บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในการศึกษาผลของวิธีการหาคะแนนจุดตัด คะแนนโดเมน และความยาวของแบบสอบที่มีต่อความน่าจะเป็นในการจำแนกความรอบรู้และความเที่ยงในการตัดสินใจ ซึ่งใช้แนวคิด 2 ทฤษฎีมาศึกษา คือ ทฤษฎีคลาสสิค และทฤษฎีการตอบสนองข้อกระทง โดยเกี่ยวข้องกับแบบสอบอิงเกณฑ์ ผู้วิจัยจึงขอเสนอแนวคิดในหัวข้อ ต่อไปนี้

1. ลักษณะและความสำคัญของการวัดแบบสอบอิงเกณฑ์
2. มโนทัศน์ทฤษฎีการตอบสนองข้อกระทงกับแบบสอบอิงเกณฑ์
3. การเลือกข้อกระทงของแบบสอบอิงเกณฑ์
4. ความหมายและความสำคัญของคะแนนโดเมน
5. จุดอ่อนของค่าสถิติตามแนวคลาสสิค
6. การกำหนดคะแนนจุดตัดให้สอดคล้องกับคะแนนโดเมน
7. ความคลาดเคลื่อนในการตัดสินและความเที่ยงในการจำแนกความรอบรู้
8. งานวิจัยที่เกี่ยวข้อง

1. ลักษณะและความสำคัญของการวัดแบบอิงเกณฑ์

การวัดผลแบบอิงเกณฑ์เป็นระบบการวัดผลที่สืบเนื่องมาจากทฤษฎีการเรียนเพื่อรอบรู้ของบลูม (Bloom 1976) โดยใช้แบบสอบอิงเกณฑ์เป็นเครื่องมือในการวัด

แบบสอบอิงเกณฑ์ เป็นแบบสอบที่ใช้วัดเพื่อแยกผู้เรียนแล้ว หรือผู้รู้แล้วออกจากผู้ที่ยังไม่เรียนหรือไม่รู้ เพื่อสำรวจความก้าวหน้าและวินิจฉัยความสามารถทางการเรียนของนักเรียนแต่ละคนว่าบรรลุถึงเกณฑ์ที่วางไว้หรือไม่ และในการวัดผลอิงเกณฑ์เป็นการวัดที่ไม่

ได้นำผลของการวัดมาเปรียบเทียบกับคะแนนคนอื่น ๆ เหมือนกับการวัดผลแบบอิงกลุ่ม ดังนั้น การวัดผลแบบอิงเกณฑ์จึงทำให้เกิดการเรียนรู้แบบเกื้อกูลช่วยเหลือซึ่งกันและกัน (สมหวัง พิธิยานุวัฒน์ 2524 ข) ซึ่งก่อให้เกิดประโยชน์ในการเรียนการสอน จึงทำให้มีงานวิจัยที่เกี่ยวข้องกับการวัดผลแบบอิงเกณฑ์เป็นจำนวนมาก เป็นผลทำให้เกิดแนวคิดเกี่ยวกับการวัดผลแบบอิงเกณฑ์ในลักษณะต่าง ๆ กัน ซึ่งแนวคิดเช่นนี้มีมาแต่ปี ค.ศ.1951 (สำเริง บุญเรืองรัตน์, 2527: 15) แต่มีการเน้นการวัดผลอิงเกณฑ์ที่แตกต่างกันของแต่ละคน ทั้งนี้เมื่อนำแนวคิดเหล่านั้นมาวิเคราะห์สรุปผลที่ได้ก็จะเป็นในลักษณะเดียวกัน เช่น อีเบล (Ebel 1962) ได้มีแนวคิดว่า การทดสอบใด ๆ ก็ตามจะมีความหมายสมบูรณ์ก็ต่อเมื่อผลการสอบนั้นต้องบอกได้ว่า ผู้สอบมีความรอบรู้อยู่ในระดับใดของเนื้อหาทั้งหมดที่ทำการสอบวัด ปี ค.ศ. 1962 แกลสเซอร์และเคลาส์ (Glasser and Klaus) เป็นผู้เสนอแนวคิดในการวัดผลแบบอิงเกณฑ์มาใช้วัดผลสัมฤทธิ์ทางการเรียนเป็นครั้งแรกเช่นกัน โดยเขากล่าวว่า การวัดผลแบบอิงเกณฑ์นั้นขึ้นอยู่กับเกณฑ์มาตรฐานที่สมบูรณ์ที่จะบอกระดับคุณภาพของสิ่งที่วัดได้ (Glasser 1963: 519) ดังนั้นจะเห็นได้ว่า การวัดผลแบบอิงเกณฑ์ให้ความหมายชัดเจนได้ต้องใช้แบบสอบที่สามารถวัดได้ครอบคลุมชัดเจนในเนื้อหา แสดงว่าแบบสอบเป็นสิ่งสำคัญอย่างยิ่งที่จะทำให้ทราบว่าผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ จึงได้มีแนวคิดเกี่ยวกับแบบสอบอิงเกณฑ์ในลักษณะต่าง ๆ เช่น โพแฟม (Popham 1975: 5) ได้เสนอว่าแบบสอบอิงเกณฑ์เป็นแบบสอบที่ใช้หาสถานภาพของผู้สอบเป็นรายบุคคลซึ่งอ้างอิงถึงคะแนนโดเมน ทั้งนี้ต้องกำหนดพฤติกรรมในโดเมนที่ชัดเจน (Hambleton and Others 1978: 2) แกลสเซอร์ และนิตโก (Glasser and Nitko 1971: 626-627) กล่าวว่า แบบสอบอิงเกณฑ์ คือ แบบสอบที่สร้างขึ้นเพื่อสามารถวัดและแปลความหมายของการกระทำ โดยกำหนดโดเมนของงานที่นักเรียนควรจะสามารถทำได้ ซึ่งการวัดจะเป็นพฤติกรรมของแต่ละคน โดยเทียบกับโดเมนที่กำหนด ส่วนมิลแมน (Millman 1974: 327) ได้ให้ความหมายได้ชัดเจนดีว่า แบบสอบนั้นต้องการวัดอะไร ซึ่งแบบสอบประกอบไปด้วยกลุ่มตัวอย่างข้อกระทงจากโดเมนที่อาจจะได้จากการสุ่มอย่างง่าย หรือสุ่มแบบแบ่งชั้น (Simple Random Sampling or Stratified Random Sampling) และจากงานวิจัยของนิศิษฐ์ ตันฑวนิช (2528: 98) ได้สรุปแนวคิดของนักการศึกษาเหล่านี้ว่า จุดเน้นของแบบสอบที่สร้างจะอยู่ที่การกำหนดเกณฑ์

มาตรฐานของแบบสอบที่สร้างขึ้นจากขอบเขตของเนื้อหาที่กำหนด และเรียกแบบสอบที่พัฒนาขึ้นมาได้ว่า แบบสอบอิงเกณฑ์ ส่วนแบบสอบที่ต้องการมีการกำหนดขอบเขตของเนื้อหาให้ชัดเจนเสียก่อนแล้วเกณฑ์มาตรฐานเป็นผลที่เกิดขึ้นทีหลัง เมื่อมีการพัฒนาแบบสอบที่มีความเป็นตัวแทนของเนื้อหาขึ้นได้แล้ว เรียกแบบสอบที่พัฒนาขึ้นมาว่า แบบสอบอิงโดเมน ส่วนชูเมคเกอร์ (Shoemaker 1975: 129-131) ได้สรุปไว้ว่า ข้อสอบประจำโดเมน หมายถึง ข้อสอบในขอบเขตเนื้อหาทั้งหมดที่สามารถสร้างขึ้นมาได้ ทั้งนี้โปรแกรมการสอนจะต้องเป็นสื่อหรือวิธีที่จะทำให้ผู้เรียนสามารถตอบข้อคำถามทุกข้อในโดเมนที่กำหนดขึ้นได้ รอยด์ และฮาลาไดนา (Roid and Haladyna 1982: 230-231) ได้ให้ความเห็นว่า โดยทั่วไปแล้วมีความเป็นไปได้ที่ในเนื้อหาบางลักษณะอาจจะกำหนดขอบเขตที่เด่นชัดได้ยาก แต่แนวคิดมีอยู่ว่าถ้าผู้สร้างแบบสอบสามารถกำหนดขอบเขตของเนื้อหาให้เด่นชัดได้เท่าใด ก็จะยิ่งมีข้อดีในการที่จะกำหนดมาตรการในการวัดผลสัมฤทธิ์ทางการเรียนที่จะเกิดขึ้นได้อย่างชัดเจน ทั้งนี้การกำหนดขอบเขตของเนื้อหาให้ชัดนั้นเป็นประโยชน์ ดังนี้

1. ก่อให้เกิดความสอดคล้องอย่างชัดแจ้งระหว่างวัตถุประสงค์ วิธีการสอนและข้อสอบในแบบสอบ

2. วิธีการสร้างข้อสอบซึ่งมีการบันทึกไว้ สามารถบรรยายให้ทุกคนทราบได้และสามารถสร้างข้อสอบในรูปแบบเดียวกันออกมาได้

3. ควบคุมให้ผู้เขียนข้อสอบต่างคนกัน สามารถเขียนข้อสอบให้วัดได้ตามจุดประสงค์ที่ต้องการ

4. ก่อให้เกิดความชัดแจ้งในการศึกษาความเที่ยง โดยการตรวจสอบความเที่ยงสรุประหว่างแบบสอบกับโดเมนได้

ด้วยลักษณะดังกล่าวนี้ผู้วิจัยใคร่เสนอหัวข้อที่เกี่ยวกับประเด็นนี้ ดังนี้

1.1 การพัฒนาวิธีการสร้างแบบสอบอิงเกณฑ์

1.2 เทคนิคการเขียนข้อสอบอิงเกณฑ์

1.3 ความยาวของแบบสอบอิงเกณฑ์

1.1 การพัฒนาวิธีการสร้างแบบสอบอิงเกณฑ์

แบบสอบอิงเกณฑ์มีแนวความคิดที่ต้องการสรุปอ้างอิงไปยังเนื้อหาที่ต้องการวัดได้ชัดเจนและเหมาะสม ดังนั้นการสร้างแบบสอบอิงเกณฑ์จึงต้องมีการระบุเซต ขอบเขตประชากรของข้อกระทงให้ชัดเจนเพื่อให้แบบสอบที่สร้างขึ้นมีความเป็นปรนัยในการแปลความหมาย โดยใช้คะแนนที่สอบได้เป็นค่าประมาณของระดับผลสัมฤทธิ์ทางการเรียน ทั้งขอบข่ายของความรู้และทักษะหรือโดเมนที่กำหนด ซึ่งเป็นสิ่งสำคัญยิ่งต่อการสอนอย่างเป็นระบบ รอยด์ และฮาลาไดนา (Roid and Haladyna 1980: 293-314) จึงได้เสนอวิธีการสร้างแบบสอบที่เรียกว่า เทคโนโลยีของการเขียนข้อสอบ โดยมีแนวคิดที่สำคัญและที่นิยมใช้กัน คือ วิธีการกำหนดวัตถุประสงค์สำคัญ (Principal Objectives) ซึ่งเทคนิคนี้สามารถแบ่งในรายละเอียด ดังนี้

1.1.1 การกำหนดวัตถุประสงค์ภาคขยาย (Amplified Objectives) วิธีการนี้เป็นการพัฒนาจากวัตถุประสงค์เชิงพฤติกรรม ทั้งนี้ในตัวของวัตถุประสงค์ภาคขยายจะมีการแจกแจงรายละเอียดเพิ่มขึ้น โดยบรรยายให้ทราบถึงสถานการณ์ของการสอบ วิธีการสร้างตัวเลือกและเกณฑ์ในการให้คะแนนกับวัตถุประสงค์เฉพาะนั้น ๆ จากวัตถุประสงค์ที่เด่นชัดดังกล่าวจะทำให้สามารถสร้างข้อสอบในแนวเดียวกันออกได้เป็นจำนวนมาก ทำให้รูปแบบการเขียนข้อสอบมีความเป็นปรนัยสูง

1.1.2 การกำหนด โครงสร้างรูปแบบเฉพาะประจำข้อสอบ (Item Form) วิธีการนี้การสร้างข้อสอบจะเป็นการเขียนบรรยายคำ โครงสร้างหลักของคำถามที่จะใช้เป็นหลักหรือแกน โดยที่บางส่วนของประโยคในโครงสร้างหลักจะถูกกำหนดให้มีลักษณะเป็นตัวแปร ลักษณะการสร้างข้อสอบนั้น ทุก ๆ ครั้งที่มีการตั้งคำถาม โครงสร้างหลักหรือแกนของประโยคจะยังคงสภาพเดิม แต่ค่าของตัวแปรจะถูกเปลี่ยนแปลง โดยวิธีการนี้จะทำให้ได้ข้อสอบเป็นจำนวนมากที่วัดพฤติกรรมการเรียนรู้เดียวกัน

นอกจากนี้ รอยด์ และฮาลาไดนา ยังได้ให้หลักการสร้างข้อสอบของแบบสอบผลสัมฤทธิ์ มี 5 ขั้น ดังนี้

ขั้นที่ 1 มโนทัศน์ที่เป็นหลักการเหตุผล และแนวคิดในการเรียนการสอน ผู้สอนต้องกำหนดพฤติกรรมที่ต้องการให้เกิดขึ้นในตัวผู้เรียนหลังจากได้รับการเรียนการสอนแล้ว และสัมพันธ์กับโดเมนของเนื้อหา ซึ่งอาจจะหาได้โดยการวิเคราะห์เนื้อหาวิชาหรือวิเคราะห์งาน

ขั้นที่ 2 เป็นขั้นที่กำหนดเนื้อหาที่จะสอน แบ่งได้ ดังนี้

ก. กำหนดจุดประสงค์การเรียนการสอน

ข. กำหนดลักษณะเฉพาะของโดเมนเนื้อหา ให้สอดคล้องกับจุดมุ่งหมายที่ให้เกิดแก่ผู้เรียน

ขั้นที่ 3 ขั้นพัฒนาข้อกระทง ซึ่งเป็นขั้นตอนการสร้างข้อกระทงโดยใช้เทคนิคการเขียนข้อกระทง ซึ่งควรเขียนไปตามโดเมนข้อสอบเพื่อให้เป็นตัวแทนการเรียนการสอนตามเป้าหมาย

ขั้นที่ 4 ปรับปรุงข้อกระทง ถ้ามีข้อบกพร่องในข้อคำถาม ต้องปรับปรุงแก้ไขให้เรียบร้อยก่อนนำไปใช้สอบ

ขั้นที่ 5 คัดเลือกข้อกระทง มักใช้การสุ่มตัวอย่างข้อกระทงจากโดเมนข้อสอบ (Roid and Haladyna 1980: 293-295)

1.2 เทคนิคการเขียนข้อสอบอิงเกณฑ์

การเขียนข้อสอบต้องเขียนให้สอดคล้องกับเนื้อหาและพฤติกรรมที่ต้องการวัด ฉะนั้นจำเป็นต้องศึกษาว่าเนื้อหาที่ต้องการวัดนั้นมีลักษณะเป็นอย่างไร เพื่อนำมาแปลงให้พฤติกรรมที่จะวัดได้เฉพาะและชัดเจน ซึ่งจะทำให้สามารถเขียนข้อสอบได้ตรงตามต้องการ โดยอาจเขียนเป็นจุดประสงค์เชิงพฤติกรรมหรือกำหนดลักษณะเฉพาะของมวลความรู้หรือพฤติกรรมหรือเป็นการกำหนดประชากรข้อสอบหรือมวลข้อสอบ ในลักษณะเช่นนี้จึงทำให้เกิดเทคนิคการเขียนข้อสอบอิงเกณฑ์ขึ้น

ผลงานชิ้นแรกในการนำเทคนิคการเขียนข้อสอบมาใช้ คือ ไฮฟลี (Hively 1974: 5-10) และคณะ ได้เสนอวิธี รูปแบบข้อสอบ (Item Form) ซึ่งข้อสอบสามารถนิยามปฏิบัติการได้ ทำให้สามารถยืนยันผลการอ้างอิงไปยังจักรวาลของเนื้อหาได้ โดยการวิเคราะห์ด้วยคุณลักษณะแฝง (Latent Trait) ซึ่งเขากล่าวว่า รูปแบบข้อสอบ มีประโยชน์ ดังนี้ (Roid and Haladyna 1982: 115)

1. รูปแบบข้อสอบช่วยประหยัดเวลาในการสร้างข้อกระทง

2. รูปแบบข้อสอบที่ระบุโดเมนไว้ชัดเจนสามารถนำไปใช้กับแบบสอบอิงเกณฑ์ที่แปลความหมายได้ชัดเจน ไม่กำกวมและให้การประมาณที่ปราศจากอคติของสัดส่วนของโดเมนทั้งหมดที่นักเรียนรอบรู้

3. การสร้างข้อกระทงจากคอมพิวเตอร์เกี่ยวกับปัญหาทางการคำนวณสามารถสร้างอย่างถูกต้องแม่นยำ

4. สามารถให้คอมพิวเตอร์สร้างแบบสอบคู่ขนานอย่างมีประสิทธิภาพได้

1.3 ความยาวของแบบสอบอิงเกณฑ์

ไคลน์ และโคเซคอฟ (Klein and Kosecoff, 1972) ทำการสำรวจงานด้านวัดผลอิงเกณฑ์เกี่ยวกับการกำหนดจำนวนข้อสอบที่จะทำให้การวัดผลและการตัดสินผู้รอบรู้และไม่รอบรู้ให้ถูกต้องแม่นยำ หรือมีความคลาดเคลื่อนน้อยนั้นควรใช้ข้อสอบจำนวน 3-5 ข้อ ในแต่ละจุดประสงค์

โนวิค และเลวิส (Novick and Lewis in Harris 1974: 31) กล่าวว่าแบบสอบอิงเกณฑ์ที่มีความยาวมากที่สุดไม่ควรเกิน 20 ข้อ สำหรับแบบสอบวัดความสามารถ และเมื่อพิศิษฐ์ ตัณฑวนิช (2528) ได้ทดลองความยาวแบบสอบเป็น 25,28 และ 30 ข้อ กับวิชาการวิเคราะห์ข้อสอบของนักศึกษาหลักสูตรประกาศนียบัตรการศึกษาชั้นสูง ปีการศึกษา 2528 เทียบกับความยาวแบบสอบ 20 ข้อ ความคลาดเคลื่อนที่ปรากฏมีค่าคงที่เกิดขึ้นร้อยละ 9.15 , 8.64 และ 8.54 ตามลำดับ จะเห็นว่าความคลาดเคลื่อนดังกล่าวไม่แตกต่างไปจากช่วงความยาวแบบสอบ 20 ข้อ ซึ่งมีความคลาดเคลื่อนร้อยละ 9.76 จากข้อค้นพบนี้ พิศิษฐ์ ตัณฑวนิช (2528 : 114) สรุปว่าแบบสอบที่พัฒนาขึ้นมานั้นความยาวแบบสอบที่เหมาะสมที่สุดในการวัดประเภทบุคคลตามลำดับการรอบรู้ คือ แบบสอบที่ประกอบด้วยข้อสอบจำนวน 20 ข้อ ที่มีความ

เป็นตัวแทนของเนื้อหา และเป็นแบบสอบด้านคิดคำนวณ ซึ่งเป็นข้อค้นพบที่มีความสอดคล้องกับ โนวิค และเลวิส

2. มโนทัศน์ของทฤษฎีการตอบสนองข้อกระทงกับแบบสอบอิงเกณฑ์

การทดสอบเท่าที่ผ่านมายังคงใช้ทฤษฎีตามคลาสสิคซึ่งประสบปัญหาตลอดมาที่สำคัญ คือ

1. ค่าพารามิเตอร์ของข้อคำถามคือความยากและค่าอำนาจจำแนกของข้อคำถามแปรเปลี่ยนไปตามกลุ่มของผู้สอบที่แตกต่างกันด้านของความสามารถ (Ability) ทำให้เกิดปัญหาในการวิเคราะห์ข้อสอบฉบับเดิม แต่เปลี่ยนกลุ่มผู้สอบ

2. การเปรียบเทียบผู้สอบที่ความสามารถระดับหนึ่งจากแบบสอบที่ไปสอบวัดต้องใช้เซตข้อสอบที่เป็นเซตเดียวกันหรือคู่ขนานกันมาสอบ จึงจะเปรียบเทียบกันได้

3. มีความคลาดเคลื่อนในการตัดสินความรอบรู้ในหน่วยการเรียน จึงเป็นข้อจำกัดของการตัดสินใจ

4. เทคนิคการสอนที่ใช้กันอยู่ไม่ได้เป็นพื้นฐานในการกำหนดว่าผู้สอบแต่ละคนจะทำข้อสอบได้เพียงใด เมื่อกำหนดข้อคำถามให้ เว้นแต่ได้มีการใช้ข้อสอบนั้นกับกลุ่มผู้สอบที่คล้ายคลึงกันกับผู้สอบ เป็นต้น

5. สามารถประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดของผู้สอบแต่ละคนได้ทุกระดับความสามารถของผู้สอบ ซึ่งค่าความคลาดเคลื่อนนี้จะไม่ขึ้นอยู่กับความคลาดเคลื่อนมาตรฐานอื่น ๆ เพิ่มเติม จึงทำให้ผลการวัดความสามารถของผู้สอบแต่ละคนมีความถูกต้องแม่นยำมาก (บุญเชิด ภิญโญอนันตพงษ์, 2527: 92)

จากข้อจำกัดดังกล่าวนักวัดผลจึงได้ค้นหาทฤษฎีแนวใหม่ คือ ทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) หรือ Item Response Theory หรือ Item Characteristic curre Theory หรือ Modern Theory ซึ่งเป็นทฤษฎีการทดสอบที่อธิบายและทำนายคุณลักษณะ (trait) ที่สังเกตเห็นไม่ได้โดยตรงของผู้สอบอธิบายแต่คุณลักษณะของความสัมพันธ์ผลการสอบกับความสามารถในรูปของฟังก์ชั่น คือ

$P = f(\theta)$

P คือ ความน่าจะเป็นในการตอบของผู้สอบ

θ คือ ความสามารถของผู้สอบ

จากความสัมพันธ์ดังกล่าวสรุปได้ว่า ความสามารถในการสอบขึ้นกับความสามารถของผู้สอบ ซึ่งความสัมพันธ์ดังกล่าวมีข้อตกลง ดังนี้

1. ความสามารถเดียวสามารถใช้อธิบายคะแนนผลการสอบได้
2. การตอบข้อสอบหนึ่งถูกไม่ขึ้นอยู่กับการตอบข้ออื่น ๆ ถูกมาก่อน
3. ข้อสอบไม่ใช่ข้อสอบแบบความเร็ว การตอบข้อสอบใช้ความสามารถตอบได้จนสุดความสามารถไม่ขึ้นกับเวลา
4. มีความเหมาะสมที่จะใช้โค้งลักษณะข้อกระทง (Item Characteristic Curve) แสดงความสัมพันธ์ระหว่างโครงสร้างที่สังเกตเห็นได้ กับโครงสร้างที่สังเกตเห็นไม่ได้ของแบบจำลอง

สำหรับโมเดลทฤษฎีคุณลักษณะแฝงหรือทฤษฎีการตอบสนองข้อกระทง ได้รับการพัฒนาโดย Lawley ในปี 1943 โดยใช้โมเดลทางคณิตศาสตร์เป็นครั้งแรก คือ Normal Ogive Model ซึ่งต่อมาได้มีการพัฒนาโมเดลเพื่อความเหมาะสมโดยลอร์ด (Lord 1952) แต่เนื่องจากโมเดลดังกล่าวมีความยุ่งยากในการประมาณค่าพารามิเตอร์ เบิร์นบอม (Birnbaum 1968) จึงได้พัฒนาโมเดลใหม่ คือ โมเดลโลจิสติค (Logistic Model) ซึ่งให้ความสะดวกในการประมาณค่าพารามิเตอร์ เพราะโมเดลให้โค้งที่สามารถปรับให้มีความใกล้เคียงโมเดลโค้งปกติ (Normal Ogive Model) ได้ง่าย ดังนั้น การวิเคราะห์แบบสอบจึงใช้โมเดลของทฤษฎีการตอบสนองข้อกระทง (IRT)

ลักษณะของฟังก์ชั่นโค้งปกติสะสม (Normal Ogive Function) ที่มีพารามิเตอร์ 3 ตัว แสดงความสัมพันธ์ได้ ดังนี้

$$P(\theta) = c + (1 - c) \int_{-\alpha}^{a(\theta - b)} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$$

(Lord 1980: 13)

a คือ ค่าอำนาจจำแนก

b คือ ค่าความยาก

c คือ ค่าการเดา

θ คือ ระดับความสามารถของผู้สอบ

t คือ คะแนนมาตรฐานที่ได้จากการแปลงคะแนนที่สังเกตหรือวัดได้จากความสัมพันธ์ที่แสดงได้ ดังนี้

$$t = (y'_g - u'_{g,\theta}) / \sigma'_{g.\theta}$$

y'_g คือ คะแนนที่วัดได้

$u'_{g,\theta}$ คือ ค่าเฉลี่ยของคะแนนที่ความสามารถ θ

$\sigma'_{g.\theta}$ คือ ส่วนเบี่ยงเบนมาตรฐานที่ความสามารถ θ

(Birnbaum 1968: 370)

ส่วนลักษณะฟังก์ชั่นโลจิสติค (Logistic Function) ที่มีพารามิเตอร์ 3 ตัว แสดงความสัมพันธ์ ดังนี้ (Lord 1980: 12)

$$P(\theta) = c + \frac{1 - c}{1 + e^{-1.7a(\theta - b)}}$$

1.7 คือ Scaling Factor ซึ่งเป็นค่าเพื่อปรับให้ Logistic Function มีค่าใกล้เคียงกับค่าจาก Normal Ogive Function มากที่สุด อาจใช้สัญลักษณ์ D ก็ได้

ลักษณะทั้ง 2 ฟังก์ชั่นยังไม่มีผู้ใดได้แสดงให้เห็นว่าฟังก์ชั่นไหนเหมาะสมกับข้อมูลจากแบบสอบทางสมองมากกว่ากันอย่างชัดเจน โดยหลักการผู้สอบที่มีความสามารถสูงควรจะตอบข้อสอบข้อง่าย ๆ ถูกหมด แต่ในทางปฏิบัติบางครั้งผู้สอบบางคนอาจทำข้อสอบประเภทดังกล่าวผิดด้วยความสะเพร่า ไม่รอบคอบ การตอบผิดในลักษณะดังกล่าวนี้มีผลต่อฟังก์ชั่นโลจิสติคน้อยกว่า Normal Ogive เพราะฟังก์ชั่นโลจิสติคมีค่าเข้าใกล้แอสซิมโทท (Asymtotes) ช้ากว่าฟังก์ชั่น Normal Ogive (Lord 1980: 14) ประกอบกับ Normal Ogive มีความยุ่งยากในการคิดคำนวณมากกว่า จึงได้มีการนำฟังก์ชั่นของโลจิสติคมาใช้กันมาก และได้มีการพัฒนาโมเดลนี้เป็นโมเดลโลจิสติคที่มีพารามิเตอร์ 2 ตัว (Two-parameter Logistic Model) โดยมีเงื่อนไขว่า c = 0 ข้อสอบไม่มีการเดาได้ถูกเป็นอันขาด ดังนั้นความสัมพันธ์ของโมเดลนี้จึงแสดงได้ ดังนี้

$$P(\theta) = \frac{1}{1 + e^{-1.7a(\theta - b)}}$$

และได้มีการพัฒนาให้โมเดลมีรูปแบบของฟังก์ชั่นที่ง่ายขึ้นเพื่อสะดวกในการคิดคำนวณเป็นโมเดลโลจิสติคที่มีพารามิเตอร์ 1 ตัว (One-parameter Logistic Model) โดยมีข้อตกลงว่าค่าการเดาเท่ากับ 0 และค่าอำนาจจำแนกเท่ากับ 1 นั่นคือ ข้อสอบไม่มีการเดาได้ถูกและข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน โมเดลนี้มีชื่อเรียกอีกอย่างว่า "ราสช์โมเดล" (Rasch Model) เพราะนักคณิตศาสตร์ชาวเดนมาร์คชื่อ "ราสช์" (Rasch) ได้เคยเสนอโมเดลคล้ายคลึงกันไว้ โดยแสดงความสัมพันธ์ของโมเดล ดังนี้

$$P(\theta) = \frac{1}{1 + e^{-1.7(\theta - b)}}$$

โมเดลโลจิสติคทั้งสามที่กล่าวมาข้างต้นมีข้อตกลงเบื้องต้นที่สำคัญ 2 ประการ คือ

1. ความเป็นมิติเดียวของแบบสอบ (Unidimentionality) หมายถึง คือ ข้อทดสอบเหล่านั้นต้องวัดคุณลักษณะ (Trait) เดียว กล่าวคือ ข้อทดสอบเหล่านั้นจะต้องเป็นเอกพันธ์ ซึ่งในการทดสอบว่าแบบสอบมีความเป็นมิติเดียวนั้น นักทดสอบโดยทั่วไปใช้เทคนิคการวิเคราะห์ตัวประกอบ (Factor Analysis) ถ้าพบว่ามีตัวประกอบสำคัญอยู่ 1 ตัวประกอบ

ก็ถือว่าตัวประกอบนั้นวัดมิติเดียว (Lord 1988 10) ซึ่งพิจารณาที่ค่าน้ำหนักตัวประกอบ (Factor Loading) ถ้าน้ำหนักตัวประกอบมีค่าที่ตัวประกอบใดมากที่สุด ในขณะที่ตัวประกอบรองลงมามีน้ำหนักตัวประกอบน้อยกว่าตัวแรกมาก ก็น่าจะประมาณได้ว่าแบบสอบนั้นวัดมิติเดียว

2. ความเป็นอิสระของการตอบข้อคำถาม (Local Independent) หมายถึง ข้อสอบแต่ละข้อเป็นอิสระไม่เกี่ยวข้องกัน คำตอบแต่ละข้อของแต่ละคนเป็นอิสระต่อกัน ซึ่งข้อสอบแต่ละข้อจะอยู่ในตำแหน่งใดของข้อสอบก็ได้ ไม่มีผลต่อการตอบของผู้สอบ ในทางปฏิบัติจะไม่ทำการทดสอบความเป็นอิสระของข้อคำถาม เพราะข้อทดสอบโดยมากจะไม่ฝ่าฝืนข้อตกลงนี้ และถ้าข้อทดสอบชุดใดทำการทดสอบความเป็นมิติเดียวแล้ว ย่อมแสดงว่าแบบสอบนั้นมีความเป็นอิสระของการตอบข้อคำถามด้วยเสมอ (Warm 1978: 107; Lord 1980: 19)

จากโมเดลทั้งสามจะเห็นว่าราสช์โมเดลเป็นโมเดลที่ง่ายที่สุดมีข้อตกลงเบื้องต้นเพิ่มขึ้นมาอีก 2 ประการ คือ ค่าอำนาจจำแนกของทุกข้อเท่ากันและไม่มีการเดา ซึ่งเป็นข้อตกลงที่จำกัดมากสำหรับข้อสอบแบบเลือกตอบ ส่วนโมเดลโลจิสติค 2 พารามิเตอร์ เป็นโมเดลที่ง่ายถัดมาเป็นอันดับที่สอง โดยมีข้อตกลงเพิ่มขึ้นประการเดียว คือ ไม่มีการเดา ส่วนโมเดลโลจิสติค 3 พารามิเตอร์เป็นโมเดลที่ซับซ้อนที่สุดและอนุญาตให้อำนาจจำแนกของข้อคำถามแปรเปลี่ยนได้ และยอมให้มีการเดาด้วย การใช้โมเดลที่ง่ายที่สุดนั้นสะดวกในการประมาณค่าพารามิเตอร์ดังได้กล่าวไว้แล้ว แต่ข้อตกลงเบื้องต้นที่เข้มงวดเกี่ยวกับค่าอำนาจจำแนกและค่าการเดา ทำให้ข้อมูลไม่เหมาะกับโมเดล ถ้าไม่ได้มีการเลือกข้อคำถามสำหรับข้อสอบไว้โดยเฉพาะสำหรับโมเดล หลักฐานอันหนึ่งที่แสดงให้เห็นถึงความล้มเหลวในการพยายามหาความเหมาะสมระหว่างข้อมูลกับราสช์โมเดล คือ การศึกษาของเมลเลนเบิร์ก (Mellenbergh 1972) เขามีความมุ่งหมายที่จะใช้ราสช์โมเดลกับข้อสอบแบบเลือกตอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ซึ่งใช้กับนักเรียนชั้นประถมศึกษาปีที่ 7 ของประเทศเนเธอร์แลนด์และอินโดนีเซีย แต่จากการวิเคราะห์ความเหมาะสมระหว่างข้อมูลกับโมเดล เขาได้พบว่าข้อมูลไม่เหมาะกับโมเดล จึงได้หยุดการวิเคราะห์ไว้ หลักฐานอีกด้านหนึ่งเป็นผลจากการศึกษาศึกษาของแฮมเบิลตันและเทราบ์ (Hambleton and Traub 1973) เขาได้ใช้โมเดลโลจิสติค

2 พารามิเตอร์และราสช์โมเดลกับข้อมูลชุดเดียวกัน (ผลการตอบข้อสอบวัดความถนัด 3 ฉบับ) ซึ่งได้ข้อสรุปว่า โมเดลที่มีจำนวนพารามิเตอร์มากกว่าให้ความเหมาะสมดีกว่า (ผจงจิต อินทสุวรรณ 2525: 63) เออรี่ (Urry 1977 cited by Birnbaum 1978: 20) ให้การสนับสนุนว่า แบบสอบอิงเกณฑ์ชนิดเลือกตอบ โมเดลที่เหมาะสมคือ โมเดลที่ใช้ 3 พารามิเตอร์ เพราะเหมาะกับคุณลักษณะของการตอบแบบสอบเลือกตอบมากที่สุด กล่าวคือ มีการเดาและค่าอำนาจจำแนกของแบบสอบมาเกี่ยวข้องด้วย จึงทำให้การวิเคราะห์ข้อสอบด้วยวิธี Three-parameter Logistic Model เป็นที่นิยมมากในการสอบแบบเลือกตอบ อย่างไรก็ดีการคำนวณข้อมูลด้วยพารามิเตอร์ 3 ตัว ควรใช้กลุ่มตัวอย่างขนาดใหญ่ เช่น ประมาณ 1,000 คน ข้อสอบควรมีจำนวนมาก เช่น ประมาณ 40 ข้อ (Wood and Others 1976: 5)

3. การเลือกข้อกระทงของแบบสอบอิงเกณฑ์

สถิติตามแนวคลาสสิคมีข้อจำกัดในหลายประการตามที่ได้กล่าวไปแล้วนั้น จึงทำให้การเลือกข้อกระทงที่ใช้กับแบบสอบอิงเกณฑ์มีปัญหาเกิดขึ้นไปด้วย ซึ่งปัญหาที่เกิดขึ้นนี้สามารถแก้ข้อจำกัดที่ใช้โดยนำทฤษฎี IRT มาแก้ไขได้ เนื่องจากความสามารถที่แท้จริงซึ่งเป็นส่วนหนึ่งของทฤษฎีคุณลักษณะแฝงมีความสำคัญในการคัดเลือกข้อกระทง ทฤษฎีนี้แตกต่างไปจากทฤษฎีทางการวัดผลแบบเดิม เพราะค่าสถิติประจำข้อสอบ เช่น ความสามารถของผู้สอบ (θ) คะแนนจุดตัด (π_0) สามารถจะแปลงจากค่าหนึ่งให้เป็นอีกค่าหนึ่งได้ จึงสามารถทำให้เลือกข้อกระทงที่สามารถจำแนกผู้สอบได้สูง ณ ที่เป็นคะแนนจุดตัด (Hambleton 1979 ; Lord 1980 ; Wright and Stone 1979) และความสามารถของข้อกระทงแต่ละข้อที่จะทำให้การวัดนั้นถูกต้อง คือ ฟังก์ชั่นสารสนเทศข้อกระทง (Item Information Function IIF) $I\{\theta, u_i\}$

$$I\{\theta, u_i\} = \frac{[P_i'(\theta)]^2}{P_i(\theta)\,Q_i(\theta)} \qquad \text{..... (1)}$$

เมื่อ $P_i'(\theta)$ คือ อนุพันธ์ที่ 1 (First Derivative) ของ $P_i(\theta)$ และ $Q_i(\theta) = 1 - P_i(\theta)$ ดังนั้น $I_i(\theta)$ อาจเขียนอยู่ในรูปของค่าพารามิเตอร์ประจำข้อกระทงได้ คือ

$$I\ \{\theta\ ,\ u_i\} = \frac{D^2 a_i^2\ (1 - c_i)}{[c_i + e^{Da_i(\theta - b_i)}]\ [1 + e^{-Da_i(\theta - b_i)}]^2} \quad \ldots\ldots(2)$$

$$D = 1.7$$

หรือเขียนในรูปที่ง่ายขึ้น คือ

$$I\ \{\theta\ ,\ u_i\} = \frac{D^2 a_i^2\ (1 - c_i)}{(c_i + e^{DL})\ (1 + e^{-DL})^2} \quad \ldots\ldots(3)$$

$$D = 1.7\ ,\quad L = a(\theta - b)$$

ค่า Item Information Function (IIF) มีค่าสูงสุดที่จุด θ_i^*

โดยที่ $$\theta_i^* = b_i + \frac{1}{Da_i} \text{ Log } \frac{1}{2}\ (1 + \sqrt{1 + 8c_i}) \quad \ldots\ldots(4)$$

จากสมการที่ 2 หรือ 3 แสดงให้เห็นว่าข้อสอบที่มีค่าอำนาจจำแนกสูง จะมีประโยชน์มากกว่าข้อสอบที่มีค่าอำนาจจำแนกต่ำ และยิ่งค่าโอกาสที่เดาถูก (c) ต่ำลง จะทำให้ข้อสอบสามารถวัดได้ถูกต้องมากขึ้น โดยไม่ได้ขึ้นกับค่าความสามารถ θ

เมื่อประมาณค่าพารามิเตอร์ของข้อสอบได้แล้ว สมการลักษณะการตอบข้อกระทง (Item Response Model) จะสามารถหาค่าความสามารถที่แท้จริงของผู้สอบได้ถูกต้อง ส่วนค่า θ_i^* และ $I_i\ \{\theta\ ,\ u_i\}$ จะเป็นค่าสถิติที่ดีสำหรับใช้ในการเลือกข้อกระทงได้ถูกต้องและเหมาะสม (Hambleton and De Gruijter 1983: 358)

ลอร์ดและโนวิค (Lord and Novick 1968) ได้กล่าวว่า แบบสอบที่สามารถแยกผู้สอบเป็นผู้รอบรู้จะประกอบด้วยข้อกระทงที่สามารถจำแนกใกล้ ๆ คะแนนจุดตัดที่มีหน่วยเดียวกับคะแนนโดเมน (Domain score scale) ซึ่งเบิร์นบอน (Birnbaum, 1968) ได้สนับสนุนเช่นเดียวกันว่า แบบสอบที่สร้างขึ้นเพื่อจำแนกผู้รอบรู้เป็น 2 ระดับ หรือมากกว่าให้สัมพันธ์กับโดเมนเนื้อหาที่ต้องการวัด ควรเลือกข้อกระทงที่จำแนกผู้สอบได้มากที่สุดที่คะแนนจุดตัด (Hambleton and De Gruijter 1983:355 citing Birnbaum 1968)

ฮาลาไดนา และรอยด์ (Haladyna and Roid 1983) ได้สรุปว่า การนำเอาทฤษฎีการตอบสนองข้อกระทงมาใช้กับแนวคิดของการกำหนดเกณฑ์มาตรฐานหรือจุดตัดของแบบสอบอิงเกณฑ์ ได้ข้อค้นพบที่สำคัญ คือ การคัดเลือกข้อกระทงโดยให้ผู้สอบได้ทำข้อสอบที่มีความยากตรงกับความสามารถของเขา ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถจะต่ำกว่าการสุ่มข้อกระทงขึ้นมาใช้เพื่อการสอบ

จากการทดสอบของแฮมเบิลตันและกรุจเตอร์ (Hambleton and De Gruijter 1983: 362) แสดงผลว่า การเลือกข้อกระทงเฉพาะข้อที่เหมาะสมกับโดเมน คือ เลือกข้อที่มีค่าพารามิเตอร์ใกล้ความสามารถ (θ_0) ที่ต้องการวัด และข้อสอบในแต่ละข้อในโดเมนเดียวกันมีความเป็นเอกพันธ์ (Homogeneity) คือความยากใกล้เคียงกัน ค่าอำนาจจำแนกแต่ละข้อใกล้เคียงกัน จะทำให้ค่าจำแนกผิดพลาดต่ำกว่าเลือกข้อกระทงที่ได้จากการสุ่ม

ซิลวา ชาร์รอน เจน (Silva Sharron Jane 1984) ได้เปรียบเทียบวิธีการเลือกข้อกระทงของแบบสอบอิงเกณฑ์โดยใช้วิธีของคลาสสิคอล 3 วิธี คือ ใช้ดัชนีของคอกซ์ วากัส (Cox Vargus) พอยท์ ไบซีเรียล (Point Biserial) และสัมประสิทธิ์ฟี (ϕ - coefficiency) กับวิธีของทฤษฎีการตอบสนองข้อกระทงที่ใช้ 3 พารามิเตอร์ ได้ข้อสรุปว่า โมเดลของทฤษฎีการตอบสนองข้อกระทงสามารถจำแนกเป็นผู้รอบรู้ได้ดีกว่าทฤษฎีคลาสสิคในการเลือกข้อกระทง

แต่ทั้งนี้ไม่ว่าการดำเนินการสอบจะจัดให้ระดับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบหรือการสุ่มข้อกระทงขึ้นมาสอบ ต่างก็ต้องการข้อสอบที่วัดในเนื้อหาเดียวกันและให้ข้อดีในแต่ละสภาพการใช้ ดังนั้นการตัดสินว่าจะใช้วิธีใดก็พิจารณาจากผลประโยชน์ที่ได้จากการทดสอบจริง

4. ความหมายและความสำคัญของคะแนนโดเมน

คะแนนโดเมน (π) หมายถึง สัดส่วนของจำนวนข้อที่ตอบถูกต่อจำนวนข้อสอบทั้งหมด เช่น 1.00 หมายถึง ผู้สอบตอบแบบสอบถูก 100% จะอ้างอิงไปสู่ปริมาณ 100% ของปริเขตของการสอบวัดนั้น ซึ่งคะแนนโดเมนนี้จะเป็นตัวคาดคะเนระดับผลสัมฤทธิ์ทางการเรียนที่ปราศจากอคติ (Unbiased estimator) และมีความพอเพียง (sufficient statistic) ก็ต่อเมื่อต้องมีการระบุเขตของข้อสอบที่เป็นไปได้ในแต่ละโดเมนไว้ล่วงหน้า และข้อสอบในแบบสอบอิงโดเมนต้องเป็นกลุ่มตัวอย่างแบบสุ่มของข้อสอบในโดเมนนั้น เป็นที่น่าสังเกตว่า คะแนนรวมใช้เป็นตัวคาดคะเนที่ดีได้ ก็ต่อเมื่อข้อสอบแต่ละข้อล้วนวัดในสิ่งเดียวกัน หรือแบบสอบนั้นมีลักษณะเป็นเอกพันธ์ (Homogeneity) และแบบสอบอิงโดเมน (ปัจจุบันการทดสอบอิงโดเมนมักจะรู้จักในชื่อของการทดสอบอิงเกณฑ์) เป็นแบบสอบที่มีคุณสมบัติในเชิงเอกพันธ์สูง ซึ่งสามารถใช้คะแนนรวมเป็นตัวประมาณค่าผลสัมฤทธิ์ทางการเรียนหรือลักษณะ (Trait) ที่มุ่งวัดได้ (สมหวัง พิธิยานุวัฒน์ 2529: 33) ซึ่งค่าพารามิเตอร์ของข้อสอบหลายข้อที่อยู่ในโดเมนเดียวกัน หรือค่าพารามิเตอร์ของข้อสอบชุดหนึ่งที่คิดว่าเป็นตัวแทนที่ดีของโดเมน จะทำให้สามารถหาความสัมพันธ์ระหว่างคะแนนโดเมน π และคะแนนความสามารถที่แท้จริง (θ) เนื่องจากคะแนนโดเมน คือ ค่าคาดหวัง (Expect Value) ของคะแนนจากข้อสอบทุกข้อในโดเมน

$$\pi = E_i P_i (\theta)$$

E_i คือ ค่าความคาดหวัง

(Hambleton and De Gruijter 1983: 358)

ด้วยเหตุนี้คะแนนโดเมนจึงสามารถสรุปผลความรอบรู้หรือไม่รอบรู้ได้ชัดเจน ซึ่งสามารถใช้การคำนวณค่าสถิติคะแนนโดเมนในแนวของคลาสสิคจากเปอร์เซนต์การตอบถูกหรือสัดส่วนการตอบถูก ตามที่แฮมเบิลตัน สวามินาทาน อัลจีนา และคูลซัน (Hambleton, Swaminathan, Algina and Coulson, 1978) ได้ใช้แบบสอบอิงเกณฑ์ศึกษาในเรื่องของการประมาณคะแนนโดเมนในการจำแนกผู้สอบ ส่วนการคำนวณค่าคะแนนโดเมนในเชิงทฤษฎีการตอบสนองข้อกระทงคำนวณจากเปอร์เซนต์ของผลรวมของโอกาสตอบถูกรายข้อของข้อสอบในแบบสอบในแต่ละค่าความสามารถ (θ)

5. จุดอ่อนของค่าสถิติตามแนวคลาสสิค

การวัดผลที่ปฏิบัติกันอยู่ในขณะนี้ ค่าสถิติที่ใช้คือ ค่าความยากของข้อสอบ (p) และค่าอำนาจจำแนก (r) ค่าความยากมีค่าอยู่ระหว่า 0 ถึง 1 คำนวณได้จากจำนวนคนที่ตอบข้อกระทงข้อนั้นถูกหารด้วยผู้เข้าสอบทั้งหมด สำหรับคะแนนโดเมน (Domain Scores) ของแต่ละคนเขียนแทนด้วยสัญลักษณ์ π มีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งคิดมาจากจำนวนข้อสอบที่ตอบถูกกับจำนวนข้อสอบทั้งหมดในโดเมน โดยที่การอ้างอิงค่าความยากของข้อสอบจะใช้ผู้เข้าสอบทั้งหมด สำหรับคะแนนโดเมนจะใช้โดเมนเนื้อหามาพิจารณา ซึ่งคะแนนจุดตัดใช้สัญลักษณ์ π_0 อยู่ในหน่วยเดียวกันกับคะแนนโดเมน π ในขณะที่มาตราวัดคะแนนโดเมนและมาตราวัดค่าความยากไม่มีความสัมพันธ์กัน แม้ว่าจะรู้คะแนนจุดตัดในรูปคะแนนโดเมน π_0 ก็ไม่สามารถนำค่าสถิติมาเลือกข้อกระทงให้เหมาะสมและถูกต้องได้ เช่น กำหนดจุดตัดที่อยู่ในรูปคะแนนโดเมน $\pi_0 = .80$ แล้วข้อสอบใดที่มีผู้ตอบถูก 80% ของประชากรผู้เข้าสอบไม่ได้หมายความว่า ข้อสอบเหล่านั้นจะสามารถจำแนกผู้สอบเป็นผู้รอบรู้ในเนื้อหาได้มากกว่า 80% หรือน้อยกว่า 80% ได้ แสดงว่าข้อสอบที่มีค่าความยาก (p) .80 ไม่สามารถจำแนกผู้สอบตรงคะแนนโดเมน .80 ออกจากกันได้

ตัวอย่าง เช่น แบ่งนักเรียนออกเป็น 5 กลุ่ม ๆ ละ 20 คน โดยที่แต่ละคนทำข้อสอบ 3 ข้อ คือ ข้อง่าย ปานกลาง และยาก ผลการสอบของนักเรียนแต่ละกลุ่มปรากฏผลดังตารางที่ 1

ตารางที่ 1 จำนวนผู้เข้าสอบในแต่ละกลุ่มที่ตอบข้อสอบแต่ละข้อได้ถูกต้อง

กลุ่ม	ขนาดของกลุ่ม	ความยากของข้อสอบ			คะแนนโดเมน
		ง่าย	ปานกลาง	ยาก	
ก	20	20	20	20	1.00
ข	20	20	20	20	1.00
ค	20	20	20	20	1.00
ง	20	20	20	0	.67
จ	20	20	0	0	.33
ค่าความยาก (p)		1.00	.80	.60	

จากตารางจะเห็นว่า ทุกกลุ่มทำข้อสอบง่ายถูกหมดมี 4 กลุ่ม ที่ทำข้อปานกลางถูก และมี 3 กลุ่มที่ทำข้อยากถูก ค่าความยาก p จากข้อง่าย ปานกลาง และยาก มีค่าเท่ากับ 1.00, .80 และ .60 ตามลำดับ ถ้ากำหนดจุดตัดของคะแนนโดเมน (π_0) เท่ากับ .80 แล้ว จะเห็นว่ามีผู้เข้าสอบ 60 คนใน 100 คน ที่มีความรอบรู้และผ่านการสอบ แต่ถ้าใช้ข้อสอบที่มีค่า p เท่ากับ .80 เท่านั้น จะมีอีก 20 คนที่จะผ่านอย่างไม่ถูกต้อง คือ ผ่านแต่ไม่รอบรู้ เพราะ p เท่ากับ .80 หมายถึงว่าจะต้องมีผู้ทำข้อสอบถูก 80 คนใน 100 คน ข้อสอบที่ดีที่สุดจึงควรเป็นข้อสอบที่มีค่าความยาก .60 เพราะข้อสอบเหล่านี้จะสามารถแบ่งกลุ่มระหว่างกลุ่มคนที่รู้จริง (True master) ออกจากกลุ่มที่ไม่รู้จริง (True Nonmaster) ได้ถูกต้อง จากตัวอย่างในตารางนี้แสดงให้เห็นชัดเจนว่า ข้อสอบที่เลือกเอาแต่เฉพาะข้อที่มีค่าความยากเท่ากับคะแนนจุดตัดของคะแนนโดเมน (π_0) ไม่ใช่ข้อสอบที่ดีที่สุดที่จะใช้ในการแบ่งกลุ่มคนให้ถูกต้อง (Hambleton and De Gruijter, 1983: 357)

จากที่ได้กล่าวไว้ในตอนต้นว่า คะแนนจุดตัดที่อยู่ในรูปคะแนนโดเมนอยู่ในหน่วยเดียวกับคะแนนโดเมน ในขณะที่คะแนนโดเมนและค่าความยากอยู่กันคนละหน่วย ดังนั้น จึงทำให้หาความสัมพันธ์ระหว่างคะแนนโดเมนและคะแนนความสามารถที่แท้จริง (θ) ได้ เมื่อทราบค่าพารามิเตอร์ของข้อสอบทั้งหมดที่อยู่ในโดเมนเดียวกันที่เป็นตัวแทนที่ดีของโดเมน ทั้งนี้เพราะความสัมพันธ์ดังกล่าว อันได้แก่ คะแนนโดเมน คือ ค่าคาดหวัง (Expected Score) ของคะแนนจากข้อสอบทุกข้อในโดเมน

$$\pi = E\ P_i(\theta)$$

ซึ่งจะได้ความสัมพันธ์ระหว่างคะแนนโดเมนและค่าความสามารถ ดังนี้

$$\pi = \frac{1}{m} \sum_{i=1}^{m} P_i(\theta)$$

เมื่อ m คือ จำนวนข้อในกลุ่มตัวอย่างของข้อสอบ (Lord and Novick 1968; Lord 1980) ดังนั้น คะแนนจุดตัดที่อยู่ในรูปคะแนนโดเมน (π_o) ซึ่งอยู่ในหน่วยเดียวกันกับคะแนนโดเมน จึงสามารถเขียนความสัมพันธ์ได้ ดังนี้

$$\pi_o = \frac{1}{m} \sum_{i=1}^{m} P_i(\theta_o)$$

จากความสัมพันธ์นี้จึงทำให้สามารถหา θ_o จากคะแนนความสามารถที่แท้จริง ที่ตรงกับ π_o ซึ่ง π_o อยู่ในหน่วยเดียวกับคะแนนโดเมน

6. การกำหนดคะแนนจุดตัดให้สอดคล้องกับคะแนนโดเมน

แฮมเบิลตันและโนวิค (Hambleton and Novick 1973: 163) ได้กล่าวว่า แบบสอบอิงเกณฑ์เป็นแบบสอบที่กำหนดเกณฑ์มาตรฐานหรือคะแนนจุดตัด (Cut-off Score) โดยที่แบบสอบประกอบด้วยข้อกระทงที่สุ่มขึ้นมาจากประชากรข้อสอบ คือ จำนวนข้อสอบทั้งหมดภายในขอบเขตเนื้อหาที่กำหนดขึ้น สัดส่วนจำนวนข้อสอบที่ผู้สอบทำได้ต่อจำนวนข้อสอบทั้งหมดจะแทนด้วยสัญลักษณ์ π ในทางทฤษฎีค่า π จะถือว่าเป็นค่าประมาณของความรู้ที่แท้จริงของผู้สอบในของเขตของเนื้อหาที่กำหนด หรือมีความหมายว่า ค่า π คือ อัตราส่วนจำนวนข้อสอบทั้งหมดในขอบเขตที่ผู้สอบทำได้ต่อจำนวนประชากรข้อสอบทั้งหมด ถ้าหากผู้สอบได้ทำข้อสอบทุก ๆ ข้อในขอบเขตของเนื้อหานั้น ๆ จุดตัดที่เป็นเกณฑ์มาตรฐานจะแทนด้วยสัญลักษณ์ π_0 จุดตัดดังกล่าวจะถือเป็นจุดแบ่งผู้สอบออกเป็นสองส่วน คือ ผู้สอบที่ได้คะแนนสูงกว่าเกณฑ์มาตรฐาน ($\pi > \pi_0$) ซึ่งจะถือว่าเป็นผู้รอบรู้ (Master) และผู้ที่ได้คะแนนต่ำกว่าเกณฑ์มาตรฐาน ($\pi < \pi_0$) ซึ่งจะต้องถือว่าเป็นผู้ไม่รอบรู้ (Nonmaster)

ในทางปฏิบัติเกณฑ์มาตรฐานหรือจุดตัด หมายถึง สัดส่วนจำนวนข้อสอบที่ผู้สอบทำได้ต่อจำนวนข้อสอบทั้งหมดในแบบสอบ ทั้งนี้ถ้าผู้สอบคนใดทำแบบสอบได้ถูกต้อง โดยมีสัดส่วนข้อสอบที่ทำถูกต่อจำนวนข้อทั้งหมดในแบบสอบสูงกว่าสัดส่วนที่กำหนด จะถือว่า ผู้สอบเป็นผู้รอบรู้ และถ้าผู้สอบทำแบบสอบได้โดยมีสัดส่วนข้อสอบที่ทำถูกต่อจำนวนข้อสอบทั้งหมดต่ำกว่าสัดส่วนที่กำหนด จะถือว่าผู้สอบเป็นผู้ไม่รอบรู้ โดยที่แบบสอบอิงเกณฑ์แต่ละข้อเป็นตัวแทนของข้อสอบทั้งหมดเท่าที่จะเป็นไปได้ในโดเมนนั้นจริง และมีความยาวพอเพียง ค่าที่ได้จากการประมาณโดยวิธีนี้จะเป็นค่าประมาณที่ไม่ลำเอียงของโดเมน ซึ่งแฮมเบิลตัน และคณะ (Hamblelton et al. 1978: 5) เรียกสัดส่วนนี้ว่า คะแนนโดเมน หาได้จากสูตร ดังนี้

$$\pi = \frac{x}{m}$$

x คือ คะแนนที่ได้จากแบบสอบ

m คือ จำนวนข้อสอบในแบบสอบ

π คือ คะแนนโดเมน

คะแนนโดเมนนี้เป็นแนวคิดตามทฤษฎีคลาสสิค ซึ่งจะเห็นว่าค่าที่ได้จะเป็นค่าคงที่ขึ้นกับจำนวนข้อที่ตอบถูก และความยาวแบบสอบ

ส่วนคะแนนโดเมนตามแนวทฤษฎีการตอบสนองข้อกระทง ซึ่งเกี่ยวข้องกับค่าความยาก ค่าอำนาจจำแนก ค่าการเดา และค่าความสามารถนั้นจะให้ค่าที่แปรเปลี่ยนไปตามความสามารถของผู้สอบ หาได้จากสูตร

$$\pi = \frac{1}{m} \sum_{i=1}^{m} P_i(\theta)$$

7. ความคลาดเคลื่อนในการตัดสินและความเที่ยงของการจำแนกความรอบรู้

ความคลาดเคลื่อนในการตัดสินความรอบรู้เป็นค่าที่ใช้สำหรับตัดสินที่จะให้โอกาสคลาดเคลื่อนที่ยอมรับได้จากการตัดสินเป็นผู้รอบรู้ ซึ่งค่านี้สามารถหาได้จากความสัมพันธ์ของคะแนนโดเมนคะแนนจุดตัด และความยาวของแบบสอบ ซึ่งงานวิจัยครั้งนี้ได้ศึกษาตามวิธีของวิลคอกซ์ ที่คะแนนจุดตัดและความยาว m ข้อ โดยใช้การหาความคลาดเคลื่อน ซึ่งใช้สัญลักษณ์จากการคำนวณในรูปของไบโนเมียล (Binomial Model)

$$P = \binom{m}{c} (\pi^*)^c \ (1 - \pi^*)^{m-c}$$

โดยที่ m คือ ความยาวแบบสอบ

c คือ คะแนนจุดตัดในรูปของคะแนน

π^* คือ คะแนนจุดตัดในรูปของคะแนนโดเมน

จากการศึกษาของแฮมเบิลตัน และกรุจเตอร์ (Hambleton and De Gruijter, 1983: 359) ยอมให้ค่าโอกาสจำแนกผิดพลาดได้ไม่เกิน .20 ($P^* < .20$) ผู้วิจัยใช้เกณฑ์ของค่าดังกล่าวเป็นค่าที่ใช้ตัดสินใจที่จะยอมรับได้

1031CCC9

ส่วนค่าความเที่ยงของแบบสอบอิงเกณฑ์เป็นดัชนีที่ใช้ในการจำแนกความรอบรู้ได้คงเส้นคงวา ซึ่งแฮมเบิลตัน สวามินาทาน อัลจีนา และ คูลซัน (Subkoviak 1976: 129-185 Citing Hambleton Swaminathan Algina and coulson 1978: 15-23) ได้ให้แนวคิดเกี่ยวกับความเที่ยงของการทดสอบอิงเกณฑ์ไว้ 3 วิธี คือ

1. ความเที่ยงในการจำแนกความรอบรู้
2. ความเที่ยงของคะแนนแบบสอบอิงเกณฑ์
3. ความเที่ยงของการประมาณคะแนนโดเมน

1. ความเที่ยงในการจำแนกความรอบรู้ หมายถึง ความคงที่ในการตัดสินให้เป็นผู้รอบรู้หรือไม่รอบรู้จากการพิจารณาแบบสอบ 2 ฟอร์ม ผู้ให้แนวคิดนี้ คือ ฮวิน (Huynh 1976) มาร์แชล และแฮร์เทล (Marshall & Haertel 1976) สับโคเวียค (Subkoviak 1976) สวามินาทาน และ อัลจีนา (Swaminathan and Algina 1974)

2. ความเที่ยงของคะแนนแบบสอบอิงเกณฑ์ เป็นความเที่ยงที่เกี่ยวกับส่วนเบี่ยงเบนของคะแนนกับคะแนนจุดตัด และพิจารณาความคงที่ของค่าที่เบี่ยงเบนจากแบบสอบ 2 ฟอร์ม ผู้ที่ให้ความหมายความเที่ยงในแนวคิดนี้ คือ เบรนเนน เคน และ แฮมเบิลตัน (Brennan Kane 1977; Hambleton 1972)

3. ความเที่ยงของการประมาณคะแนนโดเมน เป็นความเที่ยงที่เกี่ยวกับความคงที่ของคะแนนแต่ละคนกับแบบสอบฟอร์ม 1 และฟอร์มที่ 2 โดยไม่เกี่ยวข้องกับคะแนนจุดตัด ซึ่ง Hambleton, Swaminathan, Algina and coulson (1978: 17-20) ได้เสนอสัมประสิทธิ์ที่สามารถใช้วัดความเที่ยงตามแนวคิดนี้ ดัชนีนั้นเหมือนกับความคลาดเคลื่อนมาตรฐานในการวัด

ความเที่ยงในการความหมายแรก คือ ความเที่ยงในการจำแนกความรอบรู้ ซึ่งความเที่ยงในแนวคิดนี้มีผู้ให้ไว้ 5 วิธี คือ

1. วิธีของคาร์เวอร์ (Carver, 1970)
2. วิธีของสวามินาทาน แฮมเบิลตัน และอัลจีนา (1974)
3. วิธีของฮิวิน (1976)
4. วิธีของสับโคเวียค (1976)
5. วิธีของมาร์แชลและแฮร์เทล (1976)

สองวิธีแรกเป็นการหาค่าความเที่ยงจากการสอบ 2 ครั้ง ส่วน 3 วิธีหลังเป็นการหาค่าความเที่ยงจากการสอบครั้งเดียว เนื่องให้สอดคล้องกับงานวิจัยครั้งนี้ ซึ่งทำการสอบครั้งเดียว ผู้วิจัยจึงได้เสนอแนวคิดการหาค่าความเที่ยงในการตัดสินในการสอบ 1 ครั้ง วิธีที่นิยมกันคือ วิธีของสับโคเวียคและวิธีของฮิวิน

สับโคเวียค (Subkoviak 1976: 265-276) ได้เสนอวิธีการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์ที่ใช้สอบครั้งเดียว อาศัยวิธีการหาสัมประสิทธิ์ความสอดคล้องในการกำหนดความรอบรู้ของนักเรียนแต่ละคน และของกลุ่มที่ได้มาจากการสอบด้วยแบบสอบคู่ขนาน 2 ฟอร์ม ซึ่งมีคะแนน X และ X' แต่ค่าสถิติต่าง ๆ ที่คำนวณมาจากคะแนน X' นั้น สับโคเวียคได้ดัดแปลงสูตรให้สามารถคำนวณจากคะแนน X จึงสามารถหาค่าความเที่ยงจากการสอบครั้งเดียว สูตรการหาความเที่ยงของสับโคเวียต มีดังนี้

$$P_c = \frac{\sum_{i=1}^{N} P_c^{(i)}}{N} \qquad \text{.....(1)}$$

N คือ จำนวนคน

P_c คือ ค่าความเที่ยงของแบบสอบอิงเกณฑ์หรือค่าความสอดคล้องในการตัดสินความรอบรู้ของทั้งกลุ่ม

โดยที่ $P_c^{(i)}$ คือ ค่าการตัดสินความสอดคล้องแต่ละคนในการจำแนก โดยหาได้จากสูตร ดังนี้

$P_c^{(i)}$ คือ $P(X_i \geq c)\ P(X' \geq c) + P(X_i < c)\ P(X < c)$ ..(2)

X_i คือ คะแนนที่ได้จากแบบสอบ

X_i' คือ คะแนนที่ได้เสมือนจากแบบสอบฟอร์มที่ 2

c คือ คะแนนจุดตัด

สำหรับการแจกแจงคะแนน X และ X' ของแต่ละคนอยู่ในรูปของไบโนเมียล (Binomial) โดยมีข้อตกลง คือ

1. ข้อสอบแต่ละข้อในแบบสอบ n ข้อ มีการให้คะแนนตอบถูกได้ 1 ตอบผิดได้ 0
2. ผลการสอบจากข้อหนึ่งข้อใดจะไม่ส่งผลต่อการสอบข้ออื่น ๆ
3. ความน่าจะเป็นในการตอบถูกมีค่าคงที่เหมือนกันทุกข้อ

ดังนั้น จากสมการที่ (2) จัดรูปสมการใหม่ได้ ดังนี้

$$P_c^{(i)} = [P(X_i \geq c)]^2 + [P(X_i < c)]^2$$

$$= [P(X_i \geq c)]^2 + [1 - P(X_i \geq c)]^2$$

เมื่อ $P(X_i \geq c) = \sum_{x_i=c}^{n} \binom{n}{x_i} P_i^{x_i} (1 - P_i)^{n-x_i}$

P_i แทนค่า ความน่าจะเป็นจริงของการตอบข้อสอบถูก ซึ่งมีวิธีการหา 2 วิธี คือ

ก. เมื่อจำนวนข้อมาก ($n > 40$) ค่าประมาณ P_i ได้จากการคำนวณ

$$P_i = \frac{x_i}{n}$$

ข. เมื่อจำนวนข้อน้อย (n < 40)

$$P_i = \alpha_{21/x} [X_i/n] + [1 - \alpha_{21/x}][M_x/n)$$

$\alpha_{21/x}$ คือ ค่าความเที่ยงแบบคูเดอร์ริชาร์ดสัน 21

M_x คือ ค่าเฉลี่ย

ฮวิน (Huynh 1976: 253-263) ได้เสนอวิธีการประมาณค่าความเที่ยงจากการสอบครั้งเดียว โดยใช้แนวคิดการพัฒนาสูตรมาจากการใช้ฟอร์มคู่ขนาน 2 ฟอร์ม จากการสอบเพียง 1 ครั้ง และใช้คะแนนที่สอบได้จากแบบสอบอิงเกณฑ์ในรูปอิงโดเมน ซึ่งสมมติว่าใช้กระบวนการสุ่มความน่าจะเป็นในการคัดเลือกข้อกระทงจากประชากรข้อสอบที่นิยามไว้ชัดเจน ฮวินเสนอสูตรการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์ไว้ 2 สูตร คือ

$$K = (P_{11} - P_1^2) / (P_1 - P_1^2) \quad \text{เมื่อคะแนนจุดตัดใกล้ } n$$

$$K = (P_{oo} - P_o^2) / (P_o - P_o^2) \quad \text{เมื่อคะแนนจุดตัดใกล้ } 0$$

โดยที่ P_{11} คือ สัดส่วนผู้ถูกจำแนกลงในกลุ่มผู้รอบรู้ที่สอดคล้องกันทั้ง 2 ฟอร์ม

P_{oo} คือ สัดส่วนผู้ถูกจำแนกลงในกลุ่มไม่รอบรู้ที่สอดคล้องกันทั้ง 2 ฟอร์ม

P_1 คือ สัดส่วนผู้ถูกจำแนกลงในกลุ่มผู้รอบรู้ฟอร์มใดฟอร์มหนึ่ง

P_o คือ สัดส่วนผู้ถูกจำแนกลงในกลุ่มไม่รอบรู้ฟอร์มใดฟอร์มหนึ่ง

และเนื่องจากการคิดคำนวณตามวิธีของฮวินในกรณีที่จำนวนข้อมากกว่า 10 ข้อ มีความยุ่งยากในการคำนวณ คีตส์และลอร์ด (Keats and Lord, 1962) ได้เสนอแนวคิดในการแปลง X' คะแนนจากคะแนนสอบจริง X จากสูตร

$$X' = \text{Sin}^{-1} \sqrt{x/n}$$

มาใช้ประมาณการแจกแจงปกติ หาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน คือ

$$\mu_{x'} = \text{Sin}^{-1} \sqrt{\mu/n}$$

$$\sigma_{x'} = [(\alpha_{21} + 1)/ (\alpha + n)]^{\frac{1}{2}}$$

เมื่อ $\alpha = (-1 + 1 / \alpha_{21})\mu$

$$\alpha_{21} = \frac{n}{n-1} \left[1 - \frac{\mu(n-\mu)}{n\sigma^2}\right]$$

โดยที่การประมาณค่า μ ด้วย $\bar{X}$ และ σ ด้วย S

ถ้า X' และ Y' เป็นคะแนนที่แปลงจากแบบสอบฟอร์มคู่ขนานแล้ว ค่าสหสัมพันธ์ของคะแนนทั้งสองจะมีค่าประมาณ ดังนี้

$$\rho = \alpha_{21} [(n - 1)/(n + \alpha_{21})]^{\frac{1}{2}}$$

และการประมาณค่าการแจกแจงปกติจะสามารถประมาณได้ดี เมื่อ μ/n มีค่าอยู่ระหว่าง .15 และ .85 จำนวนข้อ (n) อย่างน้อย 8 ข้อ (Novick Lewis and Jackson, 1973 citing Huynh, 1976: 258-259)

เนื่องจากฟังก์ชั่น $\sin^{-1}$ มีเส้นกราฟสูงขึ้นแบบ Monotonic การกำหนดความรอบรู้ จึงนิยามด้วย $X' \geq c'$ เมื่อ

$$c' = \sin^{-1}\sqrt{(c - .5)/n}$$

และใช้ตาราง Univariate and bivariate Normal Distribution ของกูปตา (Gupta 1963) คำนวณหาค่า P_o และ P_{oo} (ดูตารางกูปตาในภาคผนวก)

ขั้นตอนในการคำนวณสัมประสิทธิ์แคปป้า (K) เมื่อจำนวนข้อมีจำนวนมาก มีดังนี้

1. คำนวณหาค่า $\bar{X}$, S^2 และ α_{21} จากคะแนนที่สอบได้

2. ประมาณค่า $\mu_{x'}$, $\sigma_{x'}$ และ ρ โดยการแทนค่า μ ด้วย $\bar{X}$ และแทนค่า σ ด้วย S

3. คำนวณหาค่า Z จากสูตร

$$Z = \frac{c' - \mu_{x'}}{\sigma_{x'}}$$

4. ประมาณค่า P_o โดยใช้ค่าความน่าจะเป็นของการกระจายปกติมาตรฐานที่น้อยกว่า Z ในการคำนวณ

5. ประมาณค่า P_{oo} โดยการเปิดตาราง ซึ่งเป็นค่าน่าจะเป็นของตัวแปรปกติมาตรฐาน 2 ค่าที่จะมีค่าสหสัมพันธ์ ρ น้อยกว่า Z ในการคำนวณ

6. คำนวณหาสัมประสิทธิ์แคปปา K จากสูตร

$$K = (P_{oo} - P_o^2) / (P_o - P_o^2)$$

สำหรับวิธีของอวินในการหาสัมประสิทธิ์แคปป้าที่เป็นดัชนีประมาณค่าความเที่ยงในการจำแนกความรอบรู้ในแนวคิดทฤษฎีคลาสสิค ถ้ามีจำนวนข้อกระทง มากกว่า 10 ข้อ จะมีขั้นตอนที่ลดความยุ่งยากในการประมาณค่า ซึ่งสับโคเวียค (Subkoviak 1978: 114) ได้กล่าวว่า วิธีของอวินเป็นวิธีที่ดี ใช้ในการประมาณการสอบครั้งเดียว และให้การประมาณที่ถูกต้องอย่างมีเหตุผล ให้การประมาณความคลาดเคลื่อนต่ำเล็กน้อยสำหรับแบบสอบสั้น

ส่วนการหาคุณภาพของแบบสอบในแนวทฤษฎีการตอบสนองข้อกระทงนั้น ใช้พิจารณาจากฟังก์ชั่นสารสนเทศแบบสอบ (Test Information Function) ซึ่งเป็นดัชนีชี้ให้เห็นคุณภาพในการวัดของแบบสอบ โดยแบบสอบจะมีประสิทธิภาพสูงสุดที่ระดับความสามารถที่ให้ฟังก์ชั่นสารสนเทศข้อกระทง (Item Information function) สูงสุด วอร์ม (Warm 1978: 76-77) ได้หาความสัมพันธ์ของความคลาดเคลื่อนมาตรฐานของการประมาณ (Standard Error of Estimate SEE) กับฟังก์ชั่นสารสนเทศแบบสอบ โดยกล่าวว่า ถ้าการประมาณค่าความสามารถ (θ) เป็นการประมาณความน่าจะเป็นสูงสุด (Maximum Likelihood of Estmate) แล้ว แต่ละความสามารถ θ ก็สามารถหาได้จากฟังก์ชั่นสารสนเทศแบบสอบ $I(\theta)$ ดังนี้

$$SEE = \frac{1}{\sqrt{I(\theta)}}$$

เนื่องจาก $I(\theta)$ แปรผันไปตามมาตราวัดความสามารถ (θ scale) จึงเกิดค่าความคลาดเคลื่อน มาตรฐานของการประมาณ (SEE) ขึ้น ยิ่ง $I(\theta)$ มีค่ามาก SEE ก็ยิ่งมีค่าน้อย และ SEE จะมีค่าน้อยที่จุดตัด วอร์มได้หาความสัมพันธ์ของค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของการประมาณ ($\overline{SEE}$) กับค่าความเที่ยง ดังนี้

$$r = 1 - \overline{SEE}^2$$

ซึ่งความสัมพันธ์นี้ให้ความหมายได้ชัดเจน เมื่อความเที่ยงได้จากการกระจายความสามารถ (θ) ถ้าผู้สอบทั้งหมดกระจายบนมาตราวัดความสามารถที่มีสารสนเทศ (Information) สูงแล้ว ค่าความเที่ยงย่อมจะสูงกว่าความสามารถที่ให้สารสนเทศต่ำ

งานวิจัยที่เกี่ยวข้อง

สำหรับงานวิจัยที่เกี่ยวข้องที่ผู้วิจัยทำการวิจัยครั้งนี้ ขอเสนอในหัวข้อ ดังนี้

1. การวิเคราะห์ข้อกระทงโดยใช้โปรแกรมโลจิสติค 3 พารามิเตอร์
2. การใช้แบบสอบอิงโดเมนในการกำหนดจุดตัด
3. การเปรียบเทียบผลของคะแนนจุดตัดโดยวิธีตามแนวทฤษฎีคลาสสิคกับแนวทฤษฎีการตอบสนองข้อกระทง
4. การกำหนดความยาวของแบบสอบอิงเกณฑ์โดยใช้ทฤษฎีการตอบสนองข้อกระทงและการหาค่าความน่าจะเป็นในการจำแนกผิดพลาด
5. การหาค่าความเที่ยงของคะแนนจุดตัด

1. การวิเคราะห์ข้อกระทงโดยใช้โปรแกรมโลจิสติค 3 พารามิเตอร์

งานวิจัยที่เกี่ยวข้องในการวิเคราะห์ ลักษณะนี้ คือ งานวิจัยของ เสนอ ภิรมจิตรผ่อง (2527) ได้ศึกษาการวิเคราะห์แบบสอบวัดความถนัดทางการเรียนด้านจำนวนชั้นประถมศึกษาปีที่ 6 จำนวน 30 ข้อ โดยคำนวณค่าพารามิเตอร์ของข้อสอบด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูปโลจิสติค 3 พารามิเตอร์ ของเบิร์นบอร์ม ประกอบด้วยการพิจารณาโค้งแสดงสารสนเทศของข้อกระทง (Item Information Curve) โค้งแสดงสารสนเทศข้อสอบ (Test Information Curve) จากผลการทดสอบกับกลุ่มนักเรียนจำนวน 1,635 คน ปรากฏผลดังนี้ ในจำนวนข้อสอบทั้งหมด 30 ข้อ แบบทดสอบนี้เป็นแบบทดสอบที่ใช้เป็นข้อสอบในการสอบคัดเลือกนักเรียนตามโครงการการศึกษาพัฒนาและส่งเสริมผู้มีปรีชาญาณทางวิทยาศาสตร์ มีจำนวน 23 ข้อ ที่มีค่าอำนาจจำแนก (a) มากกว่า .5 ค่าความยาก (b) มากกว่า -2 มีประโยชน์ที่เหมาะสมในการวัดผู้สอบที่มีระดับความสามารถในช่วง -1.70 ถึง 1.55 จำนวน 22 ข้อ และอีก 1 ข้อ มีความเหมาะสมกับผู้สอบที่มีความสามารถค่อนข้างสูงในช่วงระดับความสามารถ .50 ขึ้นไป

นอกจากนี้ก็มีผลงานการศึกษาของเออรี่ (Urry, 1977 ; cited by Warm 1978: 20) ได้ศึกษาการเปรียบเทียบผล การใช้โลจิสติคโมเดลทั้งแบบ 1, 2 และ 3 พารามิเตอร์ ในปี ค.ศ. 1970 ได้ผลสรุปว่าการวิเคราะห์โดยใช้ 3 พารามิเตอร์ เหมาะสำหรับแบบสอบชนิดเลือกตอบมากที่สุด เพราะเพิ่มค่าการเดา และค่าอำนาจจำแนก มาเกี่ยวข้องในการวิเคราะห์ มีความเป็นเหตุเป็นผลในโลกของความเป็นจริงของแบบสอบชนิดเลือกตอบ ชาร์รอน เจน (Sharron Jane 1985: 408-A) ได้เปรียบเทียบวิธีการเลือกข้อกระทงของแบบสอบอิงเกณฑ์โดยวิธีคลาสสิค 3 วิธี คือ ใช้ดัชนีของคอกซ์ วากัส (Cox Vargus) พอยท์ ไบซีเรียล (Point Biscrial) และสัมประสิทธิ์ฟี (ϕ - Coefficiency) กับวิธีของทฤษฎีการตอบสนองข้อกระทง 3 พารามิเตอร์ ได้ข้อสรุปว่า โมเดลของทฤษฎีการตอบสนองข้อกระทง สามารถจำแนกผู้รอบรู้ได้ดีกว่าทฤษฎีคลาสสิคในการเลือกข้อกระทง

ส่วนวูดและคณะ (Wood and Others 1976: 5) ใช้รูปแบบการวิเคราะห์ข้อกระทงด้วยโลจิสติค 3 พารามิเตอร์ และได้ให้ข้อสนับสนุนว่า โมเดลโลจิสติค 3 พารามิเตอร์ ควรใช้กับกลุ่มตัวอย่างขนาดใหญ่ ประมาณ 1,000 คน และข้อสอบจำนวนมาก ประมาณ 40 ข้อ จึงจะได้ผลดี

2. การใช้แบบสอบอิงโดเมนในการกำหนดคะแนนจุดตัด

เนื่องจากงานวิจัยครั้งนี้ผู้วิจัยใช้แบบสอบวัดผลสัมฤทธิ์ลักษณะอิงโดเมนทำการศึกษา ซึ่งงานวิจัยที่ใช้การสร้างแบบสอบในลักษณะดังกล่าวมาศึกษาหาคะแนนจุดตัด คือ งานวิจัยของประเทือง ทาสีแสง (2527) ได้สร้างแบบทดสอบอิงโดเมนวิชาคณิตศาสตร์ เรื่องคู่ลำดับ ชั้นมัธยมศึกษาปีที่ 1 ลักษณะการสร้างข้อสอบเป็นการสร้างขึ้นตามรูปแบบเฉพาะประจำข้อสอบที่กำหนดขึ้น สร้างแบบสอบขึ้นมา 3 ฉบับ เป็นลักษณะคู่ขนาน หาคะแนนเกณฑ์ของแบบทดสอบด้วยวิธีให้ผู้เชี่ยวชาญพิจารณาเกณฑ์นับลดจาก 100 % กับวิธีของฮวิน หาคะแนนเฉลี่ยเป็นคะแนนเกณฑ์เหมาะสม ผลปรากฏว่าแบบทดสอบคู่ขนานชุดที่ 1 ชุดที่ 2 และ ชุดที่ 3 มีคะแนนเกณฑ์คือ 8, 13 และ 17 คะแนน ตามลำดับ ถือเป็นเปอร์เซนต์ได้ 80 % 87 % และ 70 % ตามลำดับ

กำจัด เกตุสุวรรณ (กำจัด เกตุสุวรรณ 2528: 126-132) ได้สร้างแบบสอบอิงโดเมนวิชาคณิตศาสตร์ เรื่อง ลำดับ ชั้นมัธยมศึกษาปีที่ 5 เป็นแบบสอบคู่ขนาน จำนวน 4 ชุด ๆ ละ 2 ฉบับ หาคะแนนเกณฑ์โดยใช้สูตร P max ของอวิน ผลปรากฏว่า แบบทดสอบคู่ขนานชุดที่ 1, 2, 3 และ 4 มีคะแนนเกณฑ์ที่เหมาะสมเป็นร้อยละ 70, 45, 44 และ 50 ตามลำดับ หรือมีคะแนนเป็น 7, 9, 11 และ 5 คะแนนตามลำดับ

3. การเปรียบเทียบผลของคะแนนจุดตัดโดยใช้วิธีตามแนวคลาสสิคกับวิธีของทฤษฎีการตอบสนองข้อกระทง

ผู้ที่ทำการศึกษางานวิจัยด้านนี้ คือ ชวลิต โพธิ์นคร (2528) ศึกษาคะแนนจุดตัดแบบสอบอิงเกณฑ์ตามแนวทฤษฎีคลาสสิค โดยวิธีของเบอร์ก วิธีกำหนดเกณฑ์ผ่านระดับต่ำสุด ส่วนวิธีหาคะแนนจุดตัดตามแนวทฤษฎีการตอบสนองข้อกระทงใช้วิธีประยุกต์ของราชส์โมเดล พบว่าจุดตัดของแบบสอบอิงเกณฑ์ที่คำนวณโดยวิธีประยุกต์ราชส์โมเดลในกลุ่มสูง มีแนวโน้มที่ต่างจากกลุ่มปานกลางและกลุ่มต่ำ ส่วนระหว่างกลุ่มต่ำและกลุ่มปานกลางมีแนวโน้มไม่แตกต่างกัน เมื่อคิดคำนวณเปอร์เซนต์ของจุดตัดเฉพาะข้อที่เป็นไปตามข้อตกลงของราชส์โมเดล ในแต่ละกลุ่มพบว่า คะแนนจุดตัดเป็น 50 % ของทุกกลุ่ม จากผลการวิจัยนี้ชี้ให้เห็นว่าถ้าสามารถคัดเลือกข้อกระทงที่มีความเหมาะสมกับระดับความสามารถของผู้สอบแต่ละกลุ่มตามข้อตกลงของราชส์โมเดลแล้ว จุดตัดที่ได้จากการคำนวณโดยวิธีประยุกต์ราชส์โมเดลแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลง ส่วนจุดตัดที่คำนวณโดยวิธีของเบอร์กในกลุ่มตัวอย่างที่มีความสามารถสูง ปานกลาง และต่ำ ได้คะแนนจุดตัดแตกต่างกัน ทั้งนี้เพราะว่าการกำหนดจุดตัดวิธีของเบอร์กจำเป็นต้องอาศัยข้อมูลจากกลุ่มตัวอย่าง โดยธรรมชาติแล้วลักษณะของกลุ่มตัวอย่างย่อมมีระดับความสามารถที่ต่างกัน จึงทำให้จุดตัดที่ได้แตกต่างกันดังกล่าวเช่นเดียวกับการวิจัยของเบฮูเนียค (Behuniak 1981: 3998A-3999A) ที่พบว่าการใช้กลุ่มตัวอย่างที่ต่างกันกำหนดจุดตัดวิธีเดียวกัน จะทำให้ได้จุดตัดที่ต่างกัน และจากผลการวิจัยครั้งนี้สรุปได้ว่า จุดตัดของเบอร์กมีแนวโน้มที่จะแปรเปลี่ยนไปตามระดับความสามารถของกลุ่มตัวอย่าง

4. การกำหนดความยาวของแบบสอบอิงเกณฑ์โดยใช้ทฤษฎีการตอบสนองข้อกระทง และการหาค่าความน่าจะเป็นในการจำแนกผิดพลาด

ผู้ที่ให้ความสนใจศึกษางานวิจัยแนวนี้ คือ แฮมเบิลตันและกรุจเตอร์ (Hambleton and De Gruijter 1983) เขาได้ศึกษาวิธีการกำหนดความยาวแบบสอบอิงเกณฑ์ที่สร้างขึ้นจากข้อสอบที่มีการกระจายของค่าความยาก และค่าอำนาจจำแนกต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อกระทง จากแนวคิดนี้สร้างความสัมพันธ์ระหว่างคะแนนโดเมน และความสามารถที่แท้จริง (θ) ได้ คือ

$$\pi = \frac{1}{m}\sum_{i=1}^{m} P_i(\theta)$$

π คือ คะแนนโดเมน

$P_i(\theta)$ คือ โอกาสที่ผู้สอบซึ่งมีความสามารถตรงระดับความสามารถ (θ) จะทำข้อสอบข้อที่ i ได้

m คือ จำนวนข้อสอบ

และสามารถหาระดับเกณฑ์มาตรฐาน หรือจุดตัด (π_o) จากความสัมพันธ์ ดังนี้

$$\pi_o = \frac{1}{m}\sum_{i=1}^{m} P_i(\theta_o)$$

ซึ่ง $P_i(\theta_o)$ เป็นโอกาสที่ผู้สอบที่มีความสามารถตรงระดับเกณฑ์มาตรฐานหรือจุดตัดที่อยู่ในรูปคะแนนโดเมนจะทำข้อสอบข้อที่ i ได้ ใช้วิธีของวิลคอกซ์ (Wilcox 1968) กำหนดช่วงขึ้นมาช่วงหนึ่งซึ่งเรียกว่า Difference Zone ที่กระจายในมาตรวัดคะแนนโดเมน (Domain Scale) (π_1 , π_u) โดยกำหนดให้ความน่าจะเป็นในการจำแนกผิดพลาดให้ผู้ที่ไม่รอบรู้ผ่านที่ π_1 ใช้สัญลักษณ์ P_1 และความน่าจะเป็นในการจำแนกผิดพลาดให้ผู้รอบรู้ไม่ผ่านที่ π_u ใช้สัญลักษณ์ P_u ไม่ควรเกิน 20 % ซึ่งค่าความน่าจะเป็นในการจำแนกผิดพลาดเมื่อความยาว n ข้อ คะแนนจุดตัด C สามารถคำนวณโดยวิธีไบโนเมียลโมเดล (Binomial Model) และคะแนนจุดตัดที่เหมาะสมสามารถหาได้จากการคำนวณ $n\pi_o$ ผลจากการศึกษาพบว่า