



ความเป็นมาและความสำคัญของปัญหา

คุณภาพของแบบสอบเป็นหัวใจของการวัดและประเมินผลทางการศึกษา คุณสมบัตินี้สำคัญที่สุดที่จะไปถึงคุณภาพของแบบสอบ คือ ความตรง ซึ่งความตรงเป็นคุณสมบัตินี้แสดงถึงความสามารถในการวัดคุณลักษณะแ่งหรือแบบสอบนั้นทำหน้าที่ได้ตามวัตถุประสงค์ที่กำหนดไว้ การตรวจสอบความตรงของข้อสอบสามารถทำได้หลายวิธี วิธีการหนึ่งคือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบสอบหรือข้อสอบที่ไม่ได้วัดเฉพาะคุณลักษณะแ่งเป้าหมายตามที่ต้องการวัดแต่วัดคุณลักษณะแ่งแทรกซ้อนที่ไม่ต้องการวัดของผู้สอบ แสดงว่าแบบสอบหรือข้อสอบนั้นขาดความตรง (Green, 1994; Mazor et al., 1995 อ้างถึงใน รัชรินทร์ มุกดา, 2540) เพราะหากผู้สอบกลุ่มย่อยกลุ่มใดมีคุณลักษณะแ่งแทรกซ้อนสูงกว่าย่อมมีโอกาสที่จะตอบข้อสอบได้ถูกต้องมากกว่า ทั้ง ๆ ที่มีคุณลักษณะแ่งเป้าหมายเท่ากับกลุ่มผู้สอบย่อยกลุ่มอื่น จึงทำให้เกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบย่อย ลักษณะเช่นนี้เดิมเรียกว่า "ข้อสอบลำเอียง" แต่ในระยะหลังเรียกว่า "การทำหน้าที่ต่างกันของข้อสอบ" (Differential Item Functioning : DIF) วิธีการใหม่ ๆ ที่ใช้ในการตรวจสอบความลำเอียงของข้อสอบนั้นจะเน้นไปที่ความแตกต่างระหว่างกลุ่มผู้สอบที่ตอบสนองต่างกันต่อข้อสอบข้อเดียวกัน ความแตกต่างที่เกิดขึ้นดังกล่าวอาจมาจากข้อคำถาม ประสบการณ์หรือพื้นฐานเดิมที่แตกต่างกันของกลุ่มผู้สอบ ซึ่งในบางสถานการณ์ก็ไม่เหมาะสมที่จะใช้คำว่า "ลำเอียง" ด้วยเหตุนี้จึงควรใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบหรือข้อสอบทำหน้าที่ต่างกัน เพราะเป็นคำที่มีความหมายกลาง ๆ มากกว่าและเหมาะสมกว่า (Holland and Thayer, 1988; Green, 1994 อ้างถึงใน จิตมา วรรณศรี, 2539) นอกจากนี้ Narayanan (1993 อ้างถึงใน รัชรินทร์ มุกดา, 2540) ได้เสนอแนวคิดว่าการวิเคราะห์ความลำเอียงของข้อสอบเป็นกรณีย่อยหนึ่งของการวิเคราะห์ DIF เพราะเกณฑ์ในการจับคู่เปรียบเทียบที่ใช้ตัดสินว่าข้อสอบมีความลำเอียง คือ ความตรง และเกณฑ์ที่ใช้ในการเปรียบเทียบว่าข้อสอบทำหน้าที่ต่างกันหรือไม่ นั้นเมื่ออยู่หลายเกณฑ์ เช่น เพศ เชื้อชาติ ภูมิฐานะ เป็นต้น

การศึกษาเรื่องความยุติธรรมในการสอบ มีการศึกษากันอย่างจริงจังในช่วงปลายทศวรรษ 1960 มีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการต่าง ๆ ปัจจุบันวิธีการต่าง ๆ เหล่านี้เป็นส่วนสำคัญในกระบวนการพัฒนาแบบสอบและการประเมินแบบสอบเช่นเดียวกับการวิเคราะห์ข้อสอบและการตรวจสอบความเที่ยงของแบบสอบ (Hambleton and Others, 1993) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีอยู่หลายวิธี ทั้งนี้เพราะ มีการศึกษาและคิดค้น

วิธีการต่าง ๆ เพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในเงื่อนไขต่าง ๆ ได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งแต่ละวิธีมีวิธีการพอสังเขปดังนี้

กลุ่มวิธีการตอบสนองข้อสอบ (Item Response Theory : IRT) เป็นวิธีที่วิเคราะห์ความแตกต่างของฟังก์ชันการตอบสนองข้อสอบ โดยเปรียบเทียบโค้งคุณลักษณะข้อสอบ (item characteristic curves : ICCs) ของกลุ่มผู้สอบตามระดับความสามารถของผู้สอบ ถ้าโค้งคุณลักษณะข้อสอบของกลุ่มผู้สอบสองกลุ่มมีรูปร่างเหมือนกันแสดงว่าข้อสอบข้อนั้นทำหน้าที่ไม่ต่างกัน แต่ถ้าโค้งคุณลักษณะข้อสอบของผู้สอบสองกลุ่มมีรูปร่างต่างกันแสดงว่าข้อสอบนั้นทำหน้าที่ต่างกัน กระบวนการวิเคราะห์ใช้สถิติพารามตริก และมีพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบเป็นดัชนีบอกระดับการทำหน้าที่ต่างกันของข้อสอบ

การวัดพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบจะมี 2 ลักษณะคือ 1) การวัดช่วงเปิด จะเป็นการวัดในช่วงความสามารถทั้งหมดระหว่างโค้งทั้งสอง ซึ่งจะทำได้พื้นที่ที่แน่นอน 2) การวัดช่วงปิด จะวัดในช่วงความสามารถตามที่คุณวิจัยต้องการศึกษา ซึ่งจะมีดัชนีที่ใช้บอกระดับการทำหน้าที่ต่างกันของข้อสอบได้แก่ พื้นที่ชนิดไม่มีเครื่องหมาย (IRT Z-UN) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบ พื้นที่ชนิดมีเครื่องหมาย (IRT Z-SA) จะเหมือนกับพื้นที่ชนิดไม่มีเครื่องหมาย แต่จะมีเครื่องหมายบอกให้คุณวิจัยทราบว่ากลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบเป็นกลุ่มที่ได้ประโยชน์หรือเสียประโยชน์แล้วจะทำการทดสอบนัยสำคัญของพื้นที่ชนิดไม่มีเครื่องหมายและพื้นที่ชนิดมีเครื่องหมายด้วยค่าสถิติ Z นอกจากนี้แล้ววิธี IRT ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโคสแควร์ของ LORD (IRT χ^2) ได้ด้วย ซึ่งเป็นการทดสอบนัยสำคัญของความแตกต่างของค่าอำนาจจำแนก (a) และค่าความยาก (b) ระหว่างกลุ่มในคราวเดียวกัน (Raju, 1990 อ้างถึงใน รัชนีทร์ มุกดา 2540 ; Camilli and Shepard, 1994) วิธี IRT เป็นวิธีที่ให้รายละเอียดมากและมีข้อดี คือค่าสถิติของข้อสอบไม่เปลี่ยนแปลงไปตามกลุ่มตัวอย่างและใช้การประมาณค่าความสามารถที่แท้จริงของผู้สอบแทนคะแนนที่สังเกตได้ แต่วิธีนี้มีข้อจำกัดคือ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ข้อมูลต้องมีลักษณะสอดคล้องกับข้อตกลงเบื้องต้น มีการคำนวณซับซ้อน แปลผลยากและเสียค่าใช้จ่ายสูง (Osterlind, 1993 ; Narayanan and Swaminatan, 1994 อ้างถึงใน สุรศักดิ์ อมรรัตนศักดิ์, 2531)

วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel : MH) เป็นวิธีที่พัฒนามาจากวิธีโคสแควร์แบบดั้งเดิม ใช้คะแนนรวมจากแบบสอบถามเป็นตัวแทนความสามารถ โดยวิเคราะห์ทีละระดับความสามารถ มีดัชนีบอกระดับการทำหน้าที่ต่างกันของข้อสอบและการทดสอบนัยสำคัญทางสถิติ ข้อดีของวิธีนี้ คือคำนวณง่าย ใช้ได้กับกลุ่มตัวอย่างขนาดเล็กและประหยัดค่าใช้จ่าย

วิธีถดถอยโลจิสติก (Logistic Regression : LR) เป็นวิธีที่เข้าใจธรรมชาติการทำหน้าที่ต่างกันภายในแบบสอบถามได้ดี เพราะมีโมเดลพื้นฐานทางสถิติจึงสามารถตรวจสอบการทำหน้าที่ต่างกัน

ของแบบสอบได้ถูกต้องใกล้เคียงกับวิธี IRT และวิธีนี้มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบเอกรูป (uniform DIF) ได้เท่าเทียมกับวิธีริบเทสท์ (Swaminatan and Roger, 1990 ; Ackerman, 1992) นอกจากนี้ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอนเนกรูป (non-uniform DIF) ได้ดีเช่นเดียวกัน

วิธีวิเคราะห์องค์ประกอบจำกัด (Restricted Factor Analysis : RFA) เป็นวิธีที่ใช้คะแนนรวมจากการสอบเป็นตัวแทนความสามารถ วิธีนี้จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตอบเป็นลักษณะต่อเนื่องหรือเป็นมาตรวัดระดับต่างๆ ได้ดีมาก และในแบบสอบที่มีการตอบแบบ 2 ค่า (ตอบถูกได้ 1 และตอบผิดได้ 0) วิธีนี้ก็ยังคงให้ผลการวิเคราะห์ที่ดีด้วย (Oort, 1998) และวิธีนี้ไม่มีข้อจำกัดในเรื่องขนาดของกลุ่มตัวอย่างเหมือนวิธี IRT ประหยัดเวลาและค่าใช้จ่ายในการวิเคราะห์

ที่ผ่านมาได้มีการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการต่างๆ เช่น ในประเทศไทย สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ทำการศึกษาเปรียบเทียบผลการตรวจสอบด้วยวิธีค่าความยากแปลง (Transformed Item Difficulty : TID) วิธี IRT 1 พารามิเตอร์ วิธี IRT 3 พารามิเตอร์และวิธีวิเคราะห์ความแปรปรวน ผลการวิจัยพบว่าวิธี IRT 3 พารามิเตอร์ พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด ต่อมา เพ็ญพนา สุธรรม (2539) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีค่าความยากแปลง วิธี IRT 3 พารามิเตอร์และวิธี MH ซึ่งพบว่าวิธี MH ตรวจสอบพบข้อสอบทำหน้าที่ต่างกันมากที่สุด รองลงมาคือวิธี IRT 3 พารามิเตอร์ โดยวิธี MH มีความสัมพันธ์กับวิธี IRT 3 พารามิเตอร์ สูงกว่าวิธีค่าความยากแปลงและในบีเดียวกัน จิตมา วรทนต์ (2539) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี MH และวิธี SIBTESTS พบว่าทั้งสองวิธี มีอัตราตรวจสอบเท่าเทียมกันและส่วนมากวิธี SIBTESTS มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธี MH เล็กน้อย นอกจากนี้ เสรี ชัดรัมย์ (2539) ได้ศึกษาเปรียบเทียบผลการตรวจสอบด้วยวิธี MH แบบปกติกับวิธี MH แบบแบ่งกลุ่มความสามารถและความยากของข้อสอบ พบว่า วิธี MH แบบแบ่งกลุ่มความสามารถของผู้สอบสามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแบบอนเนกรูปได้ดีกว่า ซึ่งผลการวิจัยครั้งนี้สอดคล้องกับผลการวิจัยของ Mazor และคณะ (1994) ที่พบว่าวิธี MH ตรวจสอบการทำหน้าที่ต่างกันแบบอนเนกรูปได้ดีโดยไม่ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้น เมื่อมีการแบ่งกลุ่มตัวอย่างออกเป็นกลุ่มสูงและกลุ่มต่ำ และ รัชรินทร์ มุกดา (2540) ทำการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี MH กับวิธีถดถอยโลจิสติกเมื่อมีการแบ่งกลุ่มความสามารถ ค่าอำนาจจำแนก ค่าความยากของข้อสอบ พบว่าโดยภาพรวมทั้งสองวิธีให้ผลการตรวจสอบไม่ต่างกัน ในกลุ่มผู้สอบที่มีความสามารถสูง ข้อสอบที่ตรวจพบเป็นข้อสอบที่มีค่าความยากสูง ค่าอำนาจจำแนกสูง ในกลุ่ม

ผู้สอบที่มีความสามารถปานกลาง หรือสอบที่ตรวจพบเป็นข้อสอบที่มีค่าความยากปานกลาง ค่าอำนาจจำแนกสูง ในกลุ่มผู้สอบที่มีความสามารถต่ำ หรือสอบที่ตรวจพบเป็นข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง นอกจากนี้ยังมีการศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบเอกกรุปโดย Rogers และ Swaminathan (1993) พบว่าวิธี MH มีอัตราความตรวจสอบต่ำกว่าวิธีถดถอยโลจิสติก แต่อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีถดถอยโลจิสติกมีค่าสูงกว่าเล็กน้อย ในด้านวิธี IRT นั้น Kim และ Cohen (1994) ศึกษาพบว่าวิธี IRT แบบ 2 พารามิเตอร์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี IRT แบบ 3 พารามิเตอร์ อีกทั้งวิธี IRT แบบ 3 พารามิเตอร์ มีความยุ่งยากซับซ้อนในภาววิเคราะหามากกว่า

จากการศึกษานานวิจัยข้างต้น พบว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกกรุป วิธี MH เป็นที่นิยมและใช้กันอย่างกว้างขวาง เพราะเป็นวิธีที่ไม่ยุ่งยาก ใช้กลุ่มตัวอย่างขนาดเล็ก คำนวณง่าย ผลการตรวจสอบการทำหน้าที่ต่างกันแบบเอกกรุปด้วยวิธี MH ให้ผลสอดคล้องกับวิธี IRT วิธีถดถอยโลจิสติก และวิธีซิมเพลสท์ (Hambleton, 1989 ; Roger and Swaminathan, 1993 ; Narayanan and Swaminathan, 1996 อ้างถึงใน จิตินา วรณศรี, 2539) ดังนั้น การศึกษาค้นคว้าวิจัยจึงเลือกใช้วิธี MH ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกกรุป ด้วยเหตุผลดังกล่าว

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้มีการพัฒนาขึ้นเรื่อย ๆ ปัจจุบัน วิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) ถือเป็นวิธีใหม่ล่าสุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งถูกพัฒนาขึ้นโดย Oort (1998) ซึ่ง Oort ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธี IRT 1 พารามิเตอร์และวิธี RFA โดยกำหนดให้ข้อสอบมีความลำเอียง 3 ขนาด คือข้อสอบที่มีความลำเอียงสูง ข้อสอบที่มีความลำเอียงต่ำและข้อสอบที่ไม่มีความลำเอียง นอกจากนี้ยังกำหนดขนาดกลุ่มตัวอย่างเป็นกลุ่มตัวอย่างขนาดเล็กและกลุ่มตัวอย่างขนาดใหญ่ ผลการวิจัยพบว่า วิธี RFA มีประสิทธิภาพในการทำหน้าที่ต่างกันของข้อสอบได้ใกล้เคียงกับวิธี IRT แบบ 1 พารามิเตอร์ แต่ในกรณีที่มีกลุ่มตัวอย่างมีขนาดเล็กวิธี RFA มีประสิทธิภาพดีกว่าในทุกเงื่อนไข นอกจากนี้วิธี RFA ยังเสียค่าใช้จ่ายในการคำนวณไม่มากนัก ประหยัดเวลาในการคำนวณและสามารถใช้ได้กับกลุ่มตัวอย่างขนาดไม่ใหญ่มาก

นอกจากจะมีการศึกษาวิธีการใหม่ ๆ ที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแล้ว ยังมีการศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบซึ่งพบว่าปัจจัยที่ส่งผลต่อประสิทธิภาพในการตรวจสอบคือ ขนาดกลุ่มตัวอย่าง ข้อค้นพบที่ได้นี้มีความสอดคล้องกัน กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้นอัตราความตรวจสอบพบความลำเอียงของข้อสอบจะเพิ่มขึ้น (Roger and Swaminathan, 1993) แต่ในกลุ่มตัวอย่างขนาดเล็กยังไม่ชัดเจนว่า ควรใช้ขนาดเท่าใดจึงจะเพียงพอ เกี่ยวกับขนาดของกลุ่มตัวอย่าง

Hill (1990 อ้างถึงใน Mazor et al., 1992) กล่าวว่า ขนาดกลุ่มตัวอย่างที่พอเหมาะสำหรับวิธี MH ควรใช้ระหว่าง 100 และ 300 คน สำหรับกลุ่มใดกลุ่มหนึ่งหรือทั้งสองกลุ่ม (reference and focal groups) Mazor และคณะ (1992) แนะนำว่าการใช้กลุ่มตัวอย่างขนาดกลุ่มละ 200 คนก็เพียงพอแล้วและไม่ควรน้อยกว่านี้ Narayanan และ Swaminathan (1994) แนะนำว่าโดยทั่วไปใช้ขนาดกลุ่มตัวอย่าง กลุ่มละ 300 คนก็เพียงพอที่จะตรวจสอบได้อย่างมีประสิทธิภาพ และพบว่าอัตราการตรวจสอบของวิธี MH และวิธี SIBTESTS จะได้รับผลกระทบจากกลุ่มตัวอย่างขนาดเล็กของกลุ่มเปรียบเทียบ (focal group) มากกว่ากลุ่มตัวอย่างของกลุ่มอ้างอิง (reference group) นอกจากนี้ กาญจนา วัฒนสุนทร (2538) เสนอว่ากลุ่มตัวอย่างอย่างน้อยที่สุดที่จะใช้ในการตรวจสอบโดยวิธี MH ควรมีขนาด 600 คนขึ้นไป Hambleton และคณะ (1993) เสนอแนะว่าในวิธี MH นั้นกลุ่มตัวอย่างที่เพียงพอในการวิเคราะห์ควรอยู่ระหว่าง 200 ถึง 1,000 คน แต่ถ้าใช้กลุ่มตัวอย่างขนาดใหญ่จะให้ผลการตรวจสอบที่ดีกว่า หากใช้กลุ่มตัวอย่างน้อยกว่า 200 คน ในกลุ่มใดกลุ่มหนึ่งอาจจะไม่เพียงพอสำหรับการวิเคราะห์ในบางเงื่อนไข จากปัญหาที่พบจึงกล่าวได้ว่าขนาดกลุ่มตัวอย่างน่าจะมีผลกระทบต่ออัตราการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาเกี่ยวกับขนาดกลุ่มตัวอย่างว่ามีผลต่ออำนาจการตรวจสอบ (power of the test) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทั้ง 3 วิธีคือ วิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ อย่างไรก็ตาม นอกจากปัจจัยด้านกลุ่มตัวอย่างแล้วยังมีปัจจัยด้านอื่น ๆ ที่มีผลกระทบต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งจากการศึกษาของ Kim และ Cohen (1994) Uttaro และ Millsap (1994) พบว่าความยาวของแบบสอบมีผลกระทบต่ออัตราการตรวจสอบ อีกทั้ง จิตติมา วรรณศรี (2539) พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH จะมีประสิทธิภาพดีที่สุดในเมื่อแบบสอบมีความยาว 60 ข้อ และจากงานวิจัยของ Narayanan and Swaminathan (1996) พบว่าความยาวของแบบสอบ 40 ข้อ เป็นความยาวของแบบสอบที่เพียงพอสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบซึ่งจะให้ผลการตรวจสอบใกล้เคียงกับความยาวของแบบสอบ 80 ข้อ ในขณะที่ Roger และ Swaminathan (1993) พบว่าความยาวของแบบสอบไม่มีผลกระทบต่ออัตราการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และ กาญจนา วัฒนสุนทร (2537) พบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบ ดังนั้น เพื่อการยืนยันข้อค้นพบให้มีความชัดเจนมากยิ่งขึ้น จึงต้องมีการศึกษาว่าความยาวของแบบสอบมีผลต่ออัตราการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างไรบ้าง

นอกจากจะมีการศึกษามลกระทบจากปัจจัยด้านขนาดความยาวของแบบสอบและขนาดของกลุ่มตัวอย่างแล้ว นักการศึกษายังมีการศึกษาปัจจัยอื่น ๆ ที่มีผลกระทบต่อประสิทธิภาพการ

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ปัจจัยดังกล่าวคือ ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบ เช่น จากการศึกษาของ Roger and Swaminathan (1993) พบว่า วิธี MH ตรวจสอบข้อสอบทำหน้าที่ต่างกันได้ดีเฉพาะกรณีที่มีข้อสอบมีความยากต่ำและข้อสอบที่มีความยากสูงเท่านั้น แต่ผลการศึกษาของ Narayanan and Swaminathan (1996) พบว่า ส่วนใหญ่วิธี MH ตรวจสอบข้อสอบทำหน้าที่ต่างกันในกลุ่มข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง และจากผลการศึกษาในประเทศไทยของ เสรี ชัดแจ้ง (2540) พบว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปส่วนใหญ่เป็นข้อสอบที่มีความยากปานกลางและข้อสอบที่มีความยากต่ำ ดังนั้น เพื่อเป็นการยืนยันข้อค้นพบให้มีความชัดเจนมากขึ้นจึงต้องมีการศึกษาลักษณะที่เกิดจากค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบ ในข้อสอบแบบเอกรูปว่ามีผลกระทบต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีที่นำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาคั้งนี้หรือไม่

จากการศึกษางานวิจัยที่ผ่านมา จึงมีประเด็นปัญหาที่น่าสนใจ คือ

1. ยังไม่มีการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปด้วยวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ ว่าวิธีใดมีประสิทธิภาพดีกว่า
2. จากการศึกษาพบว่าปัจจัยที่ส่งผลกระทบต่ออัตราตรวจสอบคือ ขนาดของกลุ่มตัวอย่าง ดังนั้นน่าจะทำการศึกษารวามูลของขนาดกลุ่มตัวอย่างมีผลกระทบต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปด้วยวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ หรือไม่ อย่างไร
3. ความยาวของแบบสอบมีผลต่อความเที่ยงของคะแนนจากแบบสอบ ซึ่งอาจจะส่งผลกระทบต่อความคลาดเคลื่อนประเภทที่ 1 ดังนั้นความยาวของแบบสอบจะมีผลต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ หรือไม่
4. ภายได้ข้อสอบมีค่าอำนาจจำแนกและค่าความยากต่างกัน วิธี RFA วิธี MH และวิธี IRT มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปต่างกันหรือไม่ อย่างไร

จากประเด็นปัญหาข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เรล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์

วัตถุประสงค์ในการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการตรวจสอบ (power of the test) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป 3 วิธีคือ วิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เชล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ เมื่อขนาดของกลุ่มตัวอย่าง ค่าความยากของข้อสอบ อำนาจจำแนกของข้อสอบ และความยาวของแบบสอบต่างกัน

สมมุติฐานทางการวิจัย

จากการศึกษาของ Oort (1998) พบว่า ในข้อสอบแบบแบ่งเป็นสองค่า (ตอบถูกได้ 1 ตอบผิดได้ 0) เมื่อกำหนดเงื่อนไขให้ข้อสอบมีความลำเอียงสูง ข้อสอบมีความลำเอียงปานกลาง ข้อสอบมีความลำเอียงต่ำ กลุ่มตัวอย่างมีขนาดต่างกัน (กลุ่มเล็ก 200 คน, กลุ่มใหญ่ 2,000 คน) และค่าเฉลี่ยคุณลักษณะแฝงต่างกัน ในขณะที่ความยาวของแบบสอบเท่ากัน (40 ข้อ) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT แบบ 1 พารามิเตอร์และวิธี RFA ในภาพรวมผลการตรวจสอบด้วยวิธี RFA ให้ผลการตรวจสอบที่ดีกว่า กล่าวคือในกรณีข้อสอบมีความลำเอียงปานกลางและข้อสอบที่มีความลำเอียงสูง วิธี RFA ให้ผลการตรวจสอบที่ดีกว่า วิธี IRT แบบ 1 พารามิเตอร์และในกลุ่มตัวอย่างขนาดเล็กการตรวจสอบด้วยวิธี RFA ให้ผลการตรวจสอบที่ดีกว่า แต่ทั้งสองวิธีจะให้ผลการตรวจสอบดีขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่

จากงานวิจัยเกี่ยวกับการศึกษาเปรียบเทียบวิธีการต่างๆ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า วิธี IRT เป็นวิธีที่ดี แต่มีข้อจำกัดในเรื่องกลุ่มตัวอย่างต้องมีความใหญ่ และข้อมูลต้องมีลักษณะเป็นไปตามข้อตกลงเบื้องต้น การคำนวณซับซ้อนสิ้นเปลืองทั้งเวลาและค่าใช้จ่าย (Osterlind, 1993) และจากการศึกษาของ Narayanan Swaminatan (1996) พบว่า วิธี MH จะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในกรณีที่ข้อสอบมีค่าความยากต่ำและมีค่าอำนาจจำแนกสูงได้ดี นอกจากนี้ เสรี ชัดเข้ม (2539) พบว่าข้อสอบส่วนใหญ่ที่ตรวจพบโดยวิธี MH นั้นเป็นข้อสอบที่มีค่าความยากปานกลางและเป็นข้อสอบที่ง่าย ส่วนในด้านขนาดกลุ่มตัวอย่างสำหรับวิเคราะห์ด้วยวิธี MH นั้น Hill (1990) อ้างถึงใน Mazor และคณะ (1992) กล่าวว่าขนาดกลุ่มตัวอย่างที่พอเหมาะสำหรับวิเคราะห์ด้วยวิธี MH ควรอยู่ระหว่าง 100 และ 300 คนสำหรับกลุ่มใดกลุ่มหนึ่งหรือทั้งสองกลุ่ม Mazor และคณะ (1992) แนะนำว่าการใช้กลุ่มตัวอย่างขนาด 200 คนก็เพียงพอสำหรับการวิเคราะห์ด้วยวิธี MH ดังนั้น ผู้วิจัยจึงมีสมมุติฐานดังนี้

1. เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบต่างกัน ภายใต้ความยาวแบบสอบเดียวกัน วิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ มีอำนาจการตรวจสอบและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่ต่างกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2. เมื่อกลุ่มตัวอย่างมีขนาดเล็ก ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบต่างกัน ภายใต้ความยาวแบบสอบเดียวกันวิธี MH จะมีอำนาจการตรวจสอบสูงกว่าวิธี RFA และวิธี IRT แบบ 2 พารามิเตอร์ และมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี RFA และวิธี IRT แบบ 2 พารามิเตอร์ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธีการวิเคราะห์องค์ประกอบ (RFA) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์

2. ตัวแปรที่ศึกษาในครั้งนี้

2.1 ตัวแปรอิสระมีทั้งหมด 6 ตัว คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ค่าความยาก ขนาดกลุ่มตัวอย่าง ความลำเอียงของข้อสอบ ความยาวของแบบสอบ และค่าอำนาจจำแนก ดังมีรายละเอียดต่อไปนี้

2.1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ โดยใช้ค่าสถิติ IRT (Z-SA) IRT (Z-UN) และ IRT (χ^2)

2.1.2 ค่าความยากของข้อสอบมี 3 ระดับ คือ ข้อสอบที่มีค่าความยากสูง (ค่าความยากอยู่ระหว่าง 0.67 ถึง 2.5) ปานกลาง (ค่าความยากอยู่ระหว่าง -0.67 ถึง 0.67) และต่ำ (ค่าความยากอยู่ระหว่าง -2.5 ถึง -0.67) (คณิต ไข่มุกด์, 2534)

2.1.3 ขนาดกลุ่มตัวอย่างที่ศึกษามี 2 ขนาดคือ กลุ่มตัวอย่างขนาดเล็ก (300 คน) และกลุ่มตัวอย่างขนาดใหญ่ (1000 คน)

2.1.4 ความลำเอียงของข้อสอบแบ่งเป็น 2 ขนาดคือ ข้อสอบที่มีความลำเอียงสูง (0.8 SD) และข้อสอบที่มีความลำเอียงต่ำ (0.3 SD)

2.1.5 ความยาวของแบบสอบมี 2 ขนาดคือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ)

2.1.6 ค่าอำนาจจำแนกของข้อสอบแบ่งออกเป็น 3 ระดับ คือ ค่าอำนาจจำแนกสูง (ค่าอำนาจจำแนกอยู่ระหว่าง 0.85 ถึง 2.5) ปานกลาง (ค่าอำนาจจำแนกอยู่ระหว่าง 0.55 ถึง 0.85) และต่ำ (ค่าอำนาจจำแนกอยู่ระหว่าง 0.50 ถึง 0.55)

2.2 ตัวแปรตาม คือ ประสิทธิภาพในการตรวจสอบ โดยพิจารณาจาก

2.2.1 อำนาจการตรวจสอบ (Power of the test)

2.2.2 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

3. ข้อมูลที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลที่จำลองขึ้นด้วยโปรแกรม SPSS for Windows

4. เกณฑ์ที่ใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

4.1 วิธี RFA ใช้การทดสอบนัยสำคัญทางสถิติของค่าดัชนี MI และดัชนี EPC ว่าแตกต่างจาก 0 หรือไม่ที่ระดับนัยสำคัญ .05

4.2 วิธีแมนเทิล-แฮนส์เรล ได้แก่ค่า α_{MH} แตกต่างจาก 1 หรือค่า Δ_{MH} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4.3 วิธี IRT ได้แก่ค่า การทดสอบนัยสำคัญของพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบด้วยสถิติ Z และโคแอสต์รของ Lord ที่ระดับ .05

5. การศึกษาครั้งนี้เป็นการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (uniform DIF) ทั้งนี้เพราะผู้วิจัยกำหนดให้ผู้สอบที่อยู่ในกลุ่มข้างจึงมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบที่อยู่ในกลุ่มเปรียบเทียบในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างโค้งคุณลักษณะข้อสอบในทุกระดับความสามารถ

ข้อตกลงเบื้องต้น

การวิจัยในครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพของวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เรล (MH) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยข้อมูลที่ใช้ในการศึกษาครั้งนี้ผู้วิจัยใช้ข้อมูลที่จำลองขึ้นด้วยโปรแกรม SPSS for Windows ซึ่งข้อมูลดังกล่าวมีลักษณะการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

ความยาวของแบบสอบที่ศึกษามี 2 ขนาดคือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ) ในแบบสอบสั้น (25 ข้อ) จะมีข้อสอบที่มีความลำเอียงต่ำ 5 ข้อและมีข้อสอบที่มีความลำเอียง

สูง 5 ข้อ ในแบบสอบยาว (75 ข้อ) จะมีข้อสอบที่มีความลำเอียงต่ำ 5 ข้อ และมีข้อสอบที่มีความลำเอียงสูง 10 ข้อ

ค่าจำกัดความที่ใช้ในการวิจัย

ข้อสอบทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ทำให้ผู้สอบที่มีความสามารถที่ต้องการวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน ทั้งนี้เนื่องจากผู้สอบอยู่ในกลุ่มย่อยต่างกัน ในที่นี้จะแทนของผู้สอบจำลองด้วยโปรแกรม SPSS for Windows ดังนั้น ข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ทำให้ผู้สอบที่มีความสามารถเท่ากันแต่อยู่ในกลุ่มต่างกันมี โอกาสตอบถูกไม่เท่ากัน ซึ่งตรวจสอบได้ด้วยวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์

ขนาดของความลำเอียงของข้อสอบ หมายถึง ค่าการกระจายของคะแนนข้อสอบรายข้อที่อยู่ในรูปส่วนเบี่ยงเบนมาตรฐานที่ได้จากการจำลองข้อมูลด้วยโปรแกรม SPSS for Windows ซึ่งในการศึกษาค้างนี้แบ่งระดับความลำเอียงของข้อสอบออกเป็น 2 ระดับคือ กลุ่มข้อสอบที่มีความลำเอียงสูง (0.8 SD) และกลุ่มข้อสอบที่มีความลำเอียงต่ำ (0.3 SD)

ข้อสอบทำหน้าที่ต่างกันแบบเอกกรุป (uniform DIF) หมายถึงข้อสอบที่ให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างโค้งคุณลักษณะข้อสอบในทุกระดับความสามารถ

ข้อสอบทำหน้าที่ต่างกันแบบอนเอกกรุป (non-uniform DIF) หมายถึง ข้อสอบที่ให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มไม่สม่ำเสมอในทุกระดับความสามารถเมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่า ในบางระดับความสามารถจะมีปฏิสัมพันธ์กันระหว่างโค้งคุณลักษณะ ข้อสอบ

เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีต่าง ๆ หมายถึง วิธีใดมีอำนาจการตรวจสอบ (Power of the test) สูงและมีความคลาดเคลื่อนประเภทที่ 1 ต่ำ แสดงว่าวิธีนั้นมีประสิทธิภาพสูงในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ค่าความยากของข้อสอบ (b) หมายถึง ระดับความยากของข้อสอบซึ่งมีค่าเท่ากับระดับความสามารถของผู้สอบที่มีโอกาสทำข้อสอบข้อนั้นถูก ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าความยากอยู่ระหว่าง -2.5 ถึง 2.5 ในที่นี้ได้แปลงค่าความยากของข้อสอบเป็น 3 ระดับคือ ข้อสอบที่มีค่าความยากสูง (ค่าความยากอยู่ระหว่าง 0.67 ถึง 2.5) ข้อสอบที่มีค่าความ

ยากปานกลาง (ค่าความยากอยู่ระหว่าง -0.67 ถึง 0.67) และข้อสอบที่มีค่าความยากต่ำ (ค่าความยากอยู่ระหว่าง -2.5 ถึง -0.67) (คณิต ไข่มุกด์, 2534)

ค่าอำนาจจำแนก (α) หมายถึง ค่าความชันของโค้งคุณลักษณะข้อสอบที่จุดเปลี่ยนโค้ง ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าอำนาจจำแนกอยู่ระหว่าง 0.5 ถึง 2.5 ในที่นี้ ผู้วิจัยแบ่งค่าอำนาจจำแนกของข้อสอบออกเป็น 3 ระดับ คือ กลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง (ค่าอำนาจจำแนกอยู่ระหว่าง 0.85 ถึง 2.5) กลุ่มข้อสอบที่มีค่าอำนาจจำแนกปานกลาง (ค่าอำนาจจำแนกอยู่ระหว่าง 0.55 ถึง 0.85) และกลุ่มข้อสอบที่มีค่าอำนาจจำแนกต่ำ (ค่าอำนาจจำแนกอยู่ระหว่าง 0.50 ถึง 0.55) (คณิต ไข่มุกด์, 2534)

วิธี RFA หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด

วิธี MH หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เรล

วิธี IRT หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ที่พิจารณาจากความแตกต่างของฟังก์ชันการตอบข้อสอบ ระหว่างผู้สอบต่างกลุ่มที่ได้จากการตอบข้อสอบข้อเดียวกัน

วิธี IRT (Z-SA) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ที่ใช้ดัชนีแบบมีเครื่องหมายในการพิจารณาความแตกต่างของฟังก์ชันการตอบข้อสอบ ระหว่างผู้สอบต่างกลุ่มที่ได้จากการตอบข้อสอบข้อเดียวกัน

วิธี IRT (Z-UN) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ที่ใช้ดัชนีแบบไม่มีเครื่องหมายในการพิจารณาความแตกต่างของฟังก์ชันการตอบข้อสอบ ระหว่างผู้สอบต่างกลุ่มที่ได้จากการตอบข้อสอบข้อเดียวกัน

วิธี IRT (χ^2) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ที่ใช้ดัชนีค่าไคแควร์ของ LORD ในการพิจารณาความแตกต่างของฟังก์ชันการตอบข้อสอบ ระหว่างผู้สอบต่างกลุ่มที่ได้จากการตอบข้อสอบข้อเดียวกัน

เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันตามวิธี RFA หมายถึง ข้อสอบที่มีค่าดัชนี MI และดัชนี EPC แตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ .05

เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันตามวิธี MH หมายถึง ข้อสอบที่มีค่า α_{MH} แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หรือมีค่า Δ_{MH} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันตามวิธี IRT หมายถึง ข้อสอบที่มีพื้นที่อันเกิดจากความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่มซึ่งทดสอบนัยสำคัญด้วยสถิติ Z และ ไคแควร์ของ Lord ที่ระดับ .05

เกณฑ์การตัดสินข้อสอบทำหน้าที่ต่างกันแบบเอกรูป หมายถึง สิ่งที่ใช้เป็นหลักในการพิจารณาตัดสินข้อสอบทำหน้าที่ต่างกันแบบเอกรูป ในการวิจัยครั้งนี้ใช้วิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ เป็นหลักในการพิจารณาตัดสินข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

อำนาจการตรวจสอบ (Power of the test) หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้อง โดยคำนวณจากจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้องต่อจำนวนข้อที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ
อัตราความคลาดเคลื่อนของการตรวจสอบ หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ระบุผิดพลาดซึ่งมี 2 ประเภท คือ อัตราความคลาดเคลื่อนประเภทที่ 1 และอัตราความคลาดเคลื่อนประเภทที่ 2

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) คือ อัตราของการระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (False positive) ทั้งที่ความเป็นจริงข้อสอบนั้นทำหน้าที่ไม่ต่างกัน ซึ่งในที่นี้คำนวณได้จากค่าอัตราส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบ

อัตราความคลาดเคลื่อนประเภทที่ 2 (Type II error rate) คือ อัตราของการระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกัน (False negative) ทั้งที่ความเป็นจริงข้อสอบนั้นทำหน้าที่ต่างกัน ซึ่งในที่นี้คำนวณได้จากค่าอัตราส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ

ขนาดกลุ่มตัวอย่าง หมายถึง จำนวนผู้สอบแต่ละขนาดที่ใช้ในการศึกษา ซึ่งการศึกษครั้งนี้กลุ่มตัวอย่างที่ศึกษามี 2 ขนาดคือ กลุ่มตัวอย่างขนาดเล็ก (300 คน) และกลุ่มตัวอย่างขนาดใหญ่ (1,000 คน) โดยแต่ละขนาดแบ่งออกเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในอัตราส่วน 1:1

ความยาวแบบสอบ หมายถึง จำนวนข้อสอบในแบบสอบ ในการวิจัยครั้งนี้ศึกษาความยาวแบบสอบ 2 ขนาด คือ แบบสอบขนาดสั้น (25 ข้อ) และแบบสอบขนาดยาว (75 ข้อ)

กลุ่มอ้างอิง (reference group) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถเท่ากัน ในที่นี้กลุ่มอ้างอิงได้จากการสุ่มผู้สอบจากผู้สอบที่จำลองขึ้นด้วยโปรแกรม SPSS for Windows

กลุ่มเปรียบเทียบ (focal group) หมายถึง กลุ่มผู้ตอบที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ กลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้ตอบอีกกลุ่มหนึ่ง ทั้ง ๆ ที่มีความสามารถเท่ากัน ในที่นี้กลุ่มเปรียบเทียบได้จากการสุ่มผู้ตอบจากผู้ตอบที่จำลองขึ้นด้วยโปรแกรม SPSS for Windows

ประโยชน์ที่จะได้รับ

ผลการวิจัยจะช่วยให้ทราบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธีให้ผลการตรวจสอบสอดคล้องหรือแตกต่างกันอย่างไร หากผลการวิจัยพบว่า วิธีการต่าง ๆ ดังกล่าวมีความสอดคล้องกันมาก ก็สามารถเลือกใช้วิธีที่ง่าย สะดวกและประหยัดในการตรวจสอบ แต่หากผลการวิจัย พบว่าวิธีการตรวจสอบดังกล่าวมีความสอดคล้องกันน้อย ก็สามารถเลือกใช้วิธีการตรวจสอบที่ให้ผลการวิจัยที่ชัดเจนกว่า โดยสรุปเป็นข้อ ๆ ได้ดังนี้

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธีให้ค่าอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 แตกต่างกันอย่างไ
2. เป็นแนวทางในการตัดสินใจเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างเหมาะสม
3. เป็นแนวทางในการเลือกใช้นา้ขนาดกลุ่มตัวอย่างและความยาวแบบสอบ อันจะทำให้ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุด