

บทที่ 6

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่องนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบผลการตรวจสอบความเหมาะสมของบุคคล (Person-fit) ด้วยดัชนีซึ่งมีวิธีประมาณค่าแตกต่างกัน 2 วิธี ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ได้แก่ ดัชนีแอลแซด (L_2 index) ประมาณค่าด้วยฟังก์ชันไลด์ลียูด (The Likelihood-based Approach) และ ดัชนีดับเบิลยูวัน (W_1 index) ประมาณค่าดัชนีจากคะแนนส่วนที่เหลือ (The Score Residual Approach) เมื่อระดับความสามารถของผู้สอบและระดับความยาวของแบบสอบต่างกัน โดยศึกษาระดับความสามารถของผู้สอบ 3 ระดับ คือ ความสามารถสูง ($\theta = 1.5$) ความสามารถปานกลาง ($\theta = 0.0$) และความสามารถต่ำ ($\theta = -1.5$) และแบบสอบความยาว 3 ระดับ คือ ความยาว 20, 40 และ 60 ข้อ

ในส่วนของดัชนี W_1 มีค่าสถิติที่ใช้เป็นเกณฑ์ในการพิจารณาความเหมาะสมของบุคคลกับโมเดลการตอบข้อสอบ 2 ตัว คือ ค่าสถิติ Infit และค่าสถิติ Outfit เนื่องจากค่าสถิติ 2 ตัวนี้มีสูตรการคำนวณที่แตกต่างกัน ผู้วิจัยจึงแยกศึกษาในค่าสถิติแต่ละตัว สำหรับผลของการตรวจสอบความเหมาะสมของบุคคล วัดจากค่าสารสนเทศของแบบสอบ ซึ่งเป็นคุณภาพของแบบสอบตามทฤษฎีการตอบสนองข้อสอบ ความเที่ยงของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม และความตรงตามทฤษฎีซึ่งได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

ข้อมูลที่ใช้ในการศึกษา ผู้วิจัยจำลองข้อมูลด้วยโปรแกรม IRTDATA (Johanson, 1992) แล้วสุ่มข้อมูลตามเงื่อนไขที่ศึกษาด้วยโปรแกรม RANDOM ซึ่งผู้วิจัยสร้างขึ้น

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลโดย การวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของแบบสอบ (Unidimensionality) ทุกเงื่อนไขที่ศึกษา วิเคราะห์ค่าสถิติพื้นฐานและคุณภาพของแบบสอบ ด้วยโปรแกรม BILOG วิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนีแอลแซดและดัชนีดับเบิลยูวัน หาความเที่ยงของแบบสอบที่ได้จากการวิเคราะห์แบบแผนการตอบข้อสอบด้วยสูตร KR-20 ทดสอบความแตกต่างของค่าความเที่ยง ด้วยค่าสถิติพีชเชอร์-ซีคำนวณค่าสารสนเทศของแบบสอบจากค่าสารสนเทศของข้อสอบ โดยวิเคราะห์ด้วยโปรแกรม

BILOG ตรวจสอบความตรงตามทฤษฎี (Construct Validity) ด้วยการวิเคราะห์องค์ประกอบ
เชิงยืนยัน ด้วยโปรแกรม LISREL

สรุปผลการวิจัย

ผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนีแอลเซตและดัชนีดับเบิลยูวัน
สามารถสรุปได้ดังนี้

1. คุณภาพของแบบสอบตามทฤษฎีการสนองข้อสอบ ซึ่งวัดจากค่าสารสนเทศของแบบ
สอบหลังการวิเคราะห์ผู้สอบที่เหมาะสม ปรากฏผล ดังนี้

1.1 ค่าสารสนเทศของแบบสอบ ซึ่งได้จากการวิเคราะห์แบบแผนการตอบข้อ
สอบด้วยดัชนีแอลเซตมีค่าสูงขึ้น และมีค่าสูงสุดในทุกระดับความยาวของแบบสอบและทุก
ระดับความสามารถของผู้สอบ

1.2 ค่าสารสนเทศของแบบสอบ ซึ่งได้จากการวิเคราะห์แบบแผนการตอบข้อ
สอบด้วยดัชนี W, เมื่อพิจารณาค่าสถิติ Oufit มีค่าสูงกว่า การวิเคราะห์แบบแผนการตอบข้อ
สอบโดยพิจารณาค่าสถิติ Infit เมื่อผู้สอบมีระดับความสามารถสูงหรือต่ำ

1.3 ค่าสารสนเทศของแบบสอบซึ่งได้จากการวิเคราะห์แบบแผนการตอบข้อ
สอบด้วยดัชนี W, เมื่อผู้สอบมีระดับความสามารถสูงหรือต่ำ และแบบสอบสอบซึ่งมีความยาว
น้อยกว่าหรือมากกว่า 40 ข้อ มีแนวโน้มจะให้ค่าสารสนเทศต่ำกว่าแบบสอบก่อนการวิเคราะห์

2. คุณภาพของแบบสอบภายหลังการวิเคราะห์แบบแผนการตอบข้อสอบ ด้วยดัชนี
ปงชี้ความเหมาะสมของบุคคลในด้านความเที่ยงไม่สูงกว่าความเที่ยงของแบบสอบก่อนการ
วิเคราะห์แบบแผนการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนีปงชี้ความเหมาะสมของ
บุคคลทั้ง 2 วิธี เมื่อผู้สอบมีระดับความสามารถต่างกัน ปรากฏผล ดังนี้

3.1 เมื่อผู้สอบมีระดับความสามารถสูง ยังไม่มีข้อมูลเพียงพอที่จะสรุปว่า ดัชนี
แอลเซตมีผลการตรวจสอบผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบสูงกว่าดัชนีดับเบิลยู
วันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.2 เมื่อผู้สอบมีระดับความสามารถปานกลาง ไม่มีข้อมูลเพียงพอที่จะสรุปว่า
ผลการตรวจสอบผู้สอบไม่เหมาะสมกับโมเดลการตอบข้อสอบ ของวิธีประมาณค่าดัชนีปงชี้
ความเหมาะสมของบุคคลทั้ง 2 วิธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.3 เมื่อผู้เข้าสอบมีระดับความสามารถต่ำ ยังไม่มีข้อมูลเพียงพอที่จะสรุปว่า ดัชนีแอลเซดต์มีผลการตรวจสอบผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ สูงกว่าดัชนี ดัชนีเบ็ลยูวัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4. ผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคล ทั้ง 2 วิธี เมื่อแบบสอบมีความยาวต่างกันปรากฏผลว่า ยังไม่มีข้อมูลเพียงพอที่จะสรุปว่า ดัชนีแอลเซดต์ มีผลการตรวจสอบผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ สูงกว่าดัชนี ดัชนีเบ็ลยูวัน ในทุกระดับความยาวของแบบสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อภิปรายผลการวิจัย

จากผลการวิเคราะห์ข้อมูลในบทที่ 4 และสรุปผลการวิจัย ที่กล่าวข้างต้น ผู้วิจัย อภิปรายผลในประเด็นต่างๆ ดังต่อไปนี้

1. ผลการตรวจสอบความเหมาะสมของบุคคลในการวิจัยนี้ พบว่าใช้ดัชนีซึ่งประมาณ ค่าด้วยวิธีต่างกันตรวจค้นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบได้จำนวนต่างกัน ซึ่ง พบว่าดัชนีแอลเซดต์สามารถตรวจค้นผู้สอบที่ไม่เหมาะสมได้มากกว่าดัชนีดัเบ็ลยูวัน เมื่อผู้ สอบมีระดับความสามารถปานกลางและความสามารถสูง ทุกระดับความยาวของแบบสอบ และผู้สอบระดับความสามารถต่ำ แบบสอบยาว 40 และ 60 ข้อ ส่วนแบบสอบความยาว 20 ข้อ ดัชนีดัเบ็ลยูวัน (W_v) เมื่อพิจารณาค่าสถิติ Infit ระบุจำนวนผู้สอบที่ไม่เหมาะสมมากกว่า ผลการศึกษานี้สอดคล้องกับผลการศึกษาของ ดรากลโว์และคณะ (Drasgow et. al.,1987) ซึ่งทำการศึกษาค่าสถิติ Infit ระบุจำนวนผู้สอบที่ไม่เหมาะสมกับ โมเดลการตอบข้อสอบสูงกว่าดัชนีดัเบ็ลยูวัน

2. ค่าสารสนเทศของแบบสอบซึ่งได้จากการวิเคราะห์แบบแผนการตอบข้อสอบ ด้วย ดัชนีแอลเซดต์ มีค่าสูงขึ้น และสูงกว่า ค่าสารสนเทศของแบบสอบซึ่งวิเคราะห์แบบแผนการตอบ ข้อสอบด้วยดัชนีดัเบ็ลยูวันหนึ่ง ผลการวิจัยนี้สอดคล้องกับผลการวิจัยของ ดรากลโว์และคณะ (Drasgow, et.al.,1987) ซึ่งทำการศึกษาค่าดัชนีแอลเซดต์สามารถตรวจพบบุคคลที่ไม่ เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 75 และ 46 เมื่อจำลองข้อมูลให้ผู้สอบได้ คะแนนสูงและต่ำกว่าความเป็นจริง ในขณะที่ดัชนีดัเบ็ลยูวันตรวจพบร้อยละ 73 และ 21 ตาม ลำดับ ทั้งนี้เนื่องจากบุคคลที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ ด้วยสาเหตุต่าง ๆ นั้น ปร บอกว่าระดับความสามารถ (ability) ที่ประมาณค่าได้ไม่ใช่ความสามารถที่แท้จริง (Meijer,

1996; Nering,1996; Reise,1990) ดังนั้นการนำผู้สอบเหล่านี้จากการวิเคราะห์ข้อสอบ ย่อมทำให้คุณภาพของแบบสอบดีขึ้น (Schmitt: et.al, 1993; Drasgow et.al,1987)

3. คุณภาพของแบบสอบในด้านความเที่ยงของแบบสอบ หลังการวิเคราะห์แบบแผนการตอบข้อสอบ ไม่สูงขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งผลการวิจัยนี้ขัดแย้งกับการศึกษาของ ชมิต และคณะ (Schmitt; et al,1993) ที่มีผลการศึกษาว่า การวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนีแอลซัด ส่งผลให้แบบสอบมีความตรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity) สูงขึ้น ทั้งนี้ชมิต ได้อภิปรายผลไว้ว่า กลุ่มตัวอย่างที่ใช้ศึกษามีจำนวนน้อย อาจจะเป็นไปได้ว่าผลการศึกษารอง ชมิต มีจุดด้อยในเรื่องของจำนวนกลุ่มตัวอย่างและการสรุปอ้างอิง ทั้งนี้ปัจจัยซึ่งส่งผลต่อความเที่ยงมีหลายประการนอกเหนือไปจากแบบแผนการตอบข้อสอบของผู้สอบ เช่น ลักษณะของกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างมีลักษณะคล้ายคลึงกัน จะส่งผลให้แบบสอบมีความเที่ยงต่ำส่วนความตรงตามทฤษฎีนั้น ถ้าแบบสอบมีข้อสอบซึ่งวัดมิติอื่นๆ จะส่งผลให้ค่าสารสนเทศของแบบสอบลดลง และข้อมูลเชิงประจักษ์ไม่เหมาะสมกับโมเดลตามทฤษฎีหรือค่าดัชนีของความตรงมีค่าต่ำ (สุวิมล ติรกานันท์, 2537) เนื่องจากในการวิจัยครั้งนี้ผู้วิจัยทำการศึกษาโดยแบ่งระดับความสามารถของผู้สอบ 3 ระดับ คือ ระดับความสามารถสูง ปานกลาง และ ระดับความสามารถต่ำ การที่กลุ่มตัวอย่างในแต่ละเงื่อนไขมีความเป็นเอกพันธ์ (Homogenous) จึงส่งผลให้แบบสอบมีความเที่ยงต่ำ ดังนั้นคุณภาพความเที่ยงของแบบสอบภายหลังการวิเคราะห์แบบแผนการตอบข้อสอบ จึงไม่สูงขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และไม่มีความแตกต่างระหว่างวิธีประมาณค่าดัชนี

ข้อเสนอแนะในการนำไปใช้

1. ในการวิเคราะห์แบบแผนการตอบข้อสอบ โดยใช้ค่าสถิติป้องกันความเหมาะสมของบุคคล (Person-fit Statistics) ควรใช้ดัชนีแอลซัด (L_2 index) ซึ่งประมาณค่าด้วยฟังก์ชันโลคัลลิยูด เนื่องจากคุณภาพของแบบสอบในด้านค่าสารสนเทศของแบบสอบมีค่าสูงเกินกว่าการวิเคราะห์ข้อสอบโดยไม่วิเคราะห์แบบแผนการตอบข้อสอบ และมีค่าสูงกว่าการวิเคราะห์ด้วยดัชนีดับเบิลยูวัน (w_1 index) ซึ่งประมาณค่าจากคะแนนส่วนที่เหลือ

2. ถ้าใช้โปรแกรม BIGSTEPS เพื่อวิเคราะห์ความเหมาะสมของผู้สอบกับโมเดลการตอบข้อสอบ ควรใช้ค่าสถิติ Outfit เป็นเกณฑ์ในการพิจารณา คัดเลือกผู้สอบที่เหมาะสมกับโมเดลการตอบข้อสอบเนื่องจากผลการวิเคราะห์แบบแผนการตอบข้อสอบ ให้ค่าสารสนเทศสูงกว่าการใช้ค่าสถิติ Infit เป็นเกณฑ์ในการพิจารณา

ข้อเสนอแนะสำหรับการทำวิจัยต่อไป

การศึกษาวิจัยในครั้งนี้ ผู้วิจัยศึกษาผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนีซึ่งมีวิธีประมาณค่าแตกต่างกัน โดยวัดผลการตรวจสอบจากคุณภาพของแบบสอบในด้านค่าสารสนเทศของแบบสอบ ความเที่ยง และความตรงตามทฤษฎี โดยใช้แบบสอบซึ่งให้คะแนนแบบ 2 ค่า (Dichotomous score) และศึกษาปัจจัยที่ส่งผลกระทบต่อประสิทธิผลของดัชนี เพียง 2 ตัว คือ ระดับความสามารถของผู้สอบ และความยาวของแบบสอบ ในการจำลองข้อมูลผู้วิจัยจำลองเมตริกซ์คำตอบตามการสอบแบบประเพณีนิยม (Conventional test) จากการแนวคิดและผลการศึกษาในครั้งนี้ มีประเด็นที่น่าจะมีการศึกษาวิจัยเพิ่มเติม เพื่อเพิ่มสารสนเทศเกี่ยวกับการวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคล และดัชนีบ่งชี้ความเหมาะสมของข้อสอบกับโมเดลการตอบข้อสอบ ดังต่อไปนี้

1. ควรมีการศึกษาเกี่ยวกับการวิเคราะห์แบบแผนการตอบข้อสอบซึ่งมีการให้คะแนนแบบหลายค่า (Polytomous score) หรือ มาตรวัด ประเภทต่างๆ เนื่องจากในปัจจุบันมีการตรวจสอบโดยใช้คะแนนความจริงใจ หรือมาตรจับเท็จ (Lie scale) ซึ่งยังไม่มีข้อสรุปที่ชัดเจนว่าควรใช้วิธีการใดในการตรวจสอบ
2. เนื่องจากการศึกษาวิจัยครั้งนี้ ศึกษาผลการสอบของผู้สอบตามแบบประเพณีนิยม คือ ผู้สอบทุกคนตอบข้อสอบจำนวนเท่ากัน ซึ่งจะส่งผลให้เกิดการเดาสำหรับผู้สอบที่มีความสามารถต่ำ หรือผู้ตอบซึ่งมีความสามารถสูงตอบข้อสอบผิดเนื่องจากความประมาทเลินเล่อ และความสะเพร่า พฤติกรรมเหล่านี้เป็นสาเหตุของแบบแผนการตอบข้อสอบที่ผิดปกติ และทำให้ค่าประมาณความสามารถของผู้สอบที่ประมาณค่าได้ มีความคลาดเคลื่อนสูง ในปัจจุบันได้มีการจัดการทดสอบแบบปรับเหมาะสมควรกับระดับความสามารถของผู้สอบ (adaptive testing) เพื่อดำเนินการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบแต่ละคน เป็นการประหยัดเวลา และทำให้การวัดผลมีประสิทธิภาพมากขึ้น (ศิริชัย กาญจนวาสี, 2538) จึงน่าจะมีการศึกษาว่าการดำเนินการสอบแบบปรับเหมาะสมควรกับระดับความสามารถของผู้สอบหรือ การดำเนินการสอบแบบ CAT (Computerized Adaptive Testing) เป็นการทดสอบที่ผู้สอบทุกคนมีแบบแผนการตอบข้อสอบที่ปกติหรือไม่ มีผลแตกต่างจากแบบแผนการตอบข้อสอบตามแบบประเพณีนิยมอย่างไร โดยการวิเคราะห์ความเหมาะสมของบุคคลกับโมเดลการตอบข้อสอบ (Person-fit Analysis)

3. ควรศึกษาผลของปัจจัยต่างๆ ต่อประสิทธิภาพของดัชนีความเหมาะสมของบุคคล เช่น ความยากของแบบสอบ วิธีประมาณค่าพารามิเตอร์ หรือระดับความยาวของแบบสอบซึ่งแบ่งช่วงให้ตีมากกว่าการวิจัยครั้งนี้

4. ควรมีการศึกษาเกี่ยวกับ การวิเคราะห์แบบแผนการตอบข้อสอบ โดยใช้ดัชนีบ่งชี้ความเหมาะสมของบุคคล โดยพิจารณาถึงรูปแบบพฤติกรรมกรรมการตอบที่ผิดปกติ ว่าเกิดขึ้นเนื่องจากสาเหตุใด ซึ่งจะเป็นผลนำไปสู่การสรุปเกี่ยวกับ โมเดลการตอบข้อสอบแบบ 4 พารามิเตอร์ ซึ่ง แมกโดนัลด์, บาร์ตัน และ ลอร์ด (Hambleton and Swaminathan, 1985) ได้เสนอว่า พารามิเตอร์ที่ 4 นอกจากค่า ความยาว (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) คือ ความประมาทเลินเล่อของผู้สอบซึ่งมีระดับความสามารถสูง แต่เกิดความผิดพลาดในการตอบข้อสอบ ทำให้ตอบข้อสอบผิดในข้อที่มีค่าความยากต่ำกว่าระดับความสามารถ ทั้งนี้โมเดลนี้ยังไม่สามารถประยุกต์ใช้ได้

6. ควรพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อสอบ ซึ่งสามารถวิเคราะห์ผู้สอบที่เหมาะสมกับโมเดลการตอบข้อสอบ (Person-fit) และข้อสอบที่เหมาะสมกับโมเดล (Item-fit) ซึ่งประมาณค่าดัชนีด้วยฟังก์ชันไลค์ลิฮูด (Likelihood - based Approach)