การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

ปริญญานิพนธ์
ของ
มิ่ง เทพครเมือง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

ปริญญานิพนธ์

ของ

มิ่ง เทพครเมือง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

บทคัดย่อ

ของ

มิ่ง เทพครเมือง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

มิ่ง เทพครเมือง. (2554). *การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์, รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ, ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ.

การศึกษาครั้งนี้มีจุดมุ่งหมาย 1) เพื่อศึกษาการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ และ 2) เพื่อเปรียบเทียบผลการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลจำลองที่รายการคำตอบเป็นการตรวจให้คะแนนแบบสองค่า (Dichotomous Scoring) และรายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า (Polytomous Scoring) ฉบับละ 25 ข้อ ภายใต้เงื่อนไขขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ 3 ขนาด คือ ไม่มีการทำหน้าที่เบี่ยงเบน การทำหน้าที่เบี่ยงเบนขนาดเล็ก และการทำหน้าที่เบี่ยงเบนขนาดใหญ่ กลุ่มตัวอย่างแบ่งตามสัดส่วนกลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 2 สัดส่วน คือ 1:1 และ 1:2 แล้วทำการจำลองข้อมูลซ้ำในแต่ละเงื่อนไขที่ศึกษาจำนวน 100 ครั้ง

ผลการวิจัยสรุปได้ดังนี้

1. อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ ที่กลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องของการตรวจสอบสูงกว่าวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนขนาดใหญ่ ที่กลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 200:400 คน วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบต่ำกว่าวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความถูกต้องในการตรวจสอบไม่แตกต่างกัน

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ที่กลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 200:400 คน วิธีตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบสูงกว่าวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบไม่แตกต่างกัน

DETECTION OF MEASUREMENT EQUIVALENCE BASED ON CONGENERIC TEST THEORY AND ITEM RESPONSE THEORY

AN ABSTRACT

BY

MING THEPKORNMUANG

Presented in Partial Fulfillment of the Requirements for the

Doctor of Education Degree in Testing and Measurement

at Srinakharinwirot University

May 2011

Ming Thepkornmuang. (2011). *Detection of measurement equivalence based on Congeneric Test Theory and Item Response Theory*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc. Prof. Dr. Boonchird Pinyoanuntapong, Assoc. Prof. Dr. Pachongchit Intasuwan, Asst. Prof. Dr. Surachai Meechan.

The purposes of this study were to investigate the detection of measurement equivalence based on Congeneric Test Theory (CTT) and Item Response Theory (IRT), and to compare the results of detection of those two theories. The data used in the study were simulated data of 25 items for each test with dichotomous scoring and polytomous scoring, and under three differential item functioning (DIF) conditions: none, small, and large. The sample was divided into two groups of 1:1 and 1:2 according to the ratio of reference group to focal group. The replication of 100 sets of data was made for each condition.

The results of research were as follows.

1. The power rate for the detection of measurement equivalence of CTT was higher than that of IRT with statistical significance at the level of .05 when using dichotomous scoring, small DIF, large DIF, and 100:100 ratio of reference group to focal group. The power rate for the detection of measurement equivalence of CTT was lower than that of IRT with statistical significance at the level of .05 when using polytomous scoring, large DIF, and 200:400 ratio of reference group to focal group. The power rates for the detection of other conditions were not significantly different.

2. The Type I error rate for the detection of measurement equivalence of CTT was higher than that of IRT with statistical significance at the level of .05 when using dichotomous scoring and 200:400 ratio of reference group to focal group. The Type I error rates for the detection of other conditions were not significantly different.

ปริญญานิพนธ์
เรื่อง

การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบ
แบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

ของ
มิ่ง เทพครเมือง

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

…….....……คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนกุล)
วันที่.....เดือน พฤษภาคม พ.ศ. 2554

คณะกรรมการควบคุมปริญญานิพนธ์

……ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

……..กรรมการ
(รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ)

……..กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ)

คณะกรรมการสอบปากเปล่า

…….....ประธาน
(อาจารย์ ดร.สุวพร เซ็มเฮง)

…….....กรรมการ
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์)

…….....……กรรมการ
(รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ)

…….....……กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ)

…….....กรรมการ
(อาจารย์ ดร.สุนิสา จุ้ยม่วงศรี)

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัย
จาก
งบประมาณแผ่นดิน ประจำปีงบประมาณ 2552
ระดับปริญญาเอก ประจำปีการศึกษา 2551 (รอบแรก)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี เป็นเพราะผู้วิจัยได้รับความกรุณาอย่างดียิ่งจาก รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์ ประธานกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ และผู้ช่วยศาสตราจารย์ ดร.สุรชัย มีชาญ กรรมการควบคุมปริญญานิพนธ์ ท่านทั้งสามได้เสียสละเวลาอันมีค่าในการให้คำปรึกษา คำแนะนำ และข้อคิดเห็น ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ อีกทั้งให้กำลังใจแก่ผู้วิจัยตลอดมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ อาจารย์ ดร.สุวพร เซ็มเฮง และ อาจารย์ ดร.สุนิสา จุ้ยม่วงศรี กรรมการที่แต่งตั้งเพิ่มเติม ที่กรุณาให้คำแนะนำ และข้อเสนอแนะที่มีคุณค่า ทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์ อาจารย์ที่ปรึกษาที่ให้การดูแล ด้วยความห่วงใย และให้คำแนะนำอย่างดียิ่งตลอดระยะเวลาที่ศึกษาในมหาวิทยาลัย

ขอขอบพระคุณ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ให้ทุนสนับสนุนส่วนหนึ่งในการวิจัยครั้งนี้

ขอบคุณ ภรรยาและลูกทั้งสอง ที่ต้องดูแลตัวเองเป็นอย่างมาก และให้กำลังใจกับผู้วิจัยตลอดในช่วงที่ผู้วิจัยศึกษาหาความรู้และทำปริญญานิพนธ์ จนทำให้เวลาของครอบครัวหายไปบ้าง

กราบขอบพระคุณบิดา มารดา และบุคคลในครอบครัวทุกคนที่คอยให้กำลังใจ และปรารถนาจะเห็นความสำเร็จของผู้วิจัยเป็นอย่างยิ่ง คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาพระคุณ บิดา มารดา ครู-อาจารย์ทุกท่าน

มิ่ง เทพครเมือง

สารบัญ

บทที่		หน้า
1	**บทนำ**	1
	ภูมิหลัง	1
	ความมุ่งหมายของการวิจัย	6
	ความสำคัญของการวิจัย	6
	ขอบเขตของการวิจัย	8
	ข้อมูลในการศึกษา	8
	ตัวแปรที่ศึกษา	8
	นิยามศัพท์เฉพาะ	9
	กรอบแนวคิดในการวิจัย	11
	สมมติฐานในการวิจัย	17
2	**เอกสารและงานวิจัยที่เกี่ยวข้อง**	18
	ทฤษฎีการทดสอบ	18
	ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory)	18
	ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)	22
	ความเท่าเทียมกันของการวัด (Measurement Equivalence)	34
	การตรวจสอบความเท่าเทียมกันของการวัด (Detection of Measurement Equivalence)	38
	การตรวจสอบความเท่าเทียมกันด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน	39
	การตรวจสอบความเท่าเทียมกันของการวัดตามทฤษฎีการตอบข้อสอบ	46
	งานวิจัยที่เกี่ยวข้อง	51
	งานวิจัยต่างประเทศ	51
	งานวิจัยในประเทศ	54
3	**วิธีดำเนินการ**	56
	การจัดกระทำตัวแปร	56
	การจำลองข้อมูล (Simulation)	59
	การวิเคราะห์ข้อมูล	64

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล	69
ผลการจำลองข้อมูล (Simulation)	70
อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1	83
อัตราความถูกต้องการตรวจสอบความเท่าเทียมกันของการวัด	83
อัตราความคลาดเคลื่อนแบบที่ 1 ของการตรวจสอบความเท่าเทียมกันของการวัด	90
การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1	94
การเปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด	94
การเปรียบเทียบอัตราความคลาดเคลื่อนแบบที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด	97
สรุปผลการศึกษาในภาพรวม	100
5 **สรุป อภิปรายผล และข้อเสนอแนะ**	109
สรุปผลการวิจัย	113
อภิปรายผลการวิจัย	116
ข้อเสนอแนะ	121
บรรณานุกรม	123
ภาคผนวก	135
ภาคผนวก ก ภาพ Histogram และกราฟเส้นแสดงการกระจายของข้อมูล	136
ภาคผนวก ข ตัวอย่างคำสั่งในการวิเคราะห์ข้อมูล	146
ประวัติย่อผู้วิจัย	188

บัญชีตาราง

ตาราง **หน้า**

1 ระดับความสามารถของผู้สอบจำแนกตามขนาดของผู้สอบ..... 50

2 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า..... 74

3 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า..... 76

4 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษา กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า..... 78

5 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษา กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ..... 80

6 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษาที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก..... 81

7 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษาที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่..... 82

8 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) 85

9 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน(1:1) 85

10 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2.... 86

11 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2... 87

12 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1).... 88

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

13 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบมีการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)... 88

14 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 89

15 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 89

16 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)..... 92

17 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 92

18 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)..... 93

19 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 94

20 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)..... 95

21 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2..... 95

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

22 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดที่รายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)..... 96

23 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2..... 97

24 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบ แบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)..... 98

25 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดของการที่เป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2..... 98

26 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)..... 99

27 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2..... 99

28 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบสองค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)..... 101

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

29 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบสองค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 102

30 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบหลายค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)..... 103

31 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบหลายค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 104

32 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)..... 105

33 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2..... 106

34 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ..... 107

35 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกัน ของการวัด ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ..... 108

บัญชีภาพประกอบ

ภาพประกอบ **หน้า**

1 กรอบแนวคิดการวิจัย..... 17
2 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์..... 28
3 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์..... 29
4 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์..... 30
5 โค้งรายการคำตอบ (Category Response Curves: CRC)..... 33
6 โค้งลักษณะปฏิบัติการ (Operating Characteristic Curves: OCC)..... 34
7 โมเดลสมการโครงสร้าง..... 40
8 แผนภาพการจำลองข้อมูล..... 58

บทที่ 1
บทนำ

ภูมิหลัง

ทฤษฎีการทดสอบได้รับการพัฒนาจากนักจิตวิทยา นักสถิติและนักวัดผลมาอย่างต่อเนื่อง จนเป็นศาสตร์ที่ยอมรับกันอย่างกว้างขวางทั้งในวงการการศึกษา จิตวิทยา และสาขาวิชาอื่น ๆ ทฤษฎีการทดสอบทางการศึกษาเป็นองค์ความรู้ที่มีนัยทั่วไปที่นักทฤษฎีการวัดได้เสนอแบบจำลองในการวัด ข้อตกลงเบื้องต้น การพัฒนาเครื่องมือ การวิเคราะห์และการนำไปใช้ เพื่อแก้ปัญหาทางการวัด และพัฒนาแบบทดสอบให้มีคุณภาพ ทั้งนี้เพราะคุณลักษณะภายในของบุคคลเป็นสิ่งที่ไม่สามารถวัดหรือสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นที่ต้องศึกษา เนื่องจากคุณลักษณะภายใน มีอิทธิพลต่อการแสดงออกของบุคคล ถ้าเราสามารถวัดคุณลักษณะภายในของบุคคลได้จะทำให้เราเข้าใจถึงสาเหตุแห่งการเกิดพฤติกรรมของบุคคล อันจะเกิดประโยชน์อย่างยิ่งต่อการพัฒนา และควบคุมพฤติกรรมของบุคคลให้ไปในทิศทางที่สังคมต้องการได้

ฐานคิดทฤษฎีการทดสอบสามารถจำแนกกว้าง ๆ ได้เป็น 2 ฐานคิด คือ ***ฐานคิดตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory: CTT)*** เป็นทฤษฎีที่เชื่อว่าการวัดใด ๆ จะมีผลกระทบจากความคลาดเคลื่อนเสมอ ดังนั้น คะแนนสอบ (Observed Score: X) จึงประกอบด้วยคะแนนจริง (True Score: T) และคะแนนคลาดเคลื่อน (Error Score: E) เขียนสมการมูลฐานเป็น $X_i = T_i + E_i$ ภายใต้ทฤษฎีมาตรฐานเดิมนี้ยังแบ่งการทดสอบออกเป็น 3 ระดับ ตามระดับความคู่ขนานหรือความคล้ายคลึงกันของคะแนน ได้แก่ การทดสอบแบบคะแนนจริงเท่ากัน (Parallel Test) การทดสอบแบบคะแนนจริงสมมูล (Essentially Tau-Equivalent Test) และการทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test) โดยเฉพาะทฤษฎีทดสอบคะแนนจริงสัมพันธ์ (Congeneric Test Theory) เป็นรูปทั่วไปของทฤษฎีการทดสอบมาตรฐานเดิม ที่มีกรอบแนวคิดของคะแนนสอบ (x) ประกอบด้วย $aT+b+E$ และใช้ข้อตกลงของระดับความคู่ขนานหรือความคล้ายคลึงกันของคะแนนที่ผ่อนปรนมากที่สุด นั่นคือ คะแนนจริง (T) ของตัวแปรแฝงเดียวกัน สองค่าใด ๆ มีสหสัมพันธ์เป็น 1.0 หรือ $T_i = a_{ij}T_j + b_{ij}$ มีแบบจำลองคะแนนที่วัดองค์ประกอบเดียว และหลายองค์ประกอบ มีวิธีประมาณค่าความเชื่อมั่นจากค่าความแปรปรวน-ความแปรปรวนร่วม และจากค่าน้ำหนักองค์ประกอบของการทดสอบ รวมทั้งวิธีปรับเทียบคะแนนซึ่งมีประโยชน์และคุณค่าในการนำไปใช้ในเชิงปฏิบัติ (บุญเชิด ภิญโญอนันตพงษ์. 2549: 17-21) ส่วนอีกฐานคิดหนึ่งเป็น ***ฐานคิดตามทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory: MTT)*** เป็นทฤษฎีที่ยอมให้ความคลาดเคลื่อนจากการวัดมีความแตกต่างกันไป ซึ่งภายใต้ทฤษฎีการทดสอบแนวใหม่นี้สามารถจำแนกออกเป็น 2 แนว คือ *ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: G-Theory)* ที่เสนอวิธีวิเคราะห์ค่าความสัมพันธ์ของคะแนนความแปรปรวนคลาดเคลื่อนอย่างเป็นระบบจากแหล่งต่าง ๆ อันเป็นสถานการณ์

หรือเงื่อนไขของการวัด รวมทั้งปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขของการวัด ถือว่าคะแนนความคลาดเคลื่อนของการวัดสามารถเกิดขึ้นได้หลายแหล่ง (Multiple sources of error) ที่สามารถประมาณค่าแยกกันภายใต้การวิเคราะห์ครั้งเดียวกันได้ และอีกแนวหนึ่งคือ *ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)* เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล (Latent trait or ability) กับผลการตอบข้อสอบหรือข้อคำถาม โดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) มีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) ค่าอำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) ทฤษฎีการตอบข้อสอบอยู่บนฐานความคิดที่สำคัญ 2 ประการคือ 1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบ และ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในของผู้สอบ สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) ที่มีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่าฟังก์ชันโลจิสติก (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) (ศิริชัย กาญจนวาสี. 2549: 47-51) โมเดลดังกล่าวสามารถจำแนกได้เป็น 2 ประเภทตามลักษณะการตรวจให้คะแนนคำตอบ ได้แก่ โมเดลการตอบข้อสอบที่เป็นการตรวจให้คะแนนรายการคำตอบแบบสองค่า (Binary or Dichotomous) และโมเดลการตอบข้อสอบที่มีการตรวจให้ให้คะแนนแบบหลายค่า (Polytomous)

จุดมุ่งหมายหนึ่งของการทดสอบก็เพื่อที่จะทำให้สะดวกและง่ายในการเปรียบเทียบคะแนนของกลุ่มผู้สอบที่ต่างกัน การที่จะนำผลการทดสอบมาใช้พิจารณาตัดสินใจ ไม่ว่าจะเป็นการตัดสินผลการเรียน การคัดเลือกเพื่อเข้าศึกษาต่อ การคัดเลือกเพื่อบรรจุงาน การเลื่อนตำแหน่ง การออกใบรับรองหรือใบอนุญาตต่าง ๆ แบบทดสอบควรจะเป็นแบบทดสอบมาตรฐานที่นอกจากจะต้องมีความเที่ยงตรง ความเชื่อมั่น ความยากที่เหมาะสมและสามารถจำแนกระดับความสามารถผู้สอบได้แล้ว สิ่งสำคัญอีกประการหนึ่งที่ผู้เกี่ยวข้องกับการใช้แบบทดสอบควรคำนึงถึงก็คือ *ความเท่าเทียมกันของการวัด* แบบทดสอบที่ไม่มีความเท่าเทียมกันของการวัดจะทำให้ผลการวัดเกิดความลำเอียง เกิดการได้เปรียบเสียเปรียบกันของกลุ่มผู้สอบ ทั้ง ๆ ที่ผู้สอบมีความสามารถเท่า ๆ กัน ความลำเอียงของแบบทดสอบดังกล่าวนี้จัดเป็น*ความคลาดเคลื่อนอย่างเป็นระบบ* (Systematic error) (Camilli; & Shepard. 1994: 8) ความเท่าเทียมในการวัดเป็นการตรวจสอบ *“ความสัมพันธ์ระหว่างคะแนนสังเกตได้กับตัวแปรแฝงจากกลุ่มประชากรที่แตกต่างกัน”* (Drasgow. 1984: 134) จากแนวคิดที่ว่า ตัวแปรแฝงของผู้สอบมาจากกลุ่มตัวอย่างประชากรย่อย (เช่น เพศ ระดับสติปัญญา เชื้อชาติ ศาสนา วัฒนธรรม สถานภาพทางเศรษฐกิจ องค์กร พื้นที่ทางภูมิศาสตร์ ฯลฯ) ที่แตกต่างกัน ในระดับความสามารถเดียวกัน ทำการวัดด้วยเครื่องมือวัดของคุณลักษณะเดียวกันแล้วควรจะให้ค่าคาดหวังที่วัดได้เป็นค่าเดียวกัน การตรวจสอบความเท่าเทียมกันของการวัด (Measurement Equivalence) ข้ามกลุ่มจากเครื่องมือวัดเดียวกันจะเป็นวิธีการตรวจสอบอีกวิธีการหนึ่งที่จะทำให้สามารถเปรียบเทียบว่าผู้สอบที่ต่างกลุ่มกันจะมีคะแนนคล้ายกันหรือแตกต่างกันหรือไม่

การศึกษาลักษณะทางจิตวิทยาย่อมมีวัฒนธรรมกลุ่มชนปนอยู่ด้วยเสมอ การวิจัยข้ามวัฒนธรรมจึงต้องแสดงให้เห็นว่าความแตกต่างระหว่างวัฒนธรรมได้สะท้อนความแตกต่างที่แท้จริงของคุณลักษณะที่ศึกษา ไม่ใช่เกิดจากการลำเอียงหรืออคติของผู้วิจัย เช่น การใช้วิธีการ ทฤษฎี หรือเครื่องมือของโลกตะวันตก หรือเกิดจากการขาดความเท่าเทียมกัน (Equivalence) ในกระบวนการวัด หรือเกิดจากสิ่งอื่นที่มนุษย์สร้างขึ้น (ผจงจิต อินทสุวรรณ. 2544; อ้างอิงจาก Van de Vijver; & Poortinga. 1982) ในงานวิจัยข้ามวัฒนธรรมหรือข้ามกลุ่ม ส่วนใหญ่มักจะเริ่มด้วยการพัฒนาเครื่องมือขึ้นในวัฒนธรรมหนึ่งหรือกลุ่มหนึ่งแล้วจึงนำไปใช้ในอีกวัฒนธรรมหนึ่งหรือกลุ่มอื่น ผลสรุปจากการวิจัยอาจทำให้เข้าใจคนในวัฒนธรรมอื่นหรือกลุ่มอื่นผิดได้ ตัวอย่างเช่น ปัญหาของความเท่าเทียมกันของแบบทดสอบวัด IQ ในการทดสอบความสามารถ ซึ่งเป็นแบบทดสอบที่ออกแบบมาเพื่อวัดเชาว์ปัญญาของคน โดยเฉพาะอย่างยิ่งในช่วงแรก ๆ ของการทดสอบสภาวะทางจิต (Gould. 1981) แบบทดสอบเหล่านี้พัฒนาและเป็นมาตรฐานของประเทศสหรัฐอเมริกาที่ใช้ในการประเมิน “เชาว์ปัญญา” ของบุคคลและกลุ่มบุคคลที่แตกต่างกันทางเชื้อชาติ สัญชาติและภาษา ส่งผลให้ผู้ที่ได้คะแนนต่ำกว่ากลุ่มคนเหล่านั้นได้รับการแปลความหมาย โดยยืนยันว่ามีความแตกต่างในความสามารถทางปัญญา ซึ่งในบางกรณีจะมีผลร้ายแรงสำหรับคนเหล่านั้น เช่น การศึกษาของ Henry Goddard นักจิตวิทยาชาวอเมริกันที่ทำการทดสอบ IQ ผู้อพยพจากยุโรปตะวันออกเฉียงใต้และเกาะเอลลิส ในปี 1910 แล้วบอกว่าคนส่วนใหญ่ของกลุ่มที่ได้รับการทดสอบนี้ “ปัญญาอ่อน” (Goddard. 1917) ผลงานของ Henry Goddard ทำให้ผู้อพยพถูกเนรเทศออกจากประเทศเป็นจำนวนมาก (Hothersall. 1995) แม้ว่าผลงานชิ้นนี้จะโด่งดังมากที่สุด แต่มีการละเมิดระเบียบวิธีการวัดขึ้น การวัดบุคลิกภาพและการวัดทัศนคติก็พบปัญหาของความเท่าเทียมกันของการวัด แม้ว่าจะมีงานวิจัยอยู่จำนวนมากที่ศึกษาในความคล้ายคลึงกันของโครงสร้างปัจจัยบุคลิกภาพข้ามวัฒนธรรม (Church; & Lonner. 1998) แต่ก็ยังมีการศึกษาเกี่ยวกับการประเมินความเท่าเทียมกันของการวัดน้อยมาก

โดยความเชื่อในเรื่องของความเท่าเทียมกันของการวัดนั้นใกล้เคียงกับแนวคิดของความลำเอียงของข้อสอบ (Item Bias) และการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ที่มาจากกรอบทฤษฎีการตอบข้อสอบ (IRT) ที่มีการศึกษาอย่างจริงจังในด้านการทดสอบทางการศึกษาอย่างมาก (Van de Vijver; & Leung. 1997) การทำหน้าที่เบี่ยงเบนของข้อสอบหรือความลำเอียงของข้อสอบหมายถึง ความแตกต่างในความน่าจะเป็นของคำตอบสำหรับผู้ตอบแบบสอบถามด้วยแนวโน้มในการที่จะแสดงพฤติกรรมเท่ากัน ตัวอย่างเช่น ถ้าเราตั้งสมมติฐานในการวัดว่าจะวัด “ความเข้มข้นของพฤติกรรมกลุ่มคนที่มีต่อกีฬาฟุตบอล” ในประเทศไทยและประเทศอังกฤษ ซึ่งข้อคำถามมักจะสอบถามเกี่ยวกับการเข้าร่วมชมหรือร่วมกิจกรรมการแข่งขันฟุตบอล แล้วมีรายการคำตอบให้เลือกตอบ เช่น “สัปดาห์ละครั้ง” สำหรับแฟนฟุตบอลในประเทศไทยอาจจะมีต่ำกว่าแฟนฟุตบอลประเทศอังกฤษ ทั้งที่กลุ่มคนทั้งสองกลุ่มมีทัศนคติเกี่ยวกับฟุตบอลเหมือนกัน ทั้งนี้เพราะการแข่งขันฟุตบอลในประเทศไทยมีการจัดการแข่งขันน้อยกว่าประเทศอังกฤษ จะเห็นว่าข้อความต่าง ๆ

เหล่านี้จะเป็นตัวบ่งชี้ถึงความลำเอียงได้ หากใช้ผลการศึกษาไปการเปรียบเทียบความเข้มข้นของพฤติกรรมที่มีต่อกีฬาฟุตบอลในกลุ่มคนจากสองประเทศนี้

ปัจจุบันการศึกษาข้ามวัฒนธรรมมีความน่าสนใจเป็นอย่างมาก แต่ก็ยังไม่มีวิธีที่มีความชัดเจนกับการตรวจสอบความเท่าเทียมกันของการวัดในงานวิจัยลักษณะนี้ เนื่องจากเครื่องมือชนิดเดียวกัน (เช่น แบบสอบถาม แบบตรวจสอบรายการ แบบทดสอบ และอื่น ๆ) มีการใช้กับกลุ่มคนที่เกี่ยวข้องในการศึกษาทั้งหมด ทำให้เกิดข้อสันนิษฐานขึ้นบ่อยครั้งถึงผลที่นำมาเปรียบเทียบระหว่างกลุ่มบุคคล และข้อตกลงในการเปรียบเทียบผลการศึกษา ซึ่งนักวิจัยมักจะเน้นเฉพาะความแตกต่างในคะแนนเฉลี่ยของคนสองกลุ่มวัฒนธรรมหรือมากกว่า แต่ไม่ได้ทำการทดสอบความเท่าเทียมกันของการวัดข้ามกลุ่มวัฒนธรรม ผลการวิจัยจึงเป็นการสะท้อนให้เห็นถึงบริบทของวัฒนธรรมในแต่ละกลุ่มชนนั้นที่หลากหลายปัจจัย หลากหลายกระบวนการ และหลากหลายคุณลักษณะ ชุดคำถามที่เหมือนกันหรือที่ได้รับการออกแบบมา อาจมีความหมายที่แตกต่างกันสำหรับผู้คนที่ต่างวัฒนธรรมกัน นั่นคือ ผลของการวัดอาจเป็นการวัดโครงสร้างวัฒนธรรมที่ต่างกัน ถ้าผลการวัดที่เกิดขึ้นตรงกันกับข้อสรุปในการวิจัยเชิงเปรียบเทียบที่มีข้อคำถามเหล่านั้นอยู่ จะทำให้เป็นข้อกังวลเบื้องต้นของความเท่าเทียมกันในการวิจัยข้ามวัฒนธรรม ไม่ว่าจะเป็นการเปรียบเทียบหรือทดสอบความแตกต่างข้ามวัฒนธรรม (Hui; & Triandis. 1985) ในวิธีการเปรียบเทียบข้ามกลุ่มนี้เรียกว่า ความเท่าเทียมกันของการวัด หมายถึง “ไม่ว่ากลุ่มผู้สอบจะอยู่ภายใต้เงื่อนไขที่แตกต่างกันหรือไม่ก็ตาม การสังเกตและศึกษาปรากฏการณ์ต่าง ๆ เกี่ยวกับผู้สอบต่างกลุ่มกันจะต้องดำเนินการวัดประเมินของคุณลักษณะเดียวกัน” (Horn; & McArdle. 1992: 117)

ในช่วงตั้งแต่ปี ค.ศ. 1990 เป็นต้นมา นักวิจัยได้ให้ความสนใจในการตรวจสอบความเท่าเทียมกันของการวัดมากขึ้น (Vandenberg. 2002) ซึ่งการทดสอบความเท่าเทียมกันของการวัด เป็นการคำนวณผลที่ได้จากเครื่องมือวัดคุณลักษณะเดียวกันภายใต้เงื่อนไขที่แตกต่างกัน (Horn; & McArdle. 1992) ความแตกต่างของเงื่อนไขที่นักวิจัยให้ความสนใจนั้นมีหลากหลายเงื่อนไข ไม่ว่าจะเป็นเวลาในการทดสอบ (Golembiewski; Billingsley; & Yeager. 1976) การทดสอบกับกลุ่มประชากรที่แตกต่างกัน เช่น ความแตกต่างทางวัฒนธรรม (Riordan; & Vandenberg. 1994) ความแตกต่างของกลุ่มผู้ตรวจ (Facteau; & Craig. 2001) หรือในระดับความแตกต่างในการดำเนินการวัด เช่น การสอบบนเว็บเพจกับการสอบเขียนตอบบนกระดาษ (Taris; Bok; & Meijer. 1998) ซึ่งการทดสอบความเท่าเทียมกันของการวัดต่าง ๆ ที่กล่าวมาจะเป็นการวิเคราะห์องค์ประกอบเชิงยืนยันบนพื้นฐานทฤษฎีการทดสอบแบบมาตรฐานเดิม และได้มีการประยุกต์ใช้ในสถานการณ์อื่น ๆ อย่างกว้างขวาง (Vandenberg; & Lance. 2000) นอกจากนี้ก็ยังมีวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่ใช้ในจุดมุ่งหมายที่คล้ายคลึงกัน คือ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ ซึ่งให้สาระสนเทศที่มีประโยชน์มาก (Maurer; Raju; & Collins. 1998; McDonald. 1999; Raju; Laffitte; & Byrne. 2002)

การตรวจสอบความเท่าเทียมกันของการวัดได้มีการดำเนินการมาอย่างต่อเนื่อง ไม่ว่าจะอยู่บนพื้นฐานของทฤษฎีการทดสอบที่เป็น *วิธีเชิงเส้นตรง (Linear Method)* ตามทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test Theory: CTT) ซึ่งสนใจที่จะศึกษาความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มตัวอย่างที่อิสระจากกัน เป็นการทดสอบความเท่ากันของโครงสร้างองค์ประกอบ (Structural Equivalence) ความเท่าเทียมกันของค่าน้ำหนักองค์ประกอบ (Factor Loading) และการทดสอบความไม่แปรเปลี่ยนของความสัมพันธ์ระหว่างตัวแปร (Marsh; Hau; Chung; & Siu. 1998: 149) และอีกวิธีหนึ่งคือ *วิธีที่ไม่เป็นเชิงเส้นตรง (Nonlinear Method)* เป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบตามทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) วิธีการตรวจสอบความเท่าเทียมกันของการวัดนั้นมีเทคนิคและวิธีการตรวจสอบอยู่หลายวิธี ไม่ว่าจะเป็นการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) ของโจเรสคอก และซอร์บอม (Jöreskog; & Sörbom. 1989) การเปรียบเทียบค่าพารามิเตอร์ข้อสอบโดยใช้ค่าไค-สแควร์ (Chi-square) ของลอร์ด (Lord. 1980) วิธีทดสอบอัตราส่วนความเป็นไปได้ (Likelihood Ratio Test: LR) ของทิสเซน สเตียนเบิร์กและไวเนอร์ (Thissen; Steinberg; & Wainer. 1988) วีธีการวัดพื้นที่ (Area Measures) ของราชู (Raju. 1988, 1990) วิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบและแบบทดสอบ (Differential Functioning of Items and Tests) ที่พัฒนาโดยราชู แวนเดอร์ลินเด็นและเฟลียร์ (Raju; van der Linden; & Fleer. 1995) และวิธีอื่น ๆ ซึ่งล้วนเป็นการประเมินความเท่าเทียมกันของการวัดข้ามกลุ่มประชากรเช่นกัน

จากที่กล่าวมาจะพบว่าการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบในแต่วิธีเหล่านี้ก็ยังมีข้อจำกัดอยู่บ้าง ซึ่งจุดหลักสำคัญประการหนึ่งก็คือ ทฤษฎีการตอบข้อสอบต้องใช้แบบทดสอบและขนาดกลุ่มตัวอย่างจำนวนมาก ซึ่งในสถานการณ์จริงบางกรณีอาจทำให้ไม่เป็นไปตามข้อตกลงเบื้องต้นได้ ส่วนทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์เป็นกรณีทั่วไปของทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งมีข้อผ่อนปรนในเรื่องดังกล่าวจึงยังคงได้รับการพิจารณาใช้ในการตรวจสอบความเท่าเทียมกันของการวัด เพื่อให้ได้ประจักษ์พยานที่แสดงถึงความยุติธรรมของการทดสอบ ดังนั้น ในการศึกษาครั้งนี้ ผู้วิจัยจึงมุ่งที่จะศึกษาการตรวจสอบความเท่าเทียมกันของการวัดที่อยู่บนสองฐานคิดทางทฤษฎีการทดสอบที่ต่างกัน คือ ทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์กับทฤษฎีการตอบข้อสอบ และเพื่อให้สามารถเทียบเคียงความสามารถในการตรวจสอบความเท่าเทียมกันของการวัดจากทั้งสองฐานคิดได้ จึงจำเป็นต้องทำให้ความได้เปรียบเสียเปรียบจากข้อจำกัดระหว่างทฤษฎีทั้งสองเกิดขึ้นน้อยที่สุด ผู้วิจัยจึงเลือกใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ที่อยู่บนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์กับการทดสอบอัตราส่วนความเป็นไปได้ (LR) ที่อยู่บนฐานทฤษฎีการตอบข้อสอบ ทั้งนี้เนื่องจากทั้งสองวิธีต่างก็เป็นการกระทำในระดับข้อสอบเช่นเดียวกัน (Camilli; & Shepard. 1994) โดยผู้วิจัยทำการศึกษา ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 ปัจจัย ดังนี้ **ปัจจัยที่ 1** วิธีการตรวจให้คะแนนรายการคำตอบข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยศึกษาวิธีการตรวจให้คะแนน 2

แบบ คือ การตรวจให้คะแนนรายการคำตอบแบบสองค่า (0, 1) และการตรวจให้คะแนนรายการคำตอบแบบหลายค่า (4, 3, 2, 1, 0) **ปัจจัยที่ 2** ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบออกเป็น 3 ระดับ คือ 1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพื่อทำการศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดจากวิธีการตรวจสอบทั้ง 2 วิธี 2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษา ขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง และ 3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง ซึ่งการที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ เพื่อทำการศึกษาอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี และ**ปัจจัยที่ 3** ขนาดกลุ่มตัวอย่าง ในการศึกษาครั้งนี้ ผู้วิจัยแบ่งตามสัดส่วนของกลุ่มอ้างอิงกับกลุ่มสนใจ 2 สัดส่วน ได้แก่ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 และสัดส่วน 1:2 มี 4 ขนาดเช่นกัน คือ 100:200, 200:400 400:800 และ 500:1000 ผลการศึกษาครั้งนี้จะเป็นประโยชน์ในการเลือกใช้เทคนิควิธีการในการตรวจสอบความเท่าเทียมกันของการวัดให้เหมาะสมกับบริบทของข้อมูล และความเป็นไปได้ในทางปฏิบัติมากที่สุด

ความมุ่งหมายของการวิจัย

ในการศึกษาครั้งนี้ เป็นการศึกษาผลการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีที่ต่างกัน 2 ทฤษฎี คือ ทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ ภายใต้เงื่อนไขวิธีการตรวจให้คะแนน ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบและขนาดกลุ่มตัวอย่างที่แตกต่างกัน โดยมีความมุ่งหมายเฉพาะ ดังนี้

1. เพื่อศึกษาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

2. เพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ

ความสำคัญของการวิจัย

ในการวิจัยครั้งนี้มีเป้าหมายในการตรวจสอบความเท่าเทียมกันของการวัดภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 ปัจจัย ดังนี้ **ปัจจัยที่ 1** วิธีในการตรวจให้คะแนนของคำตอบข้อสอบมาตรวัดโดยทั่วไปจะมีวิธีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้คะแนนแบบสองค่า (0, 1) และการตรวจให้คะแนนแบบหลายค่า (4, 3, 2, 1, 0) **ปัจจัยที่ 2** ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบออกเป็น 3 ระดับ คือ 1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพื่อทำการศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี 2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนก ของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง และ 3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง ซึ่งการที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ เพื่อทำการศึกษาอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี และ**ปัจจัยที่ 3** ขนาดกลุ่มตัวอย่าง โดยในการศึกษาครั้งนี้ ผู้วิจัยแบ่งตามสัดส่วนของกลุ่มอ้างอิงกับกลุ่มสนใจ 2 สัดส่วน ได้แก่ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 คน และสัดส่วน 1:2 มี 4 ขนาดเช่นกัน คือ 100:200, 200:400 400:800 และ 500:1000 คน ทำการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์โดยการวิเคราะห์องค์ประกอบเชิงยืนยันและบนพื้นฐานทฤษฎีการตอบข้อสอบโดยวิธีทดสอบอัตราส่วนความเป็นไปได้

สารสนเทศที่ได้จากวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ สามารถนำไปประยุกต์ใช้ประโยชน์ในด้านการทดสอบทางการศึกษาและจิตวิทยา นักพัฒนาแบบทดสอบสามารถเลือกวิธีการในการตรวจสอบความเท่าเทียมกันของการวัดไปใช้อย่างเหมาะสมกับลักษณะข้อมูล อีกทั้งสามารถประเมินข้อดีและข้อจำกัด และความเป็นไปได้ในทางปฏิบัติในการตรวจสอบความเท่าเทียมกันของการวัด ภายใต้เงื่อนไขวิธีการตรวจให้คะแนน ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่าง นอกจากนี้ยังเป็นแนวทางในการจำลองข้อมูลตามเงื่อนไขข้อตกลงของทฤษฎีเพื่อใช้ในการศึกษาตรวจสอบความคงที่และความเชื่อถือได้ของทฤษฎีวัดผลอื่น ๆ ตามธรรมชาติของข้อมูลที่ต้องการตรวจสอบแล้วนำเสนอหลักฐานพยานที่เชื่อถือได้ และเพื่อเป็นแนวทางสำหรับนักวิจัยและผู้สนใจในการตรวจสอบและคัดกรองเครื่องมือวัดที่พัฒนามาจากวัฒนธรรมหนึ่งหรือกลุ่มประชากรหนึ่ง แล้วนำเครื่องมือเหล่านั้นมาศึกษากับกลุ่มวัฒนธรรม หรือกลุ่มประชากรอื่น หรือ

ผู้วิจัยสนใจเปรียบเทียบข้ามวัฒนธรรมหรือข้ามกลุ่มประชากร ซึ่งจะทำให้การสรุปผลมีความน่าเชื่อถือมากขึ้น

ขอบเขตของการวิจัย

ขอบเขตการวิจัย มีดังนี้

1. ข้อมูลในการศึกษาครั้งนี้เป็นข้อมูลจำลอง ซึ่งผู้วิจัยมีขั้นตอนในการจำลองข้อมูลเป็น 3 ขั้นตอน ดังนี้

1.1 จำลองความสามารถของกลุ่มตัวอย่าง โดยระบุการแจกแจงความสามารถของผู้สอบเป็นแบบปกติ (คะแนนเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1) ทำการจำลองกลุ่มตัวอย่างเพื่อใช้เป็นกลุ่มอ้างอิงและกลุ่มสนใจ จำนวน 6 ขนาด ได้แก่ 100, 200, 400, 500, 800 และ 1000 คน

1.2 จำลองค่าพารามิเตอร์ของข้อสอบในกลุ่มอ้างอิง โดยกำหนดเงื่อนไขของวิธีการตรวจให้คะแนนเป็น 2 แบบ คือ การตรวจให้คะแนนรายการคำตอบแบบสองค่าจำนวน 25 ข้อ และการตรวจให้ให้คะแนนแบบหลายค่าจำนวน 25 ข้อ ซึ่งทั้งสองรูปแบบกำหนดให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบมีการกระจายแบบปกติ ในจำนวนข้อสอบ 25 ข้อนี้จะมีข้อสอบศึกษาอยู่ 5 ข้อ คือ ข้อที่ 5, 10, 15, 20 และ 25 ส่วนข้ออื่น ๆ ที่เหลือให้เป็นข้อสอบร่วม

1.3 จำลองรายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจ โดยนำค่าความสามารถของผู้สอบที่ได้ใน 1.1 ร่วมกับค่าพารามิเตอร์ของข้อสอบที่ได้ใน 1.2 เพื่อสร้างรายการคำตอบของผู้สอบตามความสามารถของผู้สอบและพารามิเตอร์ของข้อสอบ

2. ตัวแปรที่ศึกษา ประกอบด้วย

2.1 ตัวแปรอิสระมี 4 ตัวแปร ดังนี้

2.1.1 ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษามี 3 ระดับ คือ

- ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน
- การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง
- การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง

2.1.2 วิธีการตรวจให้คะแนน 2 แบบ คือ

- การตรวจให้คะแนนรายการคำตอบแบบสองค่า
- การตรวจให้คะแนนรายการคำตอบแบบหลายค่า

2.1.3 ขนาดกลุ่มตัวอย่างมี 4 ขนาด โดยแบ่งตามสัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิง 2 สัดส่วน คือ

- สัดส่วน 1:1 ดังนี้ 100:100, 200:200 400:400 และ 500:500 คน
- สัดส่วน 1:2 ดังนี้ 100:200, 200:400 400:800 และ 500:1000 คน

2.1.4 วิธีของการตรวจสอบความเท่าเทียมกันของการวัดมี 2 วิธี คือ

- วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน
- วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยใช้วิธีทดสอบอัตราส่วนความเป็นไปได้

2.2 ตัวแปรตาม คือ ผลการตรวจสอบความเท่าเทียมกันของการวัด ได้แก่ อัตราความถูกต้อง (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate)

นิยามศัพท์เฉพาะ

1. ความเท่าเทียมกันของการวัด (Measurement Equivalence) หมายถึง ความเท่ากันของค่าคาดหวังของคะแนนรวมของกลุ่มประชากรกลุ่มย่อยที่ต่างกัน บนระดับความสามารถเดียวกัน ทำการวัดด้วยแบบวัดคุณลักษณะเดียวกัน นั่นคือ ไม่ว่าจะอยู่ภายใต้เงื่อนไขที่แตกต่างกันหรือไม่ก็ตาม การสังเกตและศึกษาปรากฏการณ์ต่าง ๆ ของกลุ่มตัวอย่างศึกษาจะต้องดำเนินการวัดประเมินคุณลักษณะเดียวกัน แบบทดสอบหรือแบบทดสอบย่อยใด ๆ หากผู้ที่มีความสามารถเดียวกัน แล้วทำแบบทดสอบฉบับนั้น ๆ ได้คะแนนโครงสร้างตัวแปรแฝงเหมือนกัน มีคะแนนสอบที่คาดหวังเดียวกัน หรือคะแนนจริงที่ทั้งระดับข้อสอบหรือระดับผลรวมทั้งหมดของข้อสอบย่อย หรือทั้งสองอย่างเท่ากัน แบบทดสอบหรือแบบทดสอบย่อยนั้นจะได้รับการกล่าวว่ามีความเท่าเทียมกันของการวัดข้ามกลุ่มวัฒนธรรมหรือกลุ่มประชากร การตรวจสอบความเท่าเทียมกันของการวัดในการวิจัยครั้งนี้จะดำเนินการด้วยวิธีการทางสถิติตรวจสอบ 2 วิธี คือ

1.1 ทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน เป็นการตรวจสอบความไม่แปรเปลี่ยนข้ามกลุ่มของโมเดล โดยทำการการทดสอบความแตกต่างของค่าไค-สแควร์ แล้วพิจารณาค่านัยสำคัญของค่าความแตกต่างขององศาความเป็นอิสระ ซึ่งในการศึกษาครั้งนี้กำหนดระดับนัยสำคัญทางสถิติในการตรวจสอบไว้ที่ระดับ .05

1.2 ทฤษฎีการตอบข้อสอบใช้วิธีทดสอบอัตราส่วนความเป็นไปได้ ซึ่งเป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยเปรียบเทียบระหว่าง 2 โมเดล คือ โมเดลพื้นฐาน ซึ่งเป็นการจัดเตรียมค่าความเป็นไปได้พื้นฐานสำหรับความกลมกลืนของพารามิเตอร์ข้อสอบ

ในโมเดลเทียบกับโมเดลเปรียบเทียบที่จัดเตรียมค่าความเป็นไปได้ที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ข้อสอบข้อที่ i ในแต่ละกลุ่มประชากร แล้วพิจารณาค่านัยสำคัญของค่าความแตกต่างขององศาความเป็นอิสระ ซึ่งในการศึกษาครั้งนี้กำหนดระดับนัยสำคัญทางสถิติในการตรวจสอบไว้ที่ระดับ .05

2. กลุ่มอ้างอิง (Reference group: R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่ากลุ่มอื่น

3. กลุ่มสนใจ (Focal group: F) หมายถึง กลุ่มผู้สอบซึ่งเป็นเป้าหมายของการศึกษา ซึ่งคาดว่าจะเสียเปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่เบี่ยงเบน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่ากลุ่มอ้างอิง

4. ข้อสอบร่วม (Anchor items) หมายถึง ชุดของข้อสอบที่กำหนดให้ค่าพารามิเตอร์ข้อสอบ ได้แก่ ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยาก ในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากันทุกค่า ซึ่งในการศึกษาครั้งนี้กำหนดให้มีข้อสอบร่วม 20 ข้อ คือข้อที่ 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 21, 22, 23 และข้อที่ 24

5. ข้อสอบศึกษา (Studied items) หมายถึง ข้อสอบที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนบนกลุ่มอ้างอิง ซึ่งในการศึกษาครั้งนี้กำหนดให้มีข้อสอบศึกษา 5 ข้อ คือข้อที่ 5, 10, 15, 20 และข้อที่ 25 แต่ละข้อจะได้รับการระบุค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบในกลุ่มสนใจต่างไปจากกลุ่มอ้างอิง ดังนี้

5.1 ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ หมายถึง การระบุค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาในกลุ่มสนใจเท่ากับกลุ่มอ้างอิง

5.2 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก หมายถึง การระบุให้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ มีค่าพารามิเตอร์อำนาจจำแนกลดลง .15 และค่าพารามิเตอร์ความยากข้อสอบระบุให้ขึ้น .25 เมื่อเทียบกับค่าของกลุ่มอ้างอิง

5.3 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ หมายถึง การระบุให้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ มีค่าพารามิเตอร์อำนาจจำแนกลดลง .40 และค่าพารามิเตอร์ความยากข้อสอบระบุให้ขึ้น .50 เมื่อเทียบกับค่าของกลุ่มอ้างอิง

6. โมเดลกระชับ (Compact Model) หมายถึง โมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกกลุ่ม เป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม

7. โมเดลขยาย (Augmented Model) หมายถึง โมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกกลุ่ม และข้อสอบศึกษาที่นำเข้าไปศึกษาจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม

8. อัตราความถูกต้อง (Power rate) หมายถึง ร้อยละของจำนวนข้อสอบที่ตรวจสอบพบว่าข้อสอบศึกษามีความไม่เท่าเทียมกันของการวัดได้อย่างถูกต้อง โดยคำนวณจากค่าอัตราส่วนของจำนวนข้อสอบศึกษาที่ตรวจสอบได้ว่ามีความไม่เท่าเทียมกันของการวัดได้อย่างถูกต้องต่อจำนวนข้อสอบศึกษาทั้งหมดในแบบทดสอบ

9. อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง ร้อยละของจำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าข้อสอบศึกษาเกิดความไม่เท่าเทียมกันของการวัดทั้งที่ความเป็นจริงข้อสอบศึกษามีความเท่าเทียมกันของการวัด โดยคำนวณจากจำนวนข้อสอบที่ตรวจสอบผิดพลาดไม่เท่าเทียมกันของการวัดต่อจำนวนข้อสอบศึกษาที่มีความเท่าเทียมกันทั้งหมดในแบบทดสอบ

10. ขนาดกลุ่มตัวอย่าง หมายถึง จำนวนของผู้ตอบข้อคำถามทั้งกลุ่มอ้างอิงและกลุ่มสนใจที่ผู้วิจัยกำหนดขึ้นเพื่อใช้ในการวิจัย โดยกำหนดไว้ 6 ขนาด ดังนี้ 100, 200, 400, 500, 800 และ 1,000 คน แบ่งสัดส่วนระหว่างกลุ่มอ้างอิงต่อกลุ่มสนใจซึ่งมี 2 สัดส่วน คือ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100 200:200 400:400 และ 500:500 คน และสัดส่วน 1:2 มี 4 ขนาด คือ 100:200 200:400 400:800 และ 500:1000 คน

11. การตรวจให้คะแนน หมายถึง การกำหนดค่าคำตอบของผู้สอบจากการตอบข้อสอบ ซึ่งในการศึกษาครั้งนี้ กำหนดวิธีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้คะแนนรายการคำตอบแบบสองค่าและที่ตรวจให้คะแนนแบบหลายค่า

11.1 การตรวจให้คะแนนรายการคำตอบแบบสองค่า (Dichotomous scoring) หมายถึง การกำหนดคะแนนให้กับคำตอบของผู้ตอบข้อคำถามแต่ละข้อเป็น 1 และ 0 คะแนน

11.2 การตรวจให้คะแนนแบบหลายค่า (Polytomous scoring) หมายถึง การกำหนดคะแนนให้กับคำตอบของผู้ตอบข้อคำถามแต่ละข้อตามระดับของความถูกต้องเหมาะสม ในการศึกษาครั้งนี้ กำหนดให้แต่ละข้อคำถามมีคะแนนเป็น 4, 3, 2, 1 และ 0 คะแนน

กรอบแนวคิดในการวิจัย

การทดสอบเป็นกระบวนการหนึ่งของการวัดผล จุดมุ่งหมายสำคัญของการทดสอบเพื่อทำให้ผลการสอบมีความเท่าเทียมกัน ไม่ว่ากลุ่มผู้สอบจะมีความแตกต่างกันหรือไม่ก็ตาม ความเท่าเทียมกันของการวัดจะทำให้การสอบมีความยุติธรรม ซึ่งความยุติธรรมของแบบทดสอบนั้นจะต้องมีการแปลความหมายของคะแนนที่ได้จากแบบทดสอบของสมาชิกในกลุ่มได้ถูกต้องและมีความเสมอภาค โดยที่ผู้สอบทุกคนมีโอกาสแสดงศักยภาพของตนเองตามโครงสร้างของแบบทดสอบที่ใช้วัด การที่จะทำให้แบบทดสอบมีความยุติธรรมนั้น นักพัฒนาแบบทดสอบจำเป็นต้องตรวจสอบความเท่าเทียมกันของการวัด เพื่อแสดงหลักฐานพยานทางสถิติ แล้วทำการคัดข้อสอบที่ไม่มีความเท่าเทียมกันของการวัดออกจากแบบทดสอบ กระบวนการตรวจสอบดังกล่าวถือเป็นส่วนหนึ่งของกระบวนการพัฒนาแบบทดสอบที่มีความสำคัญไม่ยิ่งหย่อนไปกว่าการตรวจสอบความเที่ยงตรง ความเชื่อมั่น และคุณสมบัติอื่น ๆ ของข้อสอบเลย

วิธีการตรวจสอบความเท่าเทียมกันของการวัดมีมากมายหลายวิธี แต่ละวิธีก็มีจุดแข็ง ข้อจำกัด และมาจากฐานคิดทางทฤษฎีที่ต่างกัน แต่เป้าหมายของการตรวจสอบจะเหมือนกัน นั่นคือ ต้องการทำให้ผลการวัดมีความยุติธรรม ส่วนวิธีการใดจะเหมาะสมกับลักษณะของข้อมูลแบบใดนั้น เป็นหน้าที่ของนักทฤษฎีที่จะต้องทำการทดสอบ เพื่อหาพยานหลักฐานมาเป็นหลักประกันให้กับวิธีการต่าง ๆ ว่ามีความถูกต้องในการตรวจสอบ และมีความเชื่อถือได้มากน้อยเพียงใด ในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยใช้วิธีทดสอบอัตราส่วนความเป็นไปได้ภายใต้เงื่อนไขวิธีการตรวจให้คะแนน ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่าง โดยมีรายละเอียด ดังนี้

วิธีการตรวจให้คะแนน

ปัจจุบันการประเมินผลการเรียนนั้นได้มีการขยายไปสู่การประเมินตามสภาพจริง และมีแนวโน้มของการตรวจให้คะแนนรายการคำตอบแบบหลายค่ามากขึ้น ซึ่งราชูและคณะ (Raju; & et al. 2002) ได้เสนอความเหมือนและความแตกต่างของวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่อยู่บนสองฐานทฤษฎีไว้อย่างน่าสนใจ ดังนี้ **ความเหมือนของ CFA และ IRT เป็นการตรวจสอบความสัมพันธ์ระหว่างตัวแปรแฝงและชุดของตัวแปรที่สังเกตได้** เป็นการตรวจสอบคะแนนจริงในระดับข้อสอบหรือชุดแบบทดสอบย่อยสำหรับผู้สอบแต่ละคนในสองประชากรที่มีระดับความสามารถเดียวกัน นิยามของความเท่าเทียมกันของการวัดไม่ได้หมายความว่าการแจกแจงของคะแนนภายใต้โครงสร้างของสองประชากรที่เราสนใจจะมีค่าเท่ากัน และ**ความแตกต่างของ CFA และ IRT คือ ความสัมพันธ์ระหว่างตัวแปรโครงสร้างและคะแนนจริง** CFA จะมีความสัมพันธ์เชิงเส้นตรง ส่วน IRT จะมีความสัมพันธ์ไม่เป็นเชิงเส้นตรง **วิธีการตรวจให้คะแนนการตอบข้อสอบ** CFA จะรองรับการตรวจให้คะแนนรายการคำตอบหลายค่าได้ดีกว่า ส่วน IRT จะเหมาะกับการตรวจให้คะแนนรายการคำตอบแบบสองค่า และ**ด้านข้อคำถาม** CFA รองรับตัวแปรแฝงหลายตัวและประชากรหลายกลุ่ม ส่วน IRT เหมาะกับแบบสอบถามที่มีลักษณะความเป็นมิติเดียว จะเห็นว่าวิธีการตรวจให้คะแนนก็ยังเป็นประเด็นที่ทำให้เกิดความแตกต่างใน 2 ฐานทฤษฎี ดังนั้นผู้วิจัยจึงนำวิธีการตรวจให้คะแนน 2 แบบ คือการตรวจให้คะแนนรายการคำตอบแบบสองค่า และการตรวจให้คะแนนรายการคำตอบแบบหลายค่า มาเป็นเงื่อนไขหนึ่งของการตรวจสอบความเท่าเทียมกัน

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ

โคเฮน คิม และ เบคเกอร์ (Cohen; Kim; & Baker. 1993) ตั้งระดับของการทำหน้าที่เบี่ยงเบนของข้อสอบในการศึกษาเป็น .5 และ 1.0 ข้อสอบส่วนใหญ่ที่ทำหน้าที่เบี่ยงเบนจะตรวจสอบได้

ออร์ต (Oort. 1998) ได้ศึกษาใน 3 ระดับของการทำหน้าที่เบี่ยงเบนของข้อสอบ คือ ระดับต่ำ (.2) ระดับปานกลาง (.5) และสูง (.8) พบว่า ข้อสอบที่มีระดับการทำหน้าที่เบี่ยงเบนในระดับปานกลาง และระดับสูงส่วนใหญ่จะตรวจสอบได้ แต่ในระดับการทำหน้าที่เบี่ยงเบนที่ต่ำจะไม่สามารถตรวจพบได้

จากที่กล่าวมาจะเห็นว่าการทำหน้าที่เบี่ยงเบนของข้อสอบส่วนมากสามารถตรวจสอบสมบูรณ์เมื่อพื้นที่ระหว่าง ICCs ไม่ต่ำกว่า .5 แต่เพื่อเป็นการหาประจักษ์พยานให้กว้างขึ้น ผู้วิจัยจึงกำหนดระดับของการทำหน้าที่เบี่ยงเบนของข้อสอบเป็น 3 ขนาด คือ **ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ** โดยกำหนดให้ค่าพารามิเตอร์อำนาจจำแนกและค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน **การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก** ในกลุ่มสนใจเมื่อเทียบกับค่าของกลุ่มอ้างอิงจะลดค่าพารามิเตอร์อำนาจจำแนกของข้อสอบศึกษาลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาข้อนั้นขึ้น .25 และ **การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่** ในกลุ่มสนใจเมื่อเทียบกับกลุ่มอ้างอิงจะลดค่าพารามิเตอร์อำนาจจำแนกของข้อสอบศึกษาลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาข้อนั้นขึ้น .5

ขนาดตัวอย่าง

การตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบ ระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ขนาดตัวอย่างถือว่ามีความสำคัญในการทดสอบนัยสำคัญทางสถิติ เพราะวิธีการตรวจสอบความเท่าเทียมกันของการวัดโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันอยู่บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ใช้สถิติการทดสอบภายใต้การแจกแจงแบบเชิงเส้น แต่วิธีทดสอบอัตราส่วนความเป็นไปได้อยู่บนพื้นฐานของทฤษฎีการตอบข้อสอบ และใช้สถิติการทดสอบภายใต้การแจกแจงแบบเชิงเส้นกำกับ ซึ่งการแจกแจงดังกล่าวจำเป็นต้องใช้ขนาดตัวอย่างที่เพียงพอเพื่อให้ผลการวิเคราะห์มีความถูกต้อง

จากการศึกษาวิจัยของฮูลิน ลิสแสค และ ดราสกอว์ (Hulin; Lissak; & Drasgow. 1982) ชี้ให้เห็นถึงจำนวนข้อสอบที่มากนั้นไม่มีความจำเป็นเท่ากับผู้สอบจำนวนมาก โดยได้เสนอแนะสำหรับโมเดล 2 พารามิเตอร์ว่าความยาวของแบบทดสอบควรเป็น 30 ข้อที่ทดสอบกับขนาดกลุ่มตัวอย่าง 500 คน และความยาวของแบบทดสอบจำนวน 50 ข้อกับขนาดของกลุ่มตัวอย่าง 1,000 คน ก็เพียงพอสำหรับโมเดล 3 พารามิเตอร์

สวามินาธาน และกิฟฟอร์ด (Swaminathan; & Gifford. 1983) ศึกษาการประมาณค่าของพารามิเตอร์ในโมเดล 3 พารามิเตอร์ ใช้ขนาดกลุ่มตัวอย่าง 50, 200 และ 1,000 คน พบว่าขนาดกลุ่มตัวอย่างที่เพิ่มขึ้นส่งผลน้อยมากต่อความแม่นยำในการประมาณค่าพารามิเตอร์ความยากและพารามิเตอร์การเดา แต่การเพิ่มขึ้นขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ สามารถทำให้มีความแม่นยำในการประมาณค่าพารามิเตอร์ เช่น ข้อสอบ 20 ข้อ ทดสอบผู้สอบ 1,000 คน

สามารถประมาณค่าพารามิเตอร์ความยากและพารามิเตอร์การเดาได้เป็นอย่างดี และสำหรับข้อสอบ 80 ข้อที่ทดสอบกับกลุ่มตัวอย่าง 1,000 คน สามารถประมาณค่าพารามิเตอร์ได้ดีทุกค่า

รี และเจนเซน (Ree; & Jensen. 1983) ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างในการปรับเทียบพารามิเตอร์ข้อสอบที่เป็นเชิงเส้น การปรับเทียบใช้กับแบบทดสอบหลายฟอร์ม โดยศึกษาจากการจำลองข้อมูล ที่ขนาดกลุ่มตัวอย่างเป็น 250, 500, 1,000 และ 2,000 คน ผลการศึกษาพบว่า กลุ่มตัวอย่างขนาดใหญ่ ความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ข้อสอบน้อย การประมาณค่าพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยากมีความถูกต้องและเชื่อถือได้ สำหรับการประมาณค่าพารามิเตอร์การเดานั้นต้องใช้ผู้สอบที่มีระดับความสามารถต่ำเป็นจำนวนมาก

ฟินช์ และ เฟรนช์ (Finch; & French. 2006) ศึกษาขนาดกลุ่มตัวอย่าง โดยใช้สัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิงแบบ 1:1 และ 1:2 จำนวน 3 ขนาด คือ 250:250 250:500 คน และ 500:1000 คน ระดับความสามารถค่าเฉลี่ย 2 ระดับคือ ระดับความสามารถเฉลี่ยเท่ากัน (M=0, SD=1) และระดับความสามารถเฉลี่ยกลุ่มอ้างอิงสูงกว่ากลุ่มสนใจ (M_R=0.5, M_F=0) ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชยการตอบ 3 ขนาด คือ 0% 10% และ 20% กระทำภายใต้โมเดลโลจิสติก 2 และ 3 พารามิเตอร์ แต่ละเงื่อนไขทำซ้ำ 1,000 ครั้ง ผลในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชยคำตอบใน 4 วิธีการ พบว่า ไม่มีตัวแปรหรือปฏิสัมพันธ์ร่วมใดที่มีนัยสำคัญทางสถิติ แสดงว่าระดับของเงื่อนไขวิธีการไม่มีนัยสำคัญทางสถิติอัตราความคลาดเคลื่อนประเภทที่ 1 ขนาดกลุ่มตัวอย่างและความสามารถของกลุ่มไม่ส่งผลต่อการประมาณค่าพารามิเตอร์ ส่วนการตรวจสอบอัตราความถูกต้อง พบว่า ความแปรปรวนของขนาดกลุ่มตัวอย่างส่งผลต่ออัตราความถูกต้อง ในวิธีการถดถอยโลจิสติก (LR) วิธีการทดสอบอัตราส่วนความเป็นไปได้ตามทฤษฎีการตอบข้อสอบ (LR) และวิธีซิปเทสท์ (SIBTEST) อธิบายสัดส่วนของความแปรปรวนมากที่สุด ส่วนวิธีการวิเคราะห์องค์ประกอบ (CFA) ไม่มีนัยสำคัญทางสถิติ วิธีการทดสอบอัตราส่วนความเป็นไปได้ (LR) มีปฏิสัมพันธ์ระหว่างโมเดลการประมาณค่าพารามิเตอร์และความแตกต่างของความสามารถอย่างมีนัยสำคัญทางสถิติ และพบว่าความสามารถของกลุ่มที่แตกต่างกันไม่กระทบต่ออัตราความถูกต้อง

ในการศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษาขนาดตัวอย่าง 4 ขนาด โดยใช้สัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิงแบบ 2 สัดส่วน คือ สัดส่วน 1:1 คือ 100:100, 200:200 400:400 และ 500:500 คน และสัดส่วน 1:2 คือ 100:200, 200:400 400:800 และ 500:1000 คน

สัดส่วนของข้อสอบร่วมและข้อสอบศึกษา

ความยาวของข้อสอบและจำนวนสัดส่วนของข้อสอบศึกษาต่อข้อสอบร่วมนั้น จะมีผลกระทบต่อความถูกต้องในการจับคู่เปรียบเทียบระหว่างกลุ่มผู้สอบ เนื่องจากการวิเคราะห์ความเท่าเทียมกันของการวัดจะใช้คะแนนรวมจากการสอบแทนคุณลักษณะภายในหรือระดับความสามารถของผู้สอบที่วัดได้ แบบทดสอบที่มีความยาวมากกว่าย่อมส่งผลให้มีความน่าเชื่อถือ

มากกว่า ซึ่งจะทำให้การจับคู่ระหว่างกลุ่มผู้สอบมีความถูกต้องมากขึ้น แต่ถ้าในแบบทดสอบมีข้อสอบทำหน้าที่เบี่ยงเบนปะปนอยู่จะมีผลทำให้ค่าประมาณความสามารถมีความเชื่อมั่นต่ำลง จะทำให้เกณฑ์การจับคู่ขาดความแม่นยำ ซึ่งทำให้อัตราความถูกต้องลดลง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าปกติ จากผลการศึกษาของ มอลเลอร์ (Maller. 2001) ได้ให้ข้อสังเกตว่าเศษหนึ่งส่วนสามโดยประมาณของข้อสอบใน Wechsler Intelligence Scale for Children-Third-Edition แสดงการทำหน้าที่เบี่ยงเบน โดยมีบางข้อเข้าข้างเด็กผู้ชาย และบางข้อเข้าข้างเด็กผู้หญิง ส่วนบัดเจลล์ ราชู และควอเท็ต (Budgell; Raju; & Quartett. 1995) ได้ทบทวนการศึกษาการทำหน้าที่เบี่ยงเบนในเครื่องมือการวัดประเมิน พบว่า จำนวนของข้อสอบที่มีการทำหน้าที่เบี่ยงเบนอยู่ในช่วงจาก 1.5% ถึง 64% พร้อมกับมีการศึกษาหลายครั้งที่พบมีการทำหน้าที่เบี่ยงเบน เกินกว่า 30% ของข้อสอบ ส่วนฮวง เชิร์ช และคาติกแบ็ก (Haung; Church; & Katigbak. 1977; citing Su; & Wang. 2005) ได้ทำการตรวจสอบการวัดที่เท่าเทียมกันของข้อสอบที่วัดข้ามวัฒนธรรมของ NEO Personality Inventory ซึ่งเป็นการวัดโมเดลบุคลิกภาพ 5 องค์ประกอบ ในบริบททางวัฒนธรรมที่ต่างกัน โดยใช้เครื่องมือเป็นภาษาอังกฤษ เกือบจะ 40% จากข้อสอบ 180 ข้อ ที่พบว่า แสดงการทำหน้าที่เบี่ยงเบน จะเห็นว่าแบบทดสอบที่ใช้กัน (Real test) อาจมีจำนวนข้อสอบที่ทำหน้าที่เบี่ยงเบนอยู่ในระดับที่สูง และจากผลการศึกษาของโรเจอร์สและสวามินาธาน (Rogers; & Swaminathan. 1993) พบว่า เมื่อใช้จำนวนสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ระดับคือ 0% และ 15% จำนวนสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนดังกล่าวไม่มีผลต่ออัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ยกเว้นในกรณีการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกัน และเมื่อจำนวนสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกมีค่าลดลง ส่วนนารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1994) ได้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่เป็นรูปแบบเดียวกันด้วยวิธีซิบเทสท์ และวิธีการถดถอยโลจิสติก โดยศึกษาปัจจัยสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 2 ระดับ คือ จำนวน 10% และ 20% พบว่าเมื่อสัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีซิบเทสท์และวิธีการถดถอยโลจิสติกมีค่าลดลง นอกจากนี้ยังส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีมีค่าลดลงด้วย ต่อมานารายานานและสวามินาธาน (Narayanan; & Swaminathan. 1996) ได้ตรวจสอบอีกครั้งโดยตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่เป็นรูปแบบเดียวกันด้วยวิธีแมนเทล-แฮนส์เซล วิธีโคร-ซิบ และวิธีการถดถอยโลจิสติก โดยศึกษาปัจจัยสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบน 3 ระดับ คือ จำนวน 0%, 10% และ 20% พบว่า เมื่อสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะมีผลทำให้อัตราความถูกต้องของวิธีการถดถอยโลจิสติกมีค่าลดลงแต่จะไม่มีผลต่ออัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซล และวิธีโคร-ซิบ สำหรับอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่า

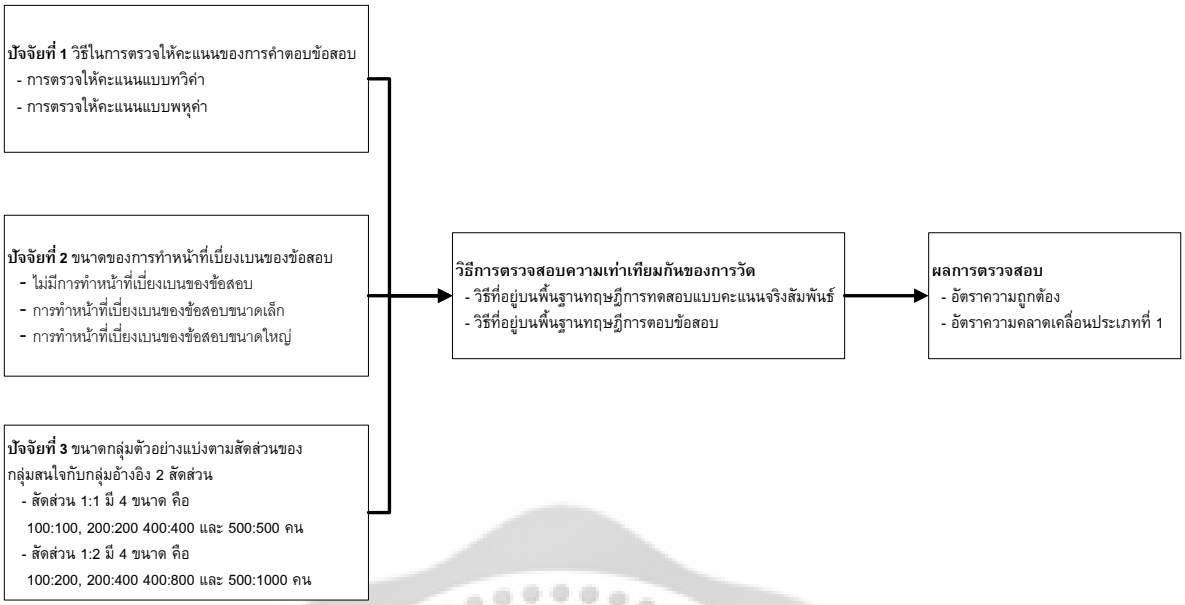
เมื่อจำนวนสัดส่วนของข้อสอบที่ทำหน้าที่เบี่ยงเบนในแบบทดสอบมีจำนวนเพิ่มขึ้นจะส่งผลให้อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสามวิธีจะมีค่าเพิ่มขึ้นด้วย โดยวิธีแมนเทล-แฮนส์เซล มีค่าต่ำสุด

สวามินาธานและโรเจอร์ส (Swaminathan; & Rogers. 1990) ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธี แมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ที่มีความยาวของแบบทดสอบ 40, 60 และ 80 ข้อ โดยใช้ข้อมูลจำลองพบว่า เมื่อใช้แบบทดสอบที่ยาวมากขึ้น อัตราความถูกต้องของทั้งสองวิธีมีค่ามากขึ้น ยกเว้นกรณีข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของวิธีแมนเทล-แฮนส์เซล ต่อมาทั้งสองคนได้ศึกษาอีกครั้ง (Rogers; & Swaminathan. 1993) โดยใช้แบบทดสอบที่มีความยาว 40 และ 80 ข้อ ตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก พบว่าความยาวของแบบทดสอบไม่มีผลต่ออัตราความถูกต้องของวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอย โลจิสติก ยกเว้นในกรณีข้อสอบทำหน้าที่เบี่ยงเบนที่ไม่เป็นรูปแบบเดียวกันของวิธีการถดถอย โลจิสติก ซึ่งให้ผลการศึกษาขัดแย้งกับครั้งแรก

โคเฮ็นและคิม (Cohen; & Kim. 1993) ใช้ความยาวของแบบทดสอบ 20 และ 60 ข้อ ในการตรวจสอบการทำหน้าที่ของข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด และวิธีการวัดพื้นที่ชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมายของราชู พบว่าเมื่อเพิ่มความยาวของแบบทดสอบแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 จะเพิ่มขึ้น ในขณะที่ผลการศึกษาของอุตตาโรและมิลล์แซบ (Uttaro; & Millsap. 1994) พบว่าเมื่อใช้แบบทดสอบที่มีความยาวมากขึ้นแล้วอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีแมนเทล-แฮนส์เซล จะลดลง

จากที่กล่าวมาจะเห็นว่ายังไม่มีความเป็นทิศทางเดียวกันของจำนวนข้อสอบ แต่ถ้าในแบบทดสอบมีข้อสอบทำหน้าที่เบี่ยงเบนปะปนอยู่นั่นหมายถึงมีสัดส่วนของข้อสอบศึกษาต่อข้อสอบร่วม ซึ่งขนาดของสัดส่วนนี้จะมีผลถึงความเชื่อมั่นในการประมาณความสามารถ จะทำให้เกณฑ์การจับคู่ขาดความแม่นยำ ซึ่งทำให้อัตราความถูกต้องในการตรวจสอบลดลง หรืออัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าปกติ ดังนั้น ในการศึกษาครั้งนี้ ผู้วิจัยจึงกำหนดให้ความยาวของข้อสอบคงที่ โดยใช้ข้อสอบจำนวน 25 ข้อ ในจำนวนนี้ให้เป็นข้อสอบร่วม 20 ข้อ และข้อสอบศึกษา 5 ข้อ คิดเป็นสัดส่วนข้อสอบศึกษาต่อข้อสอบร่วมเท่ากับ 1:4 หรือเท่ากับ 20%

จากเงื่อนไขต่าง ๆ ดังกล่าวมาข้างต้น ผู้วิจัยจึงเขียนกรอบแนวคิดในการวิจัย ดังนี้



ภาพประกอบ 1 กรอบแนวคิดการวิจัย

สมมติฐานในการวิจัย

จากกรอบแนวคิดในการวิจัยจะเห็นว่า การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และทฤษฎีการตอบข้อสอบ ภายใต้เงื่อนไขวิธีการตรวจให้คะแนน ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ขนาดกลุ่มตัวอย่างนั้น ทำให้แต่ละวิธีมีข้อได้เปรียบและข้อเสียเปรียบแตกต่างกัน ผู้วิจัยจึงตั้งสมมติฐาน ดังนี้

1. ข้อสอบที่เป็นการตรวจให้คะแนนรายการคำตอบแบบสองค่าภายใต้เงื่อนไขที่แปรเปลี่ยน 2 เงื่อนไข ประกอบด้วยขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่าง วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบจะให้ผลการตรวจสอบความเท่าเทียมกันของการวัดแตกต่างกัน

2. ข้อสอบที่เป็นการตรวจให้คะแนนรายการคำตอบแบบหลายค่าภายใต้เงื่อนไขที่แปรเปลี่ยน 2 เงื่อนไข ประกอบด้วยขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่าง วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบจะให้ผลการตรวจสอบความเท่าเทียมกันของการวัดแตกต่างกัน

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้ เป็นการศึกษาการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ เพื่อนำองค์ความรู้ที่ได้จากการศึกษาค้นคว้าไปประยุกต์ใช้ในเชิงปฏิบัติต่อไป แนวทางของการศึกษาค้นคว้าครั้งนี้ กำหนดเป็นกรอบในการศึกษา 4 ประเด็น ดังนี้

1. ทฤษฎีการทดสอบ
2. ความเท่าเทียมกันของการวัด
3. การตรวจสอบความเท่าเทียมกันของการวัด
4. เอกสารและงานวิจัยที่เกี่ยวข้อง

1. ทฤษฎีการทดสอบ

ทฤษฎีการทดสอบ เป็นองค์ความรู้เกี่ยวกับการทดสอบ วิธีการแก้ปัญหา การทดสอบและพัฒนาเครื่องมือการทดสอบ ทฤษฎีการทดสอบได้มีการพัฒนาขึ้นมาตามแนวคิดที่นักทฤษฎีพิจารณาว่าสมเหตุสมผล สำหรับสนองความต้องการในการวัดคุณลักษณะเฉพาะต่าง ๆ ของบุคคล ซึ่งมีทฤษฎีการทดสอบมาตรฐานเดิม หรือที่เรียกแบบภาษาอังกฤษว่า Classical Test Theory (CTT) ใช้เป็นพื้นฐานของทฤษฎีการวัดมานานกว่า 80 ปี และทฤษฎีที่มีการพัฒนามาอย่างต่อเนื่องอีกทฤษฎีหนึ่ง คือ ทฤษฎีการตอบข้อสอบหรือที่เรียกแบบภาษาอังกฤษว่า Item Response Theory (IRT) ซึ่งเป็นทฤษฎีการทดสอบแนวใหม่ ทฤษฎีการทดสอบแต่ละทฤษฎีมีรายละเอียด ดังนี้

1.1 ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory)

ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory: CTT) เป็นทฤษฎีที่เชื่อว่าการวัดใด ๆ จะมีผลกระทบจากความคลาดเคลื่อนเสมอ (Lord; & Novick. 1968) ดังนั้นคะแนนสอบ X (Observed Score) จึงประกอบด้วยคะแนนจริง (True Score) และคะแนนคลาดเคลื่อน (Error Score) จากความเชื่อเบื้องต้น สามารถเขียนเป็นสมการมูลฐาน ดังนี้ (บุญเชิด ภิญโญอนันตพงษ์. 2549)

$$X_i = T_i + E_i$$

นั่นคือ คะแนนที่สังเกตได้หรือวัดได้จากแบบทดสอบ = คะแนนจริง + คะแนนความคลาดเคลื่อน ซึ่งคะแนนจริงของบุคคลคนหนึ่งสามารถหาได้โดยการใช้คะแนนเฉลี่ยที่คน ๆ นั้น ได้จากการทดสอบฉบับเดิมมีจำนวนครั้งของการทดสอบที่ไม่สิ้นสุด

ข้อตกลงเบื้องต้น (Assumptions)

ทฤษฎีการทดสอบมาตรฐานเดิม ตั้งอยู่บนพื้นฐานของความเชื่อตามข้อตกลงเบื้องต้นเกี่ยวกับโมเดลการวัด และข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบคู่ขนาน (Allen; & Yen. 1979; Linn. 1989) นั่น คือ

1. คะแนนที่สังเกตได้ (Observed score) หรือคะแนนที่ได้จากการวัดของผู้สอบแต่ละคน เป็นผลรวมของคะแนนจริง (True score) และคะแนนความคลาดเคลื่อน (Error score) ของผู้สอบนั้น $X_i = T_i + E_i$

2. คะแนนจริงเป็นค่าคาดหวังหรือค่าเฉลี่ยประชากรของคะแนนสังเกตได้ E(X)=T หรือคะแนนจริง T เป็นค่าเฉลี่ยของการแจกแจงตามทฤษฎีของคะแนนสังเกตได้ (X) ที่บุคคลนั้นทำการสอบในแบบทดสอบฉบับหนึ่งหลาย ๆ ครั้ง โดยการสอบแต่ละครั้งเป็นอิสระต่อกัน เช่น เด็กชายเม่น ทำแบบทดสอบฉบับหนึ่งเป็นจำนวน ∞ ครั้ง แต่ละครั้งเป็นอิสระต่อกัน ปรากฏว่าค่าเฉลี่ยของคะแนนสังเกตได้ดังกล่าวเท่ากัน 16 แสดงว่าเด็กชายเม่นมีคะแนนจริงเท่ากับ 16 คะแนนจริง T แสดงในรูปสมการคณิตศาสตร์ ดังนี้ (Lord; & Novick. 1968)

$$T = \lim_{k \to \infty} \left[\sum_{i=1}^{k} \frac{X_i}{k} \right]$$

ในทางปฏิบัติที่เป็นจริง การที่จะให้ผู้สอบแต่ละคนทำแบบทดสอบฉบับหนึ่งจำนวนหลาย ๆ ครั้งเป็นสิ่งที่ทำได้ยาก และยิ่งจะต้องดำเนินการสอบให้การสอบแต่ละครั้งเป็นอิสระต่อกัน เป็นเรื่องที่เป็นไปได้ยากยิ่งขึ้น อีกทั้งคะแนนจริงที่ได้จะสะท้อนถึงความสามารถตามทฤษฎีของบุคคลนั้นเพียงใด ยังขึ้นอยู่กับความเที่ยงตรง (Validity) ของแบบทดสอบด้วย

3. คะแนนความคลาดเคลื่อน และคะแนนจริงที่ประชากรผู้สอบทำแบบทดสอบฉบับหนึ่งไม่มีความสัมพันธ์กัน

$$\rho_{ET} = 0$$

ข้อตกลงเบื้องต้นนี้อาจถูกฝ่าฝืนได้ ถ้าผู้สอบมีคะแนนจริงต่ำ การลอกคำตอบจากผู้สอบที่มีคะแนนจริงสูง หรือในการเรียนการสอนในชั้นเรียน ถ้าครูลำเอียงในขณะดำเนินการสอบ ให้เด็กอ่อนมานั่งสอบข้างหน้า และให้เด็กเก่งไปนั่งสอบข้างหลังห้อง ซึ่งอาจได้ยินคำสั่งของครูไม่ชัดเจน ในสถานการณ์อย่างนี้ก็จะส่งผลให้คะแนนความคลาดเคลื่อนมีความสัมพันธ์ทางลบกับคะแนนจริง

4. คะแนนความคลาดเคลื่อนในการทำแบบทดสอบฉบับหนึ่ง (E1) และคะแนนความคลาดเคลื่อนในการทำแบบทดสอบอีกฉบับหนึ่ง (E2) ไม่มีความสัมพันธ์กัน

$$\rho_{E_1E_2} = 0$$

คะแนนความคลาดเคลื่อนของแบบทดสอบสองฉบับที่แตกต่างกันไม่มีความสัมพันธ์กัน กล่าวคือ ผู้สอบคนหนึ่งมีคะแนนความคลาดเคลื่อนเป็นทางบวกในการทำ

แบบทดสอบฉบับหนึ่ง และอาจมีความคลาดเคลื่อนในทางบวกหรือทางลบในการทำแบบทดสอบอีกฉบับหนึ่งก็ได้

ข้อตกลงเบื้องต้นนี้ จะไม่สมเหตุสมผล ถ้าคะแนนการสอบได้รับอิทธิพลจากปัจจัยอื่น ๆ เช่น ความเมื่อยล้า ความคุ้นเคย อารมณ์ สิ่งแวดล้อม เป็นต้น

5. คะแนนความคลาดเคลื่อนในแบบทดสอบฉบับที่หนึ่ง (E1) ไม่มีความสัมพันธ์กับคะแนนจริงในแบบทดสอบฉบับที่สอง (T2)

$$\rho_{E_1T_2} = 0$$

ข้อตกลงเบื้องต้น อาจถูกฝ่าฝืนได้ ถ้าแบบทดสอบฉบับที่สองวัดคุณลักษณะด้านบุคลิกภาพหรือมิติความสามารถที่มีอิทธิพลต่อการเกิดความคลาดเคลื่อนในแบบทดสอบฉบับที่หนึ่ง และข้อตกลงเบื้องต้นดังกล่าวข้างต้นเป็นการนิยามความคลาดเคลื่อนของการวัดตามทฤษฎีการทดสอบแบบมาตรฐานเดิม กล่าวคือ ความคลาดเคลื่อนเป็นความคลาดเคลื่อนแบบสุ่ม (Random error) ภายใต้ทฤษฎีมาตรฐานเดิมนี้ยังแบ่งการทดสอบออกเป็น 3 ระดับ ตามระดับความคู่ขนานหรือความคล้ายคลึงกันของคะแนน ได้แก่ การทดสอบแบบคะแนนจริงเท่ากัน (Parallel Test) การทดสอบแบบคะแนนจริงสมมูล (Essentially Tau-Equivalent Test) และการทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test)

การทดสอบแบบคะแนนจริงเท่ากัน (Parallel Test)

แบบทดสอบฉบับ A และ B เป็นแบบทดสอบคู่ขนาน (Parallel tests) ถ้าคะแนนที่สังเกตได้จากแบบทดสอบ A และ B เป็น X_1 และ X_2 ที่เป็นไปตามข้อตกลงเบื้องต้นของ 2-5 เมื่อ $X_1 = T_1 + E_1$ และ $X_2 = T_2 + E_2$ แล้วสำหรับประชากร ผู้สอบทุกคน จะได้ $T_1 = T_2$ และ $\sigma^2_{E_1} = \sigma^2_{E_2}$

การทดสอบแบบคะแนนจริงสมมูล (Essentially Tau-Equivalent Test)

การทดสอบแบบคะแนนจริงสมมูล (Essentially Tau-Equivalent Test) แบบทดสอบฉบับ A และ B เป็น Essentially τ - equivalent *Test* ถ้าคะแนนที่สังเกตได้จากแบบทดสอบ A และ B เป็น X_1 และ X_2 ที่เป็นไปตามข้อตกลงเบื้องต้นข้อ 2-5 และเมื่อ $X_1 = T_1 + E_1$ และ $X_2 = T_2 + E_2$ แล้วสำหรับประชากรผู้สอบทุกคน จะได้ $T_1 = T_2 + C_{12}$ เมื่อ C_{12} เป็นค่าคงที่

การทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test)

การทดสอบแบบคะแนนจริงสัมพันธ์ หมายถึง การวัดที่วัดโครงสร้างแฝงเดียวกัน โดยที่หน่วยการวัดอาจแตกต่างกันและมีความแม่นยำไม่เท่ากัน ความแปรปรวนคลาดเคลื่อนไม่เท่ากัน (บุญเชิด ภิญโญอนันตพงษ์. 2549: 17-21) ซึ่งทฤษฎีทดสอบคะแนนจริงสัมพันธ์เป็นรูปทั่วไปของทฤษฎีทดสอบมาตรฐานเดิม ที่มีกรอบแนวคิดของคะแนนสอบ (x) ประกอบด้วย $aT + b + E$ และใช้ข้อตกลงของระดับความคู่ขนานหรือความคล้ายคลึงกันของคะแนนที่ผ่อนปรนมากที่สุด คือ คะแนนจริง (T) ของตัวแปรแฝงเดียวกันสองค่าใดๆ มีสหสัมพันธ์เป็น 1.0 หรือ $T_i = a_{ij}T_j + b_{ij}$ มีแบบจำลองคะแนนที่วัดองค์ประกอบเดียว และหลายองค์ประกอบ มีวิธีประมาณค่า

ความเชื่อมั่นจากค่าความแปรปรวน-ความแปรปรวนร่วม และจากค่าน้ำหนักองค์ประกอบของการทดสอบ ทั้งที่วัดมิติเดียวและหลายมิติ รวมทั้งวิธีปรับเทียบคะแนน ซึ่งมีประโยชน์และคุณค่าในการนำไปใช้ในเชิงปฏิบัติ

แบบจำลองคะแนนจริงสัมพันธ์

กรอบแนวคิดของแบบจำลองคะแนนจริงสัมพันธ์ (Jöreskog. 1971) กำหนดคะแนนสอบมีส่วนประกอบดังนี้

$$X_i = a_i T_i + b_i + E_i$$

เมื่อ a_i เป็นค่าอำนาจจำแนกของข้อสอบ

b_i เป็นค่าความยากของข้อสอบที่วัดโครงสร้าง T_i

คะแนนสอบสองค่าใด ๆ X_i และ X_j คะแนนจริงของมัน T_i และ T_j จะมีความสัมพันธ์เชิงเส้นตรง เมื่อหาค่า a_{ij} และ b_{ij} ที่เหมาะสมได้ ทำให้

$$T_i = a_{ij} T_j + b_{ij} \ (1 \le i, j \le k)$$

คุณสมบัตินี้อาจเขียนเป็นค่าสหสัมพันธ์คือ $\rho(T_i, T_j) = 1.0$ จากลักษณะสำคัญของ Congeneric นี้จึงเรียกว่าคะแนนจริงสัมพันธ์ทั่วไปหรือคะแนนจริงสัมพันธ์

ถ้ากำหนดให้ $j = 1$ และ $i > 1$ ค่าคงที่จะเป็น a_{i1} และ b_{i1} จะได้

$$T_i = a_{i1} T_1 + b_{i1}$$

ถ้า $a_i = a_j = 1, b_i = b_j = 1$ และ $Var(E_i) = Var(E_j)$ จะได้แบบจำลองคะแนนจริงเท่ากันคือ

$$T_1 = T_2$$

ถ้า $a_i = a_j = 1, b_i \neq b_j \neq 0$ และ $Var(E_i) \neq Var(E_j)$ จะได้แบบจำลองคะแนนจริงสมมูลคือ

$$T_1 = T_2 + b_{12}$$

ถ้า $a_i \neq a_j, b_i \neq b_j \neq 0$ และ $Var(E_i) \neq Var(E_j)$ จะได้แบบจำลองคะแนนจริงสัมพันธ์คือ

$$T_1 = a_{12} T_2 + b_{12}$$

ถ้ากำหนดให้ τ เป็นตัวแปรสุ่ม และ $T_i(T_1, T_2, ..., T_k)$ ในสมการ $T = T_1 + T_2 + ... + T_k = \sum T_i$ แต่ละค่ามีความสัมพันธ์เชิงเส้นตรงกับ τ นั่นคือ

$$T_i = \lambda_i \tau + \mu_i$$

โดย λ_i และ μ_i เป็นค่าคงที่ซึ่งมีข้อตกลงว่า $\varepsilon(\tau) = 0$ และ $Var(\tau) = 1$ นำค่า T_i สมการ $T_i = \lambda_i \tau + \mu_i$ แทนค่าในสมการ $X_i = T_i + E_i$ จะได้

$$X_i = \lambda_i \tau + \mu_i + E_i$$

ถ้าให้ X, λ, μ และ E เป็นเวกเตอร์หลัก (Column Vector) สมการ $X_i = \lambda_i \tau + \mu_i + E_i$ สามารถเขียนในรูปของเมทริกซ์ได้ดังนี้

$$X = \lambda\tau + \mu + E$$

ให้ $\sum$ เป็นเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของ X และให้ θ^2 เป็นความแปรปรวนรวมคลาดเคลื่อนและเป็นเมทริกซ์ทะแยง $\sum = \lambda\lambda$

จากที่กล่าวมาจะเห็นว่า ทฤษฎีทดสอบมาตรฐานเดิม เป็นทฤษฎีที่มีการพัฒนาเพิ่มเติมอย่างต่อเนื่อง เพื่อให้เหมาะสมกับความก้าวหน้าทางวิชาการในปัจจุบัน และสามารถนำไปใช้ปฏิบัติได้อย่างมีประสิทธิภาพ และในงานวิจัยครั้งนี้เลือกใช้การทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test) สำหรับทำการตรวจสอบความเท่าเทียมกันของการวัด ทั้งนี้เนื่องจากการทดสอบแบบคะแนนจริงสัมพันธ์เป็นลักษณะทั่วไปของการทดสอบแบบมาตรฐานเดิม และมีการผ่อนปรนในเงื่อนไขข้อตกลงเบื้องต้นของการทดสอบแบบมาตรฐานเดิมมากที่สุดและสอดคล้องกับวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน

1.2 ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการตอบข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล (Latent trait or ability) กับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) ทฤษฎีการตอบข้อสอบจึงอยู่บนฐานความคิดที่สำคัญ 2 ประการคือ 1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบสามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบ และ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) โดยโมเดล

ฟังก์ชันโลจิสติกดังกล่าว สามารถจำแนกได้เป็น 2 ประเภทตามลักษณะการตรวจให้คะแนนคำตอบ ได้แก่ โมเดลการตอบข้อสอบแบบตรวจให้คะแนนแบบสองค่า (Dichotomous) และโมเดลการตอบข้อสอบ แบบตรวจให้คะแนนแบบหลายค่า (Polytomous)

1.2.1 หลักการของทฤษฎีการตอบข้อสอบ

ทฤษฎีการทดสอบมาตรฐานเดิม (CTT) เป็นที่รู้จักกันเป็นอย่างดีในวงการวัดผลทางการศึกษาและจิตวิทยามาเป็นเวลานาน แนวคิดจากทฤษฎีการทดสอบมาตรฐานเดิมได้รับการนำไปใช้อย่างกว้างขวางในการสร้างข้อสอบ วิเคราะห์ข้อสอบ และตรวจสอบคุณภาพ สำหรับการพัฒนาแบบทดสอบต่าง ๆ ไม่ว่าจะเป็นแบบทดสอบวัดผลสัมฤทธิ์ แบบวัดความถนัด แบบวัดบุคลิกภาพ ฯลฯ ด้วยข้อจำกัดที่สำคัญของทฤษฎีการทดสอบมาตรฐานเดิม ได้แก่ การที่พารามิเตอร์ของข้อสอบผันแปรตามกลุ่มผู้สอบ และคะแนนที่สังเกตได้หรือค่าประมาณความสามารถของผู้ตอบไม่เป็นอิสระหรือขึ้นอยู่กับข้อสอบและแบบทดสอบที่นำมาใช้ นักทฤษฎีการทดสอบหลายท่านจึงได้พัฒนาระบบการวัดแนวใหม่ ซึ่งทฤษฎีการทดสอบสำหรับการวัดแนวใหม่อันเป็นที่ยอมรับกันอย่างกว้างขวางในปัจจุบัน คือ ทฤษฎีการตอบข้อสอบ (Item Response Theory) หรือที่เรียกโดยใช้ชื่อย่อว่า IRT

ทฤษฎีการตอบข้อสอบ พยายามอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด ทฤษฎีการตอบข้อสอบมีพื้นฐานความเชื่อว่า พฤติกรรมการตอบข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิดจะถูกกำหนดโดยลักษณะภายใน หรือความสามารถที่อยู่ภายในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้ ได้อธิบายความสัมพันธ์ดังกล่าวในรูปของฟังก์ชันคณิตศาสตร์ หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ คุณลักษณะของข้อสอบ และโอกาสการตอบข้อสอบได้ถูก

1.2.2 แนวคิดพื้นฐานของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบเป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคลกับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) ทฤษฎีการตอบข้อสอบจึงอยู่บนฐานความคิดที่สำคัญ 2 ประการ คือ

1) ผลการตอบข้อสอบ หรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ในตัวของผู้ตอบ

2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในตัวผู้สอบ สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ

1.2.3 ข้อตกลงของทฤษฎีการตอบข้อสอบ

ฟังก์ชันทางคณิตศาสตร์ที่ใช้ในทฤษฎีการตอบข้อสอบ กำหนดไว้ว่า ความน่าจะเป็นของการตอบข้อสอบต้องขึ้นอยู่กับความสามารถของผู้สอบและขึ้นอยู่กับคุณลักษณะของข้อสอบ

โมเดลการตอบข้อสอบจึงตั้งอยู่บนฐานความเชื่อหรือข้อตกลงเบื้องต้นหลายประการเกี่ยวกับลักษณะข้อมูลที่จะทำให้โมเดลสามารถนำไปใช้ได้อย่างเหมาะสม ถึงแม้ข้อตกลงเบื้องต้นบางประการจะตรวจสอบไม่ได้โดยตรง แต่เราสามารถเก็บรวบรวมข้อมูลหรือหลักฐานทางอ้อมมาช่วยสนับสนุนยืนยันได้ ข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการตอบข้อสอบมีดังนี้ (Lord; & Novick. 1968; Hambleton; & Swaminatan. 1985)

1. ความเป็นมิติเดียว (Unidimensionality: One trait)

ข้อตกลงเบื้องต้นที่ใช้กันทั่วไปสำหรับทฤษฎีการตอบข้อสอบ คือ ข้อคำถามจะมุ่งวัดเพียงคุณลักษณะเดียว หรือความสามารถเดียว (One ability) ถ้าความสามาถของคนคนหนึ่งมี k ด้าน ซึ่งความสามารถแต่ละด้านต่างก็ส่งผลต่อการตอบข้อสอบข้อต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลที่ได้จากการตอบแบบทดสอบ (Test Performance) หรือคะแนนของผู้ตอบสามารถอธิบายได้เพียงความสามารถเดียว จะเรียกว่าความเป็นมิติเดียว (Unidimensionality) อย่างไรก็ตามข้อตกลงเบื้องต้นนี้ไม่ได้เข้มงวดนัก ถ้าแบบทดสอบมีลักษณะเด่น (Dominant) ที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ถือเสมือนว่ามีความเป็นมิติเดียว ซึ่งการทดสอบความเป็นมิติเดียวของแบบทดสอบสามารถทำได้หลายวิธี วิธีการหนึ่งที่นิยมใช้ในการตรวจสอบ คือ การวิเคราะห์องค์ประกอบ (Factor Analysis) โดยพิจารณาจากค่า Eigen Value ว่าให้ค่าแตกต่างไปจากข้ออื่น ๆ หรือไม่

2. ความเป็นอิสระ (Independence: Local Independent)

แนวคิดเกี่ยวกับ “ความเป็นอิสระระหว่างข้อสอบและผู้สอบ” มีความเกี่ยวข้องมาจาก “ความเป็นมิติเดียวของแบบทดสอบ” ความเป็นอิสระของข้อสอบและผู้สอบ หมายถึง ความน่าจะเป็นหรือโอกาส (Probability) ในการตอบข้อสอบได้ถูกต้องเป็นอิสระจากกัน นั่นคือ เมื่อมีการควบคุมความสามารถ (θ) ที่ส่งผลต่อการตอบข้อสอบ หรือให้ความสามารถของผู้สอบ (θ) คงที่แล้ว ผลการตอบข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน หรือกล่าวได้อีกนัยหนึ่งว่า เมื่อควบคุมอิทธิพลของความสามารถ (θ) แล้วผลการตอบข้อสอบรายข้อไม่มีความสัมพันธ์กัน นั่นคือ โมเดลการตอบสนองมีเพียงความสามารถ (θ) ปัจจัยเดียวเท่านั้นที่มีผลต่อการตอบรายข้อ ความเป็นอิสระสามารถจำแนกพิจารณาเป็นความอิสระระหว่างข้อสอบและความอิสระระหว่างผู้สอบดังนี้

1) ความเป็นอิสระระหว่างข้อสอบ นั่นคือ เมื่อสุ่มผู้สอบ ซึ่งมีความสามารถ θ ขึ้นมา 1 คน ในการตอบข้อสอบ k ข้อ ให้ U_i เป็นผลการตอบหรือคะแนนข้อที่ i หลังจากควบคุมความสามารถ (θ) ของผู้สอบแล้ว คะแนนผลการตอบของผู้สอบคนนั้น ๆ ในแต่ละข้อไม่สัมพันธ์กัน ถ้าผลการตอบรายข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อที่ k ของผู้สอบที่มีความสามารถ θ จะเท่ากับ ผลคูณระหว่างความน่าจะเป็นของผลการตอบข้อสอบแต่ละข้อ

2) ความเป็นอิสระระหว่างผู้สอบ นั่นคือ เมื่อสุ่มข้อสอบขึ้นมา 1 ข้อ ในการตอบข้อสอบของผู้สอบ n คน ให้เป็นผลการตอบหรือคะแนนข้อสอบของผู้สอบคนที่ i หลังจาก

ควบคุมความสามารถ (θ) ของผู้สอบแต่ละคนแล้ว คะแนนผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน ถ้าผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อสอบของผู้สอบ n คน จะเท่ากับ ผลคูณระหว่างความน่าจะเป็นของผลการตอบข้อนั้นของผู้สอบแต่ละคน

อย่างไรก็ตาม แฮมเบิลตัน และสวามินาธาน (Hambleton; & Swaminatan. 1985) กล่าวว่า ถ้าแบบทดสอบมีความเป็นมิติเดียวกันแล้ว ความเป็นอิสระในการตอบข้อสอบก็จะเกิดขึ้นตามไปด้วย

3. โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve: Item Response Model)

โค้งคุณลักษณะข้อสอบเป็นฟังก์ชันทางคณิตศาสตร์สามารถใช้อธิบายความสัมพันธ์ระหว่างความน่าจะเป็น หรือโอกาสที่ผู้ตอบจะตอบข้อสอบได้ถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบ หรือแบบทดสอบฉบับนั้น ทั้งนี้ ความน่าจะเป็นหรือโอกาสในการจะตอบข้อสอบถูกจะขึ้นอยู่กับโค้งคุณลักษณะข้อสอบในแต่ละโมเดลที่เลือกใช้ โดยที่รูปร่าง (Shape) ของโค้งคุณลักษณะข้อสอบ ในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่างที่ใช้ ดังนั้น จึงทำให้ความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย คุณสมบัตินี้ถือเป็นลักษณะเด่นของโมเดลต่าง ๆ ในทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบมีหลายรูปแบบขึ้นอยู่กับว่าเลือกใช้โมเดลพารามิเตอร์ข้อสอบกี่ตัว

4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว

ทฤษฎีการตอบข้อสอบ (IRT) ถือว่าความสามารถ (θ) เป็นปัจจัยสำคัญต่อผลการสอบ ความเร็วในการตอบจะต้องไม่มีอิทธิพลต่อผลการตอบ การจัดการสอบจึงต้องไม่อยู่ในสถานการณ์ที่สอบแข่งขันกันด้วยเวลา การสอบจะต้องอยู่ในลักษณะที่ผู้สอบซึ่งมีความสามารถมีเวลาเพียงพอในการทำข้อสอบ

1.2.4 พารามิเตอร์ (Parameter) ในทฤษฎีการตอบข้อสอบ

พารามิเตอร์ในทฤษฎีการตอบข้อสอบ แบ่งออกเป็น 2 ชนิด คือ พารามิเตอร์ของผู้ตอบ (Person Parameter) ได้แก่ ระดับความสามารถหรือคุณลักษณะของผู้ตอบ (θ_k หรือ β_k) ส่วนพารามิเตอร์ข้อสอบ (Item Parameter) ได้แก่ ค่าความยาก (b_i) อำนาจจำแนก (a_i) สัมประสิทธิ์การเดา (c_i) และค่าความรอบคอบ (γ_i) นอกจากค่าพารามิเตอร์แล้วในทฤษฎีการตอบข้อสอบยังมีค่าคงที่ที่อยู่ในสมการอีก ซึ่งพิสัยของค่าพารามิเตอร์และค่าคงที่ต่าง ๆ มีดังนี้

พารามิเตอร์ของผู้สอบ

พารามิเตอร์ของผู้สอบ ประกอบด้วย θ คือ ระดับความสามาถของผู้สอบประมาณจากโมเดลทฤษฎีการตอบข้อสอบ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง ส่วนมากจะมีค่าอยู่ระหว่าง -3.0 ถึง +3.0 ค่าที่เป็นลบแสดงว่ามีความสามารถหรือระดับคุณลักษณะนั้นต่ำ และค่าที่เป็นบวกแสดงว่ามีค่าความสามารถหรือระดับคุณลักษณะนั้นสูง

พารามิเตอร์ของข้อสอบ

พารามิเตอร์ของข้อสอบ ประกอบด้วย

b_i คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i (Difficulty parameter) เป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ ณ จุดความสามาถของผู้สอบ θ ที่มีโอกาสตอบข้อสอบได้ถูก ในทางทฤษฎี b_i จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าระหว่าง -2.0 ถึง +2.0 ค่า b_i ที่เป็นลบแสดงว่าข้อสอบข้อนั้นง่าย ส่วนค่า b_i ที่เป็นบวกแสดงว่าข้อสอบข้อนั้นยาก

a_i คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination parameter) เป็นค่าความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i ในทางทฤษฎี a_i จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าระหว่าง 0 ถึง +2.0 เพราะค่า a_i ที่เป็นลบ แสดงว่า เป็นข้อสอบที่ไม่ดีต้องนำออกจากชุดข้อสอบนั้น ค่า a_i เป็น 0 แสดงว่าข้อสอบนั้นไม่มีอำนาจในการจำแนก หรือมีอำนาจในการจำแนกต่ำ ส่วนค่า a_i เป็น +2.0 แสดงว่าข้อสอบมีอำนาจจำแนกสูง

c_i คือ สัมประสิทธิ์การเดา เป็นค่าแสดงความน่าจะเป็นหรือโอกาสในการเดาข้อสอบข้อที่ i (Guessing parameter) ได้ถูกต้อง ซึ่งโอกาสตอบข้อสอบได้ถูกของผู้สอบที่มีความสามารถต่ำจะเป็นค่ากำกับต่ำสุด (Lower asymptote) ของ ICC ในทางทฤษฎีมีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่า c_i ไม่เกิน 0.30 ตามปกติควรมีค่าต่ำกว่าโอกาสในการตอบถูกโดยการเดาตามทฤษฎี CTT

γ_i คือ ความรอบคอบ เป็นค่าที่แสดงถึงความรอบคอบในการตอบข้อสอบแต่ละข้อ ซึ่งค่าความรอบรอบนี้จะมีค่าน้อยกว่า 1

ค่าคงที่

ค่าคงที่ ประกอบด้วย exp หรือ e เป็นค่าคงที่ของลอกาลิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.71828 และ D คือ ค่าองค์ประกอบของการปรับสเกล (Scaling factor) กำหนดให้เป็น 1 หรือ 1.70 เพื่อทำให้ฟังก์ชันโลจิสติก (Logistic function) มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) หรือมีค่าต่างกันไม่เกิน 0.01

1.2.5 โมเดลการตอบข้อสอบแบบตรวจให้คะแนนแบบสองค่า

โมเดลการตอบข้อสอบที่ตรวจให้คะแนนแบบสองค่า เป็นการตรวจให้คะแนนรายข้อสองค่า เช่น การตรวจให้คะแนนแบบ 0 หรือ 1 ถูกหรือผิด เห็นด้วยหรือไม่เห็นด้วย เป็นต้น ฟังก์ชันตอบข้อสอบสามารถที่จะนำมาใช้ศึกษาความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูก $[P_i(\theta)]$ กับระดับความสามารถของผู้สอบที่วัดได้โดยแบบทดสอบฉบับนั้น (θ) เมื่อนำมาเขียนเป็นกราฟจะได้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งโค้งลักษณะข้อสอบนี้จะมีหลายลักษณะขึ้นอยู่กับโมเดล (Model) หรือแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว โมเดลที่นิยมใช้กันได้แก่ โมเดลแบบ 1 พารามิเตอร์ (One-Parameter Model

or Rasch Model) โมเดลแบบ 2 พารามิเตอร์ (Two-Parameter Model) และโมเดลแบบ 3 พารามิเตอร์ (Three-Parameter Model) (Hambleton; Swaminathan; & Roger. 1991) โมเดลเหล่านี้เหมาะสำหรับข้อมูลการตอบข้อสอบแบบตรวจให้คะแนนแบบสองค่า ประกอบด้วยค่าพารามิเตอร์และค่าคงที่ ดังนี้

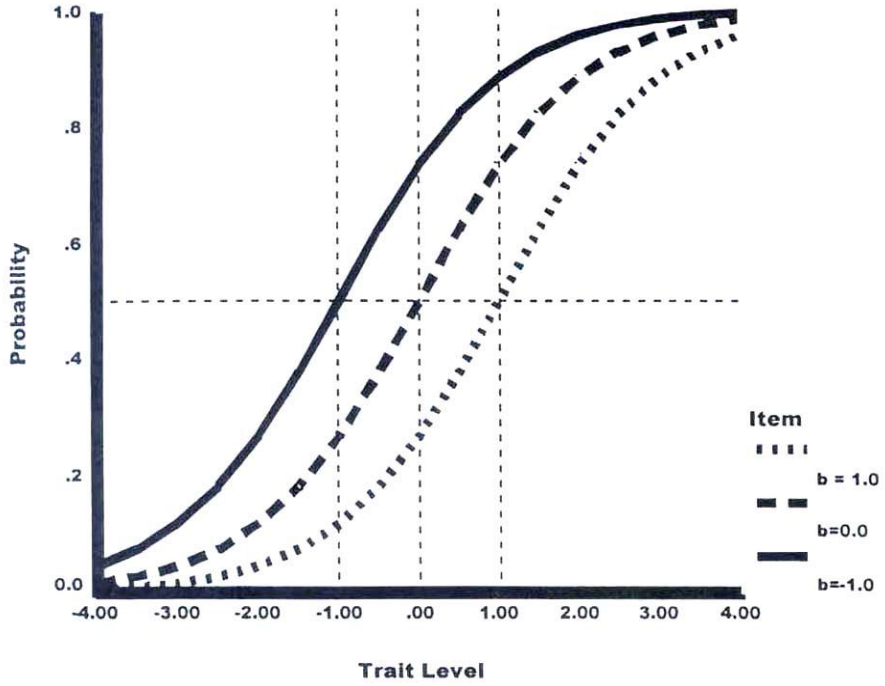
1) โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (One-parameter Logistic Model)

โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (One-parameter Logistic Model) เป็นโมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1,2,...n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง โดยทั่วไปแล้วค่าความสามารถของผู้สอบจะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ถึงแม้ว่าความสามารถจะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ผลการวิเคราะห์ส่วนใหญ่มักจะให้ค่าอยู่ในช่วง -4 ถึง +4

b_i parameter คือ พารามิเตอร์ความยากของข้อสอบ ซึ่งในทางทฤษฎี b_i จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าระหว่าง -2.0 ถึง +2.0 ค่า b_i ที่เป็นลบแสดงว่าข้อสอบข้อนั้นง่าย ส่วนค่า b_i ที่เป็นบวกแสดงว่าข้อสอบข้อนั้นยาก ค่าของความยากของข้อสอบจะแสดงตำแหน่งบนโค้งลักษณะข้อสอบ ณ จุดที่ผู้ตอบที่มีความสามารถ θ ที่มีโอกาสตอบข้อสอบถูก 0.50 ดังนั้น พารามิเตอร์ความยากของข้อสอบที่เพิ่มมากขึ้น ผู้สอบที่มีความสามารถมากขึ้นที่จะตอบข้อสอบข้อนั้นได้ถูก ตำแหน่งของโค้งลักษณะข้อสอบบนสเกลความสามารถ จะอยู่ในตำแหน่งที่สูงกว่า ซึ่งโค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์ มีลักษณะ ดังนี้



ภาพประกอบ 2 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์

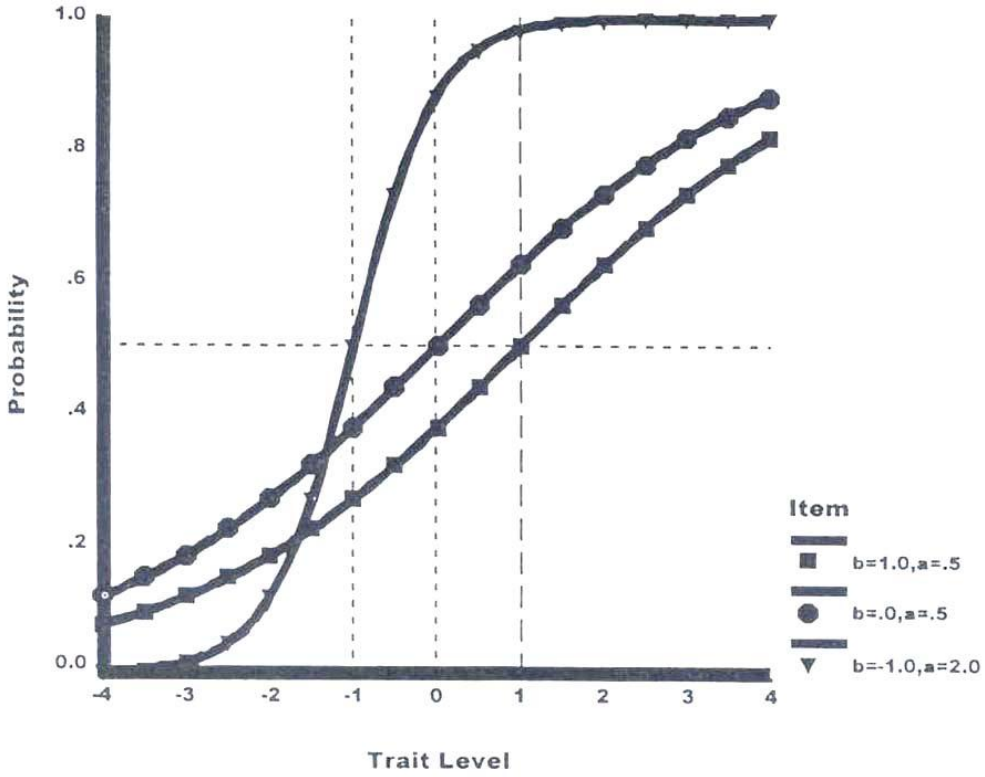
ที่มา: Embretson; & Reise. (2000). *Item response theory for psychologists*. p.68-72.

2) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two-parameter Logistic Model)

โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้ (Birnbaum. 1968)

$$P_i(\theta) = \frac{e^{Dai(\theta - b_i)}}{1 + e^{Dai(\theta - b_i)}} \quad i = 1,2,...n$$

โมเดลโลจิสติกแบบ 2 พารามิเตอร์ เมื่อเทียบกับโมเดล 1 พารามิเตอร์แล้ว จะมีตัวแปรเพิ่มเข้าไปอีก 2 ตัว คือ D เป็นค่าคงที่ ซึ่งโดยปกติแล้วจะกำหนดให้มีค่าเท่ากับ 1.7 หรือ 1.702 (Hulin; Drasgow; & Parsons. 1983) สเกลค่าคงที่นี้ได้เพิ่มเข้าไปเพื่อที่จะทำให้รูปร่างของฟังก์ชันโลจิสติกสอดคล้องหรือใกล้เคียงกับฟังก์ชันปกติสะสม ค่าคงที่นี้ไม่ได้ทำให้รูปแบบพื้นฐานของโมเดลโลจิสติกเปลี่ยนไป ซึ่งโค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์ มีลักษณะ ดังนี้



ภาพประกอบ 3 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์

ที่มา: Embretson; & Reise. (2000). *Item response theory for psychologists*. p.68-72.

a_i (a_i Parameter) คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination parameter) ซึ่งจะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ค่า a_i แสดงถึงความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i หากข้อสอบที่มีค่า a_i ติดลบจะถูกตัดออกจากแบบทดสอบวัดความสามารถ เพราะว่ามีสิ่งผิดปกติบางอย่างในตัวข้อสอบ กล่าวคือ ผู้สอบที่ความสามารถเพิ่มขึ้นมีความน่าจะเป็นที่จะตอบข้อสอบถูกน้อยลง และหากค่า a_i มากกว่า 2 ถือว่าเป็นเรื่องผิดปกติ ดังนั้น ค่า a_i ที่ควรเป็นควรจะมีพิสัยอยู่ระหว่าง 0 ถึง 2

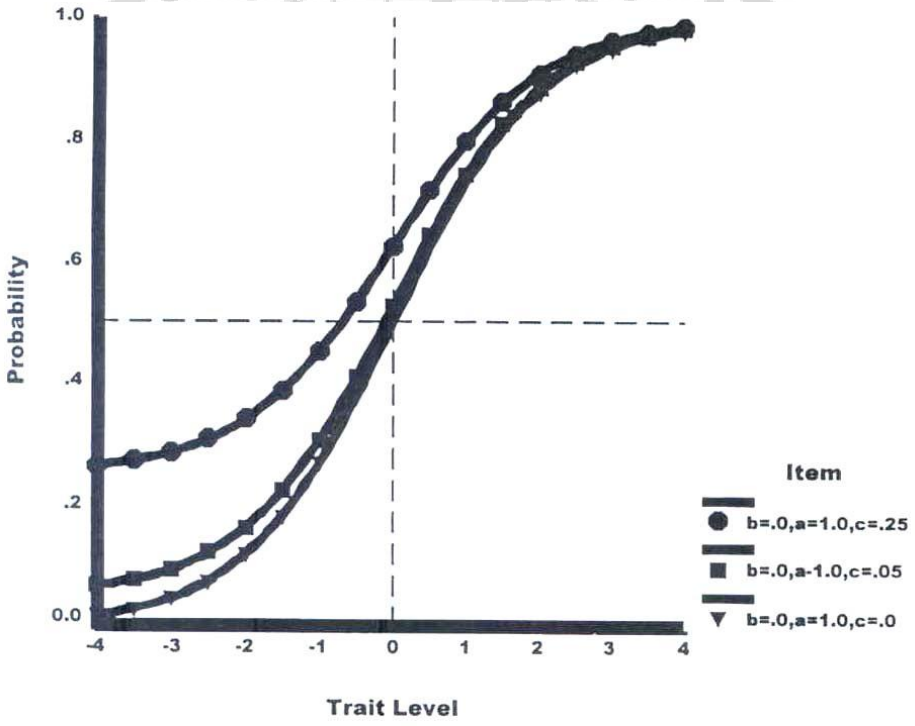
โมเดล 1 และ 2 พารามิเตอร์ จะถือว่าไม่มีพฤติกรรมการเดาในการตอบข้อสอบได้ถูกต้อง ซึ่งข้อตกลงของพฤติกรรมการเดาในการตอบข้อสอบได้ถูกต้องนี้เป็นสิ่งที่ไม่เกิดกับข้อสอบชนิดให้ตอบอย่างอิสระหรือข้อสอบชนิดเขียนตอบ แต่พฤติกรรมการเดาในการตอบข้อสอบถูกต้องมักจะพบอยู่บ้างในการสอบด้วยแบบทดสอบชนิดเลือกตอบ เมื่อแบบทดสอบนั้นไม่ยุ่งยากมากนักในการเลือกตอบของผู้สอบ

3) โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three-parameter Logistic Model)

โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad i = 1,2,..., n$$

โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ได้เพิ่มพารามิเตอร์ c_i (c_i Parameter) เข้าไปในโมเดล 2 พารามิเตอร์ ซึ่งพารามิเตอร์นี้เรียกว่า พารามิเตอร์โอกาสการเดาตอบข้อสอบได้ถูก (Pseudo-chance level parameter) หรือพารามิเตอร์การเดา (Guessing parameter) ซึ่งเป็นค่าที่ต่ำกว่าค่ากำกับของโค้งลักษณะข้อสอบ หรือความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำ (θ เป็น $-\infty$) ที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง และมีความสัมพันธ์ระหว่างพารามิเตอร์การตอบถูกกับระดับความสามารถต่ำ ซึ่งโค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์มีลักษณะ ดังนี้



ภาพประกอบ 4 โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์

ที่มา: Embretson; & Reise. (2000). *Item response theory for psychologists*. p.68-72.

1.2.6 โมเดลการตอบข้อสอบแบบตรวจให้คะแนนแบบหลายค่า

โมเดลการตอบข้อสอบที่ตรวจให้คะแนนแบบหลายค่า เป็นโมเดลความสัมพันธ์ระหว่างความสามารถของผู้ตอบกับโอกาสของการเลือกตอบแต่ละรายการคำตอบที่กำหนดให้ ซึ่งไม่เป็นลักษณะเชิงเส้นตรง ความสัมพันธ์ดังกล่าวนำไปสู่การคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ และเมื่อนำค่าที่ได้มารวมกัน ณ ตำแหน่งความสามารถเดียวกัน จะทำให้ได้ค่าฟังก์ชันสารสนเทศของแบบทดสอบ การคำนวณค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ กระทำในลักษณะเดียวกับโมเดลการตอบข้อสอบที่ตรวจให้คะแนนแบบสองค่า

สำหรับโมเดลการตอบข้อสอบที่ตรวจให้คะแนนแบบหลายค่า ได้มีผู้พัฒนาโมเดลทั้งบนพื้นฐานของโมเดลราสช์ (Rasch Model) หรือโมเดลแบบ 1 พารามิเตอร์ ได้แก่ Partial Credit Model (PCM) และ Rating Scale Model (RSM) โมเดลกลุ่มนี้เหมาะสำหรับข้อสอบหรือข้อคำถามที่แต่ละข้อมีอำนาจจำแนกเท่ากัน และโมเดลที่พัฒนาบนพื้นฐานของโมเดลแบบ 2 พารามิเตอร์ ซึ่งได้แก่ Graded-Response Model (GRM) Modified Graded-Response Model (M-GRM) Generalized Partial Credit Model (G-PCM) และ Nominal Response Model (NRM) โมเดลกลุ่มนี้เหมาะสำหรับข้อสอบหรือข้อคำถามที่มีค่าความยาก และค่าอำนาจจำแนกที่แตกต่างกัน ในการศึกษาครั้งนี้ผู้วิจัยได้นำ Graded-Response Model (GRM) (Samejima. 1969) ในการประมาณค่าพารามิเตอร์ของข้อสอบและประมาณค่าพารามิเตอร์ความสามารถผู้สอบกับข้อสอบที่เป็นการตรวจให้คะแนนแบบหลายค่า ทั้งนี้เพราะโมเดลดังกล่าวมีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ (2-Parameter Model) และไม่เข้มงวดกับข้อตกลงของความเป็นมิติเดียวของมาตรวัด และการประมาณค่าพารามิเตอร์มีค่าอำนาจจำแนกรายข้อเป็นฟังก์ชันด้วย รายละเอียดของโมเดล มีดังนี้

Graded-Response Model

ซามิจิมา (Samejima. 1969, 1996) ได้พัฒนา Graded-Response Model (GRM) สำหรับใช้กับแบบทดสอบ หรือแบบวัดที่แต่ละข้อคำถามมีรายการคำตอบเป็นแบบมาตรเรียงลำดับ (Ordered categorical response) โดยแต่ละข้ออาจมีรายการคำตอบที่แตกต่างกันได้ และวิธีการตรวจให้คะแนนความรู้บางส่วน แต่ละข้อมีจำนวนลำดับขั้นของการให้คะแนนแตกต่างกัน

Graded-Response Model มีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ (2-Parameter Model) และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละรายการคำตอบแบบ 2 ขั้นตอน (Indirect IRT Model) โดยขั้นแรกจะคำนวณค่าความชันร่วมของแต่ละข้อคำถาม จากนั้นจึงทำการคำนวณค่าพารามิเตอร์ของแต่ละรายการคำตอบในแต่ละข้อคำถาม การวิเคราะห์ตามโมเดล GRM มีเป้าหมายเพื่อประมาณค่า a_i และตำแหน่งของ b_{ki} ของผู้ตอบที่มีคุณลักษณะ (θ) บนสเกลที่ต่อเนื่องกัน โดยใช้สูตร ดังนี้

$$P_{ix}^{*}(\theta)=\frac{exp[\,D\,a_i(\theta_s-b_{ki})]}{1+exp[\,D\,a_i(\theta_s-b_{ki})]}$$

เมื่อ b_{ki} คือ ช่วง หรือ Threshold ระหว่างรายการคำตอบ k และ k+1 ของข้อ i; a_i คือ ความชัน หรือค่าพารามิเตอร์อำนาจจำแนกของข้อคำถาม และ θ_s คือ ค่าพารามิเตอร์ความสามารถของผู้ตอบ สมการนี้ยังอ้างถึงการตอบสองช่วง (BRF) โดยการตอบช่วงคล้ายกับโมเดล 2 พารามิเตอร์ที่ตรวจให้คะแนนแบบสองค่า ยกเว้น m-1 ฟังก์ชันที่ต้องการต่อข้อคำถาม (m=จำนวนของรายการคำตอบในแต่ละข้อคำถาม)

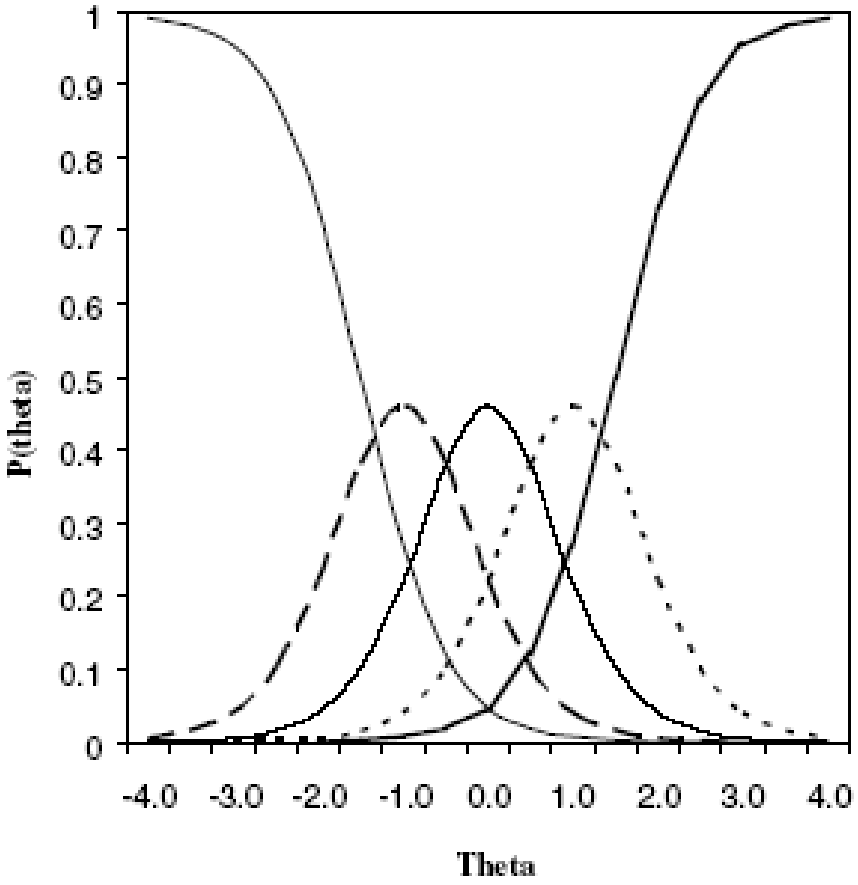
ในการคำนวณความน่าจะเป็นในการตอบแต่ละรายการคำตอบ โดยใช้การลบค่าความน่าจะเป็น ดังสมการ

$$P_{ix}(\theta)=P_{ix}^{*}(\theta)-P_{i(x+1)}^{*}(\theta)$$

เมื่อพิจารณาจากข้อเท็จจริง ความน่าจะเป็นของการเลือกตอบรายการคำตอบต่ำสุด หรือเหนือกว่าจะมีค่าเป็น 1.00 นั่นคือ $P_{i0}^{*}(\theta)=1.00$ และความน่าจะเป็นของการเลือกตอบเหนือกว่ารายการคำตอบสูงสุดจะมีค่าเป็น 0 ดังนั้น $P_{ik}^{*}(\theta)=0.00$ เช่น หากข้อคำถามมีรายการคำตอบ 5 ค่า ความน่าจะเป็นของการเลือกตอบแต่ละรายการคำตอบของผู้ที่มีคุณลักษณะ (θ) จะเป็น ดังนี้

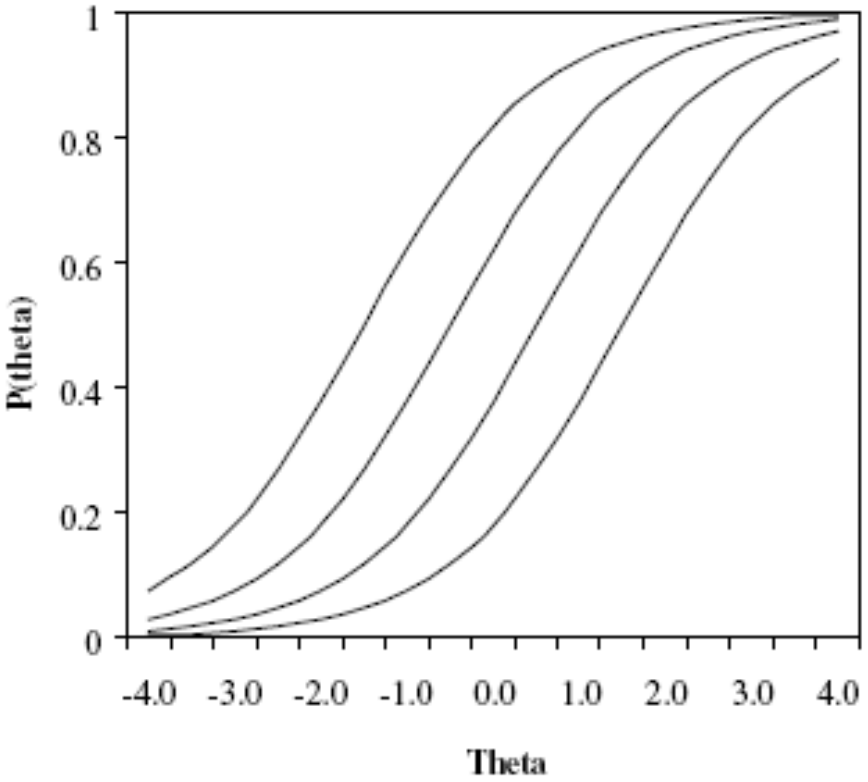
$$P_{i0}(\theta)=1.00-P_{i1}^{*}(\theta)$$
$$P_{i1}(\theta)=P_{i1}^{*}(\theta)-P_{i2}^{*}(\theta)$$
$$P_{i2}(\theta)=P_{i2}^{*}(\theta)-P_{i3}^{*}(\theta)$$
$$P_{i4}(\theta)=P_{i3}^{*}(\theta)-P_{i4}^{*}(\theta)$$
$$P_{i5}(\theta)=P_{i4}^{*}(\theta)-0$$

โค้งแสดงฟังก์ชันของความน่าจะเป็นในการเลือกรายการคำตอบต่าง ๆ ของผู้ที่มีคุณลักษณะ θ เรียกว่า โค้งรายการคำตอบ (Category Response Curves: CRC) ลักษณะของโค้งรายการคำตอบมีความเกี่ยวพันธ์กับโค้งลักษณะปฏิบัติการ (Operating Characteristic Curves: OCC) และค่าความชันร่วมของข้อคำถาม (α_i)



ภาพประกอบ 5 โค้งรายการคำตอบ (Category Response Curves: CRC)

ที่มา: Meade; & Lautenschlager (2004). *Organizational Research Methods*. 7; p.365.



ภาพประกอบ 6 โค้งลักษณะปฏิบัติการ (Operating Characteristic Curves: OCC)

ที่มา: Meade; & Lautenschlager (2004). *Organizational Research Methods*. 2004; 7; p.366.

2. ความเท่าเทียมกันของการวัด (Measurement Equivalence)

ความเท่าเทียมกันของการวัด

การศึกษาลักษณะทางจิตวิทยาย่อมมีวัฒนธรรมกลุ่มชนปนอยู่ด้วยเสมอ การวิจัยข้ามวัฒนธรรมจึงต้องแสดงให้เห็นว่าความแตกต่างระหว่างวัฒนธรรมได้สะท้อนความแตกต่างที่แท้จริงของคุณลักษณะที่ศึกษา ไม่ใช่เกิดจากการลำเอียงหรืออคติของผู้วิจัย เช่น การใช้วิธีการทฤษฎี หรือเครื่องมือของโลกตะวันตก หรือเกิดจากการขาดความเท่าเทียมกัน (Equivalence) ในกระบวนการวัด หรือเกิดจากสิ่งอื่นที่มนุษย์สร้างขึ้น (ผจงจิต อินทสุวรรณ. 2544; อ้างอิงจาก Van de Vijver; & Poortinga. 1982)

ในช่วงทศวรรษที่ผ่านมานี้ มีงานวิจัยข้ามวัฒนธรรมในทางจิตวิทยาเพิ่มขึ้น การศึกษาข้ามชาติพันธุ์ ข้ามวัฒนธรรม และความหลากหลายทางภาษาจะใช้ในการเปรียบเทียบความคิดเห็น ทัศนคติ ค่านิยม และความสามารถระหว่างกลุ่มวัฒนธรรมที่แตกต่างกันในขอบเขตที่กว้าง และได้รับความสนใจที่จะทำการศึกษาเปรียบเทียบที่เกี่ยวข้องกับการเปลี่ยนแปลงอย่างรวดเร็ว

ของสังคมเรา (Van de Vijver. 1998) ไม่ว่าจะเป็นเศรษฐกิจ สังคม การเมือง และยุคของเทคโนโลยีโลกไร้พรมแดน รวมไปถึงการย้ายถิ่นฐานที่เพิ่มขึ้น ซึ่งจะลดความเป็นชนชาติดั้งเดิมที่มีมาในหลายศตวรรษ การเปลี่ยนแปลงทั้งในระดับประเทศ และระดับภูมิภาค ตลอดจนถึงความเป็นหนึ่งเดียวของคนทั่วโลก งานวิจัยเชิงเปรียบเทียบมีเพิ่มมากขึ้นในยุโรปและกระจายไปทางสหภาพยุโรปตะวันออก

ปัจจุบันการศึกษาข้ามวัฒนธรรมมีความน่าสนใจเป็นอย่างมาก แต่ก็ยังไม่มีวิธีที่มีความชัดเจนกับการตรวจสอบความเท่าเทียมกันของการวัดในงานวิจัยลักษณะนี้ เนื่องจากเครื่องมือชนิดเดียวกัน (เช่น แบบสอบถาม แบบตรวจสอบรายการ แบบทดสอบ และอื่น ๆ) มีการใช้กับกลุ่มคนที่เกี่ยวข้องในการศึกษาทั้งหมด ทำให้เกิดข้อสันนิษฐานขึ้นบ่อยครั้งถึงผลที่นำมาเปรียบเทียบระหว่างกลุ่มบุคคล และข้อตกลงในการเปรียบเทียบผลการศึกษา ซึ่งนักวิจัยมักจะเน้นเฉพาะความแตกต่างในคะแนนเฉลี่ยของคนสองกลุ่มวัฒนธรรมหรือมากกว่า แต่ไม่ได้ทำการทดสอบความเท่าเทียมกันของการวัดข้ามกลุ่มวัฒนธรรม ผลการวิจัยจึงเป็นการสะท้อนให้เห็นถึงบริบทของวัฒนธรรมในแต่ละกลุ่มชนนั้นที่หลากหลายปัจจัย หลากหลายกระบวนการ และหลากหลายคุณลักษณะ ชุดคำถามที่เหมือนกันหรือที่ได้รับการออกแบบมา อาจมีความหมายที่แตกต่างกันสำหรับผู้คนที่ต่างวัฒนธรรมกัน นั่นคือ ผลของการวัดอาจเป็นการวัดโครงสร้างวัฒนธรรมที่ต่างกัน ถ้าผลการวัดที่เกิดขึ้นตรงกันกับข้อสรุปในการวิจัยเชิงเปรียบเทียบที่มีข้อคำถามเหล่านั้นอยู่ จะทำให้เป็นข้อกังวลเบื้องต้นของความเท่าเทียมกันในการวิจัยข้ามวัฒนธรรม ไม่ว่าจะเป็นการเปรียบเทียบหรือทดสอบความแตกต่างข้ามวัฒนธรรม (Hui; & Triandis. 1985) ในวิธีการเปรียบเทียบข้ามกลุ่มนี้เรียกว่า ความเท่าเทียมกันของการวัด หมายถึง “ไม่ว่ากลุ่มผู้สอบจะอยู่ภายใต้เงื่อนไขที่แตกต่างกันหรือไม่ก็ตาม การสังเกตและศึกษาปรากฏการณ์ต่าง ๆ เกี่ยวกับผู้สอบต่างกลุ่มกันจะต้องดำเนินการวัดประเมินของคุณลักษณะเดียวกัน” (Horn; & McArdle. 1992: 117)

คำว่า "เท่าเทียมกัน" ได้นำมาใช้ในหลากหลายศาสตร์ที่แสดงถึงแนวคิดที่ต่างกันและความหมายที่ต่างกัน จอห์นสัน (Johnson.1998) พบว่า มีคำว่า "เท่าเทียมกัน" มากกว่า 50 คำที่เป็นคำเฉพาะที่ใช้ในการระบุรูปแบบของการเท่าเทียมกันของการวัด ซึ่งเขาแบ่งออกเป็น 2 กลุ่มกว้าง ๆ คือ

1. การแปลความหมาย

การแปลความหมายความเท่าเทียมกัน ประกอบด้วย ความเท่าเทียมกันทุกอย่างที่เป็นปัญหาหลัก ๆ ที่มีความคล้ายคลึงกัน และความแตกต่างในการตีความ หรือความหมายของการวัดโครงสร้างข้ามวัฒนธรรม แนวคิด ความเท่าเทียมกัน มีความเป็นไปได้ที่จะมีความหมายที่รวมถึงโครงสร้างภายในวัฒนธรรมของแต่ละกลุ่มที่เราสนใจ ตลอดจนความเท่าเทียมกันที่แสดงถึงว่าระดับความคิดที่คล้ายกันภายในแต่ละวัฒนธรรม (Singh. 1995) การแปลความหมายของความเท่าเทียมกันเป็นสิ่งที่จำเป็นที่ต้องมีก่อนกระบวนการที่จะทำให้มีความเท่าเทียมกันของการวัดตั้งแต่ความเท่าเทียมกันในขั้นตอนการวัดแล้ว

2. ความคล้ายคลึงกันในการตีความ

ความคล้ายคลึงกันในการตีความเป็นสิ่งจำเป็นแต่ไม่ใช่เงื่อนไขที่เพียงพอสำหรับการเปรียบเทียบผล: เพื่อให้การเปรียบเทียบผลที่ถูกต้อง ยังต้องการให้ค้นหาแนวทางการวัดที่เท่าเทียมกัน วิธีพิจารณาหรือตีความของความเท่าเทียมกันของการวัด หมายถึง ประเภทของความเท่าเทียมกันที่สัมพันธ์กับมาตรการและวิธีการที่ใช้ในการศึกษาข้ามวัฒนธรรม การวัดที่เท่าเทียม (Horn; & McArdle. 1992) และข้อสอบที่เท่าเทียมกัน (Hui; & Triandis. 1985) ก็จัดอยู่ในกลุ่มความหมายนี้เช่นกัน ซึ่งทั้งสองมีความสำคัญกับระดับของความคล้ายคลึงกันในขั้นตอนของการวัดข้ามวัฒนธรรม

ดังนั้น จะเห็นได้ว่า การวัดความเท่าเทียมกัน เป็นการใช้เครื่องมือการวัดเดียวกันในโครงสร้างเดียวกัน ใช้วัดกับกลุ่มวัฒนธรรมที่แตกต่าง ความเท่าเทียมกันของผลการวัดข้ามวัฒนธรรม จะเกิดขึ้นก็ต่อเมื่อ ผลจากการตอบแบบทดสอบที่ได้จากผู้ตอบที่มีความแตกต่างกันทางวัฒนธรรมหรือข้อคำถามในแบบสอบถามที่กำหนดมีความแตกต่างในการวัดโครงสร้างข้ามวัฒนธรรม ในทางกลับกันคำตอบของผู้ตอบไม่สามารถที่จะสะท้อนให้เห็นถึงตำแหน่งของผู้ตอบในโครงสร้างที่วัดนั้นได้ นอกจากนี้ยังได้รับอิทธิพลจากปัจจัยอื่น ๆ อีก และการพิจารณาข้ามวัฒนธรรมที่แตกต่างกัน ผลที่ได้จากการวัดก็จะไม่เท่าเทียมกัน ตัวอย่างเช่น ปัญหาของความเท่าเทียมกันของแบบทดสอบวัด IQ ในการทดสอบความสามารถ ซึ่งเป็นแบบทดสอบที่ออกแบบมาเพื่อวัดเชาว์ปัญญาของคน โดยเฉพาะอย่างยิ่งในช่วงแรก ๆ ของการทดสอบสภาวะ ทางจิต (Gould. 1981) แบบทดสอบเหล่านี้พัฒนาและเป็นมาตรฐานของประเทศสหรัฐอเมริกา ที่ใช้ในการประเมิน “เชาว์ปัญญา” ของบุคคลและกลุ่มบุคคลที่แตกต่างกันเชื้อชาติ สัญชาติ และภาษา ส่งผลให้ผู้ที่ได้คะแนนต่ำกว่า กลุ่มคนเหล่านั้นได้รับการแปลความหมาย โดยยืนยันว่ามีความแตกต่างในความสามารถทางปัญญา ซึ่งในบางกรณีจะมีผลร้ายแรงสำหรับคนเหล่านั้น เช่น การศึกษาของ Henry Goddard นักจิตวิทยาชาวอเมริกันที่ทำการทดสอบ IQ ผู้อพยพจากยุโรปตะวันออกเฉียงใต้และเกาะเอลลิส ในปี 1910 แล้วบอกว่าคนส่วนใหญ่ของกลุ่มที่ได้รับการทดสอบนี้ “ปัญญาอ่อน” (Goddard. 1917) ผลงานของ Henry Goddard ทำให้ผู้อพยพถูกเนรเทศออกจากประเทศเป็นจำนวนมาก (Hothersall. 1995) แม้ว่าผลงานชิ้นนี้จะโด่งดังมากที่สุด แต่มีการละเมิดระเบียบวิธีการวัดขึ้น การวัดบุคลิกภาพ และการวัดทัศนคติก็พบปัญหาของความเท่าเทียมกันของการวัด แม้ว่าจะมีงานวิจัยอยู่จำนวนมาก ที่ศึกษาในความคล้ายคลึงกันของโครงสร้างปัจจัยบุคลิกภาพข้ามวัฒนธรรม (Church; & Lonner. 1998) แต่ก็ยังมีการศึกษาเกี่ยวกับการประเมินความเท่าเทียมกันของการวัดน้อยมาก ในช่วงต้น ๆ ของการทดสอบสภาวะทางจิต (ช่วงครึ่งปีแรกของศตวรรษที่ 20) ยังคงเป็นเรื่องปกติอยู่ที่จะพบการศึกษาที่มีข้อสอบวัด IQ ส่วนใหญ่ไม่ได้สนใจข้อคำถามว่ามีการวัดเท่าเทียมกันหรือไม่ (Lynn. 2006)

ในการวัดบุคลิกภาพและการวัดทัศนคติ ก็พบปัญหาของความเท่าเทียมกันของการวัด แม้ว่าจะมีการวิจัยอยู่จำนวนมากที่ศึกษาในความคล้ายคลึงกันของโครงสร้างปัจจัยบุคลิกภาพข้ามวัฒนธรรม (Church; & Lonner. 1998) แต่ก็ยังมีการศึกษาน้อยมากที่ทำการประเมินความเท่าเทียม

กันของการวัด โดยความเชื่อในเรื่องของความเท่าเทียมกันของการวัดนั้นใกล้เคียงกับแนวคิดของความลำเอียงของข้อสอบ (Item Bias) และการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ที่มาจากกรอบทฤษฎีการตอบข้อสอบ (IRT) ที่มีการศึกษาอย่างจริงจังในด้านการทดสอบทางการศึกษาอย่างมาก (Van de Vijver; & Leung. 1997) การทำหน้าที่เบี่ยงเบนของข้อสอบหรือความลำเอียงของข้อสอบ หมายถึง ความแตกต่างในความน่าจะเป็นของคำตอบสำหรับผู้ตอบแบบสอบถามด้วยแนวโน้มในการที่จะแสดงพฤติกรรมเท่ากัน ตัวอย่างเช่น ถ้าเราตั้งสมมติฐานในการวัดว่าจะวัด “ความเข้มข้นของพฤติกรรมกลุ่มคนที่มีต่อกีฬาฟุตบอล” ในประเทศไทยและประเทศอังกฤษ ซึ่งข้อคำถามมักจะสอบถามเกี่ยวกับการเข้าร่วมชมหรือร่วมกิจกรรมการแข่งขันฟุตบอล แล้วมีรายการคำตอบให้เลือกตอบ เช่น “สัปดาห์ละครั้ง” สำหรับแฟนฟุตบอลในประเทศไทยอาจจะมีต่ำกว่าแฟนฟุตบอลประเทศอังกฤษ ทั้งที่กลุ่มคนทั้งสองกลุ่มมีทัศนคติเกี่ยวกับฟุตบอลเหมือนกัน ทั้งนี้เพราะการแข่งขันฟุตบอลในประเทศไทยมีการจัดการแข่งขันน้อยกว่าประเทศอังกฤษ จะเห็นว่าข้อความต่าง ๆ เหล่านี้จะเป็นตัวบ่งชี้ถึงความลำเอียงได้ หากใช้ผลการศึกษาไปเปรียบเทียบความเข้มข้นของพฤติกรรมที่มีต่อกีฬาฟุตบอลในกลุ่มคนจากสองประเทศนี้

แบบทดสอบหรือแบบทดสอบย่อยใด ๆ หากผู้ที่มีความสามารถเดียวกัน ทำแบบทดสอบฉบับนั้น ได้คะแนนโครงสร้างตัวแปรแฝงเหมือนกัน มีคะแนนสอบที่คาดหวังเดียวกันหรือคะแนนจริงทั้งระดับข้อสอบหรือระดับผลรวมทั้งหมดของข้อสอบย่อยหรือทั้งสองอย่างในแบบทดสอบหรือแบบทดสอบย่อยนั้น ๆ เท่ากัน แบบทดสอบหรือแบบทดสอบย่อยนั้นจะได้รับการกล่าวว่ามีความเท่าเทียมกันของการวัดข้ามกลุ่มหรือกลุ่มประชากร (Drasgow; & Kanfer. 1985)

ปัญหาหลักในวิธีการวิจัยข้ามวัฒนธรรมอยู่ที่กระบวนการวัดคุณลักษณะที่ศึกษา การวัดที่เท่าเทียมกันในต่างวัฒนธรรมที่ศึกษาย่อมก่อให้เกิดโครงสร้างที่เปรียบเทียบกันได้ (Construct comparability) และผลการเปรียบเทียบข้ามวัฒนธรรมย่อมมีความหมายและมีความถูกต้อง ดังนั้นก่อนที่จะมีการเปรียบเทียบข้ามวัฒนธรรม ผู้วิจัยจึงต้องตรวจสอบความเท่าเทียมกัน (equivalence) ของเครื่องมือระหว่างวัฒนธรรม (ผจงจิต อินทสุวรรณ. 2544; อ้างอิงจาก Van de Vijver; & Poortinga. 1982; Hui; & Trandis. 1985)

งานวิจัยข้ามวัฒนธรรมช่วยตอบคำถามเกี่ยวกับความเป็นสากล (Universality) ซึ่งหมายถึง การนำทฤษฎี งานวิจัยและวิธีการในวัฒนธรรมหนึ่งไปใช้อีกวัฒนธรรมหนึ่ง (Keats. 1997; Van de Vijver; & Poortinga. 1982) จึงต้องใช้กระบวนการวัดเดียวกันในทุกวัฒนธรรม นั่นคือ ผู้วิจัยต้องพัฒนาความเหมาะสมของเครื่องมือวัดในเรื่องความเท่าเทียมกันของการวัด ได้แก่ 1) ความเท่าเทียมกันในความหมาย/บทบาท (Conceptual/ Functional equivalence) ความเท่าเทียมกันในความหมาย หมายถึง โครงสร้าง (Construct) มีความหมายถึงสิ่งเดียวกันในต่างวัฒนธรรมที่ศึกษาถึงแม้ว่าอาจมีการแสดงออกต่างกันในต่างวัฒนธรรม ส่วนบทบาท คือ บทบาทหน้าที่ หมายถึงโครงสร้างนั้นมีปัจจัยและผลที่ตามมาคล้ายคลึงกันในต่างวัฒนธรรม เป็นการยืนยันความเที่ยงตรงเชิงโครงสร้างในแต่ละวัฒนธรรม 2) ความเท่าเทียมกันในการปฏิบัติการ (Equivalence in construct operationalization) เป็นกระบวนการในการถ่ายโอน

จากทฤษฎีไปสู่การวัดหรือมีการปฏิบัติในแต่ละวัฒนธรรมจึงต้องมีการปฏิบัติด้วยกระบวนการเดียวกัน และมีความหมายในวัฒนธรรม เครื่องมือวัดที่ได้จะมีความเท่าเทียมกันข้ามวัฒนธรรม 3) ความเท่าเทียมกันในเครื่องมือวัด (Item equivalence) การมีเครื่องมือวัดเหมือนกัน (The same instrument) ระหว่างวัฒนธรรม หมายถึง การมีข้อคำถามที่มีความหมายเหมือนกัน (Identical) นั่นคือ ข้อสอบแต่ละข้อควรมีความหมายเหมือนกันสำหรับผู้ตอบในวัฒนธรรมต่าง ๆ ที่ศึกษา คุณลักษณะนี้ตรงกับการมีมาตราเดียวกัน ถ้าขาดคุณสมบัติข้อนี้ แสดงว่าในต่างวัฒนธรรมใช้เครื่องมือวัดที่แตกต่างกัน ดังนั้น หากนำผลการวัดที่ได้มาเปรียบเทียบกันย่อมทำให้เข้าใจผิดได้ และ 4) ความเท่าเทียมกันในสเกล (Scalar equivalence) หมายถึง ค่าตัวเลขจากสเกลมีระดับความเข้มหรือขนาดเดียวกันของโครงสร้างที่ศึกษา ไม่ว่าผู้ตอบจะอยู่ในวัฒนธรรมใด นั่นคือ การวัดด้วยมาตราเดียวกันและจุดตั้งต้นของสเกลเหมือนกันในแต่ละวัฒนธรรม คุณสมบัติข้อนี้เป็นไปได้ยากสำหรับการวัดในสังคมศาสตร์ ข้อมูลที่มีความเท่าเทียมกันในระดับนี้สามารถเปรียบเทียบทั้งภายในและระหว่างวัฒนธรรมได้ (ผจงจิต อินทสุวรรณ. 2544)

เหตุผลในความแตกต่างของความเท่าเทียมกันของการวัด ประกอบด้วย

เหตุผลทางจิตวิทยา นั่นคือ เราไม่สามารถคิดได้ว่าคำตอบของคำถามจะได้รับการประมวลผลโดยไม่มีลอกคำตอบจากผู้สอบคนอื่น คำถามที่ต้องมีการแปลความจากประสบการณ์ส่วนตัวของผู้สอบ และมีการประเมินคำตอบที่จะเขียนตอบและแสดงถึงกระบวนการหาคำตอบอาจไม่ตรงตามกรอบของเฉลย

เหตุผลทางสังคมวิทยา ทั้งหมดย่อมขึ้นอยู่กับบริบททางสังคมที่คนนั้น ๆ ฝังตัวอยู่ในชีวิตของเขามีการปฏิสัมพันธ์กับคุณลักษณะของเพศและเชื้อชาติ นอกจากนี้ในบางบริบทของข้อคำถามอาจจะยังไม่มีโอกาสแสดงออก

เหตุผลภาษาศาสตร์ กรณีของการศึกษาข้ามวัฒนธรรมยังมีอิทธิพลของภาษาที่คำถามจะถูกวางคุณภาพของการแปลและการเรียนรู้ภาษาของผู้ตอบแบบสอบถาม

การประเมินผลความเท่าเทียมกันของการวัด ความเท่าเทียมกันในความหมาย/บทบาท และความเท่าเทียมกันในการปฏิบัติการเป็นเรื่องของแนวคิดและความหมายของสิ่งที่ศึกษา รวมทั้งกระบวนการวัด ซึ่งสามารถได้มาด้วยการปรึกษาหารือและตรวจสอบซึ่งกันและกันในทีมวิจัย ซึ่งจะนำไปสู่การมีเครื่องมือวัดที่เหมาะสม ทั้งด้านเนื้อหา ภาษา และความสุภาพ และสร้างความมั่นใจว่าผู้วิจัยศึกษาสิ่งเดียวกันในแต่ละวัฒนธรรม ส่วนความเท่าเทียมกันในเครื่องมือวัด และความเท่าเทียมกันในสเกล เป็นระดับที่ละเอียดและต้องอาศัยเทคนิคทางสถิติยืนยัน ซึ่งในการวิจัยครั้งนี้สนใจเทคนิคทางสถิติสองวิธี คือ วิธีการเชิงเส้นที่อยู่บนทฤษฎีการทดสอบคะแนนจริงสัมพันธ์โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีการไม่เชิงเส้นที่อยู่บนทฤษฎีการตอบข้อสอบโดยจะเป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ซึ่งการทดสอบความเท่ากันของคะแนนจริงข้ามกลุ่มประชากร 2 กลุ่ม ทั้ง 2 วิธี จะกำหนดให้คะแนนแฝงคงที่ (Raju; Laffitte; & Byrne. 2002; Reise; Widaman; & Pugh. 1993)

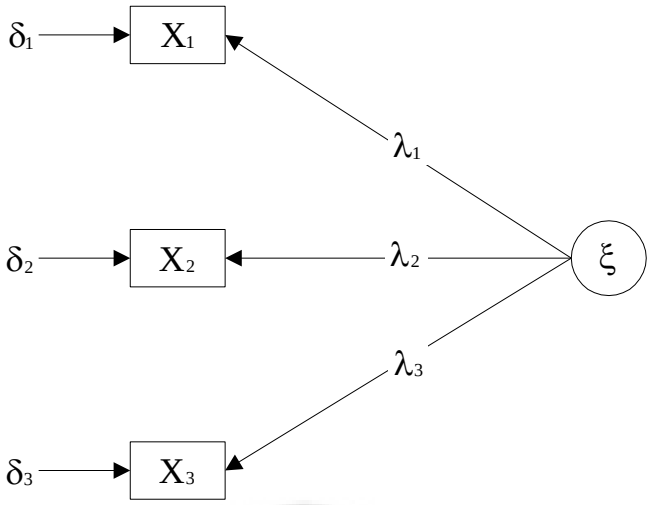
3. การตรวจสอบความเท่าเทียมกันของการวัด (Detection of Measurement Equivalence)

การตรวจสอบความเท่าเทียมกันของการวัด โดยความเชื่อในเรื่องของความเท่าเทียมกันของการวัดนั้นใกล้เคียงกับแนวคิดของความลำเอียงของข้อสอบ (Item Bias) และการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) ที่มาจากกรอบทฤษฎีการตอบข้อสอบ (IRT) การศึกษาเรื่องความเท่าเทียมกันของการวัด การทำหน้าที่เบี่ยงเบนของข้อสอบ หรือความลำเอียงของข้อสอบอาจแบ่งได้เป็นสองประเภทตามเกณฑ์ที่ใช้ คือ เกณฑ์ภายนอก โดยอาศัยสมการถดถอยเชิงเส้นตรง แต่ปัญหาสำคัญของวิธีนี้ก็คือข้อตกลงเบื้องต้นที่ว่าเกณฑ์ภายนอกต้องไม่ลำเอียง จากความยุ่งยากในการสร้างเกณฑ์ภายนอกที่ไม่ลำเอียงนี้ นักทดสอบจึงหันมาตรวจสอบโครงสร้างภายในหรือใช้เกณฑ์ภายในของแบบทดสอบ โดยเปรียบเทียบระหว่างกลุ่มย่อย ๆ วิธีนี้ช่วยให้ทราบว่าคะแนนแบบทดสอบมีความหมายเชิงจิตวิทยาอย่างเดียวกันในแต่ละกลุ่มหรือไม่ วิธีดังกล่าวนี้ใช้ข้อมูลเพียงผลที่บุคคลตอบข้อคำถามในแบบทดสอบเท่านั้น การศึกษาโครงสร้างภายในนี้ส่วนมากมุ่งตรวจสอบความลำเอียงของข้อคำถาม

ดังนั้น การตรวจสอบความเท่าเทียมกันของการวัดจึงเป็นการทดสอบคะแนนจริงจาก 2 กลุ่มประชากรที่แตกต่างกันด้วยวิธีการทางสถิติ ซึ่งมีวิธีการในการตรวจสอบหลายวิธี สำหรับการศึกษาครั้งนี้ผู้วิจัยเลือกใช้ 2 วิธี คือ *วิธีเชิงเส้น (Linear Method)* เป็นการวิเคราะห์ตามทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ (Congeneric Test Theory: CTT) และ*วิธีที่ไม่เป็นเชิงเส้น (Nonlinear Method)* เป็นการวิเคราะห์ตามทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) มีรายละเอียด ดังนี้

3.1 การตรวจสอบความเท่าเทียมกันด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA)

การตรวจสอบความเท่าเทียมกันของการวัด โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน เริ่มด้วยการกำหนดรูปแบบในการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งเป็นการระบุรูปแบบของตัวแปรทางทฤษฎีที่ผู้วิจัยต้องการยืนยันกับข้อมูลเชิงประจักษ์ ได้มาจากการกำหนดจำนวนตัวแปรแฝง (ξ) จำนวนตัวแปรสังเกต (x) ความแปรปรวนร่วมระหว่างตัวแปรแฝง ความสัมพันธ์ระหว่างตัวแปรสังเกตและตัวแปรแฝง ความสัมพันธ์ระหว่างตัวแปรสังเกตกับความคลาดเคลื่อน และความแปรปรวนและความแปรปรวนร่วมระหว่างความคลาดเคลื่อน (δ) ซึ่งข้อได้เปรียบของการวิเคราะห์องค์ประกอบเชิงยืนยัน คือ สามารถกำหนดส่วนประกอบต่าง ๆ ที่กล่าวมาตามความต้องการของนักวิจัย โดยความสัมพันธ์เชิงเส้นตรงในองค์ประกอบเชิงยืนยันมีดังนี้



ภาพประกอบ 7 โมเดลสมการโครงสร้าง

$$x = \Lambda_X \xi + \delta$$

เมื่อ x คือ เวคเตอร์ของตัวแปรสังเกตขนาด $q \times 1$

ξ คือ เวคเตอร์ของตัวแปรแฝงขนาด $n \times 1$

Λ_X คือ เมตริกของค่าสัมประสิทธิ์การถดถอย หรือน้ำหนักองค์ประกอบ (Factor loading) ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรสังเกต (x) กับตัวแปรแฝง (ξ) ขนาด $q \times n$ และ

δ คือ เวคเตอร์ของความคลาดเคลื่อนของตัวแปรสังเกตขนาด $q \times 1$

สมการ $x = \Lambda_X \xi + \delta$ เป็นโมเดลการวัดทั่วไปสำหรับตัวแปรที่เกิดขึ้นในโมเดลสมการโครงสร้าง (Jöreskog; & Sörbom. 1996) ภายในโมเดลองค์ประกอบเดียวของการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) คะแนนตัวแปรแฝงและคะแนนข้อคำถาม แสดงได้ดังนี้

$$x_i = \lambda_i \xi + \delta_i$$

เมื่อ x_i คือ คะแนนที่สังเกตได้ของข้อสอบข้อที่ i

λ_i คือ น้ำหนักองค์ประกอบ (Factor loading) ของข้อสอบข้อที่ i และ

δ_i คือ ความคลาดเคลื่อนของข้อสอบข้อที่ i

โดยค่าคาดหวัง และความแปรปรวนของตัวแปรสังเกต (x_i) ที่ i แสดงได้ ดังนี้

$$\varepsilon(x_i) = \lambda_i \xi$$

$$\sigma_{xi}^2 = \lambda_i^2 \sigma_\xi^2 + \sigma_i^2$$

ซึ่งในสมการ $x_i = \lambda_i \xi + \delta_i$ และ $\varepsilon(x_i) = \lambda_i \xi$ จะคล้ายกับสมการ $x = \Lambda_X \xi + \delta$ คือเมื่อกำหนดให้จุดตัดแกนเป็น 0 นั่นคือ ค่าเฉลี่ยของตัวแปรสังเกต (x_i) ข้อ i และตัวแปรแฝง (ξ) มีค่าเป็น 0 ซึ่งเป็นข้อตกลงเบื้องต้นทั่วไปในการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) และจะทำให้ไม่ต้องกังวลกับการประเมินความเท่าเทียมกันของการวัดภายในบริบทของการวิเคราะห์คะแนนเฉลี่ยและความแปรปรวนร่วมตัวแปรแฝง (MACS) (Sörbom. 1974)

จากสมการ $x = \Lambda_X \xi + \delta$ ค่าคาดหวังและความแปรปรวน-ความแปรปรวนร่วม ($\sum_x$) แสดงได้ดังนี้

$$\varepsilon(x) = \Lambda_X \xi$$

$$\sum_x = \Lambda_X \Phi \Lambda'_X + \Theta_\delta$$

เมื่อ $\Lambda_{\mathbf{X}}$ คือ เมตริกของ Factor loading

$\Lambda'_{\mathbf{X}}$ คือ Transpose ของเมตริก Factor loading

Φ คือ เมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปรแฝง (ξ) และ

Θ_δ คือ เมตริกซ์ความแปรปรวนที่เหลือองค์ประกอบ (Residual factor)

การทดสอบของความเท่าเทียมกันข้ามกลุ่มประชากร (Testing of equivalence across populations) ในการประเมินความเป็นโมเดลองค์ประกอบร่วมข้ามประชากร การทดสอบความไม่แปรเปลี่ยนจะได้รับการทดสอบอยู่บ่อย ๆ การทดสอบความเท่าเทียมกันของเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของค่าสังเกต (Σ) ข้ามกลุ่มประชากรจะกระทำก่อนการทดสอบความไม่แปรเปลี่ยนของพารามิเตอร์อื่น ๆ (Jöreskog; & Sörbom. 1989) และเป็นไปได้ที่จะมีความไม่แปรเปลี่ยนของความแปรปรวน-ความแปรปรวนร่วมระหว่างประชากรที่แตกต่างกัน ($\sum s$) และยังคงสมมติฐานที่แสดงถึงความไม่แปรเปลี่ยนของการวัดที่เป็นกรณีเฉพาะ หรืออาจจะปฏิเสธพารามิเตอร์โครงสร้าง (Byrne. 1998) ทางกลับกัน ความเท่าเทียมกันของความแปรปรวน-ความแปรปรวนร่วมอาจถูกปฏิเสธได้ แต่การทดสอบความไม่แปรเปลี่ยนของการวัดและความไม่แปรเปลี่ยนของโครงสร้างอาจจะคงไว้ การทดสอบต่าง ๆ ที่ใช้การพิจารณาความไม่แปรเปลี่ยนของ $\Lambda_{\mathbf{X}}$, Φ และ Θ_δ ซึ่งเป็นรูปแบบของการทดสอบครั้งนี้ เป็นการเปรียบเทียบโมเดลพื้นฐานที่สามารถกระทำได้ในช่วงเวลาหนึ่งของโมเดลที่ไม่เปลี่ยนแปลงที่ได้รับการทดสอบ

ค่าสัมประสิทธิ์การถดถอย (Lambda) ระหว่างข้อสอบกับตัวแปรแฝงเป็นสิ่งที่เราสนใจอันดับแรกในการประเมินความเท่าเทียมกันของการวัด ประเด็นสำคัญในการวัดความเท่าเทียมกันคือค่า $\Lambda_{\mathbf{X}}$ ของประชากรที่ต่างกันต้องมีค่าเท่ากัน (Drasgow; & Kanfer. 1985; Reise; & et al. 1993; Raju; & et al. 2000) จากข้อตกลงที่เคร่งครัดของความเท่าเทียมกันของการวัดนี้ จึงต้องให้ความแปรปรวนคลาดเคลื่อนของข้อสอบระหว่างกลุ่มประชากรเท่ากัน (เช่น $\sigma^2_{\delta_i} = \sigma^2_{\delta' i}$ ทุกข้อ)

ซึ่งจะทำให้เชื่อได้ว่ามีความเท่าเทียมกันของคะแนนข้อสอบข้ามกลุ่มประชากรภายใต้เงื่อนไขความแปรปรวนของ ξs ที่เท่ากันข้ามกลุ่มประชากร

มีผู้วิจัยบางคน เช่น ลิตเติล (Little. 1997) ชาน (Chan. 2000) ได้ให้เหตุผลในการตรวจสอบความเท่าเทียมกันของการวัดว่า ควรมีการทดสอบจุดตัด (Intercept) เช่นเดียวกับค่าสัมประสิทธิ์การถดถอย (Lambda) ซึ่งเป็นการทดสอบสำหรับความไม่แปรเปลี่ยนของจุดตัดคะแนนสังเกต และทดสอบความไม่แปรเปลี่ยนของจุดตัดของโครงสร้างเช่นเดียวกับค่าสัมประสิทธิ์การถดถอย (Lambda) ของตัวแปรแฝง โดยวิเคราะห์ค่าเฉลี่ยและความแปรปรวนของตัวแปรแฝง (MASC) (Byrne, 1998) จากสมการ $x_i = \lambda_i \xi + \delta_i$ เมื่อกำหนดให้จุดตัดแกนเป็น 0 ดังนั้นสมการซึ่งคล้ายกับสมการ $x_i = \lambda_i \xi + \delta_i$ ที่จุดตัดไม่เป็น 0 จะได้ดังนี้

$$x_i = \tau_i + \lambda_i \xi + \delta_i$$

เมื่อ τ_i คือ จุดตัดของข้อสอบข้อที่ i

ค่าคาดหวังและความแปรปวนของสมการ แสดงได้ดังนี้

$$\varepsilon(x_i) = \tau_i + \lambda_i \xi$$

$$\sigma_{xi}^2 = \lambda_i^2 \sigma_\xi^2 + \sigma_i^2$$

คะแนนจุดตัด (tau) แสดงได้ดังนี้

$$\tau_i = \mu\xi + \lambda_i \mu_i$$

ในขณะที่ความแปรปรวนเป็นค่าเดียวกันจะทำให้สมการ $x_i = \lambda_i \xi + \delta_i$ และสมการ $x_i = \tau_i + \lambda_i \xi + \delta_i$ แตกต่างกัน ณ ตำแหน่งจุดตัด ซึ่งในบริบทนี้หรือในบริบทค่าเฉลี่ยและความแปรปรวนของตัวแปรแฝง (MACS) ความเท่าเทียมของการวัดอาจนิยามให้เป็นความเท่าเทียมกันของค่าสัมประสิทธิ์การถดถอย หรือความชัน (Lambda) และคะแนนจุดตัด (tau) ข้ามกลุ่มประชากร ราจูและคณะ (Raju; & et al., 2000) ได้แสดงให้เห็นว่าความไม่เท่าเทียมกันของคะแนนจุดตัด (tau) อาจไม่แสดงเฉพาะความแตกต่างในค่าสัมประสิทธิ์การถดถอยหรือความชัน (Lambda) เท่านั้น แต่อาจเป็นความแตกต่างในค่าเฉลี่ยของ x_i และ ξ ข้ามกลุ่มประชากรด้วย ซึ่งสอดคล้องกับสมการ $\sum_x = \Lambda_X \Phi \Lambda'_X + \Theta_\delta$ เมื่อค่าสัมประสิทธิ์การถดถอยหรือความชัน (Lambda) เท่ากัน ความแตกต่างคะแนนจุดตัด (tau) จะสะท้อนความแตกต่างในค่าเฉลี่ยของ x_i และ ξ ข้ามกลุ่มประชากรได้ดีขึ้น กล่าวคือ ความแตกต่างของคะแนนจุดตัด (tau) อาจมีการสะท้อนผลกระทบย้อนกลับที่สูงกว่า

การทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) (Dorans; & Holland. 1993) นอกจากนี้ความแตกต่างของข้อสอบข้ามกลุ่มประชากรที่คะแนนจุดตัด (tau) ไม่เป็น 0 จะไม่มากเหมือนฟังก์ชัน ξ ดังนั้นจึงมีการแสดงค่าความแตกต่างที่คงที่ ซึ่งเป็นเหตุผลหนึ่งที่เลือกใช้ความแตกต่างของคะแนนจุดตัด (tau) เช่นเดียวกับการทดสอบความเท่าเทียมกันของการวัดในการทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) (Raju et al., 2000) ทั้ง ๆ ที่ความเกี่ยวเนื่องของความเท่าเทียมกันเฉพาะค่าสัมประสิทธิ์การถดถอย หรือความชัน (Lambda) และการตรวจสอบความเท่าเทียมกันค่าสัมประสิทธิ์การถดถอย หรือความชัน (Lambda) และคะแนนจุดตัด (tau) พร้อม ๆ กัน เป็นการตรวจสอบที่ดีสำหรับการประเมินความเท่าเทียมกันของการวัด ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA)

การทดสอบความเที่ยงตรงข้ามกลุ่ม (Cross Validation) ที่อิสระกัน เป็นวิธีการทดสอบความเที่ยงตรงของโมเดล (Jöreskog; & Sörbom. 1993: 128-129) วิธีหนึ่ง ระหว่างประชากรที่มีการวัดซ้ำหลาย ๆ ครั้ง หรือการวิเคราะห์กลุ่มพหุ (Multi-group Analysis) (นงลักษณ์ วิรัชชัย. 2542: 217) เพื่อตรวจสอบค่าพารามิเตอร์ของโมเดลของประชากรกลุ่มหนึ่งว่าจะเหมือนกับประชากรอีกกลุ่มหนึ่งหรือไม่ โดยการเปรียบเทียบโครงสร้างและค่าพารามิเตอร์ของโมเดล เรียกว่าการศึกษาความไม่แปรเปลี่ยนของโมเดล

การศึกษาความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มตัวอย่างที่มีอิสระต่อกันในการวิเคราะห์องค์ประกอบเชิงยืนยันจะมีลักษณะการทดสอบ ดังนี้ (1) การทดสอบโมเดลพื้นฐาน (Baseline Models) ของแต่ละกลุ่มตัวอย่างย่อย (2) การทดสอบความเท่าเทียมกันของโครงสร้างองค์ประกอบ (Factor Structure: Λ_{form}) และ (3) การทดสอบความเท่าเทียมกันค่าพารามิเตอร์ ซึ่งประกอบด้วย Factor Loading (Λ_x) Factor Variance/Residual (Θ_δ) และ Factor Correlation (Φ) (Marsh; Hau; Chung; & Siu. 1998: 149-150) ดังนี้

3.1.1 การทดสอบโมเดลพื้นฐาน

เบริน์ (Byrne. 1998: 262) กล่าวถึงการทดสอบโมเดลพื้นฐานว่าเป็นการประเมินค่าความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์ในแต่ละกลุ่มย่อยทีละครั้ง ว่ามีโครงสร้างองค์ประกอบที่เหมือนหรือแตกต่างกันอย่างไร เมื่อมีการปรับโมเดลให้สอดคล้องกับข้อมูลเชิงประจักษ์ ซึ่งเป็นแนวคิดที่นำไปสู่การทดสอบความไม่แปรเปลี่ยนของโครงสร้างองค์ประกอบและค่าพารามิเตอร์ เป็นการทดสอบความเท่าเทียมกันของโครงสร้างองค์ประกอบ โดยทดสอบรูปร่างของโมเดลว่าไม่แปรเปลี่ยนจริง ซึ่งจะเกิดขึ้นเมื่อข้อมูลที่เก็บมาในแต่ละกลุ่มมีจำนวนองค์ประกอบ และจำนวนข้อความสอดคล้องกันในแต่ละองค์ประกอบ (Meredith. 1993; citing Cheung; & Rensvold. 2002: 235-237) กล่าวอีกนัยหนึ่งคือ เป็นการทดสอบว่าโมเดลตามทฤษฎีที่สอดคล้องกับข้อมูลเชิงประจักษ์ของกลุ่มประชากรแต่ละกลุ่ม ซึ่งประกอบด้วยจำนวนตัวแปรในโมเดล ลักษณะโครงสร้างความสัมพันธ์ระหว่างตัวแปรเป็นแบบเดียวกันหมดทุกกลุ่ม เมตริกซ์พารามิเตอร์ในโมเดลการวัดมีขนาดเท่ากัน มีรูปแบบเมตริกซ์ (Matrix Form) และสถานะเมตริกซ์

(Matrix Mode) เป็นแบบเดียวกันทุกกลุ่ม แต่ขนาดพารามิเตอร์ไม่เท่ากัน โดยสมมติฐานสำหรับการทดสอบมีดังนี้ (Cheung; & Rensvold. 2002: 236)

$$H_{form} : \Lambda_{form}^{(1)} = \Lambda_{form}^{(2)}$$

3.1.2 การทดสอบความไม่แปรเปลี่ยนของพารามิเตอร์ในโมเดล

การทดสอบความไม่แปรเปลี่ยนของพารามิเตอร์ในโมเดล เป็นการทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ 3 เมตริกซ์ ซึ่งจะกระทำหลังการทดสอบความไม่แปรเปลี่ยนของรูปแบบของโมเดล โดยหลักการจะทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ที่มีความเข้มงวดน้อยที่สุด (Least Restriction) ไปจนถึงการทดสอบความไม่แปรเปลี่ยนของเมตริกซ์พารามิเตอร์ที่มีความเข้มงวดมากที่สุด (Most Restriction) (นงลักษณ์ วิรัชชัย. 2540: 52) แนวคิดการทดสอบทั้ง 3 พารามิเตอร์มีดังนี้ (Cheung; & Rensvold. 2002: 236) (1) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Loading (Λ_X) เป็นการทดสอบว่าในภาพรวมแล้วลักษณะความสัมพันธ์ระหว่างตัวแปรสังเกตกับตัวแปรแฝงของตัวแปรสังเกตนั้นเหมือนกันระหว่างกลุ่ม (2) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Variance/Residual (Θ_δ) เป็นการทดสอบตัวแปรสังเกตว่ามีความสอดคล้องภายในเหมือนกันระหว่างกลุ่ม หรือกล่าวได้ว่าในทุกกลุ่มตัวแปรสังเกตมีคุณภาพที่จะใช้เป็นตัวบ่งชี้ ในตัวแปรแฝงนั้นได้เหมือนกัน (3) การทดสอบความไม่แปรเปลี่ยนของเมตริกซ์ Factor Correlation (Φ) เป็นการทดสอบความสัมพันธ์ระหว่างตัวแปรแฝงว่าเหมือนกันระหว่างกลุ่ม

สำหรับโมเดลการวัดนั้น โบเลน (Bollen. 1989: 360) ได้นำเสนอสมมติฐานทางสถิติที่มีลักษณะซ้อนกันเป็นระดับลดหลั่นดังนี้

H_{form}: โครงสร้างไม่แปรเปลี่ยน (ขนาดของเมตริกซ์ และสถานะของพารามิเตอร์เป็นแบบกำหนด แบบอิสระ และแบบบังคับในเมตริกซ์ Λ, Θ_δ และ Φ เหมือนกัน)

$$H_{\Lambda_x} : \Lambda_X^{(1)} = \Lambda_X^{(2)}$$

$$H_{\Lambda_x\Theta_\delta} : \Lambda_X^{(1)} = \Lambda_X^{(2)} \qquad \Theta_\delta^{(1)} = \Theta_\delta^{(2)}$$

$$H_{\Lambda_x\Theta_\delta\Phi} : \Lambda_X^{(1)} = \Lambda_X^{(2)} \qquad \Theta_\delta^{(1)} = \Theta_\delta^{(2)} \qquad \Phi^{(1)} = \Phi^{(2)}$$

อย่างไรก็ตามในการทดสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มตัวอย่างไม่มีเกณฑ์ตายตัวว่าจะต้องทำการทดสอบความไม่แปรเปลี่ยนของค่าพารามิเตอร์ใดบ้าง แต่มีเงื่อนไขขั้นต่ำในการทดสอบคือ การทดสอบความเท่ากันของโครงสร้างองค์ประกอบและความเท่ากันของค่า Factor loading หรืออาจจะกล่าวได้ว่า สาระสำคัญที่ควรสนใจคือการทดสอบความไม่แปรเปลี่ยนของความสัมพันธ์ระหว่างตัวแปร (Marsh; Hau; Chung; & Siu. 1998: 149)

การประเมินโมเดลในการทดสอบความไม่แปรเปลี่ยนในโมเดลการวัด เป็นการทดสอบอัตราส่วนความเป็นไปได้ (Likelihood - Ratio Test: LR) หรือที่เรียกว่า การทดสอบความแตกต่างของค่าไคสแควร์ ($\Delta\chi^2$) ทีละสมมติฐาน (Bollen. 1989: 368) โดยการประเมินความกลมกลืนด้วยการพิจารณาค่านัยสำคัญของค่าความแตกต่างองศาอิสระ (Δdf) (Cheung; & Rensvold. 2002: 239) หากไม่มีนัยสำคัญทางสถิติหมายความว่าโมเดลมีความกลมกลืนระหว่างกลุ่มตัวอย่าง นอกจากนี้ยังพิจารณาถึงค่า GFI's และค่า $\Delta's$ หากมีค่าสูง หมายความว่าโมเดลมีความกลมกลืนระหว่างกลุ่มตัวอย่างเช่นกัน

3.1.3 ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีคะแนนจริงสัมพันธ์

ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.1.4 ขั้นตอนในการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีคะแนนจริงสัมพันธ์

ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์

องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.2 การตรวจสอบความเท่าเทียมกันของการวัดตามทฤษฎีการตอบข้อสอบ

ขั้นตอนในการตรวจสอบความเท่าเทียมกันของข้อสอบตามทฤษฎีการตอบข้อสอบจะประกอบด้วยขั้นตอนที่สำคัญ ๆ 2 ขั้นตอน คือ **ขั้นตอนแรก**เป็นการประมาณค่าพารามิเตอร์ของข้อสอบ และพารามิเตอร์ความสามารถจากผู้สอบ และ**ขั้นตอนที่ 2** จะเป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ จากขั้นตอนที่เกี่ยวข้องดังที่กล่าวมาประกอบกับการศึกษาครั้งนี้ ผู้วิจัยได้จำลองข้อสอบที่ตรวจให้คะแนนแบบสองค่า และที่ตรวจให้คะแนนแบบหลายค่า ดังนั้น ผู้วิจัยจึงเลือกใช้ Two-Parameter Model สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ และประมาณค่าพารามิเตอร์ความสามารถผู้สอบกับข้อสอบที่ตรวจให้คะแนนแบบสองค่า และใช้ Graded-Response Model สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ความสามารถผู้สอบกับข้อสอบที่ตรวจให้คะแนนแบบหลายค่า

การตรวจสอบความเท่าเทียมกันของการวัดตามทฤษฎีการตอบข้อสอบ โดยการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ซึ่งมีหลากหลายวิธีขึ้นอยู่กับชนิดเงื่อนไขเฉพาะตัวของหลักการที่ใช้ ได้แก่ **วิธีเปลี่ยนค่าความยาก (Difficulty shift)** เป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยวิธีเปลี่ยนค่าความยากจะใช้สถิติ Z_i ทดสอบการเท่ากันเฉพาะค่าความยาก

ของข้อสอบ ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ภายใต้โมเดลโลจิสติกแบบ 1 พารามิเตอร์ หรือโมเดลราชส์ (Rasch Model) ซึ่งกำหนดให้พารามิเตอร์การเดาเป็น 0 และพารามิเตอร์อำนาจจำแนกเท่ากันทุกข้อ ส่วนพารามิเตอร์ความยากจะแปรเปลี่ยนไปตามกลุ่มผู้สอบ **วิธีการทดสอบ F** เป็นวิธีการทดสอบทางอ้อมที่ ฮูลิน ดราสโกว์ และโคโมคาร์ (Hulin; Drasgow; & Komocar. 1982) ได้พัฒนาขึ้นเพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยการนำทฤษฎีการตอบข้อสอบ (IRT) มาประยุกต์ใช้กับการวิเคราะห์ข้อมูลทางเจตคติ และใช้โมเดลโลจิสติกแบบ 2 พารามิเตอร์ ซึ่งกำหนดให้พารามิเตอร์การเดา c เท่ากับ 0 ส่วนพารามิเตอร์ความยาก (b_i) และพารามิเตอร์อำนาจจำแนก (a_i) จะแปรเปลี่ยนไปตามผู้สอบกลุ่มย่อย **วิธีการทดสอบไค-สแควร์ของ Lord (Lord's chi-square test)** เป็นวิธีการทดสอบไค-สแควร์ที่นำเสนอโดย ลอร์ด (Lord. 1980) เพื่อใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบภายใต้ทฤษฎีการตอบข้อสอบ (IRT) โดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ ซึ่งหลักการตรวจสอบด้วยวิธีนี้จะเป็นการทดสอบความแตกต่างของค่าพารามิเตอร์ของข้อสอบระหว่างฟังก์ชันการตอบข้อสอบ (IRFs) จากผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ **วิธีการตอบข้อสอบแบบเทียม (Pseudo-IRT)** ลิน และฮาร์นิช (Linn and Harnisch. 1981) ได้พัฒนาวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบแบบเทียมภายใต้ทฤษฎีการตอบข้อสอบ โดยใช้โมเดลแบบ 3 พารามิเตอร์ เพื่อให้สามารถวิเคราะห์ได้กับกลุ่มตัวอย่างที่มีขนาดเล็ก ซึ่งจะช่วยในการประหยัดค่าใช้จ่าย และ**วิธีทดสอบอัตราส่วนความเป็นไปได้ (Likelihood ratio test: LR)** เป็นการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบโดยการทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม ซึ่งเป็นการกระทำที่ระดับข้อสอบเช่นเดียวกับวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เป็นเหตุผลที่ทำให้ผู้วิจัยเลือกวิธีดังกล่าวมาศึกษา โดยมีรายละเอียดของการตรวจสอบ ดังนี้

3.2.1 วิธีทดสอบอัตราส่วนความเป็นไปได้ (Likelihood ratio test: LR)

ทิสเซนและคณะ (Thissen; Steinberg; & Wainer. 1988) ได้เสนอวิธีการทดสอบอัตราส่วนความเป็นไปได้สำหรับใช้ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โดยการทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม วิธีทดสอบอัตราส่วนความเป็นไปได้ แบ่งออกเป็น 3 วิธีย่อยๆ คือ (1) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบ ในรูปทั่วไป (General LR) เป็นวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบซึ่งใช้การประมาณค่าพารามิเตอร์ในโมเดลการตอบข้อสอบด้วยวิธีความเป็นไปได้สูงสุดแบบมาร์จินอล (Marginal maximum likelihood; MML) (2) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปลอกลิเนียร์ (Loglinear LR) เป็นการประมาณคาพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood; ML) และ (3) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปสารสนเทศที่มีขอบเขตจำกัด (Limited information LR) สำหรับวิธีนี้จะใช้การประมาณค่าพารามิเตอร์ในโมเดลการตอบข้อสอบแบบ Normal ogive ด้วยวิธีกำลังสองน้อยที่สุดในรูปทั่วไป (Generalized least squares: GLS) ทั้ง 3 วิธีดังกล่าวจะใช้การทดสอบ

อัตราส่วนความเป็นไปได้เพื่อทดสอบนัยสำคัญของการทำหน้าที่เบี่ยงเบนของข้อสอบ ซึ่งผลความกลมกลืนของโมเดลเป็นที่รู้จักโดยทั่วไปในความกลมกลืนของฟังก์ชัน ในทฤษฎีการตอบข้อสอบ (IRT) ค่าความกลมกลืนของฟังก์ชันเป็นดัชนีบอกว่าโมเดลความเหมาะกับข้อมูลที่จะใช้กระบวนการประมาณค่าความเป็นไปได้สูงสุดในการประมาณค่าพารามิเตอร์ข้อสอบ (Camilli; & Shepard. 1994)

หลักการตรวจสอบความเท่าเทียมกันของการวัดด้วยวิธีการทดสอบอัตราส่วนความเป็นไปได้ จะเปรียบเทียบระหว่าง 2 โมเดล คือ โมเดลพื้นฐาน (Compact) และโมเดลเปรียบเทียบ (Augmented) ในโมเดลแรกสมมติให้ไม่มีกลุ่มผู้สอบที่แตกต่างกัน ดังนั้นจึงบังคับให้พารามิเตอร์ของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจมีค่าเท่ากัน (Group 1–Item 1 และ Group 2–Item 1) (Thissen; Steinberg; & Wainer. 1988, 1993) โมเดลพื้นฐานนี้เป็นการจัดเตรียมค่าความเป็นไปได้พื้นฐานสำหรับความกลมกลืนของพารามิเตอร์ข้อสอบในโมเดล จากนั้นจะทำการทดสอบข้อสอบแต่ข้อที่ทำหน้าที่เบี่ยงเบน โดยแยกข้อมูลวิเคราะห์ในแต่ข้อ ซึ่งคล้ายกับการประมาณค่าพารามิเตอร์ข้อสอบที่บังคับให้เท่ากันข้ามกลุ่ม ยกเว้นค่าพารามิเตอร์ของข้อสอบเริ่มต้น การทดสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ โมเดลเปรียบเทียบจัดเตรียมค่าความเป็นไปได้ที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ข้อสอบข้อที่ i ในแต่ละกลุ่มประชากร ค่าความเป็นไปได้นี้สามารถเปรียบเทียบทุกพารามิเตอร์ของข้อสอบทั้งหมดที่บังคับให้เท่ากันข้ามกลุ่มในโมเดลพื้นฐาน สูตรการทดสอบอัตราส่วนความเป็นไปได้ แสดงได้ดังนี้

$$LR_i = \frac{L_C}{L_{Ai}}$$

เมื่อ L_C คือ ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (จะเป็นค่าพารามิเตอร์ที่น้อยกว่า)

L_{Ai} คือ ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบที่ยอมให้พารามิเตอร์ข้อสอบของข้อสอบข้อที่ i ข้ามกลุ่มผู้สอบมีความหลากหลาย

นั่นคือในโมเดลพื้นฐาน (Compact) จะประกอบด้วยข้อสอบที่ทำหน้าที่ไม่ต่างกัน สำหรับในโมเดลเปรียบเทียบจะประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่มีค่าแปรเปลี่ยนไปตามกลุ่ม โมเดลเปรียบเทียบ (Augmented) อาจมีข้อสอบจำนวน 1 ข้อ หรือมากกว่าที่ทำหน้าที่เบี่ยงเบน นอกจากนี้ระหว่าง 2 โมเดลจะต้องมีข้อสอบร่วม ซึ่งเป็นข้อสอบที่สมมติว่าทำหน้าที่ไม่ต่างกัน จากนั้นจึงทำการเปรียบเทียบระหว่าง 2 โมเดล ด้วยสถิติอัตราส่วนความเป็นไปได้ ดังนี้

$$G_i^2 = -2log[\,L_C\,] - (-2log[\,L_A\,])$$
$$= -2log[\,L_C\,] + 2log[\,L_A\,]$$
$$= -2log\left[\frac{L_C}{L_A}\right]$$

เมื่อ G_i^2 คือ สถิติอัตราส่วนความเป็นไปได้ของข้อสอบข้อที่ i

L_C คือ ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (Compact)

L_A คือ ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบ (Augmented)

โดยทั่วไปแล้ว $L_C < L_A$ และสถิติ $G_i^2 > 0$ มีการแจกแจงแบบไค-สแควร์ ซึ่งมีระดับของความเป็นอิสระเท่ากับจำนวนพารามิเตอร์ในโมเดล L_C และโมเดล L_A ดัชนีความไม่กลมกลืนในระดับนัยสำคัญทางสถิติของผลการทดสอบจะชี้บอกว่าโมเดลพื้นฐานกลมกลืนน้อยกว่าโมเดลเปรียบเทียบอย่างมีนัยสำคัญทางสถิติ ในการใช้การทดสอบอัตราส่วนความเป็นไปได้ ค่าไค-สแควร์คำนวณมาจากข้อสอบแต่ข้อในแบบทดสอบ หากค่าไค-สแควร์ของข้อสอบเหล่านั้นมีนัยสำคัญทางสถิติ แสดงว่ามีการทำหน้าที่เบี่ยงเบนของข้อสอบ

3.2.2 ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีตอบข้อสอบ

ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.2.3 ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีตอบข้อสอบ

ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่า มีการระบุข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษาข้อนั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

4. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความเท่าเทียมกันของการวัดด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ตามทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ (CTT) และการวิเคราะห์การทำหน้าที่เบี่ยงเบนของข้อสอบ (DIF) โดยวิธีทดสอบอัตราส่วนความเป็นไปได้ (LR) ตามทฤษฎีการตอบข้อสอบ (IRT) ผู้วิจัยขอนำเสนอ ดังนี้

งานวิจัยต่างประเทศ

รัชนีวรรณ วานิชย์ถนอม (Wanichtanom, R. 2001) ได้เปรียบเทียบอัตราการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มอ้างอิงและกลุ่มสนใจด้วยวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ และวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน 3 วิธี ได้แก่ วิธีเปรียบเทียบโมเดล (MC) วิธีปรับปรุงดัชนี (MI) และวิธีปรับปรุงดัชนีด้วยสัดส่วนกลุ่มตัวอย่าง (MI-Divided) ทำการศึกษากับข้อมูลจำลองการตอบข้อคำถามโมเดลโลจิสติก 2 พารามิเตอร์ ผู้สอบจำนวน 1,000 คน ข้อสอบ 50 ข้อ โดยกำหนดให้เป็นข้อสอบร่วม 40 ข้อ และข้อสอบศึกษาจำนวน 9 ข้อ ซึ่งกำหนดรูปแบบของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษาเป็น 3 รูปแบบ ได้แก่ 1) มีการทำหน้าที่เบี่ยงเบนบนค่าพารามิเตอร์อำนาจจำแนก 2) มีการทำหน้าที่ต่างกันบนพารามิเตอร์ค่าความยาก และ 3) มีการทำหน้าที่ต่างกันบนทั้งพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยาก ที่มีขนาดของการทำหน้าที่เบี่ยงเบน 3 ระดับ ได้แก่ การทำหน้าที่เบี่ยงเบนขนาดเล็ก การทำหน้าที่เบี่ยงเบนขนาดกลาง และการทำหน้าที่เบี่ยงเบนขนาดใหญ่ และกำหนดพื้นที่ระหว่างเส้นโค้งลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มสนใจเป็นค่าคงที่ไว้ที่ .5 ผลการศึกษา พบว่า วิธีการตรวจสอบทั้งหมด ได้แก่ วิธีเปรียบเทียบโมเดล (MC) วิธีปรับปรุงดัชนี (MI) และวิธีปรับปรุงดัชนีด้วยสัดส่วนกลุ่มตัวอย่าง (MI-Divided) สามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ ได้ดีในข้อสอบที่มีการทำหน้าที่เบี่ยงเบนบนค่าพารามิเตอร์ความยาก โดยทฤษฎีการตอบข้อสอบ (IRT) และ MI-Divided ให้ผลการตรวจสอบสูงกว่าวิธี MC และ MI ทั้งในข้อสอบที่กำหนดให้ มีการทำหน้าที่เบี่ยงเบนบนค่าพารามิเตอร์อำนาจจำแนก และในข้อสอบที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนบนค่าพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยาก วิธีทฤษฎีการตอบข้อสอบ (IRT) และ MI-Divided มีอัตราการตรวจสอบไม่แตกต่างกัน และวิธี MC และ MI มีอัตราการตรวจสอบไม่แตกต่างกัน ส่วนวิธี MI-Divided มีอัตราการตรวจสอบที่สูง แต่มีความคลาดเคลื่อนในการตรวจสอบสูงกว่าค่าคาดหวัง

ฟลาวเวอร์ ราชู และโอชิมา (Flowers; Raju; & Oshima. 2002) ได้ทำการเปรียบเทียบวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน และทฤษฎีการตอบข้อสอบในการตรวจสอบความเท่าเทียมกันของการวัด โดยการจำลองข้อมูลมาตรวัดประมาณค่า 5 ระดับ จำนวน 20 ข้อ

ภายใต้ Graded Response Model และใช้ข้อสอบศึกษาใน 3 เงื่อนไข คือ เงื่อนไขที่ 1 มีข้อสอบศึกษา 2 ข้อ กำหนดให้ค่าพารามิเตอร์ความยาก (b-parameter) ในกลุ่มสนใจข้อที่ 3 เป็น .5 และข้อ 8 เป็น 1.0 เงื่อนไขที่ 2 มีข้อสอบศึกษา 4 ข้อ ได้แก่ข้อที่ 3 กำหนดค่าพารามิเตอร์ความยาก (b-parameter) คงที่เป็น 1.0 ข้อที่ 8 ลดค่าพารามิเตอร์อำนาจจำแนก (a-parameter) ลง 0.5 และเพิ่มพารามิเตอร์ความยาก (b-parameter) เป็น 0.5 ข้อที่ 13 เพิ่มพารามิเตอร์ความยาก (b-parameter) เป็น 0.5 และข้อที่ 18 ลดค่าพารามิเตอร์อำนาจจำแนก (a-parameter) ลง 0.5 ส่วนเงื่อนไขที่ 3 มีข้อสอบศึกษา 4 ข้อ โดยระบุการทำหน้าที่เบี่ยงเบนของข้อสอบด้วยการเพิ่มค่าคงที่ของพารามิเตอร์อำนาจจำแนก (a-parameter) ในข้อที่ 3 ขึ้น 0.5 ลดค่าคงที่ของพารามิเตอร์อำนาจจำแนก (a-parameter) ในข้อ 4 ลง 0.5 เพิ่มค่าคงที่ของพารามิเตอร์ความยาก (b-parameter) ในข้อที่ 12 ขึ้น 0.5 และลดค่าคงที่ของพารามิเตอร์ความยาก (b-parameter) ในข้อ 13 ลง 0.5 ส่วนปัจจัยความสามารถของผู้สอบมี 2 รูปแบบ คือ กำหนดให้กระจายความสามารถของกลุ่มสนใจเป็นแบบโค้งปกติ N(0,1) และกลุ่มสนใจมีการกระจายความสามารถต่ำกว่าการกระจายของกลุ่มอ้างอิง 1 ส่วนเบี่ยงเบนมาตรฐาน แต่ละเงื่อนไขจะทำซ้ำ 5 ครั้ง ทำการตรวจสอบความเท่าเทียมกันของการวัดข้ามกลุ่มสนใจและกลุ่มอ้างอิงด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ที่เป็นลักษณะเชิงเส้นตรง 2 วิธี คือ ตรวจสอบความชันของโครงสร้าง (Lambda structure) และ ตรวจสอบอัตราส่วนของความชันกับจุดตัดของโครงสร้าง (Slope and intercept structures) (Jöreskog & Sörbom, 1996) และวิธีการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่ตั้งอยู่บนพื้นฐานของทฤษฎีการตอบข้อสอบ (IRT) ที่ไม่เป็นเชิงเส้น ใช้วิธีการทำหน้าที่เบี่ยงเบนของข้อสอบที่ไม่มีการชดเชยการตอบ (NC-DIF) (Raju, van der Linden, & Fleer, 1995) ในการตรวจสอบความเท่าเทียมกันของการวัดข้ามกลุ่ม ผลการศึกษาพบว่า วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันที่ใช้ความชัน (Lambda) แสดงถึงความแตกต่างในค่าพารามิเตอร์อำนาจจำแนกของข้อสอบได้อย่างสมบูรณ์ วิธีที่ใช้อัตราส่วนของความชันกับจุดตัด และการทำหน้าที่เบี่ยงเบนของข้อสอบแบบไม่มีการชดเชยการตอบ แสดงถึงความแตกต่างในพารามิเตอร์ความยากของข้อสอบ และเมื่อกลุ่มสนใจและกลุ่มอ้างอิงมีการแจกแจงความสามารถแตกต่างกัน วิธีอัตราส่วนของความชันกับจุดตัดจะมีอัตราความคลาดเคลื่อนประเภทที่ 2 ต่ำ แต่จะมีอัตราความคลาดเคลื่อนแบบที่ 1 สูงขึ้น ส่วนวิธีการทำหน้าที่เบี่ยงเบนของข้อสอบแบบไม่มีการชดเชยการตอบ พบว่าส่งผลต่ออัตราความคลาดเคลื่อนแบบที่ 1 และแบบที่ 2 ส่วนการกระจายความสามารถของผู้สอบไม่ส่งผลต่ออัตราความคลาดเคลื่อนแบบที่ 1 และแบบที่ 2

มีด และ ลอเทนชลาเจอ (Meade; & Lautenschlager. 2004) ได้ทำการเปรียบเทียบทฤษฎีการตอบข้อสอบและการวิเคราะห์องค์ประกอบเชิงยืนยัน ในการแสดงความเท่าเทียมกันของการวัด โดยการจำลองข้อมูลจากข้อคำถามแบบ Likert 5 ระดับ กับขนาดกลุ่มตัวอย่าง 2 กลุ่มที่มีการแจกแจงปกติ 3 ขนาด คือ 150, 500 และ 1,000 คน แต่ละเงื่อนไขมีการทำซ้ำ 100 ครั้ง และใช้ค่าพารามิเตอร์ความยาก และค่าพารามิเตอร์อำนาจจำแนกในการประเมินความเหมาะสม ความเหมือน และความแตกต่างระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีทฤษฎีการตอบ

ข้อสอบ พบว่า กรณีที่ใช้ค่าพารามิเตอร์ความยากในการตรวจสอบความเท่าเทียมกันของการวัด วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันจะตรวจสอบได้ดีกับขนาดกลุ่มตัวอย่าง 500 และ 1,000 คน ส่วนขนาดกลุ่มตัวอย่าง 150 คนตรวจสอบได้ต่ำ และเมื่อใช้อัตราส่วนความเป็นไปได้ตรวจสอบความเท่าเทียมกันของการวัด พบว่า ขนาดกลุ่มตัวอย่าง 150 คน จะมีความถูกต้องในการตรวจสอบต่ำ การตรวจสอบความเท่าเทียมกันของการวัดด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน พบว่า ตรวจสอบได้ดีในทุกขนาดกลุ่มตัวอย่าง ส่วนการตรวจสอบด้วยวิธีอัตราความเป็นไปได้ พบว่า ในขนาดกลุ่มตัวอย่าง 1,000 ให้ผลการตรวจสอบได้ดี (ร้อยละ 97-100 ของจำนวนครั้งในการตรวจสอบข้อสอบศึกษา) ส่วนขนาดกลุ่มตัวยอย่าง 150 ให้ผลการตรวจสอบไม่คงที่ (ร้อยละ 3-80 ของจำนวนครั้งในการตรวจสอบข้อสอบศึกษา) และกลุ่มตัวอย่างขนาด 500 ให้ผลการตรวจสอบไม่แตกต่างจากขนาดกลุ่มตัวอย่าง 150 (ร้อยละ 43-98 ของจำนวนครั้งในการตรวจสอบข้อสอบศึกษา)

ฟินช์ และ เฟรนช์ (Finch; & French. 2006) ได้ตรวจสอบ การทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชยคำตอบใน 4 วิธี คือ การถดถอยโลจิสติก (LR) ซิปเทสท์ (SIBTEST) การทดสอบอัตราส่วนความเป็นไปได้ทฤษฎีการตอบข้อสอบ (LR) และการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) จากการจำลองข้อมูลภายใต้เงื่อนไขต่าง ๆ ได้แก่ ขนาดกลุ่มตัวอย่าง ใช้สัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิงแบบ 1:1 และแบบ 1:2 จำนวน 3 ขนาด คือ 250:250 250:500 และ 500:1,000 คน ระดับความสามารถค่าเฉลี่ย 2 ระดับคือ ระดับความสามารถเฉลี่ยเท่ากัน (M=0, SD=1) และระดับความสามารถเฉลี่ยกลุ่มอ้างอิงสูงกว่ากลุ่มสนใจ (M_R=0.5, M_F=0) ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบที่มีการชดเชยการตอบ 3 ขนาด คือ 0% 10% หรือ 20% กระทำภายใต้โมเดลโลจิสติก 2 และ 3 พารามิเตอร์ แต่ละเงื่อนไขทำซ้ำ 1,000 ครั้ง ผลการศึกษาพบว่า การตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ข้ามกลุ่ม ไม่มีปฏิสัมพันธ์ร่วมใด ที่มีนัยสำคัญทางสถิติ แสดงถึงระดับของเงื่อนไขวิธีการไม่มีนัยสำคัญทางสถิติ และไม่ใช่ความแตกต่างที่เด่นชัด อัตราความคลาดเคลื่อนประเภทที่ 1 ไม่มีผลกระทบต่อโมเดลการประมาณค่าพารามิเตอร์ภายใต้ขนาดกลุ่มตัวอย่าง และความสามารถของกลุ่ม ส่วนการตรวจสอบอัตราความถูกต้อง พบว่า ความแปรปรวนของขนาดกลุ่มตัวอย่างส่งผลต่ออัตราความถูกต้อง ในวิธีการถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนความเป็นไปได้ทฤษฎีการตอบข้อสอบ และวิธีซิปเทสท์ ซึ่งอธิบายสัดส่วนของความแปรปรวนมากที่สุด (มากกว่า 70%) ส่วนวิธีการวิเคราะห์องค์ประกอบไม่มีองค์ประกอบที่มีนัยสำคัญทางสถิติ วิธีอัตราส่วนความเป็นไปได้มีปฏิสัมพันธ์ภายใต้โมเดลการประมาณค่าพารามิเตอร์และความแตกต่างในความสามารถอย่างมีนัยสำคัญทางสถิติ และอธิบายความแปรปรวนได้ 4.5% โดยที่ความแตกต่างในความสามารถของกลุ่มไม่กระทบต่ออัตราความถูกต้อง และเมื่อกลุ่มมีความสามารถไม่เท่ากันของข้อมูลแบบ 3 พารามิเตอร์ ในการเปรียบเทียบวิธีถดถอยโลจิสติกจะกระทำได้ยาก นอกจากนี้ยังพบว่า วิธีซิปเทสท์มีอัตราความถูกต้องสูงที่สุดในทุกเงื่อน โดยทั่วไปทั้งวิธีการทดสอบอัตราส่วนความเป็นไปได้ตามทฤษฎีการตอบข้อสอบ และวิธีวิเคราะห์องค์ประกอบเชิงยืนยันจะมีอัตรา

ความถูกต้องต่ำ ซึ่งนำไปสู่การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า สำหรับองค์ประกอบอื่น ๆ พบว่า โมเดลโลจิสติก 3 พารามิเตอร์มีส่วนทำให้อัตราความถูกต้องของวิธีปเทสท์มีนัยสำคัญทางสถิติ และในวิธีการถดถอยโลจิสติก โมเดลโลจิสติก 3 พารามิเตอร์มีปฏิสัมพันธ์ร่วมกับความแตกต่างในความสามารถ และเมื่อมีการนำเสนอค่าการเดาจะทำให้วิธีการถดถอยโลจิสติกและวิธีซิปเทสท์ให้อัตราความถูกต้องต่ำ ส่วนในวิธีอัตราความเป็นไปได้ของทฤษฎีการตอบข้อสอบ ภายใต้โมเดลการประมาณค่าพารามิเตอร์ไม่มีผลต่ออัตราความถูกต้อง ในกรณีของการวิเคราะห์องค์ประกอบเชิงยืนยันอัตราความถูกต้องมีค่าต่ำ

จากการศึกษางานวิจัยที่เกี่ยวข้องจะพบว่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดเพื่อให้ผลการตรวจสอบมีความถูกต้องแม่นยำ และสามารถนำไปประยุกต์ใช้ให้เหมาะกับบริบทจริงของการทดสอบได้ เริ่มจากการตรวจสอบความเท่าเทียมกันของการวัดโดยใช้วิธีการตรวจสอบวิธีเดียวบนฐานคิดทางทฤษฎีการทดสอบเดียว ต่อมามีการพัฒนาเป็นการตรวจสอบความเท่าเทียมกันของการวัดที่ใช้วิธีการตรวจสอบที่หลากหลายขึ้นแต่ยังอยู่บนฐานคิดทางทฤษฎีการทดสอบเดียว จนปัจจุบันได้มีการใส่ใจในเรื่องของความหลากหลายในวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่จะสามารถนำไปใช้ได้ตามบริบทของการทดสอบจริงได้ จึงมีการศึกษาเกี่ยวกับการตรวจสอบความเท่าเทียมกันของการวัดที่หลากหลายเทคนิควิธีกับข้อสอบที่มีลักษณะการตรวจสอบที่อยู่บนฐานคิดทางทฤษฎีการทดสอบที่แตกต่างกัน ซึ่งหลัก ๆ แล้ว จะใช้เทคนิวิธีการตรวจสอบที่อยู่บน 2 ฐานคิดทางทฤษฎีที่ต่างกัน นั่นคือ วิธีที่อยู่บนฐานคิดทางทฤษฎีการทดสอบมาตรฐานเดิม (CTT) กับวิธีที่อยู่บนฐานคิดทางทฤษฎีการตอบข้อสอบ (IRT) และนักทฤษฎีส่วนใหญ่จะดำเนินการตรวจสอบภายใต้ข้อมูลจำลอง เพื่อให้ครอบคลุมเงื่อนไขต่าง ๆ ที่คาดว่าจะเกิดขึ้นในสถานการณ์ของการทดสอบจริง ทั้งนี้เพื่อเป็นการหาพยานหลักฐานที่หลากหลายและมีความน่าเชื่อถือในการที่จะนำเทคนิควิธีการตรวจสอบความเท่าเทียมกันของการวัดไปใช้ได้อย่างถูกต้องเหมาะสมกับบริบททางการทดสอบ และเพื่อเป็นแนวทางสำหรับนักวิจัยและผู้สนใจในการตรวจสอบและคัดกรองเครื่องมือวัดที่พัฒนามาจากวัฒนธรรมหนึ่งหรือกลุ่มประชากรหนึ่งแล้วต้องการนำเครื่องมือเหล่านั้นมาศึกษากับกลุ่มวัฒนธรรมหรือกลุ่มประชากรอื่น หรือสนใจเปรียบเทียบข้ามวัฒนธรรม หรือข้ามกลุ่มประชากร ซึ่งจะทำให้การสรุปผลมีความน่าเชื่อถือมากขึ้น

งานวิจัยในประเทศ

วรรณี แกมเกต; นงลักษณ์ วิรัชชัย; และสมหวัง พิธิยานุวัฒน์ (2540) ได้ศึกษาการพัฒนาตัวบ่งชี้ประสิทธิภาพการใช้ครูและการทดสอบความไม่แปรเปลี่ยนของโมเดลประสิทธิภาพการใช้ครูระหว่างกลุ่มครูโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ และกรมสามัญ โดยใช้การวิเคราะห์โมเดลสมการโครงสร้างกลุ่มพหุ และใช้ฐานข้อมูลจากโครงการวิจัยสำนักงานคณะกรรมการการศึกษาแห่งชาติ เรื่อง ประสิทธิภาพการใช้ครู: การวิเคราะห์เชิงปริมาณระดับ

มหภาค กลุ่มตัวอย่างที่ใช้ในการวิจัยคือโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ และกรมสามัญ จำนวน 625 โรงเรียน โดยศึกษากับตัวแปรสังเกต 2 กลุ่ม คือ กลุ่มตัวแปรด้านกระบวนการการใช้ครู 6 ตัวแปร และกลุ่มตัวแปรด้านผลผลิตที่เกิดกับครู 10 ตัวแปร และตัวแปรแฝง 2 ตัวคือ ประสิทธิภาพการใช้ครูด้านกระบวนการใช้ครู และด้านผลผลิตที่เกิดกับตัวครู ผลการวิจัยพบว่า ตัวแปรที่สำคัญของตัวบ่งชี้ประสิทธิภาพการใช้ครู ด้านกระบวนการใช้ครู คือ การนิเทศครู การประเมินผลการปฏิบัติงาน ลักษณะการมอบหมายงาน และการบริหารแบบมีส่วนร่วม ส่วนตัวแปรที่สำคัญของประสิทธิภาพการใช้ครูด้านผลผลิตที่เกิดกับตัวครู คือ ความพึงพอใจในการทำงาน คุณภาพของงาน ความก้าวหน้าเทียบกับเพื่อนร่วมรุ่น ความปรารถนาที่จะเป็นครูต่อไปในอนาคต ความผูกพันกับอาชีพครู และความหลากหลายของงาน ส่วนโมเดลประสิทธิภาพการใช้ครูมีความแปรเปลี่ยนของค่าพารามิเตอร์ในเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างความคลาดเคลื่อนในการวัดตัวแปรสังเกตได้ ระหว่างกลุ่มโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ และกรมสามัญ

จากการศึกษากรอบแนวคิดเชิงทฤษฎีเกี่ยวกับการตรวจสอบความเท่าเทียมกันของการวัดตลอดจนงานวิจัยต่าง ๆ ที่เกี่ยวข้องจะพบว่า ปัจจุบันได้มีการใส่ใจในเรื่องของความหลากหลายในวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่จะสามารถนำไปใช้ได้ตามบริบทของการทดสอบจริงได้ จึงมีการศึกษาเกี่ยวกับการตรวจสอบความเท่าเทียมกันของการวัดที่หลากหลายเทคนิควิธีกับข้อสอบที่มีลักษณะการตรวจสอบที่อยู่บนฐานคิดทางทฤษฎีการทดสอบที่แตกต่างกัน ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาการตรวจสอบความเท่าเทียมกันของการวัดในข้อสอบที่รายการคำตอบมีการตรวจให้คะแนนแบบสองค่า และข้อสอบที่รายการคำตอบมีการตรวจให้คะแนนแบบหลายค่า โดยใช้เทคนิควิธีการตรวจสอบที่อยู่บน 2 ฐานคิดทางทฤษฎีที่ต่างกัน นั่นคือวิธีที่อยู่บนฐานคิดทางทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ (CTT) กับวิธีที่อยู่บนฐานคิดทางทฤษฎีการตอบข้อสอบ (IRT) ทั้งนี้เพื่อเป็นการหาพยานหลักฐานที่หลากหลายและมีความน่าเชื่อถือในการที่จะนำเทคนิควิธีการตรวจสอบความเท่าเทียมกันของการวัดไปใช้ได้อย่างถูกต้องเหมาะสมกับบริบททางการทดสอบ และเพื่อเป็นแนวทางสำหรับนักวิจัยและผู้สนใจในการตรวจสอบและคัดกรองเครื่องมือวัดที่พัฒนามาจากวัฒนธรรมหนึ่งหรือกลุ่มประชากรหนึ่งแล้วต้องการนำเครื่องมือเหล่านั้นมาศึกษากับกลุ่มวัฒนธรรม หรือกลุ่มประชากรอื่น หรือสนใจเปรียบเทียบข้ามวัฒนธรรมหรือข้ามกลุ่มประชากร ซึ่งจะทำให้การสรุปผลมีความน่าเชื่อถือมากขึ้น

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้ เป็นการศึกษาวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีที่ต่างกัน 2 ทฤษฎี คือ ทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบภายใต้เงื่อนไขวิธีการตรวจให้คะแนน ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่างที่แตกต่างกัน ผู้วิจัยดำเนินการตามขั้นตอนดังนี้

1. การจัดกระทำตัวแปร
2. การจำลองข้อมูล (Simulation)
3. การวิเคราะห์ข้อมูล

1. การจัดกระทำตัวแปร

การศึกษาวิจัยครั้งนี้ มีความมุ่งหมายเพื่อศึกษาอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด ซึ่งผู้วิจัยเลือกวิธีดำเนินการในการตรวจสอบความเท่าเทียมกันของการวัด 2 วิธี คือ วิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยวิธีทดสอบอัตราส่วนความเป็นไปได้ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 ปัจจัย ดังนี้

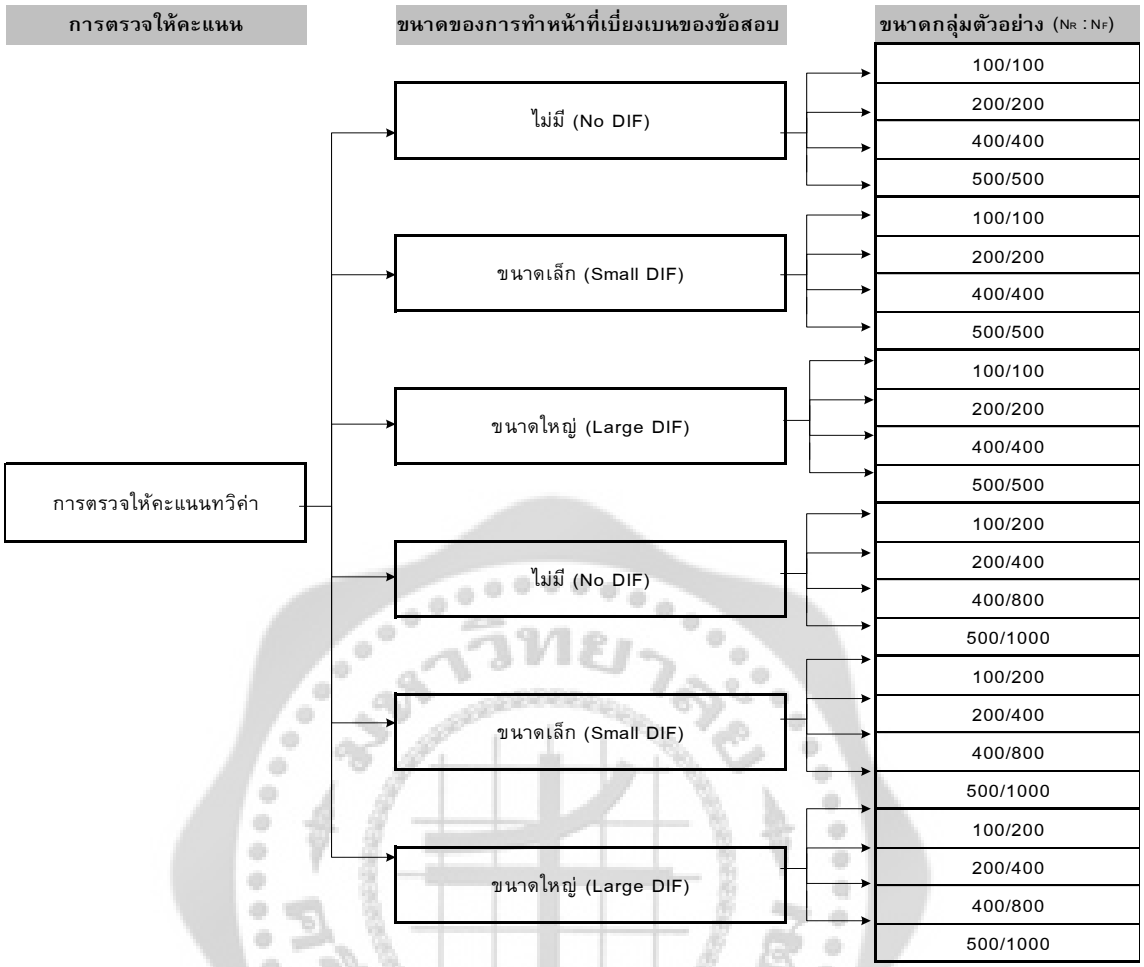
1.1 วิธีในการตรวจให้คะแนนของรายการคำตอบ มาตรวัดโดยทั่วไปจะมีวิธีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้คะแนนแบบสองค่า ซึ่งเป็นการตรวจให้คะแนนโดยคำตอบที่ถูกต้องหรือมีความเหมาะสมสูงที่สุดจะระบุการให้คะแนนเป็น 1 ส่วนตัวเลือกอื่นหรือคำตอบอื่น ๆ จะให้คะแนนเป็น 0 และการตรวจให้คะแนนแบบหลายค่า เป็นการระบุการให้คะแนนกับตัวเลือกทุกตัวตามคุณลักษณะและความเหมาะสม เช่น กำหนดให้แต่ละตัวเลือกในข้อสอบข้อ 1 เป็น 5, 3, 4, 2, 1 คะแนน เป็นต้น ซึ่งเป็นวิธีการให้คะแนนอีกวิธีหนึ่งที่พบว่านิยมใช้ในทางวัดผลการศึกษาเช่นกัน ผู้วิจัยจึงนำมาเป็นปัจจัยหนึ่งในการตรวจสอบความเท่าเทียมกันของการวัด

1.2 ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบออกเป็น 3 ระดับ คือ 1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพื่อทำการศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี 2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง และ 3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาด

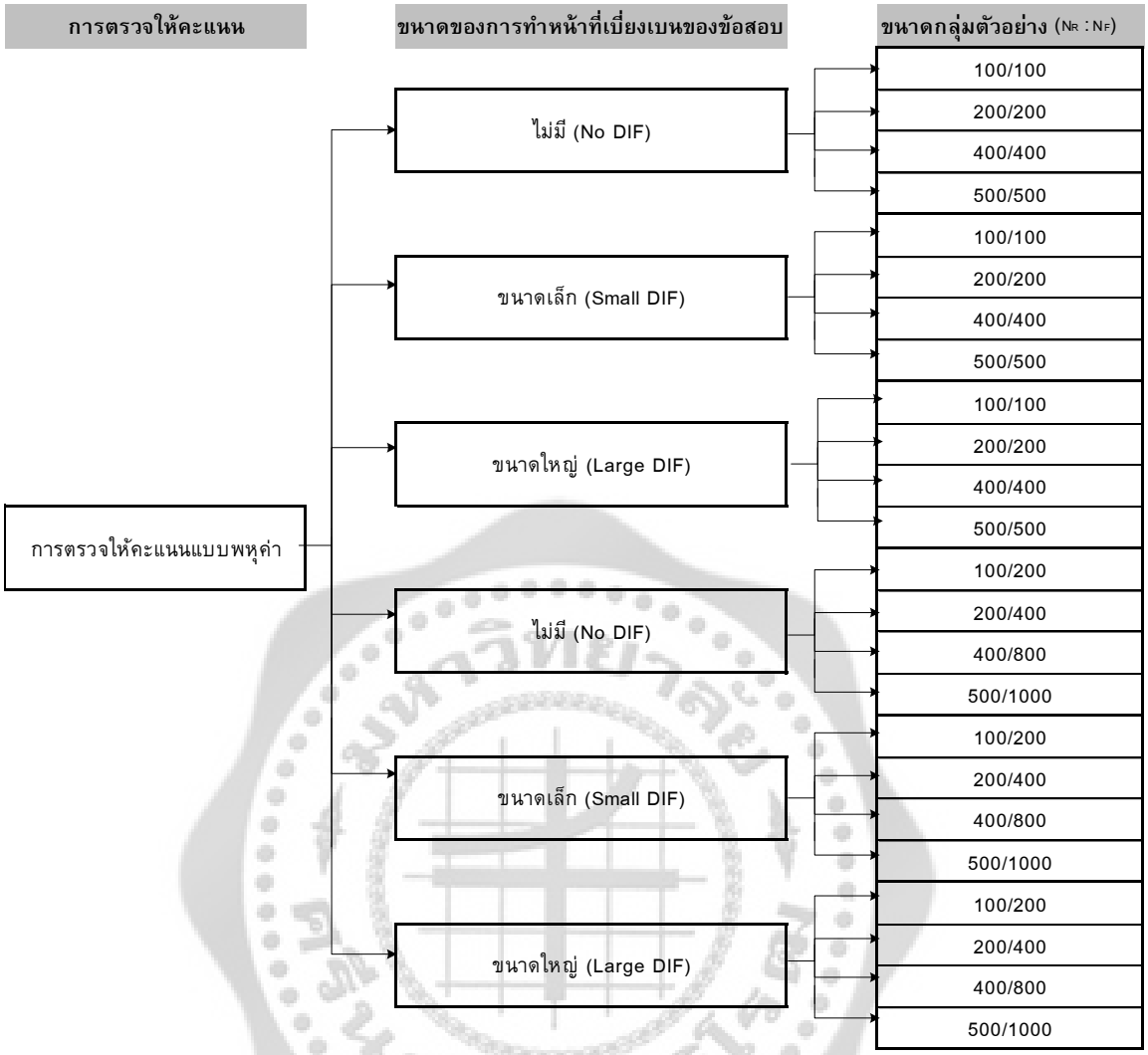
ใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง ซึ่งการที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ เพื่อทำการศึกษาอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี

1.3 ขนาดกลุ่มตัวอย่าง เนื่องจากการสอบนั้นจะต้องมีผู้สอบเข้ามาเกี่ยวข้องเสมอ ซึ่งตามสภาพโดยทั่วไปของการจัดการเรียนการสอนในโรงเรียนแล้ว จำนวนนักเรียนในระดับชั้นหนึ่งจะมีขนาดที่ไม่มากนัก แต่สำหรับการสอบในระดับเขตพื้นที่การศึกษา หรือการสอบในระดับประเทศแล้ว ขนาดของจำนวนผู้สอบจะมีจำนวนมากขึ้น ในงานวิจัยบางครั้งจะทำการศึกษาเปรียบเทียบความสามารถของนักเรียนซึ่งข้ามกลุ่มกัน ไม่ว่าจะเป็นระหว่างห้องเรียน ระหว่างโรงเรียน ระหว่างเขตพื้นที่การศึกษา จนถึงการสอบที่มีการเปรียบเทียบระหว่างประเทศ จากกลุ่มประชากรที่ต่างกันนี้ในบางครั้งก็จะมีจำนวนเท่ากันและบางครั้งก็จะมีจำนวนไม่เท่ากัน จะเห็นว่าขนาดกลุ่มตัวอย่างก็เป็นอีกปัจจัยหนึ่งที่จะเกิดขึ้นในระดับปฏิบัติการ และมีผลต่อการตรวจสอบความเท่าเทียมกันของการวัด ดังนั้น ผู้วิจัยจึงสนใจที่จะนำขนาดของกลุ่มตัวอย่างที่เป็นกลุ่มอ้างอิงและกลุ่มสนใจมาศึกษา โดยแบ่งตามสัดส่วนของกลุ่มอ้างอิงกับกลุ่มสนใจ 2 สัดส่วน ได้แก่ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 และสัดส่วน 1:2 มี 4 ขนาดเช่นกัน คือ 100:200, 200:400 400:800 และ 500:1000

ดังนั้นในการตรวจสอบความเท่าเทียมกันของการวัดครั้งนี้ ผู้วิจัยได้จำลองข้อมูลโดยจัดกระทำตามเงื่อนไขที่แปรเปลี่ยน 3 ปัจจัย คือ ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ 3 ขนาด ขนาดกลุ่มตัวอย่าง 8 ขนาด และวิธีการตรวจให้คะแนนรายการคำตอบ 2 แบบ รวมข้อมูลที่ต้องจำลองทั้งหมด 48 เงื่อนไข ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 100 ครั้ง สำหรับแผนการจำลองข้อมูลทั้ง 3 ปัจจัยแสดงดังภาพ



ภาพประกอบ 8 แผนภาพการจำลองข้อมูล



ภาพประกอบ 8 (ต่อ)

2. การจำลองข้อมูล (Simulation)

เนื่องจากการศึกษาวิจัยครั้งนี้ เป็นการศึกษาในข้อมูลจำลองรายการคำตอบของผู้สอบ ซึ่งเป็นการกระทำที่แตกต่างไปจากสถานการณ์ของการสอบจริงที่เมื่อทำการสอบแล้วจะได้รายการคำตอบของผู้สอบ แล้วจึงจะย้อนกลับไปตรวจสอบผลการตอบของผู้สอบให้ได้มาซึ่งความสามารถ และค่าพารามิเตอร์ของข้อสอบ แต่ในการศึกษาครั้งนี้ไม่ได้ทำการศึกษาจากข้อมูลจริง ดังนั้น หากจะให้ได้มาซึ่งรายการคำตอบแล้วจะต้องมีส่วนประกอบที่สำคัญ 2 ประการ นั่นคือ ค่าพารามิเตอร์ความสามารถของผู้สอบ และค่าพารามิเตอร์ของข้อสอบ แล้วนำส่วนประกอบทั้ง 2 อย่างดังกล่าวมาสัมพันธ์กันเพื่อสร้างรายการคำตอบ และครั้งนี้ผู้วิจัยได้ทำการจำลองข้อมูลจากโครงสร้างแบบทดสอบมิติเดียวที่รายการคำตอบข้อสอบมีการตรวจให้คะแนนแบบสองค่า และ

การตรวจให้คะแนนแบบหลายค่า โดยใช้โปรแกรมสำร็จรูป WinGen2 ของฮาน (Han. 2007) ซึ่งมีขั้นตอนหลัก ๆ ในการจำลองข้อมูล 3 ขั้นตอน ดังนี้

ขั้นที่ 1 การสร้างข้อมูลความสามารถของผู้สอบ

ในการศึกษาครั้งนี้ ผู้วิจัยสนใจที่จะศึกษากลุ่มตัวอย่างที่ประกอบด้วยกลุ่มอ้างอิงและกลุ่มสนใจที่มีสัดส่วนแตกต่างกัน 2 สัดส่วน คือ สัดส่วนกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) ประกอบด้วยขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 100:100 200:200 400:400 และ 500:500 คน และสัดส่วนกลุ่มอ้างอิงและกลุ่มสนใจ 1:2 ประกอบด้วยขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 100:200 200:400 400:800 และ 500:1000 คน

จากสัดส่วนของกลุ่มตัวอย่างที่ต้องการศึกษา ผู้วิจัยจำเป็นต้องทำการสร้างข้อมูลจำนวน 6 ชุด ได้แก่ จำนวนผู้สอบ 100, 200, 400, 500, 800 และ 1,000 คน เพื่อให้ครอบคลุมความสามารถของผู้สอบทั้งหมด ซึ่งโดยทั่วไปแล้วค่าความสามารถของผู้สอบจะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ถึงแม้ว่าความสามารถจะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ผลการวิเคราะห์ส่วนใหญ่มักจะให้ค่าอยู่ในช่วง -4 ถึง +4 ดังนั้น ผู้วิจัยจึงกำหนดให้การแจกแจงความสามารถมีการแจกแจงแบบปกติ โดยการระบุคะแนนเฉลี่ยความสามารถเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ผลการจำลองข้อมูลของผู้สอบได้พารามิเตอร์ของผู้สอบ ดังนี้

จำนวนผู้สอบ 100 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.12 ค่าความโด่งเท่ากับ -0.36

จำนวนผู้สอบ 200 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.51 ถึง 2.48 มีค่าเฉลี่ยเท่ากับ 0.00 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.98 ค่าความเบ้เท่ากับ -0.01 ค่าความโด่งเท่ากับ -0.11

จำนวนผู้สอบ 400 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.42 ถึง 2.97 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.93 ค่าความเบ้เท่ากับ -0.05 ค่าความโด่งเท่ากับ -0.04

จำนวนผู้สอบ 500 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -4.12 ถึง 2.82 มีค่าเฉลี่ยเท่ากับ 0.00 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.21 ค่าความโด่งเท่ากับ 0.63

จำนวนผู้สอบ 800 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.87 ถึง 3.07 มีค่าเฉลี่ยเท่ากับ -0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ 0.10 ค่าความโด่งเท่ากับ 0.02

จำนวนผู้สอบ 1,000 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -3.15 ถึง 3.35 มีค่าเฉลี่ยเท่ากับ 0.00 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ 0.02 ค่าความโด่งเท่ากับ 0.32

ขั้นที่ 2 การสร้างข้อมูลข้อสอบ

ในการศึกษาครั้งนี้ผู้วิจัยสนใจที่จะศึกษาแบบทดสอบที่มีโครงสร้างการวัดในมิติเดียว ที่รายการคำตอบข้อสอบมีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้แบบสองค่า มีจำนวน 25 ข้อ กำหนดการกระจายของคุณสมบัติข้อสอบซึ่งได้แก่ พารามิเตอร์ค่าอำนาจจำแนก และพารามิเตอร์ค่าความยาก เป็นแบบปกติ และการตรวจให้คะแนนแบบหลายค่า มีจำนวน 25 ข้อ กำหนดการกระจายของคุณสมบัติข้อสอบซึ่งได้แก่ พารามิเตอร์ค่าอำนาจจำแนก และพารามิเตอร์ค่าความยาก เป็นแบบปกติ เช่นกัน ดังนั้น ผู้วิจัยจึงกำหนดเงื่อนไขในการสร้างรายการข้อสอบให้เป็นไปตามรูปแบบรายการคำตอบ ดังนี้

1. รายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ผู้วิจัยเลือกใช้โมเดล 2 พารามิเตอร์ (Two-Parameter Model) ในการประมาณค่าพารามิเตอร์ของข้อสอบ โดยกำหนดให้ค่าอำนาจจำแนก และค่าความยาก กระจายเป็นแบบปกติ ซึ่งโดยทั่วไปแล้วค่าพารามิเตอร์อำนาจจำแนก (a_i parameter) ของข้อสอบข้อที่ i (Discrimination parameter) จะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ที่แสดงถึงความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i แต่ในทางปฏิบัติค่าพารามิเตอร์อำนาจจำแนกควรมีพิสัยอยู่ระหว่าง 0 ถึง 2 จากการสร้างพารามิเตอร์ข้อสอบพบว่า ค่าพารามิเตอร์อำนาจจำแนกมีพิสัยอยู่ระหว่าง -1.12 ถึง 3.12 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานแท่ากับ 0.95 ค่าความเบ้เป็น 0.09 และค่าความโด่งเป็น 0.17 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วนพารามิเตอร์ความยาก (b_i parameter) ของข้อสอบซึ่งมีช่วงอยู่ระหว่าง -2.0 ถึง +2.0 ค่าของความยากของข้อสอบจะแสดงตำแหน่งบนโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50 พารามิเตอร์ความยากของข้อสอบที่เพิ่มมากขึ้น ผู้สอบที่มีความสามารถมากขึ้นที่จะตอบข้อสอบข้อนั้นได้ถูกจะอยู่ในตำแหน่งที่สูงกว่า จากการสร้างพารามิเตอร์ข้อสอบ พบว่า ค่าพารามิเตอร์ความยากมีพิสัยอยู่ระหว่าง -1.27 ถึง 2.33 ค่าเฉลี่ยเท่ากับ 0.54 ส่วนเบี่ยงเบนมาตรฐานแท่ากับ 1.08 ค่าความเบ้เป็น -0.10 และค่าความโด่งเป็น -1.10

2. รายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า ผู้วิจัยเลือกใช้เลือกใช้ Graded - Response Model ซึ่งมีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบข้อสอบ 2 พารามิเตอร์ และใช้หลักในการคำนวณความน่าจะเป็นของการตอบแต่ละรายการคำตอบแบบ 2 ขั้นตอน โดยขั้นแรกจะทำการคำนวณค่าความชันร่วม (Common item slope parameter: a_i) ของแต่ละข้อคำถาม จากนั้นจึงจะทำการคำนวณ ค่า Threshold ของแต่ละรายการคำตอบ (Category threshold parameter: b_{ij}) ในแต่ละข้อคำถาม ในการประมาณค่าพารามิเตอร์ของข้อสอบครั้งนี้ ผู้วิจัยได้กำหนดให้ค่าพารามิเตอร์ของแต่ละรายการคำตอบข้อสอบ แจกแจงแบบปกติ

โดยให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความชันร่วมเป็น 1 ส่วนค่าเฉลี่ยของค่าความยากประจำขั้นแต่ละรายการคำตอบกำหนดให้เป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 จากข้อมูลจำลองที่ได้ พบว่า ค่าพารามิเตอร์ความชันร่วมมีพิสัยอยู่ระหว่าง -0.33 ถึง 3.04 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.21 ค่าความเบ้เป็น 0.47 และค่าความโด่งเป็น -0.71 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วนค่า Threshold ของแต่ละรายการคำตอบของข้อสอบมีพิสัยอยู่ระหว่าง -1.90 ถึง 2.41 ค่าเฉลี่ยเท่ากับ 0.56 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.96 ค่าความเบ้เป็น -0.08 และค่าความโด่งเป็น -0.51

2.3 การสร้างข้อมูลของข้อสอบศึกษา

เนื่องจากการศึกษาครั้งนี้เป็นการตรวจสอบความเท่าเทียมกันของการวัด จึงต้องมีการกำหนดข้อสอบศึกษาขึ้นมาเพื่อนำรายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจมาเปรียบเทียบกันในกระบวนการตรวจสอบความเท่าเทียมกันของการวัด

2.3.1 การสร้างข้อมูลของข้อสอบศึกษาที่เป็นการตรวจให้คะแนนแบบสองค่า

รายการคำตอบข้อสอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ผู้วิจัยกำหนดให้มีข้อสอบศึกษาจำนวน 5 ข้อ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.17 พารามิเตอร์ค่าความยากเท่ากับ 0.25 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.07 พารามิเตอร์ค่าความยากเท่ากับ 0.38 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.15 พารามิเตอร์ค่าความยากเท่ากับ 1.12 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.12 พารามิเตอร์ค่าความยากเท่ากับ -1.27 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.52 พารามิเตอร์ค่าความยากเท่ากับ 0.35 จากนั้นทำการกำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษาแต่ละข้อ เพื่อนำไปสร้างรายการคำตอบ เป็น 3 ขนาด คือ

1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน จากการจำลองข้อมูลได้ค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.17 พารามิเตอร์ค่าความยากเท่ากับ 0.25 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.07 พารามิเตอร์ค่าความยากเท่ากับ 0.38 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.15 พารามิเตอร์ค่าความยากเท่ากับ 1.12 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.12 พารามิเตอร์ค่าความยากเท่ากับ -1.27 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.52 พารามิเตอร์ค่าความยากเท่ากับ 0.35

2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์

ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.02 พารามิเตอร์ค่าความยากเท่ากับ 0.50 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.92 พารามิเตอร์ค่าความยากเท่ากับ 0.63 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.00 พารามิเตอร์ค่าความยากเท่ากับ 1.37 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.97 พารามิเตอร์ค่าความยากเท่ากับ -1.02 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.37 พารามิเตอร์ค่าความยากเท่ากับ 0.60

3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.57 พารามิเตอร์ค่าความยากเท่ากับ 0.75 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.47 พารามิเตอร์ค่าความยากเท่ากับ 0.88 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.55 พารามิเตอร์ค่าความยากเท่ากับ 1.62 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.52 พารามิเตอร์ค่าความยากเท่ากับ -0.77 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.92 พารามิเตอร์ค่าความยากเท่ากับ 0.85

2.3.2 การสร้างข้อมูลของข้อสอบศึกษาที่เป็นการตรวจให้คะแนนแบบหลายค่า

รายการคำตอบข้อสอบของผู้สอบมีการตรวจให้คะแนนแบบหลายค่า ผู้วิจัยกำหนดให้มีข้อสอบศึกษาจำนวน 5 ข้อ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.45 ค่าความยากประจำขั้นเป็น -0.59 -0.15 1.33 1.79 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.90 ค่าความยากประจำขั้นเป็น -0.18 0.67 0.78 2.41 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 3.04 ค่าความยากประจำขั้นเป็น -0.31 0.29 1.23 1.90 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.59 ค่าความยากประจำขั้นเป็น -0.76 -0.11 0.25 1.32 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.71 ค่าความยากประจำขั้นเป็น 0.37 1.08 1.77 1.81 ตามลำดับ จากนั้นทำการกำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษาแต่ละข้อเพื่อนำไปสร้างรายการคำตอบ เป็น 3 ขนาด คือ

1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.45 ค่าความยากประจำขั้นเป็น -0.59 -0.15 1.33 1.79 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.90 ค่าความยากประจำขั้นเป็น -0.18 0.67 0.78 2.41 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 3.04 ค่าความยากประจำขั้นเป็น -0.31 0.29

1.23 1.90 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.59 ค่าความยากประจำขั้นเป็น -0.76 -0.11 0.25 1.32 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.71 ค่าความยากประจำขั้นเป็น 0.37 1.08 1.77 1.81 ตามลำดับ

2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.30 ค่าความยากประจำขั้นเป็น -0.34 0.10 1.58 2.04 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.75 ค่าความยากประจำขั้นเป็น 0.07 0.92 1.03 2.66 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.89 ค่าความยากประจำขั้นเป็น -0.06 0.54 1.48 2.15 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.44 ค่าความยากประจำขั้นเป็น -0.51 0.14 0.50 1.57 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.56 ค่าความยากประจำขั้นเป็น 0.62 1.33 2.02 2.06 ตามลำดับ

3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.05 ค่าความยากประจำขั้นเป็น -0.09 0.35 1.83 2.29 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.50 ค่าความยากประจำขั้นเป็น 0.32 1.17 1.28 2.91 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.64 ค่าความยากประจำขั้นเป็น 0.19 0.79 1.73 2.40 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.19 ค่าความยากประจำขั้นเป็น -0.26 0.39 0.75 1.82 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.31 ค่าความยากประจำขั้นเป็น 0.87 1.58 2.27 2.31 ตามลำดับ

ขั้นที่ 3 การสร้างข้อมูลการตอบข้อสอบ

หลักในการสร้างรายการคำตอบของผู้สอบนั้นจะต้องนำความสามารถของผู้สอบไปสัมพันธ์กับค่าพารามิเตอร์ของข้อสอบ ซึ่งความสามารถของผู้สอบและพารามิเตอร์ข้อสอบนั้นผู้วิจัยได้สร้างขึ้นแล้วในขั้น 1 และ 2 ตามลำดับ และเนื่องจากการศึกษาครั้งนี้เป็นการตรวจสอบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด แล้วนำผลการตรวจสอบในแต่ละวิธีมาเปรียบเทียบกัน เพื่อเป็นการควบคุมความคลาดเคลื่อนในการสุ่มรายการคำตอบของกลุ่มตัวอย่าง ผู้วิจัยจำลองข้อมูลซ้ำจำนวน 100 ครั้ง ทำให้ได้รายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจในแต่ละเงื่อนไขจำนวน 100 ชุด

3. การวิเคราะห์ข้อมูล

การศึกษาวิจัยครั้งนี้ มีความมุ่งหมายเพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด โดยผู้วิจัยเลือกวิธีดำเนินการในการตรวจสอบความเท่าเทียมกันของการวัด 2 วิธี คือ วิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยวิธีทดสอบอัตราส่วนความเป็นไปได้ ซึ่งมีรายละเอียดของการวิเคราะห์ข้อมูล ดังนี้

3.1 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด

ในการวิเคราะห์อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด ผู้วิจัยนำรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า และที่เป็นการตรวจให้คะแนนแบบหลายค่า ที่ได้จากการจำลองข้อมูลที่มีการระบุขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ ในกลุ่มอ้างอิงและกลุ่มสนใจ มาวิเคราะห์ค่าทางสถิติเพื่อระบุถึงการตรวจสอบความเท่าเทียมกันของการวัดในข้อสอบศึกษาว่าตรวจสอบได้หรือไม่ ดังนี้

3.1.1 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปร

ความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.1.2 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ ซึ่งมีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.2 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

การวิเคราะห์อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด ผู้วิจัยนำรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า และการตรวจให้คะแนนแบบหลายค่า ที่ได้จากการจำลองข้อมูลที่ระบุไม่ให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบ นั่นคือ ทั้งกลุ่มอ้างอิงและกลุ่มสนใจจะมีค่าพารามิเตอร์ของข้อสอบศึกษาเท่ากัน มาวิเคราะห์ค่าทางสถิติ เพื่อศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดจากข้อสอบศึกษาว่ามีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดเท่าใด ดังนี้

3.2.1 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไป จะให้ค่าน้ำหนักของการเป็นสมาชิกกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัดซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.2.2 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยวิธีการทดสอบอัตราส่วนความเป็นไปได้ โมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็น

การประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยากและพารามิเตอร์อำนาจจำแนกเท่ากัน ข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าทำการตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

3.3 การเปรียบเทียบผลการตรวจสอบความเท่าเทียมกันจากวิธีที่ต่างกัน

ในการเปรียบเทียบผลการตรวจสอบความเท่าเทียมกันจากวิธีการทดสอบที่ต่างกัน ผู้วิจัยเลือกใช้สถิติ McNemar (McNemar. 1962) ดังนี้

		Classification of the Y_i	
		$Y_i = 0$	$Y_i = 1$
Classification of the X_i	$X_i = 0$	a	b
	$X_i = 1$	c	d

ในการทดสอบที่ระดับนัยสำคัญทางสถิติที่ระดับ .05 มีสูตรคำนวณ ดังนี้

$$T_i = \frac{(b-c)^2}{b+c}$$

เมื่อ Ti แทน สถิติ McNemar ซึ่งมีการแจกแจงแบบไค-สแควร์

Xi แทน ผลการตรวจสอบความเท่าเทียมกันของวิธีที่ 1 ในชุดข้อมูล i

1 แทน ผลการตรวจสอบความเท่าเทียมกันมีนัยสำคัญ

0 แทน ผลการตรวจสอบความเท่าเทียมกันไม่มีนัยสำคัญ

Yi แทน ผลการตรวจสอบความเท่าเทียมกันของวิธีที่ 2 ในชุดข้อมูล i

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยครั้งนี้ มีความมุ่งหมายเพื่อเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด ในการศึกษาครั้งนี้ ผู้วิจัยเลือกวิธีดำเนินการในการตรวจสอบความเท่าเทียมกันของการวัด 2 วิธี คือ วิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 ปัจจัย ดังนี้ **ปัจจัยที่ 1** วิธีในการตรวจให้คะแนนของการตอบข้อสอบ มาตรวัดโดยทั่วไปจะมีวิธีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้คะแนนแบบสองค่า (0, 1) และการตรวจให้คะแนนแบบหลายค่า (4, 3, 2, 1, 0) **ปัจจัยที่ 2** ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบออกเป็น 3 ระดับ คือ 1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนกและค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพื่อทำการศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี 2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนก ของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง และ 3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง ซึ่งการที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ เพื่อทำการศึกษาอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี และ**ปัจจัยที่ 3** ขนาดกลุ่มตัวอย่าง โดยในการศึกษาครั้งนี้ ผู้วิจัยแบ่งตามสัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิง 2 สัดส่วน ได้แก่ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 คน และสัดส่วน 1:2 มี 4 ขนาดเช่นกัน คือ 100:200, 200:400 400:800 และ 500:1000 คน

ในการศึกษาครั้งนี้ ผู้วิจัยได้นำเสนอผลการวิเคราะห์ข้อมูลตามขั้นตอนดังนี้

ตอนที่ 1 ผลการจำลองข้อมูล (Simulation)

ตอนที่ 2 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 3 การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

ตอนที่ 4 สรุปผลการศึกษาในภาพรวม

สัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์ข้อมูล

เพื่อให้การเสนอผลการวิเคราะห์ข้อมูลมีความสะดวกและเกิดความเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์และการแปรผลการวิเคราะห์ข้อมูล ดังนี้

CTT	แทน	วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน
IRT	แทน	วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้
Min	แทน	ค่าต่ำสุดสำหรับชุดข้อมูลนั้น ๆ
Max	แทน	ค่าสูงสุดสำหรับชุดข้อมูลนั้น ๆ
Mean	แทน	ค่าเฉลี่ยสำหรับชุดข้อมูลนั้น ๆ
SD	แทน	ส่วนเบี่ยงเบนมาตรฐานสำหรับชุดข้อมูลนั้น ๆ
Skewness	แทน	ค่าความเบ้สำหรับชุดข้อมูลนั้น ๆ
Kurtosis	แทน	ค่าความโด่งสำหรับชุดข้อมูลนั้น ๆ
a	แทน	ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ หรือค่าความชันร่วมของข้อสอบ
b	แทน	ค่าพารามิเตอร์ความยากของข้อสอบ หรือค่าความยากประจำขั้นของข้อสอบ
*	แทน	ระดับนัยสำคัญทางสถิติที่ระดับ .05

ตอนที่ 1 ผลการศึกษาการจำลองข้อมูล (Simulation)

เนื่องจากการศึกษาวิจัยครั้งนี้ เป็นการศึกษาในข้อมูลจำลองรายการคำตอบของผู้สอบ ซึ่งเป็นการกระทำที่แตกต่างไปจากสถานการณ์ของการสอบจริงที่เมื่อทำการสอบแล้วจะได้รายการคำตอบของผู้สอบแล้วจึงจะย้อนกลับไปตรวจสอบผลการตอบของผู้สอบให้ได้มาซึ่งความสามารถและค่าพารามิเตอร์ของข้อสอบ แต่ในการศึกษาครั้งนี้ไม่ได้ทำการศึกษาจากข้อมูลจริง ดังนั้น หากจะให้ได้มาของรายการคำตอบแล้วจะต้องมีส่วนประกอบที่สำคัญ 2 ประการ นั่นคือ ค่าพารามิเตอร์ความสามารถของผู้สอบ และค่าพารามิเตอร์ของข้อสอบ แล้วนำส่วนประกอบทั้ง 2 อย่างดังกล่าวมาสัมพันธ์กันเพื่อสร้างรายการคำตอบ และการศึกษาครั้งนี้ผู้วิจัยได้ทำการจำลองข้อมูลจากโครงสร้างแบบทดสอบมิติเดียวที่รายการคำตอบข้อสอบมีการตรวจให้คะแนนแบบสองค่า และการตรวจให้คะแนนแบบหลายค่า โดยใช้โปรแกรมสำร็จรูป WinGen2 ของฮาน (Han. 2007) ซึ่งมีขั้นตอนในการจำลองข้อมูล ดังนี้

1.1 การสร้างข้อมูลความสามารถของผู้สอบ

ในการศึกษาครั้งนี้ ผู้วิจัยสนใจที่จะศึกษากลุ่มตัวอย่างที่ประกอบด้วยกลุ่มอ้างอิงและกลุ่มสนใจที่มีสัดส่วนแตกต่างกัน 2 สัดส่วน คือ สัดส่วนกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) ประกอบด้วยขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 คน และสัดส่วนกลุ่มอ้างอิงและกลุ่มสนใจ 1:2 ประกอบด้วยขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 100:200, 200:400 400:800 และ 500:1000 คน

กรณีสัดส่วนของกลุ่มตัวอย่างที่ต้องการศึกษา ผู้วิจัยจำเป็นต้องทำการสร้างข้อมูลจำนวน 6 ชุด ได้แก่ จำนวนผู้สอบ 100, 200, 400, 500, 800 และ 1,000 คน เพื่อให้ครอบคลุมความสามารถของผู้สอบทั้งหมด ซึ่งโดยทั่วไปแล้วค่าความสามารถของผู้สอบจะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ถึงแม้ว่าความสามารถจะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ผลการวิเคราะห์ส่วนใหญ่มักจะให้ค่าอยู่ในช่วง -4 ถึง +4 ดังนั้น ผู้วิจัยจึงกำหนดให้การแจกแจงความสามารถมีการแจกแจงแบบปกติ การศึกษาวิจัยในครั้งนี้ ผู้วิจัยดำเนินการสร้างค่าพารามิเตอร์ความสามารถของผู้สอบจำนวน 100 คน ให้เป็นไปตามเงื่อนไขก่อน กล่าวคือ กลุ่มประชากรย่อยจะมีคะแนนความสามารถเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 เมื่อได้พารามิเตอร์ของผู้สอบจำนวน 100 คน แล้ว ผู้วิจัยได้ทำสำเนาความสามารถของผู้สอบ 100 คน ขึ้นเป็น 2 เท่า เพื่อให้ได้พารามิเตอร์ของผู้สอบจำนวน 200 คน ทำสำเนาความสามารถของผู้สอบ 100 คน ขึ้นเป็น 4 เท่า เพื่อให้ได้พารามิเตอร์ของผู้สอบจำนวน 400 คน ทำสำเนาความสามารถของผู้สอบ 100 คน ขึ้นเป็น 5 เท่า เพื่อให้ได้พารามิเตอร์ของผู้สอบจำนวน 500 คน ทำสำเนาความสามารถของผู้สอบ 100 คน ขึ้นเป็น 8 เท่า เพื่อให้ได้พารามิเตอร์ของผู้สอบจำนวน 800 คน และทำสำเนาความสามารถของผู้สอบ 100 คน ขึ้นเป็น 10 เท่า เพื่อให้ได้พารามิเตอร์ของผู้สอบจำนวน 1,000 คน ทั้งนี้เพื่อควบคุมค่าพารามิเตอร์ของประชากรผู้สอบกลุ่มย่อยทุกกลุ่มให้มีค่าพารามิเตอร์เท่ากัน เพียงแต่จำนวนประชากรผู้สอบมีขนาดต่างกันเท่านั้นเอง ผลการจำลองข้อมูลของผู้สอบได้พารามิเตอร์ของผู้สอบ ดังนี้

จำนวนผู้สอบ 100 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.12 ค่าความโด่งเท่ากับ -0.36

จำนวนผู้สอบ 200 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.38

จำนวนผู้สอบ 400 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.39

จำนวนผู้สอบ 500 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40

จำนวนผู้สอบ 800 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40

จำนวนผู้สอบ 1,000 คน มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40

จากข้อมูลค่าพารามิเตอร์ของผู้สอบดังกล่าว ชี้ให้เห็นถึงค่าพารามิเตอร์ความสามารถของผู้สอบที่จำลองขึ้นมาศึกษาทุกกลุ่มประชากรย่อยมีระดับความสามารถเดียวกัน ซึ่งเป็นไปตามเงื่อนไขที่กำหนดไว้ ผลการจำลองข้อมูลปรากฏตามราง 1 และเพื่อให้เห็นภาพของการกระจายค่าพารามิเตอร์ความสามารถของผู้สอบ ผู้วิจัยจึงนำข้อมูลการกระจายของค่าพารามิเตอร์ความสามารถของผู้สอบมาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล เมื่อพิจารณาค่าสถิติต่าง ๆ ได้แก่ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และค่าความโด่ง ประกอบกับรูปร่างของกราฟเส้นแล้วแสดงให้เห็นว่า ความสามารถของผู้สอบทุกกลุ่ม มีการกระจายค่อนข้างจะเป็นโค้งปกติ (ดังภาพประกอบ 9-14 ภาคผนวก ก)

ตาราง 1 ระดับความสามารถของผู้สอบจำแนกตามขนาดของผู้สอบ

ขนาดของผู้สอบ (คน)	ระดับความสามารถของผู้สอบ					
	Min	Max	Mean	SD	Skewness	Kurtosis
100	-2.49	2.14	0.01	1.01	-0.12	-0.36
200	-2.49	2.14	0.01	1.01	-0.11	-0.38
400	-2.49	2.14	0.01	1.01	-0.11	-0.39
500	-2.49	2.14	0.01	1.01	-0.11	-0.40
800	-2.49	2.14	0.01	1.01	-0.11	-0.40
1000	-2.49	2.14	0.01	1.01	-0.11	-0.40

1.2 การสร้างข้อมูลของข้อสอบ

ในการศึกษาครั้งนี้ผู้วิจัยสนใจที่จะศึกษาแบบทดสอบที่มีโครงสร้างการวัดในมิติเดียว ที่รายการคำตอบข้อสอบมีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้แบบสองค่า มีจำนวน 25 ข้อ กำหนดการกระจายของคุณสมบัติข้อสอบซึ่งได้แก่ พารามิเตอร์ค่าอำนาจจำแนก และพารามิเตอร์ค่าความยาก เป็นแบบปกติ และการตรวจให้คะแนนแบบหลายค่า มีจำนวน 25 ข้อ กำหนดการกระจายของคุณสมบัติข้อสอบซึ่งได้แก่ พารามิเตอร์ค่าอำนาจจำแนก และพารามิเตอร์ค่าความยาก เป็นแบบปกติ เช่นกัน ดังนั้น ผู้วิจัยจึงกำหนดเงื่อนไขในการสร้างรายการข้อสอบให้เป็นไปตามรูปแบบรายการคำตอบ ดังนี้

1.2.1 รายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ผู้วิจัยเลือกใช้โมเดล 2 พารามิเตอร์ (Two-Parameter Model) ในการประมาณค่าพารามิเตอร์ของข้อสอบ โดยกำหนดให้ค่าอำนาจจำแนก และค่าความยาก กระจายเป็นแบบปกติ ซึ่งโดยทั่วไปแล้วค่าพารามิเตอร์อำนาจจำแนก (a_i parameter) ของข้อสอบข้อที่ i (Discrimination parameter) ซึ่งจะมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ที่แสดงถึงความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i ค่าพารามิเตอร์อำนาจจำแนกที่ควรเป็น ควรจะมีพิสัยอยู่ระหว่าง 0 ถึง 2 จากการสร้างพารามิเตอร์ข้อสอบ พบว่า ค่าพารามิเตอร์อำนาจจำแนกมีพิสัยอยู่ระหว่าง -1.12 ถึง 3.12 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.95 ค่าความเบ้เป็น 0.09 และค่าความโด่งเป็น 0.17 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วนพารามิเตอร์ความยาก (b_i parameter) ของข้อสอบซึ่งมีช่วงอยู่ระหว่าง -2.0 ถึง +2.0 ค่าของความยากของข้อสอบจะแสดงตำแหน่งบนโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50 พารามิเตอร์ความยากของข้อสอบที่เพิ่มมากขึ้น ผู้สอบที่มีความสามารถมากขึ้นที่จะตอบข้อสอบข้อนั้นได้ถูกจะอยู่ในตำแหน่งที่สูงกว่า จากการสร้างพารามิเตอร์ข้อสอบ พบว่า ค่าพารามิเตอร์ความยากมีพิสัยอยู่ระหว่าง -1.27 ถึง 2.33 ค่าเฉลี่ยเท่ากับ 0.54 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.08 ค่าความเบ้เป็น -0.10 และค่าความโด่งเป็น -1.10 ซึ่งอยู่ในขอบเขตของพารามิเตอร์ความยากตามเกณฑ์ที่กำหนดไว้ รายละเอียดพารามิเตอร์ข้อสอบ ปรากฏตามตาราง 2 และเพื่อให้เห็นภาพของการกระจายค่าพารามิเตอร์ข้อสอบ ผู้วิจัยจึงนำข้อมูลการกระจายของค่าพารามิเตอร์ข้อสอบมาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล เมื่อพิจารณาค่าสถิติต่าง ๆ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง ประกอบกับรูปร่างของกราฟเส้นแล้วแสดงให้เห็นว่า พารามิเตอร์ข้อสอบทั้งค่าอำนาจจำแนกและค่าความยาก มีการกระจายค่อนข้างจะเป็นโค้งปกติ (ดังภาพประกอบ 15-16 ภาคผนวก ก)

ตาราง 2 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า

ข้อที่	พารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง	
	a	b
1	0.52	1.97
2	1.41	-0.68
3	1.08	-0.80
4	0.26	-0.12
5	2.17	0.25
6	1.40	-0.42
7	-0.04	-1.15
8	2.61	0.25
9	0.50	1.94
10	2.07	0.38
11	-1.12	1.41
12	0.91	-0.06
13	1.71	1.43
14	0.23	0.52
15	2.15	1.12
16	0.58	0.70
17	0.62	1.59
18	0.97	2.33
19	0.05	1.86
20	3.12	-1.27
21	0.74	0.11
22	0.35	1.55
23	1.97	1.30
24	1.09	-1.08
25	1.52	0.35
Min	-1.12	-1.27
Max	3.12	2.33
Mean	1.07	0.54
SD	0.95	1.08
Skewness	0.09	-0.10
Kurtosis	0.17	-1.10

1.2.2 รายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า ผู้วิจัยเลือกใช้เลือกใช้โมเดลเกรดเรสพอนส์ (Graded - Response Model) ซึ่งโมเดลเกรดเรสพอนส์ มีลักษณะเป็นโมเดลทั่วไปของโมเดลการตอบข้อสอบที่มี 2 พารามิเตอร์ และใช้หลักในการคำนวณความน่าจะเป็นของการตอบแต่ละรายการคำตอบแบบ 2 ขั้นตอน โดยขั้นแรกจะทำการคำนวณค่าความชันร่วม (Common item slope parameter: a_i) ของแต่ละข้อคำถาม จากนั้นจึงจะทำการคำนวณ ค่า Threshold ของแต่ละรายการคำตอบ (Category threshold parameter: b_{ij}) ในแต่ละข้อคำถามในการประมาณค่าพารามิเตอร์ของข้อสอบครั้งนี้ ผู้วิจัยได้กำหนดให้ค่าพารามิเตอร์ของแต่ละรายการคำตอบข้อสอบ แจกแจงแบบปกติ จึงกำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความชันร่วมเป็น 1 ส่วนค่าเฉลี่ยของค่าความยากประจำขั้นแต่ละรายการคำตอบกำหนดให้เป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 จากข้อมูลจำลองที่ได้ พบว่า ค่าพารามิเตอร์ความชันร่วมมีพิสัยอยู่ระหว่าง -0.33 ถึง 3.04 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานแท่ากับ 0.21 ค่าความเบ้เป็น 0.47 และค่าความโด่งเป็น -0.71 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วน ค่า Threshold ของแต่ละรายการคำตอบของข้อสอบมีพิสัยอยู่ระหว่าง -1.90 ถึง 2.41 ค่าเฉลี่ยเท่ากับ 0.56 ส่วนเบี่ยงเบนมาตรฐานแท่ากับ 0.96 ค่าความเบ้เป็น -0.08 และค่าความโด่งเป็น -0.51 ซึ่งอยู่ในขอบเขตของพารามิเตอร์ความยากตามเกณฑ์ที่กำหนดไว้ รายละเอียดของพารามิเตอร์ข้อสอบ ปรากฏตามตาราง 3 และเพื่อให้เห็นภาพของการกระจายค่าพารามิเตอร์ข้อสอบ ผู้วิจัยจึงนำข้อมูลการกระจายของค่าพารามิเตอร์ข้อสอบมาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล เมื่อพิจารณาค่าสถิติต่าง ๆ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง ประกอบกับรูปร่างของกราฟเส้นแล้วแสดงให้เห็นว่า พารามิเตอร์ข้อสอบทั้งค่าอำนาจจำแนกและค่าความยากประจำขั้นแต่ละรายการคำตอบ มีการกระจายค่อนข้างจะเป็นโค้งปกติ (ดังภาพประกอบ 17-28 ภาคผนวก ก)

ตาราง 3 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า

ข้อที่	พารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง					
	a	b1	b2	b3	b4	b
1	0.15	-0.72	0.09	0.13	0.48	0.00
2	-0.10	-0.93	0.08	0.71	1.27	0.28
3	-0.28	0.03	0.13	0.17	1.52	0.46
4	0.68	-0.58	0.44	0.97	1.33	0.54
5	1.45	-0.59	-0.15	1.33	1.79	0.60
6	0.64	-0.17	0.10	1.05	2.01	0.75
7	0.82	0.52	0.83	0.89	1.90	1.03
8	2.58	-0.48	0.11	1.61	1.63	0.72
9	0.36	-1.87	-0.46	0.12	1.16	-0.26
10	2.90	-0.18	0.67	0.78	2.41	0.92
11	1.07	0.89	1.10	1.43	1.48	1.23
12	-0.24	-0.42	0.58	0.70	0.95	0.45
13	0.92	0.06	0.69	0.76	1.42	0.73
14	0.94	-0.85	-0.12	1.20	1.51	0.44
15	3.04	-0.31	0.29	1.23	1.90	0.78
16	0.49	-0.44	1.17	1.27	2.30	1.08
17	0.97	-0.67	0.67	0.94	0.98	0.48
18	-0.25	-0.40	0.25	0.40	0.87	0.28
19	1.39	0.35	0.37	0.50	1.56	0.69
20	1.59	-0.76	-0.11	0.25	1.32	0.17
21	1.79	-1.39	-1.29	-0.39	2.41	-0.16
22	-0.33	-0.72	0.17	0.86	1.01	0.33
23	1.11	-1.90	-1.81	0.47	2.32	-0.23
24	2.40	0.56	1.49	1.50	1.97	1.38
25	2.71	0.37	1.08	1.77	1.81	1.26
Min	-0.33	-1.90	-1.81	-0.39	0.48	-1.90
Max	3.04	0.89	1.49	1.77	2.41	2.41
Mean	1.07	-0.42	0.26	0.83	1.57	1.87
SD	0.21	0.14	0.14	0.11	0.10	2.42
Skewness	0.47	-0.30	-1.05	-0.28	-0.07	-0.08
Kurtosis	-0.71	0.28	2.19	-0.33	-0.52	-0.51

1.3 การสร้างข้อมูลของข้อสอบศึกษา

1.3.1 กรณีเป็นการตรวจให้คะแนนแบบสองค่า

เนื่องจากการศึกษาครั้งนี้เป็นการตรวจสอบความเท่าเทียมกันของการวัด จึงต้องมีการกำหนดข้อสอบศึกษาขึ้นมาเพื่อนำรายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจมาเปรียบเทียบกันในกระบวนการตรวจสอบความเท่าเทียมกันของการวัด ดังนั้น ในการศึกษาครั้งนี้ในส่วนของรายการคำตอบข้อสอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ผู้วิจัยจึงกำหนดให้มีข้อสอบศึกษาจำนวน 5 ข้อ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.17 พารามิเตอร์ค่าความยากเท่ากับ 0.25 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.07 พารามิเตอร์ค่าความยากเท่ากับ 0.38 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.15 พารามิเตอร์ค่าความยากเท่ากับ 1.12 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.12 พารามิเตอร์ค่าความยากเท่ากับ -1.27 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.52 พารามิเตอร์ค่าความยากเท่ากับ 0.35 จากนั้นทำการกำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษาแต่ละข้อ เพื่อนำไปสร้างรายการคำตอบ เป็น 3 ขนาด คือ

1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน จากการจำลองข้อมูลได้ค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.17 พารามิเตอร์ค่าความยากเท่ากับ 0.25 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.07 พารามิเตอร์ค่าความยากเท่ากับ 0.38 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.15 พารามิเตอร์ค่าความยากเท่ากับ 1.12 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.12 พารามิเตอร์ค่าความยากเท่ากับ -1.27 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.52 พารามิเตอร์ค่าความยากเท่ากับ 0.35

2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.02 พารามิเตอร์ค่าความยากเท่ากับ 0.50 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.92 พารามิเตอร์ค่าความยากเท่ากับ 0.63 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.00 พารามิเตอร์ค่าความยากเท่ากับ 1.37 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.97 พารามิเตอร์ค่าความยากเท่ากับ -1.02 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.37 พารามิเตอร์ค่าความยากเท่ากับ 0.60

3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์

ของข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.57 พารามิเตอร์ค่าความยากเท่ากับ 0.75 ข้อ 10 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.47 พารามิเตอร์ค่าความยากเท่ากับ 0.88 ข้อ 15 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 2.55 พารามิเตอร์ค่าความยากเท่ากับ 1.62 ข้อ 20 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 3.52 พารามิเตอร์ค่าความยากเท่ากับ -0.77 และข้อ 25 พารามิเตอร์ค่าอำนาจจำแนกเท่ากับ 1.92 พารามิเตอร์ค่าความยากเท่ากับ 0.85 ผลการสร้างข้อมูลของข้อสอบศึกษาปรากฏตามตาราง 4

ตาราง 4 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษา กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า

ข้อที่	พารามิเตอร์ข้อสอบสำหรับกลุ่มอ้างอิง		พารามิเตอร์ข้อสอบสำหรับกลุ่มสนใจ					
			ไม่มีการเบี่ยงเบน		เบี่ยงเบนขนาดเล็ก		เบี่ยงเบนขนาดใหญ่	
	a	b	a	b	a	b	a	b
5	0.22	-0.31	2.17	0.25	2.02	0.50	2.57	0.75
	ผลต่าง		0.00	0.00	-0.15	0.25	0.40	0.50
10	-1.06	0.77	-1.06	0.77	-1.21	1.02	-1.46	1.27
	ผลต่าง		0.00	0.00	-0.15	0.25	0.40	0.50
15	0.68	0.52	0.68	0.52	0.53	0.77	0.28	1.02
	ผลต่าง		0.00	0.00	-0.15	0.25	0.40	0.50
20	0.03	0.23	0.03	0.23	-0.12	0.48	-0.37	0.73
	ผลต่าง		0.00	0.00	-0.15	0.25	0.40	0.50
25	-0.37	1.13	-0.37	1.13	-0.52	1.38	-0.77	1.63
	ผลต่าง		0.00	0.00	-0.15	0.25	0.40	0.50

1.3.2 กรณีเป็นการตรวจให้คะแนนแบบหลายค่า

เนื่องจากการศึกษาครั้งนี้เป็นการตรวจสอบความเท่าเทียมกันของการวัด จึงต้องมีการกำหนดข้อสอบศึกษาขึ้นมาเพื่อนำรายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจมาเปรียบเทียบกันในกระบวนการตรวจสอบความเท่าเทียมกันของการวัด ดังนั้น ในการศึกษาครั้งนี้ในส่วนของรายการคำตอบข้อสอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า ผู้วิจัยจึงกำหนดให้มีข้อสอบศึกษาจำนวน 5 ข้อ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.45 ค่าความยากประจำขั้นเป็น -0.59 -0.15 1.33 1.79 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.90 ค่าความยากประจำขั้นเป็น -0.18 0.67 0.78 2.41 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 3.04 ค่าความยากประจำขั้นเป็น -0.31 0.29 1.23 1.90 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.59 ค่าความยากประจำขั้นเป็น -0.76 -0.11 0.25 1.32 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.71 ค่าความยากประจำขั้นเป็น 0.37 1.08 1.77 1.81 ตามลำดับ จากนั้นทำการกำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบศึกษาแต่ละข้อ เพื่อนำไปสร้างรายการคำตอบ เป็น 3 ขนาด คือ

1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ จากข้อสอบทั้งหมด 25 ข้อ ได้แก่ ข้อที่ 5, 10, 15, 20 และ 25 แต่ละข้อมีค่าพารามิเตอร์ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.45 ค่าความยากประจำขั้นเป็น -0.59 -0.15 1.33 1.79 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.90 ค่าความยากประจำขั้นเป็น -0.18 0.67 0.78 2.41 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 3.04 ค่าความยากประจำขั้นเป็น -0.31 0.29 1.23 1.90 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.59 ค่าความยากประจำขั้นเป็น -0.76 -0.11 0.25 1.32 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.71 ค่าความยากประจำขั้นเป็น 0.37 1.08 1.77 1.81 ตามลำดับ ผลการสร้างข้อมูลของข้อสอบศึกษาปรากฏตามตาราง 5

ตาราง 5 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษา กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ

ข้อที่	ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มอ้างอิง					ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มสนใจ ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ				
	a	b1	b2	b3	b4	a	b1	b2	b3	b4
5	1.45	-0.59	-0.15	1.33	1.79	1.45	-0.59	-0.15	1.33	1.79
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					0.00	0.00	0.00	0.00	0.00
10	2.90	-0.18	0.67	0.78	2.41	2.90	-0.18	0.67	0.78	2.41
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					0.00	0.00	0.00	0.00	0.00
15	3.04	-0.31	0.29	1.23	1.90	3.04	-0.31	0.29	1.23	1.90
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					0.00	0.00	0.00	0.00	0.00
20	1.59	-0.76	-0.11	0.25	1.32	1.59	-0.76	-0.11	0.25	1.32
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					0.00	0.00	0.00	0.00	0.00
25	2.71	0.37	1.08	1.77	1.81	2.71	0.37	1.08	1.77	1.81
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					0.00	0.00	0.00	0.00	0.00

2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.30 ค่าความยากประจำขั้นเป็น -0.34 0.10 1.58 2.04 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.75 ค่าความยากประจำขั้นเป็น 0.07 0.92 1.03 2.66 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.89 ค่าความยากประจำขั้นเป็น -0.06 0.54 1.48 2.15 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.44 ค่าความยากประจำขั้นเป็น -0.51 0.14 0.50 1.57 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.56 ค่าความยากประจำขั้นเป็น 0.62 1.33 2.02 2.06 ตามลำดับ ผลการสร้างข้อมูลของข้อสอบศึกษาปรากฏตามตาราง 6

ตาราง 6 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษาที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก

ข้อที่	ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มอ้างอิง					ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มสนใจ มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก				
	a	b1	b2	b3	b4	a	b1	b2	b3	b4
5	1.45	-0.59	-0.15	1.33	1.79	1.30	-0.34	0.10	1.58	2.04
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.15	0.25	0.25	0.25	0.25
10	2.90	-0.18	0.67	0.78	2.41	2.75	0.07	0.92	1.03	2.66
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.15	0.25	0.25	0.25	0.25
15	3.04	-0.31	0.29	1.23	1.90	2.89	-0.06	0.54	1.48	2.15
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.15	0.25	0.25	0.25	0.25
20	1.59	-0.76	-0.11	0.25	1.32	1.44	-0.51	0.14	0.50	1.57
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.15	0.25	0.25	0.25	0.25
25	2.71	0.37	1.08	1.77	1.81	2.56	0.62	1.33	2.02	2.06
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.15	0.25	0.25	0.25	0.25

3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง จากการจำลองข้อมูลได้ค่าพารามิเตอร์ข้อสอบศึกษาในกลุ่มสนใจ ดังนี้ ข้อ 5 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.05 ค่าความยากประจำขั้นเป็น -0.09 0.35 1.83 2.29 ตามลำดับ ข้อ 10 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.50 ค่าความยากประจำขั้นเป็น 0.32 1.17 1.28 2.91 ตามลำดับ ข้อ 15 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.64 ค่าความยากประจำขั้นเป็น 0.19 0.79 1.73 2.40 ตามลำดับ ข้อ 20 พารามิเตอร์ค่าความชันร่วมเท่ากับ 1.19 ค่าความยากประจำขั้นเป็น -0.26 0.39 0.75 1.82 ตามลำดับ และข้อ 25 พารามิเตอร์ค่าความชันร่วมเท่ากับ 2.31 ค่าความยากประจำขั้นเป็น 0.87 1.58 2.27 2.31 ตามลำดับ ผลการสร้างข้อมูลของข้อสอบศึกษาปรากฏตามตาราง 7

ตาราง 7 ข้อมูลพื้นฐานของพารามิเตอร์ข้อสอบศึกษาที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า ที่มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่

ข้อที่	ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มอ้างอิง					ค่าพารามิเตอร์ข้อสอบศึกษากลุ่มสนใจ มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่				
	a	b1	b2	b3	b4	a	b1	b2	b3	b4
5	1.45	-0.59	-0.15	1.33	1.79	1.05	-0.09	0.35	1.83	2.29
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.40	0.50	0.50	0.50	0.50
10	2.90	-0.18	0.67	0.78	2.41	2.50	0.32	1.17	1.28	2.91
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.40	0.50	0.50	0.50	0.50
15	3.04	-0.31	0.29	1.23	1.90	2.64	0.19	0.79	1.73	2.40
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.40	0.50	0.50	0.50	0.50
20	1.59	-0.76	-0.11	0.25	1.32	1.19	-0.26	0.39	0.75	1.82
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.40	0.50	0.50	0.50	0.50
25	2.71	0.37	1.08	1.77	1.81	2.31	0.87	1.58	2.27	2.31
	ผลต่างค่าพารามิเตอร์ข้อสอบศึกษา					-.40	0.50	0.50	0.50	0.50

1.4 การสร้างรายการคำตอบของผู้สอบ

หลักในการสร้างรายการคำตอบของผู้สอบนั้นจะต้องนำความสามารถของผู้สอบไปสัมพันธ์กับค่าพารามิเตอร์ของข้อสอบ ซึ่งความสามารถของผู้สอบและพารามิเตอร์ข้อสอบนั้นผู้วิจัยได้สร้างขึ้นแล้วในขั้น 1.1 และ 1.2 ตามลำดับ

เนื่องจากการศึกษาครั้งนี้เป็นการตรวจสอบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด แล้วนำผลการตรวจสอบความเท่าเทียมกันของการวัดในแต่ละวิธีมาเปรียบเทียบ และเพื่อเป็นการควบคุมความคลาดเคลื่อนในการสุ่มรายการคำตอบของกลุ่มตัวอย่าง ผู้วิจัยจึงจำลองข้อมูลซ้ำจำนวน 100 ครั้ง ทำให้ได้รายการคำตอบของผู้สอบในกลุ่มอ้างอิงและกลุ่มสนใจในแต่ละเงื่อนไข กลุ่มละ 100 ชุด

ตอนที่ 2 อัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

2.1 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด

ในการวิเคราะห์ข้อมูลตอนนี้ ผู้วิจัยนำรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า และที่เป็นการตรวจให้คะแนนแบบหลายค่า ที่ได้จากการจำลองข้อมูลที่มีการระบุขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ ในกลุ่มอ้างอิงและกลุ่มสนใจ มาวิเคราะห์ค่าทางสถิติเพื่อระบุถึงการตรวจสอบความเท่าเทียมกันของการวัดในข้อสอบศึกษาว่าตรวจสอบได้หรือไม่ ดังนี้

2.1.1 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

2.1.2 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบเป็นการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่า ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่า ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

ผลจากการศึกษา พบว่า

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.40 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 86.60 ถึง 96.40 ผลการวิเคราะห์ปรากฏดังตาราง 8

ตาราง 8 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	433	86.60
200	200	500	487	97.40	463	92.60
400	400	500	500	100.00	466	93.20
500	500	500	500	100.00	482	96.40

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 91.00 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 87.20 ถึง 97.00 ผลการวิเคราะห์ปรากฏดังตาราง 9

ตาราง 9 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	436	87.20
200	200	500	457	91.40	466	93.20
400	400	500	500	100.00	485	97.00
500	500	500	500	100.00	482	96.40

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.00 ถึง 99.60 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00 ผลการวิเคราะห์ปรากฏดังตาราง 10

ตาราง 10 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	200	500	495	99.00	480	96.00
200	400	500	498	99.60	500	100.00
400	800	500	497	99.40	500	100.00
500	1,000	500	496	99.20	500	100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อย 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 92.00 ถึง 100.00 ผลการวิเคราะห์ปรากฏดังตาราง 11

ตาราง 11 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	500	100.00
200	200	500	500	100.00	460	92.00
400	400	500	500	100.00	500	100.00
500	500	500	500	100.00	500	100.00

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อยละ 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.40 ถึง 97.80 ผลการวิเคราะห์ปรากฏดังตาราง 12

ตาราง 12 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	486	97.20
200	200	500	500	100.00	489	97.80
400	400	500	500	100.00	482	96.40
500	500	500	500	100.00	487	97.40

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อยละ 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.20 ถึง 98.80 ผลการวิเคราะห์ปรากฏดังตาราง 13

ตาราง 13 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	489	97.80
200	200	500	500	100.00	493	98.60
400	400	500	500	100.00	494	98.80
500	500	500	500	100.00	486	97.20

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.80 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00 ผลการวิเคราะห์ปรากฏดังตาราง 14

ตาราง 14 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	200	500	500	100.00	500	100.00
200	400	500	500	100.00	500	100.00
400	800	500	499	99.80	500	100.00
500	1,000	500	500	100.00	500	100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในตรวจสอบอย่างสมบูรณ์ร้อย 100.00 ผลการวิเคราะห์ปรากฏดังตาราง 15

ตาราง 15 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ และสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	491	98.20	500	100.00
200	200	500	424	84.80	500	100.00
400	400	500	499	99.80	500	100.00
500	500	500	500	100.00	500	100.00

2.2 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

ในการวิเคราะห์ข้อมูลตอนนี้ ผู้วิจัยนำรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า และที่เป็นการตรวจให้คะแนนแบบหลายค่า ที่ได้จากการจำลองข้อมูลที่มีการระบุไม่ให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบ นั่นคือ ทั้งกลุ่มอ้างอิงและกลุ่มสนใจจะมีค่าพารามิเตอร์ของข้อสอบศึกษาเท่ากัน มาวิเคราะห์ค่าทางสถิติเพื่อระบุถึงความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดในข้อสอบศึกษาว่ามีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบเท่าใด ดังนี้

2.2.1 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

2.2.2 วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบเป็นการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และ

พารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่า ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่ละวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

ผลจากการศึกษา พบว่า

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 8.00 โดยที่ขนาดกลุ่มตัวอย่าง 200:200 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ พบว่า มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 4.00 ถึง 9.00 โดยที่ขนาดกลุ่มตัวอย่าง 100:100 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ผลการวิเคราะห์ปรากฏดังตาราง 16

ตาราง 16 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งของการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด	จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด
100	100	500	0	0.00	45	9.00
200	200	500	40	8.00	32	6.40
400	400	500	0	0.00	20	4.00
500	500	500	0	0.00	20	4.00

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 12.40 โดยที่ขนาดกลุ่มตัวอย่าง 200:400 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้ พบว่า ไม่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างสมบูรณ์ร้อยละ 0.00 ผลการวิเคราะห์ปรากฏดังตาราง 17

ตาราง 17 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งของการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด	จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด
100	200	500	0	0.00	0	0.00
200	400	500	62	12.40	0	0.00
400	800	500	0	0.00	0	0.00
500	1,000	500	7	1.40	0	0.00

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:1

วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 0.60 โดยที่ขนาดกลุ่มตัวอย่าง 100:100 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 1.40 ถึง 5.40 โดยที่ขนาดกลุ่มตัวอย่าง 500:500 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ผลการวิเคราะห์ปรากฏดังตาราง 18

ตาราง 18 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด	จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด
100	100	500	3	0.60	24	4.80
200	200	500	1	0.20	7	1.40
400	400	500	0	0.00	13	2.60
500	500	500	2	0.40	17	5.40

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 0.40 โดยที่ขนาดกลุ่มตัวอย่าง 200:400 มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ พบว่า ไม่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างสมบูรณ์ร้อยละ 0.00 ผลการวิเคราะห์ปรากฏดังตาราง 19

ตาราง 19 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1			
			วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด	จำนวนครั้งที่ตรวจสอบผิด	ร้อยละของการตรวจสอบผิด
100	200	500	1	0.20	0	0.00
200	400	500	2	0.40	0	0.00
400	800	500	0	0.00	0	0.00
500	1,000	500	1	0.20	0	0.00

ตอนที่ 3 การเปรียบเทียบอัตราความถูกต้อง และอัตราความคลาดเคลื่อนประเภทที่ 1

3.1 การเปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด

การวิเคราะห์ข้อมูลตอนนี้ ผู้วิจัยนำอัตราความถูกต้องจากผลการตรวจสอบความเท่าเทียมกันของการวัด ที่ได้จากวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนแบบสองค่า และที่เป็นการตรวจให้คะแนนแบบหลายค่า ที่ได้จากการตรวจสอบความเท่าเทียมกันของการวัดในข้อสอบศึกษา ด้วยวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่ฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มาสรุปลงในรูปตารางไขว้ 2x2 ข้อมูลที่ได้เป็นค่าสังเกตบนวิธีการตรวจสอบความเท่าเทียมกันของการวัด ซึ่งอยู่ในมาตรวัดในนามบัญญัติ จากนั้นทำการแจงนับผลการตรวจสอบที่ขัดแย้งกันเพื่อทำการคำนวณค่าสถิติทดสอบ McNemar ผลการเปรียบเทียบ พบว่า

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

ภายใต้เงื่อนไขของขนาดการทำหน้าที่เบี่ยงเบนของข้อสอบเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบใหญ่ ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน รูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่ารูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความถูกต้องในการตรวจสอบไม่แตกต่าง

กัน ซึ่งทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่สูง (ร้อยละ 91.40 – 100.00) ผลการทดสอบปรากฏดังตาราง 20

ตาราง 20 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ		สัดส่วนขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)			
		100:100	200:200	400:400	500:500
		วิธีแบบ IRT			
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	4.81*	0.61	1.20	0.33
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	4.38*	0.09	0.23	0.33

* มีนัยสำคัญทางสถิติที่ระดับ .05

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

ที่สัดส่วนผู้สอบกลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 1:2 ของรายการคำตอบข้อสอบแบบการตรวจให้คะแนนแบบสองค่า ทั้งวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน ซึ่งทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่สูง (ร้อยละ 92.00 – 100.00) ผลการทดสอบปรากฏดังตาราง 21

ตาราง 21 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ		สัดส่วนขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
		100:200	200:400	400:800	500:1,000
		วิธีแบบ IRT			
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	0.23	0.00	0.01	0.02
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	0.00	1.67	0.00	0.00

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่ต่างกันกับวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ ซึ่งทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่สูง (ร้อยละ 96.40 – 100.00) ผลการทดสอบปรากฏดังตาราง 22

ตาราง 22 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดที่รายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ		สัดส่วนขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)			
		100:100	200:200	400:400	500:800
		วิธีแบบ IRT			
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	0.20	0.12	0.33	0.17
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	0.12	0.05	0.04	0.20

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

ภายใต้เงื่อนไขการทำหน้าที่เบี่ยงเบนขนาดใหญ่ ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน บนวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดต่ำกว่าการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนคู่อื่นๆ มีอัตราความถูกต้องในการตรวจสอบไม่แตกต่างกัน ซึ่งทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่สูง (ร้อยละ 98.20 – 100.00) ผลการทดสอบปรากฏดังตาราง 23

ตาราง 23 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ		สัดส่วนขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
		100:200	200:400	400:800	500:1,000
		วิธีแบบ IRT			
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	0.00	0.00	0.00	0.00
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	0.08	6.25*	0.00	0.00

* มีนัยสำคัญทางสถิติที่ระดับ .05

3.2 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

การวิเคราะห์ข้อมูลตอนนี้ ผู้วิจัยนำความคลาดเคลื่อนประเภทที่ 1 จากผลการตรวจสอบความเท่าเทียมกันของการวัด ที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า และการตรวจให้คะแนนแบบหลายค่า ด้วยวิธีการตรวจสอบความเท่าเทียมกันของการวัดที่อยู่บนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มาสรุปลงในรูปตารางไขว้ 2x2 ข้อมูลที่ได้เป็นค่าสังเกตบนวิธีการตรวจสอบความเท่าเทียมกันของการวัด ซึ่งอยู่ในมาตรวัดนามบัญญัติ จากนั้นทำการแจงนับผลการตรวจสอบที่ขัดแย้งกันเพื่อทำการคำนวณค่าสถิติทดสอบ McNemar ผลการทดสอบ พบว่า

กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และวิธีที่อยู่บนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน ซึ่งทั้ง 2 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่ต่ำมาก (ร้อยละ 0.00 – 9.00) ผลการวิเคราะห์ปรากฏดังตาราง 24

ตาราง 24 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด กรณีเป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)

ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)	100:200	200:400	400:800	500:1000
วิธีการตรวจสอบ	วิธีแบบ IRT			
วิธีแบบ CTT	2.12	0.07	0.41	0.41

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน ซึ่งทั้ง 2 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่ต่ำมาก (ร้อยละ 0.00 – 1.40) ผลการวิเคราะห์ปรากฏดังตาราง 25

ตาราง 25 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดของการที่เป็นการตรวจให้คะแนนแบบสองค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2

ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2	100:200	200:400	400:800	500:1000
วิธีการตรวจสอบ	วิธีแบบ IRT			
วิธีแบบ CTT	0.00	4.10*	0.00	0.05

* มีนัยสำคัญทางสถิติที่ระดับ .05

กรณีรายการคำตอบเป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน ซึ่งทั้ง 2 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่ต่ำมาก (ร้อยละ 0.20 – 5.40) ผลการวิเคราะห์ปรากฏดังตาราง 26

ตาราง 26 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเท่ากัน (1:1)

ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)	100:100	200:200	400:400	500:500
วิธีการตรวจสอบ	วิธีแบบ IRT			
วิธีแบบ CTT	0.45	0.04	0.17	0.23

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ เป็น 1:2

วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และทฤษฎีการตอบข้อสอบโดยการทดสอบอัตราส่วนความเป็นไปได้ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน ซึ่งทั้ง 2 วิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ระดับที่ต่ำมาก (ร้อยละ 0.00 – 0.40) ผลการวิเคราะห์ปรากฏดังตาราง 27

ตาราง 27 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนแบบหลายค่า ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ ที่สัดส่วนกลุ่มตัวอย่างเป็น 1:2

ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2	100:200	200:400	400:800	500:1,000
วิธีการตรวจสอบ	วิธีแบบ IRT			
วิธีแบบ CTT	0.00	0.00	0.00	0.00

ตอนที่ 4 สรุปผลการศึกษาในภาพรวม

เพื่อเชื่อมโยงให้เห็นถึงความเกี่ยวพันกันของปัจจัยที่ศึกษา โดยการนำเสนอแยกตามประเด็นที่ศึกษา ผู้วิจัยได้นำผลการศึกษามานำเสนอในรูปของตารางสรุปรวมทั้งหมดอีกครั้ง ดังนี้

1. การศึกษาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ ผลปรากฏดังตาราง 28-31

2. การเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ ผลปรากฏดังตาราง 32-35

ตาราง 28 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบสองค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด							
			การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก				การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	433	86.60	500	100.00	436	87.20
200	200	500	487	97.40	463	92.60	457	91.40	466	93.20
400	400	500	500	100.00	466	93.20	500	100.00	485	97.00
500	500	500	500	100.00	482	96.40	500	100.00	482	96.40

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.40 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 86.60 ถึง 96.40

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 91.00 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 87.20 ถึง 97.00

ตาราง 29 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบสองค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด							
			การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก				การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	200	500	495	99.00	480	96.00	500	100.00	500	100.00
200	400	500	498	99.60	500	100.00	500	100.00	460	92.00
400	800	500	497	99.40	500	100.00	500	100.00	500	100.00
500	1,000	500	496	99.20	500	100.00	500	100.00	500	100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.00 ถึง 99.60 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อย 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 92.00 ถึง 100.00

ตาราง 30 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบหลายค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด							
			การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก				การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	500	100.00	486	97.20	500	100.00	489	97.80
200	200	500	500	100.00	489	97.80	500	100.00	493	98.60
400	400	500	500	100.00	482	96.40	500	100.00	494	98.80
500	500	500	500	100.00	487	97.40	500	100.00	486	97.20

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อยละ 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.40 ถึง 97.80

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อยละ 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.20 ถึง 98.80

ตาราง 31 อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดรายการคำตอบข้อสอบที่เป็นการตรวจให้คะแนนแบบหลายค่า บนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด							
			การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก				การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	200	500	500	100.00	500	100.00	491	98.20	500	100.00
200	400	500	500	100.00	500	100.00	424	84.80	500	100.00
400	800	500	499	99.80	500	100.00	499	99.80	500	100.00
500	1,000	500	500	100.00	500	100.00	500	100.00	500	100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.80 ถึง 100.00 ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00

ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบความเท่าเทียมกันของการวัดทั้งบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อย 100.00

ตาราง 32 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1)

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด							
			รายการคำตอบข้อสอบแบบการตรวจให้คะแนนแบบสองค่า				รายการคำตอบข้อสอบแบบการตรวจให้คะแนนแบบหลายค่า			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	100	500	0	0.00	45	9.00	3	0.60	24	4.80
200	200	500	40	8.00	32	6.40	1	0.20	7	1.40
400	400	500	0	0.00	20	4.00	0	0.00	13	2.60
500	500	500	0	0.00	20	4.00	2	0.40	17	3.40

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 0.00 ถึง 8.00 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:200 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่ามีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 4.00 ถึง 9.00 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 0.00 ถึง 0.60 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่ามีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 1.40 ถึง 5.40 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 500:500 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด

ตาราง 33 อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ กรณีที่สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2

จำนวนผู้สอบ		จำนวนครั้งในการตรวจสอบ	อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด							
			รายการคำตอบข้อสอบแบบการตรวจให้คะแนนแบบสองค่า				รายการคำตอบข้อสอบแบบการตรวจให้คะแนนแบบหลายค่า			
			วิธีแบบ CTT		วิธีแบบ IRT		วิธีแบบ CTT		วิธีแบบ IRT	
กลุ่มอ้างอิง	กลุ่มสนใจ		จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้	จำนวนครั้งที่ตรวจสอบได้	ร้อยละในการตรวจสอบได้
100	200	500	0	0.00	0	0.00	1	0.20	0	0.00
200	400	500	62	12.40	0	0.00	2	0.40	0	0.00
400	800	500	0	0.00	0	0.00	0	0.00	0	0.00
500	1,000	500	7	1.40	0	0.00	1	0.20	0	0.00

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 0.00 ถึง 12.40 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า มีอัตราความคลาดเคลื่อนประเภทที่ 1 สมบูรณ์ร้อยละ 0.00

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในช่วงร้อยละ 0.00 ถึง 0.40 โดยที่ขนาดกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 มากที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า ไม่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างสมบูรณ์ร้อยละ 0.00

ตาราง 34 เปรียบเทียบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ

วิธีการตรวจให้คะแนนแบบสองค่า		ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)				ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
		100:100	200:200	400:400	500:500	100:200	200:400	400:800	500:1,000
		วิธีแบบ IRT							
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	4.81*	0.61	1.20	0.33	0.23	0.00	0.01	0.02
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	4.38*	0.09	0.23	0.33	0.00	1.67	0.00	0.00
วิธีการตรวจให้คะแนนแบบหลายค่า		ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)				ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
		100:100	200:200	400:400	500:500	100:200	200:400	400:800	500:1,000
		วิธีแบบ IRT							
การทำหน้าที่เบี่ยงเบนขนาดเล็ก	วิธีแบบ	0.20	0.12	0.33	0.17	0.00	0.00	0.00	0.00
การทำหน้าที่เบี่ยงเบนขนาดใหญ่	CTT	0.12	0.05	0.04	0.20	0.08	6.25*	0.00	0.00

* มีนัยสำคัญทางสถิติที่ระดับ .05

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ภายใต้เงื่อนไขของขนาดการทำหน้าที่เบี่ยงเบนของข้อสอบเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบใหญ่ ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน รูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่ารูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และกรณีรายการคำตอบของผู้สอบที่มีการตรวจให้คะแนนหลายค่า ภายใต้เงื่อนไขของขนาดการทำหน้าที่เบี่ยงเบนของข้อสอบใหญ่ ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน รูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดต่ำกว่ารูปแบบการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

ตาราง 35 เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด ระหว่างวิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ

การตรวจให้คะแนนแบบสองค่า	ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)				ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
	100:100	200:200	400:400	500:500	100:200	200:400	400:800	500:1,000
	วิธีแบบ IRT							
วิธีแบบ CTT	2.12	0.07	0.41	0.41	0.00	4.10*	0.00	0.05
การตรวจให้คะแนนแบบหลายค่า	ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเท่ากัน (1:1)				ขนาดกลุ่มอ้างอิง:กลุ่มสนใจเป็น 1:2			
	100:100	200:200	400:400	500:500	100:200	200:400	400:800	500:1,000
	วิธีแบบ IRT							
วิธีแบบ CTT	0.45	0.04	0.17	0.23	0.00	0.00	0.00	0.00

* มีนัยสำคัญทางสถิติที่ระดับ .05

กรณีรายการคำตอบมีการตรวจให้คะแนนแบบสองค่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

ในศึกษาครั้งนี้ เป็นการศึกษาผลการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีที่ต่างกัน 2 ทฤษฎี คือ ทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบภายใต้เงื่อนไขวิธีการตรวจให้คะแนนรายการคำตอบ ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบและขนาดกลุ่มตัวอย่างที่แตกต่างกัน **มีความมุ่งหมายเฉพาะ** เพื่อศึกษาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ และเพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ **ทำการศึกษากับข้อมูลจำลอง** โดยทำการจำลองข้อมูลภายใต้ทฤษฎีการตอบข้อสอบที่วัดมิติเดียวด้วยโมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two-parameter Logistic Model) สำหรับรายการคำตอบที่เป็นการตรวจให้คะแนนแบบสองค่า และโมเดลเกรดเรสพอนส์ (Graded - Response Model) สำหรับรายการคำตอบที่เป็นการตรวจให้คะแนนแบบหลายค่า ฉบับละ 25 ข้อ ในจำนวน 25 ข้อ จะมีการกำหนดให้ข้อสอบข้อที่ 5, 10, 15, 20 และ 25 เป็นข้อสอบศึกษาที่มีขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ 3 ขนาด คือ ไม่มีการทำหน้าที่เบี่ยงเบน การทำหน้าที่เบี่ยงเบนขนาดเล็ก และการทำหน้าที่เบี่ยงเบนขนาดใหญ่ ขนาดกลุ่มตัวอย่าง 4 ขนาด แบ่งตามสัดส่วนของกลุ่มอ้างอิงต่อกลุ่มสนใจ 2 สัดส่วน คือ สัดส่วนเท่ากัน (1:1) ได้แก่ขนาด 100:100, 200:200, 400:400 และ 500:500 คน และสัดส่วนกลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 1:2 ได้แก่ขนาด 100:200, 200:400, 400:800 และ 500:1,000 คน สำหรับใช้ในการศึกษาอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ และเพื่อเป็นการควบคุมความคลาดเคลื่อนในการสุ่มกลุ่มตัวอย่าง ผู้วิจัยจึงทำการจำลองข้อมูลซ้ำจำนวน 100 ครั้ง ในแต่ละเงื่อนไขการศึกษา **การวิเคราะห์ข้อมูล** ผู้วิจัยนำข้อมูลจำลองตามเงื่อนไขที่กำหนดไว้แต่ละชุดมาทำการวิเคราะห์ใน 2 วิธี คือ วิธีบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน และทฤษฎีการตอบข้อสอบโดยทดสอบอัตราส่วนความเป็นไปได้ ดังนี้

การตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์

ขั้นตอนการตรวจสอบอัตราความถูกต้อง

ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ เป็นการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ ซึ่งมีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

ขั้นตอนในการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1

ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งจะกระทำใน 2 ตัวแปรแฝง คือ ความสามารถและการเป็นสมาชิกของกลุ่ม โดยใช้โปรแกรม LISREL ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าโมเดลกระชับ (Compact Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปรความสามารถ ส่วนตัวแปรสมาชิกในกลุ่มเป็นดัชนีบ่งชี้ที่จะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าโมเดลขยาย (Augmented Model) ซึ่งเป็นโมเดลที่รวมข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ ข้อสอบทุกข้อจะให้ค่าน้ำหนักของตัวแปร

ความสามารถ แต่ตัวแปรสมาชิกในกลุ่มและข้อสอบศึกษาที่นำเข้าไปจะให้ค่าน้ำหนักของการเป็นสมาชิกในกลุ่ม ผลการวิเคราะห์จะได้ค่า C(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ C(2) - C(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 1

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤต นั่นคือ โมเดลขยายมีความเหมาะเจาะพอดีกว่าโมเดลกระชับอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

ขั้นตอนการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีตอบข้อสอบ

ขั้นตอนการตรวจสอบอัตราความถูกต้อง

ขั้นตอนการตรวจสอบอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่จะวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดจริง ซึ่งเป็นการตรวจสอบที่ถูกต้อง

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1

ขั้นตอนการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดบนฐานทฤษฎีการตอบข้อสอบ โดยการทดสอบอัตราส่วนความเป็นไปได้ โดยโมเดลกระชับ (Compact Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวมกับข้อสอบศึกษา (ข้อสอบร่วมทุกข้อ (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ รวมเป็น 21 ข้อ) ส่วนโมเดลขยาย (Augmented Model) จะเป็นการประมาณค่าพารามิเตอร์ร่วมกันของทั้งกลุ่มอ้างอิงและกลุ่มสนใจ เช่น ค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมเท่านั้น (ข้อสอบร่วมทุกข้อ 20 ข้อ) โดยใช้โปรแกรม MULTILOG ในการวิเคราะห์ มีขั้นตอนดังนี้

ขั้นที่ 1 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกของข้อสอบร่วมรวม (20 ข้อ) กับข้อสอบศึกษา 1 ข้อ โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ผลการวิเคราะห์จะได้ค่า G(1) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเช่นเดียวกับขั้นที่ 1 แต่ละวิเคราะห์ข้อสอบร่วมเท่านั้น (20 ข้อ) โดยระบบจะระบุให้ทั้งกลุ่มอ้างอิงและกลุ่มสนใจมีค่าพารามิเตอร์ความยาก และพารามิเตอร์อำนาจจำแนกเท่ากัน ส่วนข้อสอบศึกษาจะนำมาแยกคำนวณระหว่างกลุ่มอ้างอิงและกลุ่มสนใจ ผลการวิเคราะห์จะได้ค่า G(2) ซึ่งเป็นค่า Chi-square goodness-of-fit

ขั้นที่ 3 คำนวณความแตกต่างของ G(2) - G(1) ค่าประมาณของผลต่างที่ได้เป็นการกระจายของไค-สแควร์ ที่ชั้นองศาความเป็นอิสระเท่ากับ 2

ขั้นที่ 4 หากความแตกต่างมากกว่าค่าวิกฤตอย่างมีนัยสำคัญ แสดงว่า มีการระบุว่าข้อสอบศึกษาที่นำเข้าในโมเดลนั้นไม่มีความเท่าเทียมกันของการวัดทั้งที่ข้อสอบศึกษานั้นมีความเท่าเทียมกันของการวัด ซึ่งเป็นการตรวจสอบที่ผิดพลาด

ขั้นที่ 5 ทำซ้ำในขั้นที่ 1-4 จนกว่าจะนำข้อสอบศึกษาเข้าตรวจสอบความเท่าเทียมกันของการวัดครบทุกข้อ

การเปรียบเทียบผลการตรวจสอบความเท่าเทียมกันของการวัดจากวิธีการตรวจสอบที่ต่างกัน ผู้วิจัยเลือกใช้สถิติ McNemar (McNemar, 1962) โดยสรุปข้อมูลในรูปตารางไขว้ 2x2 ข้อมูลเป็นค่าสังเกตบนสองทางที่อิสระ สเกลการวัดสำหรับ X และ Y อยู่ในมาตรวัดในนาม

บัญญัติ ในการทดสอบที่ระดับนัยสำคัญทางสถิติที่ระดับ .05 หากผลการทดสอบมีนัยสำคัญแสดงว่าวิธีที่ใช้ในการตรวจสอบความเท่าเทียมกันของการวัดทั้ง 2 วิธีแตกต่างกัน

สรุปผลการวิจัย

1. อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.40 ถึง 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 86.60 ถึง 96.40 ส่วนการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 91.00 ถึง 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 87.20 ถึง 97.00 และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ภายใต้เงื่อนไขของขนาดการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีอัตราความถูกต้องของการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.00 ถึง 99.60 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00 ส่วนการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องร้อยละ 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 92.00 ถึง 100.00 และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ผลการตรวจสอบความเท่าเทียมกันของการวัดทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎี

การทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบได้อย่างสมบูรณ์ร้อยละ 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.40 ถึง 97.80 ส่วนการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องร้อยละ 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 97.20 ถึง 98.80 และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่ต่างกัน

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 99.80 ถึง 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในช่วงร้อยละ 96.00 ถึง 100.00 ส่วนการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องอย่างสมบูรณ์ร้อยละ 100.00 วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอย่างสมบูรณ์ร้อยละ 100.00 และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ภายใต้เงื่อนไขการทำหน้าที่เบี่ยงเบนขนาดใหญ่ ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดต่ำกว่าวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่นๆ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 8.00 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:200 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 4.00 ถึง 9.00 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

พบว่า วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 12.40 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า ไม่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างสมบูรณ์ร้อยละ 0.00 และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า

กรณีที่ 1 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 0.60 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 1.40 ถึง 5.40 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 500:500 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีที่ 2 สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงร้อยละ 0.00 ถึง 0.40 โดยที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงที่สุด ส่วนวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ พบว่า ไม่มีอัตราความคลาดเคลื่อนประเภทที่ 1 อย่างสมบูรณ์ร้อยละ 0.00 และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์

และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีเป้าหมายเพื่อตอบคำถามในสองประเด็นหลัก คือ (1) อัตราความถูกต้องของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ และ (2) อัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์และทฤษฎีการตอบข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 3 ปัจจัย ดังนี้ **ปัจจัยที่ 1** วิธีในการตรวจให้คะแนนรายการคำตอบ มาตรวัดโดยทั่วไปจะมีวิธีการตรวจให้คะแนน 2 แบบ คือ การตรวจให้คะแนนแบบสองค่า (0, 1) และการตรวจให้คะแนนแบบหลายค่า (4, 3, 2, 1, 0) **ปัจจัยที่ 2** ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ ในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบออกเป็น 3 ระดับ คือ 1) ไม่มีการทำหน้าที่เบี่ยงเบนของข้อสอบ เป็นการระบุให้ค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากของข้อสอบศึกษาทุกข้อในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน เพื่อทำการศึกษาความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดจากวิธีการตรวจสอบทั้ง 2 วิธี 2) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก เป็นการลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .15 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .25 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง และ 3) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ โดยลดค่าอำนาจจำแนกของข้อสอบศึกษาในกลุ่มสนใจลง .40 และเพิ่มค่าพารามิเตอร์ความยากของข้อสอบศึกษาขึ้น .50 เมื่อเทียบกับค่าพารามิเตอร์ของข้อสอบศึกษาในกลุ่มอ้างอิง ซึ่งการที่กำหนดให้มีการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็กและการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ เพื่อทำการศึกษาอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด จากวิธีการตรวจสอบทั้ง 2 วิธี และ **ปัจจัยที่ 3** ขนาดกลุ่มตัวอย่าง โดยในการศึกษาครั้งนี้ ผู้วิจัยแบ่งตามสัดส่วนของกลุ่มสนใจกับกลุ่มอ้างอิง 2 สัดส่วน ได้แก่ สัดส่วน 1:1 มี 4 ขนาด คือ 100:100, 200:200 400:400 และ 500:500 คน และสัดส่วน 1:2 มี 4 ขนาดเช่นกัน คือ 100:200, 200:400 400:800 และ 500:1000 คน แตกต่างกันหรือไม่ จากผลการวิจัยที่สรุปมาข้างต้นมีประเด็นสำคัญที่นำมาอภิปรายผล ดังนี้

1. อัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัด

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในระดับที่สูง และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ภายใต้เงื่อนไขของขนาดการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ

ขนาดใหญ่ ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องของการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในระดับที่สูง และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ผลการตรวจสอบความเท่าเทียมกันของการวัดทั้ง 2 วิธี มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความถูกต้องในการตรวจสอบอยู่ในระดับสูง และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ทั้ง 2 วิธีมีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่ต่างกัน สำหรับสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 การทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีอัตราความถูกต้องในการตรวจสอบอยู่ในระดับสูง และเมื่อนำมาเปรียบเทียบระหว่างวิธีการตรวจสอบของทั้ง 2 วิธี พบว่า ภายใต้เงื่อนไขการทำหน้าที่เบี่ยงเบนขนาดใหญ่ ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดต่ำกว่าวิธีการตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่นๆ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

ทั้งนี้อาจเนื่องมาจากในการศึกษาครั้งนี้เป็นการศึกษากับข้อมูลจำลอง ทำให้ทราบถึงแหล่งหรือสาเหตุของการไม่เท่าเทียมกันของการวัดข้อสอบศึกษา โดยได้กำหนดเงื่อนไขให้ผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจทุกสัดส่วนและทุกขนาดมีความสามารถเท่ากัน แล้วให้ปัจจัยขนาดของการทำหน้าที่ต่างกันของข้อสอบศึกษาแปรเปลี่ยน ซึ่งจะเห็นได้ว่าโค้งคุณลักษณะของข้อสอบศึกษาแต่ละข้อทั้งในรูปแบบการตรวจให้คะแนนรายการคำตอบเป็นแบบสองค่าและการตรวจให้คะแนนรายการคำตอบเป็นแบบหลายค่า ในกลุ่มอ้างอิงและกลุ่มสนใจตัดกันและมีพื้นที่ใต้โค้งที่เกิดจากโค้งคุณลักษณะของข้อสอบศึกษาแต่ละข้อในกลุ่มอ้างอิงและกลุ่มสนใจตัดกันสูง ข้อสอบศึกษา

ที่กำหนดขึ้นจึงมีแนวโน้มของการทำหน้าที่เบี่ยงเบนของข้อสอบ ส่งผลให้การตรวจสอบมีอัตราความถูกต้องอยู่ในระดับที่สูง (ร้อยละ 86.60 ถึง 100.00) ซึ่งสอดคล้องกับการศึกษาของโคเฮน คิม และ เบคเกอร์ (Cohen; Kim; & Baker. 1993) ที่ศึกษาระดับการทำหน้าที่เบี่ยงเบนของข้อสอบ .5 และ 1.0 ส่วนใหญ่จะตรวจสอบได้ และออร์ต (Oort. 1998) ได้ศึกษาระดับของการทำหน้าที่เบี่ยงเบนของข้อสอบ คือ ระดับต่ำ (.2) ระดับปานกลาง (.5) และสูง (.8) พบว่า ข้อสอบที่มีระดับการทำหน้าที่เบี่ยงเบนในระดับปานกลาง และระดับสูงส่วนใหญ่จะตรวจสอบได้ แต่ในระดับการทำหน้าที่เบี่ยงเบนที่ต่ำจะไม่สามารถตรวจสอบได้ แม้ว่าจะมีบ้างในบางกรณีที่ทำให้ผลการตรวจสอบแตกต่างกัน ได้แก่ กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 100:100 คน ภายใต้เงื่อนไขของขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดเล็ก และขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบขนาดใหญ่ วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หลักฐานนี้แสดงถึง วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีแนวโน้มว่าจะรองรับรายการคำตอบของผู้สอบที่มีการตรวจให้คะแนนแบบสองค่าที่สัดส่วนของกลุ่มอ้างอิงต่อกลุ่มสนใจเท่ากัน (1:1) ส่วนกรณีรายการคำตอบของผู้สอบที่มีการตรวจให้คะแนนแบบหลายค่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 ภายใต้เงื่อนไขการทำหน้าที่เบี่ยงเบนขนาดใหญ่ การตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ มีอัตราความถูกต้องในการตรวจสอบความเท่าเทียมกันของการวัดต่ำกว่าวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 วิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบมีแนวโน้มว่าจะรองรับรายการคำตอบของผู้สอบที่มีการตรวจให้คะแนนแบบหลายค่าที่สัดส่วนของกลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 1:2 ซึ่งสอดคล้องกับการศึกษาของมีด และ ลอเทนชลาเจอ (Meade; & Lautenschlager. 2004) ที่ได้ทำการเปรียบเทียบทฤษฎีการตอบข้อสอบและการวิเคราะห์องค์ประกอบเชิงยืนยัน ในการแสดงความเท่าเทียมกันของการวัด โดยการจำลองข้อมูลจากข้อคำถามแบบ Likert 5 ระดับ กับขนาดกลุ่มตัวอย่าง 2 กลุ่มที่มีการแจกแจงปกติ 3 ขนาด คือ 150, 500 และ 1,000 คน และใช้ค่าพารามิเตอร์ความยาก และค่าพารามิเตอร์อำนาจจำแนกในการประเมินความเหมาะสม ความเหมือน และความแตกต่างระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยันและวิธีทฤษฎีการตอบข้อสอบ พบว่า กรณีที่ใช้ค่าพารามิเตอร์ความยากในการตรวจสอบความเท่าเทียมกันของการวัด วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันจะตรวจสอบได้ดีกับขนาดกลุ่มตัวอย่าง 500 และ 1,000 คน ส่วนขนาดกลุ่มตัวอย่าง 150 คน ตรวจสอบได้ต่ำ และเมื่อใช้อัตราส่วนความเป็นไปได้ตรวจสอบความเท่าเทียมกันของการวัด พบว่า ขนาดกลุ่มตัวอย่าง 150 คน จะมีความถูกต้องในการตรวจสอบต่ำ และสอดคล้องกับการศึกษาของรัชนีวรรณ วานิชย์ถนอม (Wanichtanom, R. 2001) ที่ได้เปรียบเทียบอัตราการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในกลุ่มอ้างอิงและกลุ่มสนใจด้วยวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ และ

วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน 3 วิธี ได้แก่ วิธีเปรียบเทียบโมเดล (MC) วิธีปรับปรุงดัชนี (MI) และวิธีปรับปรุงดัชนีด้วยสัดส่วนกลุ่มตัวอย่าง (MI-Divided) ทำการศึกษากับข้อมูลจำลองการตอบข้อคำถามโมเดลโลจิสติก 2 พารามิเตอร์ โดยกำหนดพื้นที่ระหว่างเส้นโค้งลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มสนใจเป็นค่าคงที่ไว้ที่ .5 ผลการศึกษาพบว่า วิธีการตรวจสอบทั้งหมดสามารถตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบได้ดีในข้อสอบที่มีการกำหนดให้ทำหน้าที่เบี่ยงเบนบนพารามิเตอร์ความยาก โดยวิธีทฤษฎีการตอบข้อสอบ (IRT) และ MI-Divided มีอัตราการตรวจสอบไม่แตกต่างกัน

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในระดับต่ำ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่แตกต่างกัน สำหรับสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในระดับต่ำ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนเงื่อนไขอื่น ๆ มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบหลายค่า สัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน (1:1) วิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในระดับต่ำ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า ทั้ง 2 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน สำหรับสัดส่วนของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจเป็น 1:2 วิธีการ

ตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีการตรวจสอบบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดอยู่ในระดับต่ำ และเมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด พบว่า วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ มีอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดไม่แตกต่างกัน

ทั้งนี้อาจเนื่องมาจากการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 เป็นการตรวจสอบข้อมูลจำลองที่กำหนดเงื่อนไขให้ผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจทุกสัดส่วนและทุกขนาดมีความสามารถเท่ากัน แล้วไม่ให้มีการทำหน้าที่ต่างกันของข้อสอบศึกษา ซึ่งจะพบว่าโค้งคุณลักษณะของข้อสอบศึกษา (ICC) ที่กำหนดโดยข้อสอบศึกษาแต่ละข้อ ได้แก่ ข้อสอบข้อที่ 5, 10, 15, 20 และข้อที่ 25 ทั้งในรูปแบบการตรวจให้คะแนนรายการคำตอบเป็นแบบสองค่าและการตรวจให้คะแนนรายการคำตอบเป็นแบบหลายค่ามีการกำหนดให้ค่าพารามิเตอร์ข้อสอบในกลุ่มอ้างอิงและกลุ่มสนใจเท่ากันทุกประการจึงทำให้โค้งแสดงคุณลักษณะข้อสอบ (ICC) เป็นเส้นเดียวกันทับกันสนิท ไม่เกิดพื้นที่ความแตกต่างโค้งแสดงคุณลักษณะข้อสอบ (ICC) จึงส่งผลให้ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบอยู่ในระดับต่ำมาก (ร้อยละ 0.00 ถึง 12.40) ซึ่งสอดคล้องกับแนวคิดของ Hui; Drasgow; & Komocar (1982) และ Lord (1980) ที่กล่าวว่า ผลการประมาณโค้งแสดงคุณลักษณะข้อสอบ (ICC) ในข้อสอบสามารถเปรียบเทียบกันได้ด้วยสถิติข้อใดที่มีโค้งแสดงคุณลักษณะข้อสอบ (ICC) แตกต่างกันแสดงว่าขาดคุณสมบัติความเท่าเทียมกันของการวัด และข้อที่โค้งแสดงคุณลักษณะข้อสอบ (ICC) ที่ประมาณได้ไม่แตกต่างกันแสดงว่ามีคุณสมบัติการเท่าเทียมกันของการวัด แม้ว่าจะมีบ้างในบางกรณีที่ทำให้ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบแตกต่างกัน ได้แก่ กรณีรายการคำตอบของผู้สอบที่เป็นการตรวจให้คะแนนแบบสองค่า ที่ขนาดของกลุ่มอ้างอิงต่อกลุ่มสนใจ 200:400 คน วิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัดสูงกว่าวิธีตรวจสอบความเท่าเทียมกันของการวัดบนพื้นฐานทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หลักฐานนี้แสดงถึงวิธีการตรวจสอบบนพื้นฐานทฤษฎีการทดสอบแบบคะแนนจริงสัมพันธ์มีแนวโน้มว่าจะมีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบความเท่าเทียมกันของการวัด สำหรับรายการคำตอบที่เป็นการตรวจให้คะแนนแบบสองค่าที่สัดส่วนของกลุ่มอ้างอิงต่อกลุ่มสนใจเป็น 1:2

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

ในงานวิจัยข้ามวัฒนธรรม หรือข้ามกลุ่มส่วนใหญ่เริ่มด้วยการพัฒนาเครื่องมือขึ้นในวัฒนธรรมหนึ่ง หรือกลุ่มหนึ่งแล้วจึงปรับใช้ในอีกวัฒนธรรมหนึ่งหรืออีกกลุ่มหนึ่ง ผลสรุปจากการวิจัยอาจทำให้เข้าใจคนในวัฒนธรรมอื่นหรือกลุ่มอื่นผิดได้ ในการตรวจสอบความเท่าเทียมกันของการวัดต่างวัฒนธรรมหรือต่างกลุ่มก่อนที่จะศึกษาย่อมก่อให้เกิดโครงสร้างที่เปรียบเทียบกันได้ (Construct comparability) และผลการเปรียบเทียบข้ามวัฒนธรรมหรือข้ามกลุ่มย่อมมีความหมายและมีความถูกต้อง ดังนั้น ก่อนที่จะมีการเปรียบเทียบข้ามวัฒนธรรมหรือข้ามกลุ่ม ผู้วิจัยควรต้องตรวจสอบความเท่าเทียมกัน (Equivalence) ของเครื่องมือระหว่างวัฒนธรรมหรือระหว่างกลุ่มก่อน

จากผลการศึกษาครั้งนี้ พบว่า อัตราการตรวจสอบความเท่าเทียมกันของการวัดที่เป็นการตรวจให้คะแนนรายการคำตอบแบบสองค่า และการตรวจให้คะแนนรายการคำตอบหลายค่า ขนาดของการทำหน้าที่เบี่ยงเบนของข้อสอบ และขนาดกลุ่มตัวอย่าง โดยภาพรวมทุกเงื่อนไขทั้งวิธีบนพื้นฐานทฤษฎีคะแนนจริงสัมพันธ์ และวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีประสิทธิภาพในการตรวจสอบใกล้เคียงกัน เพียงแต่หากกลุ่มอ้างอิงและกลุ่มสนใจเท่ากัน และจำนวนผู้สอบมีไม่มาก ที่รายการคำตอบเป็นการตรวจให้คะแนนแบบสองค่า วิธีบนพื้นฐานทฤษฎีคะแนนจริงสัมพันธ์มีแน้วโน้มที่จะรองรับสูงกว่าวิธีบนพื้นฐานทฤษฎีการตอบข้อสอบ กรณีที่กลุ่มอ้างอิงและกลุ่มสนใจไม่เท่ากัน และรายการคำตอบที่เป็นการตรวจให้คะแนนแบบสองค่า วิธีบนพื้นฐานทฤษฎีการตอบข้อสอบมีแน้วโน้มที่จะมีความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบต่ำกว่าวิธีบนพื้นฐานทฤษฎีคะแนนจริงสัมพันธ์ ส่วนรายการคำตอบที่เป็นการตรวจให้คะแนนแบบหลายค่า วิธีการตรวจสอบความเท่าเทียมกันของการวัดทั้ง 2 วิธี มีประสิทธิภาพในการตรวจสอบไม่แตกต่างกัน ดังนั้น หากผู้ที่สนใจจะนำวิธีการตรวจสอบความเท่าเทียมกันของการวัดทั้ง 2 วิธีไปใช้ จึงต้องพิจารณาถึงธรรมชาติของข้อมูลที่จะทำการตรวจสอบก่อนการเลือกใช้วิธีการตรวจสอบ

ในงานวิจัย หากผู้วิจัยนำเครื่องมือที่สร้างขึ้นในวัฒนธรรมหนึ่งหรือกลุ่มประชากรหนึ่งแล้วจึงแปลไปเป็นภาษาอื่นเพื่อใช้ในอีกวัฒนธรรมหนึ่งหรืออีกกลุ่มประชากรหนึ่ง ผลสรุปจากการวิจัยอาจทำให้เข้าใจคนในวัฒนธรรมหรือกลุ่มประชากรอื่นผิดได้ และจะทำให้ไม่สามารถค้นพบความจริงในสังคมนั้นได้ ดังนั้นผู้ที่จะทำวิจัยควรมีการตรวจสอบคุณภาพการแปลเพื่อกระจายอิทธิพลของเครื่องมือวัดออกมาจากฉบับที่เป็นภาษาเดิม เพื่อให้ได้ฉบับในภาษาที่สองที่ราบรื่นและดูเป็นธรรมชาติ จะทำให้เครื่องมือไม่จำกัดอยู่ที่วัฒนธรรมเดียวหรือภาษาเดียว แต่ละกลุ่มภาษาที่ร่วมศึกษามีส่วนสนับสนุนต่อเครื่องมือวัดที่ใช้รวบรวมข้อมูลในงานวิจัย เครื่องมือวัดฉบับที่ใช้ในแต่ละภาษาจึงสะท้อนแนวคิดเดียวกันและมีความหมายเหมือนกันต่อผู้สอบ เครื่องมือวัดที่ได้มาจากการกระจายอิทธิพล ทำให้มีข้อคำถามร่วมกันระหว่างวัฒนธรรมหรือกลุ่มประชากร และข้อคำถามเฉพาะวัฒนธรรมหรือเฉพาะกลุ่มประชากร ผลการวิจัยสามารถสรุปได้สองส่วน คือ ส่วนที่

เป็นหลักฐานเฉพาะวัฒนธรรมหรือเฉพาะกลุ่มประชากร และส่วนที่เป็นหลักฐานข้ามวัฒนธรรมหรือข้ามกลุ่มประชากร

ข้อเสนอแนะในการวิจัยครั้งต่อไป

จากข้อค้นพบในการวิจัยครั้งนี้ วิธีการในการตรวจสอบภายใต้เงื่อนไขต่าง ๆ ยังมีบางประเด็นที่น่าสนใจที่คาดว่าจะส่งผลกระทบต่อการตรวจสอบความเท่าเทียมกันของการวัดภายใต้การจำลองข้อมูล ดังนี้

1. สัดส่วนของข้อสอบศึกษากับข้อสอบร่วม เนื่องจากในการวิจัยครั้งนี้ กำหนดให้ข้อสอบศึกษาและข้อสอบร่วมมีจำนวนข้อคงที่ หากปล่อยให้ข้อสอบศึกษามีสัดส่วนต่อข้อสอบร่วมเปลี่ยนแปลงไป จะส่งผลกระทบต่อการตรวจสอบความเท่าเทียมกันของการวัดหรือไม่

2. ควรทำการศึกษาลักษณะทำนองเดียวกับการศึกษาในครั้งนี้ โดยใช้ข้อมูลจริง อาจใช้ตัวแปรเพศ เชื้อชาติ ศาสนา ภูมิลำเนา อายุ หรือประสบการณ์ เป็นเกณฑ์ในการแบ่งกลุ่มผู้สอบ สำหรับแบบทดสอบที่ใช้ในระดับประเทศ โดยให้ผู้เชี่ยวชาญเป็นผู้ตรวจสอบว่าข้อสอบข้อใดที่จะนำมาเป็นข้อสอบศึกษาแล้วทำการตรวจสอบ ส่วนความสามารถของผู้สอบใช้การจับคู่ความสามารถ แล้วพิจารณาผลการตรวจสอบว่าแตกต่างหรือสอดคล้องกับผลการศึกษาครั้งนี้

บรรณานุกรม

บรรณานุกรม

นงลักษณ์ วิรัชชัย. (2540). *ความไม่แปรเปลี่ยนของแบบจำลองการเป็นสมาชิกด้วยใจรักของครูระหว่างบุคลากรครู 2 กลุ่ม: การประยุกต์ใช้การสร้างแบบจำลองสมการโครงสร้างชนิดกลยุทธ์กลุ่มพหุ*. กรุงเทพฯ: ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

_____. (2542). *โมเดลลิสเรล สถิติวิเคราะห์สำหรับการวิจัย*. กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

ผจงจิต อินทสุวรรณ. (2544, ม.ค.-เม.ย.2544). วิธีการวัดในงานวิจัยข้ามวัฒนธรรม: การอนุมานสาเหตุของวิกฤตเศรษฐกิจ. *วารสารวิธีวิทยาการวิจัย*. 14(1): 1-18.

บุญเชิด ภิญโญอนันตพงษ์. (2549). *ทฤษฎีทดสอบคะแนนจริงสัมพันธ์ (Congeneric Test Theory): จากทฤษฎีสู่การปฏิบัติเพื่อพัฒนาคุณภาพการศึกษา*. ภาควิชาวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. 17-44.

วรรณี แกมเกต; นงลักษณ์ วิรัชชัย; และสมหวัง พิธิยานุวัฒน์. (2540). การพัฒนาตัวบ่งชี้ประสิทธิภาพการใช้ครูและการทดสอบความไม่แปรเปลี่ยนของโมเดลประสิทธิภาพการใช้ครูโดยใช้การวิเคราะห์โมเดลสมการโครงสร้างกลุ่มพหุ. *วารสารวิธีวิทยาการวิจัย*. 10(2): 19-45.

ศิริชัย กาญจนวาสี. (2549). *ทฤษฎีการวัด: พรมแดนของการศึกษาวิจัย (Measurement Theories: The State of the Art)*. ภาควิชาวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. 45-66.

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*. 29(1): 67-91.

Agresti, A. (2002). *Categorical data analysis.* Hoboken. NJ: John Wiley.

Allen, M.; & Yen, W. (1979). Introduction to measurement theory. Long Grove: Waveland Press.

Anastasi, A. (1988). *Psychological testing (6th ed.)*. New York: Macmillan.

Angoff, W. H. (1972, September). *A technique for the investigation of cultural differences.* Paper presented at the annual meeting of the American Psychological Association. Honolulu. (ERIC Document Reproduction Service No. ED 069 686)

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability: In: Lord F. M.; & Novick M. R. (eds.). *Statistical Theories of Mental Test Scores.* (pp. 397-472). Reading. MA: Addison-Wesley.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brislin, R. W. (1970). Back-translations for cross-cultural Research. *Journal of Cross-Cultural Psychology*. 1(3): 185-216.

Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*. 29: 289-311.

_____. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Camilli, G.; & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Candell, G. L.; & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement*. 12: 253-260.

Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research*. 35: 169-199.

Cheung, G. W.; & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 9: 233–255.

Church, T. A.; & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality: Rationale and current research. *Journal of Cross-Cultural Psychology*. 29: 32–62.

Cohen, A. S.; Kim, S. H., & Baker, F. B. (1993). Detection of differential item function of DIF. *Applied Psychological Measurement.* 17: 335-350.

Cohen, A. S.; Kim, S. H.; & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*. 20: 15-26.

Collins, W. C., Raju, N. S., & Edwards, J. E. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology,* 85: 451-461.

Donoghue, J. R.; & Allen, N. L. (1993). Thin versus thick matching in the Mantel-Haenszel procedure for detecting DIF. *Journal of Educational Statistics*. 18: 131-154.

Donovan, M. A.; Drasgow, F.; & Probst, T. M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference New IRT analyses offer insight. *Journal of Applied Psychology*. 85: 305-313.

Dorans, N. J.; & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In: Holland, P. W.; & Wainer H. (Eds.)*: Differential item functioning*: 35-66. Hillsdale. NJ: Erlbaum.

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*. 95: 134-135.

_____. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*. 72: 19-29.

Drasgow, F.; & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology.* 70: 662-680.

Drasgow, F.; & Parsons, C. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement.* 7: 189-199.

Ebel, R. L. (1965). *Measuring educational achievement.* Englewood Cliffs. NJ: Prentice-Hall.

Ellis, B. E.; & Kimmel, H. D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology*. 77: 177-184.

Ellis, B. E.; & Mead, A. D. (2000). Assessment of the measurement equivalence of a Spanish translation of the 16PF questionnaire. *Educational and Psychological Measurement.* 60(5): 787-807.

Embretson S. E.; & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah. New Jersey: Lawrence Erlbaum Associates.

Embretson, S. (1999). The new rules of measurement. *Psychological Assessment.* 8: 341-49.

Embretson, S. E.; & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates, Inc.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable. *Journal of Applied Psychology.* 86: 215-227.

Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST and the IRT likelihood ratio. *Applied Psychological Measurement*. 29: 278-295.

Finch, W. H.; & French, B. F. (2006, June). Type I error rate for one type of DIF detection in the presence of the other type: A Monte Carlo study. *Paper presented at the International Meeting of the Psychometric Society.* Montreal: QC, Canada.

_____. (2007). Detection of Crossing Differential Item Functioning A Comparison of Four Methods. *Educational and Psychological Measurement*. 67(4): 565-582.

Flowers, C. P.; Raju, N. S.; & Oshima, T. C. (2002, April). A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory. *Paper presented at the annual meeting of the National Council on Measurement in Education,* New Orleans, LA.

Flowers, C.P., Oshima, T.C., & Raju, N.S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement,* 23: 309-326.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling,* 13: 378-402.

Goddard, P. E. (1917). The Beaver Indians. American Museum of Natural History. *Anthropological Papers*. 10(3): 201-93.

Golembiewski, R.; Billingsley, K.; & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*. 12: 133-157.

Gould, S.J. (1981). *The Mismeasure of Man*. Norton. New York.

_____. (May, 1981). *Evolution as Fact and Theory*. New York: W. W. Norton & Company. 1994: 253-262.

Graen, G. B.; Novak, M. A.; & Sommerkamp, P. (1982). The effects of leader-member exchange and job design on productivity and job satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Performance*. 30: 109-131.

Graen, G. B.; & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multilevel multi-domain perspective. *Leadership Quarterly*. 6: 219-247.

Hambelton, R.; & Jones, R. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*. 12: 38-47.

Hambleton, R. K.; & Swaminathan, H. (1985). *Item response theory.* Boston: Kluwer Nijhoff.

Hambleton, R. K.; Swaminathan, H.; & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park. CA: Sage.

Hambleton, R. K.; & Rovinelli, R. (1973). A Fortran IV program for generating examinee response data from logistic test models. *Behavioral Science,* 17: 73-74. (Software was revised, September 1990)

Han, K. T. (2007). WinGen2: *Windows software that generates IRT parameters and item responses [computer program]*. Amherst, MA: University of Massachusetts, Center for Educational Assessment. Retrieved May 13, 2007, from http://www.umass.edu/remp/software/wingen/

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. *In H. Holland & H. I. Braun (Eds.), Test validity*: 129-145. Hillsdale, NJ: Lawrence Erlbaum.

Holmes Finch. (2005). The MIMIC Model as a Method for Detecting DIF: Comparison With Mantel-Haenszel, SIBTEST,and the IRT Likelihood Ratio. *Applied Psychological Measurement.* 29: 278-295.

Horn, J. L.; & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research.* 18: 117-144.

Horn, J. L.; McArdle, J. J.; & Mason, R. (1983). When is invariance not invariant: A practical scientist's book at the ethereal concept of factor invariance. *The Southern Psychologist.* 1: 179-188.

Hothersall, D. (1995). *History of Psychology (3rd Ed.)*. McGraw Hill.

Huang, C.; Church, A.T.; & Katigbak, M.S. (1997). Identifying cultural differences in items and traits: Differential function in the NEO Personality Inventory. Journal of Cross-Cultural Psychology. 28: 192-218.

Hui, C. H.; & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*. 16(2): 131-152.

Hui, C. H.; & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross- Cultural Psychology*. 20: 296-309.

Hulin, C. L. (1987). A psychometric theory of evaluations of item and scale translations: Fidelity across languages. *Journal of Cross-Cultural Psychology*. 18(2): 115-142.

Hulin, C. L.; Drasgow, F.; & Parsons, C. K. (1983). Item response theory: Application to psychological measurement. Homewood. IL: Dow Jones-Irwin.

Hulin, C. L.; Drasgow, F.; & Komocar, J. (1982). Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology*. 67: 818-825.

Hulin, C. L.; Lissak, R. I.; & Drasgow, F. (1982). Recovery of two- and Three-parameter logistic item characteristic curve: A monte carlo study. *Applied Psychological Measurement.* 6: 249-260.

Johnson, D.E. (1998). *Applied Multivariate Methods for Data Analysis*. Pacific Grove. CA: Brooks/Cole Publishing.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika.* 36: 409-426.

_____. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika.* 59: 381-389.

Jöreskog, K. G.; & Sörbom, D. (1988). *PRELIS: A program for multivariate data screening and data summarization. A predecessor for LISREL* (2nd ed.). Mooresville. IN: Scientific Software.

_____. (1989). *LISREL 7: User's reference guide.* Chicago: Scientific Software.

_____. (1993). LISREL 8: User's guide. Chicago: Scientific Software International.

Keats, J. A. (1997). *Classical Test Theory*. In: Keeves J. P. (Ed.). Educational Research, Methodology, and Measurement: An International Handbook (2nd ed., pp. 713-19). Oxford: Pergamon Press.

Kim, S. H.; & Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement*. 29: 551-566.

_____. (1997, March). An investigation of the likelihood ratio test for detection of differential item functioning under the graded response model. *Paper presented at the annual meeting of the American Educational Research Association*. Chicago.

_____. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement*. 22: 345-355.

Linn, R.L. (Ed.) (1989). *Educational measurement (3rd Ed.)*. New York: American Council on Education/Macmillan.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*. 32: 53-76.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale. NJ: Lawrence Erlbaum.

Lord, F. M.; & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading. MA: Addison-Wesley.

Lubke, G. H.; & Muthe´n, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*. 11: 514-534.

Lynn, M. (2006). Geo-demographic differences in knowledge about the restaurant tipping norm. Journal of Applied Social Psychology, 36(3), 740-750.

MacIntosh, R.; & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement.* 27(5): 372-379.

Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. *Educational and Psychological Measurement.* 61: 793-817.

Marsh, H. W.; & Hau, Kit-Tai. (1996). Assessing Goodness of Fit: Is Patrimony Always Desirable?. *Journal of Experimental Education*. 64(4): 364-391.

Marsh, H. W.; Balla, J. R.; & McDonald, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*. 103(3): 391-410.

Marsh, H. W.; Hau, Kit-Tai.; Chung, Choi-Man.; & Sui, T. L. P. (1998). Confirmatory Factor Analyses of Chinese Students' Evaluations of University Teaching. *Structural Equation Modeling.* 5(2): 143-164.

Maurer, T. J.; Raju, N. S.; & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology.* 83: 693-702.

McDonald, R. P. (1999). Test theory : a unified treatment. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

McDonald, R. P.; & Paunonen, S. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*. 62: 921-943.

McLaughlin, M. E.; & Drasgow, F. (1987). Lord's chi-square test of item bias with estimated and known person parameters. *Applied Psychological Measurement*. 21: 161-173.

McNemar, Q. (1962). *Psychological Statistics (3rd ed.).* New York: Wiley.

Meade, A. W.; & Lautenschlager, G. L. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods.* 7(4): 361-388.

_____. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling.* 11: 60-72.

Meade, A. W.; Lautenschlager, G. J.; Michels, L. C.; & Gentry, W. (2003, April). The equivalence of online and paper and pencil assessments. *Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology*. Orlando, FL.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525–543.

Narayanan, P.; & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement.* 18(4): 315-328.

Narayanan, P.; & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement.* 20: 257-274.

Nunnally, J. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.

Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation modeling.* 5: 107-124.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika.* 53: 495-502.

_____. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*. 14: 197-207.

Raju, N. S.; Laffitte, L. J.; & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on Confirmatory Factor Analysis and item response theory. *Journal of Applied Psychology*. 87: 517–529.

Raju, N. S.; van der Linden, W.; & Fleer, P. (1995). An IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement.* 19(4): 353-368.

Ree, M. J.; & Jensen, H. E. (1983). Effects of sample size on linear equating of item characteristic curve parameters. In: Weiss, D.J. (Edito) 1983. New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing. *Academic Press*. New York. 135–146.

Reise, S. P.; Widaman, K. F.; & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*. 114: 552-566.

Riordan, C. M.; & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner. *Journal of Management.* 20: 643-671.

Rogers, H. J.; & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement.* 17: 105-116.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement.* 34: 100-114.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology.* 27: 229-239.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology.* 3: 271-95.

Stage, C. (1998a). A comparison between item analysis based on item response theory and classical test theory: A study of the SweSAT ERC. *Educational Measurement.* No 30. Umea University, Department of Educational Measurement.

Stage, C. (1998b). A comparison between item analysis based on item response theory and classical test theory: A study of the SweSAT test WORD. *Educational Measurement*. No 29. Umea University, Department of Educational Measurement.

Stage, C. (1999). A comparison between item analysis based on item response theory and classical test theory: A study of the SweSAT test READ. *Educational Measurement.* No 31. Umea University, Department of Educational Measurement.

Su, Y. H.; & Wang, W. C. (2005). Efficiency of the Mantel, generalized Mantel-Haenszel, and logistic discriminant function analysis methods in detecting differential item functioning in polytomous items. *Applied Measurement in Education*. 18: 313-350.

Swaminathan, H.; & Gifford J. A. (1983). *Estimation of parameters in the three-parameter latent trait model*. In: Weiss, D. J. (Ed.) New Horizons in Testing: Latent Traits Test Theory and Computerized Adaptive Testing. New York: Academic Press.

Swaminathan, H.; & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*. 27: 361-370.

Taris, T. W.; Bok, I. A.; & Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *Journal of Psychology*. 132: 301-316.

Thissen, D. (2003). *MULTILOG 7: Multiple categorical item analysis and test scoring using item response theory [computer program]*. Chicago. IL: Scientific Software.

Thissen, D.; Steinberg, L.; & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In: Wainer, H.; & Braun H. I. (Eds.). *Test validity*. 147-169. Hillsdale. NJ: Lawrence Erlbaum.

_____. (1993). Detection of differential item functioning using the parameters of item response models. In: Holland, P. W.; & Wainer, H. (Eds.). *Differential item functioning*, 67-113. Hillsdale. NJ: Lawrence Erlbaum.

Uttaro, T.; & Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement*. 18: 15-25.

Van de Vijver, F. (1998). Towards a theory of bias and equivalence: *ZUMA Nachrichten Spezial*. No. 3: Cross-cultural survey equivalence: 41-65.

Van de Vijver, F.; & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks. CA: Sage.

Van de Vijver, F.; & Leung, K. (1997a). Methods and Data Analysis for Cross-Cultural Research. Thousand Oaks: Sage.

Van de Vijver, F.; & Leung, K. (1997b). *Methods and data analysis of comparative research*. 257-300. In: Handbook of cross-cultural psychology (2nd ed.; vol. 1). edited by Berry, J.W.; Poortinga, Y.H.; & Pandey, J. Boston: Allyn & Bacon.

Van de Vijver, F.J.; & Poortinga ,Y.H. (1982). Cross-Cultural Generalization And Univerality. *Journal of Cross-Cultural Psychology*. 13(4): 387–408.

van der Linden, W. J.; & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*. 5: 139-158.

Vandenberg, R. J.; & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 3(1): 4-70.

Vandenberg, R. J.; & Self, R. M. (1993). Assessing newcomer's changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology.* 78: 557-568.

Wang, W. C.; & Yeh, L. Y. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*. 27: 479-498.

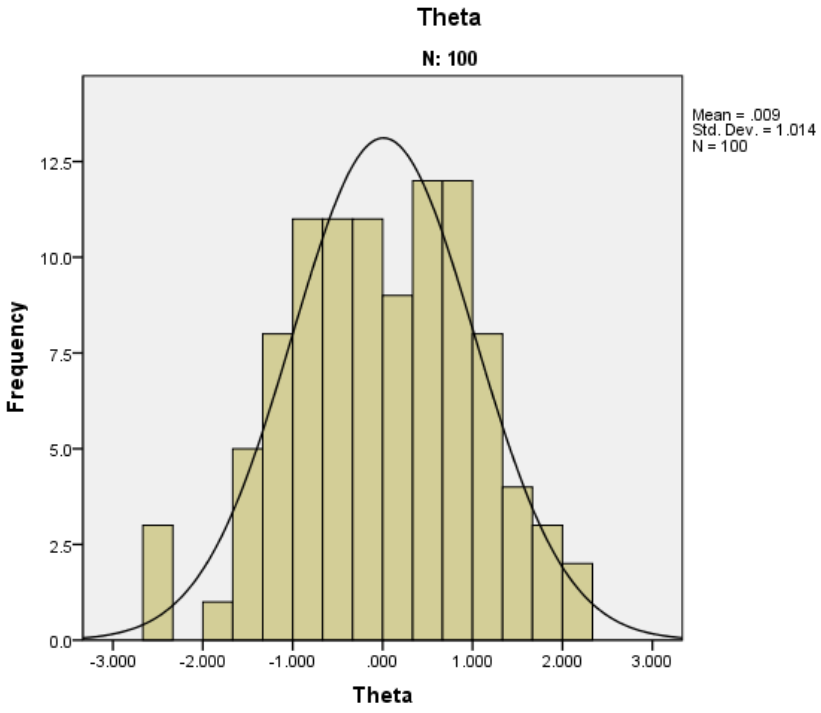
Wanichtanom, R. (2001). Methods of detecting differential item functioning: *A comparison of item response theory and confirmatory factor analysis.* Unpublished doctoral dissertation.

ภาคผนวก

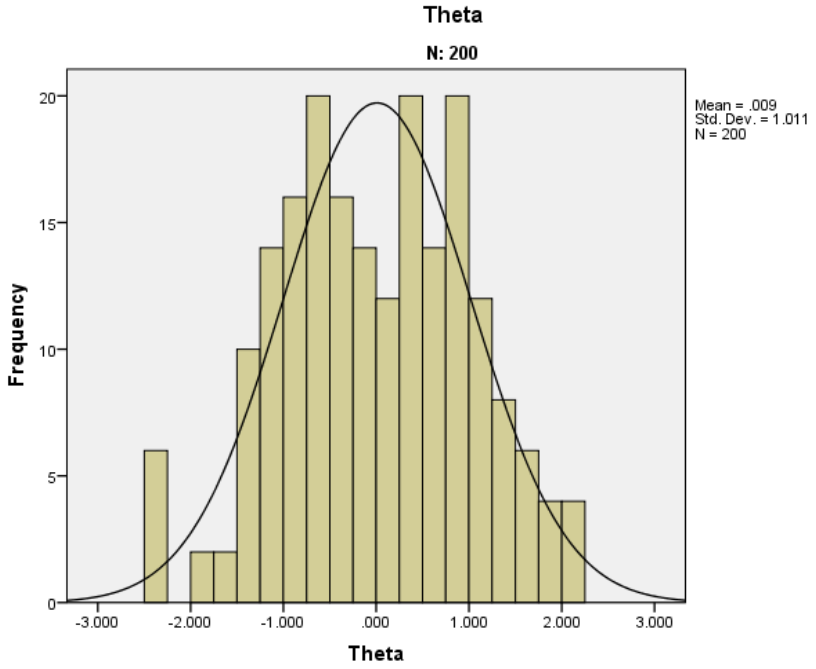
ภาคผนวก ก
ภาพ Histogram และกราฟเส้นแสดง การกระจายของค่าพารามิเตอร์ความสามารถของผู้สอบ และการกระจายค่าพารามิเตอร์ของข้อสอบ

การวิเคราะห์การกระจายค่าพารามิเตอร์ความสามารถของผู้สอบ

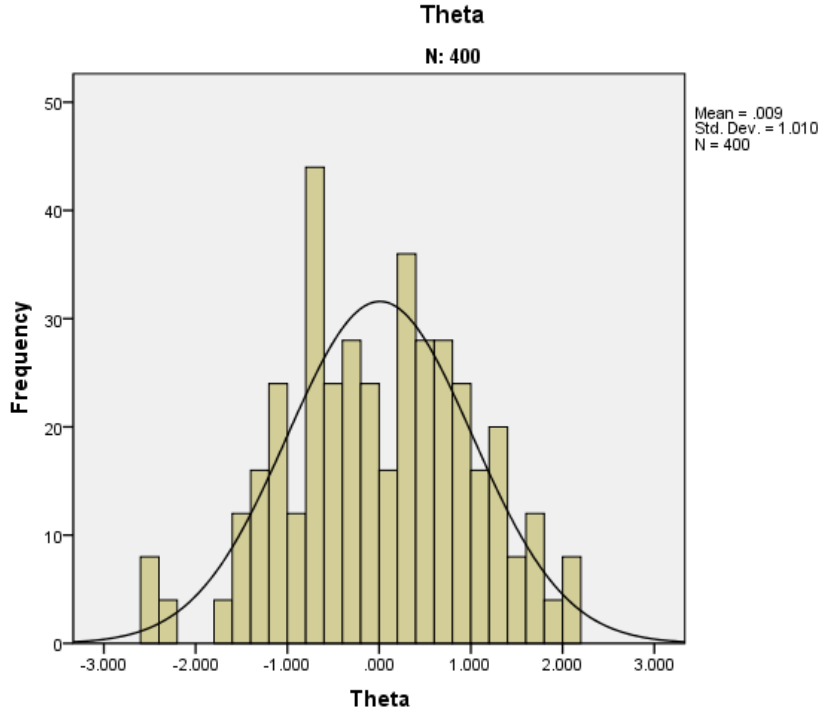
การวิเคราะห์ขตอนนี้ ผู้วิจัยได้นำข้อมูลค่าพารามิเตอร์ความสามารถของผู้สอบมาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล พบว่า **จำนวนผู้สอบ 100 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.12 ค่าความโด่งเท่ากับ -0.36 **จำนวนผู้สอบ 200 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.38 **จำนวนผู้สอบ 400 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.39 **จำนวนผู้สอบ 500 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40 **จำนวนผู้สอบ 800 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40 และ**จำนวนผู้สอบ 1,000 คน** มีค่าพารามิเตอร์ความสามารถอยู่ในช่วงตั้งแต่ -2.49 ถึง 2.14 มีค่าเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าความเบ้เท่ากับ -0.11 ค่าความโด่งเท่ากับ -0.40 เมื่อพิจารณาค่าสถิติต่าง ๆ ประกอบกับรูปร่างของกราฟเส้นแสดงให้เห็นว่า ความสามารถของผู้สอบทุกกลุ่ม มีการกระจายค่อนข้างจะเป็นโค้งปกติ แสดงดังภาพประกอบ 9-14



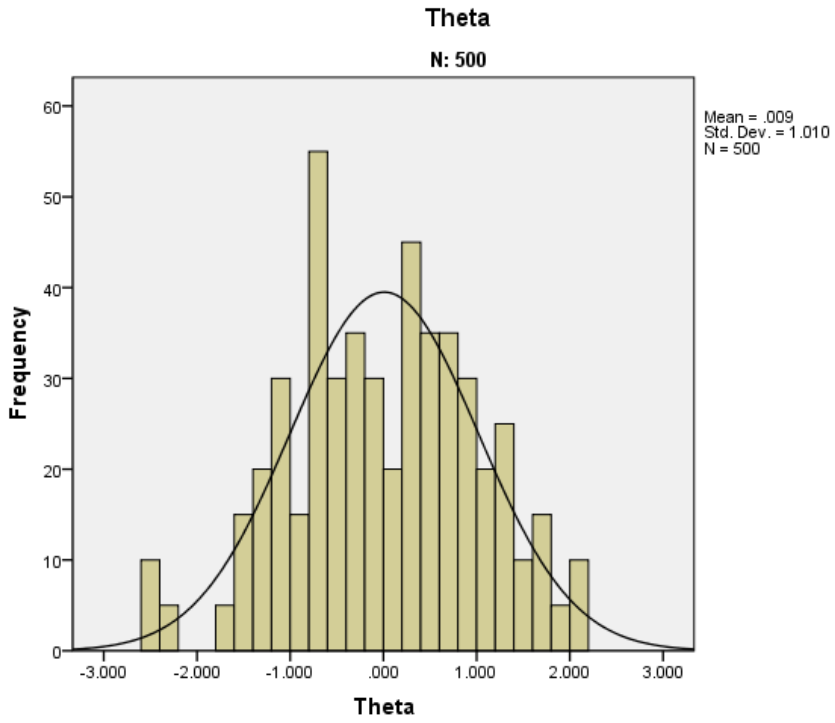
ภาพประกอบ 9 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 100 คน



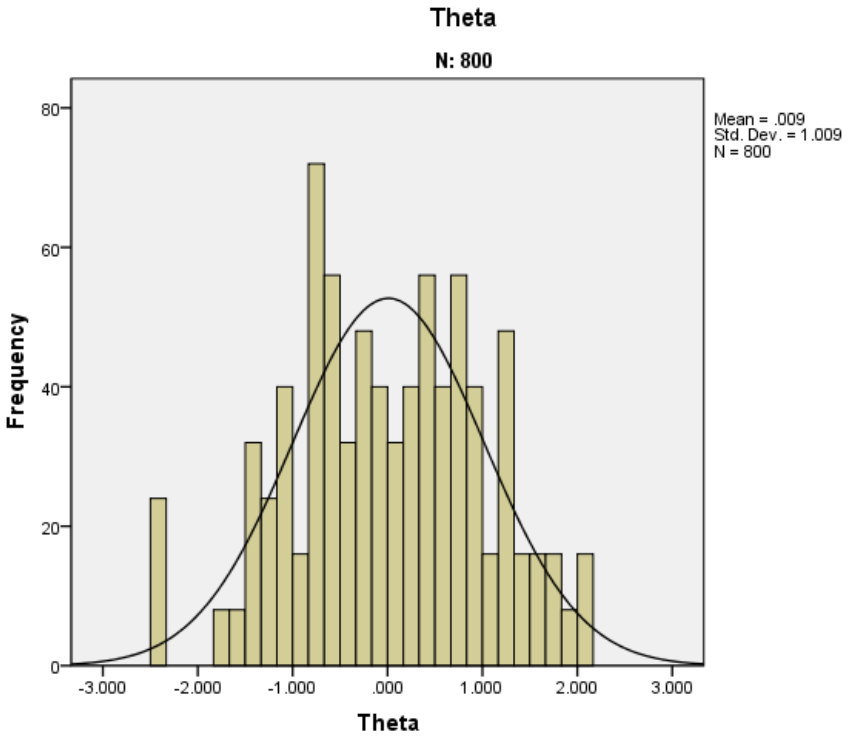
ภาพประกอบ 10 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 200 คน



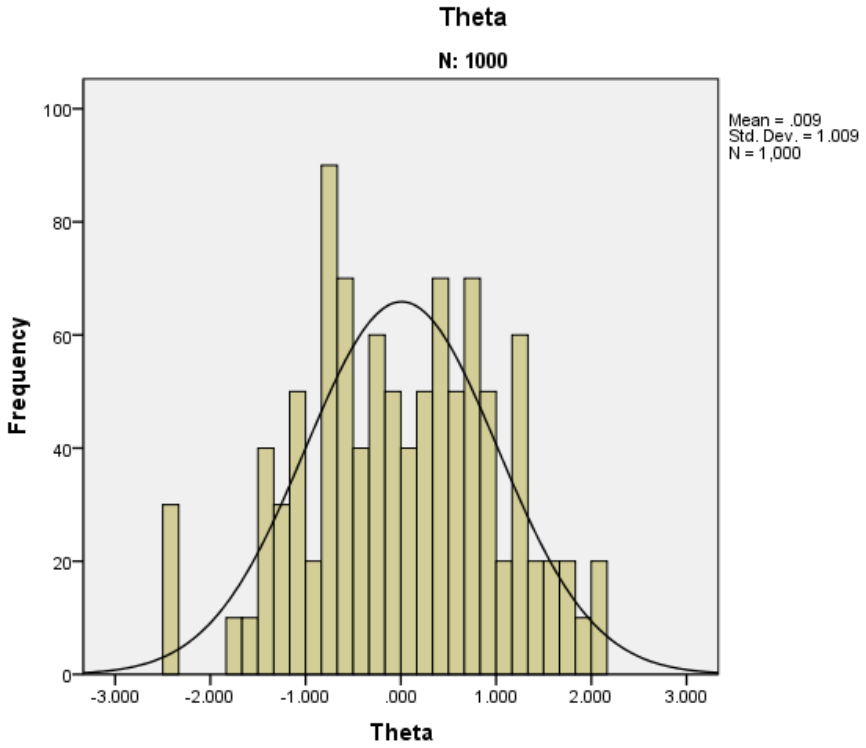
ภาพประกอบ 11 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 400 คน



ภาพประกอบ 12 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 500 คน



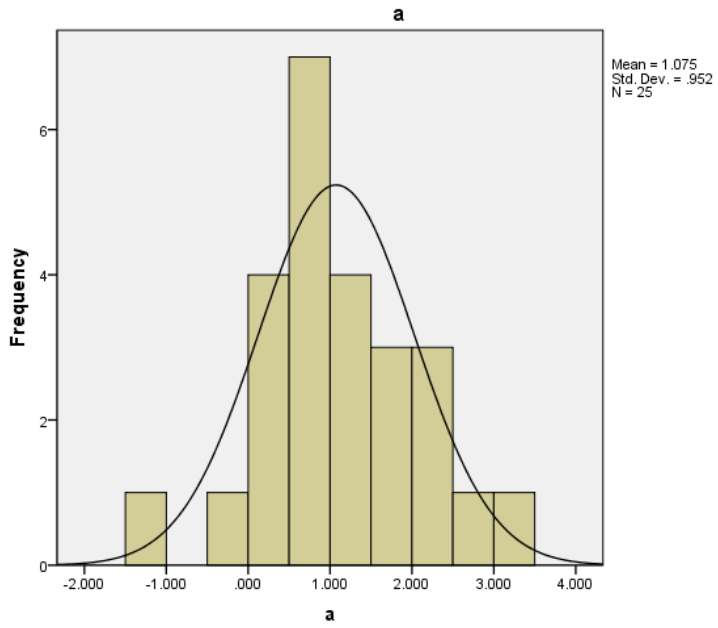
ภาพประกอบ 13 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 800 คน



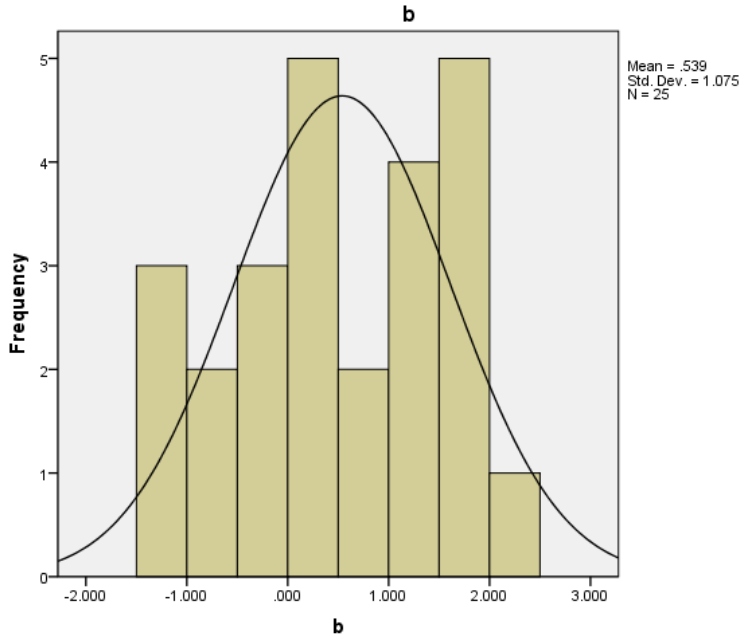
ภาพประกอบ 14 การกระจายความสามารถของค่าพารามิเตอร์ของผู้สอบ 1000 คน

การวิเคราะห์การกระจายค่าพารามิเตอร์ของข้อสอบที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า

การวิเคราะห์ขตอนนี้ ผู้วิจัยได้นำข้อมูลค่าพารามิเตอร์ของข้อสอบที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า มาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล พบว่า ค่าพารามิเตอร์อำนาจจำแนกมีพิสัยอยู่ระหว่าง -1.12 ถึง 3.12 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.95 ค่าความเบ้เป็น 0.09 และค่าความโด่งเป็น 0.17 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วนพารามิเตอร์ความยาก (b_i parameter) ของข้อสอบซึ่งมีช่วงอยู่ระหว่าง -2.0 ถึง +2.0 ค่าของความยากของข้อสอบจะแสดงตำแหน่งบนโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50 พารามิเตอร์ความยากของข้อสอบที่เพิ่มมากขึ้น ผู้สอบที่มีความสามารถมากขึ้นที่จะตอบข้อสอบข้อนั้นได้ถูกจะอยู่ในตำแหน่งที่สูงกว่า จากการสร้างพารามิเตอร์ข้อสอบ พบว่า ค่าพารามิเตอร์ความยากมีพิสัยอยู่ระหว่าง -1.27 ถึง 2.33 ค่าเฉลี่ยเท่ากับ 0.54 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.08 ค่าความเบ้เป็น -0.10 และค่าความโด่งเป็น -1.10 ซึ่งอยู่ในขอบเขตของพารามิเตอร์ความยากตามเกณฑ์ที่กำหนดไว้ เมื่อพิจารณาค่าสถิติต่าง ๆ ประกอบกับรูปร่างของกราฟเส้นแสดงให้เห็นว่าความสามารถของผู้สอบทุกกลุ่ม มีการกระจายค่อนข้างจะเป็นโค้งปกติ แสดงดังภาพประกอบ 15-16



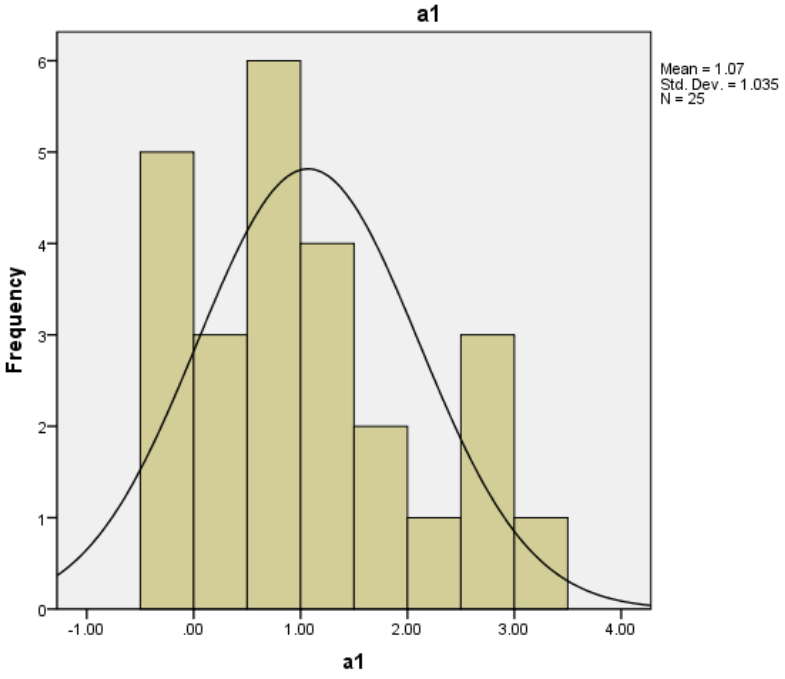
ภาพประกอบ 15 การกระจายความสามารถของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า



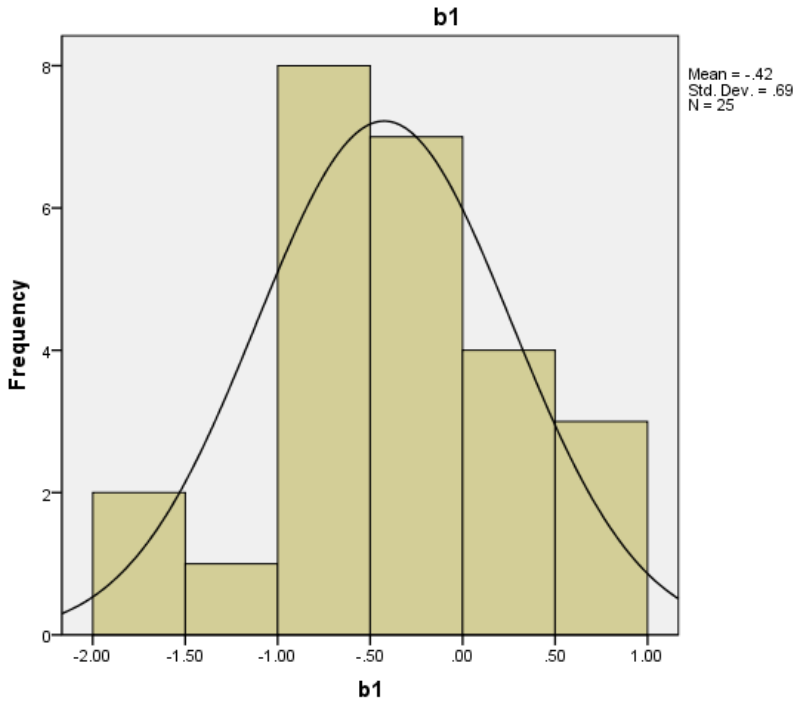
ภาพประกอบ 16 การกระจายความสามารถของค่าพารามิเตอร์ความยากของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบสองค่า

การวิเคราะห์การกระจายค่าพารามิเตอร์ของข้อสอบที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า

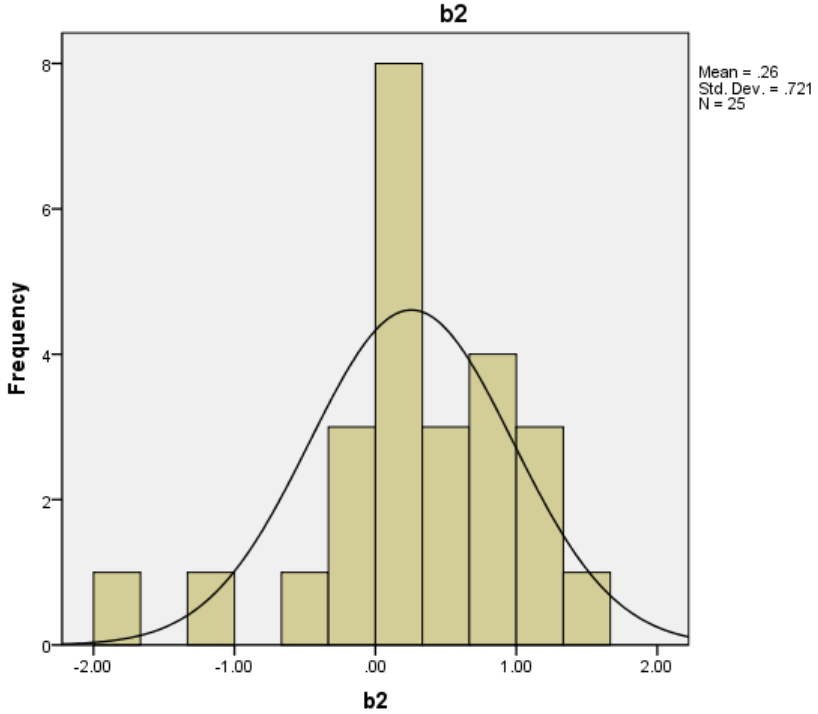
การวิเคราะห์ขตอนนี้ ผู้วิจัยได้นำข้อมูลค่าพารามิเตอร์ของข้อสอบที่รายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า มาวิเคราะห์ความถี่แล้วนำเสนอในรูปของ Histogram และกราฟเส้นของการกระจายข้อมูล พบว่า ค่าพารามิเตอร์ความชันร่วมมีพิสัยอยู่ระหว่าง -0.33 ถึง 3.04 ค่าเฉลี่ยเท่ากับ 1.07 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.21 ค่าความเบ้เป็น 0.47 และค่าความโด่งเป็น -0.71 ซึ่งอยู่ในขอบเขตของพารามิเตอร์อำนาจจำแนกตามเกณฑ์ที่กำหนดไว้ ส่วน ค่า Threshold ของแต่ละรายการคำตอบของข้อสอบมีพิสัยอยู่ระหว่าง -1.90 ถึง 2.41 ค่าเฉลี่ยเท่ากับ 0.56 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.96 ค่าความเบ้เป็น -0.08 และค่าความโด่งเป็น -0.51 ซึ่งอยู่ในขอบเขตของพารามิเตอร์ความยากตามเกณฑ์ที่กำหนดไว้ เมื่อพิจารณาค่าสถิติต่าง ๆ ประกอบกับรูปร่างของกราฟเส้นแสดงให้เห็นว่า ความสามารถของผู้สอบทุกกลุ่ม มีการกระจายค่อนข้างจะเป็นโค้งปกติ แสดงดังภาพประกอบ 17-28



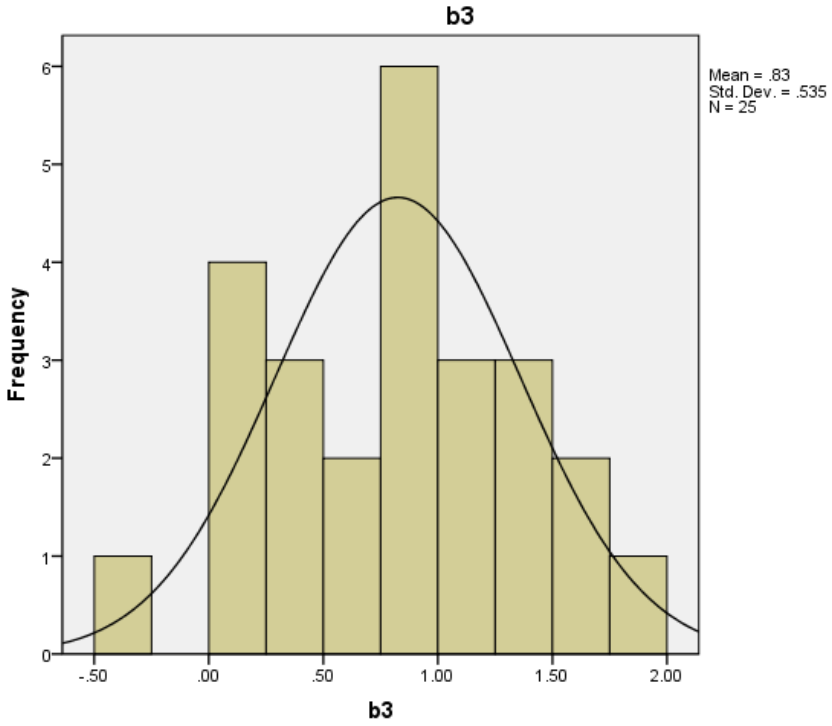
ภาพประกอบ 17 การกระจายความสามารถของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า



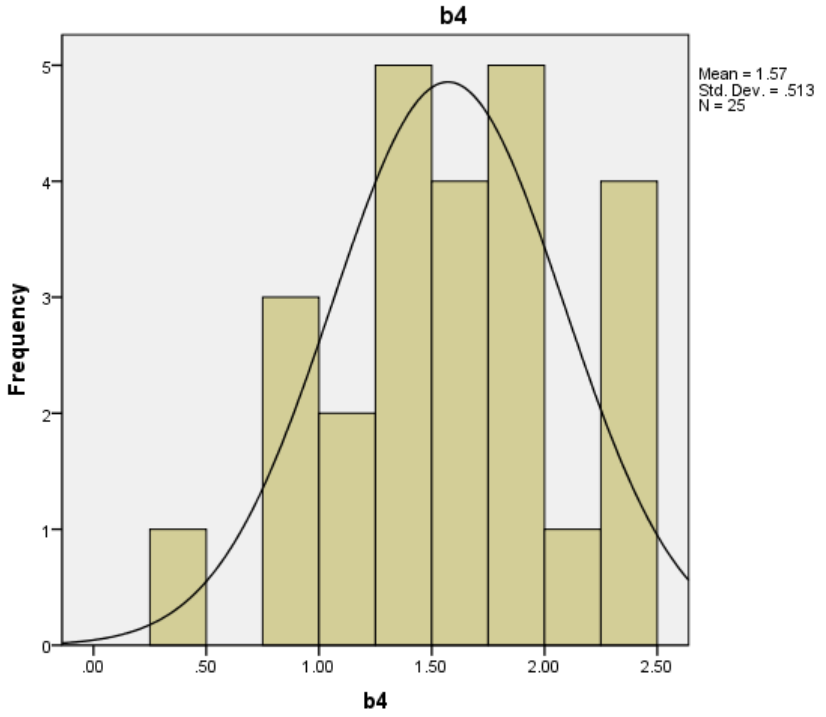
ภาพประกอบ 18 การกระจายความสามารถของค่าพารามิเตอร์ความยากประจำขั้นที่ 1 ของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า



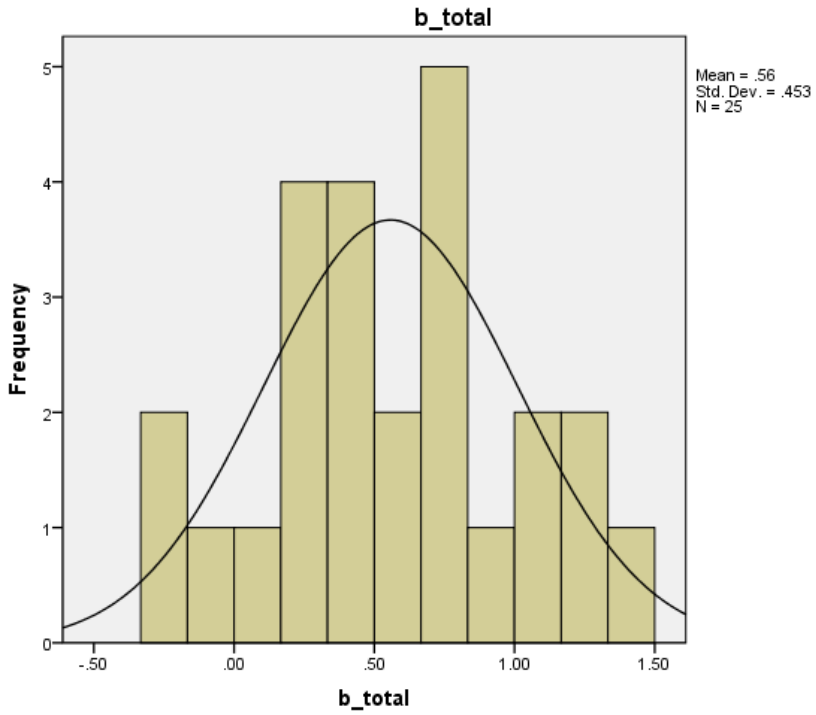
ภาพประกอบ 19 การกระจายความสามารถของค่าพารามิเตอร์ความยากประจำขั้นที่ 2 ของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า



ภาพประกอบ 20 การกระจายความสามารถของค่าพารามิเตอร์ความยากประจำขั้นที่ 3 ของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า



ภาพประกอบ 21 การกระจายความสามารถของค่าพารามิเตอร์ความยากประจำขั้นที่ 4 ของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า



ภาพประกอบ 22 การกระจายความสามารถของค่าพารามิเตอร์ความยากประจำขั้นรวม ของข้อสอบสำหรับกลุ่มอ้างอิง กรณีรายการคำตอบข้อสอบเป็นการตรวจให้คะแนนแบบหลายค่า

ภาคผนวก ข
ตัวอย่างคำสั่งในการวิเคราะห์ความเท่าเทียมกันของการวัด

ตัวอย่างคำสั่ง MULTILOGและ OUTPUT MULTILOG
สำหรับรายการคำตอบของผู้สอบ 100 คน
ที่เป็นการตรวจให้คะแนนแบบสองค่า
ข้อสอบศึกษาข้อที่ 5

```
MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
N100_I5.MLG -'I5' DIF
            20-ITEM ANCHOR , 2PL    >PROBLEM RANDOM,
          INDIVIDUAL,
          DATA = '100_01-I05_001',
          NITEMS = 42,
          NGROUPS = 2,
          NEXAMINEES = 200;
   DATA FILE NAME IS
          D\100\L\100_01-I05_001.TXT
  TYPE OF INPUT:
    INDIVIDUAL RESPONSE VECTORS
  >TEST ALL,
       L2;
  ---> FIND WARNING:    1 RECORDS NOT CONTAINING > IN COLUMN 1 HAVE BE
  NUMBER OF CODES  2
1
  VECTOR OF CORRECT RESPONSES
1.11111E+41
  CODES FOR MISSING OR NOT-REACHED
  N
 (I1,42A1)
  MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
   N100_I5.MLG -'I5' DIF
               20-ITEM ANCHOR , 2PL
  DATA PARAMETERS:
   NUMBER OF LINES IN THE DATA FILE:  200
   NUMBER OF CATEGORICAL-RESPONSE ITEMS:  42
   NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS:   2
   TOTAL NUMBER OF ITEMS (INCLUDING GROUPS):  44
   NUMBER OF CHARACTERS IN ID FIELDS:  0
   MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM:  2
   THE MISSING VALUE CODE FOR CONTINUOUS DATA:  9.0000
   THE DATA WILL BE STORED IN MEMORY
```

ESTIMATION PARAMETERS:

THE ITEMS WILL BE CALIBRATED--

BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

MAXIMUM NUMBER OF EM CYCLES PERMITTED: 25

NUMBER OF PARAMETER-SEGMENTS USED IS: 43

NUMBER OF FREE PARAMETERS IS: 85

MAXIMUM NUMBER OF M-STEP ITERATIONS IS 4 TIMES

THE NUMBER OF PARAMETERS IN THE SEGMENT

NUMBER OF QUADRATURE POINTS IS: 19

THE M-STEP CONVERGENCE CRITERION IS: 0.000100

THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000

THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS: 0.9000

THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS: 1.0000

THE MAXIMUM ACCELERATION PERMITTED IS: 0.0000

THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED

QUADRATURE POINTS FOR MML,

AT THETA:

-4.5

-4

-3.5

-3

-2.5

-2

-1.5

-1

-0.5

0

0.5

1

1.5

2

2.5

3

3.5

```
4
4.5
 N100_I5.MLG -‘I5’ DIF
 READING DATA...
 KEY-
 CODE  CATEGORY
  0     1111111111111111111111111111111111111111
  1     2222222222222222222222222222222222222222
 FORMAT FOR DATA-
 (I1,42A1)
 FIRST OBSERVATION AS READ-
 ITEMS 101011010110001111100
 NORML       0.000      9.000
 FINISHED CYCLE  25
 MAXIMUM INTERCYCLE PARAMETER CHANGE=  23.98401 P(  26)
 ITEM SUMMARY
 N100_I5.MLG -‘I5’ DIF
 ITEM   1:        2 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A          1    2.12  (0.61)
 B( 1)      2    0.01  (0.16)
 @THETA:       INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.008  0.012  0.018  0.027  0.041  0.061  0.092  0.138
 -1.4 -  0.0  0.203  0.296  0.419  0.575  0.753  0.929  1.064  1.119
  0.2 -  1.6  1.078  0.952  0.780  0.600  0.440  0.312  0.215  0.146
  1.8 -  3.0  0.098  0.065  0.043  0.028  0.019  0.012  0.008
  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2
  OBS. FREQ.      52      48
  OBS. PROP.  0.5200 0.4800
  GROUP  1:
  EXP. PROP.  0.5175 0.4825
  GROUP  2:
  EXP. PROP.  0.5045 0.4955
```

ITEM 2: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 3 0.00 (0.23)

B(1) 4 83.41 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 54 46

OBS. PROP. 0.5400 0.4600

GROUP 1:

EXP. PROP. 0.5399 0.4601

GROUP 2:

EXP. PROP. 0.5399 0.4601

ITEM 3: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 5 0.05 (0.39)

B(1) 6 -16.48 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 31 69

OBS. PROP. 0.3100 0.6900

GROUP 1:

EXP. PROP. 0.3216 0.6784

GROUP 2:

EXP. PROP. 0.3212 0.6788

ITEM 4: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 7 0.95 (0.30)

B(1) 8 0.36 (0.29)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.034 0.041 0.048 0.057 0.067 0.079 0.091 0.105

-1.4 - 0.0 0.121 0.137 0.154 0.170 0.186 0.200 0.212 0.221

0.2 - 1.6 0.226 0.228 0.225 0.218 0.208 0.195 0.180 0.163

1.8 - 3.0 0.147 0.130 0.114 0.099 0.086 0.074 0.063

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 58 42

OBS. PROP. 0.5800 0.4200

GROUP 1:

EXP. PROP. 0.5794 0.4206

GROUP 2:

EXP. PROP. 0.5712 0.4288

ITEM 5: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 9 0.91 (0.35)

B(1) 10 -1.26 (0.50)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.117 0.131 0.145 0.159 0.172 0.184 0.194 0.201

-1.4 - 0.0 0.205 0.206 0.203 0.197 0.189 0.178 0.165 0.151

0.2 - 1.6 0.137 0.123 0.109 0.096 0.083 0.072 0.062 0.053

1.8 - 3.0 0.046 0.039 0.033 0.028 0.023 0.020 0.017

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 28 72

OBS. PROP. 0.2800 0.7200

GROUP 1:

EXP. PROP. 0.2801 0.7199

GROUP 2:

EXP. PROP. 0.2735 0.7265

ITEM 6: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 11 0.01 (0.35)

B(1) 12 68.00 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 61 39

OBS. PROP. 0.6100 0.3900

GROUP 1:

EXP. PROP. 0.6094 0.3906

GROUP 2:

EXP. PROP. 0.6094 0.3906

ITEM 7: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 13 0.58 (0.27)

B(1) 14 0.71 (0.54)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.032 0.035 0.038 0.041 0.045 0.048 0.052 0.056

-1.4 - 0.0 0.060 0.063 0.067 0.071 0.074 0.077 0.079 0.082

0.2 - 1.6 0.083 0.084 0.085 0.085 0.084 0.083 0.082 0.080

1.8 - 3.0 0.077 0.074 0.071 0.067 0.064 0.060 0.056

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 60 40

OBS. PROP. 0.6000 0.4000

GROUP 1:

EXP. PROP. 0.5998 0.4002

GROUP 2:

EXP. PROP. 0.5943 0.4057

ITEM 8: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 15 -0.81 (0.30)

B(1) 16 -0.16 (0.33)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.054 0.062 0.070 0.079 0.088 0.098 0.109 0.119

-1.4 - 0.0 0.129 0.138 0.146 0.153 0.159 0.163 0.164 0.163

0.2 - 1.6 0.161 0.156 0.150 0.142 0.133 0.123 0.113 0.103

1.8 - 3.0 0.093 0.083 0.074 0.065 0.057 0.050 0.044

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 52 48

OBS. PROP. 0.5200 0.4800

GROUP 1:

EXP. PROP. 0.5203 0.4797

GROUP 2:

EXP. PROP. 0.5277 0.4723

ITEM 9: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 17 0.05 (0.25)

B(1) 18 3.30 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001

-1.4 - 0.0 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001

0.2 - 1.6 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001

1.8 - 3.0 0.001 0.001 0.001 0.001 0.001 0.001 0.001

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 54 46

OBS. PROP. 0.5400 0.4600

GROUP 1:

EXP. PROP. 0.5400 0.4600

GROUP 2:

EXP. PROP. 0.5395 0.4605

ITEM 10: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 19 -0.37 (0.25)

B(1) 20 0.07 (0.66)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.025 0.026 0.027 0.028 0.029 0.030 0.031 0.032

-1.4 - 0.0 0.032 0.033 0.033 0.034 0.034 0.034 0.035 0.035

0.2 - 1.6 0.035 0.035 0.034 0.034 0.034 0.033 0.033 0.032

1.8 - 3.0 0.031 0.031 0.030 0.029 0.028 0.027 0.026

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 49 51

OBS. PROP. 0.4900 0.5100

GROUP 1:

EXP. PROP. 0.4901 0.5099

GROUP 2:

EXP. PROP. 0.4938 0.5062

ITEM 11: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 21 -0.81 (0.26)

B(1) 22 0.07 (0.31)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.047 0.053 0.061 0.069 0.078 0.087 0.097 0.108

-1.4 - 0.0 0.118 0.128 0.138 0.146 0.154 0.159 0.163 0.165

0.2 - 1.6 0.165 0.162 0.158 0.152 0.144 0.135 0.125 0.115

1.8 - 3.0 0.104 0.094 0.084 0.075 0.066 0.058 0.051

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 48 52

OBS. PROP. 0.4800 0.5200

GROUP 1:

EXP. PROP. 0.4804 0.5196

GROUP 2:

EXP. PROP. 0.4878 0.5122

ITEM 12: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 23 1.71 (0.43)

B(1) 24 -0.22 (0.18)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.025 0.035 0.049 0.067 0.093 0.128 0.173 0.232

-1.4 - 0.0 0.305 0.391 0.486 0.580 0.662 0.717 0.732 0.706

0.2 - 1.6 0.644 0.557 0.462 0.368 0.286 0.216 0.161 0.118

1.8 - 3.0 0.086 0.062 0.045 0.032 0.023 0.016 0.012

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 45 55

OBS. PROP. 0.4500 0.5500

GROUP 1:

EXP. PROP. 0.4483 0.5517

GROUP 2:

EXP. PROP. 0.4366 0.5634

ITEM 13: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 25 0.00 (0.27)

B(1) 26 252.84 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 52 48

OBS. PROP. 0.5200 0.4800

GROUP 1:

EXP. PROP. 0.5199 0.4801

GROUP 2:

EXP. PROP. 0.5199 0.4801

ITEM 14: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 27 1.76 (0.49)

B(1) 28 -0.47 (0.19)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.036 0.050 0.070 0.098 0.135 0.185 0.249 0.329

-1.4 - 0.0 0.423 0.525 0.626 0.710 0.761 0.767 0.728 0.652

0.2 - 1.6 0.554 0.450 0.353 0.269 0.201 0.147 0.107 0.077

1.8 - 3.0 0.055 0.039 0.028 0.020 0.014 0.010 0.007

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 38 62

OBS. PROP. 0.3800 0.6200

GROUP 1:

EXP. PROP. 0.3787 0.6213

GROUP 2:

EXP. PROP. 0.3673 0.6327

ITEM 15: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 29 1.46 (0.41)

B(1) 30 0.30 (0.21)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.017 0.022 0.030 0.040 0.052 0.069 0.091 0.118

-1.4 - 0.0 0.152 0.193 0.242 0.297 0.357 0.416 0.470 0.510

0.2 - 1.6 0.532 0.532 0.510 0.470 0.416 0.357 0.297 0.242

1.8 - 3.0 0.193 0.152 0.118 0.091 0.069 0.052 0.040

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 59 41

OBS. PROP. 0.5900 0.4100

GROUP 1:

EXP. PROP. 0.5887 0.4113

GROUP 2:

EXP. PROP. 0.5780 0.4220

ITEM 16: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 31 0.00 (0.29)

B(1) 32 129.00 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 53 47

OBS. PROP. 0.5300 0.4700

GROUP 1:

EXP. PROP. 0.5297 0.4703

GROUP 2:

EXP. PROP. 0.5297 0.4703

ITEM 17: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 33 0.02 (0.24)

B(1) 34 -17.72 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 41 59

OBS. PROP. 0.4100 0.5900

GROUP 1:

EXP. PROP. 0.4212 0.5788

GROUP 2:

EXP. PROP. 0.4210 0.5790

ITEM 18: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 35 0.03 (0.26)

B(1) 36 -16.86 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 39 61

OBS. PROP. 0.3900 0.6100

GROUP 1:

EXP. PROP. 0.3954 0.6046

GROUP 2:

EXP. PROP. 0.3952 0.6048

ITEM 19: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 37 0.00 (0.34)

B(1) 38 93.63 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 54 46

OBS. PROP. 0.5400 0.4600

GROUP 1:

EXP. PROP. 0.5395 0.4605

GROUP 2:

EXP. PROP. 0.5395 0.4605

ITEM 20: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 39 0.11 (0.95)

B(1) 40 34.63 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 98 2

OBS. PROP. 0.9800 0.0200

GROUP 1:

EXP. PROP. 0.9796 0.0204

GROUP 2:

EXP. PROP. 0.9795 0.0205

ITEM 21: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 41 0.11 (0.28)

B(1) 42 1.43 (4.57)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

-1.4 - 0.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

0.2 - 1.6 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003

1.8 - 3.0 0.003 0.003 0.003 0.003 0.003 0.003 0.003

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 54 46

OBS. PROP. 0.5400 0.4600

GROUP 1:

EXP. PROP. 0.5400 0.4600

GROUP 2:

EXP. PROP. 0.5389 0.4611

ITEM 22: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 43 2.36 (0.64)

B(1) 44 -0.12 (0.15)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.006 0.010 0.016 0.026 0.041 0.065 0.103 0.161

-1.4 - 0.0 0.249 0.378 0.555 0.780 1.030 1.252 1.378 1.360

0.2 - 1.6 1.205 0.970 0.722 0.508 0.343 0.225 0.145 0.092

1.8 - 3.0 0.058 0.037 0.023 0.014 0.009 0.006 0.004

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 46 54

OBS. PROP. 0.4600 0.5400

GROUP 1:

EXP. PROP. 0.4736 0.5264

GROUP 2:

EXP. PROP. 0.4602 0.5398

ITEM 23: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 45 0.00 (0.25)

B(1) 46 -137.00 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 47 53

OBS. PROP. 0.4700 0.5300

GROUP 1:

EXP. PROP. 0.4702 0.5298

GROUP 2:

EXP. PROP. 0.4702 0.5298

ITEM 24: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 47 0.09 (0.31)

B(1) 48 -11.80 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002

-1.4 - 0.0 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002

0.2 - 1.6 0.002 0.002 0.002 0.002 0.002 0.001 0.001 0.001

1.8 - 3.0 0.001 0.001 0.001 0.001 0.001 0.001 0.001

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 24 76

OBS. PROP. 0.2400 0.7600

GROUP 1:

EXP. PROP. 0.2557 0.7443

GROUP 2:

EXP. PROP. 0.2550 0.7450

ITEM 25: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 49 0.91 (0.41)

B(1) 50 1.09 (0.61)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.019 0.023 0.027 0.032 0.038 0.044 0.052 0.061

-1.4 - 0.0 0.071 0.082 0.094 0.107 0.121 0.135 0.149 0.163

0.2 - 1.6 0.176 0.188 0.197 0.203 0.206 0.206 0.203 0.196

1.8 - 3.0 0.187 0.175 0.162 0.148 0.133 0.119 0.105

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 70 30

OBS. PROP. 0.7000 0.3000

GROUP 1:

EXP. PROP. 0.7065 0.2935

GROUP 2:

EXP. PROP. 0.6996 0.3004

ITEM 26: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 51 1.16 (0.37)

B(1) 52 -0.72 (0.29)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.083 0.101 0.122 0.147 0.174 0.203 0.233 0.263

-1.4 - 0.0 0.290 0.313 0.329 0.337 0.336 0.327 0.309 0.285

0.2 - 1.6 0.257 0.228 0.198 0.169 0.142 0.119 0.098 0.080

1.8 - 3.0 0.065 0.053 0.043 0.034 0.027 0.022 0.017

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 34 66

OBS. PROP. 0.3400 0.6600

GROUP 1:

EXP. PROP. 0.3498 0.6502

GROUP 2:

EXP. PROP. 0.3410 0.6590

ITEM 27: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 53 0.01 (0.34)

B(1) 54 19.01 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 56 44

OBS. PROP. 0.5600 0.4400

GROUP 1:

EXP. PROP. 0.5551 0.4449

GROUP 2:

EXP. PROP. 0.5550 0.4450

ITEM 28: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 55 0.26 (0.30)

B(1) 56 2.24 (2.87)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.011 0.011 0.012 0.012 0.012 0.013 0.013 0.013

-1.4 - 0.0 0.014 0.014 0.014 0.015 0.015 0.015 0.015 0.016

0.2 - 1.6 0.016 0.016 0.016 0.016 0.017 0.017 0.017 0.017

1.8 - 3.0 0.017 0.017 0.017 0.017 0.017 0.017 0.017

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 64 36

OBS. PROP. 0.6400 0.3600

GROUP 1:

EXP. PROP. 0.6425 0.3575

GROUP 2:

EXP. PROP. 0.6400 0.3600

ITEM 29: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 57 0.01 (0.29)

B(1) 58 26.74 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 54 46

OBS. PROP. 0.5400 0.4600

GROUP 1:

EXP. PROP. 0.5366 0.4634

GROUP 2:

EXP. PROP. 0.5365 0.4635

ITEM 30: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 59 0.00 (0.59)

B(1) 60 66.18 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 55 45

OBS. PROP. 0.5500 0.4500

GROUP 1:

EXP. PROP. 0.5492 0.4508

GROUP 2:

EXP. PROP. 0.5492 0.4508

ITEM 31: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 61 0.00 (0.23)

B(1) 62 -159.74 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 48 52

OBS. PROP. 0.4800 0.5200

GROUP 1:

EXP. PROP. 0.4804 0.5196

GROUP 2:

EXP. PROP. 0.4804 0.5196

ITEM 32: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 63 -0.60 (0.30)

B(1) 64 0.07 (0.45)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.043 0.046 0.050 0.054 0.059 0.063 0.067 0.071

-1.4 - 0.0 0.075 0.078 0.082 0.085 0.087 0.089 0.090 0.091

0.2 - 1.6 0.090 0.090 0.088 0.086 0.084 0.081 0.078 0.074

1.8 - 3.0 0.070 0.066 0.062 0.057 0.053 0.049 0.045

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 49 51

OBS. PROP. 0.4900 0.5100

GROUP 1:

EXP. PROP. 0.4840 0.5160

GROUP 2:

EXP. PROP. 0.4898 0.5102

ITEM 33: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 65 0.48 (0.28)

B(1) 66 -0.35 (0.62)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.039 0.041 0.043 0.045 0.047 0.049 0.051 0.052

-1.4 - 0.0 0.054 0.055 0.056 0.056 0.057 0.057 0.057 0.057

0.2 - 1.6 0.056 0.055 0.054 0.053 0.052 0.050 0.048 0.046

1.8 - 3.0 0.044 0.042 0.040 0.038 0.036 0.034 0.032

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 46 54

OBS. PROP. 0.4600 0.5400

GROUP 1:

EXP. PROP. 0.4650 0.5350

GROUP 2:

EXP. PROP. 0.4602 0.5398

ITEM 34: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 67 0.00 (0.37)

B(1) 68 -77.69 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 44 56

OBS. PROP. 0.4400 0.5600

GROUP 1:

EXP. PROP. 0.4405 0.5595

GROUP 2:

EXP. PROP. 0.4404 0.5596

ITEM 35: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 69 1.43 (0.49)

B(1) 70 -0.68 (0.24)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.069 0.090 0.116 0.148 0.187 0.232 0.284 0.339

-1.4 - 0.0 0.394 0.444 0.483 0.505 0.507 0.490 0.454 0.407

0.2 - 1.6 0.352 0.297 0.244 0.197 0.156 0.123 0.095 0.073

1.8 - 3.0 0.056 0.043 0.033 0.025 0.019 0.014 0.011

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 33 67

OBS. PROP. 0.3300 0.6700

GROUP 1:

EXP. PROP. 0.3410 0.6590

GROUP 2:

EXP. PROP. 0.3312 0.6688

ITEM 36: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 71 1.47 (0.45)

B(1) 72 0.08 (0.20)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.023 0.031 0.041 0.054 0.071 0.093 0.121 0.156

-1.4 - 0.0 0.199 0.249 0.305 0.365 0.425 0.477 0.517 0.536

0.2 - 1.6 0.534 0.509 0.466 0.411 0.351 0.291 0.236 0.188

1.8 - 3.0 0.147 0.114 0.088 0.067 0.051 0.038 0.029

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 52 48

OBS. PROP. 0.5200 0.4800

GROUP 1:

EXP. PROP. 0.5310 0.4690

GROUP 2:

EXP. PROP. 0.5200 0.4800

ITEM 37: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 73 0.00 (0.49)

B(1) 74 73.62 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 58 42

OBS. PROP. 0.5800 0.4200

GROUP 1:

EXP. PROP. 0.5797 0.4203

GROUP 2:

EXP. PROP. 0.5797 0.4203

ITEM 38: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 75 0.03 (0.34)

B(1) 76 -22.58 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 31 69

OBS. PROP. 0.3100 0.6900

GROUP 1:

EXP. PROP. 0.3157 0.6843

GROUP 2:

EXP. PROP. 0.3154 0.6846

ITEM 39: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 77 0.01 (0.42)

B(1) 78 -66.20 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 38 62

OBS. PROP. 0.3800 0.6200

GROUP 1:

EXP. PROP. 0.3807 0.6193

GROUP 2:

EXP. PROP. 0.3807 0.6193

ITEM 40: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 79 0.00 (0.37)

B(1) 80 -131.07 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

0.2 - 1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

1.8 - 3.0 0.000 0.000 0.000 0.000 0.000 0.000 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 37 63

OBS. PROP. 0.3700 0.6300

GROUP 1:

EXP. PROP. 0.3705 0.6295

GROUP 2:

EXP. PROP. 0.3704 0.6296

ITEM 41: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 81 0.19 (1.29)

B(1) 82 20.10 (****)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.001 0.001 0.001

-1.4 - 0.0 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001

0.2 - 1.6 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001

1.8 - 3.0 0.001 0.001 0.001 0.001 0.001 0.001 0.001

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 98 2

OBS. PROP. 0.9800 0.0200

GROUP 1:

EXP. PROP. 0.9794 0.0206

GROUP 2:

EXP. PROP. 0.9792 0.0208

ITEM 42: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 83 0.26 (0.28)

B(1) 84 0.47 (1.17)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.014 0.014 0.015 0.015 0.015 0.015 0.016 0.016

-1.4 - 0.0 0.016 0.016 0.016 0.017 0.017 0.017 0.017 0.017

0.2 - 1.6 0.017 0.017 0.017 0.017 0.017 0.017 0.017 0.017

1.8 - 3.0 0.017 0.016 0.016 0.016 0.016 0.016 0.015

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 53 47

OBS. PROP. 0.5300 0.4700

GROUP 1:

EXP. PROP. 0.5328 0.4672

GROUP 2:

EXP. PROP. 0.5301 0.4699

ITEM 43: GRP1, N[MU: -0.04 SIGMA: 1.00]

P(#);(S.E.): 85; (0.13) 87; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

-1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000

ITEM 44: GRP2, N[MU: 0.00 SIGMA: 1.00]

P(#);(S.E.): 89; (0.00) 90; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

-1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000

TOTAL TEST INFORMATION

FOR GROUP 1:

@THETA: INFORMATION:

-3.0 - -1.6 1.707 1.841 2.004 2.205 2.452 2.755 3.126 3.576

-1.4 - 0.0 4.115 4.745 5.455 6.208 6.930 7.506 7.807 7.744

0.2 - 1.6 7.326 6.657 5.876 5.101 4.403 3.809 3.318 2.920

1.8 - 3.0 2.598 2.339 2.129 1.958 1.818 1.702 1.605

@THETA: POSTERIOR STANDARD DEVIATION:

-3.0 - -1.6 0.765 0.737 0.706 0.673 0.639 0.603 0.566 0.529

-1.4 - 0.0 0.493 0.459 0.428 0.401 0.380 0.365 0.358 0.359

0.2 - 1.6 0.369 0.388 0.413 0.443 0.477 0.512 0.549 0.585

1.8 - 3.0 0.620 0.654 0.685 0.715 0.742 0.767 0.789

MARGINAL RELIABILITY: 0.8025

TOTAL TEST INFORMATION

FOR GROUP 2:

@THETA: INFORMATION:

-3.0 - -1.6 1.707 1.841 2.004 2.205 2.452 2.755 3.126 3.576

-1.4 - 0.0 4.115 4.745 5.455 6.208 6.930 7.506 7.807 7.744

0.2 - 1.6 7.326 6.657 5.876 5.101 4.403 3.809 3.318 2.920

1.8 - 3.0 2.598 2.339 2.129 1.958 1.818 1.702 1.605

@THETA: POSTERIOR STANDARD DEVIATION:

-3.0 - -1.6 0.765 0.737 0.706 0.673 0.639 0.603 0.566 0.529

-1.4 - 0.0 0.493 0.459 0.428 0.401 0.380 0.365 0.358 0.359

0.2 - 1.6 0.369 0.388 0.413 0.443 0.477 0.512 0.549 0.585

1.8 - 3.0 0.620 0.654 0.685 0.715 0.742 0.767 0.789

MARGINAL RELIABILITY: 0.8018

GROUP 1

NEGATIVE TWICE THE LOGLIKELIHOOD= 1709.1

(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

GROUP 2

NEGATIVE TWICE THE LOGLIKELIHOOD= 1707.0

(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

TOTAL, NEGATIVE TWICE THE LOGLIKELIHOOD, ALL GROUPS= 3416.1

NORMAL PROGRAM TERMINATION

START DATE: 03-26-2008

START TIME: 16:15:12

END TIME: 16:15:20

ตัวอย่างคำสั่ง LISREL และ OUTPUT LISREL
สำหรับรายการคำตอบของผู้สอบ 100 คน
ที่เป็นการตรวจให้คะแนนแบบสองค่า
ข้อสอบศึกษาข้อที่ 5

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

GROUP: Reference

!DA NI=21 NO=100 NG=2 MA=CM

SY='N100_R_1' NG=2

SE

2 4 3 5 6 7 8 9 10 12 13 14 15 17 18 19 20 22 23 24 25/

MO NX=21 NK=1

FR LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1

FR LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1

FR LX 19 1 LX 20 1 LX 21 1

VA 1 LX 1 1

OU

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

GROUP: Focal

!DA NO=200

SY='N200_F_1' NG=2

SE

2 4 3 5 6 7 8 9 10 12 13 14 15 17 18 19 20 22 23 24 25/

OU

DATE: 5/15/2008

TIME: 13:43

L I S R E L 8.54

BY

Karl G. J”reskog & Dag S”rbom

This program is published exclusively by

Scientific Software International, Inc.

7383 N. Lincoln Avenue, Suite 100

Lincolnwood, IL 60712, U.S.A.

Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140

Copyright by Scientific Software International, Inc., 1981-2002

Use of this program is subject to the terms specified in the

Universal Copyright Convention.

Website: www.ssicentral.com

The following lines were read from file

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

GROUP: Reference

!DA NI=21 NO=200 NG=2 MA=CM

SY='L200ab_R_1' NG=2

SE

2 4 3 5 6 7 8 9 10 12 13 14 15 17 18 19 20 22 23 24 25/

MO NX=21 NK=1

FR LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1

FR LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1

FR LX 19 1 LX 20 1 LX 21 1

VA 1 LX 1 1

OU

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

Number of Input Variables 26

Number of Y - Variables 0

Number of X - Variables 21

Number of ETA - Variables 0

Number of KSI - Variables 1

Number of Observations 200

Number of Groups 2

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

GROUP: Focal

!DA NO=200

SY='L200ab_F_1' NG=2

SE

2 4 3 5 6 7 8 9 10 12 13 14 15 17 18 19 20 22 23 24 25/

OU

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

Number of Input Variables 26

Number of Y - Variables 0

Number of X - Variables 21

Number of ETA - Variables 0

Number of KSI - Variables 1

Number of Observations 200

Number of Groups 2

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

Covariance Matrix

	I01	I03	I02	I04	I05	I06
I01	0.25					
I03	-0.09	0.18				
I02	-0.03	0.02	0.25			
I04	0.05	-0.06	-0.01	0.24		
I05	0.03	-0.02	-0.01	0.00	0.25	
I06	0.07	-0.04	-0.01	0.02	0.03	0.21
I07	-0.02	0.01	0.02	-0.02	-0.01	-0.03
I08	0.00	-0.03	-0.03	0.04	-0.01	0.00
I09	-0.07	0.05	0.03	-0.04	-0.01	-0.04
I11	-0.03	0.00	0.01	-0.02	0.01	0.00
I12	-0.03	0.02	0.00	-0.01	0.01	-0.02
I13	-0.04	0.02	0.01	-0.03	-0.01	-0.04
I14	0.08	-0.06	-0.02	0.05	0.01	0.05
I16	-0.06	0.03	-0.01	-0.04	-0.03	-0.04
I17	0.10	-0.07	-0.04	0.08	0.02	0.03
I18	0.07	-0.06	0.01	0.08	0.03	0.06
I19	0.03	-0.02	0.01	0.04	- -	-0.01
I21	-0.06	0.06	-0.01	-0.02	-0.02	-0.06
I22	-0.07	0.08	0.01	-0.07	-0.01	-0.05
I23	0.00	0.00	0.00	-0.01	0.00	0.00
I24	-0.02	0.01	0.00	-0.01	-0.02	0.00

Covariance Matrix

	I07	I08	I09	I11	I12	I13
I07	0.25					
I08	-0.01	0.25				
I09	0.03	-0.02	0.25			
I11	0.01	0.03	0.02	0.25		
I12	0.01	-0.03	-0.02	0.00	0.25	

I13	0.00	0.00	0.01	0.03	-0.01	0.25
I14	-0.01	-0.01	-0.02	-0.04	0.00	-0.02
I16	0.00	0.01	0.02	0.01	0.00	0.01
I17	-0.05	0.01	-0.08	-0.02	-0.03	-0.04
I18	-0.02	0.02	-0.06	0.00	0.03	-0.03
I19	- -	0.02	-0.02	0.01	-0.01	-0.02
I21	0.03	0.03	0.02	0.00	0.01	0.01
I22	0.01	-0.03	0.05	0.01	0.02	0.03
I23	0.01	-0.01	0.00	0.00	-0.01	0.00
I24	0.02	-0.01	0.00	-0.01	0.01	-0.01

Covariance Matrix

	I14	I16	I17	I18	I19	I21
I14	0.25					
I16	-0.04	0.24				
I17	0.06	-0.05	0.23			
I18	0.05	-0.05	0.06	0.24		
I19	0.01	0.01	0.01	0.01	0.25	
I21	-0.05	0.03	-0.03	-0.03	0.01	0.24
I22	-0.06	0.04	-0.07	-0.05	0.02	0.07
I23	-0.01	0.01	0.02	0.01	-0.03	0.01
I24	-0.02	0.01	-0.03	-0.01	-0.01	0.01

Covariance Matrix

	I22	I23	I24
I22	0.24		
I23	0.02	0.25	
I24	0.02	-0.01	0.04

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

Covariance Matrix

	I01	I03	I02	I04	I05	I06
I01	0.25					
I03	-0.10	0.20				

I02	0.00	0.00	0.25			
I04	0.07	-0.08	0.01	0.24		
I05	-0.01	0.02	-0.01	-0.02	0.25	
I06	0.07	-0.05	0.01	0.05	0.00	0.20
I07	-0.01	0.03	-0.02	-0.02	0.00	-0.03
I08	0.01	-0.04	0.00	0.02	0.01	0.01
I09	-0.08	0.04	0.00	-0.05	0.03	-0.05
I11	-0.02	0.00	0.00	-0.01	0.01	0.00
I12	-0.01	0.01	0.01	0.00	-0.01	-0.01
I13	-0.05	0.03	0.03	-0.04	0.00	-0.05
I14	0.04	-0.06	0.03	0.04	-0.02	0.04
I16	-0.03	0.02	-0.03	-0.03	0.00	-0.04
I17	0.08	-0.05	0.00	0.06	-0.02	0.06
I18	0.08	-0.06	0.00	0.04	0.00	0.04
I19	-0.01	-0.01	-0.02	-0.01	0.01	0.01
I21	-0.05	0.04	0.00	-0.04	0.00	-0.02
I22	-0.05	0.05	0.00	-0.08	0.02	-0.05
I23	0.00	-0.03	0.02	-0.01	0.01	-0.03
I24	-0.01	0.00	0.00	0.00	0.00	0.00

Covariance Matrix

	I07	I08	I09	I11	I12	I13
I07	0.25					
I08	-0.02	0.24				
I09	0.01	0.00	0.25			
I11	-0.01	0.01	0.00	0.23		
I12	0.01	0.01	0.02	0.02	0.25	
I13	0.01	-0.04	0.00	0.01	0.01	0.25
I14	0.00	0.02	-0.05	-0.01	-0.02	-0.03
I16	0.02	0.02	0.05	-0.01	0.01	0.01
I17	0.00	0.02	-0.05	-0.02	-0.01	-0.05
I18	0.01	0.03	-0.05	-0.01	0.01	-0.03
I19	0.00	0.04	0.00	-0.02	0.01	-0.02
I21	-0.01	-0.01	0.03	0.00	-0.04	0.04

I22	0.01	-0.04	0.05	0.00	-0.02	0.04
I23	-0.02	0.01	-0.01	-0.02	0.02	0.02
I24	0.00	0.00	0.00	0.00	0.00	0.00

Covariance Matrix

	I14	I16	I17	I18	I19	I21
I14	0.25					
I16	-0.05	0.25				
I17	0.06	-0.04	0.22			
I18	0.05	0.01	0.03	0.25		
I19	0.01	-0.02	-0.02	0.00	0.25	
I21	0.00	0.01	-0.03	-0.05	-0.03	0.23
I22	-0.03	0.01	-0.07	-0.03	0.01	0.03
I23	0.01	0.01	-0.01	-0.01	0.00	-0.03
I24	0.00	0.00	0.00	-0.01	0.01	0.01

Covariance Matrix

	I22	I23	I24
I22	0.22		
I23	-0.02	0.25	
I24	0.00	0.00	0.01

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

Parameter Specifications

LAMBDA-X EQUALS LAMBDA-X IN THE FOLLOWING GROUP

PHI

	KSI 1
21	

THETA-DELTA

	I01	I03	I02	I04	I05	I06
	22	23	24	25	26	27

THETA-DELTA

	I07	I08	I09	I11	I12	I13

-------- -------- -------- -------- -------- --------

28 29 30 31 32 33

THETA-DELTA

I14 I16 I17 I18 I19 I21

-------- -------- -------- -------- -------- --------

34 35 36 37 38 39

THETA-DELTA

I22 I23 I24

-------- -------- --------

40 41 42

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

Parameter Specifications

LAMBDA-X

KSI 1

I01 0

I03 1

I02 2

I04 3

I05 4

I06 5

I07 6

I08 7

I09 8

I11 9

I12 10

I13 11

I14 12

I16 13

I17 14

I18 15

I19 16

I21 17

I22 18

I23 19

I24 20

PHI

KSI 1

43

THETA-DELTA

I01 I03 I02 I04 I05 I06

--- --- --- --- --- ---

44 45 46 47 48 49

THETA-DELTA

I07 I08 I09 I11 I12 I13

--- --- --- --- --- ---

50 51 52 53 54 55

THETA-DELTA

I14 I16 I17 I18 I19 I21

--- --- --- --- --- ---

56 57 58 59 60 61

THETA-DELTA

I22 I23 I24

--- --- ---

62 63 64

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonR

Number of Iterations = 9

LISREL Estimates (Maximum Likelihood)

LAMBDA-X EQUALS LAMBDA-X IN THE FOLLOWING GROUP

PHI

KSI 1

0.11

-0.02

5.67

THETA-DELTA

I01 I03 I02 I04 I05 I06

0.14	0.11	0.25	0.19	0.25	0.18
(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
8.09	8.38	9.97	9.25	9.97	9.43

THETA-DELTA

I07	I08	I09	I11	I12	I13
0.24	0.24	0.21	0.25	0.25	0.24
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
9.93	9.93	9.49	9.95	9.96	9.79

THETA-DELTA

I14	I16	I17	I18	I19	I21
0.20	0.22	0.15	0.19	0.25	0.21
(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
9.48	9.78	8.83	9.34	9.97	9.67

THETA-DELTA

I22	I23	I24
0.17	0.25	0.04
(0.02)	(0.03)	(0.00)
9.18	9.97	9.95

Squared Multiple Correlations for X - Variables

I01	I03	I02	I04	I05	I06
0.45	0.41	0.00	0.24	0.00	0.19

Squared Multiple Correlations for X - Variables

I07	I08	I09	I11	I12	I13
0.02	0.02	0.18	0.01	0.01	0.08

Squared Multiple Correlations for X - Variables

I14	I16	I17	I18	I19	I21
0.18	0.08	0.33	0.22	0.00	0.12

Squared Multiple Correlations for X - Variables

I22	I23	I24
0.26	0.00	0.01

Group Goodness of Fit Statistics

Contribution to Chi-Square = 218.22

Percentage Contribution to Chi-Square = 51.93

Root Mean Square Residual (RMR) = 0.014

Standardized RMR = 0.063

Goodness of Fit Index (GFI) = 0.91

TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS C. GROUP NonF

Number of Iterations = 9

LISREL Estimates (Maximum Likelihood)

LAMBDA-X

	KSI 1
I01	1.00
I03	-0.79
	-0.08
	-9.65
I02	-0.09
	-0.09
	-1.08
I04	0.72
	-0.09
	8.01
I05	0.05
	-0.09
	0.59
I06	0.61
	-0.08
	7.32
I07	-0.20
	-0.09

-2.36

I08 0.21

-0.09

2.46

I09 -0.63

-0.09

-6.95

I11 -0.14

-0.08

-1.65

I12 -0.11

-0.09

-1.33

I13 -0.41

-0.09

-4.71

I14 0.62

-0.09

6.87

I16 -0.41

-0.09

-4.7

I17 0.81

-0.09

9.02

I18 0.67

-0.09

7.43

I19 0.08

-0.09

0.96

I21 -0.49

-0.09

-5.73

I22 -0.72

-0.09

-8.15

I23 0.02

-0.09

0.19

I24 -0.06

-0.03

-2.26

PHI

KSI 1

0.1

-0.02

5.58

THETA-DELTA

I01	I03	I02	I04	I05	I06
0.14	0.13	0.25	0.18	0.25	0.16
(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)
8.13	8.69	9.97	9.20	9.97	9.36

THETA-DELTA

I07	I08	I09	I11	I12	I13
0.24	0.23	0.20	0.23	0.25	0.23
(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
9.93	9.93	9.47	9.95	9.96	9.78

THETA-DELTA

I14	I16	I17	I18	I19	I21
0.21	0.24	0.16	0.21	0.25	0.21
(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
9.51	9.79	8.90	9.41	9.97	9.67

THETA-DELTA

I22	I23	I24
0.17	0.25	0.01
(0.02)	(0.03)	(0.00)
9.19	9.97	9.91

Squared Multiple Correlations for X - Variables

I01	I03	I02	I04	I05	I06
0.41	0.33	0.00	0.23	0.00	0.19

Squared Multiple Correlations for X - Variables

I07	I08	I09	I11	I12	I13
0.02	0.02	0.16	0.01	0.01	0.07

Squared Multiple Correlations for X - Variables

I14	I16	I17	I18	I19	I21
0.15	0.07	0.29	0.18	0.00	0.11

Squared Multiple Correlations for X - Variables

I22	I23	I24
0.23	0.00	0.02

Global Goodness of Fit Statistics

Degrees of Freedom = 398

Minimum Fit Function Chi-Square = 420.24 (P = 0.21)

Normal Theory Weighted Least Squares Chi-Square = 396.26 (P = 0.52)

Estimated Non-centrality Parameter (NCP) = 0.0

90 Percent Confidence Interval for NCP = (0.0 ; 48.76)

Minimum Fit Function Value = 1.06

Population Discrepancy Function Value (F0) = 0.0

90 Percent Confidence Interval for F0 = (0.0 ; 0.12)

Root Mean Square Error of Approximation (RMSEA) = 0.0

90 Percent Confidence Interval for RMSEA = (0.0 ; 0.025)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.50

Expected Cross-Validation Index (ECVI) = 1.32

90 Percent Confidence Interval for ECVI = (1.32 ; 1.44)

ECVI for Saturated Model = 1.16

ECVI for Independence Model = 4.16

Chi-Square for Independence Model with 420 Degrees of Freedom = 1615.42

Independence AIC = 1699.42

Model AIC = 524.26

Saturated AIC = 924.00

Independence CAIC = 1909.06

Model CAIC = 843.71

Saturated CAIC = 3230.06

Normed Fit Index (NFI) = 0.74

Non-Normed Fit Index (NNFI) = 0.98

Parsimony Normed Fit Index (PNFI) = 0.70

Comparative Fit Index (CFI) = 0.98

Incremental Fit Index (IFI) = 0.98

Relative Fit Index (RFI) = 0.73

Critical N (CN) = 442.87

Group Goodness of Fit Statistics

Contribution to Chi-Square = 202.02

Percentage Contribution to Chi-Square = 48.07

Root Mean Square Residual (RMR) = 0.014

Standardized RMR = 0.061

Goodness of Fit Index (GFI) = 0.92

DATE: 5/15/2008

TIME: 13:43

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ-สกุล	นายมิ่ง เทพครเมือง
วัน เดือน ปีเกิด	14 มิถุนายน 2512
สถานที่เกิด	จังหวัดเชียงราย
สถานที่อยู่ปัจจุบัน	323/435 ถนนสายไหม แขวงสายไหม เขตสายไหม กรุงเทพฯ 10220
สถานที่ทำงาน	โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร (ฝ่ายประถม)

ประวัติการศึกษา

พ.ศ. 2535	การศึกษาบัณฑิต (กศ.บ.) (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2539	การศึกษามหาบัณฑิต (กศ.ม.) (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2554	การศึกษาดุษฎีบัณฑิต (กศ.ด.) (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ