



บทที่ 2

วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้องในการศึกษาวิจัยครั้งนี้ ผู้วิจัยได้เสนอเป็นหัวข้อย่อย ดังนี้

1. การวัดแบบอิงเกณฑ์ และการสร้างแบบสอบอิงเกณฑ์
2. แบบสอบปรนัย เลือกตอบและตอบสั้น
3. การวิเคราะห์แบบสอบอิงเกณฑ์
4. การกำหนดคะแนนจุดตัดหรือเกณฑ์
5. งานวิจัยที่เกี่ยวข้อง

โดยมีรายละเอียดแยกแต่ละหัวข้อดังนี้

การวัดแบบอิงเกณฑ์และการสร้างแบบสอบอิงเกณฑ์

ความสำคัญของการวัดและการวัดแบบอิงเกณฑ์

การเรียนการสอนในห้องเรียนนั้น การวัดผลและประเมินผลเป็นกระบวนการสำคัญที่ทำให้ครู และนักเรียน ได้ทราบผลของการเรียนการสอน การวัดผลและประเมินผลแต่เดิมนั้น เป็นการวัดผลแบบอิงกลุ่ม (Norm - Referenced Measurement) ซึ่งมีแนวคิดและความเชื่อในเรื่องของความแตกต่างระหว่างบุคคล ซึ่งแต่ละบุคคลในกลุ่มย่อมมีความสามารถต่าง ๆ กัน การวัดผลแบบอิงกลุ่มเป็นที่รู้จักแพร่หลายในหมู่นักการศึกษา ครู-อาจารย์ แต่การวัดผลประเภทนี้ไม่ช่วยให้ครู ได้ทราบว่า "นักเรียนมีความสามารถหรือไม่สามารถอะไรบ้าง" (Glaser 1963; Hambleton and others 1978; Popham and Husek, 1969) ปัจจุบันการศึกษาได้เปลี่ยนไปจากแนวเดิม โดยเน้นการเรียนการสอนแบบโปรแกรมมากขึ้น เนื่องจากพบว่า นักเรียนแต่ละคนมีความแตกต่างกันมากในค่านต่าง ๆ เช่น ความถนัด

สติปัญญา ความสนใจ และสภาพแวดล้อมต่าง ๆ จุดประสงค์ที่สำคัญของการเรียนแบบโปรแกรมคือเพิ่มโอกาสการปฏิบัติ ทักษะหรือความสามารถในการเรียนของนักเรียนให้มากที่สุด แบบสอบที่ใช้ควรให้ข่าวสารที่แสดงถึงความก้าวหน้าในช่วงความต่อเนื่องของความสามารถสมบูรณ์ อีกทั้งแสดงถึงจุดเด่นจุดด้อยของนักเรียน ทั้งนี้เพื่อให้สามารถจัดสอนซ่อมเสริมได้อย่างมีประสิทธิภาพมากที่สุด ซึ่งในแบบสอบอิงกลุ่มนั้นสร้างขึ้นเพื่อเปรียบเทียบระหว่างนักเรียนในกลุ่ม ไม่เหมาะสำหรับตัดสินการเรียนการสอนแบบโปรแกรม อันเป็นการเรียนการสอนรายบุคคล (Hambleton, R.K. 1974 : 372) แบบสอบที่เหมาะสมสำหรับการวัดความก้าวหน้า (Formative) และการวินิจฉัย (Diagnostic) นักเรียนเป็นรายบุคคลนั้น ควรเป็นแบบสอบอิงเกณฑ์ (Criterion - Referenced Tests) (Gränlund, 1977; Hambleton and others 1978 : 2)

การวัดอิงกลุ่ม ขึ้นอยู่กับเกณฑ์มาตรฐานสัมพัทธ์ (Relative Standard) หรือพฤติกรรมของกลุ่มเป็นสำคัญ จึงมุ่งหาความแตกต่างระหว่างผู้เรียน เพราะเชื่อว่าความสามารถของบุคคลในกลุ่มใด ๆ ในเรื่องใดเรื่องหนึ่งมีไม่เท่ากัน บางคนมีความสามารถเด่น บางคนมีความสามารถด้อย แต่บุคคลที่มีความสามารถเด่นและด้อยจะมีน้อย ส่วนใหญ่จะมีความสามารถปานกลาง การกระจายความสามารถของผู้เรียนจึงคล้ายโค้งรูปประฆัง (Normal curve) (สมศักดิ์ สิ้นธุระเวช 2521 : 74) ทั้งนี้จึงยึดคนส่วนใหญ่หรือกลุ่มเป็นหลักในการเปรียบเทียบ โดยพิจารณาคะแนนของนักเรียนผู้นั้น เปรียบเทียบกับคะแนนของนักเรียนคนอื่น ๆ ภายในกลุ่ม จึงเป็นแรงกระตุ้นก่อให้เกิดการเรียนรู้แบบแข่งขัน

การวัดอิงเกณฑ์ เป็นวิธีการวัดที่รู้จักกันมานานแล้ว Glaser (1963) และ Popham และ Husek (1969) เป็นบุคคลแรกที่นำเสนอและเชี่ยวชาญในเรื่องการวัดอิงเกณฑ์ (Criterion - Referenced Testing) แรงจูงใจที่ทำให้นักการศึกษาสนใจการวัดอิงเกณฑ์ ซึ่งเนื่องมาจากการใช้การวัดแบบอิงกลุ่มนั้นเกิดข้อวิพากษ์วิจารณ์ ตามที่แอร์ราเซียน และมาคอส (Airasian and Madaus, 1974 : 76-77) เสนอไว้ 4 ประการ ดังนี้

1. การใช้แบบสอบถามให้ข้อมูลไม่เพียงพอ ในเรื่องความหมายของการวัด และไม่เหมาะสมกับเนื้อหาบางประการที่ตองการวัด

2. การให้ระดับคะแนน (Grade) ทำให้จุดหมายของการศึกษาเปลี่ยนไป กลายเป็นการแข่งขัน เพื่อไต่คะแนนสูง ๆ และกลายเป็นสิ่งตองการเข้าศึกษา การสมัครงาน ฯลฯ ซึ่งระดับคะแนนไม่ให้ข้อมูลว่า เด็กทำอะไรได้บ้าง นอกเพียงว่าเขาดีหรือค้อยกว่าผู้อื่นในกลุ่มเท่านั้น

3. การวิวัฒนาการทางเทคนิคการสอนทำให้พบว่า แบบสอบอิงกลุ่มไม่เพียงพอที่จะประเมินพฤติกรรมเป็นรายบุคคล หรือประเมินวิธีสอนว่าดีหรือไม่เพียงใด

4. นักการศึกษาส่วนมากเชื่อว่า ทุกคนหรือคนส่วนมาก สามารถเรียนรู้ หรือสัมฤทธิ์ผลในวิชาต่าง ๆ ได้ถ้ามีวิธีการเรียนการสอนที่ดีกว่า จึงจะทำให้การทดสอบเปลี่ยนไปจากการเปรียบเทียบในกลุ่ม เป็นการตรวจสอบและให้รางวัลในการเรียนในรูปของการกระทำ เช่น ถ้าทุกคนทำได้ถึงเกณฑ์ที่ยอมรับแล้วอาจจะได้ "A"

เขาวที วิบูลย์ศรี (2526) ได้เสนอข้อจำกัดของการใช้แบบสอบอิงกลุ่มไว้ดังนี้

1. ขาดประสิทธิภาพในการวินิจฉัยจุดอ่อน ทั้งในค้ำนผลสัมฤทธิ์ของนักเรียน และค้ำนการจึคกระบวนกรเรียนการสอนของครู

2. การประเมินผลระดับความรู้ของนักเรียน จากขอบข่ายเนื้อหาที่จำกัด ไม่สามารถตรวจสอบพฤติกรรมที่ครอบคลุมจุดประสงค์ของการเรียนการสอนที่จะนำไปสู่ผลิผลกรเรียนรู้ที่แท้จริงได้

3. การประเมินผลหลักสูตรก็เป็นเพียงการตรวจสอบจากผลสัมฤทธิ์ของนักเรียนที่เปลี่ยนแปลงไปเท่านั้น ไม่สามารถวิเคราะห์ไปถึงจุดประสงค์ของหลักสูตรว่า ได้บรรลุตามพฤติกรรมหรือไม่เพียงใด

4. การประเมินผลจุดเด่น จุดอ่อนของโปรแกรมการศึกษา จำกัดเฉพาะนักเรียนในกลุ่มโรงเรียนที่ใช้แบบสอบเท่านั้น ไม่สามารถอ้างอิงถึงสถานะภาพที่แท้จริงของโปรแกรมการศึกษานั้นว่ามีประสิทธิภาพเพียงใด

โดยสรุปแล้วจะเห็นว่า แบบสอบอิงกลุ่ม หรือการวัดแบบอิงกลุ่มไม่เหมาะที่จะนำมาใช้วัดผลสัมฤทธิ์ทางการศึกษาในบางสภาพการณ์ โดยเฉพาะอย่างยิ่งเพื่อการประเมินความก้าวหน้า และการวินิจฉัยจุดเด่น จุดด้อย เพื่อหาข้อบกพร่อง และปรับปรุงแก้ไข ในระหว่างการเรียนการสอนแบบโปรแกรมซึ่งเน้นหนักการเรียนการสอนเป็นรายบุคคล และเปิดโอกาสให้ผู้เรียนซึ่งมีความแตกต่างกันอยู่แล้วได้มีโอกาสเรียนรู้ได้ตามจุดประสงค์การเรียนการสอน ทั้งนี้แบบสอบและการวัดที่เหมาะสมที่สุดควรจะเป็น แบบอิงเกณฑ์ (Criterion - Referenced)

ในการวัดอิงเกณฑ์ เกลเซอร์ (Glaser 1963 : 519-520) กล่าวว่า มโนทัศน์ที่เน้นในการวัดผลสัมฤทธิ์ คือ ความคิดเกี่ยวกับความต่อเนื่องของความรู้ที่ได้เรียนมา ซึ่งจะเรียงลำดับจากการไม่มีเลยไปจนถึงการมีอย่างสมบูรณ์ ระดับผลสัมฤทธิ์ของบุคคลจะตกอยู่ที่จุดใดจุดหนึ่งในช่วงของความต่อเนื่องนี้ ซึ่งจะชี้ได้โดยพฤติกรรมที่แสดงออกในระหว่างการสอบ ระดับผลสัมฤทธิ์ที่จะแสดงว่าเป็นการปฏิบัติที่น่าพอใจหรือไม่ จะประเมินได้ด้วยวิธีการวัดแบบอิงเกณฑ์... เกณฑ์ที่ใช้ในลักษณะนี้ไม่จำเป็นต้องอ้างอิงถึงพฤติกรรมสุดท้ายหลังจากเรียนจบแล้ว ระดับของเกณฑ์สามารถกำหนดที่จุดใด ๆ ของการสอบก็ได้ถ้าจำเป็น เพื่อให้ได้ข้อมูลเกี่ยวกับการปฏิบัติของแต่ละคนอย่างพอเพียง... คะแนนของนักเรียนที่ได้จากการวัดแบบอิงเกณฑ์จะให้ข้อมูลที่แสดงไว้อย่างชัดเจนว่า แต่ละคนสามารถหรือไม่สามารถทำอะไรได้บ้าง

การวัดอิงเกณฑ์ (Criterion - Referenced Measurement)

สืบเนื่องมาจากทฤษฎีการเรียนรู้แบบ การเรียนเพื่อรอบรู้ (Mastery Learning) ของบลูม (Bloom) ที่กล่าวว่า "การเรียนรู้ทั้งหลาย ควรจะเป็นการเรียนเพื่อรอบรู้ ในเนื้อเรื่องต่าง ๆ ครูจึงต้องการให้นักเรียนทุกคน บรรลุสู่ความเป็นผู้รอบรู้ หรือควรจะเป็น 80-90% ของเนื้อหาที่เรียนรู้" (สมศักดิ์ สินธุวะเวช 2521 : 74-75) การประเมินผลแบบอิงเกณฑ์ ถือว่าเนื้อหาวิชาที่สอนเป็นสิ่งสำคัญ และจำเป็นที่นักเรียนควรจะรู้ และการที่จะชี้ชัดลงไปว่านักเรียนคนใดเรียนรู้หรือไม่รู้จึงขึ้นอยู่กับเกณฑ์มาตรฐาน ที่ให้ความหมายไว้ชัดเจนล่วงหน้า ซึ่งมุ่งที่จะประเมินว่านักเรียนสามารถหรือไม่สามารถทำอะไรได้บ้าง มากกว่าที่จะเปรียบเทียบว่านักเรียนดี หรือด้อยอย่างไรในกลุ่ม จึงผลักดันให้เกิดการเรียนรู้แบบเกื้อกูลกัน

โพเพม (Popham 1975) ใ้ค้้นิยามแบบสอบอิงเกณฑ์ไว้ว่า "แบบสอบอิง-
เกณฑ์เป็นแบบสอบที่โยงถึงสถานภาพของผู้สอบเป็นรายบุคคล ซึ่งอ้างอิงถึงคะแนน
โคเมน ทั้งนี้ต้องกำหนดพฤติกรรมในโคเมนที่ชัดเจน (Hambleton and others
1978 : 2)

เกลเซอร์และนิตโก (Glaser and Nitko 1971) ได้กล่าวถึงแบบสอบ-
อิงเกณฑ์ไว้ว่า แบบสอบอิงเกณฑ์เป็นแบบสอบที่สร้างขึ้นโดยมีเจตนาให้เป็นการวัดที่
สามารถแปลความหมายออกมาได้ตรง ๆ ในรูปของมาตรฐานการกระทำที่เฉพาะ ซึ่ง
ในที่นี้บ่งชี้ได้โดยกำหนดโคเมนของงานที่นักเรียนหรือผู้สอบควรทำได้ จากนั้นนำตัว-
อย่างที่เป็นตัวแทนของงานจากโคเมนนี้มาจัดเป็นแบบสอบ ผลการวัดจะแสดงถึง
พฤติกรรมของแต่ละคนโดยเทียบกับโคเมนที่กำหนดไว้

มิลแมน (Millman 1974) กล่าวถึง แบบสอบอิงเกณฑ์ในรูปของแบบสอบ
อิงโคเมน (Domain - Referenced Tests) ซึ่ง สามารถบรรยายถึงการกระทำ
ได้ดีกว่า โดยกล่าวว่า การสอบอิงโคเมน เป็นการวัดที่เป็นปัจจุบันของนักเรียนแต่ละ
บุคคล โดยอ้างอิงถึงชุดของงานที่กระทำได้ ซึ่งบรรยายไว้อย่างชัดเจนเรียกว่า
Domain แบบสอบอิงโคเมนคือ แบบสอบที่ประกอบไปด้วยกลุ่มตัวอย่างของข้อกระทง
จากโคเมน ซึ่งอาจจะได้มาจากการสุ่มอย่างง่ายหรือสุ่มแยกประเภท (Stratified
Sampling)

จากคำกล่าวข้างต้นจะเห็นว่า การวัดแบบอิงเกณฑ์เป็นการวัดที่เน้นการแปล
ความหมายของแต่ละบุคคล โดยเทียบกับเกณฑ์ หรือ โคเมนของงานที่ได้บรรยายไว้
อย่างชัดเจน เพื่อที่จะบ่งชี้ว่า นักเรียนแต่ละคนสามารถหรือไม่สามารถทำอะไรได้บ้าง

กรอนสันด์ (Gronlund, N.E. 1976 : 139) ได้เสนอแนวคิดเกี่ยวกับการ
การตัดสินใจ ในเรื่องการใช้แบบสอบที่เหมาะสม โดยพิจารณาจุดประสงค์การสอบและ
ลักษณะการวัดผลสัมฤทธิ์ 4 ชนิด ดังตารางที่ 19



ตารางที่ 1 สรุป รูปแบบการทดสอบและการใช้แบบสอบ

	วางตัวบุคคล		วัดความก้าวหน้า	วินิจฉัย	วัดรวมสรุป
	ความพร้อม	จัดตำแหน่ง	Formative	Diagnostic	Summative
จุดมุ่งหมายของการวัด	วัดทักษะเบื้องต้น	วัดจุดประสงค์ของงานหรือเนื้อหาวิชา	วัดถึงการเปลี่ยนแปลงการเรียนการสอน	วัดเพื่อหาความบกพร่องในการเรียน	วัดจุดประสงค์ของงานหรือเนื้อหาวิชา
ความยากของข้อสอบ	ปกติข้อสอบมีระดับความยากต่ำ	ข้อสอบมีระดับความยากกระจายกว้าง	ความยากของข้อสอบเปลี่ยนแปลงตามการเรียนการสอน	ข้อสอบมีระดับความยากต่ำ	ข้อสอบมีระดับความยากกระจายกว้าง
การบริหารการสอบ	สอบเมื่อเริ่มงานหรือรายวิชา	สอบเมื่อเริ่มงานหรือรายวิชา	สอบระหว่างการเรียนการสอน	สอบระหว่างการเรียนการสอน	สอบเมื่อจบหน่วยหรือรายวิชา
ลักษณะเครื่องมือ	เป็นแบบสอบอิงเกณฑ์	เป็นแบบสอบอิงกลุ่ม	เป็นแบบสอบอิงเกณฑ์เพื่อวัดความรอบรู้	เป็นแบบสอบอิงเกณฑ์เพื่อวัดจุดบกพร่อง	เป็นแบบสอบอิงกลุ่ม
ผลของการใช้	ช่วยแก้ไขส่วนที่สูญเสียหรือไม่ปกติหรือกำหนดกลุ่มก่อนเรียน	ช่วยวางแผนการเรียนการสอนวางตัวบุคคล	ให้ข้อมูลในการเรียนการสอนปรับปรุงการเรียนการสอน	ช่วยแก้ไขจุดบกพร่องที่สังเกตเห็นถึงความยากในการเรียน	ให้ระดับคะแนนรับรองความรู้ประเมินผลการศึกษาสอน

การสร้างแบบสอบอิงเกณฑ์

Millman (1973) ได้กล่าวถึงลักษณะของข้อสอบอิงเกณฑ์ ซึ่งสรุปได้ว่าข้อสอบอิงเกณฑ์นั้นความแปรปรวนไม่ใช่สิ่งจำเป็นเหมือนข้อสอบแบบอิงกลุ่ม ในการสร้างข้อสอบแบบอิงเกณฑ์นั้น จะสร้างให้แต่ละข้อวัดได้ตามจุดประสงค์เชิงพฤติกรรม โดยไม่คำนึงถึงความยากง่าย และอำนาจจำแนก เพื่อให้แบบสอบมีความเที่ยง จะต้องสร้างให้ข้อสอบอิงเกณฑ์มีความคงที่ภายในสูง และวัดความรู้ในเรื่องเดียวกัน

ไพศาล หวังพานิช (2526 : 189-190) ได้เสนอลักษณะและแนวการสร้างข้อสอบอิงเกณฑ์ไว้ว่า แบบสอบอิงเกณฑ์นั้น สิ่งที่น่าสนใจที่สุดคือ ข้อสอบที่ใช้ข้อวัดตามจุดมุ่งหมายอย่างแท้จริง อีกประการหนึ่ง การสอบกับการสอนจะประสานสัมพันธ์กันตลอดเวลา ดังนั้นความยุ่งยากในการสร้างข้อสอบจึงมีน้อยคงมีเพียงเรื่องความตรง (Validity) ตามจุดมุ่งหมาย และความเป็นปรนัย หรือความแจ่มชัดของข้อคำถามเท่านั้น งานสร้างข้อสอบจึงขึ้นอยู่กับจุดมุ่งหมายของการสอนเป็นสำคัญ

Berk (1980) และ Hambleton and others (1978) ก็ได้กล่าวถึงการสร้างข้อสอบอิงเกณฑ์ที่สำคัญที่สุดคือ ความตรงเชิงเนื้อหา (Content Validity) คือ ความสอดคล้องของข้อสอบกับจุดประสงค์การเรียนการสอน ซึ่งในการสร้างข้อสอบอิงเกณฑ์ครั้งนี้ผู้วิจัยเน้นความสำคัญของการสร้างข้อสอบให้สอดคล้องกับจุดประสงค์เชิงพฤติกรรมหลักที่กำหนดไว้ ซึ่งมีนักการศึกษาหลายคนได้ให้แนวคิดและหลักการสร้างแบบสอบอิงเกณฑ์โดยทั่วไป ดังต่อไปนี้

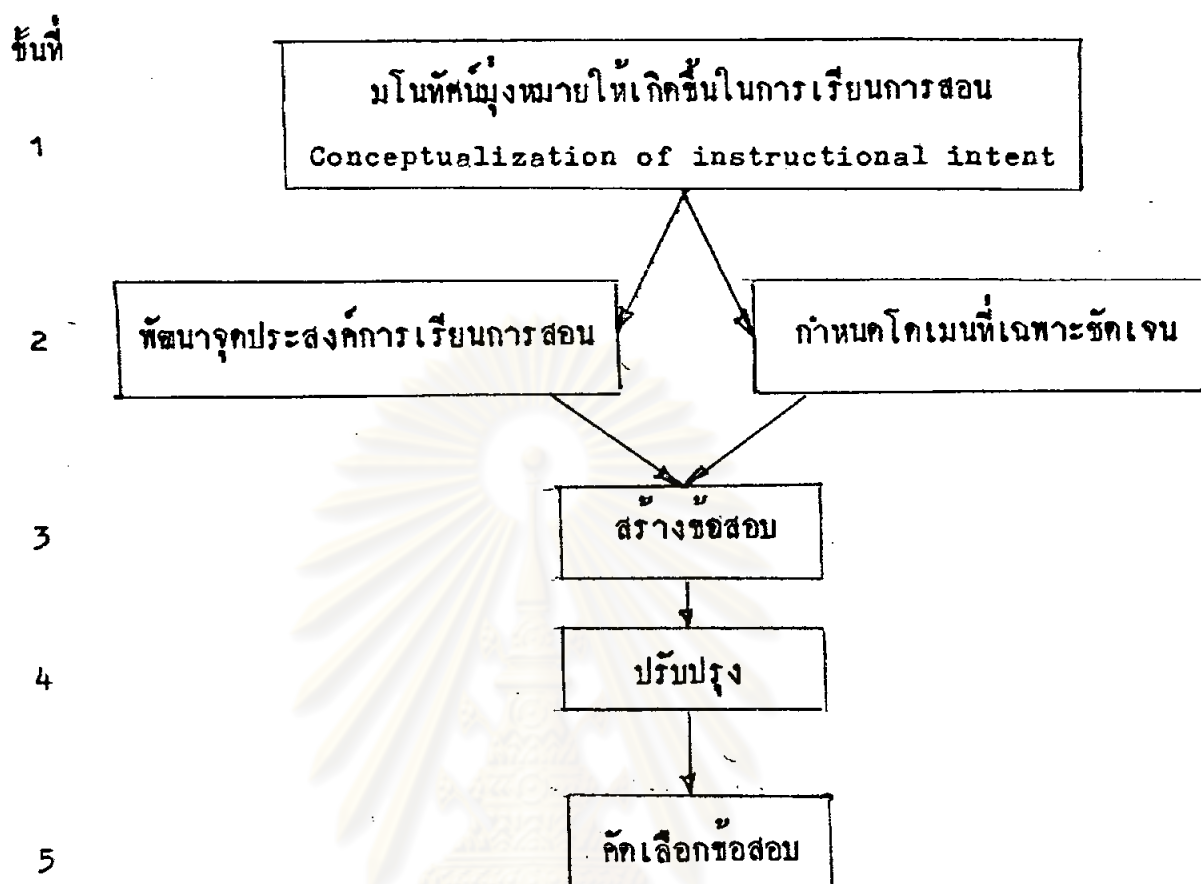
กรอนลันด์ (Gronlund) ให้ความสำคัญในการสร้างแบบสอบอิงเกณฑ์
ไว้พิจารณา 4 ประการคือ (เขาวที วิบูลย์ศรี 2526 : 31)

1. ตัดสินว่าการเรียนรู้ใดที่ควรจะมีผลในการเรียนรู้รอบ
2. จำกัดขอบข่ายของเนื้อหา และทักษะที่คองการทดสอบ
3. กำหนดชนิดผลการเรียนรู้ให้ชัดเจน
4. กำหนดมาตรฐานในการปฏิบัติ

สุภาพ วาณิช (2525 : 275-276) ได้เสนอหลักการสร้างข้อสอบ
อิงเกณฑ์ 4 ประการคือ

1. กลุ่มพฤติกรรมที่ระบุความแตกต่างของผลสัมฤทธิ์ จะต้องเขียนไว้อย่าง
ชัดเจน ก่อนสร้างแบบสอบ
2. แต่ละกลุ่มพฤติกรรม จะต้องกำหนดด้วย ชุดคำถาม หรืองานที่กำหนด
ให้ทำ ซึ่งพฤติกรรมหรือความแตกต่างแม้แต่เพียงเล็กน้อยของพฤติกรรมที่สำคัญก็จะต้อง
ระบุไว้
3. ต้องมีการวางแผนเลือก กลุ่มตัวแทน ของข้อคำถามในข้อ 2 มาใช้ใน
แบบสอบ
4. คะแนนที่ได้รับมาจะเป็นลักษณะการกระทำของกลุ่มพฤติกรรมดังกล่าว
รอยด์ และฮาธาไคนา (Roid and Haladyna 1980 : 293-295)
ได้เสนอแนวคิดในหลักการสร้างข้อสอบของแบบสอบผลสัมฤทธิ์ (The Role of
Items Writing in Achievement Test Development) มี 5 ขั้นตอน

จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 1 ขั้นตอนในการพัฒนาแบบสอบอิงเกณฑ์

- ขั้นที่ 1 มโนทัศน์อันเป็นหลักการ เหตุผล และแนวความคิดในการเรียนการสอน หรือ การทดสอบในขั้นแรกนี้ ผู้สอนต้องกำหนดพฤติกรรมที่ต้องการให้เกิดขึ้นในตัว ผู้เรียนหลังจากได้รับการเรียนการสอนแล้ว และสัมพันธ์กับโคเมนของเนื้อหา ซึ่งอาจจะหาได้โดยการวิเคราะห์เนื้อหาวิชาหรือวิเคราะห์งาน
- ขั้นที่ 2 เป็นขั้นที่ต้องกำหนดว่าสอนอะไร ซึ่งแบ่งออกดังนี้
- ก. กำหนดจุดประสงค์ในการเรียนการสอน
 - ข. กำหนดลักษณะเฉพาะของโคเมนเนื้อหา อันสอดคล้องกับมโนทัศน์ที่มุ่งหมายให้เกิดแก่ผู้เรียน

- ขั้นที่ 3 ขั้นการพัฒนาข้อสอบ เป็นขั้นหนึ่งของเทคนิคการเขียนข้อสอบ ซึ่งข้อสอบที่สร้างนั้น ควรให้เป็นไปตามโคเมนของข้อสอบ อันจะเป็นตัวแทนของการเรียนการสอนที่ต้องการ
- ขั้นที่ 4 ปรับปรุงข้อสอบ ในขั้นนี้หากมีข้อคำถามที่บกพร่อง ต้องปรับปรุงให้เรียบร้อยก่อนนำไปใช้ทดสอบต่อไป
- ขั้นที่ 5 คัดเลือกข้อสอบ ปัจจุบันนิยมคัดเลือกข้อสอบ โดยการสุ่มตัวอย่างข้อสอบจากโคเมนข้อสอบ (Roid and Haladyna 1980 citing Hambleton et. al. 1978 and Popham 1975)

เฟรเมอร์ (Fremer 1974) ได้เสนอขั้นตอนการสร้างแบบสอบอิงเกณฑ์

ดังนี้

- ขั้นที่ 1 เตรียมจุดประสงค์การเรียนการสอนให้สอดคล้องกับเนื้อหาการเรียนรู้อย่างเหมาะสม
- ขั้นที่ 2 สร้างข้อสอบ
- ขั้นที่ 3 วิเคราะห์ข้อสอบ
- ขั้นที่ 4 คัดเลือกข้อสอบ, กำหนดความยาวและคะแนนจุดตัด
- ขั้นที่ 5 วิเคราะห์ความเที่ยง และความตรงของแบบสอบ

บุญเชิด ภิญโญนนท์พงษ์ (2527) ได้ดัดแปลงแนวคิดการสร้างข้อสอบของ Roid และ Haladyna (1980) และ Fremer (1974) มาสรุปเป็นขั้นตอนการสร้างแบบทดสอบอิงเกณฑ์ได้ 7 ขั้นตอน ดังแสดงไว้ในแผนภาพที่ 2



ขั้นที่

1

วิเคราะห์เนื้อหาและจุดประสงค์

2

แปลงจุดประสงค์ให้เฉพาะเจาะจง

เขียนจุดประสงค์เชิงพฤติกรรม

กำหนดลักษณะมวลความรู้

3

เขียนหรือผลิตข้อสอบ

4

ตรวจทานข้อสอบ

5

วิเคราะห์ข้อสอบ

ความตรงเชิงเนื้อหา
Content Validity

ความตรงเชิงโครงสร้าง
Construct Validity

6

คัดเลือกข้อสอบ

กำหนดความยาว
และคะแนนจุดตัด

7

วิเคราะห์แบบทดสอบ
ความเที่ยง

แผนภาพที่ 2 แสดงขั้นตอนการสร้างแบบสอบอิงเกณฑ์

รายละเอียดของแต่ละขั้นตอนในการสร้างข้อสอบข้างต้นมีดังต่อไปนี้

1. การวิเคราะห์เนื้อหาและจุดประสงค์ เป็นการวิเคราะห์ที่ให้ทราบถึงพฤติกรรมที่ต้องการให้เกิดในหัวข้อเนื้อหาวิชานั้น ๆ อาจต้องใช้วิธีการวิเคราะห์กิจกรรมการเรียนรู้ (Task analysis) โดยวิเคราะห์พฤติกรรมตามลำดับขั้นของการเรียนรู้ ซึ่งจะช่วยให้ได้ทราบพฤติกรรมย่อยอันเป็นผลนำไปสู่พฤติกรรมที่ต้องการหรือจุดประสงค์หลักของเนื้อหาวิชานั้น ๆ

2. การแปลงจุดประสงค์หลักให้เฉพาะเจาะจง ซึ่งจุดประสงค์หลักที่วิเคราะห์ได้ยังไม่มียุติลักษณะเฉพาะเจาะจง จึงต้องแปลงจุดประสงค์หลักโดยแตกพฤติกรรมหลักเป็นพฤติกรรมย่อย ๆ ซึ่งในแต่ละพฤติกรรมหลัก อาจจะมีมากกว่า 1 พฤติกรรมย่อย ทั้งนี้ขึ้นอยู่กับลักษณะเนื้อหาวิชา พฤติกรรมย่อยเหล่านี้จะเป็นแนวทางในการเขียนข้อสอบ โดยเขียนพฤติกรรมย่อยให้เป็นจุดประสงค์เชิงพฤติกรรมหรือโดยการเขียนเป็นลักษณะเฉพาะของมวลความรู้

3. การเขียนข้อสอบหรือยลิตข้อสอบ โดยการเขียนจากกฎเกณฑ์ต่าง ๆ ที่ได้ระบุไว้ในจุดประสงค์เชิงพฤติกรรมซึ่งได้ระบุลักษณะเฉพาะที่ต้องการและแตกย่อยมาจากพฤติกรรมหลักนั้น ๆ

4. การตรวจทานข้อสอบ เป็นการสำรวจตรวจสอบข้อสอบเป็นรายข้อหลังจากเขียนข้อสอบเสร็จ เพื่อแก้ไขปรับปรุงข้อสอบให้เหมาะสมกับกลุ่มตัวอย่างนักเรียนที่จะใช้สอบ และให้มีความถูกต้องตามหลักวิชา ทั้งด้านเนื้อหาและพฤติกรรมที่ต้องการทดสอบ ภาษาที่ใช้เขียนข้อสอบมีความชัดเจน รัดกุม ถูกต้องและเหมาะสม

5. การวิเคราะห์ข้อสอบ โดยเน้นคุณภาพของข้อสอบในด้านความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ในการเรียนการสอน และความสามารถของข้อสอบในการจำแนกผู้รอบรู้และไม่รอบรู้ นั่นคือข้อสอบมีความตรงเชิงเนื้อหาและความตรงเชิงทฤษฎีตามลำดับ

6. การคัดเลือกข้อสอบ โดยคัดเลือกข้อสอบที่มีความตรงเชิงเนื้อหา 50%-100% และความตรงเชิงทฤษฎี 20%-100% (บุญเชิด ภิญโญอนันตพงษ์ 2527 : 43)

7. การวิเคราะห์แบบสอบถามทั้งฉบับ โดยการวิเคราะห์หลังจากได้ข้อสอบที่มีคุณภาพในค่าความตรงมารวมเป็นฉบับ และเน้นการวิเคราะห์ความเที่ยงของแบบสอบถาม

เทคนิคการเขียนข้อสอบอิงเกณฑ์

ไฮฟลี (Hively 1974) ได้เสนอเทคนิคการเขียนข้อสอบจากการกำหนดลักษณะเฉพาะของข้อสอบ ซึ่งจะบรรยายถึงรูปแบบและเหตุการณ์ที่ข้อสอบกำหนด การกำหนดลักษณะเฉพาะของข้อสอบเรียกว่า "ฟอร์มข้อสอบ" (Item Forms) (Roid and Haladyna 1980 : 298 citing Hively 1974) กลุ่มข้อสอบที่เป็นตัวแทนของประชากรข้อสอบซึ่งสร้างขึ้นจากฟอร์มข้อสอบนี้จะสามารถวัดโคเมนของเนื้อหาได้อย่างครอบคลุม

ออสเบิร์น (Osburn 1968) ได้ขยายส่วนประกอบของฟอร์มข้อสอบซึ่งประกอบด้วย 3 ส่วนดังนี้

1. โครงสร้างที่คงที่ เป็นเสมือนตัวฟอร์มของข้อสอบซึ่งใช้ผลิตข้อสอบ
2. ส่วนที่แปรผัน เป็นส่วนประกอบของโครงสร้างข้อความที่วางเว้นไว้ไม่น้อยกว่าหนึ่งแห่ง สำหรับที่จะนำกลุ่มของคำหรือข้อความที่ต้องการถามใส่ลงไปแทน
3. กำหนดคำหรือข้อความ เป็นส่วนที่ใช้เติมลงไปในส่วนที่แปรผันส่วนนี้จะ เป็นกลุ่มของคำ หรือข้อความที่กำหนดไว้อย่างชัดเจน

โพพแฮม (Popham 1975) แห่งศูนย์ IOX (Item-Objective Exchange) มหาวิทยาลัยแคลิฟอร์เนียซึ่งตั้งขึ้นในปี ค.ศ. 1968 เป็นศูนย์สำหรับศึกษาค้นคว้าด้านการประเมินผล ซึ่งมุ่งที่จะรวบรวมและพัฒนาเทคนิคการวัดที่เหมาะสม Popham พยายามแก้ปัญหาการเขียนจุดประสงค์ใหม่ที่มีความชัดเจนยิ่งขึ้น โดยนำจุดประสงค์เชิงพฤติกรรม ผสมกับ Item Forms ของ Hively เรียกว่าจุดประสงค์ขยายความ (Amplified Objectives) ซึ่งจะช่วยให้เห็นรูปร่างของข้อสอบ การเขียนข้อสอบจากจุดประสงค์ขยายความ จึงเป็นการเขียนข้อสอบจากมวลความรู้ (Domain) และสามารถนิยามประชากรได้ชัดเจน โดยมีส่วนประกอบ 5 ส่วนดังนี้ (Roid and Haladyna 1980)

1. คำบรรยายสิ่งที่ต้องการวัด ส่วนนี้ระบุโมโนทัศน์ (Concept) ของความรู้ที่ต้องการวัด ซึ่งเป็นเสมือนจุดประสงค์หลัก
2. จุดประสงค์ ส่วนนี้จะแปลงโมโนทัศน์ที่ต้องการให้เป็นจุดประสงค์เชิงพฤติกรรม หรือจุดประสงค์ที่เฉพาะเจาะจง
3. ส่วนขยายความจุดประสงค์ ส่วนที่จะอธิบายรายละเอียดของจุดประสงค์ โดยอธิบายถึงรูปร่างของข้อสอบ (Item Form) ทั้งภาคคำถาม และภาคคำตอบ
4. เกณฑ์การตอบถูก ส่วนนี้จะระบุถึงรายละเอียดและเงื่อนไขของการตอบถูก
5. ตัวอย่างข้อสอบ เป็นการเสริมการอธิบายการเขียนข้อสอบให้ชัดเจนยิ่งขึ้น

การสร้างข้อสอบอิงเกณฑ์ในการวิจัยครั้งนี้ ผู้วิจัยได้อาศัยเทคนิคการเขียนข้อสอบ จากการกำหนดลักษณะเฉพาะของข้อสอบตามวิธีของ สงบ ลักษณะ (2525) ซึ่งพัฒนาเทคโนโลยีการเขียนข้อสอบโดยอาศัยแนวคิดมาจากจุดประสงค์ขยายความ และการกำหนดลักษณะเฉพาะของมวลความรู้ของ Popham (1975) เรียกว่า "ลักษณะเฉพาะของข้อสอบ" ซึ่งเชื่อว่าการกำหนดลักษณะของข้อสอบจะช่วยจัดความยุ่งยากของการเขียนข้อสอบตามจุดประสงค์เชิงพฤติกรรมที่มีอยู่จำนวนมากต่อหนึ่งจุดประสงค์การสอน จุดประสงค์เหล่านี้ไม่สามารถนำมารวมกัน และตีความหมายได้อย่างชัดเจน เพราะว่ามันเกินทำข้อสอบในบางจุดประสงค์เชิงพฤติกรรมได้ แต่กลับทำข้อสอบผิดในบางจุดประสงค์เชิงพฤติกรรม นอกจากนั้นบางเนื้อหาวิชาสามารถเขียนจุดประสงค์เชิงพฤติกรรมได้จำนวนมากข้อ แต่บางเนื้อหาวิชาเขียนจุดประสงค์เชิงพฤติกรรมได้จำนวนน้อยข้อ หรือไม่ได้เลย จุดประสงค์เชิงพฤติกรรมย่อยเหล่านี้ จึงไม่สามารถบ่งชี้ถึงผลสัมฤทธิ์ในเนื้อหาวิชาในลักษณะรวมได้ จุดอ่อนนี้สามารถแก้ไขได้โดยใช้การกำหนดลักษณะเฉพาะของข้อสอบ อันเป็นการจัดระเบียบหรือสร้างกฎเกณฑ์สำหรับการเขียนข้อสอบให้รัดกุม รอบคอบ เค้นชัด สมบูรณ์ด้วยคุณภาพต่าง ๆ เช่น มีเหตุผล เป็นปรมันย์ วัดในสิ่งที่ต้องการวัดได้

ลักษณะเฉพาะของข้อสอบ จะเป็นแบบฟอร์มที่กำหนดขึ้น เพื่ออำนวยความสะดวกในการเขียนข้อสอบหรือผลิตข้อสอบ ซึ่งมีส่วนประกอบที่สำคัญ 5 ส่วน ดังนี้ (บุญเชิด ภิญโญอนันต์พงษ์, 2527 : 62-64 อ้างถึง สงบ ลักษณะ, 2525)

1. ทดพิกรรรมที่ตองการวัด ส่วนนี้จะระบุพิกรรรมหลักที่วิเคราะห์ได้จาก รายวิชาที่จะผลิตข้อสอบ
2. ทดพิกรรรมย่อย ส่วนนี้จะระบุพิกรรรมเฉพาะที่แตกมาจากพิกรรรมหลัก
3. คำอธิบาย ส่วนนี้จะเขียนขยายพิกรรรมย่อย โดยเขียนในรูปจุด- ประสงค์เชิงพิกรรรมที่ครอบคลุม สถานการณ์เงื่อนไขอย่างละเอียด และมีพิกรรรม คาคหวังที่สังเกตเห็นได้
4. ลักษณะคำถามและคำทอม ส่วนนี้จะกำหนดฟอร์มข้อสอบทั้งส่วนที่เป็น คำถามและคำทอม

ส่วนที่เป็นคำถาม จะกำหนดสถานการณ์ เงื่อนไข และคำสั่งที่จะให้ ผู้สอบทำ
5. ตัวอย่างข้อสอบ ส่วนนี้จะช่วยให้เข้าใจคำอธิบายชัดเจนยิ่งขึ้น

จำนวนข้อของแบบสอบอิงเกณฑ์

Klein and Kosecoff (1972) ได้ทำการสำรวจคำถามการวัดผลแบบ อิงเกณฑ์เกี่ยวกับการ กำหนดจำนวนข้อสอบที่จะทำให้การวัดและการตัดสินผู้รอบรู้และไม่ รอบรู้ให้ถูกต้องแม่นยำ หรือมีความคลาดเคลื่อนน้อยนั้นควรใช้ข้อสอบจำนวน 3-5 ข้อ ในแต่ละจุดประสงค์

Berk, (1980) ได้ศึกษาวิจัยถึงความคลาดเคลื่อนประเภทที่ I (Type I error) คือ ผู้รอบรู้แต่ถูกตัดสินให้เป็นผู้ไม่รอบรู้ (false non-mastery) และ ความคลาดเคลื่อนประเภทที่ II (Type II error) คือ ผู้ไม่รอบรู้ แต่ถูกตัดสินให้ เป็นผู้รอบรู้ (false mastery) โดยสรุปได้ว่า ถ้าใช้ข้อสอบจำนวน 15 ข้อ ในแต่ละ จุดประสงค์ กับการสอบที่ส่งผลย้อนกลับทันทีที่สอบเสร็จ จะทำให้การตัดสินถูกต้องแม่นยำ แต่ถ้าใช้ 2-3 ข้อ ในแต่ละจุดประสงค์จะเกิดความคลาดเคลื่อนในการตัดสินมาก และ เบอร์กได้เสนอแนะว่า การกำหนดจำนวนข้อสอบสำหรับการประเมินผลความก้าวหน้า (Formative Evaluation) หรือการประเมินผลเพื่อจัดตำแหน่ง (Placement) นั้น ควรใช้ข้อสอบ จำนวน 5-10 ข้อ แต่ถ้าเป็นการประเมินผลรวมสรุป

(Summative Evaluation) หรือการตัดสินใจแบบถาวร (Permanent Decision) จำเป็นต้องใช้ข้อสอบจำนวนมากขึ้นเป็นสองเท่า นั่นคือ จำนวน 10-20 ข้อ

การคัดเลือกข้อสอบ

Berk (1980 : 49) ได้กล่าวถึงการสร้างข้อสอบจากจุดประสงค์ที่เฉพาะเจาะจงนั้น ควรครอบคลุมจุดประสงค์ และข้อสอบที่ถูกคัดเลือกนั้นควรได้มาจากการสุ่มข้อสอบในโคเมนของข้อสอบ ซึ่งอาจสุ่มอย่างง่ายหรือสุ่มแบบแยกชั้น (Stratified random sampling) ข้อสอบเหล่านี้ก็จะเป็นตัวแทนที่ดีของโคเมน และมีคุณค่าต่อการแปลความหมายคะแนนซึ่งอ้างอิงคะแนนโคเมนนั้น

Millman (1974) Popham (1975) และ Hambleton and others (1978) ก็กล่าวไว้ในทำนองเดียวกันว่าข้อสอบในแบบสอบอิงเกณฑ์นั้นควรได้มาจากการสุ่มอย่างง่ายหรือการสุ่มแบบแยกชั้นจากโคเมนข้อสอบ ทั้งนี้เพื่อให้ข้อสอบเหล่านั้นเป็นตัวแทนของโคเมนและสามารถแปลความหมายอ้างอิงคะแนนโคเมน

Berk (1980) ได้กล่าวถึงการคัดเลือกข้อกระทงของแบบสอบอิงเกณฑ์ก่อนที่จะได้รับการสุ่มให้เป็นตัวแทนของโคเมนนั้น ข้อสอบควรมีลักษณะดังนี้

1. ความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ อันหมายถึง ความตรงเชิงเนื้อหาของข้อสอบ (Content Validity) ซึ่งเป็นคุณลักษณะที่สำคัญที่สุดของข้อสอบอิงเกณฑ์
2. ความยากง่ายของข้อสอบซึ่งพิจารณาก่อนการเรียนการสอน มีความยากมากกว่าหลังการเรียนการสอน Klien และ Kosecoff เรียกลักษณะเช่นนี้ว่าดัชนีความไวในการเรียนการสอน (Sensitive to Instruction)
3. คำนี้อ่านง่ายแจ่มแจ้ง ข้อสอบควรมีความตรงต่อการจำแนกผู้รอบรู้และไม่รอบรู้ หรือสามารถบ่งชี้ได้ว่าผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ ซึ่งคุณลักษณะภายในตัวผู้สอบ Hambleton เรียกลักษณะเช่นนี้ว่าเป็นความตรงเชิงทฤษฎี (Construct Validity) (Hambleton and others 1978)

การพิจารณาคุณภาพของข้อสอบอิงเกณฑ์ นักวัดผลการศึกษา เช่น Bloom (1971) กล่าวว่าความยากง่ายของข้อสอบอิงเกณฑ์ไม่ได้ถือเป็นเรื่องสำคัญ สิ่งสำคัญคือต้องเขียนข้อสอบให้สอดคล้องกับจุดประสงค์เชิงพฤติกรรม (สำเร็จบุญเรืองรัตน์ 2525 : 3) สมศักดิ์ อินทร์เวชญ์ (2521) และ สงบ ลักษณะ (2523 : 21-22) ได้กล่าวไว้ในทำนองเดียวกันว่า ความยากง่ายของข้อสอบอิงเกณฑ์นั้นขึ้นอยู่กับเนื้อหาวิชาและจุดมุ่งหมายเชิงพฤติกรรม ถ้าสิ่งที่ต้องการวัดง่าย ข้อสอบก็จะง่าย ถ้าสิ่งที่ต้องการวัดยากข้อสอบก็จะยากเช่นกัน

จากแนวความคิดของนักวัดผลการศึกษาที่กล่าวมานี้ ผู้วิจัยเห็นว่าความยากง่ายของข้อสอบไม่ใช่เป็นเรื่องที่ควรพิจารณา ข้อสอบจะยากหรือง่ายนั้นควรจะขึ้นอยู่กับลักษณะเนื้อหาวิชาหรือจุดประสงค์เชิงพฤติกรรมที่ต้องการจะวัด ดังนั้น การคัดเลือกข้อสอบอิงเกณฑ์ควรคำนึงถึงหลัก 3 ประการ โดยสรุปได้ดังนี้

1. ข้อสอบมีความตรงเชิงเนื้อหา เป็นความสอดคล้องระหว่างข้อสอบกับพฤติกรรมในจุดประสงค์ ซึ่งเป็นคุณลักษณะสำคัญที่สุดของแบบสอบอิงเกณฑ์
2. ข้อสอบมีความตรงเชิงทฤษฎี เป็นความสามารถของข้อสอบที่จะแสดงว่าผู้สอบคนใดเป็นผู้รอบรู้ (ตอบถูก) หรือไม่รอบรู้ (ตอบผิด) บางครั้งเรียกว่า ความตรงในการแปลความหมาย
3. ข้อสอบในแบบสอบอิงเกณฑ์ มีความเป็นตัวแทนของโคเมนหรือประชากรข้อสอบ ซึ่งกระทำได้โดยการสุ่มตัวอย่างข้อสอบด้วยวิธีสุ่มอย่างง่ายหรือสุ่มแบบแยกชั้นจากโคเมนข้อสอบ

ข้อสอบปรนัยแบบเลือกตอบกับตอบสั้น

แบบทดสอบทั้ง 2 ประเภทนี้ นักวิชาการหลายท่านให้ทัศนะต่าง ๆ กัน อีเบล (Ebel, R.L., 1972) กล่าวถึงแบบสอบทั้งสองประเภทนี้มีความสัมพันธ์กันสูงมาก เมื่อทั้ง 2 แบบ วัดในสิ่งเดียวกัน และเป็นแบบสอบคู่ขนาน (Parallel Form) คือ ถ้านักเรียนสามารถเขียนคำตอบได้ถูกต้องก็น่าที่จะจำแนก (Identifying) คำเลือกได้ถูกต้องเช่นกัน แต่อย่างไรก็ตามการจำแนกตัวเลือกน่าจะง่ายกว่าการสร้างคำตอบขึ้นมา เพราะโดยรูปแบบของแบบเลือกตอบมีการนำเสนอสิ่งเร้า และการตอบ-

สนองให้เสร็จ ย่อมง่ายกว่าการที่จะตอบถูกโดยรู้แต่สิ่งเร้าหรือคำถามเท่านั้น ซึ่งผู้ตอบต้องใช้ความสามารถในการระลึกได้ (Recall) จึงทำให้ยากที่จะเดาคำตอบที่ถูกต้อง ถ้าจะกล่าวถึงประโยชน์ของแบบสอบทั้งสองแล้ว (Gronlund, N.E., 1976) แบบสอบแบบเลือกตอบสามารถวัดความรู้อย่างมีประสิทธิภาพได้หลาย ๆ ชนิด และสามารถวัดผลการเรียนรู้ได้ละเอียดซับซ้อน ส่วนคำตอบสั้นนั้นสามารถตอบได้หลายแนวทาง แต่แบบเลือกตอบจะทำให้คำตอบของนักเรียนมีความเฉพาะเจาะจง

มิเชลล์ (Michell, W.J., 1950) มีความเห็นว่า แบบสอบเลือกตอบเป็นแบบสอบที่มีคุณค่ามากที่สุดไยบรรดาแบบสอบทั้งหลาย เพราะสามารถออกให้นักเรียนศึกษา วิเคราะห์ความ และนำสิ่งที่เรียนมาแล้วใช้ให้เป็นประโยชน์ สามารถวัดความเข้าใจ การตัดสินใจ การวินิจฉัยเหตุผล และวัดความจำของนักเรียนได้อย่างกว้างขวางกว่าที่ใช้แบบเติมคำ

สมหวัง พิธิยานุวัฒน์ มีความเห็นว่า ข้อสอบแบบตอบสั้น ๆ เป็นข้อสอบที่เหมาะสมสำหรับวัดความรู้เกี่ยวกับคำนิยาม และศัพท์เฉพาะอย่างมีประสิทธิภาพ และความสามารถทางหลักไวยากรณ์ ข้อสอบนี้เป็นประโยชน์ในวิชาคณิตศาสตร์และวิทยาศาสตร์ที่ต้องการคำตอบในรูปของการคำนวณ หรือให้เขียนเป็นสูตรหรือสมการ ซึ่งมุ่งวัดความรู้ในสิ่งที่ต้องการให้ผู้เรียนระลึกได้ (recall) มิใช่เห็นสิ่งนั้นแล้วจึงจำได้ (recognition)

สมหวัง พิธิยานุวัฒน์ และเพ็ญศิริ คำนชนะ (2524 : 44) ได้ศึกษาวิจัยเพื่อเปรียบเทียบคุณภาพของแบบสอบอิงกลุ่มประเภทเลือกตอบกับตอบสั้นนั้น ได้เสนอข้อสรุปที่น่าสนใจคือ ถ้าสร้างแบบสอบแบบเลือกตอบให้มีคุณภาพสูงแล้วแบบสอบเลือกตอบก็สามารถทำหน้าที่แทนแบบตอบสั้นได้ โดยเฉพาะกลุ่มตัวอย่างที่มีความรู้ความสามารถในระดับปานกลางหรือต่ำ แต่ถ้ากลุ่มตัวอย่างที่มีความรู้ความสามารถในระดับสูง ถ้าใช้แบบเลือกตอบต้องเป็นแบบสอบที่มีความแหลมคมจริง ๆ ผู้ที่ยังไม่เชี่ยวชาญในการสร้างแบบสอบแบบเลือกตอบควรใช้แบบตอบสั้น ซึ่งได้ผลดีทุกกรณี แต่มีข้อเสียในเรื่องการตรวจที่อาจต้องใช้เวลานานขึ้น แต่ถ้าเราทำให้คำตอบมีความเป็นปรนัยสูงแล้ว ก็อาจใช้เครื่องคอมพิวเตอร์ช่วยตรวจได้เช่นกัน นอกจากนี้ Stanley และ

Hopkins (1978) ได้เสนอว่า แบบทดสอบเป็นแบบสอบที่มีคุณค่า และเหมาะสม เป็นพิเศษสำหรับวิชาคณิตศาสตร์ และฟิสิกส์ เพราะคำถามของวิชาเหล่านี้อยู่ในรูปของปัญหาที่ต้องการคำตอบโดยการคำนวณ

การวิเคราะห์แบบสอบอิงเกณฑ์

การวิเคราะห์ความตรงของข้อสอบอิงเกณฑ์

1. ความตรงเชิงเนื้อหา (Content Validity)

ความตรงเชิงเนื้อหา หมายถึง ความสอดคล้องระหว่างพฤติกรรมที่ข้อสอบวัดได้ กับพฤติกรรมที่ระบุไว้ในจุดประสงค์ ซึ่งต้องอาศัยดุลพินิจของผู้เชี่ยวชาญเนื้อหาวิชา ในการตัดสินใจว่าข้อสอบแต่ละข้อมีความสัมพันธ์ สอดคล้องกับจุดประสงค์ที่ต้องการวัดหรือไม่ และใช้สูตรดัชนีความสอดคล้อง ระหว่างข้อสอบกับจุดประสงค์ (Item - Objective Congruence) เสนอโดย โรวินเนลลี และ แฮมเบิลตัน (Rovinelli and Hambleton, 1976) ซึ่งได้ปรับปรุงมาจากสูตร Hemphill - Westie Index (Matuzak, V.R. 1977 : 283-285 citing Rovinelli and Hambleton 1976)

การพิจารณาให้คะแนน

- | | | |
|-----|---------|--|
| + 1 | หมายถึง | แน่ใจว่าข้อสอบข้อนั้นวัดจุดประสงค์นั้น |
| 0 | หมายถึง | ไม่แน่ใจว่าข้อสอบข้อนั้นวัดจุดประสงค์นั้นหรือไม่ |
| - 1 | หมายถึง | แน่ใจว่าข้อสอบข้อนั้นไม่ได้วัดจุดประสงค์นั้น |

$$\text{สูตร } I_{io} = \frac{(M - 1) S_o - S_o}{2 N (M - 1)}$$

เมื่อ I_{io} คือ ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ Q

M คือ จำนวนจุดประสงค์ที่ต้องการวัด

N คือ จำนวนผู้เชี่ยวชาญเนื้อหาวิชา

S_o คือ ผลรวมคะแนนจากผู้เชี่ยวชาญที่ให้กับจุดประสงค์ 0

S'_o คือ ผลรวมคะแนนจากผู้เชี่ยวชาญที่ให้กับจุดประสงค์
ทั้งหมด ยกเว้นจุดประสงค์ 0

และอีกสูตรหนึ่งของ Rovinelli และ Hambleton ซึ่งเป็น
สูตรที่ขึ้นกับความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ (IOC) เช่นกัน แต่ใช้คำนวณ
เป็นรายจุดประสงค์ โดยมีสูตรคำนวณดังนี้ (บุญเชิด วิทยุอินันท์พงษ์ 2527 : 68-
69 อ้างถึง Hambleton and others 1978)

$$IOC = R/N$$

เมื่อ IOC คือ ทัศนคติความสอดคล้องระหว่างข้อสอบกับจุดประสงค์

R คือ ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญ

N คือ จำนวนผู้เชี่ยวชาญที่ใช้พิจารณาตัดสินทั้งหมด

ตามสูตรนี้จะพบว่า ค่า IOC ที่ถือว่าเป็นเกณฑ์ที่สอดคล้องกันนั้นควร
ไม่น้อยกว่า 0.5 แสดงว่าข้อสอบวัดได้ตรงกับจุดประสงค์นั้น

2. ความตรงเชิงทฤษฎี (Construct Validity)

ความตรงเชิงทฤษฎี หรือนักการศึกษาบางคนเรียกว่า ความ
ตรงเชิงโครงสร้างซึ่งอาจจะหาได้จากความสามารถของข้อสอบที่บ่งชี้ผู้สอบได้ว่าเป็น
ผู้รอบรู้ (Mastery) หรือไม่รอบรู้ (Non - mastery) โดยอาศัย วิธีทดลองเชิง-
ประจักษ์ กับกลุ่มตัวอย่างที่มีความรอบรู้ หรือได้รับการเรียนการสอนแล้ว สามารถ
ตอบข้อนั้นถูกต้อง และกลุ่มตัวอย่างที่ไม่รอบรู้หรือไม่ได้รับการเรียนการสอน จะตอบ
ข้อสอบข้อนั้นผิด ถ้าจะพิจารณาแบบส้อมซึ่งประกอบด้วยข้อสอบหลายข้อก็สามารถปฏิบัติ
ได้เช่นเดียวกัน นักเรียนที่ได้เรียนรู้ครบถ้วนตามจุดประสงค์แล้วจะเป็นผู้รอบรู้ และ
นักเรียนที่ยังไม่ได้เรียนจะเป็นผู้ไม่รอบรู้ ผลการสอบ และความสามารถของ
ข้อสอบที่จำแนกได้ว่า ผู้สอบเป็นผู้รอบรู้ (ตอบถูก) หรือไม่รอบรู้ (ตอบผิด) ลักษณะ
เช่นนี้ แสดงว่า แบบส้อมมี Construct Validity ด้วยเหตุนี้ นักการศึกษาแต่ละ
ท่านเรียก Construct Validity แตกต่างกันไป เช่น เรียกว่า ความตรงใน

การแปลความหมายคะแนน (Berk 1980) คำนี้อำนาจจำแนก หรือดัชนีความไว เป็นต้น ในการวิจัยครั้งนี้ผู้วิจัยใช้ Construct Validity ในความหมายของ คำนี้อำนาจจำแนกของข้อกระทง (Discrimination Index) คือ ความสามารถของ ข้อสอบในการจำแนกผู้สอบเป็นผู้รอบรู้ หรือไม่รอบรู้ ตามวิธีของ Crehan (1974 : 256) ซึ่งเขาได้ปรับการคำนวณมาจากสูตรหาความเที่ยงของแบบสอบ อิงเกณฑ์ของ Carver (1970) มาใช้ประมาณค่าความตรงของแบบสอบอิงเกณฑ์ Hambleton and others (1978) เรียกว่า การหาคำนี้อำนาจจำแนกของข้อสอบ การประมาณค่าความตรงตามสูตรนี้ อาศัยเทคนิคการทดลองเชิงประจักษ์จากผลการ สอบก่อนสอนและหลังสอน แล้วนำจำนวนนักเรียนที่มีคะแนนสอบผ่านหรือตก (รอบรู้ หรือไม่รอบรู้) มาแจกแจงในตาราง 2×2 ดังนี้

	ก่อนสอน	หลังสอน
ผ่าน	b	a
ตก	c	d

$$\text{ประมาณค่าความตรง} = \frac{a + c}{N}$$

$$\text{เมื่อ } N = a + b + c + d$$

ถ้านำมาวิเคราะห์ความสามารถในการจำแนกผู้เรียน ระหว่างก่อน-สอบและหลังสอบ โดยพิจารณาเป็นรายข้อ เมื่อ การตอบถูกหมายถึง รอบรู้หรือผ่าน และตอบผิด หมายถึง ไม่รอบรู้หรือตก นำจำนวนนักเรียนตอบถูก-ผิด มาแจกแจงใน ตาราง 2×2 ได้ดังนี้

	ก่อนสอน	หลังสอน
ถูก	b	a
ผิด	c	d

ประมาณค่าความตรง = $\frac{a + c}{N}$

จากสูตรที่เสนอโดย Crehen นี้เป็นการแสดงสัดส่วนการจำแนก ถูกกับจำนวนผู้สอบทั้งหมด การจำแนกถูกนั้น หมายถึง ก่อนการเรียนการสอนผู้สอบตอบข้อนั้นผิด และหลังการเรียนการสอนผู้สอบจะตอบข้อนั้นถูก นั่นคือ ถ้าข้อใดที่ผู้สอบตอบผิดมากในเหตุการณ์ก่อนการเรียนการสอน และหลังการเรียนการสอนแล้วตอบถูกมาก ข้อนั้นจะแสดงถึงความสามารถในการจำแนกผู้เรียน หรือมีความตรง ค่าความตรงที่ได้จากการคำนวณตามสูตรนี้ ไม่ควรน้อยกว่า .50 จากเหตุผลที่ค่าความตรง .50 นั้น แสดงว่า ผลการสอบก่อนสอนและหลังสอนมีผู้ตอบถูกจำนวนเท่ากัน ซึ่งเป็นความสามารถในการจำแนกค่า แต่ในลักษณะการวัดเชิง เกณฑ์ที่ดีถือว่า มีความตรง ไรก็ได้

นอกจากนี้ลักษณะความตรงเชิงทฤษฎียังสามารถคำนวณได้ด้วยแนวความคิดอื่น ๆ ดังต่อไปนี้

Kryspin and Feldhusen (1974) (อ้างถึงใน โกวิท ประวาลพฤษษ์ 2523 : 204) ได้เสนอดัชนีความไว (Sensitive Index) ซึ่งเป็นความแตกต่าง ความยากก่อนสอนกับหลังสอน เช่นกัน โดยมีสูตร

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

เมื่อ S คือ ดัชนีความไว

Rpos คือ จำนวนนักเรียนที่ตอบถูกหลังการเรียนการสอน

Rpre คือ จำนวนนักเรียนที่ตอบถูกต้องก่อนการเรียนการสอน

T คือ จำนวนนักเรียนทั้งหมด

Cox and Vagas ได้เสนอดัชนี Pre - to - Post Difference Index (PPDI) (Haladyna and Roid 1981 : 40 citing Cox and Vagas, 1966)

ซึ่ง PDDI = $\frac{\text{เปอร์เซ็นต์ความยากหลังสอน} - \text{เปอร์เซ็นต์ความยากก่อนสอน}}{\text{เปอร์เซ็นต์ความยากก่อนสอน}}$

หรือ PDDI = $\frac{\text{ตอบถูกต้องหลังสอน} - \text{ตอบถูกต้องก่อนสอน}}{\text{ตอบถูกต้องก่อนสอน}}$

Breanan and Stolurow (1971) ได้ประยุกต์ PPDI ของ Cox and Vagas เป็น Percent of Possible Gain (PPG) โดยมีสูตร ดังนี้

$$PPG = \frac{PPDI}{1 - \text{Pretest Difficult}}$$

เมื่อ PPG คือ เปอร์เซ็นต์ความบรรลุผล

PPDI คือ ความแตกต่างความยากหลังสอนกับก่อนสอน

ความเที่ยง(Reliability)ของแบบสอบอิงเกณฑ์

เบอร์ก (Berk, R.A. 1980 : 127) ได้สรุปมโนทัศน์ของความเที่ยงของแบบสอบอิงเกณฑ์ ไว้ 3 มโนทัศน์ คือ

1. ความเที่ยงเป็นความคงที่ในการตัดสินผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ โดยการสอบซ้ำ หรือสอบด้วยแบบสอบคู่ขนาน

2. ความเที่ยงเป็นความคงที่ของความแตกต่างกำลังสองของคะแนนแต่ละคนที่เบี่ยงเบนไปจากคะแนนจุดตัด โดยการทดสอบด้วยแบบสอบคู่ขนานหรือแบบสอบคู่ขนานอย่างสุ่ม

3. ความเที่ยงเป็นความคงที่ของคะแนนของนักเรียนแต่ละคน จากการทดสอบด้วยแบบสอบคู่ขนาน หรือแบบสอบคู่ขนานอย่างสุ่ม

จากข้อสรุปมโนทัศน์ความเที่ยงของ Berk นั้น จะเห็นว่า ตามแนว
มโนทัศน์ที่ 1 และ 2 เป็นลักษณะความเที่ยงของแบบอิงเกณฑ์อย่างชัดเจน ทั้งนี้เป็น
ความเกี่ยวข้องกับคะแนนจุดตัดของแบบสอบ ในมโนทัศน์ที่ 3 นั้น สามารถใช้หาความ-
เที่ยงได้ทั้งแบบสอบอิงเกณฑ์และแบบสอบอิงกลุ่ม ซึ่งไม่ได้เกี่ยวข้องกับเกณฑ์หรือคะแนน
จุดตัดของแบบสอบนั้นเลย ดังนั้นผู้วิจัยจึงขอเสนอการประมาณค่าความเที่ยงตามแนว
มโนทัศน์ที่ 1 และ 2 เท่านั้น

สำหรับการวิจัยครั้งนี้ ผู้วิจัยประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์
ประเภทเลือกตอบและตอบสั้น ตามแนวมโนทัศน์ที่ 2 ซึ่งเป็นความคงที่ของความแตก-
ต่างกำลังสองของคะแนนแต่ละคนที่เบี่ยงเบนไปจากคะแนนจุดตัด โดยที่โลเวตต์
(Lovett 1977 : 21-29 and Lovett 1978 : 239-251) ได้พัฒนาหลักสูตร
การคำนวณหาจากความเที่ยงของแบบอิงกลุ่ม ซึ่งเป็นสัดส่วนความแปรปรวนของคะแนน
ดิบ (คะแนนสังเกต) กับคะแนนจริงแต่ลักษณะการกำหนดหรือการวัดความแปรปรวน
ต่างกันออกไป โดยที่ ความแปรปรวน หมายถึง ความแปรปรวนของคะแนน ซึ่งคะแนน
นั้นมีการย้ายเบนจากมาตรฐานการปฏิบัติ (คะแนนจุดตัด) สัมประสิทธิ์ของความเที่ยง จึง
เป็นการวัดความแปรปรวนคะแนนของแต่ละบุคคลซึ่งเบี่ยงเบนไปจากคะแนนจุดตัด (C) ถ้า
C เท่ากับค่าเฉลี่ยแล้ว ก็จะเป็นการหาความเที่ยงแบบอิงกลุ่มนั่นเอง

การพัฒนาสูตร

ฟอร์มคะแนนของการวัดอิงเกณฑ์ คือ

$$x_{ij} = c + (\mu_{ij} - c) + (x_{ij} - \mu_{ij}) \quad (1)$$

เมื่อ x_{ij} คือ คะแนนดิบของคนที่ i ในการวัดครั้งที่ j

c คือ คะแนนเกณฑ์หรือคะแนนจุดตัด

μ_{ij} คือ คะแนนจริงของคนที่ i ในการวัดครั้งที่ j

$$\text{แล้ว } \mu_{ij} = \lim_{l \rightarrow \infty} \left[\left(\sum_{k=1}^l x_{ijk} \right) / l \right], \quad (k = 1, 2, 3, \dots, l)$$

เมื่อ l เป็นการสังเกตอย่างอิสระ กระทำจากคนที่ i ในการวัดครั้งที่ j



ความแปรปรวนคะแนนดิบ เป็น

$$D_{X_{1j}}^2 = (X_{1j} - c)^2 \quad (2)$$

เมื่อ $D_{X_{1j}}^2$ คือ ความแปรปรวนของคะแนนดิบของคนที่ 1 ในการวัดครั้งที่ j

ในทำนองเดียวกันความแปรปรวนคะแนนจริงคือ

$$D_{\mu_{1j}}^2 = (\mu_{1j} - c)^2 \quad (3)$$

เมื่อ $D_{\mu_{1j}}^2$ เป็นความแปรปรวนคะแนนจริงของคนที่ 1 ในการวัดครั้งที่ j และความแปรปรวนคะแนนความคลาดเคลื่อน คือ

$$\sigma_{e_{1j}}^2 = (X_{1j} - \mu_{1j})^2 \quad (4)$$

เมื่อ $\sigma_{e_{1j}}^2$ เป็นกำลังสองของคะแนนความคลาดเคลื่อนของคนที่ 1 ในการวัดครั้งที่ j

e คือ ความคลาดเคลื่อนของการวัด โดยมีข้อตกลงว่า เป็นอิสระจากกัน ความแปรปรวนการสุ่มเป็น เอกพันธ์ (e, σ_e^2)

จากสมการ 1, 2, 3 และ 4 จะได้ ความแปรปรวนคะแนนที่คาดหวัง

$$E(D_{X_{1j}}^2) = E(D_{\mu_{1j}}^2) + \sigma_{e_{1j}}^2 \quad (5)$$

จากสมการที่ (5) และหลักการที่สนาสูตรความเที่ยงมาจากอิงกลุ่มซึ่งเป็น สักส่วนความแปรปรวนคะแนนจริง กับคะแนน คำนวณ

$$R_{cc} = \frac{E(D_{\mu_{1j}}^2)}{E(D_{X_{1j}}^2)}$$

เมื่อ R_{cc} เป็นค่าความเที่ยงของการวัดอิงเกณฑ์ โดยวิธีการสอบเที่ยง ครั้งเดียว

โดยทั่วไปแล้วการวัดทางสมองไม่สามารถวัดได้เพียงครั้งเดียว ส่วนใหญ่
คะแนนแต่ละคนได้จากการรวมหรือเฉลี่ยมาจากจำนวนข้อสอบ Livingston (1972)
ได้พิสูจน์ว่า สูตรของ Spearman - Brown สามารถประยุกต์ใช้ในการวัดแบบ
อิงเกณฑ์ได้

ความเที่ยงของการวัดอิงเกณฑ์ จากการวัด k ครั้ง คือ

$$\begin{aligned}
 R_{ck} &= \frac{KR_{cc}}{1+(k-1)R_{cc}} \\
 &= \frac{K E(D_{\mu_i}^2)}{E(D_{x_i}^2)} \\
 &= \frac{(k-1)E(D_{\mu_i}^2)}{1 + \frac{E(D_{x_i}^2)}{E(D_{\mu_i}^2)}} \\
 &= \frac{K E(d_{\mu_i})}{K E(D_{\mu_i}^2) + \sigma_e^2} \quad (6)
 \end{aligned}$$

เมื่อ R_{ck} คือ ความเที่ยงแบบอิงเกณฑ์ ของผลรวมหรือค่าเฉลี่ยต่อการวัด
จำนวน k ครั้ง

ข้อตกลงในการวัด k ครั้ง นั้น

$$a) \mu_{i1} = \mu_{i2} = \mu_{i3} = \dots = \mu_{ik}$$

นั่นคือ ในการวัดทุกครั้งของแต่ละคนค่าคะแนนจริงเท่ากันหมด

$$b) \sigma_{e.1}^2 = \sigma_{e.2}^2 = \sigma_{e.3}^2 = \dots = \sigma_{e.k}^2 = \sigma_e^2$$

นั่นคือ ความแปรปรวนของความคลาดเคลื่อน เท่ากันหมด

$$\text{และ } c) c_1 = c_2 = c_3 = \dots = c_k$$

นั่นคือ คะแนนจุดตัด ทุกครั้งเท่ากันหมด

จากข้อตกลง a, b แทนค่า j ที่ dot (.) ในสมการที่ 6 แล้ว
คะแนนจริงที่คาดหวังกับความแปรปรวนคะแนนดิบจะคงที่ ในการวัดทุก ๆ k ครั้ง

การประยุกต์ ANOVA

ถ้าผู้สอบ n คนทำข้อสอบจำนวน k ข้อ เราจะได้คะแนนทั้งหมด
kxn จำนวน (รูปเมทริกซ์ kxn) สามารถวิเคราะห์โดยใช้การวิเคราะห์ความ-
แปรปรวนด้วยวิธีแบบกลุ่มสุ่มสมบูรณ์ (Randomized complete - block design)
ที่ปราศจากปฏิสัมพันธ์ต่อกัน

ถ้ารูปฟอร์มคะแนนดิบคือ

$$x_{ij} = c + (\bar{x}_{1.} - c) + (\bar{x}_{.j} - \bar{x}_{..}) + e_{ij} \quad (7)$$

เมื่อ $\bar{x}_{1.}$ คือ ค่าเฉลี่ยของคนที่ i จากการทำข้อสอบ k ข้อ

$\bar{x}_{.j}$ คือ ค่าเฉลี่ยของข้อที่ j จากการทำข้อสอบของผู้สอบ n คน

$\bar{x}_{..}$ คือ ค่าเฉลี่ยรวม (Grand Mean) ซึ่งเป็นค่าประมาณของค่า
เฉลี่ยประชากร

ผลบวกกำลังสองและชั้นความเป็นอิสระ (df)

ถ้าหักลบ c ออกจากสมการ (7) แล้วความแตกต่างกำลังสองของผลบวก
จากข้อสอบ k ข้อ และผู้สอบ n คน ดังนี้

$$\sum_{i=1}^n \sum_{j=1}^k (x_{ij} - c)^2 = k \sum_{i=1}^n (\bar{x}_{1.} - c)^2 + n \sum_{j=1}^k (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum_{i=1}^n \sum_{j=1}^k e_{ij}^2 \quad (8)$$

$$\text{จาก } SS_{Tot} = \sum_{i=1}^n \sum_{j=1}^k (\bar{x}_{ij} - c)^2 \quad (9)$$

เมื่อ SS_{Tot} คือ ผลบวกกำลังสองทั้งหมด

$$SS_p = k \sum_{i=1}^n (\bar{x}_{1.} - c)^2 \quad (10)$$

เมื่อ SS_p คือ ผลบวกกำลังสองระหว่างบุคคล

$$SS_I = n \sum_{j=1}^k (\bar{X}_{.j} - \bar{X}_{..})^2 \quad (11)$$

เมื่อ SS_I คือ ผลบวกกำลังสองระหว่างข้อสอบ

$$SS_e = \sum_{i=1}^n \sum_{j=1}^k e_{ij}^2 \quad (12)$$

เมื่อ SS_e คือ เศษเหลือหรือผลบวกกำลังสองของความคลาดเคลื่อน

SS_{Tot} มี df คือ $k \times n$ ซึ่งโดยปกติจะเป็น $k \times n - 1$

SS_p มี df คือ n ซึ่งโดยปกติจะเป็น $n - 1$

ระดับความเป็นอิสระของข้อสอบและความคลาดเคลื่อนเป็น $df_I = k - 1$

และ $df_e = (k-1)(n-1)$

กำลังสองเฉลี่ย (MS)

ในสมการที่ 10, 11, 12 หากด้วย df แล้วจะได้

$$MS_p = SS_p/n \quad (13)$$

$$MS_I = SS_I/(k - 1) \quad (14)$$

$$MS_e = SS_e / (k - 1)(n - 1) \quad (15)$$

ค่าคาดหวัง MS คือ

$$E(MS_p) = \sigma_e^2 + k E(D\mu_i) \quad (16)$$

$$E(MS_p) = \sigma_e^2 + n \sigma_{\mu_j}^2 \quad (17)$$

เมื่อ $\sigma_{\mu_j}^2$ เป็นความแปรปรวนค่าเฉลี่ยของข้อสอบที่กระจายออกจากค่าเฉลี่ยรวม (Grand Mean) จะเท่ากับ 0

ถ้า $\mu_{i1} = \mu_{i2} = \dots = \mu_{ik}$ สำหรับทุก ๆ ค่าของ i

$$E(MS_e) = \sigma_e^2 \quad (18)$$

ค่าความคาดหวังของ MS_e, MS_I พิจารณา โดยผู้เขียนหลายท่าน

(Dayton, 1970; Lindman, 1974; Winner, 1971)

$$\begin{aligned}
 \text{โดยที่ } E (MS_p) &= kE (\bar{X}_{i.} - c)^2 \\
 &= \lim_{n \rightarrow \infty} \frac{k \sum_{i=1}^n (\bar{X}_{i.} - c)^2}{n} \\
 &= k [E(D_{\mu_i}^2) + \sigma_e^2] \quad (19)
 \end{aligned}$$

แทนค่าสมการ (16), (18) ลงในสมการ (6) ได้

$$R_{ck} = \frac{E (MS_p) - E (MS_e)}{E (MS_p)} \quad (20)$$

ถ้าค่าประมาณพารามิเตอร์ในสมการ (20) ถูกต้องแล้ว การประมาณค่าความเที่ยงดังนี้

$$r_{ck} = \frac{MS_p - MS_e}{MS_p}$$

เมื่อ r_{ck} คือ ค่าประมาณค่าของ R_{ck} ซึ่งใช้การสอบเพียงครั้งเดียว

หรือถ้าให้ k เป็นจำนวนข้อสอบ

n เป็นจำนวนผู้สอบ

การประมาณค่าความเที่ยงจากสูตร ANOVA

$$r_{ck} = 1 - \frac{MS_e}{MS_p} \quad (22)$$

$$MS_p = \frac{k \sum_{i=1}^n (\bar{X}_{i/k} - c)^2}{n}$$

$$MS_e = \frac{\sum_{i=1}^n \sum_{j=1}^k (X_{ij} - c)^2 - n \sum_{j=1}^k (X_{j/n} - \bar{X}_{ij})^2 - k \sum_{i=1}^n (\bar{X}_{i/k} - c)^2}{(n-1)(k-1)}$$

c คือ คะแนนจุดตัด

\bar{X}_{ij} คือ ค่าเฉลี่ยรวม

MS_p คือ ผลบวกกำลังสองเฉลี่ยคะแนนของบุคคล

MS_e คือ ผลบวกกำลังสองเฉลี่ยคะแนนความคลาดเคลื่อน

สูตรในสมการ (22) จะเป็นค่าลำเอียงโลเวทท์ที่เสนอสูตรปรับแก้
ดังนี้ (บุญเชิด วิทยุโณนันต์พงษ์ 2527 : 201 อ้างถึง Lovett, 1978)

$$r_{ck} \text{ (ANOVA Corrected)} = 1 - \frac{n(k-1)}{n(k-1)-2} \cdot \frac{MS_e}{MS_p} \quad (23)$$

ในการคำนวณความเที่ยงของแบบสอบอิงเกณฑ์ อาจใช้สูตร (22) หรือ
(23)

นอกจากการคำนวณตามวิธีของโลเวทท์ แล้ว ยังมีนักการศึกษาอีกหลาย
ท่าน เสนอสูตรการคำนวณที่สอดคล้องกับมโนทัศน์ที่ 1 และ 2 เช่น

ลิฟวิงสตัน (Livingston 1970) ได้พัฒนา สูตรหาค่าความเที่ยงของแบบ
สอบอิงเกณฑ์ โดยมีแนวคิดดั้งเดิมมาจาก ความเที่ยงแบบคลาสสิก (Classical
Reliability) ซึ่ง

$$r_{xx} = \frac{s_t^2}{s_x^2} = \frac{s_t^2}{s_t^2 + s_e^2}$$

r_{xx} เป็นค่าความเที่ยงแบบคลาสสิก

s_t^2 เป็นความแปรปรวนคะแนนจริง

s_e^2 เป็นความแปรปรวนคะแนนคลาดเคลื่อน

s_x^2 เป็นความแปรปรวนคะแนนดิบ

แล้วพัฒนาเป็นสูตรหาค่าความเที่ยงของแบบสอบอิงเกณฑ์ คือ (Mehren and Lehman,
1973 : 121 - 122)

$$r_{cc} = \frac{r_{xx} s_x^2 + (\bar{X} - c)^2}{s_x^2 + (\bar{X} - c)^2}$$

เมื่อ r_{cc} คือ ความเที่ยงของแบบสอบอิงเกณฑ์

r_{xx} คือ ความเที่ยงแบบคลาสสิก

s_x^2 คือ ความแปรปรวนของคะแนน

c คือ คะแนนเกณฑ์

จะเห็นว่า ถ้าคะแนนเกณฑ์ (c) เท่ากับค่าเฉลี่ย (\bar{X}) แล้วสูตรของ Livingston ก็คือ สูตรความเที่ยงแบบคลาสสิกนั่นเอง

Livingston (1972) ได้พัฒนาสูตรให้สอดคล้องกับการจำแนกคะแนน โทเมนของผู้สอบจากการพิจารณาคะแนนจุดตัดของโทเมน π_0 โดยเสนอสูตร ประมาณค่าความเที่ยงดังนี้ (Hambleton and others 1978 : 16)

$$k^2 (\hat{\pi}, \pi) = \frac{\sigma^2(\pi) + (\bar{\pi} - \pi_0)^2}{\sigma^2(\hat{\pi}) + (\bar{\pi} - \pi_0)^2}$$

เมื่อ $\hat{\pi}$ คือ ค่าประมาณคะแนนโทเมน

π คือ คะแนนโทเมน

π_0 คือ คะแนนจุดตัด

$\bar{\pi}$ คือ ค่าเฉลี่ยคะแนนโทเมน

$\sigma^2(\pi)$ คือ ความแปรปรวนคะแนนโทเมนที่เบี่ยงเบนจากคะแนนจุดตัด

$\sigma^2(\hat{\pi})$ คือ ความแปรปรวนคะแนนโทเมนโดยประมาณที่เบี่ยงเบนจากคะแนนจุดตัด

Harris (1972) เสนอ การหาค่าสัมประสิทธิ์ความเที่ยงในกรณีจำแนก ผู้รอบรู้ 2 ระดับ ($k = 2$) โดยการวิเคราะห์ความแปรปรวน ซึ่งมีการให้คะแนน 1 หมายถึง รอบรู้ และ 0 หมายถึงไม่รอบรู้ โดยมีสูตรคำนวณดังนี้ (Subkoviak 1976 : 265 - 266)

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

สูตรนี้ใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way analysis) ค่าสัมประสิทธิ์ μ_c^2 จะมีค่าสูงสุด เมื่อ $c = \mu$

การประมาณค่าความเที่ยงตามแนวโน้ตที่ 1 เป็นความคงที่ในการตัดสินใจในการจำแนกนักเรียนเป็นผูรรอบรู้และไม่รอบรู้ โดยมีความเชื่อว่าแบบสอบจะมีความเที่ยงสูง ถ้าสอบซ้ำหรือสอบด้วยแบบสอบคู่ขนานแล้วจำแนกผูรรอบรู้และไม่รอบรู้ได้คงที่ถูกต้อง ผู้เสนอสูตรในแนวนี้ได้แก่

คาร์เวอร์ (Carver, 1970) โดยใช้กลุ่มตัวอย่างเกี่ยวตบแบบสอบตามคู่ขนาน 2 ฉบับ แล้วประมาณค่าความเที่ยงจากสัดส่วนของความคงที่ในการจำแนกผูรรอบรู้ และไม่รอบรู้ดังนี้

ฉบับ ก.

		ไม่รอบรู้	รอบรู้
ฉบับ ข.	รอบรู้	b	a
	ไม่รอบรู้	c	d

$$\text{ความเที่ยง} = \frac{a + c}{n}$$

$$\text{เมื่อ } n = a + b + c + d$$

สับโคเวียค (Subkoviak) ได้กล่าวถึงปัญหาสูตรของคาร์เวอร์ว่า ความคงที่ในการจำแนกนักเรียนเป็นรายบุคคลไม่มีความชัดเจน สูตรนี้ยังไม่เหมาะสมนัก (Berk, 1980 : 131)

Hambleton and Novick (1973) เสนอ สูตรความเที่ยงโดยการบริหารการสอบ 2 ครั้ง จากแบบสอบฉบับเดียว หรือ แบบสอบคู่ขนาน จำแนกผู้สอบเป็นผูรรอบรู้ m ระดับ แล้วคำนวณจากสูตร (Hambleton and others 1978 : 21)

$$P_o = \sum_{k=1}^m P_{kk}$$

เมื่อ P_0 คือ ทัศนคติความเที่ยง

P_{kk} คือ สัดส่วนผู้ตอบที่ถูกจำแนกเป็นผู้รอบรู้ระดับ k

Swaminathan-Hambleton-Algina (1974) เสนอ ทัศนคติความเที่ยง จากสัมประสิทธิ์ แคปป่า ($Kappa : k$) ของ Cohen (1960) โดยใช้การบริหาร การสอบ 2 ครั้ง โดยมีสูตรดังนี้ (Hambleton and others 1978; Subkoviak 1980)

$$k = (P - P_c) / (1 - P_c)$$

เมื่อ P คือ สัดส่วนความคงที่ของการตัดสินใจจากการทดสอบซ้ำ

P_c คือ สัดส่วนความคงที่ของการตัดสินใจที่คาดหวังโอกาสจะเกิดขึ้น

โดยที่
$$P_c = \sum_{k=1}^m (P_{k.}) (P_{.k})$$

เมื่อ $P_{k.}$ แทน สัดส่วนผู้เข้าสอบที่จัดให้เป็นผู้รอบรู้ระดับ k ในการสอบ ครั้งที่ 1

$P_{.k}$ แทน สัดส่วนผู้เข้าสอบที่จัดให้เป็นผู้รอบรู้ระดับ k ในการสอบ ครั้งที่ 2

m คือ ระดับของการรอบรู้

Subkoviak (1976 : 265-276) เสนอ การประมาณค่าความเที่ยง ของแบบสอบอิงเกณฑ์ จากการสอบครั้งเดียว โดยอาศัยหลักการหาสัมประสิทธิ์ความ สอดคล้องในการกำหนดความรู้ของนักเรียนแต่ละคนในกลุ่มที่สอบแบบสอบคู่ขนาน 2 ฉบับ ซึ่งมีคะแนน X และ X' โดยดัดแปลงสูตรให้สามารถคำนวณคะแนน X' ได้ จากคะแนน X จึงสามารถคำนวณค่าความเที่ยงจากการสอบเพียงครั้งเดียวโดยใช้สูตร

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

เมื่อ P_c แทน ค่าความน่าจะเป็นของการตัดสินสอดคล้องกันทั้ง 2 กลุ่ม

$\sum_{i=1}^n P_c^{(i)}$ แทน ผลรวมของความน่าจะเป็นของการตัดสินที่สอดคล้องกัน
สำหรับแต่ละบุคคล

การกำหนดเกณฑ์หรือคะแนนจุดตัดของแบบสอบอิงเกณฑ์

คำว่าเกณฑ์ (Criterion) ในการวัดผลอิงเกณฑ์ต่างจากคำว่าเกณฑ์ที่ใช้ในการวัดผลอิงกลุ่ม ซึ่งหมายถึง ความสามารถของกลุ่มที่เรียกว่ามาตรฐานของกลุ่ม (Normative Standard) แต่คำว่า "เกณฑ์" ในแบบสอบอิงเกณฑ์นั้น หมายถึง กลุ่มงานเฉพาะที่ผู้สอบจะต้องสามารถปฏิบัติได้ (Performance) ก่อนที่จะบรรลุระดับผลสัมฤทธิ์ที่กำหนดขึ้น ตามความหมายนี้ เกลเซอร์ (Glaser, 1963) กล่าวว่า การทดสอบด้านประสิทธิภาพ (Proficiency) นั้น คือ การทดสอบอิงเกณฑ์ ดังนั้น เกณฑ์ สำหรับแบบสอบอิงเกณฑ์โดยทั่วไปแล้ว ควรมีคุณสมบัติ 2 ประการ คือ ประการแรกต้องมีลักษณะที่จะบรรยายถึงวัตถุประสงค์ของการเรียนการสอนได้อย่างชัดเจน ประการที่สอง ต้องมีลักษณะที่จะบ่งชี้ถึงประสิทธิภาพของพฤติกรรมอันจะเป็นเกณฑ์ที่ยอมรับ นั่นคือ คุณลักษณะของการทดสอบอิงเกณฑ์นั้นจะต้องสามารถบ่งชี้ถึงระดับของการปฏิบัติที่เป็นมาตรฐาน (Performance Standard) สามารถนำไปสู่การตีความหมายของกลุ่มงานที่บุคคลทำได้อย่างอิสระ (เขาวที วิบูลย์ศรี 2526 : 27)

คะแนนหรือเกณฑ์ที่ใช้ในการเปรียบเทียบกับคะแนนที่ผู้สอบทำได้เพื่อจัดประเภทผู้สอบออกเป็นผู้รอบรู้และไม่รอบรู้นั้น มีนักการศึกษาเรียกต่างกัน เช่น กรอนลินค์ ใช้คำว่า มาตรฐาน (Standard) เมเจอร์ (Mager, 1962) และ มาร์ทูซา (Martuza, 1977) ใช้คำว่า มาตรฐานการปฏิบัติ (Performance Standard) เช่นเดียวกับ เกลเซอร์ (Glaser, 1963) อย่างไรก็ตามคำว่า มาตรฐาน, มาตรฐานการปฏิบัติ หรือเกณฑ์ที่ใช้ในการเปรียบเทียบนี้ อาจจะมีความหมายกว้างกว่าคะแนนจุดตัด (Cut-off Score) แต่นักการศึกษาหลายท่าน เช่น มาร์ทูซา (Martuza) แกลส (Glass, 1976) แฮมเบิลตันและคณะ (Hambleton and others, 1978) ต่างก็หมายถึง คะแนนจุดตัดนั่นเอง

การกำหนดคะแนนจุดตัดมีนักวัดผลการศึกษาหลาย ๆ ท่านให้แนวคิดและวิธีการแตกต่างกันไป ทั้งนี้ทางก็มีจุดประสงค์เพื่อให้การกำหนดคะแนนจุดตัด สามารถใช้เป็นเกณฑ์ตัดสินผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ได้ถูกต้องที่สุด มีความเป็นปรนัยและมาตรฐานมากที่สุด

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้อาศัยแนวคิดและวิธีการกำหนดคะแนนจุดตัด โดยใช้ทฤษฎีการตัดสินใจ (Decision Theoretic Approach) ของแกลส (Glass, G.V. 1978 : 251-253) ซึ่งต้องอาศัยวิธีการทดลองเชิงประจักษ์คะแนนจุดตัดที่เหมาะสมนั้น ควรทำให้เกิดการจำแนกนิคน้อยที่สุด นั่นคือ นักเรียนที่ไม่ได้รับการเรียนการสอนจะถูกตัดสินให้เป็นผู้รอบรู้ หรือสอบผ่านเกณฑ์มีจำนวนน้อยที่สุด และนักเรียนที่ได้รับการเรียนการสอนจะถูกตัดสินให้เป็นผู้ไม่รอบรู้ หรือสอบไม่ผ่านเกณฑ์มีจำนวนน้อยที่สุดเช่นกัน แกลสได้สร้างฟังก์ชัน ซึ่งเป็นความสัมพันธ์ของคะแนนจุดตัดกับเกณฑ์ภายนอก ซึ่งกำหนดไว้ก่อนเพื่อจำแนกผู้สอบเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ได้รับการเรียนการสอน กับกลุ่มที่ไม่ได้รับการเรียนการสอน สัดส่วนของบุคคลทั้งสองกลุ่มนี้แทนด้วย P_E และ $1 - P_E$ ตามลำดับ และจากการสมมติคะแนนจุดตัด (C_x) เพื่อแสดงการจำแนกผู้สอบเป็นผ่าน (รอบรู้) หรือไม่ผ่าน (ไม่รอบรู้) สัดส่วนของบุคคลที่จำแนกด้วย C_x นี้ แทนด้วย P_C และ $1 - P_C$ ตามลำดับ จากเกณฑ์ทั้งสองคือ เกณฑ์ภายนอก และคะแนนจุดตัดสมมติ (C_x) ทำให้เกิดการจำแนกร่วมกันเป็น 4 กลุ่ม ดังนี้

เกณฑ์ภายนอก

		เกณฑ์ภายนอก		
		ได้รับการเรียนการสอน	ไม่ได้รับการเรียนการสอน	
เกณฑ์จากแบบทดสอบ	ไม่ผ่าน	P_A	P_B	$1 - P_C$
	ผ่าน	P_C	P_D	P_C
		P_E	$1 - P_E$	1

- เมื่อ P_A แทน สัดส่วนจำนวนคนที่ไม่ผ่านเกณฑ์ของแบบทดสอบแต่ผ่านเกณฑ์ภายนอกเป็นการจำแนกผิดพลาด (false negative)
- P_D แทน สัดส่วนจำนวนคนที่ผ่านเกณฑ์ของแบบทดสอบ แต่ไม่ผ่านเกณฑ์ภายนอกเป็นการจำแนกผิดพลาด (false positive)
- P_B แทน สัดส่วนจำนวนคนที่ไม่ผ่านเกณฑ์ของแบบทดสอบและไม่ผ่านเกณฑ์ภายนอกเป็นการจำแนกถูก
- P_C แทน สัดส่วนจำนวนคนที่ผ่านเกณฑ์ของแบบทดสอบ และผ่านเกณฑ์ภายนอก เป็นการจำแนกถูก

จากสัดส่วนดังกล่าวมาสร้างเป็นฟังก์ชัน C_x ดังนี้

$$f(C_x) = (P_A + P_D) / (P_B + P_C) \dots\dots\dots(1)$$

ค่าฟังก์ชัน C_x จะแปรตามสัดส่วนที่จำแนกไว้ ถ้าสัดส่วนการจำแนกผิดพลาด ($P_A + P_D$) มีค่าต่ำสุด และสัดส่วนการจำแนกถูก ($P_B + P_C$) มีค่าสูงสุด และถ้า $f(C_x)$ มีค่าต่ำสุดแล้ว ค่า C_x นี้ จะเป็นคะแนนจุดตัดที่เหมาะสม

ฟังก์ชันในสมการ (1) มีข้อดลวงว่า การจำแนกผิดพลาดและการจำแนกผิดพลาดมีความสำคัญเท่ากัน ถ้าให้ความสำคัญของการจำแนกผิดพลาดทั้งสองไม่เท่ากัน ค่าฟังก์ชันจะเปลี่ยนแปลงไป ดังนั้นจึงต้องมีการปรับแก้ โดยกำหนดค่าถ่วงสองค่าขึ้นมา ให้ α เป็นค่าความสำคัญของการจำแนกผิดพลาด และ β เป็นค่าความสำคัญของการจำแนกผิดพลาด ดังนั้นจะได้ฟังก์ชันใหม่ดังนี้

$$f(C_x) = (\alpha P_A + \beta P_D) / (P_B + P_C) \dots\dots\dots(2)$$

ค่า α และ β นี้จะกำหนดขึ้นจากการใช้ดุลพินิจว่ามีความสำคัญมากน้อยเท่าใด ซึ่งมีพิสัยความสำคัญ ตั้งแต่ 0 ถึง 1 โดยใช้การตอบคำถามสองคำถามต่อไปนี้

"นักเรียนที่คะแนนผ่านเกณฑ์ของแบบสอบ แต่ถูกตัดสินให้ตก หรือเรียนไม่สำเร็จต้องออกจากโรงเรียน ควรให้ความสำคัญเท่าใด" เป็นการกำหนด α

"นักเรียนที่ไคคะแนนไม่ผ่านเกณฑ์ของแบบสอบ แต่ตัดสินให้ผ่านหรือเรียนจบหลักสูตร ควรให้ความสำคัญเท่าใด" เป็นการกำหนด β

สำหรับค่า α , β นี้ต้องอาศัยคุณลักษณะของผู้เชี่ยวชาญเนื้อหาวิชา ทักษะ ความสำคัญของการจำแนกผิด เพื่อให้ค่าฟังก์ชัน C_x มีความถูกต้องต่อการกำหนดคะแนนจุดตัด

นอกจากทฤษฎีการตัดสินใจแล้ว แกลสยังได้กล่าวถึงวิธีการกำหนดคะแนนจุดตัดตามแนวคิดและวิธีการของนักการศึกษาอื่น ๆ ดังนี้

1. การใช้ผลการสอบของคนอื่นเป็นเกณฑ์ (Performance of Others as a Criterion)

ระดับเกณฑ์แบบนี้สร้างขึ้นโดยอ้างอิงค่าพารามิเตอร์ของประชากรผู้เข้าสอบ ทั้งนี้เกณฑ์จึงอาจกำหนดด้วยคะแนนมัชฌิมฐาน (Median) ของคะแนนจากแบบสอบ หรืออาจกำหนดเป็นตำแหน่งเปอร์เซ็นต์ไทล์ วิธีนี้สอดคล้องกับแนวคิดของ Millman (1973) โดยอาศัยผู้เชี่ยวชาญเนื้อหาวิชา หรือครูประจำวิชาพิจารณาแบบทดสอบแล้วให้ระบุจำนวนเปอร์เซ็นต์ผู้สอบผ่านว่าควรเป็นเท่าไร แล้วจึงหาคะแนนที่สอดคล้องกับเปอร์เซ็นต์ที่ระบุนั้น หรืออาจกระทำได้โดยให้ครูผู้เชี่ยวชาญเนื้อหาวิชาเลือกตำแหน่งเปอร์เซ็นต์ที่เหมาะสมขึ้นมาก่อน แล้วนำไปเทียบกับคะแนนผลการสอบที่สอดคล้องกับตำแหน่งเปอร์เซ็นต์ไทล์

วิธีนี้เป็นที่น่าสังเกตว่าไม่สอดคล้องกับปรัชญาของการทดสอบอิงเกณฑ์ซึ่งยึดหลักการตัดสินว่า ผู้สอบมีความสามารถหรือไม่มีความสามารถที่จะบรรลุตามจุดประสงค์ที่กำหนด มากกว่าที่จะนำผลการสอบไปตัดสินเปรียบเทียบกับคะแนนผลการสอบของคนอื่น ๆ อันเป็นลักษณะอิงกลุ่มนั่นเอง

2. การนับลจจาก 100% (Counting Backwards from 100%)

วิธีการนี้มีหลักการเช่นเกี่ยวกับการกำหนดค่าเกณฑ์ หรือระดับมาตรฐานของจุดประสงค์เชิงพฤติกรรม คือค่าเกณฑ์ต้องกำหนดตามความสำคัญของจุดประสงค์ ถ้าจุดประสงค์ใดมีความสำคัญมาก ค่าเกณฑ์ที่ควรจะเป็น 100% ถ้าจุดประสงค์ใดมีความสำคัญน้อยลงมาก ค่าเกณฑ์ที่กำหนดก็ต้องลดลงมาจาก 100% ตามลำดับ แต่ตามหลักเหตุผลและประสบการณ์ ส่วนมากแล้วการกำหนดคะแนนจุดตัด หรือระดับเกณฑ์ที่สมบูรณ์ คือ 100% ย่อมเป็นไปได้ เพราะนักเรียนอาจมีความบกพร่องในด้านต่าง ๆ เช่น มีความสะเพร่าในการคิดหรือเขียนตอบ ไขข้อมูลผิดพลาดหรือขาดความตั้งใจ สิ่งเหล่านี้ย่อมทำให้ระดับเกณฑ์คลาดเคลื่อนไปจากอุดมคติที่คาดหวัง ดังนั้นการกำหนดคะแนนจุดตัด มักมีค่าต่ำกว่า 100% ประมาณ 5%-20% หรืออาจลดต่ำมากกว่านี้ การยินยอมเช่นนี้มีข้อเสียอยู่ที่ น่าจะเป็นการกำหนดตามความพอใจเสียมากกว่า

3. การปรับคะแนนตามเกณฑ์อื่น ๆ (Bootstrapping on other Criterion Score)

วิธีการนี้อาศัยเกณฑ์ภายนอกมาระบุเพิ่มขึ้นอีกว่า สำเร็จหรือรอบรู้ เช่นอาจจะระบุว่าได้รับใบขับขี่แล้ว หรือมีใบประกอบอาชีพเป็นต้น ผู้ที่อาศัยแนวคิดนี้ได้แก่

เบอร์ก (Berk, R.A., 1976: 4-9) ได้นำวิธีนี้มาประยุกต์ใช้โดยกำหนดเกณฑ์ภายนอกว่า ได้รับการสอนหรือไม่ได้รับการสอนตามวัตถุประสงค์ที่จะวัดแบ่งนักเรียนออกเป็น 2 กลุ่ม แล้วคำนวณหาคะแนนจุดตัดที่เหมาะสมที่สุด จากแบบสอบชุดเดียวกัน แล้วนำคะแนนที่ได้ไปเขียนกราฟพร้อมกัน จุดที่กราฟหรือฟังก์ชันทั้งสองตัดกันจะกำหนดให้เป็นคะแนนพยากรณ์ (Predicted Score) คะแนนนี้จะแบ่งนักเรียนออกเป็น 2 กลุ่ม คือกลุ่มที่ได้คะแนนเท่ากันหรือสูงกว่าคะแนนพยากรณ์ เรียกว่า พวกรอบรู้จากการพยากรณ์ (Predicted Master) แทนด้วย PM อีกพวกหนึ่ง คือกลุ่มที่ได้คะแนนต่ำกว่าคะแนนพยากรณ์ เรียกว่า พวกไม่รอบรู้จากการพยากรณ์ (Predicted Non-master) แทนด้วย PN ซึ่งเมื่อพิจารณาแจกแจงกับกลุ่มที่ได้รับการสอนและไม่ได้รับการสอนแล้วนำมาเขียนเป็นตาราง 2×2 ก็จะแบ่งนักเรียนออกเป็น 4 กลุ่ม ดังนี้

เกณฑ์ภายนอก

รอบรู้ คะแนนเกณฑ์พยากรณ์	ได้รับการศึกษา	ไม่ได้รับการศึกษา
	รอบรู้จริง TM	รอบรู้ไม่จริง FM
ไม่รอบรู้	ไม่รอบรู้ไม่จริง FN	ไม่รอบรู้จริง TN

N

M

โดยใช้สูตร $P(TM) = TM / (M + N)$

$$P(TN) = TN / (M + N)$$

$$P(FM) = FM / (M + N)$$

$$P(FN) = FN / (M + N)$$

เมื่อ M คือ จำนวนนักเรียนที่ไม่ได้รับการศึกษา

N คือ จำนวนนักเรียนที่ได้รับการศึกษา

ถ้าคะแนนพยากรณ์ที่ทำให้ $P(TM) + P(TN)$ มีค่าสูงสุดหรือ $P(FM) + P(FN)$ มีค่าต่ำสุดแล้ว คะแนนนั้นจะเป็นคะแนนจุดตัดที่เหมาะสมของแบบสอบ

วิธีนี้เป็นการศึกษาเกณฑ์ภายนอกแบบเพิ่มขึ้นอีกว่า สำเร็จหรือรอบรู้ เช่น อาจารย์ว่า ต้องมีใบขับขี่ หรือมีใบประกอบอาชีพเป็นต้น แต่การกำหนดคะแนนจุดตัดวิธีนี้อาจมีปัญหา 2 ประการคือ

1. ถ้าแบบสอบที่เราต้องการกำหนดจุดตัดมีค่าสัมพันธ์กับเกณฑ์ภายนอกต่ำ เราก็ไม่สามารถกำหนดจุดตัดให้สอดคล้องกับเกณฑ์ภายนอกได้อย่างสมบูรณ์ เช่น อาจจะมีผู้ผ่านเกณฑ์ภายนอกบางคนได้คะแนนสูงกว่าจุดตัด แต่อาจมีอีกหลายคนที่ได้คะแนนต่ำกว่าจุดตัด และเราก็ไม่สามารถกำหนดจุดตัดให้ผู้ที่ผ่านเกณฑ์ภายนอกทุกคนสามารถสอบผ่านแบบสอบที่เราต้องการกำหนดจุดตัดนั้นได้ด้วย

2. ถ้าจะกำหนดจุดตัดให้สอดคล้องกับเกณฑ์ภายนอกก็จะมีปัญหาว่า เกณฑ์ภายนอกที่ระบุขึ้นมา นั้น เลือกมาอย่างไร

4. พิจารณาตัดสินจากสมรรถภาพต่ำสุด (Judging Minimal)

วิธีนี้ได้จากการให้ผู้เชี่ยวชาญเนื้อหาวิชา หรือครูประจำวิชาศึกษาข้อสอบแล้วระบุว่า นักเรียนที่มีสมรรถภาพขั้นต่ำสุดที่สามารถยอมรับได้ ควรทำข้อสอบนั้นได้ วิธีนี้ผู้เชี่ยวชาญมาใช้คือ Nedelsky, Ebel และ Angoff

Nedelsky (1954) เสนอเทคนิคในการหาคะแนนผ่านต่ำสุดของแบบสอบปรนัย เรียกว่า "เทคนิคการหาคะแนนการเคาะหว่างนักเรียนเกรด F และ D" (F - D Guess Score Techniques) ซึ่งแบ่งนักเรียนออกเป็น 2 พวก คือ พวกที่อยู่เหนือจุดแบ่งเป็นพวกที่อยู่ระดับต่ำสุดที่ยอมรับให้ผ่าน (D-students) และพวกที่อยู่ต่ำกว่าจุดแบ่งคือพวกที่ไม่ผ่าน (F-student) ส่วนพวกที่อยู่ระหว่างคาบเส้น (Border - line) (F - D student)

วิธีการปฏิบัติโดยให้ครูพิจารณาว่า ในข้อสอบเลือกตอบแต่ละข้อมีตัวเลือกใดที่ D - student บอกว่าผิด แล้วเขียนส่วนกลับของจำนวนตัวเลือกที่เหลือ เช่น ข้อสอบ 5 ตัวเลือก และมี D - student ระบุได้ว่า ผิด 1 ตัวเลือก ดังนั้นส่วนกลับของข้อนี้คือ $\frac{1}{4}$ เป็นต้น นำเศษส่วนเหล่านี้รวมกันเป็นค่า MFD เพื่อคำนวณจากสูตร (Glass 1978 citing Nedelsky 1954)

$$D = M_{FD} + k\sigma_{FD}$$

เมื่อ D คือ คะแนนสอบผ่านขั้นต่ำ

M_{FD} คือ ค่าเฉลี่ยคะแนนการเคาะตอบถูก

σ_{FD} คือ ส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยการเคาะตอบถูก

k คือ ค่าคงที่ถ้าตัวเลือกผิดเกินครึ่งจะใช้ 0.5 ถ้าตัวเลือกแตกต่างกันเล็กน้อย จะใช้ค่า 1.0 ถ้าตัวเลือกใกล้เคียงกันมากใช้ค่า 1.5

Guilbert (1976) ได้คิดแปลงวิธีของ Nedelsky คำนวณหาระดับผ่านต่ำสุด (The Minimum Pass Level : MPL) ของแบบสอบชนิดเลือกตอบ โดย



ค่าความคาดหวังความสามารถยอมรับ (Acceptability Index : AI) ของแต่ละข้อจากการตัดสินของครูผู้สอน ซึ่งพิจารณาถึงความยากและทัศนคติที่คิดว่านักเรียนเห็นว่าผิดชัดเจนและไม่เลือก เช่นเดียวกับวิธีของ Nedelsky คะแนน MPL คือผลบวกของดัชนีความสามารถยอมรับของแต่ละข้อ

Ebel (1972) พิจารณาถึงจุดอ่อนของการนิยามคะแนนจุดตัด หรือคะแนนผ่านโดยใช้เปอร์เซ็นต์ของคะแนนรวมนั้นยังมีความไม่แน่นอน ขึ้นอยู่กับความบังเอิญมากกว่า ข้อสอบที่เขียนขึ้นอาจยากหรือง่ายเกินไป หรือมีอำนาจจำแนกต่ำเกินไป การที่จะสอบผ่านหรือไม่ จึงเป็นการตัดสินโดยคำถามในแบบสอบถามมากกว่าระดับความสามารถของผู้สอบก็ได้ เขาจึงเสนอวิธีการแก้ไขโดยการวิเคราะห์เนื้อหาของแบบสอบในแง่ที่เกี่ยวข้องกับความยากของแต่ละข้อ แล้วกำหนดเปอร์เซ็นต์ความคาดหวัง ซึ่งเป็นปริมาณที่คาดว่าผู้มีความสามารถต่ำสุดจะสอบผ่าน ดังตารางต่อไปนี้ (Glass 1978 citing Eble 1972)

ตารางที่ 2 แสดงเปอร์เซ็นต์ความคาดหวังสอบผ่าน ตามระดับความยากโดยแยกตามระดับความสำคัญ

ระดับความสำคัญ	ระดับความยาก		
	ง่าย	ปานกลาง	ยาก
ตรงกับปัญหาและสำคัญมาก	100%		
ตรงกับปัญหาและสำคัญ	90%	70%	
ยอมรับว่าตรงกับปัญหา	80%	60%	40%
ไม่แน่ใจว่าตรงกับปัญหา	70%	50%	30%

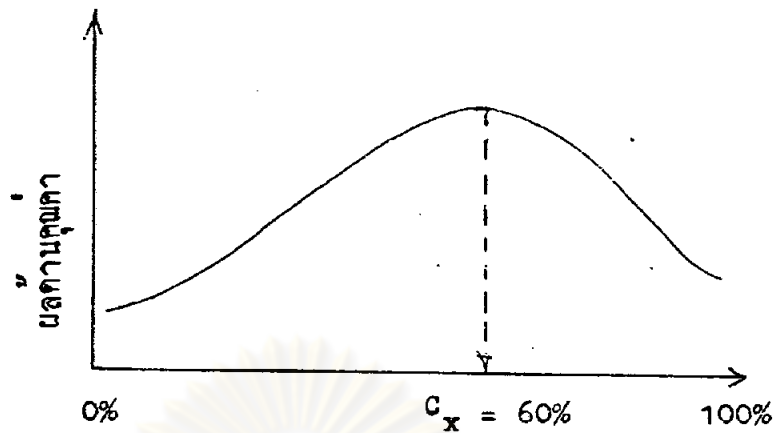
วิธีการของ Ebel พิจารณาจำนวนข้อสอบในแต่ละระดับความยาก ควบคู่กับเปอร์เซ็นต์ของความคาดหวัง และค่าเฉลี่ยของผลคูณที่ได้นี้ คือ คะแนนจุดตัดหรือคะแนนที่ผู้มีความสามารถต่ำสุดจะสอบผ่าน

Angoff (1971) เสนอวิธีที่คล้ายกับของ Ebel ซึ่งเป็นการหาความสัมพันธ์ของความยากแต่ละข้อ (Glass 1978 citing Angoff 1971) โดยใช้ความน่าจะเป็นของนักเรียนที่ยอมรับความสามารถต่ำสุดสามารถตอบข้อสอบได้อย่างถูกต้อง ซึ่งต้องอาศัยผู้เชี่ยวชาญเนื้อหาวิชา เสนอค่าความน่าจะเป็นของผู้อยู่ในระดับที่ยอมรับต่ำสุด สามารถตอบข้อสอบนั้นได้ถูกต้อง และประมาณค่าสัดส่วนผลรวมของความน่าจะเป็นหรือสัดส่วนเหล่านี้จะใช้เป็นคะแนนต่ำสุดที่ยอมรับความสามารถของผู้สอบได้ หรือคะแนนจุดตัด

สงขม ลักษณะ (2523 : 21) กล่าวว่าปัญหาของ Angoff คือ จะหลีกเลี่ยงอัตวิสัย (Subjective) ในการกำหนดค่าความน่าจะเป็น (Probability) ของแต่ละคนในแต่ละข้อได้อย่างไร

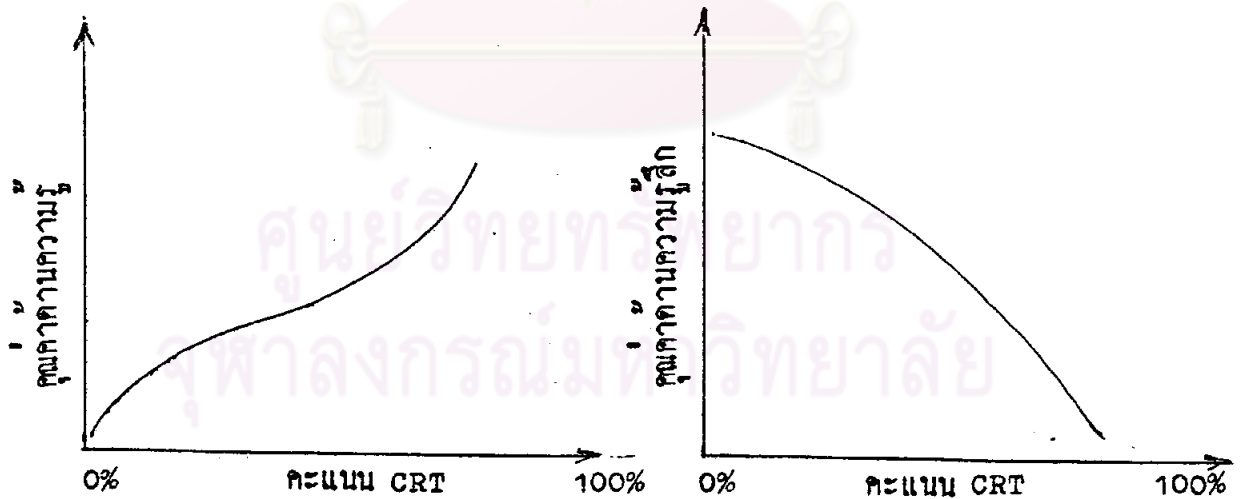
5. การใช้วิธีวิจัยเชิงปฏิบัติ (Operations Research Method)

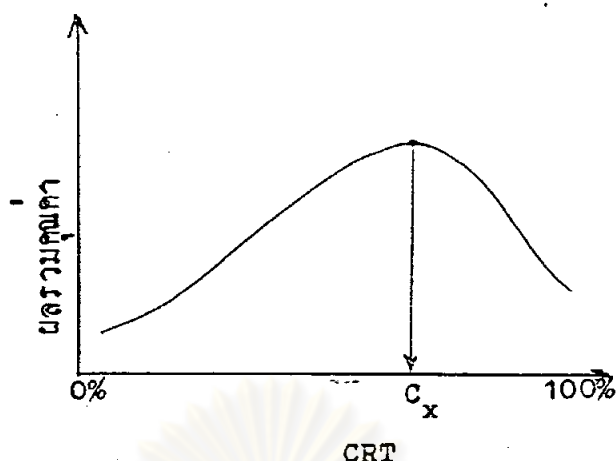
วิธีนี้อาศัยผลค่านับค่า อย่างใดอย่างหนึ่งมาช่วยในการพิจารณาเกณฑ์ โดยใช้ผลค่านับค่าสูงสุดมาเป็นเครื่องชี้กำหนดคะแนนเกณฑ์ ซึ่งพิจารณาจากลักษณะกราฟคะแนนที่ตรงกับจุดสูงสุดของกราฟจะเป็นคะแนนเกณฑ์ ซึ่ง บล็อก (Block) ได้ประยุกต์วิธีนี้มาใช้ดังนี้ (Glass, 1978 : 253 - 255 citing Block, 1972) บล็อกได้ทำการสอนกลุ่มต่าง ๆ ที่มีความเท่าเทียมกันจนมีผลสัมฤทธิ์จากการทำแบบสอบอิงเกณฑ์ระดับต่าง ๆ กันเช่น 10%, 15%, 20%...95%, 100% แล้ววัดผลค่านับค่า (Value) อย่างใดอย่างหนึ่งที่สัมพันธ์กับคะแนนจากแบบสอบอิงเกณฑ์ของแต่ละกลุ่มที่มีระดับผลสัมฤทธิ์ต่าง ๆ กันนั้น นำผลที่ได้มาเขียนกราฟแสดงความสัมพันธ์เพื่อพิจารณาค่าของคะแนนเกณฑ์ ถ้ากราฟไม่เป็นโมนोटอนิก (Non - monotonic) คะแนนเกณฑ์จะอยู่ระหว่าง 0%-100%



ผลปฏิบัติในแบบสอบอิงเกณฑ์

ถ้ากราฟที่ได้มีลักษณะเป็นโมโนโทนิก (Monotonic) คือ ไม่มีจุดที่กราฟจะโค้งกลับกะแนนเกณฑ์จะเป็น 100% ซึ่งเป็นไปไม่ได้ แต่ถ้าหากเป็นเช่นนี้ควรวัดผลค่านค่าอื่น ๆ ที่สัมพันธ์กับกะแนนจากแบบสอบอิงเกณฑ์มาประกอบการพิจารณา เช่น วัดความสนใจหรือทัศนคติต่อวิชานั้น ๆ เมื่อรวมผลค่านคุณค่า ค่านผลสัมฤทธิ์กับความรู้อีกเข้าด้วยกันก็จะเป็นกะแนนเกณฑ์





การพิจารณาผลรวม ควรต้องมีการกำหนดน้ำหนักของ ค่าการจำแนก
ผิดทางลบ (α) และการจำแนกผิดทางบวก (β) เช่นเดียวกับทฤษฎีการตัดสินใจ
ของแกส (Glass, 1978) แล้ว แทนค่าในสมการ

$$\text{ผลลัพธ์ผลรวม} = \alpha (\text{คุณค่าด้านความรู้}) + \beta (\text{คุณค่าด้านความรู้ลึก})$$

ในการพิจารณาความคลาดเคลื่อนในการตัดสินใจ (Decision error)

Emrick (1971) เสนอว่า ควรพิจารณาองค์ประกอบอื่น ๆ นอกเหนือจาก
ตัวแบบสอบ ในการสอบที่เกี่ยวข้องกับการใช้ผลการสอบเป็นเครื่องตัดสินใจ ควรพิจารณา
สิ่งที่เกี่ยวข้องกับความคลาดเคลื่อนในการตัดสินใจ 3 ประการ คือ (อ้างถึงใน สมภู
จันทร์อมพร 2523 : 22-23)

1. ทางด้านสถิติ (Statistical) เช่น ความเที่ยงของข้อสอบ (item reliability) ความยาวของแบบสอบ และอื่น ๆ ที่เกี่ยวข้องกับค่านี ถ้าเป็นการ
ตัดสินใจสอบของแต่ละคนที่มีความสำคัญยิ่ง การพิจารณาค่านีจะมีความสำคัญมาก
ถ้าการสอบใช้เพื่อวินิจฉัย การพิจารณาค่านีก็ลดความสำคัญลงมา
2. ความสำคัญของเนื้อหา (Centrality of content) การสอบ
เพื่อความรู้ (mastery testing) ควรเน้นวัตถุประสงค์ที่มีความสำคัญต่อกระบวนการ
ศึกษา วัตถุประสงค์ของการเรียนบางข้อมีความสำคัญค่านีเนื้อหา หากวัตถุประสงค์
เหล่านั้น ไม่ได้รับการวัดอย่างเพียงพอ และเที่ยงตรงแล้ว จะเกิดความเสียหาย
เนื่องจากความคลาดเคลื่อนขึ้นได้
3. คุณค่าทางจิตวิทยาที่เป็นผลมาจากความคลาดเคลื่อนในการตัดสินใจ
เหล่านี้เป็นสิ่งที่ต้องพิจารณา ไม่ว่าจะ เป็นกระบวนการประเมินผลใด ๆ ในการ

ประเมินผลเกี่ยวกับเด็ก คุณค่าด้านจิตวิทยามีความสำคัญมาก สำหรับผู้ใหญ่ ความสำคัญก็จะลดลง

การกำหนดคะแนนจุดตัดจากการใช้ทฤษฎีการตัดสินใจของเบส์ (Bayesian Decision Theoretic Procedure)

วิธีนี้ พิจารณาจักประเภทผู้สอบโดยใช้คะแนนโคเมน (\mathcal{A}) ของผู้สอบมาเปรียบเทียบกับคะแนนจุดตัด (\mathcal{A}_0) ที่กำหนดขึ้นมาใช้ โดยที่ผู้สอบมีคะแนนโคเมนมากกว่าหรือเท่ากับคะแนนจุดตัด จักเป็นผู้รอบรู้ (Master) และผู้สอบที่มีคะแนนโคเมนน้อยกว่าคะแนนจุดตัด จักเป็นผู้ไม่รอบรู้ (Nonmaster) (Swaminathan, Hambleton and Algina, 1975 : 88 - 93)

ในทางปฏิบัติเราไม่สามารถหาค่า \mathcal{A} ที่แท้จริงได้เนื่องจากไม่สามารถนำข้อสอบในโคเมนนั้น ๆ มาใช้ได้หมด จึงหาค่า \mathcal{A} ได้ด้วยวิธีประมาณมาจากคะแนนสอบซึ่งแปลงให้เป็นคะแนนมาตรฐาน แล้วพิจารณาคะแนนจุดตัด โดยอาศัยฟังก์ชันการสูญเสียที่คาดหวังของการตัดสินใจให้ผ่านและไม่ผ่านคะแนนระดับต่าง ๆ ความสูญเสียที่คาดหวังของการตัดสินใจให้ผ่านและไม่ผ่านที่น้อยที่สุดจะเป็นตัวกำหนดคะแนนจุดตัด

$$\text{ความสูญเสียที่คาดหวังของการตัดสินใจให้ผ่าน} = 1_{12} \text{ Prob} [Z < Z_{oi}/data]$$

$$\text{ความสูญเสียที่คาดหวังของการตัดสินใจให้ไม่ผ่าน} = 1_{21} \text{ Prob} [Z \geq Z_{oi}/data]$$

เมื่อ 1_{12} และ 1_{21} เป็นความสูญเสียที่เกิดขึ้นจากความคลาดเคลื่อนในการยอมรับที่ผิด และปฏิบัติที่ผิด ตามลำดับ

$\text{Prob} [Z \geq Z_{oi}/data]$ และ $\text{Prob} [Z < Z_{oi}/data]$ แทนความน่าจะเป็นที่จะสอบผ่านและไม่ผ่านตามลำดับ

ในการพิจารณาคะแนนจุดตัดของแบบสอบอิงเกณฑ์นั้น Millman (1973) ได้กล่าวว่าควรพิจารณาคำนี้ถึงสิ่งต่าง ๆ ต่อไปนี้

1. คะแนนสอบของผู้อื่น (Performance of others)

กำหนดโดยพิจารณา จำนวนนักเรียนที่สอบผ่านไว้ก่อน แล้วเลือกจำนวนเปอร์เซ็นต์ของผู้ผ่านไว้ หรืออาจใช้คะแนนดิบที่ตรงกับเปอร์เซ็นต์ไทล์ (Percentile) ที่เราเลือกหลังจากใช้แบบสอบกับกลุ่มผู้เข้าสอบที่มีความเข้าใจอย่างแจ่มแจ้งในเนื้อหาเป็นอย่างดี แต่วิธีนี้ขัดแย้งกับปรัชญาการใช้แบบสอบอิงเกณฑ์ที่ว่า ผู้สอบควรได้รับการตัดสินว่าผ่านหรือไม่ผ่าน โดยพิจารณาจากความสามารถที่เขาทำได้ตามเกณฑ์ที่วางไว้

2. เนื้อหาของข้อสอบ (Item Content)

โดยพิจารณาจากองค์ประกอบของเนื้อหาและความสำคัญในตัวข้อสอบว่านักเรียนจะตอบถูกหรือไม่ และให้ผู้เชี่ยวชาญตัดสินว่า ข้อสอบมีมากน้อยเท่าใดในแต่ละด้าน และให้เสนอสัดส่วน จำนวนข้อสอบที่คาดว่าผู้สอบมีความสามารถชั้นต่ำสุด ที่ควรทำได้เป็นเกณฑ์ผ่าน

3. ผลทางการศึกษาที่ตามมา (Educational Consequences)

จะทำให้เห็นความสัมพันธ์ของการทำแบบสอบ และการวัดโดยใช้เกณฑ์มากขึ้น เช่น การวัดวัตถุประสงค์เรื่องแรก ๆ หรือวัตถุประสงค์ที่เป็นพื้นฐานของวัตถุประสงค์ข้อต่อไป คะแนนจุดตัดควรจะสูง แต่ถ้าสิ่งที่วัดไม่มีความจำเป็นต่อการเรียนรายวิชาต่อไป คะแนนจุดตัดควรจะต่ำหรืออาจไม่มีเลยก็ได้ ฉะนั้นควรต้องพิจารณาถึงจุดมุ่งหมายในการสอนและผลที่ตามมาเป็นสำคัญด้วย เนื่องจากถ้าเราตั้งเกณฑ์ไว้ต่ำเกินไป นักเรียนที่ผ่านเพื่อจะไปเรียนบทเรียนต่อไป อาจเรียนโน้ตค้นและทักษะใหม่ได้อย่างขาดประสิทธิภาพ แต่ถ้าตั้งเกณฑ์ไว้สูงเกินไปก็จะทำให้เสียเวลาในการเรียนซ่อมเสริมโดยไม่จำเป็นต้องเสียทรัพยากรไปโดยเปล่าประโยชน์ สรุปก็คือควรพิจารณาให้ถี่ถ้วนว่าคะแนนเกณฑ์ควรจะเป็นเท่าใดจึงจะทำให้เกิดประโยชน์ทางการศึกษามากที่สุด

4. คุณค่าทางจิตวิทยาและทางด้านการเศรษฐกิจ (Psychological and Financial Costs)

ถ้าคุณค่าทางจิตวิทยาและทางด้านการเศรษฐกิจที่เกี่ยวข้องกับโปรแกรมการ-สอนซ่อมเสริมสูงมาก ก็ควรจะกำหนดคะแนนจุดตัดให้ต่ำลง แต่ในสภาพการณ์ที่ค่าใช้จ่ายในเรื่องนี้ต่ำ หรืออาจจะสูง แต่มีความคลาดเคลื่อนในการพิจารณาให้ผู้ที่ยังไม่ได้เรียนสอบผ่าน ซึ่งจะทำให้เกิดผลในทางลบเช่น ทำให้ประสิทธิภาพในการเรียนลดลง ก็ควรกำหนดคะแนนจุดตัดให้สูงขึ้น

5. ความคลาดเคลื่อนเนื่องมาจากการเดาและการสุ่มข้อสอบ (Error Due to Guessing and Item Sampling)

ถ้ารูปแบบของข้อสอบเป็นแบบชี้แนะหรือเปิดโอกาสให้นักเรียนตอบถูกได้ โดยการเดา เราอาจแก้ไขได้โดยการกำหนดคะแนนจุดตัดไว้สูง ๆ ทั้งนี้เชื่อว่าการกำหนดคะแนนจุดตัดไว้สูง การเดาจะไม่มีผลต่อผู้ที่ได้คะแนนใกล้เคียงกับคะแนนจุดตัด ส่วนความคลาดเคลื่อนในการสุ่มข้อสอบ เป็นความล่าช้าซึ่งเนื่องมาจากการไม่พิจารณาชนิดของคำถาม และเนื้อหาในโคเมน ทำให้การเลือกตัวอย่างไม่ครอบคลุมวัตถุประสงค์ที่ตั้งไว้ การตัดข้อความที่วัตถุประสงค์อื่นไป อาจเป็นเพราะโครงสร้างยากไม่สะดวกในการบริหารการสอบ หรือข้อสอบมีความคลุมเครือก็ควรจะได้มีการปรับคะแนนจุดตัดเพื่อแก้ไขจุดบกพร่องนี้

สงบ ลักษณะ (2523 : 21-22) ได้ให้ข้อเสนอแนะในการกำหนดคะแนนจุดตัดดังนี้

1. การหาคะแนนจุดตัดที่คั่น ต้องแน่ใจว่าข้อสอบทุกข้อมีคุณภาพวัตถุประสงค์ที่กำหนดไว้จริง ถ้าครูไม่ตั้งจุดประสงค์เชิงพฤติกรรม ไม่ได้สอบตามจุดประสงค์ และไม่ได้ออกข้อสอบตามจุดประสงค์นี้ คะแนนหรือจุดตัดก็ไม่มี ความหมาย

2. ข้อสอบ ๑ ฉบับ อาจวัดหลายจุดประสงค์ แต่ละจุดประสงค์อาจออกข้อสอบได้หลายข้อ การหาจุดตัดจุดเดียวของคะแนนรวมทั้งฉบับอาจไม่เหมาะสม แม้ว่าบังเอิญทุกจุดประสงค์มีความสำคัญเท่า ๆ กัน และแต่ละจุดประสงค์มีจำนวนข้อสอบ

เท่ากันก็ตาม เพราะหน้าที่ของครูต้องตรวจสอบทีละจุดประสงค์ วิธีที่คืออย่างหนึ่งคือ อาจต้องมีการจัดน้ำหนักของจุดประสงค์ตามความสำคัญของการนำไปใช้ แล้วหาคะแนน จุดตัดของแต่ละจุดประสงค์ จุดประสงค์ใดจำเป็นมากก็อาจตั้งไว้สูง จุดประสงค์ใดจำเป็นน้อยก็ตั้งเกณฑ์ไว้พอควร

3. ควรใช้กลุ่มผู้มีประสบการณ์ช่วยกันพิจารณากำหนดความสำคัญของ แต่ละจุดประสงค์ และกำหนดจุดตัดของแต่ละจุดประสงค์ ผลที่ได้คือแต่ละจุดประสงค์ มีจุดตัดต่างกัน วิชาต่างกันอาจมีจุดตัดไม่เหมือนกัน วิชาเดียวกันอาจมีจุดตัดต่างกัน สำหรับกลุ่มครูผู้สอนแตกต่างกัน ซึ่งขึ้นอยู่กับสภาพที่ตามมา (Consequence) ของ เกณฑ์ผู้สอน

4. มีการตรวจสอบปรับปรุงจุดตัดเป็นการทบทวนเกณฑ์ที่เคยตั้งไว้ผู้-เสมอ โดยนำผลต่อเนื่องที่ตามมาของการตัดสินผ่าน-ไม่ผ่าน มาเป็นข้อมูลของการปรับปรุงเกณฑ์ให้เหมาะสม

5. ปัญหาความยาก-ง่ายของข้อสอบ ซึ่งมีผู้กล่าวว่าผลทำให้เด็กสอบผ่านได้ "ง่าย" หรือ "ยาก" นั้น ควรพิจารณายึดจุดประสงค์เชิงพฤติกรรมเป็นเกณฑ์ ข้อสอบที่ออกตามจุดประสงค์ที่แตกต่างกันย่อมมีความยากง่ายไม่เท่ากันเป็นธรรมดา แต่สิ่งสำคัญคือ เราต้องการให้เด็กทำพฤติกรรมต่าง ๆ ตามจุดประสงค์เหล่านั้นให้ครบทุกจุดประสงค์ใช่หรือไม่ ถ้าตอบว่า "ใช่" ก็เป็นอันว่าหมดปัญหาเรื่องความยาก-ง่ายของข้อสอบ ก็ควรจะใช้เวลามาพิจารณาว่าการสอนได้ทั้งจุดประสงค์และสอบตามจุดประสงค์หรือไม่ และออกข้อสอบตามจุดประสงค์เพียงใดแล้วใช้เกณฑ์การตัดสินที่เหมาะสม

งานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคะแนนจุดตัดและความเที่ยงของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ระหว่างแบบสอบประเภทเลือกตอบกับตอบสั้นโดยการ ใช้ทฤษฎีการตัดสินใจของแกลส กำหนดคะแนนจุดตัดและใช้การวิเคราะห์ความแปรปรวน (ANOVA) ตามวิธีของโลเวทซ์ประมาณค่าความเที่ยงของแบบสอบทั้งสองประเภท ซึ่งมีงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

✓ สมหวัง พิธิยานุวัฒน์ และเพ็ญศิริ กำนธนะ (2524) ได้ศึกษาเปรียบเทียบคุณภาพของแบบสอบถามอิงกลุ่ม แบบเลือกตอบกับแบบตอบสั้น กับกลุ่มตัวอย่างที่เป็นนิสิตปริญญาโทที่มีพื้นฐานความรู้ทางคณิตศาสตร์จำนวน 30 คน และไม่มีพื้นฐานความรู้ทางคณิตศาสตร์ จำนวน 58 คน โดยใช้ข้อสอบฉบับละ 25 ข้อ ปรากฏว่า ความแปรปรวนของคะแนนจากแบบสอบทั้ง 2 ประเภท ไม่แตกต่างกันในกลุ่มที่มีพื้นฐานความรู้ทางคณิตศาสตร์ แต่ความแปรปรวนของคะแนนจากแบบตอบสั้นในกลุ่มตัวอย่างที่มีพื้นฐานความรู้ทางคณิตศาสตร์สูงกว่ากลุ่มตัวอย่างที่ไม่มีพื้นฐานความรู้ทางคณิตศาสตร์ แบบสอบแบบเลือกตอบง่ายกว่าแบบตอบสั้น ทั้งนี้ไม่ขึ้นอยู่กับกลุ่มตัวอย่าง ความเที่ยงของแบบสอบทั้งสองประเภทไม่แตกต่างกัน ในกลุ่มตัวอย่างที่ไม่มีพื้นฐานความรู้ทางคณิตศาสตร์ แต่กลุ่มตัวอย่างที่มีพื้นฐานความรู้ทางคณิตศาสตร์พบว่าความเที่ยงของแบบสอบแบบตอบสั้นสูงกว่าแบบเลือกตอบอย่างมีนัยสำคัญ และแบบสอบทั้งสองประเภทนี้สามารถจำแนกกลุ่มตัวอย่างได้ตามทฤษฎี

สำหรับงานวิจัยที่เกี่ยวข้องโดยตรงกับทฤษฎีของแกลส และโลเวคท์ นั้นไม่ปรากฏว่ามีใครนำไปวิจัย ดังนั้นผู้วิจัยจึงขอเสนองานวิจัยอื่น ๆ ที่เกี่ยวข้องกับการหาคะแนนจุดตัด และความเที่ยงของแบบสอบถามอิงเกณฑ์ทั่วไปดังต่อไปนี้

สมหวัง พิธิยานุวัฒน์ (อ้างถึงใน ประภา แก่นเพิ่ม, 2524) ใ้ค้นคว้าวิธีการของเบอร์คมาประยุกต์ใช้กับนิสิตปริญญาโทแผนกจิตวิทยาที่เรียนวิชาสถิติประยุกต์ จำนวน 18 คน เพื่อหาจุดแบ่งที่พอดี ของแบบสอบถามอิงเกณฑ์วิชาสถิติ เรื่อง สหสัมพันธ์ซึ่งมีความยาว 14 ข้อ โดยให้ทดสอบก่อนและหลังการเรียน แล้วนำคะแนนที่ได้มาเขียนกราฟของการกระจายของคะแนน แล้วพิจารณากำหนดคะแนนจุดตัดจากจุดที่เส้นกราฟทั้งสองตัดกันให้เป็นจุดตัดขั้นต่ำอย่างสูง และประมวลความคิดของ Bloom, Block, Glaser เกี่ยวกับการกำหนดระดับความรอบรู้ของผู้เรียนมากำหนดเป็นคะแนนจุดตัดขั้นสูงอย่างต่ำ จากคะแนนจุดตัดที่กำหนดขั้นทั้ง 2 นี้ ได้นำมาหาคะแนนจุดตัดที่พอดีปรากฏผลว่าได้คะแนน 7.5 เป็นจุดตัดขั้นต่ำอย่างสูงและ 9.8 เป็นจุดตัดขั้นสูงอย่างต่ำ

แอมเบิลตัน, สวามินาธาน และอัลกินา (1975) ได้ใช้วิธีการของเบส เพื่อหาเกณฑ์การจกประเภทผู้เข้าสอบ โดยพิจารณาจากความสูญเสียที่น้อยที่สุด (Threshold loss) จากการใช้แบบสอบที่มีความยาว 10 ข้อ นักเรียน 25 คน ทกลงกำหนดคะแนนจุดตัด $T_{10} = .80$ ความสูญเสีย เนื่องมาจากความคลาดเคลื่อนในการยอมรับที่ผิด (1_{12}) เป็น 1 หน่วย และความสูญเสียเนื่องมาจากความคลาดเคลื่อนในการปฏิเสธที่ผิด (1_{21}) เป็น 2 หน่วย พบว่าเมื่อแบ่งระดับการรอบรู้ 2 ระดับ ($k = 2$) ค่าความสูญเสียจะน้อยที่สุดเมื่อตัดสินให้ผู้สอบทำข้อสอบได้ 9 ข้อขึ้นไป เป็นผู้รอบรู้ และเมื่อแบ่งผู้สอบเป็นการรอบรู้ 3 ระดับ ($k = 3$) กำหนดจุดตัดเป็น .6 และ .8 พบว่าความสูญเสียจะน้อยที่สุด เมื่อตัดสินให้ผู้ที่ทำได้ 10 ข้อ เป็นผู้รอบรู้ ทำได้ 7-9 ข้อ เป็นผู้รอบรู้บางส่วนคือต้องพบทวนอีกเล็กน้อย ส่วนผู้ที่ทำได้ต่ำกว่า 7 ข้อ เป็นผู้ไม่รอบรู้ต้องเรียนใหม่

Poggio, Glasnapp และ Eros (1981) ได้ศึกษาเปรียบเทียบการกำหนดคะแนนจุดตัดหรือมาตรฐานการปฏิบัติ จากการใช้วิธีของ Angoff, Ebel และ Nedelsky โดยใช้กลุ่มตัวอย่างที่มีระดับการรอบรู้ตรงข้ามกัน (Contrasting Groups) ปรากฏว่า วิธีของ Nedelsky ก็ยังคงได้คะแนนเกณฑ์ต่ำที่สุดตามด้วย วิธีของ Angoff และ Ebel ตามลำดับ (Haledyna and Roid 1983 : 284)

กาญจนา วัชรสุนทร (2521) ได้สร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้น ม.1 เรื่องสมการจำนวน 40 ข้อ แยกเป็น 4 ฉบับ ๑ละ 10 ข้อ แล้ววิเคราะห์หาคุณภาพข้อสอบที่สร้างขึ้นโดยใช้ดัชนีเอสคำนวณหาค่าอำนาจจำแนก หากค่าความยาก โดยใช้สัดส่วนผู้ตอบถูก และกำหนดจุดตัดเป็น .6, .7 และ .8 หากค่าความเที่ยงโดยใช้สูตรของ Livingston หากค่าความตรงโดยใช้สูตร Carver ผลการศึกษาพบว่า เมื่อใช้จุดตัดเป็น .6 จะให้ค่าความเที่ยงและความตรงสูงสุด

ชมพู จันทรมรพร (2523) ได้ศึกษาหาจุดตัดของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ เรื่องสมการชั้น ม.1 โดยปรับปรุงข้อสอบของกาญจนา (2521) และหาจุดตัดโดยใช้ทฤษฎีการตัดสินใจของเบส หากค่าความเที่ยงและความตรงโดยใช้สูตร

Subkoviak และ Carver ตามลำดับ พบว่า จุดตัดแต่ละฉบับเป็น 6, 5, 5 และ 5 ตามลำดับ ความเที่ยงเป็น .76, .72, .65 และ .67 ส่วนความตรงเป็น .84, .80, .79 และ .78

ประภา แก่นเพิ่ม(2524) ได้วิจัยเกี่ยวกับความถูกต้องในการกำหนดจุดตัดด้วยวิธีการกำหนดเกณฑ์ระดับผ่านต่ำสุด (MPL) โดยเปรียบเทียบจุดตัดที่ได้จากผลการสอบแบบเลือกตอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ของนักเรียนที่อยู่ระดับความเส้นจำนวน 82 คน ปรากฏว่า จุดตัดที่ได้จากการกำหนดเกณฑ์ระดับผ่านต่ำสุด เมื่อคำนวณด้วยดัชนีความสามารถยอมรับโอกาสของการกระทำข้อสอบนั้นถูกตามวิธีของกิลเบิร์ต (MPL₁) มีค่าเท่ากับ 15.07 หรือเท่ากับ 37.68 เปอร์เซ็นต์ และถ้าคำนวณด้วยดัชนีความสามารถยอมรับตามวิธีของมหาวิทยาลัยอิลลินอยส์ (MPL₂) มีค่าเท่ากับ 12.61 หรือ 31.52 เปอร์เซ็นต์ และจุดตัดที่ได้จากการทดสอบของนักเรียนความเส้น ซึ่งถือเป็นเกณฑ์กับจุดตัดที่ได้จากการกำหนดเกณฑ์ระดับผ่านต่ำสุด มีความแตกต่างกันน้อยมาก

④ Knapp (1969) ได้ศึกษาเปรียบเทียบความเที่ยงของแบบทดสอบแบบเลือกตอบและแบบเติมคำ (open - end) ในวิชาคณิตศาสตร์โดยทดลองกับนักเรียนเกรด 5 จำนวน 83 คน ใช้ข้อสอบจำนวน 25 ข้อ คำนวณค่าความเที่ยงโดยไม่ใช้ความสัมพันธ์ (Correlationless Approach) เขานิยามความเที่ยงเป็นรายข้อภายใต้ความเชื่อที่ว่า แบบทดสอบในแต่ละข้อจะเป็น 1 เสมอ ถ้านักเรียนตอบถูก เขาใช้การบริหารการสอบ 2 ครั้ง ห่างกัน 1 สัปดาห์ ผลการวิจัยพบว่า บางข้อแบบเลือกตอบก็สูงกว่าข้อที่ต่ำกว่าแบบเติมคำตอบ เมื่อพิจารณาทั้ง 25 ข้อแล้วพบว่า แบบเลือกตอบมีความเที่ยง .79 แบบตอบสั้นมีความเที่ยง .78 เขาสรุปว่าเรื่องนี้ไม่ทราบว่า จะเกิดความล่าเอียงมากน้อยอย่างไรในการประมาณค่าความเที่ยง และสรุปว่า เรื่องความเที่ยงนี้เป็นเรื่องที่มีความซับซ้อนมาก