



บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยจะได้กล่าวถึงวรรณคดีที่เกี่ยวข้องกับการวัดผลแบบอิงเกณฑ์ที่สำคัญ ๆ 6 ประเด็นด้วยกันคือ

1. แนวคิดพื้นฐานของการทดสอบอิงเกณฑ์
2. ความหมายและลักษณะของแบบสอบอิงเกณฑ์
3. การวิเคราะห์ข้อกระทงแบบอิงเกณฑ์
 - 3.1 ความสอดคล้องของข้อกระทงกับจุดประสงค์
 - 3.2 ค่าสถิติของข้อกระทง
 - 3.3 การคัดเลือกข้อกระทง
 - 3.4 การปรับปรุงข้อกระทง
4. ความเที่ยง
5. ความตรง
6. งานวิจัยที่เกี่ยวข้อง

แนวคิดพื้นฐานของการทดสอบอิงเกณฑ์

แนวคิดเกี่ยวกับการเรียนรู้ให้บรรลุจุดมุ่งหมายหรือเรียนเพื่อรู้ ซึ่งมีความเชื่อว่า ผู้เรียนทั้งหมดหรือเกือบทั้งหมดควรจะเรียนรู้ได้ในสิ่งที่เราจัดให้เขาเรียน ถ้ามีการจัดประสบการณ์และเวลาให้เหมาะสมกับความแตกต่างของผู้เรียน ในการวัดผลแบบอิงกลุ่มจะให้ข้อมูลที่บอกได้แต่เพียงว่า ผู้เรียนมีความสามารถมากหรือน้อยกว่าคนอื่น แต่ไม่บอกให้ทราบว่าผู้เรียนสามารถทำอะไรได้บ้าง ซึ่งข้อมูลประเภทหลังนี้ จะเป็นพื้นฐานสำหรับปรับปรุงการเรียนการสอนให้เหมาะสมกับผู้เรียนแต่ละคนได้ (กมล ภูประเสริฐ 2518: 7 อ้างจาก Tyler 1970 and Glazer and Nitko 1971)

การสอบระหว่างสอนเป็นการสอบถามจุดประสงค์เชิงพฤติกรรมที่วางไว้ ปัจจุบันจะเขียนเรียงกันจากเรื่องง่ายไปหาเรื่องยาก เรื่องที่สอนให้เกิดความรู้อีกก่อนที่จะไปสอนให้เกิดทักษะ

หน้าที่ของครูคือ ไม่ทำให้คนเก่งและคนอ่อนห่างกันออกไป แต่สอนให้คนอ่อนรอบรู้ได้ คือ พยายามให้คนอ่อนเก่งขึ้น การที่จะทำอย่างนี้ได้ต้องมีการประเมินผลเป็นระยะ นั่นก็คือ ครูต้องสอบวัดอยู่ตลอดเวลา (กมล ภูประเสริฐ 2520: 26-28) จากผลการวิจัยทั้งในประเทศและต่างประเทศต่างให้ผลลัพท์ออกมาตรงกันอย่างหนึ่งคือ เมื่อต้นเทอมนักเรียนเก่งและอ่อนมีความรู้ไม่ห่างกันมากนัก แต่พอถึงปลายเทอม ความรู้ของนักเรียนเก่งและอ่อนกลับห่างกันมากขึ้น สรุปได้ว่าการสอนจะให้ได้ ต้องสอนให้คะแนนที่ได้จากการสอบระหว่างสอนเข้าไปทางซ้าย ไม่ใช่แจกแจงอย่างโค้งงอ

บลูม (Bloom 1971: 45-46) มีความคิดว่า การสอนที่มีประสิทธิภาพสูง จะทำให้การกระจายของคะแนนผลสัมฤทธิ์ทางการเรียนเข้าไปทางซ้ายมือ ซึ่งมีความหมายว่า นักเรียนส่วนใหญ่สามารถสัมฤทธิ์ผลในเนื้อหาวิชานั้นเป็นไปตามจุดมุ่งหมายที่สำคัญของการจัดการศึกษา เพื่อให้ผู้เรียนส่วนใหญ่มี "การรอบรู้" ได้

การเรียนรู้ที่เป็นรากฐานของการวัดและประเมินผลแบบอิงเกณฑ์ ในทัศนะของการเรียนเพื่อรอบรู้ ผู้เรียนจะต้องเรียนรู้ทั้งหมดในสิ่งที่เรียน โดยต้องมีความสามารถถึงระดับที่คาดหวังไว้ กล่าวคือถึงมาตรฐานของเกณฑ์ที่กำหนดไว้ เนื่องจากแนวคิดเกี่ยวกับการเรียนเพื่อรอบรู้มีความคิดพื้นฐานอยู่ว่า คนปกติธรรมดาสามารถเรียนรู้เรื่องที่เรียนทั่ว ๆ ไปได้ แต่อัตราเร็วต่างกัน และด้วยวิธีการต่างกัน และการเรียนรู้เรื่องใดเรื่องหนึ่งต้องเป็นการเรียนรู้จนถึงมาตรฐานของเกณฑ์ที่กำหนดไว้ในจุดมุ่งหมาย ดังนั้นหลักการสำคัญของแนวคิดนี้จึงมีลักษณะดังนี้ (เอนก เพ็ชรอนุกุลบุตร 2522)

1. นักเรียนแต่ละคนมีความแตกต่างกัน ทั้งในแง่ความถนัดและความสามารถในการเข้าใจเรื่องราว ดังนั้น "อัตราเร็ว" การเรียนของแต่ละคนจะไม่เท่ากัน นักเรียนเกือบทุกคนจะสามารถเรียนรู้ จนถึงระดับ "รู้รอบ" ได้ในระดับเดียวกัน ถ้าให้เวลาแก่พวกเขาอย่างเพียงพอ และใช้วิธีการเรียน วิธีสอน และอุปกรณ์ให้สอดคล้องกับผู้เรียน

2. เนื่องจากผู้เรียนมีความแตกต่างกันในแง่ต่าง ๆ อีกมาก เช่น ในด้านเจตคติ แรงจูงใจไม่สัมฤทธิ์ แบบของการคิด ฯลฯ ดังนั้นการสอนทุกคนด้วยวิธีเดียวกัน ใช้วัสดุอุปกรณ์ชนิดเดียวกันย่อมจะไม่เป็นผลดีที่สุดสำหรับทุกคน บางคนเรียนได้ไม่ดี อาจเป็นผลจากวิธีสอน วัสดุอุปกรณ์ วิธีเรียนที่มอบหมายให้ไม่เหมาะสมแก่เขาก็ได้ ดังนั้นจึงควรจัดแบบของวิธีสอน เนื้อหาอุปกรณ์ให้เหมาะสมกับนักเรียน เพื่อให้แต่ละคนได้พัฒนาถึงระดับสูงสุด

3. กิจกรรมการเรียนของบุคคลมีการเปลี่ยนแปลงเข้าสู่ "การเรียนรู้อุป" ทั้งสิ้น และการตัดสินใจว่า "ใครเรียนรู้อุปแล้ว" จะต้องพิจารณาจากพฤติกรรมของบุคคลว่าสอดคล้องกับมาตรฐานที่ตั้งไว้หรือไม่

ดังนั้น การเรียนการสอนที่สอดคล้องหรือเป็นไปตามหลักการเรียนเพื่อรู้อุป จึงเป็น "การสอนแบบรายบุคคล" (Individualized Instruction) และการประเมินผลจากการเรียนการสอนต้องเป็นแบบที่เปรียบเทียบพฤติกรรมกับจุดมุ่งหมายหรือเกณฑ์ ซึ่งเป็นการวัดแบบอิงเกณฑ์

เนื่องจากการเรียนเพื่อรู้อุปหลักความคิดพื้นฐานอยู่ที่จัดสภาพการเรียนให้สอดคล้องกับแต่ละบุคคลซึ่งต้องใช้เวลาต่างกัน ใช้วิธีที่เหมาะสม และต้องกำหนดเกณฑ์ให้ชัดเจน แนวทางการเรียนเพื่อรู้อุปดังนี้

1. กำหนดจุดมุ่งหมายในการเรียนการสอนให้ชัดเจน โดยกำหนดเป็นจุดมุ่งหมายที่เฉพาะเจาะจง ในรูปแบบของจุดมุ่งหมายเชิงพฤติกรรม แล้วกำหนดมาตรฐานของการปฏิบัติที่จะบรรลุจุดมุ่งหมายเชิงพฤติกรรมนั้น

2. ใช้แนวทางการเรียนการสอนแบบ "การสอนรายบุคคล" (Individualized Instruction)

3. ใช้การประเมินผลการเรียนการสอน 2 แบบ คือ ประเมินผลย่อย (Formative Evaluation) เป็นการประเมินผลครั้งย่อย ๆ ตามจุดมุ่งหมายเชิงพฤติกรรมครั้งละไม่กี่ข้อ อาจจะประเมินเมื่อจบสิ้นการเรียนแต่ละหน่วยย่อย เพื่อพิจารณาว่าใครบรรลุแล้ว ใครยังไม่บรรลุ เพื่อนำไปสู่การตัดสินใจ เช่น การสอนซ่อมเสริม หรือการเรียนในหน่วยต่อไปซึ่งเรื่องอันดันดับต่อเนื่องกันไว้ จึงใช้การสอบย่อยนี้ สำหรับวินิจฉัยผลการเรียนได้

ความหมายและลักษณะของแบบลอบอิงเกณฑ์

การทดสอบอิงเกณฑ์ (Criterion-Referenced Testing) เป็นการทดสอบซึ่งแปลความหมายของผลการสอบโดยนำเอาผลการสอบของผู้สอบไปเทียบกับมาตรฐานที่แท้จริง (Absolute Standard) ว่าอยู่ในระดับใด ถึงมาตรฐานที่ยอมรับหรือไม่

แบบสอบอิงเกณฑ์เป็นแบบสอบที่ เกลเซอร์และเคลาส์ (Glaser and Klaus 1962) เริ่มนำมาใช้ในปี ค.ศ.1962 แต่ได้ใช้แนวคิดนี้มาตั้งแต่ปี ค.ศ.1949 แล้วในการประเมินผล การเรียน และมีนักทดสอบอีกจำนวนมากได้พัฒนาแนวคิดมาเรื่อย ๆ ได้มีการให้นิยามไว้แตกต่างกัน หรือใช้คำอื่นแทนคำว่า Criterion-referenced test เช่น

โพนแฟม : Objective-based tests

วิลสัน : Criterion-referenced testing

เชกเกอร์ : Objective-based assessment systems

นิตโก : Criterion-referenced score interpretations

ความหมายของแบบสอบอิงเกณฑ์ที่เป็นที่ยอมรับโดยทั่วไป คือ ของเกลเซอร์และนิตโก ได้ให้ความหมายไว้ว่า แบบสอบอิงเกณฑ์หมายถึง แบบสอบที่สร้างขึ้นอย่างมีนิตินัยเพื่อการวัดผล ซึ่งสามารถที่จะแปลความหมายได้โดยตรงตามมาตรฐานของความสามารถที่กำหนดไว้ และอีก ความหมายหนึ่งของ โพนแฟมและฮุเซค (Popham and Husek 1978) กล่าวไว้ว่า แบบสอบ อิงเกณฑ์ หมายถึง แบบสอบซึ่งใช้เพื่อวัดสภาพที่แน่นอนของแต่ละบุคคล โดยอาศัยเกณฑ์บางอย่าง เช่น มาตรฐานของความสามารถหรือพฤติกรรมที่ได้นิยามไว้อย่างดีแล้ว

แบบสอบอิงเกณฑ์ (Criterion-Referenced Test) เป็นแบบสอบที่บรรจุเนื้อหา สารที่เฉพาะเจาะจงสอดคล้องกับจุดประสงค์การเรียนการสอนและมีคะแนนจุดตัด (Cut-off score) หรือคะแนนเกณฑ์สำหรับใช้เป็นเครื่องตัดสินว่า ผู้สอบมีความรู้ตามเกณฑ์ที่กำหนดไว้หรือไม่ สาเหตุที่เกิดการวัดผลและการใช้แบบสอบอิงเกณฑ์ขึ้นมาก็เนื่องจากความต้องการข้อสนเทศ (Information) ที่ได้จากคะแนนซึ่งเป็นผลจากการสอบของนักเรียนมาประกอบการพิจารณาตัดสิน ว่านักเรียนที่สอบไปแล้วนั้นมีความรอบรู้หรือไม่ โดยการวัดผลดังกล่าวจะครอบคลุมจุดประสงค์ การเรียนการสอน (Hambleton, Swaminathan, Algens and Coulson 1978: 1) และ ในช่วงปี ค.ศ.1960 ถึง 1970 การเรียนการสอนเน้นรูปแบบที่เป็นรายบุคคลมากขึ้น ดังนั้น แบบสอบที่จะนำมาเป็นเครื่องมือสำหรับวัดผลก็ต้องมีลักษณะของการวัดด้วย นั่นคือ จะต้องมิเกณฑ์ ที่จะตัดสินว่าผู้สอบรอบรู้หรือไม่

กระบวนการทดสอบอิงเกณฑ์จะแปลความหมายคะแนนโดยนำผลการสอบไปเทียบกับ มาตรฐานที่แท้จริง (Absolute Standard) ซึ่งเป็นเกณฑ์ภายนอกที่กำหนดเอาไว้อย่างรอบคอบ

โดยจะไม่เปรียบเทียบผลการสอบดังกล่าวกับผู้สอบคนอื่น ๆ ในกลุ่ม ดังนั้นผลงานของนักเรียนจะอยู่ในระดับมาตรฐานหรือไม่ก็ต้องพิจารณาหรือเปรียบเทียบกับเกณฑ์หรือมาตรฐานที่แท้จริงเท่านั้น และการรายงานผลการสอบจะรายงานในรูปของการอ้างอิงไปยังมาตรฐานที่กำหนดเอาไว้ล่วงหน้า ซึ่งอาจเสนอในพจน์ของจำนวนเปอร์เซ็นต์การตอบถูกของแต่ละบุคคล ซึ่งจะเป็นดัชนีบ่งชี้ว่าจำนวนเปอร์เซ็นต์นั้นบรรลุมาตรฐานการปฏิบัติที่กำหนดเอาไว้หรือยัง นอกจากนั้นคะแนนอาจจะอยู่ในรูปของจำนวนเวลาที่ทำงานนั้นเสร็จสมบูรณ์ (นายเชิด ภิญโญอนันตพงษ์ 2527: 14)

การวิเคราะห์ข้อกระตงแบบอิงเกณฑ์

ข้อกระตงที่ได้มาจากลักษณะเฉพาะของมวลความรู้ (Domain specification) จำเป็นที่จะต้องตรวจสอบความบกพร่องเชิงโครงสร้างและพิจารณาว่าข้อกระตงเหล่านั้นมีคุณสมบัติตรงกับจุดมุ่งหมายของการสร้างหรือไม่ เมื่อสร้างข้อกระตงจากจุดประสงค์โดยใช้กฎการสร้างข้อกระตงแบบเก่าแล้วนำมารวมเป็นแบบสอบชุดหนึ่ง ปริเขตของข้อกระตงแบบนี้จะเรียกว่า ปริเขตข้อกระตงไม่จำกัด (infinite item domain) ตัวอย่างเช่น ข้อกระตง 5 ข้อ ซึ่งสร้างมาจากจุดประสงค์ข้อหนึ่งที้อาจจะเขียนข้อกระตงได้ถึง 50 หรือ 500 ข้อ ตรงกันข้ามกับ ปริเขตข้อกระตงที่จำกัด (finite item domain) ข้อกระตงทั้งหมดที่สร้างขึ้นมาได้โดยใช้วิธีการเขียนข้อกระตงแบบ item transformation, item forms และ Algorithms และนำข้อกระตงมารวมกันเป็นแบบสอบโดยการลุ่มแบบลุ่ม หรือการลุ่มแบบแบ่งชั้น (random or stratified random sampling plan) วิธีการแบบนี้จะทำให้มั่นใจได้ว่า ข้อกระตงที่ได้มาเป็นตัวแทนของปริเขตข้อกระตง และจากเหตุนี้สามารถแปลความหมายคะแนนของแต่ละบุคคลในปริเขตที่กำหนดได้

การวิเคราะห์ข้อกระตงแต่ละข้อนั้นเพื่อประเมินว่า

1. วัดได้ตามลำดับการสอนและจุดประสงค์เชิงพฤติกรรมหรือไม่
2. มีความแตกต่างระหว่างผู้รอบรู้และไม่รอบรู้ ในจุดประสงค์ที่สอนหรือไม่

ข้อกระตงที่ได้มานั้นต้องปราศจากข้อผิดพลาดเชิงโครงสร้าง ซึ่งอาจจะแนะนำคำตอบให้นักเรียนหรืออาจจะทำให้นักเรียนสับสนก็ได้ ในการวิเคราะห์ข้อกระตงเพื่อที่จะให้ได้ความตรงนั้นเกี่ยวข้องกับ การวิเคราะห์ 2 ประการคือ

1. การสร้างระดับความเท่าเทียมกันของข้อกระทงกับจุดประสงค์ หรือระดับของ ความสัมพันธ์ระหว่างข้อกระทงแต่ละข้อและจุดประสงค์ซึ่งเป็นจุดมุ่งหมายที่จะวัด
2. เกี่ยวข้องกับกระบวนการทางสถิติที่แสดงถึงระดับของความสัมพันธ์ระหว่าง จุดมุ่งหมายและการตอบของนักเรียนแต่ละข้อ

ตามข้อเท็จจริงแล้วการวิเคราะห์ข้อกระทงอาจจะไม่จำเป็น เมื่อตัวอย่างข้อกระทงถูก ลุ่มมาอย่างจริง ๆ จากปริเขตที่จำกัด (finite domain) หรือคลังของข้อกระทง (pool of item) ซึ่งการวิเคราะห์ข้อกระทงนั้นถูกโต้แย้งว่า ในการทดสอบอิงปริเขต (Domain-referenced testing) การใช้ค่าสถิติในการคัดเลือกและพิจารณาข้อกระทงให้ดีขึ้นจะไป ทำลายการนิยามลักษณะของแบบสอบเชิงทฤษฎี ด้วยเหตุนี้จะมีจุดอ่อนในการแปลความหมาย ค่ะแนในปริเขตนั้น

ดังนั้นการวิเคราะห์ข้อกระทงควรจะนำมาใช้ในแบบสอบอิงเกณฑ์ ซึ่งข้อกระทงสร้างมา จากรายการของจุดประสงค์ที่ไม่เข้มงวดกับการเขียนจุดประสงค์มากนักและการลุ่มตัวอย่างข้อกระทง จากปริเขตที่จำกัด (finite domain) ไม่สามารถทำได้ นอกจากนี้การสร้างปริเขตข้อกระทง โดยทั่วไปสำหรับจุดมุ่งหมายในการลุ่มข้อกระทงนั้นมีความยุ่งยากเสียเวลา ถ้าใช้คอมพิวเตอร์ ช่วยในการสร้างจะทำได้แต่แพงเกินไป ในทางปฏิบัติแนวโน้มการเขียนข้อกระทงจากจุดประสงค์ สามารถนำไปปฏิบัติได้ง่ายกว่า ดังนั้นการวิเคราะห์ข้อกระทงในแบบสอบอิงเกณฑ์จึงเป็นสิ่งจำเป็น ซึ่งแนวทางในการวิเคราะห์จะได้กล่าวในรายละเอียดต่อไป

กระบวนการในการวิเคราะห์ข้อกระทงแบบอิงเกณฑ์จะพิจารณา 4 ลักษณะด้วยกันคือ

1. ความสอดคล้องของข้อกระทงกับจุดประสงค์ (item-objective congruence)
2. ค่าสถิติของข้อกระทง (item-statistics)
3. การคัดเลือกข้อกระทง (item-selection)
4. การปรับปรุงข้อกระทง (item-revision)

1. ความสอดคล้องของข้อกระทงกับจุดประสงค์ (item-objective congruence)

ลักษณะของข้อกระทงที่สำคัญที่สุดคือ ความสอดคล้องของข้อกระทงกับจุดประสงค์ ซึ่งเกี่ยวข้องกับขอบเขตของข้อกระทงวัดได้ตามจุดประสงค์ที่มุ่งหมายให้ข้อกระทงวัด ความสอดคล้องกันนี้ใช้กระบวนการพิจารณาตัดสิน โดยปกติข้อกระทงสร้างมาจากความล้มพันธ์ตามรายการของจุดประสงค์การสอน จุดประสงค์เชิงพฤติกรรมและตารางลักษณะเฉพาะของมวลความรู้ (table of Domain specification) มีเกณฑ์อยู่ 3 ประการ ที่จะพิจารณาความสอดคล้องกันหรือการจับคู่กันระหว่างข้อกระทงและจุดประสงค์ คือ พฤติกรรม (Behavior) เนื้อหา (Content) และการแบ่งแยกตามลำดับชั้น (hierarchical classification)

ในการสร้างข้อกระทง พฤติกรรม (กริยา) และเนื้อหาที่รวมกันเป็นจุดประสงค์ ควรจะสะท้อนถึงรูปแบบและเนื้อหาของข้อกระทง ความสอดคล้องในเกณฑ์ 2 ข้อแรก คือ พฤติกรรม และเนื้อหา สร้างขึ้นมาเพื่อให้เป็นแบบอย่างในระหว่างขั้นตอนการสร้างข้อกระทง อย่างไรก็ตาม ลักษณะของความสอดคล้องกัน (Congruence) ยังคงต้องนำมาพิจารณาก่อนการบริหารการสอบ บางครั้งกริยาในจุดประสงค์เชิงพฤติกรรม เสนอแนะคำถามที่เป็นอ้อมนัย เช่นคำว่า อธิบาย ซึ่งไม่สามารถจะจับคู่กับแบบสอบเลือกตอบได้ ดังนั้นจำเป็นต้องเลือกอย่างใดอย่างหนึ่งว่า จุดประสงค์ ถูกต้อง หรือควรเขียนข้อกระทงใหม่

การสร้างความสอดคล้องกันควรจะนำเกณฑ์ข้อที่ 3 มาพิจารณาคด้วย ความไม่สอดคล้องกันจะเกิดขึ้นในการพิจารณาข้อกระทงและจุดประสงค์ที่แบ่งแยกตามพุทธิปริเขต (Cognitive domain) จิตปริเขต (Affective domain) และทักษะปริเขต (Psychomotor domain) มากกว่าเกณฑ์ที่กล่าวมา เมื่อตรวจสอบตามรายการลำดับชั้นปกติ จะพบความคลาดเคลื่อน โดยเฉพาะอย่างยิ่งในระดับกลางและสูงกว่าระดับกลางของพุทธิปริเขต การวิเคราะห์ข้อกระทง จะทำให้ทราบว่าข้อกระทงที่สร้างนั้นวัดจุดประสงค์ได้จริงหรือไม่ ครุมีแนวโน้มที่จะสร้างข้อกระทงในระดับต่ำของพุทธิปริเขต เช่น ในระดับความรู้-ความจำ (Knowledge) เท่านั้นซึ่งจะทำให้เกิดความล้าสนเมื่อนำข้อกระทงไปวัดระดับที่แตกต่างกันของลำดับชั้นการเรียนรู้ หรือในการสร้างจุดประสงค์

เมื่อเขียนข้อกระทงจากจุดประสงค์ชนิดของพฤติกรรมที่ได้จะจำกัดอยู่กับความรู้ในการเขียนข้อกระทงของครู ขอบเขตของลักษณะเฉพาะของมวลความรู้ ได้จากการนิยามในบริบทของสิ่งที่มีมาก่อนว่าข้อกระทงจะเป็นอย่างไร และการแบ่งแยกตาม พหุธิปริเขต จิตปริเขต และทักษะปริเขต สร้างขึ้นเพื่อให้เหมาะกับข้อกระทงมากกว่าจะใช้ปรับปรุงข้อกระทง ซึ่งจะสร้างปัญหาในการรักษาความลุ่มนัยกับปริเขตเนื้อหา การประเมิน ความสอดคล้องจะไม่คงที่ระหว่างข้อกระทงกับรายการที่กำหนดไว้ ถ้าเป็นแบบนี้ผู้สร้างแบบสอบมีทางเลือก 2 ทางคือ ขยายจุดประสงค์และการแบ่งแยกตามพหุธิปริเขต จิตปริเขต และทักษะปริเขต เพื่อให้สอดคล้องกับข้อกระทงหรือเขียนข้อกระทงใหม่ให้สอดคล้องกับการแบ่งแยกที่มีอยู่

กระบวนการที่กล่าวมาสามารถนำไปใช้ในการสร้างแบบสอบระดับจังหวัดและระดับประเทศได้ แต่ความแตกต่างขึ้นอยู่กับจำนวนคนในการพิจารณาตัดสิน แทนที่ครูเพียงคนเดียวจะตัดสินในระดับชั้นเรียน การรวมการตัดสินของคณะบุคคลที่ชำนาญในด้านเนื้อหา จะทำให้ได้แบบสอบที่มีคุณภาพมากยิ่งขึ้น โรวินเนลลีและแฮมเบิลตัน (Rovinelli and Hambleton 1977) ได้เสนอวิธีการประเมินความสอดคล้องกันโดยใช้ผู้เชี่ยวชาญด้านเนื้อหาวิชา 3 วิธี คือ คำนวณที่ได้มาจากความเป็นเอกพันธ์ของการแทนที่ของ Hamphill and Westie's (1950) การประเมินในรูปแบบของสเกลซีแมนติคดิฟเฟอเรนเชียล (Semantic-Differential Scale) และการจับคู่โดยใช้การทดสอบความเป็นอิสระด้วยไคสแควร์ (Chi-square) การเปรียบเทียบเชิงประจักษ์ตามวิธีต่าง ๆ แสดงให้เห็นว่า คำนวณความสอดคล้องที่มีความหมายมากที่สุด คือ ความสอดคล้องตามการตัดสินของแต่ละบุคคล อย่างไรก็ตาม การไม่พิจารณาความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์จะทำให้เกิดความคลาดเคลื่อนในด้านความตรงของแบบสอบ ความถูกต้องในการสร้างแบบสอบเพียงอย่างเดียวไม่เพียงพอ ถ้าไม่สามารถกำหนดได้อย่างตรงไปตรงมาได้ว่า ข้อกระทงนั้นวัดได้ตามจุดประสงค์หรือไม่แล้ว ลักษณะอื่น ๆ เช่น ความเที่ยงก็ไม่มี ความหมาย ความสามารถในการแปลความหมายของแบบสอบ จะขึ้นอยู่กับความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ หรือความตรงเชิงเนื้อหานั้นเอง

การวิเคราะห์ความตรงเชิงเนื้อหาของข้อกระทงแบบอิงเกณฑ์จำแนกได้เป็น 2 วิธีดังนี้ (บุญเชิด วิทยุอนันตพงษ์ 2527: 68 อ้างจาก Hambleton, et al. 1978: 34-37)

1.1 อาศัยคุณนิมิตของผู้เชี่ยวชาญเนื้อหาวิชา

1.2 อาศัยเทคนิคการตรวจสอบจากการทดลองหรือเทคนิคเชิงประจักษ์ (Empirical Techniques)

1.1 การวิเคราะห์ความตรงเชิงเนื้อหาโดยอาศัยคุณนิมิตของผู้เชี่ยวชาญเนื้อหาวิชาตัดสิน เป็นการพิจารณาว่าข้อกระทงกับจุดประสงค์เชิงพฤติกรรม หรือลักษณะเฉพาะของมวลความรู้ที่ต้องการวัด มีความสอดคล้องกันหรือไม่ การพิจารณาเช่นนี้ ต้องมั่นใจเสียก่อนว่าจุดประสงค์หรือลักษณะเฉพาะของมวลความรู้เขียนไว้ชัดเจนแล้ว โดยใช้การตรวจสอบกับตารางเฉพาะที่ได้จากการวิเคราะห์เนื้อหาและจุดประสงค์ การวิเคราะห์ความตรงเชิงเนื้อหาโดยอาศัยคุณนิมิตของผู้เชี่ยวชาญเนื้อหาวิชาตัดสินนี้ โรวีเนลลีและแอมเบลตัน ได้เสนอไว้ 3 วิธีดังนี้

1.1.1 ใช้ดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์

1.1.2 ใช้ดัชนีความเหมาะสมระหว่างข้อกระทงกับจุดประสงค์

1.1.3 ใช้ดัชนีการจับคู่ระหว่างข้อกระทงกับจุดประสงค์

1.1.1 การวิเคราะห์ความตรงเชิงเนื้อหาโดยใช้ดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ การหาค่าดัชนีความตรงเชิงเนื้อหาหรือดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์เชิงพฤติกรรม โดยการตัดสินของผู้เชี่ยวชาญที่เสนอโดยแอมฟีลล์และเวสตี (เฮาเวตี วิบูลย์ศรี 2528: 40) คำนวณได้จากสูตรดังนี้

$$IOC = \frac{\sum_{i=1}^N R_i}{N}$$

IOC คือดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์

R_i คือคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชาคนที่ i

N คือจำนวนผู้เชี่ยวชาญเนื้อหาวิชา

โรวีเนลลีและแอมเบลตันได้พบว่า Hamphill-Westie Index จากวิธีการดังกล่าวมีข้อจำกัด 2 ประการคือ อิทธิพลเนื่องมาจากจำนวนจุดประสงค์ และจำนวนผู้เชี่ยวชาญในการตัดสินความตรง จึงได้ปรับปรุง Hamphill-Westie index ให้เป็น Index of Item-Objective Congruence คำนวณได้ดังนี้

$$I_{10} = \frac{(M-1) S_0 - S'_0}{2N(M-1)}$$

เมื่อ

- I_{10} คือ ดัชนีความสอดคล้องของข้อกระทง i และจุดประสงค์ o
 M คือ จำนวนจุดประสงค์ที่วัด
 N คือ จำนวนผู้เชี่ยวชาญที่เป็นผู้ตัดสิน
 S_0 คือ คะแนนรวมจากผู้ตัดสินทั้งหมดที่ให้กับจุดประสงค์ o
 S'_0 คือ คะแนนรวมจากผู้ตัดสินทั้งหมดที่ให้กับจุดประสงค์ทั้งหมด ยกเว้นจุดประสงค์ o

การแปลความหมายดัชนีความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ ถ้าดัชนีที่คำนวณได้มีมากกว่าหรือเท่ากับ .5 แสดงว่าข้อกระทงวัดหรือเป็นตัวแทนจุดประสงค์เชิงพฤติกรรมข้อนั้น ถ้าค่าดัชนีน้อยกว่า .5 แสดงว่าข้อกระทงไม่วัดหรือไม่เป็นตัวแทนจุดประสงค์เชิงพฤติกรรมข้อนั้น คัดเลือกข้อที่มีความตรงเชิงเนื้อหาไว้ ข้อกระทงที่ขาดความตรงเชิงเนื้อหาจะถูกกำจัดออกไปหรือไม่ก็ปรับแต่งข้อกระทงใหม่

1.1.2 การวิเคราะห์ความตรงเชิงเนื้อหาโดยใช้ดัชนีความเหมาะสมระหว่างข้อกระทงกับจุดประสงค์ การหาดัชนีความตรงเชิงเนื้อหาของข้อกระทงโดยใช้ดัชนีความเหมาะสมระหว่างข้อกระทงกับจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะของมวลความรู้ คำนวณได้จากสูตรค่าเฉลี่ยของคะแนนและค่าความแปรปรวนของคะแนนดังนี้ (บุญเชิด วิทยุอินันตพงษ์ 2527: 71)

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$$

$$S = \sqrt{\left[N \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2 \right] / N(N-1)}$$

เมื่อ

- \bar{X} คือค่าเฉลี่ยคะแนนความคิดเห็น
 S คือค่าความเบี่ยงเบนมาตรฐานของคะแนนความคิดเห็น
 X_i คือคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชาคนที่ i
 N คือจำนวนผู้เชี่ยวชาญเนื้อหาวิชา

การหาค่านี้ดังกล่าวอาศัยมาตราส่วนประมาณค่าดังนี้

- 4 เหมาะสมมากที่สุด
 3 เหมาะสมมาก
 2 เหมาะสมปานกลาง
 1 เหมาะสมน้อย
 0 เหมาะสมน้อยที่สุดหรือไม่เหมาะสมเลย

การแปลความหมายค่าเฉลี่ยของคะแนนความคิดเห็น แปลความหมายได้ดังนี้

ค่าเฉลี่ย	ข้อกระทงมีความเหมาะสมกับจุดประสงค์
3.50-4.00	มากที่สุด
2.50-3.49	มาก
1.50-2.49	ปานกลาง
0.50-1.49	น้อย
0.00-0.49	น้อยที่สุด

คัดเลือกข้อกระทงที่มีค่าเฉลี่ยตั้งแต่ 2.5-4.0 และต้องมีค่าส่วนเบี่ยงเบนมาตรฐานไม่เกิน 1.0 ไว้ ข้อกระทงที่ไม่เป็นไปตามเกณฑ์จะถูกกำจัดออกไป หรือนำไปแก้ไขปรับแต่งข้อกระทงใหม่

1.1.3 การวิเคราะห์ความตรงเชิงเนื้อหาโดยใช้ดัชนีการจับคู่ระหว่างข้อกระทงกับจุดประสงค์ การหาค่าดัชนีความตรงเชิงเนื้อหา โดยใช้ดัชนีการจับคู่ระหว่างข้อกระทงกับจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะของมวลความรู้ คำนวณได้จากสูตรดังนี้

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

เมื่อ

M คือเปอร์เซ็นต์ความถี่ของการจับคู่แทนดัชนีการจับคู่

f คือจำนวนความถี่ของผู้เชี่ยวชาญเนื้อหาวิชาที่ลงความคิดเห็นว่าข้อกระทงจับคู่กับจุดประสงค์

คัดเลือกข้อกระทงที่มีเปอร์เซ็นต์ความถี่ตั้งแต่ 80% ขึ้นไปไว้ ซึ่งแสดงว่าข้อกระทงวัดพฤติกรรมตรงตามที่ระบุไว้ในจุดประสงค์ที่จับคู่จริง

1.2 การวิเคราะห์ความตรงเชิงเนื้อหาโดยอาศัยเทคนิคการตรวจสอบจากการทดลองหรือเทคนิคเชิงประจักษ์ เป็นการพิจารณาว่าข้อกระทงเป็นตัวแทนของจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะของมวลความรู้หรือไม่ โดยนำข้อกระทงไปสอบกับนักเรียนแล้วนำผลการสอบที่ได้มาพิจารณาตัดสิน

1.2.1 การวิเคราะห์ความตรงเชิงเนื้อหาโดยอาศัยเทคนิคการทดลองตามวิธีของครอนบาค การวิเคราะห์ความตรงเชิงเนื้อหาที่ได้กล่าวมาแล้วจะถือว่าจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะของมวลความรู้ มีความชัดเจน สามารถนิยามประชากรหรือมวลของข้อกระทงได้ แต่โดยความเป็นจริงแล้วมีบางวิชาเท่านั้นที่สามารถนิยามประชากรหรือมวลข้อกระทงได้ชัดเจน เช่น วิชาคณิตศาสตร์ และการสะกดคำ ส่วนวิชาอื่น ๆ จะไม่สามารถกำหนดลักษณะเฉพาะของมวลความรู้ให้สมบูรณ์ชัดเจนได้ ครอนบาค (Cronbach 1971 อ้างถึงใน บุญเชิด วิทยุไถอนันตพงษ์ 2527: 75) ได้เสนอวิธีการวิเคราะห์ความตรงเชิงเนื้อหาโดยอาศัยเทคนิคการทดลอง ซึ่งถือว่าประชากรข้อกระทงได้นิยามไว้ชัดเจนแล้วข้อกระทงที่ลุ่มมาจากประชากรข้อกระทงเดียวกันต้องให้ผลตรงกัน การวิเคราะห์ตามวิธีของครอนบาคมีดังนี้

- 1.2.1.1 เลือกผู้เขียนข้อกระทงที่มีความสามารถและ
ประสบการณ์ที่ทัดเทียมกันสองกลุ่ม
- 1.2.1.2 กำหนดจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะ
ของมวลความรู้แล้วให้ผู้เขียนข้อกระทงทั้งสองกลุ่มเขียนข้อกระทงจำนวนหนึ่งแล้วตรวจปรับแต่งจน
ดีที่สุด
- 1.2.1.3 นำข้อกระทงทั้งสองชุดไปสอบกลุ่มนักเรียนที่มี
ความรู้ตามจุดประสงค์นั้น
- 1.2.1.4 นำคะแนนสอบของผู้สอบแต่ละชุดมาพิจารณา
เปรียบเทียบ ถ้าค่าเฉลี่ยคะแนนของข้อกระทงสองชุดใกล้เคียงกันแสดงว่า ข้อกระทงสองชุดนั้น
เป็นตัวแทนของจุดประสงค์เชิงพฤติกรรม หรือลักษณะเฉพาะของมวลความรู้ที่กำหนดให้ และแสดง
ว่าจุดประสงค์เชิงพฤติกรรมหรือลักษณะเฉพาะของมวลความรู้เขียนไว้ชัดเจน การเปรียบเทียบ
ใช้สูตร t-test ทำการทดสอบดังนี้

$$t = \frac{\sum D}{\sqrt{\frac{N \sum D^2 - (\sum D)^2}{N - 1}}}$$

เมื่อ

D คือผลต่างระหว่างคะแนนข้อกระทงชุดที่ 1 และข้อกระทงชุดที่ 2
N คือจำนวนนักเรียนที่สอบทั้งหมด

การแปลความหมายเมื่อค่า t-test ที่คำนวณได้น้อยกว่าค่าวิกฤติ แสดงว่า
ข้อกระทงทั้งสองชุดให้ผลไม่แตกต่างกัน สรุปได้ว่า จุดประสงค์เชิงพฤติกรรมที่กำหนดให้ นั้นชัดเจน
จึงทำให้ได้ข้อกระทงสองชุดที่ลุ่มมามีความสอดคล้องกัน

2. ค่าสถิติของข้อกระทง (Item Statistic)

เมื่อพิจารณาความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์แล้ว ผู้สร้างแบบสอบควร
จะวางแผนนำข้อกระทงไปสอบกับกลุ่มของนักเรียน ข้อมูลที่ได้จากการตอบของนักเรียนสามารถ

จะนำมาประเมินประสิทธิภาพของข้อกระทงว่า ข้อกระทงเหล่านั้น มีลักษณะต่าง ๆ ที่เราคั้งใจไว้หรือไม่ ขึ้นตอนต่าง ๆ ประกอบไปด้วย การเลือกกลุ่มที่เป็นเกณฑ์ การรวบรวมผลย้อนกลับของนักเรียน การคำนวณค่าความยากและค่าอำนาจจำแนก และการคำนวณค่าความเป็นเอกพันธ์

2.1 การเลือกกลุ่มที่เป็นเกณฑ์ (Selection of Criterion Groups)

การพิจารณาครั้งแรกในการใช้ค่าสถิติของข้อกระทง คือ การเลือกกลุ่มนักเรียนที่จะทำแบบสอบ ส่วนสำคัญของค่าสถิติมีพื้นฐานมาจากการสอบก่อนสอน-สอบหลังสอน โดยใช้การสอบซ้ำจากนักเรียนเพียงกลุ่มเดียว และจากนักเรียนสองกลุ่มที่อิสระจากกัน

การเลือกกลุ่มขึ้นอยู่กับจุดมุ่งหมายของแบบสอบและประโยชน์ที่จะได้รับ เนื่องจากแบบสอบอิงเกณฑ์ ส่วนมากใช้ในกรณีที่จะพิสูจน์ว่า การรอบรู้และไม่รอบรู้ของนักเรียนในจุดประสงค์ข้อหนึ่ง คำถามก็คือครูและนักประเมิน ต้องการจะตอบคำถามว่า ใครควรจะสามารรถตอบข้อกระทงอย่างถูกต้อง และใครควรจะตอบข้อกระทงไม่ถูกต้อง กลุ่มของนักเรียนจะประกอบไปด้วยนักเรียนที่มีความรู้กับพวกที่ไม่มีความรู้ หลังจากที่ได้รับ การสอนแล้ว เหมือนกับว่าก่อนการสอนนักเรียนส่วนมากแต่ไม่ทั้งหมดจะทำข้อกระทงไม่ถูกต้อง ถ้าการกระทำแบบนี้ไม่เกิดขึ้น ข้อกระทงจะผิดคนลาดในหน้าที่ที่คาดหวังไว้ หรือจุดประสงค์ได้ถูกสอนมาก่อนแล้ว เมื่อจุดประสงค์ได้สอนไปแล้วก็พิจารณาการสอนในจุดประสงค์นั้น และเมื่อทดสอบหลังจากได้สอนเรียบร้อยแล้ว นักเรียนส่วนมากแต่ไม่ทั้งหมดควรจะตอบข้อกระทงอย่างถูกต้อง การนิยามกลุ่มที่เป็นเกณฑ์มี 3 วิธีคือ

2.1.1 วิธีสอบกับกลุ่มตัวอย่างกลุ่มเดียว โดยสอบก่อนสอนและสอบหลังสอน (preinstruction-post instruction)

2.1.2 วิธีสอบกับกลุ่มตัวอย่างสองกลุ่ม โดยสอบกับกลุ่มที่ไม่ได้รับการสอนและกลุ่มที่ได้รับการสอน (uninstruction-instruction groups)

2.1.3 วิธีสอบกับกลุ่มตัวอย่างกลุ่มเดียว โดยพิจารณาจากคะแนนจุดตัด

2.1.1 วิธีสอบกับกลุ่มตัวอย่างกลุ่มเดียว โดยสอบก่อนสอนและสอบหลังสอน (preinstruction-post instruction) เป็นการทดสอบนักเรียนกลุ่มเดียวกัน 2 ครั้ง ก่อนการสอน (pre-instruction) และหลังการสอน (post-instruction) ปกตินักเรียนจะถูกทดสอบด้วยข้อกระทงที่เหมือนกัน 2 ครั้ง หรือข้อกระทงที่คู่ขนานกัน ประโยชน์ของการสอบนักเรียนก่อนและหลังการสอนคือ จะทำให้ครูตรวจสอบนักเรียนแต่ละคนได้คั้ง ๆ กับตรวจสอบทั้ง

กลุ่ม โดยเฉพาะการทำแบบนี้มีประโยชน์ในการประเมินความก้าวหน้าจากหน่วยการเรียนหนึ่งไปยังหน่วยการเรียนถัดไป

ข้อเสียของวิธีนี้ คือ ไม่สามารถจะวิเคราะห์ข้อกระทำได้ถ้าจุดประสงค์ที่สอนนั้น นักเรียนได้เรียนมาแล้ว ข้อเสียประการที่สองเกี่ยวกับเวลาที่ใช้ในการสอบครั้งแรกและครั้งหลัง เมื่อการสอบใกล้กันมากเกินไปเป็นไปได้ว่า การทำข้อกระของนักเรียนก่อนการสอบมีอิทธิพลในการทำข้อกระหลังการสอบ การแก้ปัญหานี้ทำได้โดยใช้แบบสอบคู่ขนาน แต่ถ้าการสร้างแบบสอบคู่ขนานไม่สะดวก ควรจะขยายเวลาในการสอบสองครั้งให้มากขึ้น แต่เวลาที่ห่างกันไม่ควรมากกว่า 4 เดือน เพื่อจะหลีกเลี่ยงผลของวุฒิภาวะ ซึ่งรู้จักเป็นอย่างดีว่าเด็กที่มีอายุมากขึ้นสามารถจะทำข้อกระได้ดีขึ้น แต่ครูผู้สอนปรารถนาแต่เพียงผลของการสอนและการวิเคราะห์ข้อกระที่ต้องการเนืองการเปลี่ยนแปลงการตอบข้อกระที่ขึ้นอยู่กับการสอนเท่านั้น

ในการให้คะแนนของแบบสอบอิงเกณฑ์ข้อถูกจะได้ 1 ส่วนข้อผิดจะได้ 0 เหตุผลซึ่งข้อกระไม่ตอบให้ศูนย์เนื่องมาจาก การทดสอบอิงเกณฑ์จะประเมินนักเรียนว่ามีความรู้มากน้อยเพียงใด โดยไม่จำกัดเวลาในการตอบ เมื่อนักเรียนมีเวลาในการตอบอย่างมากพอที่จะทำแบบสอบได้เสร็จสมบูรณ์ ดังนั้นจะสรุปได้ว่า ถ้านักเรียนไม่ตอบก็เป็นเพราะว่า เขาไม่มีความรู้ในคำตอบที่ถูกต้อง

2.1.2 วิธีสอบกับกลุ่มตัวอย่างสองกลุ่ม โดยสอบกับกลุ่มที่ไม่ได้รับการสอนและกลุ่มที่ได้รับการสอน (uninstruction - instruction groups approach) เป็นรูปแบบของ know-groups technique ที่ใช้หาความจริงของแบบสอบขลุกลินแล้วนำมาประยุกต์ใช้ในการวัดแบบอิงเกณฑ์ เกี่ยวกับการทดสอบนักเรียนสองกลุ่มที่แตกต่างกันในเวลาเดียวกัน กลุ่มแรกไม่ได้รับการสอน (The uninstructed group) และกลุ่มที่สองได้รับการสอน (The instructed group)

นักเรียนที่ถูกเลือกเป็นกลุ่มที่ได้รับการสอนควรมีระดับความสามารถและลักษณะอื่น ๆ ที่เหมือน ๆ กันกับกลุ่มที่ไม่ได้รับการสอน ถึงแม้ว่าขนาดจะไม่เท่ากัน การกระจายตามสัดส่วนของลักษณะต่าง ๆ ควรจะเท่าเทียมกัน ตัวอย่างเช่น ถ้ากลุ่มแรกมี

2.3 ค่าความยากของข้อกระถง (Item difficulty) ค่าความยากของข้อกระถง คือ เปอร์เซนต์ของจำนวนนักเรียนซึ่งตอบข้อกระถงอย่างถูกต้อง คำนวณโดยสูตร

$$\text{ค่าความยาก} = \frac{C}{N} \times 100$$

เมื่อ

C คือ จำนวนนักเรียนซึ่งตอบข้อกระถงอย่างถูกต้อง

N คือ จำนวนนักเรียนทั้งหมด

ค่าที่ได้ก็คือ ค่าเฉลี่ยของข้อกระถงคูณด้วย 100 นั่นเอง ค่าความยากนี้จะมีนิลยตั้งแต่ 0-100 ค่าความยากสูงข้อกระถงจะง่าย เมื่อนักเรียนทุกคนตอบข้อกระถงถูก ค่าความยากจะเป็น 100 ถ้าทุกคนตอบผิด ค่าที่ได้จะเป็น 0 ในการพิจารณาค่าความยากนี้ได้มาจากการนิยามกลุ่มที่เป็นเกณฑ์ ถ้าระดับความยากพิจารณาเพียงกลุ่มนักเรียนที่ได้รับการสอน เราจะไม่สามารถทราบได้เลยว่า นักเรียนทำข้อสอบคล้ายกับกลุ่มที่ไม่ได้รับการสอนหรือไม่ นั่นก็คือข้อกระถงที่ง่ายหลังการสอน อาจจะง่ายก่อนสอนก็ได้

ในการแปลความหมายค่าความยาก ค่าความยากก่อนสอนควรอยู่ระหว่าง 0-50% และค่าความยากหลังสอนควรอยู่ระหว่าง 70-100% (Berk 1980: 66)

2.4 ค่าอำนาจจำแนกของข้อกระถง (Item discrimination) ค่าอำนาจจำแนกวัดการเปลี่ยนแปลงที่ได้จากการปฏิบัติจากการสอบก่อนสอนและสอบหลังสอน หรือความแตกต่างของกลุ่มที่ได้รับการสอนกับกลุ่มที่ไม่ได้รับการสอน ในการทดสอบอิงเกณฑ์ ค่าอำนาจจำแนกควรจะสูงสุทธระหว่างกลุ่มที่เป็นเกณฑ์ และค่าอำนาจจำแนกต่ำสุดภายในกลุ่มเดียวกัน ความแตกต่างที่เกิดขึ้นระหว่างกลุ่มที่เป็นเกณฑ์เนื่องมาจากผลของการสอน การแปลความหมายค่าอำนาจจำแนกที่ได้ในพจน์ของโปรแกรมการสอน จะเรียกว่า ความไวในการสอน (Instructional Sensitivity) วิธีวิเคราะห์ค่าอำนาจจำแนก เมื่อพิจารณาจากจำนวนกลุ่มตัวอย่างนักเรียนและจำนวนครั้งของการสอบแล้วอาจแบ่งได้ 4 วิธีดังนี้ (บุญเชิด วิทยุโณนันทพงษ์ 2527: 83)

- 2.4.1 วิธีสอบครั้งเดียวจากกลุ่มตัวอย่างกลุ่มเดียว ซึ่งจะหาค่าอำนาจ
จำแนกของเบรนนอน (Brennan Index)
- 2.4.2 วิธีสอบครั้งเดียวจากกลุ่มตัวอย่างสองกลุ่ม ใช้ค่าอำนาจจำแนก
ของเบรนนอนเช่นกัน
- 2.4.3 วิธีสอบซ้ำจากกลุ่มตัวอย่างกลุ่มเดียว หรือเรียกว่า วิธีสอบก่อน
และหลังสอน ซึ่งจะหาค่าอำนาจจำแนกแบบต่าง ๆ คือ
- 2.4.3.1 วิธีของคอกซ์และวาร์กาส (Cox and Vargas)
- 2.4.3.2 วิธีของรูดาบช (Roudabush)
- 2.4.3.3 วิธีของโคซีคอฟฟ์และไคลน์ (Kosecoff and Klein)
- 2.4.3.4 วิธีของเบย์ (Bayes)
- 2.4.4 วิธีสอบครั้งเดียวหรือสอบซ้ำจากกลุ่มตัวอย่างกลุ่มเดียว โดย
อาศัยแบบจำลองคุณลักษณะแฝง (Latent trait Model)

2.4.1 การหาค่าอำนาจจำแนกโดยวิธีสอบครั้งเดียวจากกลุ่มตัวอย่าง
กลุ่มเดียว เบรนนอน (Brennan) ได้อาศัยดัชนีอำนาจจำแนกแบบกลุ่มสูงและกลุ่มต่ำของจอห์นสัน
(Johnson 1951 อ้างถึงใน นุญเชิด วิทยุอนันตพงษ์ 2527: 83)

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

เมื่อ

- B คือ ดัชนีเบรนนอนหรือดัชนีอำนาจจำแนก
- N_1 คือ จำนวนนักเรียนที่สอบได้คะแนนสูงกว่าจุดตัด
- N_2 คือ จำนวนนักเรียนที่สอบได้คะแนนต่ำกว่าจุดตัด
- U คือ จำนวนนักเรียนกลุ่ม N_1 ตอบข้อกระทงถูก
- L คือ จำนวนนักเรียนกลุ่ม N_2 ตอบข้อกระทงถูก

2.4.2 การหาค่าอำนาจจำแนกโดยวิธีสอบครั้งเดียวจากกลุ่มตัวอย่าง
สองกลุ่ม เครเฮน (Crehen 1974: 257) ได้ดัดแปลงดัชนีบีของเบรนนอนมาใช้กับกลุ่มรอบรู้
และไม่รอบรู้ที่นิยามจากกลุ่มที่ได้รับการเรียนรู้ตามจุดประสงค์และกลุ่มที่ไม่ได้เรียนรู้ตามจุดประสงค์

เนื่องมาจากปัญหาของการกำหนดคะแนนจุดตัด ดังนั้นสูตรที่ใช้ในการคำนวณก็ยังคงใช้แบบเดิม แต่
นิยาม N_1 และ N_2 ใหม่ ดังนี้

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

เมื่อ

- B คือ ดัชนีเบรนนอนหรือดัชนีอำนาจจำแนก
 N_1 คือ จำนวนนักเรียนในกลุ่มที่ได้รับการเรียนรู้ตามจุดประสงค์
 N_2 คือ จำนวนนักเรียนในกลุ่มที่ไม่ได้รับการเรียนรู้ตามจุดประสงค์
U คือ จำนวนนักเรียนกลุ่ม N_1 ตอบข้อกระทงถูก
L คือ จำนวนนักเรียนกลุ่ม N_2 ตอบข้อกระทงถูก

นอกจากจะใช้ดัชนีอำนาจจำแนกของเบรนนอนเพื่อตรวจหาความตรงเชิงจำแนกแล้ว
อาจใช้สูตรอื่น ๆ อีก เช่น ใช้สัมประสิทธิ์สหสัมพันธ์แบบนิ (ϕ) ตรวจหาความสอดคล้องที่ว่า กลุ่ม
นักเรียนที่ได้รับการสอนควรตอบถูก และกลุ่มที่ไม่ได้รับการสอนควรตอบผิดหรืออาจคัดแปลง ดัชนี
ความตรงทั้งฉบับของคาร์เวอร์ (Carver 1970 อ้างถึงใน บุญเชิด วิทยุอินันตพงษ์ 2527 :
86) มาใช้ก็ได้ ซึ่งจะอาศัยการแจกแจงการตอบถูก-ผิดของทั้งสองกลุ่มดังนี้

		กลุ่มนักเรียน		
		ไม่ได้รับการสอน	ได้รับการสอน	
ตอบถูก		b	a	a+b
	ตอบผิด	c	d	c+d
		b+c	a+d	

แล้วคำนวณค่าอำนาจจำแนกจากสูตร

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

2.4.3 การหาค่าอำนาจจำแนกโดยวิธีสลับข้างจากกลุ่มตัวอย่าง กลุ่มเดียวหรือเรียกว่า วิธีสลับก่อนและสลับหลังสอน เบอร์ก (Berk 1980: 61) ได้กล่าวถึงวิธีหาค่าอำนาจจำแนก โดยพิจารณาจากการสลับก่อนและสลับหลังสอน ซึ่งเป็นวิธีที่คำนวณง่าย มีความหมายในการพัฒนาแบบสลับ 3 วิธีคือ วิธีของคอกซ์และวาร์กาส วิธีของรูคาบซุชและวิธีของโคชีคอฟฟ์และโคลน์ วิธีของคอกซ์และวาร์กาสนั้นคำนวณง่ายที่สุด วิธีของโคชีคอฟฟ์และโคลน์มีความไวมากที่สุด ค่าอำนาจจำแนกของโคชีคอฟฟ์และโคลน์ใช้ข้อสังเกตมาก นิสัยกว้างกว่าค่าที่เป็นไปได้ตามวิธีของรูคาบซุช ในทางสถิติแล้วค่าอำนาจจำแนกของโคชีคอฟฟ์และโคลน์และของรูคาบซุช มีความไวมากกว่า ค่าอำนาจจำแนกของคอกซ์และวาร์กาส เพราะค่าที่ได้นี้คำนึงถึงคะแนนของแต่ละบุคคลทั้งที่ทำถูกและทำผิด ซึ่งเป็นผลมาจากการสอนมากกว่าการพิจารณาเพียงรวม ๆ กล่าวโดยสรุปแล้วค่าอำนาจจำแนกตามวิธีของของโคชีคอฟฟ์และโคลน์จะคงที่มากกว่า ค่าอำนาจจำแนกของคอกซ์และวาร์กาสและรูคาบซุช

2.4.3.1 วิธีหาค่าอำนาจจำแนกของคอกซ์และวาร์กาส

(Cox and Vargas) พิจารณาจากตารางต่อไปนี้

สลับก่อน	สลับหลัง		สัดส่วนแยก
	ผิด=0	ถูก=1	
ถูก=1 ผิด=0	P_{10}	P_{11}	$P_{1.}$
สัดส่วนแยก	$P_{.0}$	$P_{.1}$	

คอกซ์และวาร์กาลได้เสนอวิธีการวิเคราะห์ข้อกระทบเพื่อ
 ตรวจสอบความตรงเชิงจำแนกโดยใช้ค่าอำนาจจำแนกที่คัดแปลงมาจากสถิติ D ของจอนห์สันคล้ายกับ
 ของเบรนนอน ซึ่งเป็นการหาผลต่างค่าความยากของการสอบจากกลุ่มตัวอย่างสองชุด โดยอาศัย
 คะแนนจากการสอบก่อนและสอบหลังสอน แล้วหาค่าความแตกต่างระหว่างสัดส่วนของผู้ตอบถูก
 ตอนสอบก่อนและสอบหลังสอน ดังนี้ (Van der Linden 1981: 381)

$$DIS_{FFD} = P.1 - P1.$$

- DIS_{FFD} คือค่าอำนาจจำแนกของคอกซ์และวาร์กาล
 $P.1$ คือสัดส่วนของจำนวนผู้ตอบถูกตอนสอบหลังสอน
 $P1.$ คือสัดส่วนของจำนวนผู้ตอบถูกตอนสอบก่อนสอน

พิสัยค่าอำนาจจำแนกของคอกซ์และวาร์กาลอยู่ระหว่าง
 -1.00 ถึง 1.00 ข้อจำกัดของวิธีนี้คือ ค่าที่ได้จะไม่ให้ความไวกับการแปรเปลี่ยนการสอบราย
 บุคคล แต่เป็นผลจากความก้าวหน้าของกลุ่มผู้สอบทั้งหมด

2.4.3.2 วิธีหาค่าอำนาจจำแนกของรูดาบุช (Roudabush) รูดาบุช
 ได้เสนอวิธีหาค่าอำนาจจำแนกโดยยึดหลักคะแนนความก้าวหน้ารายบุคคล ค่าอำนาจจำแนกของ
 รูดาบุชคำนวณจากสัดส่วนของผู้ตอบผิดก่อนสอนและตอบถูกหลังสอน ค่าอำนาจจำแนกตามวิธีของ
 รูดาบุชนี้เป็นผลจากการเรียนรู้ที่เพิ่มขึ้นโดยตรง

$$DIS_{RB} = P_{01}$$

- DIS_{RB} คือ ค่าอำนาจจำแนกของรูดาบุช
 P_{01} คือ สัดส่วนของการตอบผิดก่อนสอน และตอบถูกหลังสอน

ถ้ามีการแก้การเดาพิสัยของค่าอำนาจจำแนกจะอยู่ระหว่าง -1.00 ถึง
 1.00 แต่ถ้าไม่นำวิธีแก้การเดามาใช้จะมีพิสัยอยู่ระหว่าง 0 ถึง 1.00 ข้อจำกัดของการหาค่า
 อำนาจจำแนกตามวิธีของรูดาบุชคือ ถ้านำวิธีแก้การเดามาใช้ ข้อตกลงของการแก้การเดายัง
 เป็นสิ่งที่น่าสงสัยอยู่ (Berk 1980: 60)

2.4.3.3 วิธีหาค่าอำนาจจำแนกของโคซีคอฟฟ์และไคลน์

(Kosecoff and Klein) โคซีคอฟฟ์และไคลน์ ได้เสนอวิธีหาค่าอำนาจจำแนกรายข้อโดยเสนอสูตรที่เป็นผลมาจากความก้าวหน้ารายบุคคล คำนวณจาก สัดส่วนของการตอบผิดครั้งแรกและตอบผิดครั้งหลังลบด้วย สัดส่วนของการตอบผิดทั้งสองครั้ง สัดส่วนตัวแรกเป็นผลจากการเรียนรู้ที่เพิ่มขึ้น สัดส่วนหลังที่นำไปลบนั้นเนื่องจากความคลาดเคลื่อนที่เกิดจากการสร้างแบบสอบ และการเรียนรู้ (สมถวิล วิจิตรวรรณ 2524 อ้างจาก โคซีคอฟฟ์และไคลน์ 1974) ดังนั้น ค่าอำนาจจำแนกในสูตรนี้ จึงมีค่าต่ำกว่าสูตรอื่น ๆ สูตรการหาค่าอำนาจจำแนกตามวิธีของโคซีคอฟฟ์และไคลน์ มีดังนี้

$$DIS_{ng} = P_{01} - P_{00}$$

DIS_{ng} คือ ค่าอำนาจจำแนกตามวิธีของโคซีคอฟฟ์และไคลน์
 P_{01} คือ สัดส่วนของการตอบผิดก่อนสอนและตอบถูกหลังสอน
 P_{00} คือ สัดส่วนของการตอบผิดทั้งสองครั้ง

2.4.3.4 วิธีหาค่าอำนาจจำแนกของเบย์ส (Bayes) ทฤษฎีของเบย์ส

มีวิธีหาค่าอำนาจจำแนกหรือดัชนีความไวในการสอนอยู่ 3 วิธี ค่าอำนาจจำแนกทั้ง 3 วิธีนี้ต้องการข้อมูลสักอ้อยมาก วิธีที่เหมาะสมที่สุดพิจารณาจากค่าความยากจากนักเรียนที่ได้รับการสอนและไม่ได้เรียน ค่าอำนาจจำแนกของเบย์สทั้ง 3 วิธีคือ (Haladyna and Roid 1981: 42-43)

2.4.3.4.1 B1 ความน่าจะเป็นของนักเรียนที่มีความรู้

และตอบข้อกระทงถูก

2.4.3.4.2 B2 ความน่าจะเป็นของนักเรียนที่ไม่มีความรู้

และตอบข้อกระทงผิด

2.4.3.4.3 B3 ความน่าจะเป็นของการตัดสินใจอย่าง

ถูกต้อง โดยขึ้นอยู่กับกลุ่มผู้รอบรู้และไม่รอบรู้

ในการนำไปใช้ ค่าอำนาจจำแนก B3 จะทำให้ความน่าจะเป็นของการแบ่งแยกผู้รอบรู้และไม่รอบรู้ได้ถูกต้องเพิ่มขึ้น สูตรการคำนวณแต่ละวิธีเกี่ยวข้องกับค่าความยากในการสอบก่อนสอนและหลังสอน ดังนี้

$$B1 = \frac{(POSTDIFF) (COMDIFF)}{(POSTDIFF)(COMDIFF) + (PREDIFF)(1-COMDIFF)}$$

$$B2 = \frac{(1 - PREDIFF) (1 - COMDIFF)}{(1-PREDIFF) (1-COMDIFF) + (1-POSTDIFF)(COMDIFF)}$$

$$B3 = (POSTDIFF-COMDIFF)+(1-PREDIFF)+(COMDIFF-POSTDIFF)$$

เมื่อ

- B1, B2, B3 คือ ค่าอำนาจจำแนกของเบต้า
 PREDIFF คือ ค่าความยากจากกลุ่มตัวอย่างก่อนสอน
 POSTDIFF คือ ค่าความยากกลุ่มตัวอย่างหลังสอน
 COMDIFF คือ ค่าเฉลี่ยของ PREDIFF + POSTDIFF หรือค่าความยาก
 ของกลุ่มตัวอย่างรวมกัน (COMBINED-SAMPLES'
 DIFFICULTY)

2.4.4 การหาค่าอำนาจจำแนกโดยวิธีสอบครั้งเดียวหรือสอบซ้ำจากกลุ่มตัวอย่าง
 กลุ่มเดียวโดยอาศัยแบบจำลองคุณลักษณะแฝง (Latent trait Model) แบบจำลองคุณลักษณะ
 แฝงเป็นความสัมพันธ์ของผลการสอบกับความสามารถในรูปของฟังก์ชัน ซึ่งเกี่ยวกับพารามิเตอร์
 ของข้อระทงบางค่า แต่ค่าพารามิเตอร์เหล่านี้นิยามแตกต่างจากแบบจำลองคลาสสิก ดังนี้

ความยากของข้อระทงแทนด้วย b_i หมายถึง ระดับความสามารถของ
 ผู้สอบที่จะตอบข้อระทงข้อนั้นถูกต้อง ระดับความยากของข้อระทงมีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ แต่ทางปฏิบัติ
 จะมีค่าอยู่ในช่วง -2 ถึง $+2$ ค่า -2 แสดงว่าข้อระทงนั้นง่ายมาก ค่า $+2$ แสดงว่า
 ข้อระทงข้อนั้นยากมาก

ค่าอำนาจจำแนกของข้อระทงแทนด้วย s_i หมายถึงความลาดเอียง
 (Slope) ของโค้งลักษณะของข้อระทง ณ จุดที่ $\theta = b_i$ ระดับค่าอำนาจจำแนกของข้อระทง
 มีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ ในทางปฏิบัติ ค่าที่เป็นลบจะถูกกำจัดออกไปเพราะเป็นข้อระทงที่ไม่ดี

ค่าอำนาจจำแนกที่นำมาใช้จะมีค่าตั้งแต่ 0 ถึง +3 ค่า 0 แสดงว่าไม่มีอำนาจจำแนกเลย ค่า +3 แสดงว่ามีอำนาจจำแนกสูง

การแจกแจงแทนด้วย C_i ความน่าจะเป็นของผู้สอบที่มีความสามารถ
ต่ำสุดจะตอบข้อกระทงถูก

แบบจำลองคุณลักษณะแฝงสามารถแสดงความสัมพันธ์ได้ 2 แบบ คือ
แบบจำลอง Normal-Ogive มีฟังก์ชันทางคณิตศาสตร์ดังนี้

$$P_i(\theta) = \left(1 / (2\pi)^{1/2} \right) \int_{-\infty}^u e^{-u^2/2} du$$

เมื่อ $P_i(\theta)$ แทนความน่าจะเป็นของผู้สอบที่ตอบข้อกระทงข้อที่ i ถูกด้วย
ความสามารถ θ

และแบบจำลอง Logistic มีฟังก์ชันทางคณิตศาสตร์ดังนี้

$$P_i(\theta) = 1 / [1 + e^{-u}]^{-1}$$

$$U_i = a_i(\theta - b_i)$$

แบบจำลองนี้เป็นความสัมพันธ์ของความน่าจะเป็นในการตอบข้อกระทงถูกซึ่งแปร
ตามฟังก์ชันของความสามารถกับคุณสมบัตินี้ของข้อกระทง ได้แก่ ความยาก ค่าอำนาจจำแนก และ
การแจกแจง ดังนั้นแบบจำลองนี้จึงแบ่งได้เป็น 3 แบบดังนี้

แบบจำลอง Logistic ที่มี 3 พารามิเตอร์ คือ แบบจำลองของความน่าจะเป็น
ในการตอบข้อกระทงข้อที่ i ถูก ด้วยความสามารถ θ โดยข้อกระทงมีความยาก b_i มีอำนาจ
จำแนก a_i และมีการแจกแจง C_i มีความสัมพันธ์ที่ตัดแปลงแล้วดังนี้ (Van der Linden
1981: 390)

$$P_i(\theta) = C_i + (1 - C_i) [1 + \exp[-a_i(\theta - b_i)]]^{-1}$$

แบบจำลอง Logistic ที่มี 3 พารามิเตอร์ ถ้าไม่มีการเอาตอบถูกแล้ว $C_i = 0$ และจะกลายเป็นแบบจำลองที่มี 2 พารามิเตอร์ คือ ความยาก กับ ค่าอำนาจจำแนก และถ้ากำหนดว่า ข้อกระทงมีค่าอำนาจจำแนกเท่ากับ 1 ($a_i = 1$) แบบจำลองนี้จะเหลือเพียง 1 พารามิเตอร์ คือ ความยากของข้อกระทง โดยถือว่า ค่าความยากนี้เป็นค่าสถิติที่จะสามารถประมาณค่าพารามิเตอร์ได้ แบบจำลองนี้เรียกว่าแบบจำลองของแรลช์ (Rasch Model) มีความสัมพันธ์ดังนี้

$$P_i(\theta) = [1 + \exp[-(\theta - b_i)]]^{-1}$$

ในแต่ละค่าความสามารถ θ ของสมการที่กล่าวมา จะให้สัดส่วนของการตอบถูกที่คาดหวังได้ ดังนั้นค่า $P_i(\theta)$ จึงนำมาใช้คำนวณค่าความยากแบบเท่าที่นิยามจากสัดส่วนของการตอบถูก ซึ่งค่าความยากแบบเท่าของแต่ละประชากรที่ได้จากฟังก์ชันความน่าจะเป็น คำนวณได้ดังนี้

$$P_i(\theta) = \int_{-\infty}^{+\infty} P_i(\theta) g(\theta) d\theta$$

เมื่อ $P_i(\theta)$ แทนความยากของข้อกระทงที่นิยามจากสัดส่วนการตอบถูก

จากสมการ $P_i(\theta) = \int_{-\infty}^{+\infty} P_i(\theta) g(\theta) d\theta$ สามารถนำมาใช้หาค่าอำนาจจำแนกของข้อกระทงแบบอิงเกณฑ์โดยอาศัยการแจกแจงของคะแนนสอบก่อนสอบ $g_u(\theta)$ และการแจกแจงของคะแนนสอบหลังสอบ $g_l(\theta)$ เหมือนกับดัชนีของคอกซ์และวาร์กาส ดังนี้

$$D_{uu} = \int_{-\infty}^{+\infty} P_i(\theta) [g_l(\theta) - g_u(\theta)] d\theta$$

ถ้าการวิเคราะห์อาศัยผลการสอบจากกลุ่มตัวอย่างนักเรียนเพียงครั้งเดียวแล้ว กำหนดจุดตัด การพิจารณาคุณภาพข้อกระทงจะอาศัยฟังก์ชันข้อสนเทศของข้อกระทง (Item

Information Function) มาเขียนเป็นกราฟซึ่งเรียกว่า โค้งข้อสนเทศของข้อกระถง (Item Information Curve) หรือโค้งลักษณะของข้อกระถง

ถ้าข้อกระถงให้คะแนนถูก-ผิด เป็น 1-0 และให้ θ_c เป็นความสามารถในการตอบถูกหรือผ่านเกณฑ์แล้ว จะได้ฟังก์ชันข้อสนเทศของข้อกระถงที่ $\theta = \theta_c$ ซึ่งเป็นแบบจำลอง Logistic 3 พารามิเตอร์ ดังนี้

$$I_i(\theta_c) = [P'_i(\theta_c)]^2 / P_i(\theta_c) [1 - P_i(\theta_c)]$$

เมื่อ $I_i(\theta_c)$ แทนฟังก์ชันข้อสนเทศของข้อกระถงที่ $\theta = \theta_c$
 $P'_i(\theta_c)$ แทนค่าอนุพันธ์ที่หนึ่งของ $P_i(\theta_c)$ ที่ $\theta = \theta_c$

เมื่อทำให้อยู่ในรูปแบบจำลองแรลส์ จะได้ดังนี้

$$I_i(\theta_c) = P_i(\theta_c) [1 - P_i(\theta_c)]$$

ค่า $I_i(\theta_c)$ เป็นดัชนีบอกคุณภาพของข้อกระถง ถ้า I มีค่าสูงเพียงใดก็แสดงว่าข้อกระถงข้อนั้นมีคุณภาพในการบ่งชี้ผู้รอบรู้และไม่รอบรู้ได้ดีมากที่ $\theta = \theta_c$

นอกจากวิธีหาค่าอำนาจจำแนกที่กล่าวมาแล้ว เบอร์ก (Berk 1980: 62) ได้เสนอแนะการหาค่าอำนาจจำแนกซึ่งการคำนวณยุ่งยากซับซ้อน ดัชนีบางค่าต้องใช้โปรแกรมคอมพิวเตอร์ช่วยในการคำนวณมีรายละเอียดตามตารางที่ 1 ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 วิธีหาค่าอำนาจจำแนกของข้อกระทบแบบอิงเกณฑ์

สถิติ	แหล่งที่มา	ข้อเสนอแนะ
สหสัมพันธ์ระหว่าง คะแนนของข้อกระทบ และคะแนนรวม โดย การรวมกลุ่ม (Combined group item total r)	ฮาไลดาไคนาและ แอมสเต็ดเตอร์ (Haladyna and Helmstedter)	ดัชนีประยুক্তสหสัมพันธ์ระหว่างคะแนนของข้อกระทบ และคะแนนรวมโดยรวมกลุ่มที่เป็นเกณฑ์ มีความสัมพันธ์ กับดัชนีของคอกซ์และวาร์กัส (DIS _{pp}) อยู่ระหว่าง .64-.86 การคำนวณสามารถใช้โปรแกรมการวิเคราะห์ ข้อกระทบแบบอิงเกณฑ์ ข้อเสีย ดัชนีไม่ไวพอเมื่อนักเรียนได้คะแนนสูง และต่ำและดัชนีนี้อาจจะไม่วัดการจำแนกระหว่าง กลุ่มที่เป็นเกณฑ์
สหสัมพันธ์ระหว่าง คะแนนของข้อกระทบ กับกลุ่มที่เป็นเกณฑ์ (Item-criterion partial r)	คาร์ลิงตัน บิชอป และมิลล์แมน (Darlington Bishop and Millman)	เป็นการหาสหสัมพันธ์ระหว่างคะแนนของข้อกระทบกับ กลุ่มที่เป็นเกณฑ์ โดยกลุ่มที่เป็นเกณฑ์แบ่งเป็น (0,1) และให้คะแนนรวมเป็นค่าคงที่ สหสัมพันธ์ระหว่าง ข้อกระทบนำไปใช้ในการคัดเลือกข้อกระทบ ซึ่งวิธี การแบบนี้จะให้ค่าความตรงของแบบสอบสูงสุด ข้อเสีย การคำนวณยุ่งยากและซับซ้อน เนื้อหาและ ความสำคัญของจุดประสงค์ไม่นำมาพิจารณา
สหสัมพันธ์ระหว่าง คะแนนของข้อกระทบ กับคะแนนรวมที่ เปลี่ยนแปลงไป ในการสอบก่อนสอน และสอบหลังสอน (Change-item r)	ซัวปี (Saupe)	สหสัมพันธ์ของคะแนนข้อกระทบ (-1,0,+1) กับคะแนน รวมที่ได้จากการสอบก่อนสอน และการสอบหลังสอน การหาค่าอำนาจจำแนกแบบนี้เพื่อให้มีความเที่ยงหรือ ความตรงสูงสุดกับตัวแปรเกณฑ์ ข้อเสีย วิธีการคำนวณซับซ้อนมาก โดยการคำนวณ ความแปรปรวนร่วม 6 แบบ ที่แตกต่างกัน ข้อจำกัดของ ความแปรปรวนจากคะแนนสอบก่อนสอนและสอบหลังสอน

ตารางที่ 1 (ต่อ)

สถิติ	แหล่งที่มา	ข้อเสนอแนะ
มัลติเพิลรีเกรสชัน (Multiple regression)	มิลล์แมน (Millman)	จะทำให้ค่าอำนาจจำแนกมีขอบเขตจำกัด และไม่ได้ นำเนื้อหาและความสำคัญของ จุดประสงค์มานิเคราะห์ ในการคัดเลือกข้อกระตง การวิเคราะห์ข้อกระตง ต้องใช้โปรแกรมที่เรียกว่า CHANGE-SCORE โดยใช้การวิเคราะห์การถดถอยสหสัมพันธ์แบบมีขั้นคอง (Stepwise regression analysis) มีข้อกระตง เป็นตัวแปรทำนายและการแบ่งกลุ่มที่เป็นเกณฑ์เป็นตัวแปร เกณฑ์ วิเคราะห์รวมกันกับสหสัมพันธ์ระหว่างข้อกระตง ข้อเสีย ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ไม่มีดัชนี ค่าอำนาจจำแนกของข้อกระตง การคัดเลือกข้อกระตง ต้องนิเคราะห์สมการทำนายที่วิเคราะห์ได้ สัมประสิทธิ์ การถดถอยของข้อกระตง นำมาใช้เพื่อให้หน้าหนักคณน ของข้อกระตงและการวิเคราะห์แบบนี้ไม่ได้นิเคราะห์ ว่า ข้อกระตงวัตถุประสงค์เชิงพฤติกรรมหรือไม่

2.5 ความเป็นเอกพันธ์ของข้อกระตง (Item homogeneity) การวัดความเป็น
เอกพันธ์เพื่อจะนิสุจน์ว่าความสอดคล้องของข้อกระตงกับจุดประสงค์เหมือนกันหรือไม่โดยสอบเพียง
1 ครั้งหรือการสอบซ้ำ มีข้อตกลงอยู่ว่า ข้อกระตงควรจะมีค่าความยากเหมือนกัน ซึ่งข้อตกลงนี้ยัง
เป็นสิ่งที่น่าสงสัยอยู่ ดังนั้นจุดประสงค์เฉพาะหรือจุดประสงค์ทั่วไปอาจจะไม่จำเป็นต้องนิเคราะห์ค่า
สถิติความเป็นเอกพันธ์

เบอร์ก (Berk 1980: 65) ได้เสนอวิธีนิเคราะห์ความเป็นเอกพันธ์ของข้อกระตง
สรุปได้ตามตารางที่ 2 ดังนี้

ตารางที่ 2 วิธีพิจารณาความเป็นเอกพันธ์ของข้อกระทง

สถิติ	แหล่งที่มา	ข้อเสนอแนะ
ไคสแควร์	เพทท์ และออสเทอร์ฮอฟ (Pettie and Osterhof)	เป็นดัชนีที่ขยายมาจากวิธีไคสแควร์ของโพแฟมใช้ในกรณี ที่สอบเพียงครั้งเดียว ค่าความยาก ที่ได้จากการวัดข้อ กระทงแต่ละข้อนำไปเปรียบเทียบกับค่าความยากมัธยฐาน ของกลุ่มของข้อกระทง โดยใช้ไคสแควร์ ข้อเสีย เหมือนกับวิธีของโพแฟม
ไคสแควร์	โพแฟม (Popham)	ความถี่ของข้อกระทงจากการสอบก่อนสอนและหลังสอน เปรียบเทียบกับความถี่มัธยฐาน (ค่าคาดหวัง) ของ ข้อกระทงเดิม โดยการใช้ไคสแควร์ ค่ามัธยฐานได้ มาจากความถี่ของข้อกระทงในการวัดจุดประสงค์ เดียวกัน ข้อกระทงที่มีความถี่แตกต่างจากข้อกระทง เดิมต้องนำไปปรับปรุง ข้อเสีย ไม่มีค่าวิกฤตที่จะนิสฺจฺนข้อกระทงที่ไม่ดี
—	แอลเจด และลาดูคา (Engel and Laduca), โลวิงเจอร์ (Loevinger)	ขยายดัชนี ความเป็นเอกพันธ์ของข้อกระทงของลอง (Long) ดัชนีนี้อธิบายความสัมพันธ์ระหว่างคะแนน รายข้อและคะแนนทั้งหมดโดยทั่วไปมีมโนทัศน์คล้ายกับ ดัชนีค่าอำนาจจำแนกแบบเก่า ข้อเสีย การแปลความหมายดัชนีที่มีค่าเป็น 0 ไม่ชัดเจน ไม่ทราบการกระจายการสุ่มของค่าสถิติ ดังนั้นจะไม่มีระดับความมีนัยสำคัญที่จะพิจารณาได้
—	แอฟสไตน์ (Epstein)	ดัชนีความคลาดเคลื่อนเฉลี่ยยกกำลังสอง สร้างขึ้นมา เพื่อนิสฺจฺนข้อกระทงซึ่งให้ผลที่ไม่คงที่ในรูปแบบการตอบ มีพื้นฐานมาจากความคลาดเคลื่อนในแต่ละคนของคะแนน

ตารางที่ 2 (ต่อ)

สถิติ	แหล่งที่มา	ข้อเสนอแนะ
		<p>ในแต่ละข้อ การแปลความหมายดัชนีนี้มีประโยชน์มาก ในข้อกระทงแบบอิงเกณฑ์ ข้อเลือก การคำนวณยุ่งยากซับซ้อน ไม่มีค่าวิกฤติที่จะ พิสูจน์ข้อกระทงที่ไม่ดี</p>

3. การคัดเลือกข้อกระทง (Item Selection)

กระบวนการรวบรวมและวิเคราะห์ข้อมูลทั้งหมดที่กล่าวมา จะนำมาใช้ในการพิจารณาคุณภาพหรือประสิทธิภาพของข้อกระทงว่า ข้อกระทงมีคุณสมบัติตามจุดมุ่งหมายที่สร้างหรือไม่ ข้อกระทงที่ได้วิเคราะห์ความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ ค่าความยาก และค่าอำนาจจำแนก แล้วจะนำมาพิจารณา 3 กรณีคือ นำข้อกระทงไปรวมเป็นแบบสอบขั้นสุดท้าย ปรับปรุงก่อนนำไปรวมเป็นแบบสอบ หรือตัดทิ้งไปแล้วสร้างขึ้นมาใหม่

ลักษณะของข้อกระทงที่มีน้ำหนักมากที่สุดในการคัดเลือกคือ ความสอดคล้องระหว่างข้อกระทงกับจุดประสงค์ ค่าความยาก และค่าอำนาจจำแนก การพิจารณา อาจจะแยกจากกันหรือรวมกันก็ได้ การแนะนำสำหรับการคัดเลือกข้อกระทงทั้ง 3 ลักษณะแสดงไว้ในตารางที่ 3

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3 การแนะนำสำหรับการคัดเลือกข้อระทงแบบอิงเกณฑ์

ลักษณะของข้อระทง	เกณฑ์	ค่าดัชนี
ความสอดคล้องระหว่างข้อระทงกับจุดประสงค์	วิเคราะห์ความตรงเชิงเนื้อหาโดยคลินิจของผู้เชี่ยวชาญเนื้อหาวิชาตัดสิน	เป็นค่าบวกค่อนข้างสูง
ความยาก	ข้อระทงยากสำหรับกลุ่มที่ไม่ได้รับการสอน	0-50%
	ข้อระทงง่ายสำหรับกลุ่มที่ได้รับการสอน	70-100%
ค่าอำนาจจำแนก	ค่าอำนาจจำแนกเป็นบวกระหว่างกลุ่มที่เป็นเกณฑ์	เป็นค่าบวกค่อนข้างสูง

โดยทั่วไปข้อระทงที่เหมาะสมตามเกณฑ์ที่กำหนดไว้ก็จะนำไปรวมเป็นแบบสอบอย่างไรก็ตาม เมื่อข้อระทงมีเกณฑ์ไม่เหมาะสม การพิจารณาว่าจะยังคงไว้ ปรับปรุงหรือตัดทิ้งมีรายละเอียดที่จะพิจารณาดังต่อไปนี้

3.1 ความสอดคล้องระหว่างข้อระทงกับจุดประสงค์ หน้าที่หรือบทบาทของความสอดคล้องระหว่างข้อระทงและจุดประสงค์ในการคัดเลือกข้อระทง เป็นส่วนสำคัญของประสิทธิภาพของแบบสอบ และมีประโยชน์ในการแปลผล ข้อระทงที่ไม่สอดคล้องกับจุดประสงค์ ไม่ควรจะนำมาไว้ในแบบสอบ ในกรณีที่ข้อระทงได้รับการพิจารณาว่าตรงกับจุดประสงค์ แต่ปรากฏว่าไม่มี ความตรงตามค่าสถิติของข้อระทง หรือมีข้อผิดพลาดที่จะจำแนกระหว่างกลุ่มที่เป็นเกณฑ์ ข้อระทงนั้นควรจะรักษาสถานการณ์ต่อไปนี้

3.1.1 ความเฉพาะเจาะจงของจุดประสงค์เชิงพฤติกรรม เมื่อไม่มีข้อกระทงอื่น ๆ ที่จะสร้างขึ้นเพื่อวัดจุดประสงค์ดังกล่าว

3.1.2 ดัชนีค่าอำนาจจำแนกที่ต่ำหรือเป็นศูนย์ควรจะเป็นข้อกระทงที่วัดได้โดยตรง โดยไม่มีอิทธิพลของการสอนเข้ามาเกี่ยวข้อง ถ้าไม่เช่นนั้นควรจะปรับปรุงหรือสร้างข้อกระทงใหม่

3.2 ค่าความยากและค่าอำนาจจำแนก (Item difficulty and discrimination)

เหมือนกับความสัมพันธ์ระหว่างข้อกระทงกับจุดประสงค์ เพื่อให้แบบสอบมีคุณสมบัติตรงตามเนื้อหา ค่าอำนาจจำแนกของข้อกระทงที่นำมาพิจารณา เพื่อให้แบบสอบมีคุณสมบัติตรงตามการตัดสินใจ ชนิดของความตรงนี้สะท้อนถึงความถูกต้องที่แบ่งแยกนักเรียนออกเป็นผู้รอบรู้และไม่รอบรู้ในจุดประสงค์ การสอนหนึ่ง ข้อกระทงที่มีค่าอำนาจจำแนกสูงส่งเสริมความถูกต้องของการตัดสินใจ มีข้อโต้แย้งว่าการคัดเลือกข้อกระทงเพียงค่าอำนาจจำแนกที่ดีที่สุด จะสร้างแบบสอบที่ดีที่สุด ในเรื่องของการตัดสินใจ ซึ่งวิธีการแบบนี้มีแนวโน้มที่จะให้ความตรงในการตัดสินใจลดลง แต่จะเสียหายในด้านความตรงเชิงเนื้อหา ข้อกระทงที่เกี่ยวข้องกับจุดประสงค์เชิงพฤติกรรมที่เฉพาะเจาะจงจะถูกคัดทิ้งออกไปเมื่อข้อกระทงนั้นมีค่าอำนาจจำแนกต่ำหรือเป็นศูนย์ ดังนั้นในการคัดเลือกข้อกระทงควรพิจารณาเนื้อหาของจุดประสงค์ให้มากพอที่จะครอบคลุมค่าสถิติ และไม่ควรประเมินประสิทธิภาพของข้อกระทงโดยพื้นฐานจากค่าสถิติเพียงอย่างเดียว การแปลความหมายค่าดัชนีควรคำนึงถึงจุดประสงค์ที่วัด นักเรียนที่สอบ และโปรแกรมการสอนด้วย เกณฑ์ในการแปลความหมายค่าความยากและค่าอำนาจจำแนกดังแสดงไว้ในตารางที่ 3

4. การปรับปรุงข้อกระทง (Item revision)

เมื่อข้อมูลย้อนกลับหรือการวิเคราะห์ทางสถิติพบว่าข้อกระทงมีความผิดพลาด ผู้ร่างแบบสอบควรวิเคราะห์ข้อกระทงเพื่อหาข้อผิดพลาดนั้น โดยพิจารณาจากการตอบของนักเรียนทั้งก่อนสอนและหลังสอน (หรือพิจารณาจากกลุ่มที่ไม่ได้รับการสอนกับกลุ่มที่ได้รับการสอน) ข้อกระทงเหล่านี้จะมีค่าอำนาจจำแนกต่ำ เป็นศูนย์ หรือมีค่าอำนาจจำแนกเป็นลบ เกณฑ์ในการพิจารณาตัวเลือกและตัวลวงมีดังต่อไปนี้

4.1 การสอบก่อนการเรียน

- 4.1.1 ตัวเลือกถูกที่คิดควรมีจำนวนนักเรียนเลือกตบน้อยกว่าหรือเท่ากับตัวลวง
- 4.1.2 ตัวลวงที่คิดควรมีนักเรียนเลือกตบมากกว่าหรือเท่ากับตัวเลือกถูก รวม
ทั้งมีการกระจายของผู้ตอบในตัวลวงแต่ละตัวเท่าๆกัน

4.2 การสอบหลังเรียน

- 4.2.1 ตัวเลือกถูกที่คิดคือตัวเลือกถูกที่มีจำนวนนักเรียนเลือกตบมากกว่าตัวลวง
- 4.2.2 ตัวลวงที่คิดคือตัวลวงที่มีจำนวนนักเรียนเลือกตบน้อยกว่าตัวเลือกถูกยิ่ง
น้อยกว่ามากยิ่งเป็นตัวลวงที่คิด

ถ้าไม่เป็นไปตามเกณฑ์ดังกล่าว จะมีความผิดพลาดเกิดขึ้นได้ ควรจะปรับปรุงหรือสร้าง
ข้อกระทงใหม่ แต่ผู้สร้างแบบทดสอบควรจะต้องเลือกกรณีการสร้างข้อกระทงใหม่เป็นกรณีสุดท้ายจะดีที่สุด

ความเที่ยงของแบบสอบอิงเกณฑ์

ความเที่ยง (Reliability) เป็นดัชนีตัวหนึ่งที่สามารถบ่งบอกคุณภาพของแบบสอบได้
เป็นอย่างดี ดังนั้นในการสร้างแบบสอบทุกครั้งจึงจำเป็นต้องมีการตรวจหาความเที่ยงของแบบสอบ

ความเที่ยงของแบบสอบอิงเกณฑ์หมายถึง

1. ความคงที่ในการตัดสินใจจำแนกผู้รอบรู้และไม่รอบรู้ ในการสอบซ้ำด้วยแบบสอบ
ฉบับเดียวกัน หรือแบบสอบคู่ขนาน
2. ความคงที่ของคะแนนของผู้สอบแต่ละคนที่ เบี่ยงเบนไปจากจุดตัดในการสอบซ้ำด้วย
แบบสอบคู่ขนาน
3. ความคงที่ของคะแนนโดเมนของผู้สอบแต่ละคนในการสอบซ้ำด้วยแบบสอบคู่ขนาน

จากนิยามความเที่ยงของแบบสอบอิงเกณฑ์ทั้ง 3 รูปแบบดังกล่าว เราสามารถเลือก
นิยามใดนิยามหนึ่งในการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ก็ได้ อย่างไรก็ตาม รูปแบบ
แรก คือ รูปแบบความคงที่ ในการตัดสินใจจำแนกผู้รอบรู้ดูเหมือนว่าจะได้รับความนิยมมากกว่ารูป
แบบอื่นๆ ดังจะเห็นได้จากจุดประสงค์ในการสร้างแบบสอบอิงเกณฑ์ดังกล่าวแต่แรกแล้วว่า เป็น

แบบสอบที่สร้างขึ้นเพื่อใช้ผลการสอบจำแนกผู้สอบออกเป็นผู้รอบรู้และไม่รอบรู้ ดังนั้นความคงที่ในการตัดสินใจจำแนกผู้รอบรู้จึงมีความเหมาะสมที่จะใช้เป็นนิยามความเที่ยงของแบบสอบอิงเกณฑ์

วิธีการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์โดยพิจารณาจากความเที่ยงในการตัดสินใจจำแนกผู้รอบรู้ (Reliability of mastery classification decisions) เป็นการหาดัชนีความคงที่ในการตัดสินใจจำแนกผู้รอบรู้และไม่รอบรู้โดยการสอบซ้ำด้วยแบบสอบฉบับเดียวกัน หรือแบบสอบคู่ขนาน มีวิธีการประมาณค่าวิธีการต่าง ๆ ดังนี้

1. วิธีของคาร์เวอร์ (Carver)
2. วิธีของแฮมเบิลตันและโนวิก (Hambleton and Novick)
3. วิธีของสวามินาทาน แฮมเบิลตัน และอัลจินา (Swaminathan, Hambleton and Algena)
4. วิธีของสับโคเวียค (Subkoviak)
5. วิธีของฮวน (Huynh)

วิธีของสวามินาทาน แฮมเบิลตัน และอัลจินา (Swaminathan, Hambleton and Algena 1974)

สวามินาทาน แฮมเบิลตัน และอัลจินา ได้เสนอสูตรสำหรับการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์คล้ายกับวิธีของแฮมเบิลตันและโนวิก (Hambleton and Novick) การประมาณค่าความเที่ยงของแบบสอบตามวิธีนี้ จะมีการปรับแก้โดยหักเอาค่าความสอดคล้องโดยบังเอิญออกไป วิธีคำนวณอาจใช้แบบสอบฉบับเดียวสอบสองครั้ง หรือใช้แบบสอบคู่ขนานสองฉบับสอบนักเรียนกลุ่มเดียวกัน

$$K = [P_o - P_c] / [1 - P_c]$$

เมื่อ K แทนค่าความเที่ยงของแบบสอบถามซึ่งได้หักเอาค่าความคงที่
โดยบังเอิญออกไปแล้ว

P_c แทนสัดส่วนของความคงที่ที่คาดหวังว่าเกิดขึ้นโดยบังเอิญ

P_o แทนสัดส่วนของความคงที่ในการตัดสินใจจำแนกผู้รอบรู้

$$P_o = \sum_{i=1}^2 P_{i1}$$

$$P_c = \sum_{i,j=1}^2 P_{i.} P_{.j}$$

เมื่อ

P_{11} แทนสัดส่วนของผู้สอบที่ถูกตัดสินว่ารอบรู้ตรงกัน 2 ฉบับหรือ 2 ครั้ง

P_{12} แทนสัดส่วนของผู้สอบที่ถูกตัดสินว่ารอบรู้ในการสอบครั้งที่ 1 แต่ไม่รอบรู้
ในการสอบครั้งที่ 2

P_{21} แทนสัดส่วนของผู้สอบที่ถูกตัดสินว่า ไม่รอบรู้ครั้งที่ 1 แต่รอบรู้ครั้งที่ 2

P_{22} แทนสัดส่วนของผู้สอบที่ถูกตัดสินว่า ไม่รอบรู้ทั้งการสอบครั้งที่ 1 และครั้งที่ 2

$P_{1.}$ แทนสัดส่วนแยกของผู้สอบที่รอบรู้ในการสอบครั้งที่ 1

$P_{2.}$ แทนสัดส่วนแยกของผู้สอบที่ไม่รอบรู้ในการสอบครั้งที่ 1

$P_{.1}$ แทนสัดส่วนแยกของผู้สอบที่รอบรู้ในการสอบครั้งที่ 2

$P_{.2}$ แทนสัดส่วนแยกของผู้สอบที่ไม่รอบรู้ในการสอบครั้งที่ 2

ความตรงของแบบสอบถาม

ความตรงของแบบสอบถามมีรากฐานอยู่บนมโนคติ เกี่ยวกับทฤษฎีการวัดอิงเกณฑ์ที่
ว่าเครื่องมือที่ใช้ควรเป็นเครื่องมือที่สามารถ แยกผู้ที่เรียนแล้ว และผู้ที่ยังไม่ได้เรียนออกจาก
กัน เช่น ถ้าเครื่องมือที่ใช้เป็นแบบสอบ เมื่อนำแบบสอบไปทดสอบคนสองกลุ่ม คือกลุ่มผู้ที่ยังไม่เคย

เรียนเนื้อหาในแบบสอบและกลุ่มผู้ที่เคยเรียนมาแล้ว ผู้สอบกลุ่มหลังควรจะทำข้อนั้นได้มากกว่ากลุ่มผู้สอบแรก จากความคิดเห็น ความตรงในความหมายของการวัดอิงเกณฑ์ ก็คือ ความตรงเชิงจำแนก หรือความตรงในการตัดสินผู้สอบนั่นเอง คุณภาพของแบบสอบอิงเกณฑ์ในด้านนี้มีผู้ศึกษาไม่มากนัก จึงไม่ยุ่งยากซับซ้อน เหตุผลส่วนหนึ่งอาจเนื่องมาจากดัชนีที่ใช้คัดเลือกข้อกระทง ในการวิเคราะห์รายข้อก็เป็นตัวชี้ถึงความตรงโดยส่วนรวมของข้อกระทงเป็นข้อๆ อยู่แล้วก็อาจเป็นได้ ดัชนีแสดงความตรงเชิงจำแนกตัวแรกเสนอโดย คาร์เวอร์ (Carver 1970 อ้างถึงใน Crehen 1974) ได้ค่าความตรงจากข้อมูลที่ได้จากแบบสอบของกลุ่มผู้เรียนแล้วและผู้ยังไม่เคยเรียนดังตาราง

	ยังไม่เรียน	เรียนแล้ว
ผ่าน	b	a
ไม่ผ่าน	d	c

ค่าสัมประสิทธิ์ความตรงได้จาก $(a+d) / (a+b+c+d)$

สัมประสิทธิ์ความตรงของคาร์เวอร์จะมากหรือน้อยขึ้นอยู่กับจำนวนผู้สอบที่สอบไม่ผ่านก่อนเรียน และสอบผ่านหลังจากเรียน ดังนั้น องค์ประกอบที่ควรจะมีผลต่อความตรงของคาร์เวอร์ก็น่าจะเป็นความตรงตามเนื้อหาของข้อกระทงแต่ละข้อและประสิทธิภาพในการสอน

ดัชนีความตรงอีกตัวได้มาจากค่าสัมประสิทธิ์สหสัมพันธ์แบบพี ซึ่งมีลักษณะรวมทั้งองค์ประกอบที่มีผลต่อค่าที่ได้คล้ายกับของคาร์เวอร์ มีสูตรว่า

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

เมื่อ

- a คือจำนวนผู้ที่สอบผ่านหลังจากเรียน
- b คือจำนวนผู้ที่สอบผ่านก่อนเรียน
- c คือจำนวนผู้ที่สอบไม่ผ่านหลังจากเรียน
- d คือจำนวนผู้ที่สอบไม่ผ่านก่อนเรียน

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อกระตงแบบอิงเกณฑ์ในด้านความตรงเชิงเนื้อหา ได้แก่ การศึกษาความสอดคล้องระหว่างจุดประสงค์การเรียนรู้และข้อกระตงของคอห์ล (Dahl 1971: 6187 - A) โดยศึกษาตัวบ่งชี้ความสอดคล้อง 2 วิธีคือการวิเคราะห์และการตัดสิน ข้อกระตงสร้างมาจากจุดประสงค์การอ่าน แล้วนำไปสอบกับนักเรียนเกรด 2 และ 3 ในโรงเรียนระดับประถมศึกษา 3 โรงเรียนที่แตกต่างกันผลจากการสอบนำไปวิเคราะห์ โดยใช้วิธีการวิเคราะห์ตัวประกอบแบบโมโนโทนิค (Monotonic factor analysis method) และให้ผู้เชี่ยวชาญทางด้าน การอ่านตัดสินว่า ข้อกระตงสอดคล้องกับจุดประสงค์หรือไม่ แล้วนำผลการวิเคราะห์และการตัดสินของผู้เชี่ยวชาญมาเปรียบเทียบกัน ผลการวิจัยพบว่า วิธีการวิเคราะห์ที่สำคัญและเป็นเครื่องมือที่มีประโยชน์ สำหรับการวัดความสอดคล้องเพราะมีบางอย่างของความสอดคล้องที่ไม่สามารถตรวจสอบได้โดยการตัดสิน ข้อมูลจากการตัดสินจะเป็นการรับรองผลจากการวิเคราะห์อีกทีหนึ่ง

ส่วนการศึกษาค่าอำนาจจำแนกในการวิเคราะห์ข้อกระตงนั้น ได้มีการศึกษาค่าอำนาจจำแนกแบบอิงกลุ่มและแบบอิงเกณฑ์ ในการคัดเลือกข้อกระตง มีงานวิจัยดังต่อไปนี้

โอลสัน (Olson 1974: 5933-A) ได้ศึกษาเปรียบเทียบเทคนิคสำหรับการคัดเลือกข้อกระตงแบบอิงเกณฑ์ 3 วิธีคือ ดัชนีความตรงของข้อกระตงโดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียล ดัชนีอำนาจจำแนกแบบเก่า (The traditional item discrimination index) และดัชนีความไวในการสอนซึ่งได้มาจากสัดส่วนของจำนวนนักเรียนเกรด 4 ซึ่งตอบข้อกระตงได้อย่างถูกต้อง ลบด้วย สัดส่วนของนักเรียนเกรด 2 ซึ่งตอบข้อกระตงถูก ดัชนีความตรงของกระตงโดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียลศึกษาโดยใช้ตัวแปรเกณฑ์ 3 ตัว และนอกจากนี้ยังเปรียบเทียบ

ดัชนีความไวในการสอนกับดัชนีเกณฑ์ภายในซึ่งได้จาก ความแตกต่างของคะแนนในการสอบระหว่างกลุ่มที่รอรับรู้ คะแนนได้มาจากแบบสอบ 2 ชุดคือ แบบสอบโคลส และแบบสอบย่อยความเข้าใจในการอ่านจาก Comprehensive Tests of Basic Skills โดยวัดทักษะการอ่านทั้งหมด 16 ทักษะผลการวิจัยพบว่าดัชนีความตรงของข้อกระถงโดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียลจากตัวแปรเกณฑ์ภายนอกทั้ง 3 ตัว ไม่แตกต่างกัน ค่าสหสัมพันธ์ของดัชนีความตรงระหว่าง 2 แบบค่อนข้างสูง และแบบสอบที่มีความยากของข้อกระถงแตกต่างกัน (ง่าย ปานกลาง และยาก) ค่าของดัชนีความตรงจะคงที่

การเปรียบเทียบดัชนีความตรงของข้อกระถง ดัชนีเกณฑ์ภายใน และความไวในการสอน ได้ข้อค้นพบว่า เทคนิคทั้ง 3 แตกต่างกัน เป็นไปได้ว่า การตัดสินใจลักษณะของข้อกระถงจะแตกต่างกัน ขึ้นอยู่กับเทคนิคที่ใช้ และความสัมพันธ์ระหว่างดัชนีอำนาจจำแนกแบบเก่ากับดัชนีความตรงของข้อกระถง โดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียลมีค่าคงที่สรุปได้ว่า ดัชนีความตรงของข้อกระถงโดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียล นำมาใช้คัดเลือกข้อกระถงแบบอิงเกณฑ์ได้ และดัชนีอำนาจจำแนกแบบเก่ากับดัชนีความตรงของข้อกระถงโดยใช้สหสัมพันธ์แบบพอยท์ไบซีเรียลนำมาใช้ได้เหมือนกันเมื่อวัดทักษะเดียวกัน

เดวิส (Davis 1974: 4245-A) ได้ศึกษาเปรียบเทียบเทคนิคการคัดเลือกข้อกระถงแบบอิงเกณฑ์ 2 วิธี โดยการจำลองข้อมูลด้วยคลังของข้อกระถง (item pool) ซึ่งแตกต่างตามระดับค่าความยาก ข้อกระถงทั้งหมด 25 ข้อเรียงลำดับตามค่าอำนาจจำแนกจากเทคนิคที่ใช้คัดเลือกข้อกระถง คือ ดัชนีความแตกต่างของการสอบก่อนสอนและสอบหลังสอน (Dpp) นิจณาจากสัดส่วนของนักเรียนซึ่งตอบข้อกระถงถูกต้องก่อนสอบ หลังสอบ ลบด้วยสัดส่วนของนักเรียนซึ่งตอบข้อกระถงถูกต้องก่อนสอบ และค่าสถิติสหสัมพันธ์ (r_{Dpp}) คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนทั้งหมด แล้วนำมาสร้างเป็นข้อกระถงแบบอิงเกณฑ์จำนวน 12 ข้อ การจำลองคะแนนได้ตกจำนวน 1,000 ครั้ง จากตัวอย่าง 40, 70 และ 100 คน ผลการวิจัยพบว่า

1. ผู้เชี่ยวชาญด้านวัดผลและนักประเมินผลที่จะนำวิธีการคัดเลือกข้อกระถงไปใช้วัดประสิทธิภาพการสอนควรจะใช้ ดัชนีความแตกต่างของการสอบก่อนสอนและสอบหลังสอน (Dpp) ซึ่งนิจณาจากสัดส่วนของนักเรียนที่ตอบข้อกระถงถูกต้องก่อนสอบ หลังสอบ ลบด้วยสัดส่วนของนักเรียนที่ตอบข้อกระถงถูกต้องก่อนสอบ มากกว่าที่จะเลือกใช้สัมประสิทธิ์สหสัมพันธ์ (r_{Dpp}) คือสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนทั้งหมด

2. การเรียงลำดับข้อกระทงทั้ง 25 ข้อจากค่าอำนาจจำแนกทั้ง 2 วิธีแล้วหาค่าสหสัมพันธ์พบว่ามีความสัมพันธ์ทางลบอย่างมีนัยสำคัญ

3. ข้อกระทงที่คัดเลือกโดยใช้ดัชนีความแตกต่างของการสอบก่อนสอนและสอบหลังสอน (Dpp) ซึ่งพิจารณาจากสัดส่วนของนักเรียนที่ตอบข้อกระทงถูกต้องสอบหลังสอน ลบด้วย สัดส่วนของนักเรียนที่ตอบข้อกระทงถูกต้องสอบก่อนสอน จะแตกต่างกับข้อกระทงที่คัดเลือกโดยใช้สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนทั้งหมด (r_{xy})

และในปีค.ศ. 1977 สมิธ (Smith 1977: 4759-A) ได้ศึกษาผลของวิธีการคัดเลือกข้อกระทงแบบอิงเกณฑ์หลายวิธีในเรื่องของความถูกต้องและความคงที่ของการแบ่งแยกการศึกษาใช้วิธีมอนติคาร์โล กับโมเดลคุณลักษณะแฝงแบบโลจิสติก 2 พารามิเตอร์ จำลองการตอบ 1,000 ครั้ง ในกลุ่มข้อกระทง 70 ข้อ ข้อมูลที่ได้จะแตกต่างกันไปตามประสิทธิภาพการสอบ ซึ่งใช้การจำลองข้อลือกเหมือนกัน วิธีการคัดเลือกข้อกระทง 5 วิธีที่นำมาศึกษาคือ

1. สัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล (RPB)
2. ดัชนีของคอกซ์และวาร์กาส (CV)
3. ดัชนีของเบรนนอน (BR)
4. สัมประสิทธิ์แบบพี (PHI)
5. การเลือกแบบสุ่ม (RDM)

แล้วเรียงลำดับข้อกระทงจากค่าดัชนีที่ต่ำที่สุดตามลำดับ จะได้แบบสอบ 5 ชุดแต่ละชุดมี 40 ข้อกระทง การศึกษาความคงที่และความถูกต้องในการแบ่งแยกจากแบบสอบ 10 ชุดที่คู่ขนานกัน แต่ละชุดประกอบด้วยข้อกระทง 20 ข้อ โดยสร้างข้อกระทงจากที่คัดเลือกไว้ ผลการวิจัยพบว่า การคัดเลือกข้อกระทงแบบสุ่ม (RDM) ไม่คงที่ถึงแม้ว่าจะแบ่งแยกได้ถูกต้อง เมื่อเปรียบเทียบกับวิธีอื่น ๆ ส่วนวิธีการคัดเลือกข้อกระทงโดยใช้ดัชนีของเบรนนอน (BR) ซึ่งพิจารณาจากสัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกต้องของกลุ่มที่ได้รับการสอน ลบด้วย สัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกต้องของกลุ่มที่ไม่ได้รับการสอนและสัมประสิทธิ์แบบพี (PHI) เป็นวิธีที่ไม่ดี อาจจะมีสาเหตุเนื่องมาจาก ความผิดพลาดของวิธีการเหล่านี้ในการคัดเลือกข้อกระทงที่มีค่าอำนาจจำแนกปานกลางและสูง ส่วนวิธีการคัดเลือกข้อกระทงโดยใช้ สัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล (RPB) จะให้การแบ่งแยกที่ถูกต้องและคงที่เมื่อคะแนนสอบมีความเปลี่ยนแปลงสูง ถ้า

ไม่มีการเปลี่ยนแปลงคะแนนสอบ วิธีการคัดเลือกข้อกระทงโดยใช้ดัชนีของคอกซ์และวาร์กาสซึ่งพิจารณาจากสัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกในการสอบหลังสอน ไปด้วย สัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกในการสอบก่อนสอน จะดีกว่า

สรุปได้ว่าเมื่อนำค่าอำนาจจำแนกแบบอิงกลุ่ม และค่าอำนาจจำแนกแบบอิงเกณฑ์มาศึกษาร่วมกันในการคัดเลือกข้อกระทงแบบอิงเกณฑ์แล้ว ค่าอำนาจจำแนกแบบอิงเกณฑ์โดยเฉพาะค่าอำนาจจำแนกของคอกซ์และวาร์กาสซึ่งพิจารณาจาก สัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกในการสอบหลังสอน ไปด้วย สัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกในการสอบก่อนสอน จะเหมาะสมกับแบบสอบอิงเกณฑ์ และได้มีการนำมาศึกษามากที่สุด งานวิจัยต่อมาได้นำค่าอำนาจจำแนกแบบอิงเกณฑ์มาศึกษาร่วมกันในการวิเคราะห์ข้อกระทง มีงานวิจัยดังต่อไปนี้

แวดแมน (Wedman 1974: 64) ศึกษาดัชนีอำนาจจำแนก 5 แบบ กับนักเรียนมัธยมปลาย 285 คน และนักศึกษาคณะครู 20 คน ในวิชาทฤษฎีความน่าจะเป็นค่าอำนาจจำแนกที่นำมาศึกษา คือ ความแตกต่างของกลุ่มสอบก่อนและสอบหลังสอน (D:PP) ความแตกต่างของกลุ่มสูงกลุ่มต่ำ 27% (D:27%) ความแตกต่างของกลุ่มผู้ชำนาญและไม่ชำนาญ (D:ENE) ความแตกต่างของกลุ่มสูงกลุ่มต่ำ 13% (D:13%) และค่าสถิติไคสแควร์ของโฟลแฟม พบว่า ส่วนใหญ่มีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างต่ำ ที่นับว่าสูงได้แก่ สัมประสิทธิ์สหสัมพันธ์ระหว่าง ค่าสถิติไคสแควร์ กับความแตกต่างของกลุ่มสูงต่ำ 27% (D:27%) ค่าสถิติไคสแควร์กับความแตกต่างของกลุ่มผู้ชำนาญและไม่ชำนาญ (D:ENE) และค่าสถิติไคสแควร์กับความแตกต่างของกลุ่มสูงกลุ่มต่ำ 13% (D:13%) คือ 58 .52 และ .75 ตามลำดับ สัมประสิทธิ์สหสัมพันธ์ระหว่าง ความแตกต่างของกลุ่มสูงกลุ่มต่ำ 27% (D:27%) กับความแตกต่างของกลุ่มสอบก่อนและสอบหลังสอน (D:PP) เท่ากับ .62 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง ค่าอำนาจจำแนกที่พิจารณาจาก ความแตกต่างของกลุ่มสูงกลุ่มต่ำ 27% (D:27%) กับความแตกต่างของกลุ่มสูงกลุ่มต่ำ 13% (D:13%) มีค่าสูงสุดคือ .91

ค.ศ. 1977 แธรช (Thrash 1977: 247-A) เปรียบเทียบค่าอำนาจจำแนก 3 ชนิด คือ ค่าอำนาจจำแนกของคอกซ์และวาร์กาส (C-V) ค่าอำนาจจำแนกของรูดาบุช (R) และค่าอำนาจจำแนกของเบรนนอนและสจิวท (B-S) โดยศึกษากับจุดประสงค์ 16 ข้อ ในวิชาคณิตศาสตร์และการอ่าน ทดสอบกับนักเรียน 2 ระดับคือ ชั้นมัธยมศึกษาและระดับที่สูงขึ้นไป มีการทดลอง 2 แบบคือ แบบที่ผู้วิจัยกำหนดจุดประสงค์ กับแบบที่นักเรียนเป็นผู้เลือกจุดประสงค์ในการสอบ พบว่า

ดัชนีอำนาจจำแนกทั้ง 3 ชนิด นำมาใช้วิเคราะห์แบบสอบวิชาคณิตศาสตร์ได้ดีกว่าการอ่าน เมื่อพิจารณาจากการทดลองพบว่า การทดลองแบบที่ผู้วิจัยกำหนดจุดประสงค์ จะดีกว่า การทดลองแบบที่ให้นักเรียนเป็นผู้เลือกจุดประสงค์ในการสอบ ส่วนระดับชั้นให้ผลคล้ายกัน สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าอำนาจจำแนกของคอกซ์และวาร์กาส (C-V) และค่าอำนาจจำแนกของรูดาบซ (R) เท่ากับ .80 ค่าอำนาจจำแนกของคอกซ์และวาร์กาส (C-V) กับค่าอำนาจจำแนกของเบรนนอนและสจีวิต (B-S) เท่ากับ .70 และค่าอำนาจจำแนกของเบรนนอนและสจีวิต (B-S) กับค่าอำนาจจำแนกของรูดาบซ (R) เท่ากับ .36 สำหรับกลุ่มตัวอย่างขนาดใหญ่ค่าอำนาจจำแนกของคอกซ์และวาร์กาส (C-V) ให้ค่าประมาณดีกว่า ค่าอำนาจจำแนกของรูดาบซ (R) และในการวิเคราะห์รายข้อของการสอบก่อนเรียนและหลังเรียน แทรชแนชนะนำให้ใช้วิธีของคอกซ์และวาร์กาส (C-V)

งานวิจัยการวิเคราะห์ข้อกระทงที่กล่าวมาอาศัยสถิติแบบคลาสสิก ซึ่งมีข้อจำกัด (บุญเชิด วิทยุโณนัตตพงษ์ 2527: 91) เช่น ค่าสถิติของข้อกระทงที่ซับซ้อนคุณภาพ อันได้แก่ค่าความยาก และค่าอำนาจจำแนก ขึ้นอยู่กับกลุ่มตัวอย่างนักเรียน การเปรียบเทียบผู้สอบแต่ละคนต้องใช้เซตของข้อกระทงเดียวกันหรือค่านานกันจึงจะเปรียบเทียบกันได้ และความแปรปรวนของความคลาดเคลื่อนในการวัดจากผู้สอบแต่ละคน จะสมมติไว้ล่วงหน้าว่ามีค่าเท่ากันหมด จากข้อจำกัดดังกล่าว ทำให้สถิติคลาสสิกไม่สามารถนำมาใช้แก้ปัญหาการทดสอบได้ในบางแง่มุมปัญหา จึงมีผู้พัฒนาทฤษฎีคุณลักษณะแฝงขึ้นมา และสามารถนำมาใช้ในการวิเคราะห์ข้อกระทงเพื่อตรวจหาคุณภาพเป็นรายข้อและทั้งฉบับได้ งานวิจัยที่ศึกษาค่าอำนาจจำแนกแบบอิงเกณฑ์ และค่าอำนาจจำแนกตามทฤษฎีคุณลักษณะแฝง มีดังนี้

ฮาลาไดนา และรอยด์ (Haladyna and Roid 1981: 39-51) ได้ศึกษาดัชนีความไวในการสอนหรือค่าอำนาจจำแนก ในการวิเคราะห์ข้อกระทงแบบอิงเกณฑ์มีวัตถุประสงค์เพื่อตรวจสอบลักษณะของข้อกระทงที่แตกต่างกัน ซึ่งคำนวณค่าดัชนีของข้อกระทงจากกลุ่มตัวอย่าง ที่มีประสิทธิภาพการสอนที่แตกต่างกัน ดัชนีความไวในการสอนที่ศึกษาได้แก่

1. ดัชนีของคอกซ์และวาร์กาสซึ่งพิจารณาจากค่าความยากก่อนสอบและสอบหลังสอบ (DISppd)
2. ดัชนีของเบรนนอนและสโตรโลว์ (Brennan and Strolurow 1971)

3. ดัชนีค่าอำนาจจำแนกสหสัมพันธ์แบบพอยต์ไบซีเรียลโดยรวมกลุ่มตัวอย่าง (COMPBI)
4. ดัชนีความไวในการสอนจากทฤษฎีการตอบข้อกระทง (Item Response Theory)
5. ดัชนีความไวในการสอนจากทฤษฎีของเบส์ (Bayes's theorem)

พบว่าดัชนีของคอกซ์และวาร์กาล (DISppd) คำนวณและแปลความหมายได้ง่ายที่สุด และผลจากการศึกษาแสดงว่า ดัชนีของคอกซ์และวาร์กาล (DISppd) มีสหสัมพันธ์สูงกับดัชนีความไวในการสอนที่กล่าวมาทั้งหมด

แฮร์ริส (Harris 1983:2741-A) ได้ศึกษาการคัดเลือกข้อกระทงสำหรับแบบสอบรอบรู้ โดยเปรียบเทียบกระบวนการ 3 วิธีซึ่งประกอบด้วย

1. DISppd ของคอกซ์และวาร์กาลพิจารณาจากสัดส่วนของการตอบถูกต้องหลังสอนลบด้วยสัดส่วนของการตอบถูกต้องก่อนสอน
2. $I(0c)$ เป็นดัชนีที่มีพื้นฐานมาจากทฤษฎีคุณลักษณะแฝง
3. $P(Xc)$ เป็นดัชนีรวมกันระหว่าง DISppd และ $I(0c)$

การศึกษานี้ได้จำลองลักษณะของข้อมูลหลายๆแบบ ข้อกระทงเรียงลำดับตามค่าดัชนีทั้ง 3 แล้วหาสหสัมพันธ์แบบ Spearman rank order correlations ผลการวิจัยพบว่า $P(Xc)$ เป็นดัชนีที่รวมกันระหว่าง DISppd ของคอกซ์และวาร์กาลซึ่งพิจารณาจากสัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกต้องในการสอบหลังสอน ลบด้วย สัดส่วนของจำนวนนักเรียนที่ตอบข้อกระทงถูกต้องในการสอบก่อนสอน กับ $I(0c)$ ซึ่งเป็นดัชนีที่มีพื้นฐานมาจากทฤษฎีคุณลักษณะแฝง เป็นดัชนีที่เหมือนกับ $I(0c)$ และในการนำไปใช้คัดเลือกข้อกระทงในการสร้างแบบสอบอิงเกณฑ์ในชั้นเรียน ควรใช้ดัชนี $P(Xc)$

สำหรับผู้ศึกษาเปรียบเทียบการวิเคราะห์รายข้อของแบบสอบอิงเกณฑ์ ที่มีผลต่อค่าความเที่ยงและความตรงของแบบสอบคือ เกรเฮน (Crehan 1974: 255-269) ซึ่งใช้วิธีการคัดเลือกข้อกระทงเพื่อแบ่งเป็นแบบสอบคู่ขนาน 6 วิธีคือ

1. เรียงลำดับข้อกระทงตามวิธีหาค่าอำนาจจำแนกของคอกซ์และวาร์กาล ซึ่งพิจารณาจากสัดส่วนของการตอบถูกหลังสอน ลบด้วยสัดส่วนของการตอบถูกก่อนสอน
2. เรียงลำดับข้อกระทงตามวิธีหาค่าอำนาจจำแนกของเบรนนอน ซึ่งพิจารณาจากสัดส่วนการตอบถูกของคะแนนที่สูงกว่าจุดตัด ลบด้วยสัดส่วนของการตอบถูกของคะแนนที่ต่ำกว่าจุดตัด
3. เรียงลำดับตามสัดส่วนการตอบถูกของผู้สอบหลังเรียน
4. เรียงลำดับตามการเลือกของครูผู้สร้างแบบสอบ
5. เรียงลำดับตามค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล และ
6. เรียงแบบสุ่ม

พบว่าวิธีการสร้าง 6 วิธี ให้ผลต่อความเที่ยงของแบบสอบไม่แตกต่างกัน ส่วนค่าอำนาจจำแนกของคอกซ์และวาร์กาล และดัชนีอำนาจจำแนกของเบรนนอน ให้ค่าความตรงของแบบสอบสูงสุด

ในประเทศไทยงานวิจัยเกี่ยวกับการวัดผลแบบอิงเกณฑ์ส่วนมากจะเป็นการสร้างแบบสอบ การศึกษาความเที่ยงจากคะแนนจุดตัด ในการศึกษาจะนำค่าอำนาจจำแนกแบบอิงเกณฑ์มาศึกษาเพียงวิธีใดวิธีหนึ่งเท่านั้น งานวิจัยเกี่ยวกับการเปรียบเทียบความตรงและความเที่ยงของแบบสอบอิงเกณฑ์ โดยการคัดเลือกข้อกระทงด้วยวิธีหาค่าอำนาจจำแนกที่แตกต่างกันยังไม่มีผู้ใดศึกษาวิจัยมาก่อน ดังนั้นการวิจัยที่ศึกษาในครั้งนี้จึงวิเคราะห์ประเด็นของความตรงและความเที่ยงของแบบสอบอิงเกณฑ์ โดยการคัดเลือกข้อกระทงด้วย วิธีหาค่าอำนาจจำแนกที่แตกต่างกัน 3 วิธีคือ วิธีหาค่าอำนาจจำแนกของคอกซ์และวาร์กาล วิธีหาค่าอำนาจจำแนกของรูตาบรู และวิธีหาค่าอำนาจจำแนกของโคชีคอฟฟ์และโคลน์