการนำเทคนิคการวิเคราะห์ตามลำดับขั้นมาใช้เป็นเกณฑ์การจำแนกระดับความรอบรู้ที่มีความยาวต่างๆ แล้วนำผลที่ได้จากการตัดสินระดับความรอบรู้ด้วยการวิเคราะห์ตามลำดับขั้นมาเทียบกับ ผลการตัดสินระดับความรอบรู้จากแบบสอบเต็มฉบับของแบบสอบทั้งสองประเภท โดยจะมุ่งวิเคราะห์ความคลาดเคลื่อนที่เกิดขึ้นจากแบบสอบทั้งสองประเภท