

สรุป อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ของวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการทดสอบของข้อสอบ (IRT) แบบ 2 พารามิเตอร์โดยมีวัตถุประสงค์ดังนี้

1. เพื่อเปรียบเทียบอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการทดสอบของข้อสอบ (IRT) แบบ 2 พารามิเตอร์ โดยเทียบกับเกณฑ์ที่กำหนดเมื่อขนาดกลุ่มตัวอย่าง ลักษณะคุณภาพและความยาวของข้อสอบต่างกัน

ข้อมูลที่ใช้ในการศึกษาได้จากการจำลองขึ้นด้วยโปรแกรม SPSS for Windows โดยกำหนดกลุ่มตัวอย่างออกเป็น 2 กลุ่ม คือ กลุ่มตัวอย่างขนาดเล็ก 300 คน และกลุ่มตัวอย่างขนาดใหญ่ 1,000 คน ส่วนคุณภาพของแบบสอบผู้วิจัยได้กำหนด ความยาวของข้อสอบแบ่งออกเป็น 2 ขนาด คือ ขนาดความยาว 25 ข้อ และขนาดความยาว 75 ข้อ ในแต่ละแบบสอบมีการกำหนดขนาดของความลำเอียงของข้อสอบออกเป็น 2 ระดับคือ ข้อสอบที่มีความลำเอียงสูงและข้อสอบที่มีความลำเอียงต่ำ ค่าความยากของข้อสอบแบ่งออกเป็น 3 ระดับ คือ กลุ่มข้อสอบที่มีค่าความยากสูง กลุ่มข้อสอบที่มีค่าความยากปานกลาง กลุ่มข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกของข้อสอบแบ่งออกเป็น 3 ระดับ คือ กลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง กลุ่มข้อสอบที่มีค่าอำนาจจำแนกปานกลาง กลุ่มข้อสอบที่มีค่าอำนาจจำแนกต่ำ รวมเงื่อนไขที่ศึกษาทั้งหมด 36 เงื่อนไข ภายในเงื่อนไขเดียวกันที่จำลองขึ้นนั้นใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธี RFA วิธี MH และวิธี IRT

กลุ่มตัวอย่างที่ใช้วิเคราะห์ในแต่ละเงื่อนไขได้จากการสุ่มจำนวนผู้สอบด้วยวิธีสุ่มอย่างง่ายจากข้อมูลที่ได้จำลองขึ้น โดยสุ่มจำนวนผู้สอบกลุ่มข้างอิงและกลุ่มเปรียบเทียบมีสัดส่วนเท่ากันคือ 1:1 ซึ่งมีขนาดกลุ่มตัวอย่างกลุ่มละ 150 คน ในเงื่อนไขที่ 1 ถึงเงื่อนไขที่ 18 ส่วนเงื่อนไขที่ 19 ถึงเงื่อนไขที่ 36 ใช้กลุ่มตัวอย่างกลุ่มละ 500 คน

ในการวิจัยครั้งนี้ทำการวิเคราะห์ข้อมูลดังนี้

1. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการตอบสนองข้อสอบ (TRT)
 - 1.1 วิเคราะห์องค์ประกอบของแบบสอบในแต่ละเงื่อนไขเพื่อหาความเป็นเอกมิติของแบบสอบ ซึ่งเป็นข้อตกลงเบื้องต้นของวิธี IRT โดยใช้โปรแกรม SPSS for Windows
 - 1.2 ประมาณค่าพารามิเตอร์ และค่าความแปรปรวนร่วมของข้อมูลโดยใช้โปรแกรม BILOG 3.04 ซึ่งข้อมูลแยกกันในแต่ละพื้นที่ระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ
 - 1.3 เปรียบมาตรฐานของค่าพารามิเตอร์ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยใช้โปรแกรม EQUATE 2.0
 - 1.4 คำนวณพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบ และทดสอบค่านัยสำคัญด้วยสถิติ Z ที่ระดับ 0.05 โดยใช้โปรแกรม AREA
2. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนด์เรล โดยใช้โปรแกรมสำเร็จรูป MH-DIF เพื่อคำนวณค่าดัชนี α_{MH} และดัชนี Δ_{MH} ทดสอบนัยสำคัญด้วยไคสแควร์
3. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด
 - 3.1 เตรียมข้อมูลในรูปแบบเมตริกสหสัมพันธ์โดยใช้โปรแกรม PELIS 2 ซึ่งเป็นโปรแกรมย่อยในโปรแกรม LISREL
 - 3.2 วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยโปรแกรม LISREL เพื่อค่าดัชนี MI และดัชนีค่า EPC ที่แตกต่างจาก 0 ที่ระดับนัยสำคัญทางสถิติ .05 ตัดข้อสอบที่มีค่า MI สูงที่สุดออกไปทีละข้อจนกระทั่งไม่มีข้อสอบที่มีค่าดัชนี MI และดัชนีค่า EPC แตกต่างจาก 0 อยู่ในแบบสอบ
4. คำนวณอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทิล-แฮนด์เรล (MH) และ วิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์

สรุปผลการวิจัย

ผลการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี RFA วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ เมื่อจัดกลุ่มความยาวของแบบสอบ ขนาดกลุ่มตัวอย่างและลักษณะของแบบสอบที่ต่างกัน ผลการวิจัยสรุปได้ดังนี้

1. เมื่อเปรียบเทียบอำนาจการตรวจสอบ ระหว่างวิธี IRT ทั้ง 3 วิธีคือ การตรวจสอบด้วยวิธี IRT ที่ใช้ค่าดัชนีในการพิจารณาการทำหน้าที่ต่างกันของข้อสอบด้วยค่าพื้นที่แบบคิดเครื่องหมาย (Z-SA) ค่าพื้นที่ไม่คิดเครื่องหมาย IRT(Z-UN) และการทดสอบไคแอสควร์ของ Lord IRT(χ^2) พบว่า โดยภาพรวมการตรวจสอบด้วยดัชนีค่าพื้นที่ไม่คิดเครื่องหมาย (Z-UN) มีอำนาจการตรวจสอบสูงกว่า การใช้ดัชนีค่าพื้นที่ที่มีเครื่องหมาย (Z-SA) และการทดสอบไคแอสควร์ของ Lord (χ^2) ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างวิธีทั้ง 3 วิธี พบว่าวิธี IRT(Z-UN) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี IRT(Z-SA) และวิธี IRT (χ^2) ตามลำดับ

2. สำหรับกลุ่มข้อสอบที่มีความยากสูง โดยภาพรวมข้อมูลไม่เหมาะสำหรับการตรวจสอบด้วยวิธีทั้ง 5 เพราะเมื่อนำข้อมูลผลการตอบข้อสอบของกลุ่มตัวอย่างไปวิเคราะห์องค์ประกอบเพื่อทดสอบความเป็นเอกมิติ ซึ่งเป็นข้อตกลงเบื้องต้นของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT แบบ 2 พารามิเตอร์ พบว่า ข้อมูลส่วนใหญ่ไม่มีความเป็นเอกมิติ วิธี RFA ก็ไม่สามารถตรวจสอบได้เช่นกัน เนื่องจากข้อมูลไม่มีความผันแปร (variation) มากพอ ส่วนวิธี MH แม้จะสามารถใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในเงื่อนไขนี้ได้ แต่พบว่าจะให้อำนาจการตรวจสอบที่ต่ำมาก

3. สำหรับกลุ่มข้อสอบที่มีความยากปานกลาง เมื่อเปรียบเทียบอำนาจการตรวจสอบระหว่างวิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือวิธี MH และวิธี IRT ทั้ง 3 วิธี ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT(Z-UN) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า วิธี MH และวิธี RFA ตามลำดับ

4. สำหรับกลุ่มข้อสอบที่มีความยากต่ำ เมื่อเปรียบเทียบอำนาจการตรวจสอบระหว่างวิธี RFA วิธี MH และวิธี IRT(Z-UN) พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือวิธี MH และวิธี IRT(Z-UN) ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT(Z-UN) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า วิธี MH และวิธี RFA ตามลำดับ

สรุป เมื่อพิจารณาจากความยากทั้ง 3 ระดับ วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธี MH และวิธี IRT ทั้ง 3 วิธี ตามลำดับ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 นั้นวิธี IRT ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี MH และวิธี RFA ตามลำดับ

5. สำหรับกลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง เมื่อเปรียบเทียบอำนาจการตรวจสอบระหว่างวิธี RFA วิธี MH และวิธี IRT(Z-UN) พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือวิธี MH และวิธี IRT(Z-UN) ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT(Z-UN) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า วิธี MH และวิธี RFA ตามลำดับ

6. สำหรับกลุ่มข้อสอบที่มีค่าอำนาจจำแนกปานกลาง เมื่อเปรียบเทียบอำนาจการตรวจสอบระหว่างวิธี RFA วิธี MH และวิธี IRT(Z-UN) พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือวิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT(Z-UN) และวิธี MH มีอัตราความคลาดเคลื่อนประเภทที่ 1 พอ ๆ กัน ในขณะที่วิธี RFA มีอัตราความคลาดเคลื่อนต่ำที่สุด

7. สำหรับกลุ่มข้อสอบที่มีค่าอำนาจจำแนกต่ำ เมื่อเปรียบเทียบอำนาจการตรวจสอบระหว่างวิธี RFA วิธี MH และวิธี IRT(Z-UN) พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือวิธี MH และวิธี IRT(Z-UN) ตามลำดับ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT(Z-UN) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า วิธี MH และวิธี RFA ตามลำดับ

สรุป เมื่อพิจารณาจากค่าอำนาจจำแนกทั้ง 3 ระดับ วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธี MH และวิธี IRT ทั้ง 3 วิธี ตามลำดับ ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 นั้นวิธี IRT ทั้ง 3 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี MH และวิธี RFA ตามลำดับ

8. ผลการเปรียบเทียบอำนาจการตรวจสอบ เมื่อเพิ่มขนาดความยาวของแบบสอบจาก 25 ข้อเป็นแบบสอบที่มีความยาวของแบบสอบ 75 ข้อ พบว่า วิธี MH และวิธี IRT ทั้ง 3 วิธีจะมีอำนาจการตรวจสอบลดลงเมื่อกลุ่มตัวอย่างขนาดเล็ก แต่ในกลุ่มตัวอย่างขนาดใหญ่ วิธี MH และวิธี IRT ทั้ง 3 วิธี จะมีอำนาจการตรวจสอบสูงขึ้น ในขณะที่อำนาจการตรวจสอบของวิธี RFA ยังคงมีอำนาจการตรวจสอบสูงที่สุดตลอดทุกเงื่อนไข และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี MH และวิธี IRT ทั้ง 3 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 ลดลงด้วย ในขณะที่วิธี RFA ยังคงมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดตลอดทุกเงื่อนไข

9. ผลการเปรียบเทียบอำนาจการตรวจสอบ เมื่อเพิ่มขนาดกลุ่มตัวอย่างจาก 300 คน เป็นขนาดกลุ่มตัวอย่าง 1000 คน ในทุกเงื่อนไขของแบบสอบ พบว่า วิธี MH และวิธี IRT ทั้ง 3 วิธี มีอำนาจการตรวจสอบของเพิ่มขึ้น ในขณะที่อำนาจการตรวจสอบของวิธี RFA ยังมีอำนาจการตรวจสอบสูงที่สุด และเมื่อพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี MH และวิธี IRT ทั้ง 3 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้น ในขณะที่วิธี RFA ยังมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุด แต่โดยภาพรวมสรุปได้ว่า เมื่อกลุ่มตัวอย่างมากขึ้นอำนาจการตรวจสอบของวิธี MH และวิธี IRT ทั้ง 3 วิธี จะสูงขึ้นด้วย

10. ผลการเปรียบเทียบอำนาจการตรวจสอบในกลุ่มข้อสอบที่มีความลำเอียงสูงกับกลุ่มข้อสอบที่มีความลำเอียงต่ำ พบว่า โดยภาพรวมข้อสอบที่มีความลำเอียงสูงวิธีการตรวจสอบทั้ง 5 วิธี คือ วิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี จะมีอำนาจการตรวจสอบสูงกว่าการในกลุ่มข้อสอบที่มีความลำเอียงต่ำ

แต่เมื่อแยกเปรียบเทียบอำนาจการตรวจสอบเฉพาะในกลุ่มข้อสอบที่มีความลำเอียงสูงด้วยวิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธี IRT ทั้ง 3 วิธี และวิธี MH ตามลำดับ

เมื่อเปรียบเทียบอำนาจการตรวจสอบในกลุ่มข้อสอบที่มีความลำเอียงต่ำด้วยวิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี พบว่า วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธี MH และวิธี IRT ทั้ง 3 วิธี ตามลำดับ

กล่าวโดยสรุป เมื่อพิจารณาที่ค่าอำนาจการตรวจสอบของแต่ละวิธี จะพบว่า วิธี IRT ทั้ง 3 วิธี จะมีอำนาจการตรวจสอบสูงที่สุด เฉพาะกรณีที่เป็นแบบสอบสั้น (25 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากปานกลาง อำนาจจำแนกต่ำ ส่วนวิธี MH จะมีอำนาจการตรวจสอบสูงที่สุด เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกต่ำ ในขณะที่วิธี RFA มีอำนาจการตรวจสอบสูงที่สุดทุกเงื่อนไข

กล่าวโดยสรุป เมื่อพิจารณาค่าอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธี IRT ทั้ง 3 วิธี จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูง ในกรณีที่เป็นแบบสอบสั้น (25 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากปานกลาง อำนาจจำแนกต่ำ วิธี MH จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 สูง ในกรณีที่เป็นแบบสอบสั้น (25 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากปานกลาง อำนาจจำแนกปานกลางและอำนาจจำแนกต่ำ ในขณะที่วิธี RFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดทุกเงื่อนไข

ดังนั้น ลักษณะข้อมูลที่เหมาะสมสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแต่ละวิธี มีดังนี้ วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี ทุกเงื่อนไข ในขณะที่วิธี MH จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกสูง วิธี IRT จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ ทุกค่าอำนาจจำแนก แต่เมื่อแยกพิจารณาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธี IRT ทั้ง 3 วิธี พบว่า วิธี IRT(Z-SA) จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกต่ำ วิธี IRT(Z-UN) จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกปานกลาง และวิธี IRT(χ^2) จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี เฉพาะกรณีที่เป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกต่ำ

สรุป ทั้ง 5 วิธี จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงเมื่อเป็นแบบสอบยาว (75 ข้อ) และกลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) ข้อสอบมีความยากต่ำ อำนาจจำแนกสูง

ผลการตรวจสอบครั้งนี้ทำให้ปฏิเสธสมมติฐานการวิจัยข้อที่ 1 ที่กล่าวว่า เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบต่างกัน ภายใต้อำนาจการตรวจสอบและมิติความคลาดเคลื่อนประเภทที่ 1 ไม่ต่างกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และสมมติฐานการวิจัยข้อที่ 2 ที่กล่าวว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็ก ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบต่างกัน ภายใต้อำนาจการตรวจสอบเดียวกัน วิธี MH จะมีอำนาจการตรวจสอบสูงกว่าวิธี RFA และวิธี IRT แบบ 2 พารามิเตอร์ และมีมิติความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี RFA และวิธี IRT แบบ 2 พารามิเตอร์ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตารางที่ 16 สรุปวิธีที่มีอำนาจการตรวจสอบสูงที่สุดและวิธีที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุดในแต่ละเงื่อนไขการศึกษา

เงื่อนไขที่	ขนาดผู้ สอบ	ความยาว ของแบบ สอบ	ค่า ความ ยาก	อำนาจ จำแนก	ขนาด ความ คำเหียง	วิธีที่มีอำนาจการตรวจสอบสูง	วิธีที่มีอัตราความคลาด เคลื่อนประเภทที่ 1 ต่ำ
1	150	25	สูง	สูง	สูง	MH	MH
					ต่ำ	MH	
2				ปาน กลาง	สูง	RFA	MH, IRT(χ^2)
					ต่ำ	RFA	
3				ต่ำ	สูง	RFA	MH, IRT(χ^2)
					ต่ำ	RFA	
4			ปาน กลาง	สูง	สูง	RFA	RFA, MH, IRT(χ^2)
					ต่ำ	RFA	
5				ปาน กลาง	สูง	RFA, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)	RFA
					ต่ำ	RFA	
6				ต่ำ	สูง	RFA, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)	RFA
					ต่ำ	RFA	
7			ต่ำ	สูง	สูง	RFA	RFA, MH, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	
8				ปาน กลาง	สูง	RFA	RFA, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	
9				ต่ำ	สูง	RFA	RFA
					ต่ำ	RFA	
10	150	75	สูง	สูง	สูง	MH	MH
					ต่ำ	-	
11				ปาน กลาง	สูง	MH	MH
					ต่ำ	MH	
12				ต่ำ	สูง	RFA, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)	MH, IRT(χ^2)
					ต่ำ	RFA	
13			ปาน กลาง	สูง	สูง	RFA	RFA
					ต่ำ	RFA	
14				ปาน กลาง	สูง	RFA	RFA, MH, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	
15				ต่ำ	สูง	RFA	RFA, MH
					ต่ำ	RFA	
16			ต่ำ	สูง	สูง	RFA	RFA, MH, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	
17				ปาน กลาง	สูง	RFA	RFA, MH, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	
18				ต่ำ	สูง	RFA	RFA, IRT(Z-SA), IRT(Z-UN), IRT(χ^2)
					ต่ำ	RFA	

ตารางที่ 16 (ต่อ)

เงื่อนไข ที่	ขนาดผู้ สอบ	ความยาว ของแบบ สอบ	ค่า ความ ยาก	อำนาจ จำแนก	ขนาด ความ ลำเอียง	วิธีที่มีอำนาจการตรวจสอบสูง	วิธีที่มีอัตราความคลาด เคลื่อนประเภทที่ 1 ต่ำ
19	500	25	สูง	สูง	สูง	-	MH
					ต่ำ	-	
20				ปาน กลาง	สูง	-	MH
					ต่ำ	-	
21				ต่ำ	สูง	-	MH)
					ต่ำ	-	
22			ปาน กลาง	สูง	สูง	RFA	RFA,IRT(UN),IRT (χ^2)
					ต่ำ	RFA	
23				ปาน กลาง	สูง	RFA,MH	RFA
					ต่ำ	RFA	
24				ต่ำ	สูง	RFA,MH,IRT(Z-SA),IRT(Z-UN),IRT(χ^2)	RFA
					ต่ำ	RFA	
25	ต่ำ	สูง	สูง	RFA	RFA,MH,IRT(χ^2)		
			ต่ำ	RFA			
26		ปาน กลาง	สูง	RFA,MH,IRT(Z-SA),IRT(Z-UN),IRT(χ^2)	RFA,IRT(χ^2)		
			ต่ำ	RFA			
27		ต่ำ	สูง	RFA,MH,IRT(Z-UN),IRT(χ^2)	RFA,IRT(χ^2)		
			ต่ำ	RFA			
28	500	75	สูง	สูง	สูง	MH	MH
					ต่ำ	-	
29				ปาน กลาง	สูง	MH	MH
					ต่ำ	-	
30				ต่ำ	สูง	MH	MH
					ต่ำ	MH	
31			ปาน กลาง	สูง	สูง	RFA,MH	RFA,MH,IRT(Z-SA), IRT(Z-UN),IRT(χ^2)
					ต่ำ	RFA	
32				ปาน กลาง	สูง	RFA,MH	RFA,MH,IRT(Z-N), IRT(χ^2)
					ต่ำ	RFA	
33				ต่ำ	สูง	RFA,MH	RFA
					ต่ำ	RFA	
34	ต่ำ	สูง	สูง	RFA	RFA,MH,IRT(Z-SA), IRT(Z-UN),IRT(χ^2)		
			ต่ำ	RFA			
35		ปาน กลาง	สูง	RFA,MH	RFA,MH,IRT(Z-SA), IRT(Z-UN),IRT(χ^2)		
			ต่ำ	RFA			
36		ต่ำ	สูง	RFA,MH,IRT(Z-UN),IRT(χ^2)	RFA,MH,IRT(Z-SA), IRT(Z-UN),IRT(χ^2)		
			ต่ำ	RFA			

อภิปรายผล

1. ขนาดกลุ่มตัวอย่างมีผลกระทบต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบด้วยวิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี กล่าวคือ

ในกรณีที่กลุ่มตัวอย่างขนาดเล็ก (300 คน) วิธี RFA สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีที่สุด รองลงมาคือ วิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ ซึ่งข้อค้นพบนี้ไม่สอดคล้องกับสมมุติฐานข้อที่ 2 ที่ตั้งเอาไว้ ทั้งนี้อาจเป็นเพราะกลุ่มตัวอย่างขนาดเล็กที่ใช้ไม่เพียงพอสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH ดังนั้นข้อค้นพบนี้จึงเป็นการยืนยันได้ว่า กลุ่มตัวอย่างขนาดเล็ก (เฉพาะกลุ่มอ้างอิง หรือกลุ่มเปรียบเทียบ) ที่เหมาะสำหรับการตรวจสอบด้วยวิธี MH น่าจะมากกว่ากลุ่มละ 150 คน ซึ่งผลการวิจัยครั้งนี้จะสอดคล้องกับการศึกษาของ Mazor และคณะ (1993) ที่พบว่าการใช้กลุ่มตัวอย่างในแต่ละกลุ่มน้อยกว่ากลุ่มละ 200 คน (เฉพาะกลุ่มอ้างอิง หรือกลุ่มเปรียบเทียบ) จะมีผลกระทบต่อประสิทธิภาพในการตรวจสอบของวิธี MH

ในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ (1000 คน) วิธี RFA มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธี MH และวิธี IRT ทั้ง 3 วิธี ข้อค้นพบนี้จึงไม่สอดคล้องกับสมมุติฐานข้อที่ 1 ที่ตั้งไว้ว่าเมื่อขนาดกลุ่มตัวอย่างมีขนาดใหญ่ วิธี RFA วิธี MH และวิธี IRT จะมีอำนาจการตรวจสอบไม่แตกต่างกัน แต่การศึกษาครั้งนี้สอดคล้องกับผลการศึกษาของ Mazor (1992) Sawaminathan และ Roger (1990) ที่พบว่าเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นจะทำให้อำนาจการตรวจสอบของแต่ละวิธีสูงขึ้นด้วย

ด้านอัตราความคลาดเคลื่อนประเภทที่ 1 เมื่อใช้กลุ่มตัวอย่าง 300 คน วิธี MH และวิธี IRT ทั้ง 3 วิธี จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในระดับหนึ่ง แต่เมื่อเพิ่มขนาดกลุ่มตัวอย่างเป็น 1000 คน อัตราความคลาดเคลื่อนประเภทที่ 1 ก็เพิ่มขึ้นด้วย ทั้งนี้อาจเป็นเพราะเมื่อกลุ่มตัวอย่างมากขึ้น โอกาสในการตรวจพบข้อสอบทำหน้าที่ต่างกันมีมากขึ้น ส่งผลให้อัตราความคลาดเคลื่อนมากขึ้นตามไปด้วย ในขณะที่วิธี RFA สามารถควบคุมไม่ให้มีอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีมาก แสดงว่าการตรวจสอบด้วยวิธี RFA ไม่ขึ้นอยู่กับขนาดกลุ่มตัวอย่าง ซึ่งผลการศึกษาครั้งนี้สอดคล้องกับการศึกษาของ Oort (1998) ที่พบว่าอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี RFA ไม่ขึ้นอยู่กับขนาดกลุ่มตัวอย่าง

2. ระดับความยาวของแบบสอบไม่มีอิทธิพลต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี RFA แต่ความยาวของแบบสอบมีอิทธิพลต่ออำนาจการตรวจสอบของ

วิธี MH และวิธี IRT กล่าวคือ เมื่อเพิ่มความยาวของแบบสอบจาก 25 ข้อ เป็น 75 ข้อ ในกลุ่มตัวอย่างขนาดใหญ่ ทั้งวิธี MH และวิธี IRT จะมีอำนาจการตรวจสอบสูงขึ้น ซึ่งข้อค้นพบนี้สอดคล้องกับการศึกษาของ Roger และ Swaminathan (1990) Uttaro และ Millsap (1994) ที่พบว่าแบบสอบที่มีขนาดยาวและกลุ่มตัวอย่างขนาดใหญ่ ทำให้มีโอกาสตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากขึ้น

ด้านอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี MH และวิธี IRT ทั้ง 3 วิธี เมื่อเพิ่มความยาวของแบบสอบจาก 25 ข้อ เป็น 75 ข้อ อัตราความคลาดเคลื่อนประเภทที่ 1 จะลดลง แสดงว่า แบบสอบที่ยาวขึ้นและกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ทำให้มีโอกาสตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากขึ้น มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำลง ส่งผลให้ประสิทธิภาพในการตรวจสอบสูงขึ้น

3. ผลการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกที่มีอิทธิพลต่ออำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า ข้อสอบที่มีความยากสูง ข้อมูลไม่เหมาะสมสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี RFA วิธี MH และวิธี IRT ทั้ง 3 วิธี ในกรณีการวิเคราะห์ด้วยวิธี RFA และวิธี IRT ทั้ง 3 วิธี นั้นอาจเป็นเพราะข้อสอบที่มีความยากสูง ผู้สอบทำข้อสอบไม่ค่อยได้เหมือนกันทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ทำให้ข้อมูลไม่มี variation ไม่สามารถนำไปวิเคราะห์ได้ และในกรณีของวิธี MH อาจเป็นเพราะ ในกลุ่มข้อสอบที่มีความยากสูง ผู้สอบทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตอบข้อสอบไม่ค่อยถูก เมื่อมีการแจกแจงจำนวนผู้สอบตามระดับความสามารถแล้ว จะพบว่าสัดส่วนของจำนวนผู้ตอบข้อสอบถูกในระดับความสามารถสูง ๆ มีจำนวนน้อยมาก ทำให้ค่าสถิติ χ^2 ที่คำนวณได้ในการตรวจสอบด้วยวิธี MH มีค่าต่ำ โอกาสมีนัยสำคัญจึงมีน้อย เป็นผลให้ตรวจพบข้อสอบทำหน้าที่ต่างกันได้น้อยมากหรืออาจไม่พบเลย

4. ความลำเอียงของข้อสอบมีผลกระทบต่ออำนาจการตรวจสอบ กล่าวคือ ในกรณีที่ข้อสอบมีความลำเอียงสูงทั้งวิธี MH และวิธี IRT แบบ 2 พารามิเตอร์ ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้สูงกว่าข้อสอบมีความลำเอียงต่ำ ซึ่งข้อค้นพบนี้สอดคล้องกับการศึกษาของอ็อด (Oort, 1998) ที่พบว่า ในกรณีที่ข้อสอบมีความลำเอียงสูง วิธี IRT จะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่า ในกรณีข้อสอบที่ข้อสอบมีความลำเอียงต่ำ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการวิจัยพบว่า โดยภาพรวม วิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) มีอำนาจการตรวจสอบสูงที่สุด รองลงมาคือ วิธีแมนเทิล-แฮนส์เซล (MH) และวิธีการทดสอบของข้อสอบ (IRT) ตามลำดับ แต่วิธีการทดสอบของข้อสอบ (IRT) มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) และวิธีแมนเทิล-แฮนส์เซล (MH) นอกจากนี้ยังพบว่า

1.1 วิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) ไม่ยุ่งยากและประหยัดเวลาในการวิเคราะห์ในเงื่อนไขแบบสอบมีความยาว 25 ข้อ แต่ในแบบสอบที่มีขนาดความยาว 75 ข้อ จะใช้เวลานานในการเตรียมแฟ้มข้อมูลในรูปแบบตรรกศาสตร์สัมพันธ์

1.2 วิธีการทดสอบของข้อสอบ (IRT) มีความยุ่งยากในการวิเคราะห์ ซึบซ้อนหลายขั้นตอน แม้จะมีโปรแกรมสำเร็จรูปช่วยในการวิเคราะห์ก็ตาม แต่ยังมีข้อเสียที่พบจากการศึกษาคั้งนี้ คือในบางเงื่อนไขต้องใช้เครื่องคอมพิวเตอร์ที่มีหน่วยความจำ 64 MB ขึ้นไปในการวิเคราะห์ข้อมูล ในเงื่อนไขดังกล่าวทำให้สิ้นเปลืองและเสียเวลา

2. จากการศึกษาพบว่าเงื่อนไขขนาดกลุ่มตัวอย่างที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้อำนาจการตรวจสอบสูงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำคือ 1000 คน

3. ความยาวของแบบสอบ เมื่อพิจารณาถึงอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ขนาดความยาวของแบบสอบ 25 ข้อ และ 75 ข้อ พบว่าการเพิ่มความยาวของแบบสอบให้มากขึ้นจะทำให้อำนาจการตรวจสอบเพิ่มขึ้นด้วย

4. ค่าความยากของแบบสอบ โดยภาพรวมเมื่อพิจารณาถึงอำนาจการตรวจสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ทั้งในข้อสอบที่มีขนาดความลำเอียงต่ำและขนาดความลำเอียงสูงของทุกเงื่อนไขที่ศึกษาพบว่า แบบสอบที่มีค่าความยากต่ำที่มีค่าอำนาจจำแนกสูง ทั้ง 5 วิธีจะให้ค่าอำนาจการตรวจสอบสูงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ

ข้อเสนอแนะในการศึกษาเพิ่มเติม

1. ในการศึกษาครั้งนี้อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ศึกษาคืออัตราส่วน 1:1 ซึ่งยึดอัตราส่วน จากการศึกษาของ จิตติมา วรรณศรี ที่พบว่า อัตราส่วน 1:1 เป็นอัตราส่วนที่ให้ประสิทธิภาพในการตรวจสอบดีมาก แต่จากการศึกษาของ Hambleton (1993) ได้เสนอแนะว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH หากเพิ่มจำนวนผู้สอบกลุ่มอ้างอิงขึ้นและคงจำนวนผู้สอบกลุ่มเปรียบเทียบให้เท่าเดิม จะทำให้อำนาจการตรวจสอบสูงขึ้น และไม่ทำให้อัตราความคลาดเคลื่อนประเภทที่ 1 สูงขึ้นด้วย จึงน่าจะศึกษาประสิทธิภาพในการทำหน้าที่ต่างกันของข้อสอบในกรณีที่อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน ตามคำแนะนำของ Hambleton เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบ เพราะในทางปฏิบัติในการสอบคัดเลือกเข้าศึกษาต่อในสถาบันการศึกษาต่างๆ มักใช้อัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ต่างกันมาก

2. ในปัจจุบันวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ได้รับการพัฒนาขึ้นมาก และจากการติดต่อทาง E-mail กับ Oort ได้รับคำแนะนำจาก Oort ว่าน่าจะได้มีการศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี RFA กับวิธีการต่างๆ เช่น วิธี Logistic Regression เพื่อเปรียบเทียบกับผลที่ได้จากการศึกษาครั้งนี้และเพื่อเป็นทางเลือกใหม่ในการนำไปใช้ด้วย