

ประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบไล่ตามทฤษฎีการตอบสนองข้อสอบ

จกมล บัวแก้ว

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

สาขาวิชาวิจัย วัฒนผลและสถิติการศึกษา

คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

สิงหาคม 2560

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมคุษฎีนิพนธ์และคณะกรรมการสอบคุษฎีนิพนธ์ ได้พิจารณา
คุษฎีนิพนธ์ของ จงกต บัวแก้ว ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปรัชญาคุษฎีบัณฑิต สาขาวิชาวิจัย วัฒนผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมคุษฎีนิพนธ์



.....อาจารย์ที่ปรึกษาหลัก
(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)



.....อาจารย์ที่ปรึกษาร่วม
(ดร.สมพงษ์ ปั่นหุ่น)

คณะกรรมการสอบคุษฎีนิพนธ์



.....ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ จัดกระโทก)



.....กรรมการ
(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)



..... กรรมการ
(ดร.สมพงษ์ ปั่นหุ่น)



..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ไชยฐ์ ศิริสวัสดิ์)

คณะศึกษาศาสตร์อนุมัติให้รับคุษฎีนิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปรัชญาคุษฎีบัณฑิต สาขาวิชาวิจัย วัฒนผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพา



..... คณบดีคณะศึกษาศาสตร์
(รองศาสตราจารย์ ดร.วิชิต สุรัตน์เรืองชัย)

วันที่.....๑.....เดือน.....สิงหาคม.....พ.ศ. 2560

กิตติกรรมประกาศ

ดุชนิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความสามารถจากศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม อาจารย์ที่ปรึกษาหลัก ดร.สมพงษ์ ปั่นหุ่น อาจารย์ที่ปรึกษาร่วมที่กรุณาให้คำปรึกษา แนะนำแนวทางที่ถูกต้องตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนและเอาใจใส่เสมอมา ผู้วิจัยรู้สึกซาบซึ้งเป็นอย่างยิ่งจึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังคกระโทก ประธานสอบปากเปล่า ผู้ช่วยศาสตราจารย์ ดร.เชษฐ ศิริสวัสดิ์ กรรมการสอบปากเปล่า และผู้ช่วยศาสตราจารย์ ดร.สุวีพร อนุศาสนนันท์ กรรมการสอบเค้าโครงดุชนิพนธ์ที่ได้กรุณาให้ข้อเสนอแนะในการปรับปรุงแก้ไข ทำให้ดุชนิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณผู้อำนวยการ รองผู้อำนวยการ และเจ้าหน้าที่ผู้ประสานงานสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่อนุเคราะห์ข้อมูลในการดำเนินการทำดุชนิพนธ์ และขอขอบคุณ Dr.Howard Wainer ที่กรุณาขอโปรแกรม SCORIGHT เพื่อใช้ในการศึกษาการประมาณค่าพารามิเตอร์ของข้อมูลที่มีอิทธิพลของเทสต์เลท

ขอขอบคุณคณะผู้บริหารมหาวิทยาลัยราชภัฏสงขลาที่อนุมัติทุนสนับสนุนการศึกษาคงทุนพัฒนาบุคลากรและบุคคลทั่วไป คณะครุศาสตร์ที่ให้การสนับสนุนการศึกษาต่อ และเจ้าหน้าที่ผู้ประสานงานดำเนินการระหว่างการศึกษาทำให้การศึกษาของผู้วิจัยสำเร็จลุล่วงไปด้วยดี

ขอขอบคุณคณาจารย์ และเพื่อน ๆ พี่ ๆ น้อง ๆ ภาควิชาวิจัยและจิตวิทยาประยุกต์ สำหรับการช่วยเหลือในขั้นตอนต่าง ๆ ระหว่างการดำเนินการวิจัยผู้วิจัยพบความรักและน้ำใจมากมายจากดุชนิพนธ์ฉบับนี้

สุดท้ายขอกราบขอบพระคุณคุณพ่อ คุณแม่ และครอบครัวที่ให้การสนับสนุนและเป็นกำลังใจแก่ผู้วิจัยเสมอมา

คุณค่าและประโยชน์ของดุชนิพนธ์ฉบับนี้ผู้วิจัยขอมอบเป็นกตัญญูกตเวทิตาแด่ บุพการีบูรพาจารย์และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบันที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษาและประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

57810104: สาขาวิชา: วิจัย วัดผลและสถิติการศึกษา; ปร.ด. (วิจัย วัดผลและสถิติการศึกษา)

คำสำคัญ: ทฤษฎีการตอบสนองข้อสอบ/ ทฤษฎีการตอบสนองเทสต์เลท/ แบบสอบเทสต์เลท

จกมล บัวแก้ว: ประสิทธิภาพของการเชื่อมโยงคะแนนสำหรับแบบสอบเทสต์เลทตามทฤษฎีการตอบสนองข้อสอบ (EFFECTIVENESS OF LINKING SCORE FOR TESTLETS-TEST BASED ON ITEM RESPONSE THEORY) กรรมการควบคุมคุชฎินิพนธ์: ไพรัตน์ วงษ์นาม, ค.ด., สมพงษ์ ปันนุณ, ค.ด. 152 หน้า. ปี พ.ศ. 2560.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบระหว่างการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) และการประมาณค่าตามโมเดลทฤษฎีการตอบสนองเทสต์เลท (TRT) 2) เปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามโมเดล IRT และโมเดล TRT และ 3) เปรียบเทียบประสิทธิผลการเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ (TS IRT, TS TRT) และการเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีออคิเปอร์เซ็นต์ไทล์ (OS IRT, OS TRT) ใช้คะแนนจากการสอบ O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,500 คน ในวิชาภาษาไทยและวิชาภาษาอังกฤษ จากนั้นวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เปรียบเทียบด้วยสถิติ Paired *t*-test และเปรียบเทียบประสิทธิผลการเชื่อมโยงจากค่า bias SEE และ RMSE

ผลการวิจัยสรุปได้ ดังนี้

1. ค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) ระหว่างวิธีการประมาณค่าตามโมเดล IRT และโมเดล TRT มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทั้งในวิชาภาษาไทยและวิชาภาษาอังกฤษ ส่วนพารามิเตอร์ความสามารถผู้สอบพบความสัมพันธ์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ในวิชาภาษาอังกฤษ
2. ความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามโมเดล IRT และโมเดล TRT แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ในวิชาภาษาอังกฤษ ส่วนวิชาภาษาไทยพบว่าค่าพารามิเตอร์ยากเท่านั้นที่แตกต่างอย่างมีนัยสำคัญที่ระดับ 0.05
3. ประสิทธิภาพของการเชื่อมโยงพบว่า ค่า bias น้อยที่สุดเมื่อเชื่อมโยงคะแนนด้วยวิธี OS TRT ค่า SEE น้อยที่สุดเมื่อเชื่อมโยงด้วยวิธี OS TRT และค่า RMSE น้อยที่สุดเมื่อเชื่อมโยงด้วยวิธี TS TRT

57810104: MAJOR: EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS;
Ph.D. (EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS)
KEYWORDS: ITEM RESPONSE THEORY/ TESTLET RESPONSE THEORY/ TESTLET-TEST
JONGKON BUAKAEW: EFFECTIVENESS OF LINKING SCORE FOR
TESTLETS-TEST BASED ON ITEM RESPONSE THEORY. DISSERTATION ADVISORY
COMMITTEE: PAIRATTANA WONGNAM, Ph.D., SOMPONG PANHOON, Ph.D. 152 P.
2017.

This study aimed to; 1) investigate the relationship between the test parameters (discrimination, difficulty, and guessing) and ability parameters by IRT and TRT, 2) compare the standard error of the test parameters and ability parameters by IRT and TRT, and 3) compare the effectiveness of linking of true scores (TS IRT, TS TRT) and linking of observed scores (OS IRT, OS TRT). This study employed the O-NET scores of 1,500 grade 9 students in two subjects; Thai Language and English Language. The data were analyzed for correlation by Pearson's correlation coefficient, Paired *t*-test, and the effectiveness of the linking by the bias SEE and RMSE.

Results were that;

1. The item parameters (discrimination, difficulty, and guessing) of both Thai and English languages by IRT and TRT were significantly correlated ($p < .01$), while the ability parameters were found significant correlation only in English language ($p < .01$).
2. The standard error of the item parameters (discrimination, difficulty, and guessing) and the ability parameters by IRT and TRT were significantly different only in English language ($p < .05$), while Thai language it was found significantly different only in difficulty parameter ($p < .05$).
3. The effectiveness of linking of both true and observed scores with TS IRT and TS TRT were found to be least bias with TS TRT, least SEE with OS TRT, and least RMSE with OS IRT.

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ญ
บทที่	
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
คำถามการวิจัย.....	7
วัตถุประสงค์ของการวิจัย.....	7
ขอบเขตการวิจัย.....	8
ประโยชน์ที่ได้รับจากการวิจัย.....	10
นิยามศัพท์เฉพาะ.....	10
กรอบแนวคิดในการวิจัย.....	12
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	14
ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบถามทดสอบ.....	14
ตอนที่ 2 แนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ	25
ตอนที่ 3 แนวคิดเกี่ยวกับประสิทธิผลของการเชื่อมโยงคะแนน.....	44
ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง.....	49
3 วิธีดำเนินการวิจัย.....	64
ประชากรและกลุ่มตัวอย่าง.....	64
เครื่องมือที่ใช้ในการวิจัย.....	66
การเก็บรวบรวมข้อมูล.....	66
การวิเคราะห์ข้อมูล.....	67

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล.....	77
ตอนที่ 1 ข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียน.	78
ตอนที่ 2 ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบ	80
ตอนที่ 3 ประสิทธิภาพของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบ.....	88
5 สรุป อภิปรายผล และข้อเสนอแนะ	94
สรุปผลการวิจัย	95
อภิปรายผลการวิจัย	97
ข้อเสนอแนะ	101
บรรณานุกรม.....	103
ภาคผนวก	111
ภาคผนวก ก.....	112
ภาคผนวก ข.....	114
ภาคผนวก ค	124
ภาคผนวก ง	137
ภาคผนวก จ.....	143
ประวัติย่อของผู้วิจัย	152

สารบัญตาราง

ตารางที่	หน้า	
2-1	รูปแบบการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy/ Linn และแนวคิดระดับ ความเหมือนของ Kolen and Brennan.....	29
2-2	สรุปรูปแบบการออกแบบการเก็บรวบรวมข้อมูล	32
2-3	สรุปประเด็นที่เกี่ยวข้องกับประสิทธิผลของการเชื่อมโยงคะแนน.....	60
3-1	จำนวนประชากรและกลุ่มตัวอย่าง.....	65
3-2	โครงสร้างและข้อสอบที่มีลักษณะทดสอบ.....	66
4-1	ข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียน.....	79
4-2	ความเชื่อมั่นของแบบสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ.....	80
4-3	สถิติพื้นฐานพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และ พารามิเตอร์ความสามารถของผู้สอบ	81
4-4	ค่าความแปรปรวนของอิทธิพลทดสอบจำแนกตามวิชา.....	82
4-5	ค่าสัมประสิทธิ์สหสัมพันธ์พารามิเตอร์ระหว่างการประมาณค่าตามโมเดลทฤษฎี การตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนองทดสอบ แบบ 3 พารามิเตอร์.....	83
4-6	ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ระหว่าง การประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และทฤษฎี การตอบสนองทดสอบแบบ 3 พารามิเตอร์ ในรายวิชาภาษาไทย.....	84
4-7	ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ระหว่าง การประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และทฤษฎี การตอบสนองทดสอบแบบ 3 พารามิเตอร์ ในรายวิชาภาษาอังกฤษ.....	84
4-8	คะแนนจริงภาษาไทย คะแนนจริงภาษาอังกฤษ คะแนนเชื่อมโยงด้วยวิธีโค้ง คุณลักษณะแบบสอบ.....	88
4-9	คะแนนสังเกตได้ภาษาไทย คะแนนสังเกตได้ภาษาอังกฤษ คะแนนเชื่อมโยงด้วย วิธีอควิเปอร์เซ็นต์ไทล์.....	90
4-10	ประสิทธิผลของการเชื่อมโยงคะแนน.....	91

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
ภาคผนวก ข-1 ค่าพารามิเตอร์ข้อสอบวิชาภาษาไทย	115
ภาคผนวก ข-2 ค่าพารามิเตอร์ข้อสอบวิชาภาษาอังกฤษ	119
ภาคผนวก ค-1 คะแนนดิบ พารามิเตอร์ความสามารถ คะแนนจริง วิชาภาษาไทย	125
ภาคผนวก ค-2 พารามิเตอร์ความสามารถแต่ละเทสต์เลขท วิชาภาษาไทย	128
ภาคผนวก ค-3 คะแนนดิบ พารามิเตอร์ความสามารถ คะแนนจริง วิชาภาษาอังกฤษ	131
ภาคผนวก ค-4 พารามิเตอร์ความสามารถแต่ละเทสต์เลขท วิชาภาษาอังกฤษ	134
ภาคผนวก ง-1 ผลการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ.....	138
ภาคผนวก จ-1 ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ที่ประมาณ ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (IRT 3PL)	144
ภาคผนวก จ-2 ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ที่ประมาณ ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองเทสต์เลขทแบบ 3 พารามิเตอร์ (TRT 3PL)	148

สารบัญภาพ

ภาพที่	หน้า
1-1 กรอบแนวคิดในการวิจัย	13
2-1 รูปแบบการเชื่อมโยงคะแนนและวัตถุประสงค์ในการใช้	26
2-2 ระดับความคล้ายคลึงกันของการเชื่อมโยงคะแนน	28
3-1 ขั้นตอนการดำเนินการวิเคราะห์ข้อมูล.....	76
4-1 สารสนเทศของแบบสอบภาษาไทยตามโมเดลทฤษฎีการตอบสนองข้อสอบ	86
4-2 สารสนเทศของแบบสอบภาษาไทยตามโมเดลทฤษฎีการตอบสนองทดสอบทดสอบ.....	86
4-3 สารสนเทศของแบบสอบภาษาอังกฤษตามโมเดลทฤษฎีการตอบสนองข้อสอบ	87
4-4 สารสนเทศของแบบสอบภาษาอังกฤษตามโมเดลทฤษฎีการตอบสนองทดสอบ.....	87
4-5 เปรียบเทียบประสิทธิผลของการเชื่อมโยงคะแนน	92
ภาคผนวก ง-1 ความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยง ที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ	146
ภาคผนวก ง-2 ความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยง ที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองทดสอบ.....	147

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การศึกษาไทยตั้งแต่อดีตถึงปัจจุบันจะให้ความสำคัญกับคะแนนเป็นอย่างมาก คะแนนเป็นเครื่องมือช่วยในการตัดสินผู้เรียนว่าสอบผ่านหรือไม่ผ่าน หรือตัดสินว่าได้ศึกษาต่อในระดับที่สูงขึ้นหรือไม่ หรือนักเรียนมีพัฒนาการมากน้อยเพียงใด นอกจากการนำคะแนนมาประเมินด้านการศึกษาแล้วยังมีการนำคะแนนมาใช้ในการประเมินหน่วยงานที่เกี่ยวข้องกับการศึกษา หรือนำคะแนนมาใช้ในการประเมินบุคคลเพื่อเข้ารับการทำงานอีกด้วย ดังนั้นคะแนนที่ได้จะต้องวัดได้ตรงกับสิ่งที่ต้องการวัดหรือวัตถุประสงค์ของการวัดมากที่สุด ซึ่งคะแนนที่ได้จะมีความตรงนั้นส่วนหนึ่งมาจากเครื่องมือที่ใช้ในการวัดผล และกระบวนการเก็บรวบรวมข้อมูลที่มีคุณภาพและถูกต้องเหมาะสมมากที่สุด

การทดสอบมาตรฐานในปัจจุบันมีกระบวนการที่มีความหลากหลายมาก เช่น การทดสอบด้วยดินสอดและกระดาษ (Paper-pencil) ที่มีการสอบเป็นมาตรฐานทั้งระดับนานาชาติ (PISA) และระดับชาติ (O-NET) การทดสอบที่นำคอมพิวเตอร์มาใช้หรือการทดสอบแบบปรับเหมาะ เช่น การสอบวัดระดับภาษาอังกฤษ การสอบใบอนุญาตขับขี่รถยนต์หรือจักรยานยนต์ ซึ่งการสอบในแต่ละครั้งแบบสอบที่ใช้ก็จะต่างฉบับกัน แต่มีโครงสร้างของเนื้อหาที่เหมือนกัน อาจจะต่างกันบ้างในกรณีของความยากของแบบสอบ และในปัจจุบันแบบสอบที่นำมาใช้ในการทดสอบมีความหลากหลายมากขึ้นเพื่อเพิ่มความซับซ้อนหรือความยากของข้อสอบ เช่น แบบสอบหนึ่งฉบับมีทั้งข้อสอบที่เป็นรูปแบบอัตนัยและปรนัย หรือในหนึ่งฉบับข้อสอบแต่ละข้อจะมีรูปแบบการให้คะแนนไม่เหมือนกัน ดังนั้นผลการสอบแต่ละครั้งจะมีความไม่เท่าเทียมกัน ไม่สามารถนำมาเปรียบเทียบกันได้ ดังนั้นนักวัดผลได้ศึกษาและเสนอเกี่ยวกับหลักการปรับเทียบคะแนนระหว่างแบบสอบ มาช่วยในการปรับเทียบคะแนนระหว่างแบบสอบให้เท่าเทียมกัน เพื่อนำคะแนนจากแบบสอบทั้งสองฉบับมาเปรียบเทียบกันได้ แต่วิธีการนี้จะไม่เหมาะสมเมื่อเรานำมาใช้กับแบบสอบสองฉบับที่มีโครงสร้างเนื้อหา หรือลักษณะของแบบสอบที่ต่างกัน จึงมีวิธีการปรับคะแนนให้เท่าเทียมกันที่เรียกว่า การเชื่อมโยงคะแนน (Linking) (Kolen & Brennan, 2014, p. 486)

การเชื่อมโยงคะแนนนั้นจะมีรูปแบบต่าง ๆ ตามลักษณะโครงสร้างของแบบสอบสองฉบับที่นำมาเชื่อมโยงกัน สรุปได้ 4 รูปแบบ คือ 1) การปรับเทียบคะแนน (Equating) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างและเนื้อหาเดียวกัน มีความยาก ความเชื่อมั่น

คล้ายคลึงกัน คะแนนเชื่อมโยงที่ได้สามารถนำมาใช้แทนกันได้ 2) การปรับให้เป็นคะแนนมาตรฐาน (Calibration) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบทดสอบสองแบบสอบที่มีโครงสร้างหลักเหมือนกัน แต่ต่างกันในเรื่องคุณลักษณะแบบสอบ (เช่น ความยาวข้อสอบ) และต่างกันในเรื่องคุณลักษณะทางสถิติ 3) การทำนาย (Predicting) หรือการฉายภาพ (Projection) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบทดสอบสองแบบสอบที่มีคุณลักษณะต่างกัน เป็นการทำนายหรือฉายภาพคะแนนจากแบบสอบหนึ่งไปยังแบบสอบอื่น โดยผลที่ได้จากการเชื่อมโยงจะไม่สมมาตรกัน กล่าวคือ ความสัมพันธ์ที่ได้ระหว่างคะแนนแบบสอบ A กับแบบสอบ B ไม่เท่ากับความสัมพันธ์ระหว่างคะแนนจากแบบสอบ B กับแบบสอบ A นอกจากนี้วิธีการนี้ใช้กับการออกแบบในการเก็บข้อมูลกับผู้สอบกลุ่มเดียวและไม่คำนึงคุณลักษณะของแบบสอบว่าต้องเหมือนกัน และ 4) การปรับด้วยค่าสถิติ (Statistical moderation) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบที่มีคุณลักษณะต่างกันและกลุ่มผู้สอบไม่เท่าเทียมกัน วิธีนี้ถือว่าเป็นวิธีการเชื่อมโยงคะแนนที่อ่อนแอที่สุด (Feuer, Holland, Green, Bertenthal, & Hemphill, 1999, pp. 15-23; Holland, 2007, pp. 5-30; Kolen & Brennan, 2014, pp. 498-500) จากรูปแบบการเชื่อมโยงคะแนนทั้งสี่รูปแบบจะมีลักษณะเฉพาะตามโครงสร้างของแบบสอบในแต่ละฉบับที่จะนำมาใช้ในการเชื่อมโยง ซึ่งการที่ผู้ใช้จะเลือกรูปแบบการเชื่อมโยงใดนั้นขึ้นอยู่กับวิธีการออกแบบการเชื่อมโยงคะแนน

การออกแบบการเชื่อมโยงคะแนนมีหลักการออกแบบเหมือนการปรับเทียบคะแนนงานวิจัยบางเล่มมีการใช้คำสองคำนี้ทดแทนกัน ซึ่งการออกแบบแบ่งออกเป็น 2 ประเด็นหลัก คือ 1) การออกแบบการเก็บรวบรวมข้อมูล ซึ่งการออกแบบการเก็บรวบรวมข้อมูลนั้นมีวิธีการที่หลากหลายขึ้นอยู่กับลักษณะของกลุ่มผู้สอบและแบบสอบ เช่น ผู้สอบกลุ่มเดียวทำแบบสอบสองฉบับที่เหมือนกันหรือแบบสอบสองฉบับต่างกัน ผู้สอบสองกลุ่มเท่าเทียมกันทำแบบสอบหนึ่งฉบับต่างกันหรือทำแบบสอบสองฉบับเหมือนกันแต่ลำดับลำดับที่การทำหรือทำแบบสอบฉบับเดียวแต่มีข้อสอบร่วม กลุ่มผู้สอบสองกลุ่มไม่เท่าเทียมกันทำแบบสอบฉบับเดียวแต่มีข้อสอบร่วมภายในหรือแบบสอบร่วมภายนอก เป็นต้น 2) การออกแบบการเชื่อมโยงคะแนน ซึ่งการออกแบบการเชื่อมโยงคะแนนนั้นจะยึดตามหลักการทฤษฎีการวัด แบ่งเป็น 2 หลักการ คือ การเชื่อมโยงคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม และการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ ซึ่งการเชื่อมโยงคะแนนที่ยึดตามทฤษฎีการตอบสนองข้อสอบมีข้อได้เปรียบในทางปฏิบัติมากกว่าตามทฤษฎีการทดสอบแบบดั้งเดิมดังนี้ 1) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบให้ผลการเชื่อมโยงคะแนนที่ดีกว่าการเชื่อมโยงคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม กล่าวคือคะแนนเต็มของมาตรวัดมักจะมีส่วนนำมาใช้ในการการตัดสินใจเป็นสำคัญ ในขณะที่การเชื่อมโยงคะแนนด้วยทฤษฎีการทดสอบแบบดั้งเดิมนั้นสามารถใช้ได้เฉพาะข้อมูลคะแนนที่ได้รับจริงของ

ผู้สอบเท่านั้น 2) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบจะมีความยืดหยุ่นมากกว่า เพราะสามารถเชื่อมโยงคะแนนระหว่างแบบสอบให้มีสเกลเดียวกัน แบบสอบใหม่สามารถปรับได้ทั้งบางส่วนหรือทั้งหมดของแบบสอบเก่า 3) การเชื่อมโยงทำได้ง่าย ถ้าข้อสอบถูกตัดทิ้ง หรือ จำนวนข้อสอบน้อยลงสามารถปรับได้ง่ายบนพื้นฐานของสารสนเทศข้อสอบจากข้อสอบที่มีอยู่ และ 4) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบของการปรับเทียบก่อน (Pre-equating) สามารถทำได้ โดยที่วิธีการ Pre-equating เป็นการเตรียมความพร้อมตารางแปลงคะแนนดิบไปเป็นคะแนนมาตรฐานก่อนที่มีการทำข้อสอบ สามารถทำได้เมื่อข้อมูลข้อสอบรายข้อเป็นตัวแทนที่มีอยู่และสามารถปรับสเกลได้ พารามิเตอร์จะไม่แปรเปลี่ยนเมื่อถูกประยุกต์ใช้กับกลุ่มใหม่ นอกจากนี้แบบสอบที่ใช้ในการเชื่อมโยงคะแนนสามารถใช้ได้ทั้งลักษณะที่เป็นแบบสอบคู่ขนาน และแบบสอบไม่คู่ขนาน จากข้อได้เปรียบทั้งสี่ประการดังกล่าวสามารถสรุปได้ว่าผลที่ได้จากการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบจะมีความเสมอภาค (Equity) ความสมมาตร (Symmetry) และความไม่ผันแปรตามกลุ่ม (Invariance) (Cook & Eignor, 1983; Kolen, 1981 cited in Tain, 2011, p.10)

การออกแบบการเชื่อมโยงตามทฤษฎีการตอบสนองข้อสอบมีวิธีการเชื่อมโยงคะแนน 2 วิธี คือ การเชื่อมโยงคะแนนจริง (Linking true score) การเชื่อมโยงคะแนนสังเกตได้ (Linking observed score) ซึ่งทั้งสองวิธีนี้จะมีการแปลงคะแนนให้สมมูลกันระหว่างแบบสอบสองฉบับที่แตกต่างกัน เช่น การเชื่อมโยงคะแนนจริงสามารถแปลงคะแนนด้วยวิธีค่าเฉลี่ย/ ค่าเฉลี่ย ค่าเฉลี่ย/ ส่วนเบี่ยงเบน โค้งคุณลักษณะของแบบสอบ กระบวนการปรับเทียบด้วยคะแนนจริง วิธีการของ Newton-Raphson การเชื่อมโยงคะแนนสังเกตได้สามารถแปลงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ วิธีค่าเฉลี่ย วิธีเชิงเส้นตรง วิธีเชิงเส้นคู่ขนาน (Kolen & Brennan, 2014, pp. 198-214, 501-505) แต่จากการศึกษาที่ผ่านมาพบว่าวิธีการแปลงคะแนนที่นิยมใช้สำหรับการเชื่อมโยงคะแนนจริงคือวิธีโค้งคุณลักษณะแบบสอบ (อดิศร ศรีบุญวงษ์, 2545; อัญชลี ศรีภักชญา, 2552; จิระนาฏ จิวพัฒน์, บัณฑิตา อินสมบัติ และธีรยุทธ ภูเขา, 2554; Li, 2009; Rui, Shu-Liang, & Deng-Wen, 2010) และวิธีการแปลงคะแนนที่นิยมใช้สำหรับการเชื่อมโยงคะแนนสังเกตได้คือวิธีอิกวิเปอร์เซ็นต์ไทล์ (Zhang, 2007; He, Li, & Wolfe, 2012; Cao, Lu, & Tao, 2014; Tao & Cao, 2016)

การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบนั้น เริ่มจากการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถผู้สอบ โดยวิธีการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบนั้นจะขึ้นอยู่กับลักษณะของการตรวจให้คะแนน คือ แบบสอบมีการตรวจให้คะแนนสองค่า การประมาณค่าพารามิเตอร์ที่รู้จักกันมี 3 รูปแบบ คือ โมเดล 1 พารามิเตอร์ (ความยาก) โมเดล 2 พารามิเตอร์ (ความยาก อำนาจจำแนก) และโมเดล 3 พารามิเตอร์

(ความยาก อำนาจจำแนก การเดา) และแบบสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า การประมาณค่าพารามิเตอร์ที่รู้จักกัน คือ การให้คะแนนความรู้บางส่วน การตรวจให้คะแนนตามลำดับขั้นของหลายคำตอบ เป็นต้น แต่การจะใช้วิธีการประมาณค่าพารามิเตอร์ความสามารถผู้สอบ และพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ ข้อมูลหรือคะแนนที่จากแบบสอบจะต้องอยู่ภายใต้ข้อตกลงเบื้องต้น 2 ประการคือ แบบสอบมุ่งวัดคุณลักษณะเดียว (Unidimensionality) และข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน (Independence)

แบบสอบมาตรฐานในปัจจุบันถึงแม้ว่าในแบบสอบแต่ละฉบับจะมุ่งวัดคุณลักษณะเดียว แต่จะขาดความเป็นอิสระในการตอบข้อสอบแต่ละข้อ อันเนื่องมาจากมีการนำข้อสอบแบบเลือกตอบมาประยุกต์ใช้ในการวัดทักษะความสามารถที่มีความซับซ้อนมากขึ้น เช่น ทักษะการให้เหตุผลเชิงตรรกะ การให้เหตุผลเชิงวิเคราะห์ ทำให้ข้อสอบเดี่ยว (Single item) ไม่สามารถวัดทักษะที่ซับซ้อนได้อย่างครอบคลุม จึงมีแนวคิดเกี่ยวกับการจัดกลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน ข้อสอบลักษณะนี้จึงถูกมองว่าเป็นกลุ่มของข้อสอบ ซึ่งรวมอยู่ภายในแบบสอบฉบับเต็ม เรียกข้อสอบลักษณะนี้ว่า เทสต์เลท (Testlet) เช่น แบบสอบคณิตศาสตร์ แบบสอบวิทยาศาสตร์ แบบสอบภาษาอังกฤษ แบบสอบภาษาไทย ซึ่งลักษณะแบบสอบด้านคำนวณหรือวิทยาศาสตร์จะเป็นลักษณะของการนำผลจากข้อแรกเป็นฐานของการตอบโจทย์ข้อถัดไป หรือการอ่านค่าจากตารางกราฟ ผลจากการทดลอง เป็นต้น ส่วนด้านภาษานั้นส่วนใหญ่จะเน้นเกี่ยวกับการอ่านเพื่อทำความเข้าใจ โดยให้บทความหรือข้อความมาแล้วใช้ตอบคำถาม 3-5 ข้อ หรือบางครั้ง 6-12 ข้อ ขึ้นอยู่กับระดับความสามารถของผู้สอบ โดยแบบสอบเทสต์เลทมีทั้งที่เป็นปรนัยและอัตนัย จึงทำให้การประเมินทางการศึกษาในทางปฏิบัตินั้นมีการลดหรือขยายโครงสร้างการตอบสนองเทสต์เลทมาใช้มากขึ้นและนำมาใช้ในการสร้างแบบสอบมาตรฐานแบบดั้งเดิม โดยส่วนใหญ่มีรูปแบบข้อสอบแบบหลายตัวเลือก (Multiple choice: MC) ซึ่งมีประโยชน์มากในการประเมินกับคนกลุ่มขนาดใหญ่ (Large scale assessment) ได้แก่ 1) ผู้ประเมินไม่มีอคติ เมื่อตรวจข้อสอบแบบหลายตัวเลือก หรือที่เรียกว่ามีความเป็นปรนัย 2) การทดสอบหนึ่งครั้งสามารถใช้ข้อสอบแบบเลือกตอบได้หลายข้อ และครอบคลุมเนื้อหา 3) การให้คะแนนมีความเป็นปรนัย และจำนวนข้อสอบที่มากเป็นการเพิ่มความเชื่อมั่นและความตรงของแบบสอบ และ 4) การให้คะแนนของข้อสอบแบบหลายตัวเลือกสามารถทำได้ง่ายและประหยัดเวลา (Chen, 2014, p.11) แต่การขาดความเป็นอิสระในการตอบข้อสอบแต่ละข้อก็จะส่งผลต่อการตอบข้อสอบถูกของผู้สอบ และยังส่งผลกระทบต่อคุณภาพของแบบสอบหรือข้อสอบคลาดเคลื่อนไปจากความเป็นจริง

Sereci, Thissen, and Wainer (1991, pp 237-247) ได้ศึกษาเกี่ยวกับแบบสอบที่มีอิทธิพลของเทสต์เลท แล้วพบว่า การละเลยความเป็นอิสระระหว่างข้อสอบ ทำให้การประมาณค่า

ความเที่ยงจะเกินจริงประมาณ 10-15 % จึงทำให้นักวิจัยสนใจในประเด็นนี้มากขึ้นจนสรุปเป็นแนวทางในการวิเคราะห์ที่มีอิทธิพลของเทสต์เลท จำนวน 3 แนวทาง คือ 1) การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ เป็นรูปแบบทั่วไปที่นิยมใช้มากในโมเดลทฤษฎีการตอบสนองข้อสอบที่มีการวัด 2 ค่าและวัดในมิติเดียว ซึ่งเหมาะสมกับข้อมูลที่มาจากการทดสอบกับคนจำนวนมาก และแบบสอบเป็นแบบหลายตัวเลือก ข้อมูลที่มาจากแบบสอบที่มีเทสต์เลทนี้มีความไม่อิสระกันของข้อสอบแต่ผลของเทสต์เลทจะถูกละเว้น ดังนั้นข้อสอบจะพิจารณาให้เป็นอิสระจากกันในแต่ละข้อ 2) การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบ Graded response model เป็นการรวมข้อสอบที่ตรวจให้คะแนนแบบสองค่าในเทสต์เลทให้เป็นข้อสอบหนึ่งข้อที่มีการตรวจให้คะแนนมากกว่าหนึ่งค่า ซึ่งการทำลักษณะเช่นนี้จะมีข้อจำกัดเกี่ยวกับสารสนเทศของข้อสอบที่เป็นรายข้อจะไม่สามารถวิเคราะห์ได้ และ 3) การวิเคราะห์ตามทฤษฎีการตอบสนองของเทสต์เลท เป็นโมเดลที่วิเคราะห์โดย การนำอิทธิพลของเทสต์เลทเข้ามาเกี่ยวข้องอย่างชัดเจน (Chen, 2014, pp. 17-22) จากทั้งสามวิธีพบว่า การใช้คะแนนเทสต์เลทหรือการประยุกต์ใช้โมเดลการตรวจให้คะแนนมากกว่าสองค่าแบบมิติเดียว ข้อมูลจากแบบสอบที่มีเทสต์เลทดูเหมือนว่าจะไม่มีปัญหาความไม่เป็นอิสระกันของข้อสอบ แต่การนำ การวิเคราะห์วิธีนี้รูปแบบของคะแนนภายในแต่ละเทสต์เลทถูกละเลย และสารสนเทศภายในเทสต์เลทก็หายไป ดังนั้นวิธีการนี้จึงไม่เป็นที่นิยมนำมาใช้กันในทางปฏิบัติ (Lee, Kolen, Frisbie, & Ankenmann, 2001; Zhang, 2007; Wainer & Wang, 2000; Wainer, Bradlow, & Du, 2000; Wainer, Bradlow, & Wang, 2007)

การออกแบบการเชื่อมโยงคะแนน งานวิจัยส่วนใหญ่จะใช้วิธีการจำลองข้อมูล มีกำหนดสถานการณ์หรือเงื่อนไขไว้ล่วงหน้าแล้วเปรียบเทียบประสิทธิภาพของการเชื่อมโยงคะแนน ซึ่งผู้วิจัยมีความเห็นว่าถ้าข้อมูลที่ได้มาจากสถานการณ์หรือจากการเก็บข้อมูลจริง ประสิทธิภาพของการเชื่อมโยงคะแนนจะเป็นอย่างไร จึงทำการสำรวจแบบสอบถามมาตรฐานระดับชาติขั้นพื้นฐาน (O-NET) ว่ามีลักษณะแบบสอบเทสต์เลทหรือไม่ ผลปรากฏว่าแบบสอบในระดับชั้นมัธยมศึกษาปีที่ 3 มีแบบสอบวิชาภาษาไทยและแบบสอบภาษาอังกฤษที่มีลักษณะดังกล่าว อีกทั้งนักภาษาศาสตร์มองว่าภาษาไทยและภาษาอังกฤษมีความเหมือนกัน กล่าวคือลักษณะทั่วไปของภาษาเหมือนกัน แบ่งการศึกษาเป็นระดับต่าง ๆ ได้แก่ ระดับเสียง ระดับคำ ระดับวากยสัมพันธ์ และระดับความหมาย (ภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, ม.ป.ป.) และการศึกษาของ อรทัย เจริญสิทธิ์ (2559, หน้า 109) พบความสัมพันธ์ของคะแนน O-NET ระหว่างวิชาภาษาไทยและวิชาภาษาอังกฤษในขนาดปานกลางและมีนัยสำคัญทางสถิติที่ระดับ 0.01 ส่วนนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 3 นั้นเป็นช่วงขั้นสุดท้ายของการศึกษาภาคบังคับและเป็นวัยที่ทุกคนให้ความสำคัญสำหรับเตรียมความพร้อมด้านการศึกษาต่อสายสามัญหรือสายอาชีวะ

เพื่อประกอบอาชีพในอนาคต ดังนั้นผู้วิจัยจึงนำผลการสอบของนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 3 ทั้งสองวิชามาใช้ในการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ โดยเลือกวิชาภาษาไทย เป็นฐานการเชื่อมโยงเนื่องจากวิชาภาษาไทยเป็นวิชาพื้นฐานที่นักเรียนทุกคนต้องเรียน อีกทั้งยังเป็นภาษากลางที่นักเรียนทุกคนสามารถสื่อสารได้เข้าใจตรงกันและมีความเท่าเทียมกันในการเรียนรู้

การตรวจสอบประสิทธิผลของการเชื่อมโยงคะแนน ซึ่งการประเมินประสิทธิผลของการเชื่อมโยงคะแนนนั้นพิจารณาได้จากค่าต่าง ๆ เช่น ความเสมอภาค ความเพียงพอ การปรับเทียบกลับสู่แบบสอบเดิม ดัชนีความแตกต่าง ความคลาดเคลื่อนของการเชื่อมโยง เป็นต้น โดยการหาประสิทธิผลที่ได้รับความนิยมในปัจจุบันคือ ความคลาดเคลื่อนของการเชื่อมโยงเป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างสุ่มและเป็นระบบ และค่าที่คำนวณได้นั้นจะต้องมีค่าต่ำ ๆ ถึงจะทำให้การเชื่อมโยงมีประสิทธิผล พิจารณาจาก 3 ค่า คือ ความลำเอียง (Bias) ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) และดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error) (Zhang, 2007; Li, 2009; Rui et al., 2010; He et al., 2012; Cao et al., 2014; Tao & Cao, 2016)

จากความเป็นมาที่กล่าวมาข้างต้นการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบมีส่วนช่วยให้คะแนนระหว่างแบบสอบที่มีความเป็นคู่ขนานกัน หรือไม่คู่ขนานกันมีความเท่าเทียมกัน แม้ว่าแบบสอบที่ใช้มีขนาดคุณลักษณะความเป็นอิสระของข้อสอบ หรือแบบสอบมีลักษณะของเทสต์เลทจึงมีผลต่อความสามารถในการตอบถูกของผู้สอบก็ตาม แต่เรามีแนวทางในการประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบได้ 2 แนวทางที่เป็นที่นิยมนำมาใช้ในทางปฏิบัติคือ การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองของ เทสต์เลท แบบ 3 พารามิเตอร์ (Testlet Response Theory: TRT 3 PL) และตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (Item Response Theory: IRT 3 PL) นอกจากนี้ การออกแบบการเชื่อมโยงคะแนนนั้นผู้วิจัยได้ออกแบบเก็บข้อมูลกับคนกลุ่มเดียวทำแบบสอบสองฉบับต่างกัน ซึ่งสอดคล้องกับรูปแบบการเชื่อมโยงคะแนนแบบการทำนาย และยังได้ออกแบบวิธีการเชื่อมโยงคะแนนสองวิธีคือ การเชื่อมโยงคะแนนจริง (Linking true score) ด้วยวิธีโค้งคุณลักษณะแบบสอบและการเชื่อมโยงคะแนนสังเกตได้ (Linking observed score) ด้วยวิธีอควิเปอร์เซ็นต์ไทล์ และตรวจสอบประสิทธิผลของการเชื่อมโยงคะแนนจากค่าความคลาดเคลื่อนของการเชื่อมโยงคะแนนประกอบด้วย ความลำเอียง (Bias) ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard Error of Equating: SEE) และดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ดังนั้นผู้วิจัยสนใจศึกษาประสิทธิผลของ

การเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบตามทฤษฎีการตอบสนองข้อสอบ เมื่อวิธีการเชื่อมโยงคะแนนต่างกัน 2 วิธีคือ วิธีการเชื่อมโยงคะแนนจริงและวิธีการเชื่อมโยงคะแนนสังเกตได้ และการประมาณค่าพารามิเตอร์ข้อสอบ 2 วิธี คือ การประมาณค่าพารามิเตอร์ข้อสอบตามโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์ (Testlet Response Theory: TRT 3 PL) และตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Item Response Theory: IRT 3 PL) ซึ่งผลที่ได้จากการศึกษาครั้งนี้เป็นการขยายองค์ความรู้เกี่ยวกับการประมาณค่าพารามิเตอร์ข้อสอบเมื่อมีอิทธิพลของทดสอบ และได้นำวิธีการเชื่อมโยงคะแนนเมื่อมีแบบสอบทดสอบสำหรับนักวัดผล นักวิชาการ หรือผู้ที่เกี่ยวข้องกับการศึกษานำไปใช้ต่อยอดองค์ความรู้หรือพัฒนาแนวทางการสร้างและหาคุณภาพเครื่องมือให้สอดคล้องกับลักษณะของแบบสอบที่มีลักษณะทดสอบแบบทดสอบรวมทั้งวิธีการเชื่อมโยงคะแนนที่เหมาะสมเพื่อให้เกิดประโยชน์แก่ผู้เรียนมากที่สุด

คำถามการวิจัย

1. การประมาณค่าพารามิเตอร์ของข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบตามโมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองทดสอบแบบให้ค่าพารามิเตอร์เหมือนหรือต่างกันอย่างไร
2. ประสิทธิภาพของการเชื่อมโยงคะแนนเหมือนหรือต่างกันอย่างไร เมื่อวิธีการเชื่อมโยงคะแนนต่างกัน (วิธีการเชื่อมโยงคะแนนจริง และวิธีการเชื่อมโยงคะแนนสังเกตได้) และโมเดลการประมาณค่าพารามิเตอร์ข้อสอบต่างกัน (โมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองทดสอบ)

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบระหว่างโมเดลการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ และทฤษฎีการตอบสนองทดสอบ
2. เพื่อเปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ของข้อสอบ (อำนาจจำแนก ความยาก การเดา) พารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองทดสอบ
3. เพื่อเปรียบเทียบประสิทธิภาพของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีการเชื่อมโยงคะแนนจริง และวิธีการเชื่อมโยงคะแนนสังเกตได้ภายใต้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน

ขอบเขตของการวิจัย

ขอบเขตของการวิจัยครั้งนี้ ประกอบด้วย 3 ด้าน มีรายละเอียดดังนี้

ขอบเขตด้านประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2558 จำนวน 42 เขตพื้นที่การศึกษา รวมจำนวนนักเรียนเป็น 414,287 คน (กระทรวงศึกษาธิการ, 2559)

ขอบเขตด้านเนื้อหา

ในการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบสอบเทสต์เลท ด้วยวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบต่างกัน และมีวิธีการประมาณค่าพารามิเตอร์ข้อสอบต่างกัน ซึ่งข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นคะแนนทฤษฎีจากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ได้รับการทดสอบการศึกษาระดับชาติขั้นพื้นฐานวิชาภาษาไทย และวิชาภาษาอังกฤษ ปีการศึกษา 2558 มีรายละเอียดของแบบสอบในแต่ละวิชาดังนี้

1. แบบสอบวิชาภาษาไทย จำนวน 50 ข้อ เป็นข้อสอบแบบเลือกตอบที่มีรูปแบบการให้คะแนนต่างกัน คือ ข้อที่ 1-40 เป็นข้อสอบเลือกตอบ 4 ตัวเลือกมีคำตอบถูกที่สุดเพียงข้อเดียว เกณฑ์ให้คะแนนเป็นตอบถูกได้ 2 คะแนน และตอบผิดได้ 0 คะแนน ส่วนข้อที่ 41-50 เป็นข้อสอบเลือกตอบ 5 ตัวเลือกมีคำตอบถูก 2 คำตอบ โดยในแต่ละข้อจะมีคะแนนเต็ม 2 คะแนนเมื่อเลือกคำตอบถูกต้องทั้งสองคำตอบ 1 คะแนนเมื่อเลือกคำตอบถูกเพียงข้อใดข้อหนึ่ง และ 0 คะแนนเมื่อไม่เลือกคำตอบถูกเลย นอกจากนี้ภายในแบบสอบประกอบด้วยข้อสอบที่ไม่มีเทสต์เลท จำนวน 20 ข้อ และมีเทสต์เลท จำนวน 5 เทสต์เลท ประกอบด้วย เทสต์เลทที่ 1 (ข้อ 1-10) เทสต์เลทที่ 2 (ข้อ 11-15) เทสต์เลทที่ 3 (ข้อ 16-20) เทสต์เลทที่ 4 (ข้อ 21-25) และเทสต์เลทที่ 5 (ข้อ 26-30)

2. แบบสอบวิชาภาษาอังกฤษ จำนวน 50 ข้อ เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือกมีคำตอบถูกเพียงข้อเดียว เกณฑ์ให้คะแนนเป็นตอบถูกได้ 2 คะแนน และตอบผิดได้ 0 คะแนน และภายในแบบสอบประกอบด้วยข้อสอบที่ไม่มีเทสต์เลท จำนวน 18 ข้อ และมีเทสต์เลท จำนวน 8 เทสต์เลท คือ เทสต์เลทที่ 1 (ข้อ 2-3) เทสต์เลทที่ 2 (ข้อ 4-5) เทสต์เลทที่ 3 (ข้อ 11-15) เทสต์เลทที่ 4 (ข้อ 16-20) เทสต์เลทที่ 5 (ข้อ 21-23) เทสต์เลทที่ 6 (ข้อ 26-30) เทสต์เลทที่ 7 (ข้อ 31-35) และเทสต์เลทที่ 8 (ข้อ 36-40)

3. โมเดลการวัดที่ใช้ในการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้สอบ ประกอบด้วย 2 โมเดล คือ โมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Item

response theory: IRT 3 PL) และโมเดลทฤษฎีการตอบสนองของเทสต์เลทแบบ 3 พารามิเตอร์ (Testlet Response Theory: TRT 3 PL)

4. การเชื่อมโยงคะแนนระหว่างแบบสอบตามทฤษฎีการตอบสนองข้อสอบ แบ่งเป็น 2 วิธี คือ การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะของแบบสอบ และการเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิควิเปอร์เซ็นต์ไทล์ โดยแบบสอบที่นำมาเชื่อมโยงมีโครงสร้างและเนื้อหาต่างกัน

ขอบเขตด้านตัวแปร

ตัวแปรอิสระ ประกอบด้วย

1. โมเดลการประมาณค่าพารามิเตอร์ข้อสอบ
 - 1.1 ตามทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (Item response theory: IRT 3 PL)
 - 1.2 ตามทฤษฎีการตอบสนองของเทสต์เลท แบบ 3 พารามิเตอร์ (Testlet response theory: TRT 3 PL)
2. วิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ
 - 2.1 วิธีเชื่อมโยงคะแนนจริง (Linking true score: TS)
 - 2.2 วิธีเชื่อมโยงคะแนนสังเกตได้ (Linking observed score: OS)

ตัวแปรตาม ประกอบด้วย

1. ความสัมพันธ์ แบ่งออกเป็น
 - 1.1 ความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ
 - 1.2 ความสัมพันธ์ของค่าพารามิเตอร์ความสามารถผู้สอบ
2. ความคลาดเคลื่อนมาตรฐาน แบ่งออกเป็น
 - 2.1 ความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ข้อสอบ
 - 2.2 ความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ความสามารถผู้สอบ
3. ประสิทธิภาพของการเชื่อมโยงคะแนน
 - 3.1 ค่าความลำเอียง (Bias)
 - 3.2 ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating: SEE)
 - 3.3 ค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error: RMSE)

ขอบเขตด้านข้อมูล

ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลทุติยภูมิ เป็นข้อมูลคะแนนสอบรายข้อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในวิชาภาษาไทยและวิชาภาษาอังกฤษ จากการทดสอบการศึกษา ระดับชาติขั้นพื้นฐาน (O-NET) ปีการศึกษา 2558

ประโยชน์ที่ได้รับจากการวิจัย

ประโยชน์ที่จะได้รับแบ่งออกเป็น 2 ด้าน ดังนี้

1. ประโยชน์ด้านวิชาการ

1.1 ได้ขยายองค์ความรู้ทั้งทางทฤษฎีและการปฏิบัติเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ ด้วยวิธีการเชื่อมโยงคะแนนจริง และคะแนนสังเกตได้ เมื่อแบบสอบที่ใช้มีอิทธิพลจากเทสต์เลท

1.2 ได้ขยายองค์ความรู้ทั้งทางทฤษฎีและการปฏิบัติเกี่ยวกับการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ และทฤษฎีการตอบสนองเทสต์เลท เมื่อแบบสอบที่ใช้มีอิทธิพลของเทสต์เลท

2. ประโยชน์ด้านการนำไปใช้

2.1 ผลการวิจัยนำไปใช้ประกอบการตัดสินใจเกี่ยวกับการเลือกใช้วิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ เมื่อแบบสอบที่พัฒนาขึ้นมีอิทธิพลของ เทสต์เลท

2.2 ผลการวิจัยนำไปใช้ประกอบการตัดสินใจเกี่ยวกับการเลือกใช้วิธีการประมาณค่าพารามิเตอร์ข้อสอบ เมื่อแบบสอบที่พัฒนาขึ้นมีอิทธิพลของเทสต์เลท

2.3 ให้สารสนเทศสำหรับผู้ที่เกี่ยวข้องกับการสร้างข้อสอบ ผู้ใช้ผลการสอบ และผู้บริหารการสอบ ในการตัดสินใจดำเนินการใด ๆ กับการทดสอบ เพื่อให้ผลของคะแนนมีความถูกต้อง และให้สารสนเทศที่ถูกต้องน่าเชื่อถือเกี่ยวกับผู้เรียน

นิยามศัพท์เฉพาะ

แบบสอบเทสต์เลท (Testlet test) หมายถึง แบบสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาไทย และวิชาภาษาอังกฤษ ในระดับชั้นมัธยมศึกษาปีที่ 3 ที่มีกลุ่มข้อสอบใช้สถานการณ์หรือข้อความร่วมกัน ทำให้ไม่เป็นอิสระในการตอบข้อสอบของแต่ละข้อ

แบบสอบเชื่อมโยง (Linking test) หมายถึง แบบสอบการศึกษาระดับชาติขั้นพื้นฐาน วิชาภาษาอังกฤษที่ได้รับการเชื่อมโยงคะแนนให้เท่าเทียมกับแบบสอบภาษาไทย

แบบสอบฐาน (Base test) หมายถึง แบบสอบการศึกษาระดับชาติขั้นพื้นฐานวิชาภาษาไทยที่เป็นหลักในการเชื่อมโยงคะแนนให้เท่าเทียมกัน

การประมาณค่าพารามิเตอร์ หมายถึง กระบวนการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบ ตามโมเดลทฤษฎีการทดสอบจำนวน 2 โมเดล คือ การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (Item response theory model: IRT 3PL) และการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองเทสต์เลท แบบ 3 พารามิเตอร์ (Testlet response theory model: TRT 3PL)

ความสัมพันธ์ของค่าพารามิเตอร์ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation coefficient) ของพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบที่ได้รับการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ และตามโมเดลทฤษฎีการตอบสนองเทสต์เลท แบบ 3 พารามิเตอร์

การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ หมายถึง กระบวนการในการเชื่อมโยงคะแนนของผู้สอบจากแบบสอบภาษาอังกฤษให้เท่าเทียมกับแบบสอบภาษาไทยซึ่งคะแนนจากแบบสอบทั้งสองฉบับถูกนำมาเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบจำนวน 2 วิธี คือ

1. วิธีการเชื่อมโยงคะแนนจริง (Linking true score) เป็นการเชื่อมโยงคะแนนจริงของนักเรียนที่ได้จากการหาผลรวมความน่าจะเป็นในการตอบข้อสอบถูกของนักเรียนแต่ละคนในแบบสอบวิชาภาษาไทย และวิชาภาษาอังกฤษ แล้วนำคะแนนจริงที่ได้จากแบบสอบภาษาอังกฤษมาแปลงให้สมมูลกับคะแนนจริงแบบสอบวิชาภาษาไทยด้วยวิธีโค้งคุณลักษณะแบบสอบ (Test characteristic curve method) โดยการเชื่อมโยงพารามิเตอร์ความสามารถผู้สอบของแบบสอบฐานและแบบสอบเชื่อมโยง

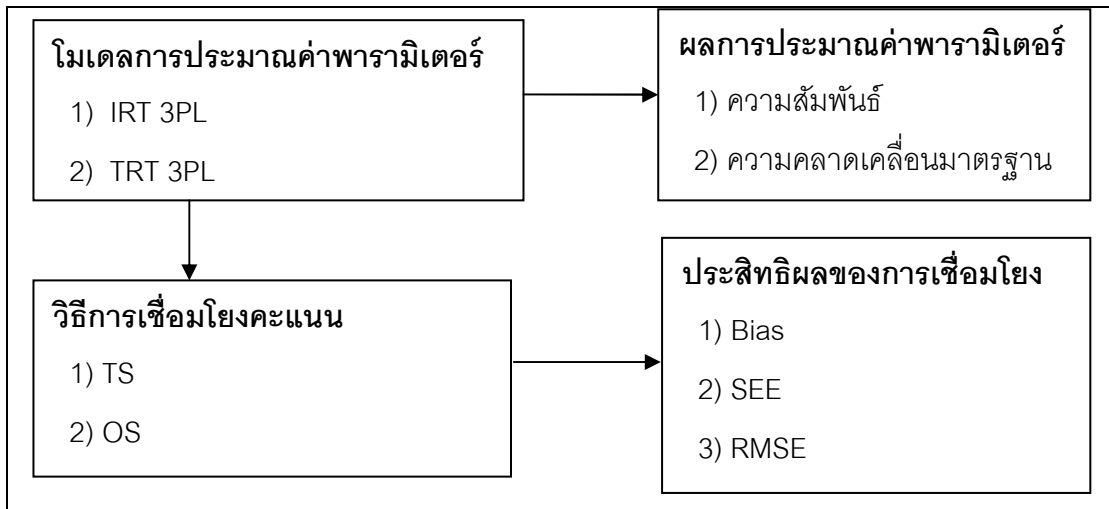
2. วิธีการเชื่อมโยงคะแนนสังเกตได้ (Linking observed score) เป็นการเชื่อมโยงคะแนนของนักเรียนโดยการนำความน่าจะเป็นในการตอบข้อสอบมาแปลงเป็นคะแนนสังเกตได้ แล้วหาความถี่ จากนั้นนำผลที่ได้มาแปลงคะแนนวิชาภาษาอังกฤษให้สมมูลกับวิชาภาษาไทยด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ (Equipercetile method) โดยการเชื่อมโยงตำแหน่งเปอร์เซ็นต์ไทล์ที่ตรงกันทั้งสองฉบับ

คะแนนเชื่อมโยง หมายถึง คะแนนภาษาอังกฤษที่ผ่านกระบวนการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบให้เท่าเทียมกับคะแนนภาษาไทย

ประสิทธิผลของการเชื่อมโยงคะแนน หมายถึง ความถูกต้องแม่นยำในการเชื่อมโยงคะแนน ซึ่งพิจารณาจาก ค่าความลำเอียง (Bias) เป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างเป็นระบบ ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating: SEE) เป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างสุ่ม และดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error: RMSE) เป็นความคลาดเคลื่อนที่เกิดขึ้นทั้งหมดในการเชื่อมโยง 4 เหตุการณ์ คือ การเชื่อมโยงคะแนนจริงด้วยการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (TS-IRT 3PL) การเชื่อมโยงคะแนนจริงด้วยการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองของเทสต์เลท แบบ 3 พารามิเตอร์ (TS-TRT 3PL) การเชื่อมโยงคะแนนสังเกตได้ด้วยการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (OS-IRT 3PL) และการเชื่อมโยงคะแนนสังเกตได้ด้วยการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองของเทสต์เลท แบบ 3 พารามิเตอร์ (OS-TRT 3PL)

กรอบแนวคิดของการวิจัย

ผู้วิจัยศึกษาประสิทธิผลของการเชื่อมโยงคะแนนแบบสอบเทสต์เลทตามทฤษฎีการตอบสนองข้อสอบ โดยออกแบบการเก็บรวบรวมข้อมูลกับคนกลุ่มเดียวใช้แบบสอบต่างฉบับไม่คู่ขนานกัน ทำการประมาณค่าพารามิเตอร์ข้อสอบ 2 โมเดล คือ การประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ (Item response theory: IRT 3PL) และการประมาณค่าตามโมเดลทฤษฎีการตอบสนองของเทสต์เลท แบบ 3 พารามิเตอร์ (Testlet response theory: TRT 3PL) ค่าพารามิเตอร์ที่ได้จากการประมาณค่า คือ ค่าพารามิเตอร์ข้อสอบ (อำนาจ จำแนก ความยาก ค่าการเดา) และค่าพารามิเตอร์ของผู้สอบ ตรวจสอบผลการประมาณค่าด้วยการหาความสัมพันธ์ของค่าพารามิเตอร์และเปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ จากนั้นจัดกระทำข้อมูลที่ได้ให้เป็นคะแนนจริงและคะแนนสังเกตได้ เพื่อนำไปใช้ในการเชื่อมโยงคะแนน ซึ่งวิธีการเชื่อมโยงคะแนนที่ใช้ในการศึกษาครั้งนี้มี 2 วิธี คือ การเชื่อมโยงคะแนนจริง (Linking true score: TS) ด้วยวิธีโค้งคุณลักษณะแบบสอบ และวิธีการเชื่อมโยงคะแนนสังเกตได้ (Linking observed score: OS) ด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ จากนั้นเปรียบเทียบประสิทธิผลของการเชื่อมโยงในแต่ละเงื่อนไขโดยพิจารณาจากค่าความลำเอียง (Bias) ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating: SEE) และค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error: RMSE) เขียนเป็นแผนภาพได้ดังนี้



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ศึกษาเกี่ยวกับประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบสอบ
ทดสอบตามทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยได้ศึกษาเอกสาร งานวิจัยทั้งในและต่างประเทศที่
เกี่ยวข้อง โดยนำเสนอแยกเป็นประเด็นได้ดังนี้

ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบทดสอบ

ตอนที่ 2 แนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 3 แนวคิดเกี่ยวกับประสิทธิผลของการเชื่อมโยงคะแนน

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบทดสอบ

1. มโนทัศน์เกี่ยวกับแบบสอบ

Wainer and Kiely (1987, pp. 190) ได้อธิบายความหมายของแบบสอบว่า “กลุ่มของ
ข้อสอบที่เกี่ยวข้องกับเนื้อหาเดียวกัน โดยได้รับการพัฒนาเป็นหนึ่งหน่วยและมีการกำหนด
โครงสร้างในการทำข้อสอบที่คงที่” และยังกล่าวอีกว่าแบบสอบถูกนำมาใช้ได้อย่างเหมาะสมใน
การสร้างแบบทดสอบ และมีประสิทธิภาพในการแก้ปัญหาเกี่ยวกับผลการสร้างข้อสอบที่เกิดจาก
เนื้อหา การเรียงลำดับตัวเลือก และความเท่าเทียมกันของเนื้อหา ซึ่งแนวคิดเกี่ยวกับแบบสอบนั้น
แรกเริ่มเดิมทีมาจากการทดสอบแบบปรับเหมาะ กล่าวคือ ข้อสอบภายในแบบสอบ แบบสอบ
ภายในแบบสอบสามารถเชื่อมโยงทั้งแบบระดับขั้นในการทดสอบแบบปรับเหมาะ หรือสามารถ
เชื่อมโยงในแบบเชิงเส้นตรงของการทดสอบแบบกระดาษและดินสอ แบบสอบสามารถแบ่งได้
เป็น 2 ประเภทตามวิธีการจัดกลุ่มข้อสอบ คือ 1) กลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน เช่น แบบทดสอบ
การอ่านเพื่อทำความเข้าใจ ซึ่งข้อสอบอ้างถึงเนื้อหาเดียวกัน และ 2) กลุ่มข้อสอบที่ใช้เนื้อหาเดียวกัน
แต่ไม่ต้องใช้สิ่งเร้าร่วมกัน เช่น แบบทดสอบย่อย ดังนั้นการศึกษานี้จะยึดหลักการของแบบสอบ
ประเภทที่ 1 คือ กลุ่มข้อสอบในแต่ละแบบสอบจะใช้สิ่งเร้าร่วมกัน และเป็นการทดสอบแบบ
กระดาษและดินสอ

ข้อสอบที่มีลักษณะแบบสอบที่ใช้สิ่งเร้าร่วมกัน ที่พบเห็นกันโดยทั่วไป คือ แบบสอบ
คณิตศาสตร์ แบบสอบวิทยาศาสตร์ ในการประเมินผลร่วมกับนานาชาติ (PIZA) แบบสอบวิชา
ภาษาไทย แบบสอบวิชาภาษาอังกฤษ ในการทดสอบระดับชาติขั้นพื้นฐาน (O-NET) ยกตัวอย่าง

เช่น

“เมื่อเคราะห์ร้ายกายเราก็เท่านี้ ไม่มีที่พสุธาจะอาศัย”

1. คำใดมีความหมายเหมือนกับคำว่า “พสุธา”

1. หาว 2. นภา 3. เวหา 4. ธรณี

2. สำนวนใดมีความหมายตรงกับบทร้อยกรองข้างต้น

1. คับที่อยู่ได้ คับใจอยู่ยาก 2. บ้านเคยอยู่ คู่เคยนอน

3. เจ้าไม่มีศาล สมภารไม่มีวัด 4. ทำคุณแก่ใครเหมือนไฟตกน้ำ

(ข้อสอบ O-NET วิชาภาษาไทย ระดับประถมศึกษาปีที่ 6 ปีการศึกษา 2558)

At a seaside resort:

Sara: What shall we do this afternoon?

Sonia: _____(1)_____.

Sara: Umm, it's such a boring activity. Why don't we go diving?

Sonia: _____(2)_____.

1) 1. I still have no idea what to do

2. Let's go sunbathing on the beach

3. I have to ask the rest of us first

4. Some of our friends won't go diving with us

2) 1. I'm afraid we can't

2. Because it's dangerous

3. That sounds like a great idea

4. We still have enough of time to go

(ข้อสอบ O-NET วิชาภาษาอังกฤษระดับมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558)

ขับรถ

เช้าวันหนึ่ง น้ำฝนออกไปขับรถเล่น ระหว่างทางแมตัวหนึ่งวิ่งตัดหน้ารถ น้ำฝนต้องเหยียบเบรคอย่างกะทันหันเพื่อไม่ให้ชนแมว

น้ำฝนตกใจมากจนมือสั่น จึงตัดสินใจขับรถกลับบ้าน

กราฟต่อไปนี้จะแสดงการบันทึกอัตราเร็วของรถอย่างคร่าว ๆ



1. อัตราเร็วสูงสุดของรถที่น้ำฝนขับในเช้าวันนั้น เป็นเท่าไร
2. ขณะที่น้ำฝนเหยียบเบรคเพื่อไม่ให้ชนแมว ขณะนั้นเป็นเวลาเท่าใด
3. สามารถบอกได้หรือไม่ว่า เส้นทางที่น้ำฝนขับรถกลับบ้านสั้นกว่าเส้นทางที่เธอออกจากบ้านถึงจุดชนแมว จงใช้ข้อมูลจากกราฟอธิบายเพื่อสนับสนุนคำตอบของนักเรียน

(ข้อสอบ PISA วิชาคณิตศาสตร์ ปี 2006)

จากตัวอย่างแบบสอบทดสอบเลขที่ทั้งที่เป็นปรนัยและอัตนัย จึงทำให้การประเมินทางการศึกษาในทางปฏิบัตินั้นมีการลดหรือขยายโครงสร้างการตอบสนองของทดสอบเลขที่เข้ามาใช้มากขึ้นและนำมาใช้ในการสร้างแบบสอบมาตรฐานแบบดั้งเดิม โดยส่วนใหญ่มีรูปแบบข้อสอบแบบหลายตัวเลือก (Multiple choice: MC) ซึ่งมีประโยชน์มากในการประเมินกับคนกลุ่มขนาดใหญ่ (Large scale assessment) ได้แก่ 1) ผู้ประเมินไม่มีอคติ เมื่อตรวจข้อสอบแบบหลายตัวเลือก หรือที่เรียกว่ามีความเป็นปรนัย 2) การทดสอบหนึ่งครั้งสามารถใช้ข้อสอบแบบเลือกตอบได้หลายข้อและครอบคลุมเนื้อหา 3) การให้คะแนนมีความเป็นปรนัย และจำนวนข้อสอบที่มากเป็นการเพิ่มความเชื่อมั่นและความตรงของแบบสอบ และ 4) การให้คะแนนของข้อสอบแบบหลายตัวเลือกสามารถทำได้ง่ายและประหยัดเวลา (Chen, 2014, p.11)

ข้อสอบแบบหลายตัวเลือกที่อิสระกัน เหมาะสำหรับการวัดพฤติกรรมด้านพุทธิพิสัยหรือผลการเรียนรู้ที่คาดหวังในระดับต่ำ ซึ่งข้อสอบลักษณะนี้มีนักวิจัยบางท่านได้วิพากษ์ไว้ว่า (เช่น

Bowman & Peng, 1972) ดีน้อยกว่าการประเมินความรู้ในระดับสูงของผู้สอบ นอกจากนี้ Nickerson (1989, p. 3) เน้นย้ำว่าการคิดขั้นสูงควรเป็นเป้าหมายที่สำคัญในการศึกษา และชี้ให้เห็นว่าไม่มีเครื่องมือประเมินสำหรับการคิดขั้นสูง “เราควรจะต้องตัดสินใจเกี่ยวกับการศึกษาในภาพรวม” ส่วนประเด็นการใช้ข้อสอบแบบหลายตัวเลือกอาจเกิดปัญหาด้านความตรง ซึ่งในความเป็นจริงจะต้องมีความรู้ความเข้าใจที่ซับซ้อน ในขณะที่ข้อสอบเหล่านี้ยังเป็นข้อสอบที่มีความเป็นอิสระกล่าวคือ ข้อสอบควรพิจารณาเกี่ยวกับการแยกเนื้อหา และการแยกออกจากโดเมนของการอ้างอิงเพื่อให้การใช้งานมีประสิทธิภาพมากขึ้น (Wainer et al., 2000, p. 245) ด้วยเหตุผลหลายประการที่กล่าวนี้ พบว่ากลุ่มข้อสอบแบบหลายตัวเลือกในเทสต์เลทสามารถช่วยแก้ปัญหาดังกล่าวได้ คือ 1) การรวมกันของเทสต์เลทในแบบทดสอบจะยังคงความเป็นแบบสอบหลายตัวเลือกทั้งหมด เช่น การให้คะแนนที่มีความเป็นปรนัย การครอบคลุมเนื้อหาที่ออกสอบ 2) เทสต์เลทประกอบด้วยกลุ่มของข้อคำถามที่มีสิ่งเร้าเดียวกัน เป็นการรวมเนื้อหาในการทดสอบเพิ่มขึ้นและช่วยลดความกังวลเกี่ยวกับการใช้ข้อคำถามที่เป็นอิสระกันหรือคำถามเดียว 3) เทสต์เลทจะมีข้อคำถามหลายข้อที่ใช้เนื้อหาที่เป็นสิ่งเร้าเดียวกัน เป็นการเพิ่มประสิทธิภาพในการพัฒนาแบบสอบและดีสำหรับการใช้เวลาการทำข้อสอบของผู้เข้าสอบ และ 4) การใช้ชุดข้อสอบที่มีความซับซ้อนและมีความสัมพันธ์กันภายในแบบสอบทำให้แบบสอบที่สร้างขึ้นมีความยืดหยุ่นมากขึ้นในการเขียนข้อสอบที่มีการประเมินต่างกันเกี่ยวกับลักษณะและระดับขั้นของการวัดพุทธิพิสัย ดังนั้นข้อสอบแบบเทสต์เลทจึงได้รับการพิจารณาในการใช้จริงมากขึ้นและมีความเหมาะสมกับการวัดทักษะในระดับสูงที่ดีกว่า (Wainer et al., 2000; DeMars, 2006 cited in Chen, 2014, p.12)

แบบสอบเทสต์เลทที่มีข้อจำกัดเหมือนกัน โดยเฉพาะประเด็นเกี่ยวกับความไม่อิสระกันของข้อสอบ (Local item dependence) และมีอิทธิพลต่อหลายลักษณะเกี่ยวกับการวิเคราะห์ข้อมูลในแบบสอบที่มีเทสต์เลท

2. ความเป็นอิสระกันของข้อสอบ

ความเป็นอิสระของข้อสอบเป็นการกล่าวถึงในบริบทของข้อตกลงที่สำคัญในทฤษฎีการตอบสนองของข้อสอบ (Item response theory: IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างความสามารถที่อยู่ภายในบุคคล (Ability) กับผลการตอบข้อสอบ ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) ดังนั้น IRT จึงอยู่บนพื้นฐานความคิดที่สำคัญ 2 ประการ คือ 1) ผลการตอบข้อสอบหรือข้อคำถามของผู้ตอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบ และ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (Item characteristic curve: ICC) (ศิริชัย กาญจนวาสี, 2555: หน้า 53)

โดยมีข้อตกลงเบื้องต้น 3 ประการคือ 1) Unidimensionality เป็นการสมมติว่าข้อสอบในแบบสอบมีลักษณะเป็นเอกพันธ์ นั่นคือ แบบสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียว และ 2) Local independence เป็นการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบสนองข้อสอบ กล่าวคือ การตอบสนองต่อข้อสอบข้อใดข้อหนึ่งของผู้สอบไม่มีผลต่อการตอบสนองต่อข้อสอบข้ออื่น ๆ ในแบบสอบ (สุรศักดิ์ อมรรัตนศักดิ์, 2544, หน้า 180-192)

จากข้อตกลงเบื้องต้นในข้อที่ 2 เกี่ยวกับความเป็นอิสระในการตอบสนองข้อสอบของผู้ตอบ ไม่สอดคล้องกับสถานการณ์การทดสอบในปัจจุบัน ที่มีการนำข้อสอบแบบเลือกตอบมาประยุกต์ใช้ในการวัดทักษะความสามารถที่มีความซับซ้อนมากขึ้น เช่น ทักษะการให้เหตุผลเชิงตรรกะ การให้เหตุผลเชิงวิเคราะห์ ทำให้ข้อสอบเดี่ยว (Single item) ไม่สามารถวัดทักษะที่ซับซ้อนได้อย่างครอบคลุม จึงมีแนวคิดเกี่ยวกับการจัดกลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน ข้อสอบลักษณะนี้จึงถูกมองว่าเป็นกลุ่มของข้อสอบ (Testlet) ซึ่งรวมอยู่ภายในแบบสอบฉบับเต็ม ตัวอย่างสถานการณ์ที่นำชุดข้อสอบมาใช้ ซึ่งส่วนใหญ่จะเป็นแบบสอบเกี่ยวกับการอ่านเพื่อความเข้าใจ ได้แก่ ในส่วนที่เป็น การอ่านเพื่อความเข้าใจ (Reading comprehension) ของการสอบ TOFEL ข้อสอบในส่วนนี้จะมีลักษณะเป็นชุดข้อสอบ โดยกลุ่มข้อสอบประมาณ 6 ถึง 12 ข้อจะใช้บทความ (Passage) ที่มีความยาวประมาณ 250 ถึง 400 คำ เป็นสิ่งเร้าร่วมกันเพื่อให้ผู้สอบตอบข้อสอบ ซึ่งการตอบภายในชุดข้อสอบถูกหรือไม่นั้น ไม่ได้ขึ้นอยู่กับความสามารถของผู้สอบเพียงอย่างเดียวแต่ขึ้นอยู่กับความเข้าใจในการตีความหรือการแปลความจากสิ่งเร้า และอาจจะเป็นไปได้ว่าผู้สอบเกิดการเรียนรู้จากการตอบข้ออื่น ๆ ในชุดข้อสอบนั้น ดังนั้นการตอบข้อสอบก่อนหน้าอาจส่งผลต่อการตอบในข้อถัดไป ถ้าผู้สอบเข้าใจสิ่งเร้าได้ถูกต้อง โอกาสในการตอบข้อสอบถูกต้องสูง ส่งผลให้การตอบข้อต่อไปถูกต้องสูงด้วย แต่ถ้ามองในทางกลับกัน ถ้าผู้สอบไม่เข้าใจสิ่งเร้า แล้วทำข้อสอบข้อแรก ๆ ผิด โอกาสในการตอบข้อต่อไปก็อาจจะผิดด้วย

Yen (1993, pp. 188-189) กล่าวว่า ความเป็นอิสระกันของข้อสอบ เป็นการนำเสนอปัจจัยที่สำคัญเกี่ยวกับ “ผลที่เกิดขึ้นบ่อยครั้งในความสามารถของนักเรียนบางคนเกี่ยวกับการทำข้อสอบบางข้อดีกว่าคนอื่น ๆ” ซึ่งอธิบายเพิ่มเติมได้ว่า ปัจจัยอาจจะเป็นที่น่าพอใจเมื่อได้สารสนเทศที่สำคัญเกี่ยวกับความสามารถที่ถูกวัด หรือปัจจัยอาจจะสร้างความรำคาญโดยไม่ได้ตั้งใจ เมื่อเป็นเช่นนี้ปัจจัยที่เกิดขึ้นโดยไม่ได้ตั้งใจมีผลคงที่ หรือเฉพาะกับนักเรียนหรือข้อสอบทั้งหมด ซึ่งสิ่งที่เกิดขึ้นนี้ว่าความไม่เป็นอิสระของข้อสอบ (Local item dependence) และ Yen ยังกล่าวถึงสาเหตุของการเกิดความไม่เป็นอิสระของข้อสอบ คือ การได้รับความช่วยเหลือจากภายนอกหรือสิ่งรบกวน (External assistance or interference) ความเร็ว (Speededness)

ความเมื่อยล้า (Fatigue) การฝึกปฏิบัติ (Practice) รูปแบบข้อสอบหรือการตอบข้อสอบ (Item or response formats) ความอิสระของข้อความหรือการตั้งค่า (Passage/ Setting dependence) ความเกี่ยวกันของข้อสอบ (Item chaining) คำอธิบายของคำตอบก่อนหน้า (Explanation of previous answer) ความเหมือนกันของเกณฑ์การให้คะแนนหรือผู้ประเมิน (Same scoring rubric or raters) และเนื้อหา ความรู้ และความสามารถ (Content, knowledge and ability)

สำหรับในแบบสอบที่มีเทสต์เลท ซึ่งข้อสอบถูกแบ่งเป็นกลุ่ม ๆ ภายใต้เทสต์เลทที่เกี่ยวข้องกับจำนวนของสิ่งเร้า ซึ่งสามารถนำเสนอเป็นรูปแบบต่าง ๆ เช่น บทความ แผนภาพ ตาราง และกราฟ โดยแบบสอบลักษณะนี้จะละเมิดข้อตกลงเบื้องต้นของความเป็นอิสระกันของข้อสอบ เพราะว่าการเพิ่มขึ้นของคุณลักษณะแฝงหลักที่ต้องการจะวัด ผลของเทสต์เลทเกิดขึ้นอย่างสม่ำเสมอตามลักษณะเฉพาะของแต่ละบุคคลภายในเทสต์เลทนั้น และการสร้างแผนผังข้อรายการของแหล่งความไม่เป็นอิสระของข้อสอบ มักจะเกิดจากข้อความหรือการตั้งค่าที่ไม่อิสระกัน เมื่อข้อมูลที่ได้จากการตอบข้อสอบต่างกันภายในเทสต์เลทเป็นความสัมพันธ์ภายในที่เกิดขึ้นจากสิ่งเร้าของเทสต์เลท หรือเมื่อนักเรียนมีระดับความสนใจหรือความรู้พื้นฐานเดิมเกี่ยวกับสิ่งเร้าต่างจากคนอื่นหรือผิดปกติ ดังนั้นถ้าเกิดลักษณะสถานการณ์เหล่านี้ นั้นหมายความว่า ข้อสอบไม่มีความเป็นอิสระกัน (Bradlow, Wainer, & Wang, 1999; Wainer & Wang, 2000; Yen, 1993 cited in Chen, 2014, pp.14-15)

ดังนั้น แบบสอบที่มีลักษณะนี้จะทำให้มีแหล่งความแปรปรวนและปัจจัยอันเนื่องมาจากความไม่เป็นอิสระในการตอบข้อสอบเพิ่มขึ้น ซึ่งถ้าวิเคราะห์คุณภาพข้อสอบหรือแบบสอบโดยไม่คำนึงถึงอิทธิพลส่วนนี้ อาจจะทำให้คุณภาพของแบบสอบหรือข้อสอบคลาดเคลื่อนไปจากความเป็นจริง จากการศึกษาของ Sireci, Thissen, and Wainer (1991, pp 237-247) สามารถอธิบายได้ว่า การละเลยความเป็นอิสระระหว่างข้อสอบและผู้สอบ ทำให้การประมาณค่าความเที่ยงจะเกินจริงประมาณ 10-15% จากปัญหาความไม่เป็นอิสระในการตอบข้อสอบนี้ มีนักวิจัยได้เสนอแนวทางในการแก้ปัญหาไว้ดังนี้

1. นำการวิเคราะห์ของคะแนนแบบหลายค่ามาประยุกต์ใช้ (Polytomous scoring) โดยการรวมข้อสอบที่ตรวจให้คะแนนแบบสองค่าในแบบสอบเทสต์เลท ให้เป็นข้อสอบหนึ่งข้อที่มีการตรวจให้คะแนนมากกว่าหนึ่งค่า (เช่น Samejima's, 1969; graded response model, Bock's 1972; nominal response model, and Muraki's, 1992; generalized partial credit model) ซึ่งการทำลักษณะเช่นนี้จะมีข้อจำกัดเกี่ยวกับสารสนเทศของข้อสอบที่เป็นรายข้อจะไม่สามารถวิเคราะห์ได้

2. การกำหนดให้ชุดข้อสอบถือเป็นหน่วยในการวิเคราะห์ด้วย โดยคำนวณคะแนนที่ตอบถูกในแต่ละชุดข้อสอบแบ่งในชุดข้อสอบ (เช่น Lee & Frisbie, 1999; Lee et al., 2001; Sireci, Thissen, & Wainer, 1991; Thissen, Steinberg, & Mooney, 1989; Wainer & Thissen, 1996) ซึ่งวิธีนี้จะทำให้ข้อสอบที่อยู่ระหว่างชุดข้อสอบมีความเป็นอิสระในการตอบ แต่จะไม่ใช่เป็นอิสระในการตอบภายในชุดข้อสอบนั้น ๆ

3. โมเดลทฤษฎีการตอบสนองของเทสต์เลท (Testlet response theory models: TRT) โดยคำนึงถึงอิทธิพลของความไม่เป็นอิสระในการตอบ ด้วยการเพิ่มส่วนที่เป็น Random Effect ($Y_{jg(i)}$) เข้าไปในสมการของโมเดลการตอบสนองของข้อสอบแบบ IRT ซึ่งเป็นการผ่อนคลายข้อตกลงที่เกี่ยวกับความเป็นอิสระระหว่างข้อสอบ

3. การจัดการความไม่เป็นอิสระของแบบสอบเทสต์เลทด้วยทฤษฎีการตอบสนองของข้อสอบ

จากความไม่เป็นอิสระกันของข้อสอบทำให้เกิดปัญหาหรือผลกระทบต่อทฤษฎีการวิเคราะห์ข้อมูล ดังนั้นจึงมีรูปแบบที่ใช้สำหรับการวิเคราะห์ข้อมูลลักษณะดังกล่าว และเป็นที่ยอมรับในปัจจุบัน (Chen, 2014, pp. 17-22) มีรายละเอียดดังนี้

3.1 ทฤษฎีการตอบสนองของข้อสอบแบบสามพารามิเตอร์ (3PL IRT)

รูปแบบ 3PL IRT เป็นรูปแบบทั่วไปที่นิยมใช้มากในโมเดลทฤษฎีการตอบสนองของข้อสอบ ที่มีการวัด 2 ค่าและวัดในมิติเดียว ซึ่งเหมาะสมกับข้อมูลที่ได้มาจากแบบสอบกับคนจำนวนมากและแบบสอบเป็นแบบหลายตัวเลือก อาจจะมีปัญหาเกี่ยวกับการเดาคำตอบ เมื่อมีการประยุกต์ใช้รูปแบบ 3PL IRT กับข้อมูลที่ได้มาจากแบบสอบที่มีเทสต์เลท การมีความไม่เป็นอิสระกันของข้อสอบและผลของเทสต์เลทจะถูกละเว้น ดังนั้นข้อสอบจะพิจารณาให้เป็นอิสระจากกันในแต่ละข้อ โดยมีฟังก์ชันการตอบสนองของข้อสอบ (Birnbaum, 1968) ดังนี้

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[1.7a_i(\theta_j - b_i)]}}{1 + e^{[1.7a_i(\theta_j - b_i)]}}$$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกของคนที่ j

θ_j คือ พารามิเตอร์ความสามารถของผู้สอบคนที่ j

a_i คือ พารามิเตอร์อำนาจจำแนกข้อสอบ i

b_i คือ พารามิเตอร์ความยากข้อสอบ i

c_i คือ พารามิเตอร์การเดาข้อสอบ i

3.2 ทฤษฎีการตอบสนองของข้อสอบแบบ Graded response model

จากการเกิดความไม่เป็นอิสระกันของข้อสอบ การใช้โมเดลทฤษฎีการตอบสนองของข้อสอบ ที่มีการวัด 2 ค่าและวัดในมิติเดียวอาจจะไม่เหมาะสม นักวิจัย (เช่น Lee et al., 2001; Thissen et al., 1989; Zhang, 2007) ได้เสนอการให้คะแนนแบบสอบทดสอบหลายระดับ โดยการรวมคะแนนของข้อสอบภายในทดสอบเดียวกัน คะแนนทดสอบที่ได้จะเป็นแบบการตรวจให้คะแนนมากกว่า 2 ค่า และโมเดลการตอบสนองของข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า ถูกนำมาประยุกต์ใช้ในกรณีนี้ (Thissen et al., 1989) Graded response model: GRM ของ Samejima (1969) เป็นหนึ่งในโมเดลการตอบสนองของข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า สามารถนำมาใช้กับข้อมูลที่ได้จากแบบสอบที่มีหลายระดับ GRM ของ Samejima (1969) เกี่ยวข้องกับการตอบสนองแบบเรียงลำดับ เมื่อคะแนนข้อสอบภายในทดสอบถูกรวมให้เป็นคะแนนทดสอบหลายระดับ ซึ่งต้องพิจารณาคะแนนทดสอบหลายระดับคือ “การเรียงลำดับ (Ordered)” ในแง่ความเป็นไปได้ของผู้สอบที่จะตอบข้อสอบถูกต้องภายในทดสอบหลายระดับ กับแนวโน้มความสามารถของผู้สอบระดับสูง นั่นคือความน่าจะเป็นคนที่ j กับความสามารถ θ_j ทำข้อสอบข้อที่ i หรือสูงกว่ารายการคำตอบ k แสดงดังสมการ

$$P_{ijk}^*(\theta_j, a_i, b_{i2}, \dots, b_{iK}) = \begin{cases} 1 & k = 1 \\ \frac{e^{[a_i(\theta_j - b_{ik})]}}{1 + e^{[a_i(\theta_j - b_{ik})]}} & 2 \leq k \leq K \\ 0 & k > K \end{cases}$$

จากสมการ ข้อสอบข้อที่ i มีทั้งหมด K รายการคำตอบ a_i คือ พารามิเตอร์ค่าอำนาจจำแนกของข้อสอบข้อที่ i และ b_{ik} คือพารามิเตอร์ความยากของข้อสอบข้อที่ i รายการคำตอบ 2 ถึง K บทฐานของฟังก์ชันการตอบสนองรายการคำตอบสะสม (Cumulative category response function) ซึ่งเป็นความน่าจะเป็นของการให้คะแนนในรายการคำตอบ k สามารถคำนวณได้โดยการหาความแตกต่างระหว่างฟังก์ชันการตอบสนองรายการคำตอบสะสม ดังสมการ

$$P_{ijk} = P_{ijk}^*(\theta_j, a_i, b_{i2}, \dots, b_{iK}) - P_{ij(k+1)}^*(\theta_j, a_i, b_{i2}, \dots, b_{iK})$$

3.3 โมเดลทฤษฎีการตอบสนองของทดสอบหลายระดับ

ในโมเดลทฤษฎีการตอบสนองของทดสอบหลายระดับ (Testlet response theory model: TRT) เป็นโมเดลที่แสดงผลที่เกิดจากอิทธิพลของทดสอบหลายระดับอย่างชัดเจน (Wainer, et al., 2000; Wainer,

et al., 2007) โดยมีฟังก์ชันที่ใช้ในการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้สอบดังนี้

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}}{1 + e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}}$$

เมื่อ P_{ij} เป็นความน่าจะเป็นในการตอบถูกของคุณลักษณะหลักและพารามิเตอร์อิทธิพลของเทสต์เลทของผู้สอบคนที่ j ส่วน a_i , b_i , c_i เป็นพารามิเตอร์ข้อสอบข้อที่ i ประกอบด้วย ค่าอำนาจจำแนก ค่าความยาก และค่าการเดาตามลำดับ พารามิเตอร์ความสามารถของผู้สอบคนที่ j (θ_j) เป็นคุณลักษณะแฝงหลัก ซึ่งในเทอมของ $\gamma_{jd(i)}$ เป็นตัวชั้อิทธิพลสุ่มของเทสต์เลทสำหรับคนที่ j ในเทสต์เลท $d(i)$ การแจกแจงของ θ_j ถูกกำหนดให้มีการแจกแจงแบบปกติ ($N(0,1)$) และอิทธิพลสุ่มของเทสต์เลทถูกกำหนดให้เหมือนการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานของ $d(i)$ ($\sigma_{d(i)}$) ค่าของ $\gamma_{jd(i)}$ เป็นค่าคงที่ภายในเทสต์เลทสำหรับผู้สอบเฉพาะและแตกต่างกันระหว่างผู้สอบ

จากที่กล่าวมาทั้งสามวิธี พบว่าการใช้คะแนนเทสต์เลทและการประยุกต์ใช้โมเดลการตรวจให้คะแนนมากกว่าสองค่าแบบมิติเดียว ข้อมูลจากแบบสอบที่มีเทสต์เลทดูเหมือนว่าจะไม่มีปัญหาความไม่เป็นอิสระกันของข้อสอบ (Lee, et al., 2001, Zhang, 2007) อย่างไรก็ตามมีข้อจำกัดหลายประการเมื่อนำคะแนนเทสต์เลทมาใช้เป็นหน่วยของการวิเคราะห์ และการประยุกต์ใช้โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า Wainer et al. (2000) and Wainer et al. (2007) ซึ่งชี้ให้เห็นว่าการประยุกต์ใช้โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า รูปแบบที่ชัดเจนของคะแนนภายในแต่ละเทสต์เลทถูกละเลย และสารสนเทศภายในเทสต์เลทก็หายไป Wainer (1995) ยังชี้ให้เห็นว่า เมื่อระดับความไม่อิสระกันมีค่าสูง แบบสอบที่มีเทสต์เลทอาจจะไม่เหมาะสมกับโมเดลที่มีการตรวจให้คะแนนหลายค่า ยิ่งไปกว่านั้น เมื่อเทสต์เลทในแบบสอบมีจำนวนข้อสอบมาก การประยุกต์ใช้โมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่าข้อมูลที่ได้ อาจจะมีความซับซ้อน และโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองยังเป็นวิธีที่ใช้กันอย่างไม่แพร่หลายในทางปฏิบัติ (Wainer & Wang, 2000) จากข้อจำกัดเหล่านี้ ผู้วิจัยจึงใช้โมเดลทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ (3PL IRT) และโมเดลทฤษฎีการตอบสนองเทสต์เลทแบบสามพารามิเตอร์ (3PL TRT) ในการประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถผู้สอบ

4. การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ที่นิยมใช้กันในปัจจุบันมี 3 วิธี คือ วิธี Joint maximum likelihood วิธี Marginal maximum likelihood และ วิธี Bayesian estimation ทั้ง 3 วิธีใช้สำหรับการประมาณพารามิเตอร์ข้อสอบและความสามารถของผู้สอบ แต่วิธี Bayesian จะขจัดปัญหาบางอย่าง เช่น ความไม่เหมาะสมของวิธีการ และการไม่เข้าสู่เกณฑ์ที่พบในวิธี Joint และ marginal maximum likelihood โดยใช้การแจกแจง prior distribution แทนที่พารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถ (Swaminathan & Gifford, 1982; 1985; 1986) นอกจากนี้แบบสอบที่ละเมิดความเป็นอิสระในการตอบข้อสอบแต่ละข้อหรืออิทธิพลของเทสต์เลท จึงมีนักการศึกษา ได้เพิ่มเติมอิทธิพลของเทสต์เลทเข้าไปในการประมาณค่าด้วย พบว่าการประมาณค่าพารามิเตอร์วิธีเบย์สแบบมีอิทธิพลเทสต์จะให้ความแม่นยำมากกว่าวิธีอื่น ๆ เมื่อข้อสอบไม่อิสระจากกัน (Purya & Hamdollah, 2016; อริศฟ้า เตหฺลิม, 2559) ซึ่งการประมาณค่าพารามิเตอร์ที่คำนึงถึงอิทธิพลของเทสต์เลทนั้น ได้มีโปรแกรมช่วยในการประมาณค่าพารามิเตอร์ คือ SCORIGHT (version 3.0) ที่พัฒนาโดย Wang, Bradlow and Wainer (2005) โดยมีรายละเอียดในการประมาณค่าดังนี้

4.1 วิธีประมาณค่าพารามิเตอร์

การประมาณค่าของโปรแกรม SCORIGHT จะประมาณค่าด้วยวิธีของเบย์ส (Bayesian) โดยใช้เทคนิคมาร์คอฟเชนมอนติคาร์โล (Markov Chain Monte Carlo: MCMC) สามารถประมาณค่าพารามิเตอร์ของข้อสอบตามโมเดลทฤษฎีการตอบสนองข้อสอบที่มีการตรวจให้คะแนน 2 ค่า และการตรวจให้คะแนนมากกว่าสองค่า อีกทั้งยังประมาณค่าพารามิเตอร์ข้อสอบตามโมเดลทฤษฎีการตอบสนองเทสต์เลทได้อีกด้วย โดยมีฟังก์ชันในการประมาณค่าพารามิเตอร์ข้อสอบดังนี้ กรณีที่มีการตรวจให้คะแนนแบบ 2 ค่า จะใช้โมเดล 3PL

$$P(Y_{ij} = 1) = c_i + (1 - c_i)\text{logit}^{-1}(t_{ij})$$

กรณีที่ตรวจให้คะแนนมากกว่า 2 ค่า จะใช้โมเดล ordinal response ของ Samejima (1969)

$$P(Y_{ij} = r) = \Phi(d_r - t_{ij}) - \Phi(d_{r-1} - t_{ij})$$

โดยที่ $t_{ij} = a_i(\theta_j - b_i - \gamma_{jd(i)})$ เมื่อ Y_{ij} คือการตอบสนองข้อสอบข้อที่ i ของคนที่ j

การประมาณค่าพารามิเตอร์ด้วยโปรแกรม SCORIGHT นั้นมีการกำหนดลักษณะการแจกแจงของข้อมูล ดังนี้

$$\theta_j \sim N(0, 1)$$

$$\log(a_i) \sim N(\mu_a, \sigma_a^2)$$

$$b_i \sim N(\mu_b, \sigma_b^2)$$

$$\log(c_i) \sim N(\mu_c, \sigma_c^2)$$

$$\gamma_{jd(i)} \sim N(0, \sigma_\gamma^2)$$

4.2 ขนาดกลุ่มตัวอย่างและจำนวนข้อสอบที่ใช้ในการประมาณค่าพารามิเตอร์

มีนักวิจัยได้ศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการประมาณค่าพารามิเตอร์ข้อสอบแบบ 3PL พบว่าขนาดกลุ่มตัวอย่างจำนวน 1,000 คน ก็เพียงพอสำหรับการประมาณค่าพารามิเตอร์ข้อสอบแบบ 3PL (Swaminathan & Gifford, 1983 cited in Foley, 2010, p. 2; Hambleton, 1993; Jones et. al., 2006; Yen & Fitzpatrick, 2006 cited in Foley, 2010, p. 16; de Ayala, 2009, p. 131) และมีนักวิจัยอีกกลุ่มหนึ่งให้ข้อเสนอแนะว่าขนาดกลุ่มตัวอย่าง 700-1,500 คน เหมาะสมกับการวิเคราะห์ IRT 3 พารามิเตอร์ (Lam, 2009, p. 14) แต่ทั้งนี้การพิจารณาว่าขนาดกลุ่มตัวอย่างที่เหมาะสมหรือเพียงพอเป็นจำนวนเท่าไรนั้นอาจจะต้องพิจารณาปัจจัยอื่นร่วมด้วย เช่น วัตถุประสงค์ของการทดสอบ วิธีการประมาณค่าการแจกแจงของข้อมูล ข้อมูลสูญหาย จำนวนข้อสอบ โดยจำนวนข้อสอบที่เหมาะสมในการประมาณพารามิเตอร์แบบ 3PL เมื่อมีขนาดกลุ่มตัวอย่างขั้นต่ำ 1,000 คนนั้น พบว่า จำนวนข้อสอบขั้นต่ำเท่ากับ 50 ข้อ (Lord, 1968) จำนวน 30 หรือ 60 ข้อ (Hulin, Lissak, & Drasgow, 1982)

จากแนวคิดเกี่ยวกับเทสต์เลทผู้วิจัยสามารถสรุปประเด็นเกี่ยวกับการศึกษาครั้งนี้ได้ว่าแบบสอบที่ใช้ในการศึกษาครั้งนี้จะมีลักษณะความไม่อิสระกันของข้อสอบ โดยใช้แบบสอบการศึกษาขั้นพื้นฐานระดับชาติ (O-NET) ของระดับชั้นมัธยมศึกษาปีที่ 3 ประกอบด้วยแบบสอบ 2 ฉบับ คือ แบบสอบวิชาภาษาไทย และแบบสอบวิชาภาษาอังกฤษ โดยแต่ละฉบับจะมีจำนวนข้อฉบับละ 50 ข้อ และการประมาณค่าพารามิเตอร์ของข้อสอบนั้นจะพิจารณาจาก 2 โมเดล คือ โมเดลทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ และทฤษฎีการตอบสนองเทสต์เลทแบบสามพารามิเตอร์ ประมาณค่าพารามิเตอร์ข้อสอบด้วยโปรแกรม SCORIGHT กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,500 คน โดยสุ่มนักเรียนมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558

จากแบบสอบทั้งสองฉบับที่ใช้ในการศึกษานั้น เราสามารถทำอะไรได้บ้างที่จะทำให้คะแนนทั้งสองฉบับนี้มีความเท่าเทียมกัน ดังนั้นผู้วิจัยจึงศึกษาเอกสารและงานวิจัยเกี่ยวกับทฤษฎีการวัด พบว่า มีหลักการหนึ่งที่สามารถทำให้คะแนนมีความเท่าเทียมกันคือ การเชื่อมโยงคะแนนระหว่างแบบสอบ โดยมีรายละเอียดในหัวข้อถัดไป

ตอนที่ 2 แนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองของข้อสอบ

การศึกษาแนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองของข้อสอบนั้น ผู้วิจัยได้ศึกษาในประเด็นต่าง ๆ ดังนี้

มโนทัศน์เกี่ยวกับการเชื่อมโยงคะแนน

นิยามการเชื่อมโยงคะแนนและรูปแบบการเชื่อมโยงคะแนน ได้มีนักวิชาการได้นำเสนอไว้หลายท่านด้วยกัน อาทิเช่น Flanagan (1951); Angoff (1971); Mislevy (1992); Linn (1993); Feuer et al. (1999); Dorans (2000, 2004); Kolen (2004); Holland (2007); Kolen and Brennan (2014) ซึ่งผู้วิจัยได้นำเสนอเฉพาะบางแนวคิดที่สอดคล้องกับการวิจัยดังนี้

Feuer et al. (1999, pp.15-23) ได้สรุปเกณฑ์การจำแนกรูปแบบการเชื่อมโยงออกเป็น 3 เกณฑ์ คือ เกณฑ์ที่ 1 แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบเดียวกัน เกณฑ์ที่ 2 แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบที่มีลักษณะเดียวกันแต่มีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน และ เกณฑ์ที่ 3 แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน จากนั้นนำเกณฑ์ที่ได้มาผนวกกับแนวคิดของ Mislevy (1992) และ Linn (1993) ได้รูปแบบการเชื่อมโยงคะแนน 4 รูปแบบ พร้อมทั้งนิยามของทั้ง 4 รูปแบบ ดังนี้

รูปแบบที่ 1 การปรับเทียบคะแนน (Equating) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบสองแบบสอบที่ถูกสร้างขึ้นภายใต้โครงสร้างข้อสอบ (Test blueprint) เดียวกัน กล่าวคือ มีโครงสร้างหรือรูปแบบข้อสอบเหมือนกัน ความยาก ความเชื่อมั่นเท่ากัน และถูกใช้ในวัตถุประสงค์การสอบเดียวกัน

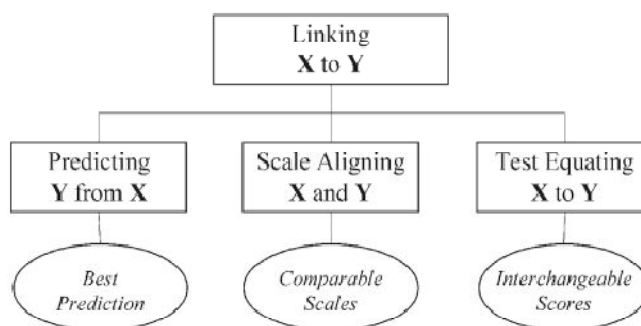
รูปแบบที่ 2 การทำให้เป็นคะแนนมาตรฐาน (Calibration) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบสองแบบสอบที่มีโครงสร้างหลักเหมือนกัน แต่ต่างกันในเรื่องคุณลักษณะแบบสอบ (เช่น ความยาวข้อสอบ) และต่างกันของคุณลักษณะทางสถิติ

รูปแบบที่ 3 การฉายภาพ (Projection) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบสองแบบสอบที่มีคุณลักษณะต่างกัน โดยการทำนายหรือฉายภาพคะแนนจากแบบสอบหนึ่งไปยังแบบสอบอื่น โดยผลที่ได้จากการเชื่อมโยงจะไม่สมมาตรกัน กล่าวคือ ความสัมพันธ์ที่ได้ระหว่าง

คะแนนแบบสอบ A กับแบบสอบ B ไม่เท่ากับความสัมพันธ์ระหว่างคะแนนจากแบบสอบ B กับแบบสอบ A นอกจากนี้วิธีการนี้ใช้กับการออกแบบในการเก็บข้อมูลกับผู้สอบกลุ่มเดียวและไม่น่าหนึ่งคุณลักษณะของแบบสอบว่าต้องเหมือนกัน

รูปแบบที่ 4 การปรับค่า (Moderation) หรือบางครั้งเรียก การปรับค่าทางสถิติ (Statistical moderation) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบที่มีคุณลักษณะต่างกัน และกลุ่มผู้สอบไม่เท่าเทียมกัน วิธีนี้ถือว่าเป็นวิธีการเชื่อมโยงคะแนนที่อ่อนแอที่สุด

Holland (2007, pp. 5-30) ได้สรุปหลักการการเชื่อมโยงคะแนนของ Holland and Dorans (2006) ที่ขยายและปรับมาจากแนวคิดของ Mislevy/ Linn ออกเป็น 3 รูปแบบ คือ



ภาพที่ 2-1 รูปแบบการเชื่อมโยงคะแนนและวัตถุประสงค์ในการใช้ (Holland, 2007, p. 5)

รูปแบบที่ 1 การทำนาย (Predicting) หมายถึงการเชื่อมโยงคะแนนที่นำวิธีการถดถอยมาใช้ในการทำนายคะแนนจากแบบสอบฉบับหนึ่งไปยังแบบสอบฉบับอื่น เช่น คะแนนที่ได้รับ การทดสอบในระดับมัธยมสามารถนำไปใช้ทำนายคะแนนจากแบบทดสอบในระดับมหาวิทยาลัย และอีกประเภทของการเชื่อมโยงนี้คือ การฉายภาพ (Projection) การแจกแจงคะแนนของแบบสอบหนึ่งบนอีกแบบสอบหนึ่ง

รูปแบบที่ 2 สเกลสอดคล้อง (Scale aligning) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบ แบ่งออกเป็น 2 ลักษณะ คือ แบบสอบที่มีโครงสร้างเหมือนกัน ประกอบด้วย 3 รูปแบบ คือ 1) ความสอดคล้อง (Concordance) เป็นการเชื่อมโยงในกรณีที่มีโครงสร้าง ความยาก และความเชื่อมั่นคล้ายกัน อีกทั้งประชากรมีคุณลักษณะที่คล้ายกันด้วย 2) การปรับสเกลแนวตั้ง (Vertical scaling) เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างและความเชื่อมั่นคล้ายคลึงกัน แต่จะต่างกันในเรื่องของความยากและคุณลักษณะของประชากร และ 3) การทำให้

เป็นคะแนนมาตรฐาน (Calibration) เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างเหมือนกัน แต่ความเชื่อมั่นต่างกัน เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีจำนวนข้อน้อยไปยังแบบสอบที่มีจำนวนข้อมากกว่า ส่วนแบบสอบที่มีโครงสร้างต่างกัน เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างต่างกัน โดยมักจะมีการปรับสเกลให้มีค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการกระจายในประชากรที่เท่ากัน เพื่อเพื่อความสะดวกในการเชื่อมโยงคะแนน หรือที่เรียกว่า Battery scaling

รูปแบบที่ 3 การปรับเทียบ (Equating) หมายถึงการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างและเนื้อหาเดียวกัน มีความยาก ความเชื่อมั่นคล้ายคลึงกัน คะแนนเชื่อมโยงที่ได้สามารถนำมาใช้แทนกันได้

Kolen and Brennan (2014, pp. 486, 498-500) กล่าวถึงสถานการณ์เกี่ยวกับการนำกระบวนการทางสถิติมาปรับเปลี่ยนคะแนนระหว่างแบบสอบในกรณีที่แบบสอบมีความแตกต่างกันในด้านเนื้อหา หรือความยาก หรือทั้งสองอย่าง ซึ่งบางการทดสอบอาจมีความแตกต่างกันเพียงเล็กน้อย แต่การทดสอบส่วนใหญ่มีความแตกต่างด้านเนื้อหาหรือโครงสร้างอย่างชัดเจน เรียกความสัมพันธ์ระหว่างคะแนนว่า การเชื่อมโยง และพวกเขาได้เสนอแนวคิดเกี่ยวกับการเชื่อมโยงคะแนนในเรื่องของระดับความคล้ายคลึงในสิ่งที่เรียกว่า คุณสมบัติ หรือธรรมดาสัมฤทธิ์ มาเป็นเกณฑ์ อีกทั้งประโยชน์และความสมเหตุสมผลของการเชื่อมโยงคะแนนนั้นขึ้นอยู่กับคุณสมบัตินั้นๆ สรุปได้ 4 คุณสมบัติ ดังนี้

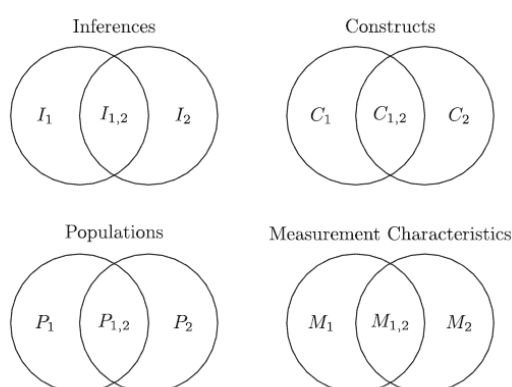
1. การสรุปอ้างอิง (Inferences) เป็นการตอบคำถามว่า “ขอบเขตด้านการสรุปอ้างอิงของคะแนนสำหรับแบบสอบสองฉบับนำมาใช้บอกความเหมือนกันได้หรือไม่” กล่าวคือ คำถามในแบบทดสอบทั้งสองฉบับมีวัตถุประสงค์ในการวัดร่วมกัน นั่นคือกระบวนการในการแปลงสเกลจากวัตถุประสงค์ไปยังผลที่เหมือนกัน

2. โครงสร้าง (Constructs) เป็นการตอบคำถามว่า “ขอบเขตด้านโครงสร้างของแบบสอบทั้งสองฉบับเหมือนกันหรือไม่” กล่าวคือ คะแนนจริงของแบบสอบสองฉบับมีความเกี่ยวข้องสัมพันธ์กัน ในบริบทของการเชื่อมโยงเนื้อหาซึ่งแบบสอบอาจจะมีโครงสร้างร่วมกัน แต่ไม่สามารถประเมินได้ว่าแบบสอบมีโครงสร้างเดียวกัน

3. กลุ่มประชากรผู้สอบ (Populations) เป็นการตอบคำถามว่า “ขอบเขตของแบบสอบทั้งสองฉบับที่ออกแบบมาเพื่อใช้วัดประชากรกลุ่มเดียวกันหรือไม่” กล่าวคือ แบบทดสอบอาจจะสร้างมาจากโครงสร้างเดียวกัน แต่ไม่เหมาะสมกับประชากรกลุ่มเดียวกัน

4. คุณลักษณะหรือเงื่อนไขของการวัด (Measurement characteristics/ Conditions) เป็นการตอบคำถามว่า “ขอบเขตของแบบสอบสองฉบับที่มีคุณลักษณะหรือเงื่อนไขของการวัด

ร่วมกัน (เช่น ความยาวของแบบสอบ รูปแบบแบบสอบ การบริหารการสอบ) หรือที่เรียกว่า ฟาเซต (Facet) ในทฤษฎีการสรุปอ้างอิง ของทั้งสองแบบสอบอาจจะต่างกันขึ้นอยู่กับจำนวนเงื่อนไข” กล่าวคือ คุณลักษณะของแบบสอบนั้นอาจจะเป็นส่วนหนึ่ง (ถึงแม้จะสำคัญมาก) ของคุณลักษณะหรือเงื่อนไขการวัด เช่น คุณลักษณะของแบบสอบไม่ได้ทำให้คะแนนมีความมั่นคงพอที่จะอ้างอิงไปยังสถานการณ์อื่น แต่จะวัดได้ดีในสิ่งที่เราสนใจ สามารถสรุปเป็นรูปภาพของทั้ง 4 คุณสมบัติดังนี้



ภาพที่ 2-2 ระดับความคล้ายคลึงกันของการเชื่อมโยงคะแนน (Kolen & Brennan, 2014, p. 449)

จากระดับความเหมือนนี้ Kolen and Brennan ได้นำไปผนวกเข้ากับแนวคิดของ Mislevy (1992) ในกรณีของแบบสอบ 2 ฉบับ วัดคุณลักษณะต่างกัน และ Linn (1993) ได้นำแนวคิดดังกล่าวมาใช้ในการจำแนกชนิดของการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับที่วัดคุณลักษณะต่างกัน โดยเรียกว่า รูปแบบการเชื่อมโยงคะแนน ซึ่งใช้วิธีการทางสถิติและความแกร่งจากผลของคะแนนที่ได้จากการเชื่อมโยงเป็นเครื่องมือในการจำแนกแบ่งออกเป็น 4 ชนิด คือ 1) การปรับเทียบคะแนน (Equating) 2) การทำคะแนนให้เป็นมาตรฐาน (Calibration) 3) การฉายภาพ (Projection) และ 4) การปรับด้วยค่าสถิติ (Statistical moderation) โดยการเชื่อมโยงชนิดการปรับด้วยค่าสถิติ (Statistical moderation) และการฉายภาพ (Projection/ Prediction) มีความเป็นอิสระจากกลุ่มผู้สอบและการแปลงคะแนนด้วยวิธีการปรับด้วยค่าสถิติ (Statistical Moderation) มีคุณสมบัติของความสมมาตร (Symmetric) แต่วิธีการฉายภาพ (Projection) ไม่มีคุณสมบัติของความสมมาตร (Symmetric) สรุปรายละเอียดดังตาราง

ตารางที่ 2-1 รูปแบบการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy/ Linn และแนวคิดระดับ
ความเหมือนของ Kolen and Brennan (2014, p. 449)

ประเภท	ระดับความคล้ายคลึงกัน			
	การสรุป อ้างอิง	โครงสร้าง	ประชากร	คุณลักษณะ/ เงื่อนไข
Equating	เหมือนกัน	เหมือนกัน	เหมือนกัน	เหมือนกัน
Vertical scaling	เหมือนกัน	เหมือน/ คล้ายคลึงกัน	ไม่คล้ายคลึง กัน	เหมือน/ คล้ายคลึงกัน
Concordance/ Moderation	เหมือนกัน	คล้ายคลึงกัน	เหมือน/ คล้ายคลึงกัน	ไม่คล้าย/ คล้ายคลึงกัน
Projection	ไม่คล้าย/ คล้ายคลึงกัน	ไม่คล้าย/ คล้ายคลึงกัน	คล้ายคลึงกัน	ไม่คล้ายคลึงกัน
Statistical moderation	ไม่คล้าย/ คล้ายคลึงกัน	ไม่คล้าย/ คล้ายคลึงกัน	ไม่คล้าย/ คล้ายคลึงกัน	ไม่คล้ายคลึงกัน

จากรูปแบบของการเชื่อมโยงคะแนนที่กล่าวมาข้างต้น สามารถสรุปได้ 4 รูปแบบ คือ
1) การปรับเทียบคะแนน (Equating) 2) การปรับสเกลคะแนนมาตรฐาน (Calibration)
3) การทำนายหรือการฉายภาพ (Predicting or Projection) และ 4) การปรับค่า (Moderation)
จากทั้ง 4 รูปแบบพบว่ารูปแบบที่มีความสอดคล้องกับการศึกษาครั้งนี้ คือ วิธีการทำนาย
(Predicting) หรือวิธีการฉายภาพ (Projection) เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มี
คุณลักษณะต่างกัน ในการศึกษาครั้งนี้ผู้วิจัยจะใช้คำว่า การเชื่อมโยงคะแนนด้วยวิธีการทำนาย
โดยมีรายละเอียดเกี่ยวกับการเชื่อมโยงคะแนนด้วยวิธีการทำนายเพิ่มเติมดังนี้

ความเป็นมาของการเชื่อมโยงคะแนนด้วยวิธีการทำนาย

การทำนายเป็นรูปแบบการเชื่อมโยงคะแนนที่เก่าแก่ที่สุดและเกิดความสับสนกับ
การเชื่อมโยงคะแนนวิธีอื่นตั้งแต่เริ่มมีการจัดประเภท เริ่มตั้งแต่ศตวรรษที่ 19 Legendre, Gauss,
Laplace และนักวิทยาศาสตร์ทั้งหลายเข้าใจวิธีกำลังสองน้อยสุด (Least squares method) เพื่อ
ตรวจสอบความพอดีของเส้นโค้งในการแก้ปัญหาทางดาราศาสตร์ จนตอนท้ายศตวรรษที่ 19
วิธีการถดถอยเชิงเส้นตรง (Linear regression method) มีการนำไปประยุกต์ใช้กับสถานการณ์
ทางสังคมและจิตวิทยาได้เป็นอย่างดี ซึ่ง Galton (1888 cite in Stigler, 1986) เป็นคนแรกที่

สังเกตเห็นผลกระทบของการถดถอยกับค่าเฉลี่ย ดังนั้นการใช้วิธีการถดถอยเชิงเส้นตรงที่จะทำนายคะแนนแบบสอบหนึ่งหรือการวัดจากวิธีการอื่น ๆ มีความเป็นไปได้ที่จะเก่าแก่ที่สุดสำหรับการเชื่อมโยงคะแนน

รูปแบบของการเชื่อมโยงคะแนนด้วยวิธีการทำนาย

รูปแบบการเชื่อมโยงคะแนนด้วยวิธีการทำนายของคะแนนแบบสอบ Y จากคะแนนแบบสอบ X มี 2 รูปแบบ คือ

1. การทำนายคะแนนสังเกตได้ (Predicting observed score)

วัตถุประสงค์ของการทำนายคือ การทำนายคะแนนของผู้สอบของแบบสอบหนึ่งจากข้อมูลอื่น ๆ ของผู้สอบ ซึ่งข้อมูลอื่น ๆ อาจหมายถึงคะแนนจากแบบสอบอื่นเพียงฉบับเดียวหรือหลาย ๆ ฉบับ และอาจจะรวมไปถึงคุณลักษณะประชากรหรือสารสนเทศอื่นที่เกี่ยวข้องกับผู้สอบ ดังนั้นวิธีการนี้จึงไม่มีความสมมาตรระหว่างสิ่งที่ทำนายกับสิ่งที่ถูกนำมาทำนาย การทำนายและคุณภาพการทำนายอาจจะต่างกันทั้งในด้านจำนวนและคุณลักษณะ เห็นได้ชัดจากการทำนายคะแนนแบบสอบ Y จากแบบสอบ X ในกรณีนี้ เป็นที่รู้จักกันในศตวรรษที่ 19 ว่าฟังก์ชันการถดถอยเชิงเส้นตรงสำหรับการทำนายคะแนน Y จาก X ไม่ผูกผันกับฟังก์ชันการถดถอยเชิงเส้นตรงสำหรับการทำนายคะแนน X จาก Y (Galton, 1888) ค่าคาดหวังอย่างมีเงื่อนไข (หรือเฉลี่ยอย่างมีเงื่อนไข) ของการทำนาย Y จาก X คือ

$$E(Y | X=x, P)$$

โดย X และ Y คือคะแนนจากแบบสอบสองฉบับของผู้สอบที่มาจากประชากร (P) ถ้า X มีค่าเท่ากับ x แล้ว $y = E(Y | X=x, P)$ ทำนาย y จากค่าของ Y การทำนาย Y จาก X เป็นการเชื่อมโยงคะแนนทางตรงระหว่างคะแนนของแบบสอบสองฉบับ จะยกเว้นก็ต่อเมื่อ Y ขึ้นอยู่กับ X นั่นคือ เกิดความคลาดเคลื่อนหรือความไม่แน่นอนในการทำนาย ความคลาดเคลื่อนในการทำนายนี้เป็นความแตกต่างระหว่าง $E(Y | X=x, P)$ กับ Y ที่เกิดขึ้นจริง

$$Y - E(Y | X=x, P)$$

ค่าคาดหวังอย่างมีเงื่อนไขเป็นตัวทำนายที่ดีที่สุดของ Y ในแง่การทำนายอื่น ๆ ของ Y จาก X หรือที่เรียกว่า $y = \text{Pred}(x)$ ที่มีค่าคาดหวังของความคลาดเคลื่อนยกกำลังสอง (Expected squared error) หรือความแปรปรวนของค่าคาดหวังขนาดใหญ่ ดังสมการ

$$E[(Y - \text{Pred}(x))^2 | X=x, P] \geq E[(Y - E(Y | X=x, P))^2 | X=x, P] = \text{Var}(Y | X=x, P)$$

การใช้วิธีการถดถอย ทั้งค่าคาดหวังอย่างมีเงื่อนไข และความคลาดเคลื่อนของการทำนายอย่างมีเงื่อนไขสามารถประมาณค่าได้จากข้อมูลของผู้สอบที่สุ่มมาจากประชากร และสอบทั้งแบบสอบ X และแบบสอบ Y และวิธีการถดถอยสามารถใช้ได้อย่างกว้างขวาง แม้ว่าไม่มีรายละเอียดมากพอเกี่ยวกับความหลากหลายของความเป็นได้ (Holland, 2007, pp.7-9)

2. การทำนายคะแนนจริง (Predicting true score)

การทำนายคะแนนจริงจากคะแนนสังเกตได้ที่เก่าแก่ที่สุดเป็นการคำนวณด้วยสูตรของ Kelley (1927) แนวคิดนี้ได้ถูกนำมาใช้ในงานของ Wainer et al. (2001) เกี่ยวกับการทำนายคะแนนจริงแบบสอบหนึ่งจากคะแนนสังเกตได้จากแบบสอบนั้นและจากการทดสอบอื่น ๆ ซึ่งเรียกรการทำนายคะแนนจริงว่า คะแนนขยาย (Augmented score) ต่อมา Holland and Hoskens (2003) พิจารณาปัญหาการทำนายคะแนนจริงจากคะแนนสังเกตได้ เมื่อคะแนนจริงมาจากแบบสอบ Y และคะแนนสังเกตได้มาจากแบบสอบ X โดยใช้ฟังก์ชันการถดถอยเชิงเส้นตรง ซึ่งทำนายคะแนนสังเกตได้ของแบบสอบ Y จากคะแนนสังเกตได้ของแบบสอบ X เป็นการทำนายที่เหมาะสมของคะแนนจริงจากแบบสอบ Y แต่โดยปกติแล้วความคลาดเคลื่อนของการทำนายจากการถดถอยเชิงเส้นตรงมีมากด้วย และต้องมีการปรับด้วยความเชื่อมั่นของแบบสอบทั้งสองฉบับ (Holland, 2007, p. 10)

การออกแบบการเชื่อมโยงคะแนน

ในความเป็นจริงแล้วการออกแบบการเชื่อมโยงคะแนน จะมีวิธีการออกแบบการเก็บรวบรวมข้อมูล และวิธีการทางสถิติหาความสัมพันธ์ระหว่างคะแนนสองแบบสอบไม่ต่างจากการออกแบบการปรับเทียบคะแนน ดังนั้นสองค่านี้สามารถใช้แทนกันได้ในกระบวนการของคะแนนเชื่อมโยงระหว่างแบบสอบหรือรูปแบบข้อสอบที่ต่างกัน (Tian, 2011, pp.6) ดังนั้นในการวิจัยนี้จะใช้คำว่า การเชื่อมโยงคะแนน

1. การออกแบบการเก็บรวบรวมข้อมูล

การออกแบบการเก็บรวบรวมข้อมูลในการเชื่อมโยงคะแนน จะมีลักษณะทั่วไปเหมือนกับการออกแบบการปรับเทียบคะแนน จากการศึกษาเอกสารเกี่ยวกับการออกแบบการเก็บรวบรวมข้อมูลนั้น ผู้วิจัยสรุปรูปแบบได้ดังนี้

ตารางที่ 2-2 สรุปรูปแบบการออกแบบการเก็บรวบรวมข้อมูล

ลักษณะ แบบสอบ	กลุ่มผู้สอบ							
	กลุ่มเดี่ยว		สองกลุ่มเท่าเทียม		สองกลุ่มไม่เท่าเทียม		สองกลุ่มมีผู้สอบรวม	
	สองฉบับเท่าเทียม	สองฉบับต่างกัน	หนึ่งฉบับต่างกัน	สองฉบับ สลับกัน	ข้อสอบรวม	ข้อสอบรวมภายใน	ข้อสอบรวมภายนอก	หนึ่งฉบับต่างกัน
1	/	/				/		
2	/							
3		/		/		/		
4		/	/	/		/		
5		/	/		/	/	/	/

1 = Kolen and Brennan, 2014, pp. 12-21

2= Grant, 2006 cited in Puhan et al., 2009

3 = Crocker and Algina, 2008, pp. 457-458

4 = von Davier, Holland and Thayer, 2004, pp.19-43

5 = Load, 1975b; Angoff, 1984; Petersen et al., 1989 (อ้างถึงใน ศิริชัย กาญจนวาสี, 2555, หน้า 163-166)

รายละเอียดของการออกแบบการปรับเทียบคะแนนในแต่ละรูปมีดังนี้

1. การออกแบบกลุ่มผู้สอบเดี่ยวทำแบบสอบสองฉบับเท่าเทียมกัน

ผู้สอบ

ลำดับการทำแบบสอบฉบับ

กลุ่ม 1

A, A'

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะมีผู้สอบเพียงกลุ่มเดียว ทำแบบสอบสองฉบับ ซึ่งแบบสอบสองฉบับมีลักษณะที่สมมูลกัน กล่าวคือ ข้อสอบของแบบสอบทั้งสองฉบับได้รับการสุ่มมาจากประชากรเนื้อหาเดียวกัน และแบบสอบทั้งสองฉบับจะมีข้อสอบที่ร่วมกันมากกว่า 80% ของข้อสอบทั้งหมด

2. การออกแบบกลุ่มผู้สอบเดี่ยวทำแบบสอบสองฉบับต่างกัน

รูปแบบ	ผู้สอบ	ลำดับการทำแบบสอบฉบับ
ไม่สมดุล	1 กลุ่ม	A , B
สมดุล	สุ่มเป็น กลุ่มย่อย 1	A , B
	กลุ่มย่อย 2	B , A

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะมี 2 ลักษณะคือ 1) รูปแบบที่ผู้สอบไม่ได้รับการจัดกลุ่มให้สมดุล เป็นการจัดให้ผู้สอบแต่ละคนทำแบบสอบทั้งสองฉบับ โดยทำแบบสอบฉบับหนึ่งแล้วตามด้วยแบบสอบอีกหนึ่งฉบับ แต่รูปแบบนี้อาจจะมีอิทธิพลต่อการเปรียบเทียบคะแนน เพราะเป็นผลที่เกิดมาจากการทำแบบสอบฉบับแรกส่งผลต่อการทำแบบสอบฉบับสอง เช่น ความเมื่อยล้า การเรียนรู้ และ 2) รูปแบบที่ผู้สอบได้รับการจัดกลุ่มให้สมดุล เป็นการสุ่มผู้สอบแยกออกเป็น 2 กลุ่มย่อย แล้วแต่ละกลุ่มย่อยก็ทำแบบสอบทั้ง 2 ฉบับที่สลับกัน กล่าวคือ กลุ่มย่อย 1 ทำแบบสอบฉบับที่ A ก่อนแล้วตามด้วยแบบสอบฉบับที่ B ส่วนกลุ่มย่อยที่ 2 ทำแบบสอบฉบับที่ B ก่อนแล้วตามด้วยฉบับที่ A ซึ่งการใช้รูปแบบที่ 2 นี้จะได้รับอิทธิพลจากลำดับการทดสอบ การเรียนรู้ ความเมื่อยล้าที่สมดุลกันทั้งสองกลุ่ม

3. การออกแบบกลุ่มผู้สอบสองกลุ่มเท่าเทียมกันทำแบบสอบหนึ่งฉบับต่างกัน

ผู้สอบ	ลำดับการทำแบบสอบฉบับ
กลุ่ม 1	A
กลุ่ม 1'	B

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้ เป็นการจัดกลุ่มผู้สอบให้มีความเท่าเทียมกันหรือใกล้เคียงกันมากที่สุด (Equivalent group) แล้วให้ผู้สอบแต่ละกลุ่มจะได้รับแบบสอบเพียงฉบับเดียว ซึ่งอาจจะได้มาโดยการสุ่ม วิธีนี้อาจจะมีปัญหาเกี่ยวกับการแจกแจงความสามารถที่ต่างกันของกลุ่มผู้สอบ แต่แก้ไขได้โดยใช้กลุ่มตัวอย่างที่มีขนาดใหญ่

4. การออกแบบกลุ่มผู้สอบสองกลุ่มเท่าเทียมกันทำแบบสอบสองฉบับสลับที่กัน

ผู้สอบ	ลำดับการทำแบบสอบฉบับ
กลุ่ม 1	A , B
กลุ่ม 1'	B , A

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะคล้ายกับรูปแบบที่ 3 ซึ่งเป็นการจัดกลุ่มผู้สอบให้มีความเท่าเทียมกันหรือใกล้เคียงกันมากที่สุด (Equivalent group) เหมือนกัน แต่จะต่างที่ผู้สอบแต่ละกลุ่มจะได้รับแบบสอบสองฉบับ แต่สลับฉบับการทำก่อนหลัง ซึ่งอาจจะได้มาโดยการสุ่ม คือ กลุ่มที่ 1 ทำแบบสอบฉบับที่ A ก่อนแล้วตามด้วยแบบสอบฉบับที่ B ส่วนกลุ่มที่ 2 ทำแบบสอบฉบับที่ B ก่อนแล้วตามด้วยฉบับที่ A วิธีนี้จะช่วยลดปัญหาเกี่ยวกับการแจกแจง

ความสามารถที่ต่างกันของกลุ่มผู้สอบได้

5. การออกแบบกลุ่มผู้สอบสองกลุ่มเท่าเทียมกันทำข้อสอบร่วม

ผู้สอบ	แบบสอบร่วมภายใน	แบบสอบร่วมภายนอก
กลุ่ม 1	AC	A , C
กลุ่ม 1'	BC	B , C

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะสุ่มผู้สอบจำนวนสองกลุ่ม ที่มี
ความสามารถเท่าเทียมกัน จากนั้นให้ทำแบบสอบร่วมโดยมี 2 กรณี ดังนี้ กรณีที่ 1 การใช้แบบสอบ
ร่วมภายใน นั่นคือ ผู้กลุ่มแต่ละกลุ่มจะทำแบบสอบเพียงฉบับเดียวที่ต่างกัน แต่แบบสอบทั้งสอง
ฉบับนั้นจะมีข้อบางส่วนที่เหมือนกันหรือมีข้อสอบร่วมกัน ส่วนกรณีที่ 2 การใช้แบบสอบร่วม
ภายนอก นั่นคือ ผู้สอบจะต้องทำแบบสอบสองฉบับโดยที่แต่ละกลุ่มจะทำแบบสอบที่เหมือนกัน
หนึ่งฉบับ และต่างกันหนึ่งฉบับ ซึ่งวิธีการนี้จะมีอิทธิพลของจำนวนข้อสอบร่วม คุณภาพของ
ข้อสอบร่วม ความสัมพันธ์ระหว่างข้อสอบร่วมกับแบบสอบที่ต้องการ

6. การออกแบบกลุ่มผู้สอบสองกลุ่มไม่เท่าเทียมกันทำข้อสอบร่วมภายใน

ผู้สอบ	ลำดับการทำแบบสอบฉบับ
กลุ่ม 1	AC
กลุ่ม 2	BC

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะเก็บข้อมูลจากผู้สอบสองกลุ่ม ที่มาจาก
ประชากรต่างกัน นั่นคือ ผู้สอบมีความสามารถไม่เท่าเทียมกัน จากนั้นให้ทำแบบสอบกลุ่มละฉบับ
โดยภายในแต่ละฉบับจะมีข้อสอบจำนวนหนึ่งที่เหมือนกันหรือมีข้อสอบร่วมที่ใช้ร่วมกันระหว่าง
แบบสอบสองฉบับ ซึ่งวิธีการนี้จะมีอิทธิพลของจำนวนข้อสอบร่วม คุณภาพของข้อสอบร่วม
ความสัมพันธ์ระหว่างข้อสอบร่วมกับแบบสอบที่ต้องการ และความสามารถของผู้สอบเข้ามา
เกี่ยวข้อง

7. การออกแบบกลุ่มผู้สอบสองกลุ่มไม่เท่าเทียมกันทำข้อสอบร่วมภายนอก

ผู้สอบ	ลำดับการทำแบบสอบฉบับ
กลุ่ม 1	A,C
กลุ่ม 2	B,C

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะมีลักษณะที่เหมือนกับรูปแบบที่ 5 จะ
ต่างกันตรงที่จำนวนของแบบสอบ ซึ่งวิธีนี้ผู้สอบจะต้องทำแบบสอบสองฉบับโดยที่แบบสอบฉบับ
หนึ่งผู้สอบทั้งสองกลุ่มจะต้องทำเหมือนกัน และอีกฉบับหนึ่งต่างกัน

8. การออกแบบกลุ่มผู้สอบสองกลุ่มมีผู้สอบร่วมทำแบบสอบหนึ่งฉบับต่างกัน

ผู้สอบ	ลำดับการทำแบบสอบฉบับ
กลุ่ม 1,3	A
กลุ่ม 2,3	B

การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้จะมีลักษณะพิเศษต่างจากรูปแบบอื่น ๆ คือ ผู้สอบที่แบ่งออกเป็นสองกลุ่ม แล้วทำแบบทดสอบต่างฉบับกัน แต่ในกลุ่มผู้สอบทั้งสองกลุ่มนั้น จะมีผู้สอบร่วมกันจำนวนหนึ่ง ดังนั้น ผู้สอบที่อยู่ในทั้งสองกลุ่มจะต้องทำแบบสอบทั้งสองฉบับ ซึ่งรูปแบบนี้จะมีอิทธิพลของความเมื่อยล้า การเรียนรู้จากแบบสอบของผู้สอบร่วม

จากการออกแบบการเก็บรวบรวมข้อมูลทั้ง 8 รูปแบบนี้ พบว่า รูปแบบที่ 2) การออกแบบกลุ่มผู้สอบเดี่ยวทำแบบสอบสองฉบับต่างกัน มีลักษณะไม่สมดุล สอดคล้องกับข้อมูลของการวิจัยครั้งนี้ คือ นักเรียน 1 คน ทำแบบสอบสองฉบับ คือ แบบสอบวิชาภาษาไทย และแบบสอบวิชาภาษาอังกฤษโดยมีลำดับการทำแบบสอบที่เหมือนกัน

การออกแบบวิธีการเชื่อมโยงคะแนน

การออกแบบวิธีการเชื่อมโยงคะแนนระหว่างแบบสอบมีเทคนิควิธีการที่หลากหลาย ขึ้นอยู่กับรูปแบบที่นำมาใช้ในการเชื่อมโยง Kolen and Brennan (2014, pp. 29-45, 171-201) สรุปรูปแบบการเชื่อมโยงคะแนนยึดหลักการตามทฤษฎีการวัด 2 รูปแบบใหญ่ ๆ คือ 1) การเชื่อมโยงคะแนนโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) ซึ่งเป็นการนำคะแนนสังเกตได้มาใช้โดยตรง มีเทคนิควิธีการเปรียบเทียบ 3 วิธี คือ เทคนิควิธีการเปรียบเทียบคะแนนอิควีเปอร์เซ็นต์ไทล์ วิธีการเปรียบเทียบคะแนนเชิงเส้นตรง และเทคนิควิธีการเปรียบเทียบคะแนนโดยใช้ค่าเฉลี่ย และ 2) รูปแบบการเปรียบเทียบคะแนนโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item response theory) เป็นการนำค่าความสามารถ ที่ได้จากการประมาณค่ามาประยุกต์ใช้ ประกอบด้วยเทคนิควิธีการเปรียบเทียบคะแนนด้วยเทคนิควิธีโค้งลักษณะข้อสอบ วิธีค่าเฉลี่ยและซิกมา วิธีปรับค่าเฉลี่ยและซิกมา วิธีสมการถดถอย วิธีการปรับเทียบคะแนนจริง วิธีการปรับเทียบคะแนนสังเกตได้ เป็นต้น

การเชื่อมโยงคะแนนที่ยึดหลักการวัดตามทฤษฎีการตอบสนองข้อสอบมีข้อได้เปรียบในทางปฏิบัติมากกว่าทฤษฎีการทดสอบแบบดั้งเดิม (Cook & Eignor, 1983) คือ 1) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบให้ผลการเชื่อมโยงคะแนนที่ดีกว่าการเชื่อมโยงคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม กล่าวคือ คะแนนเต็มของมาตรฐานวัดมักจะมีส่วนนำมาใช้ในการการตัดสินใจเป็นสำคัญ ในขณะที่การเชื่อมโยงคะแนนด้วยทฤษฎีการทดสอบนั้นสามารถใช้ได้เฉพาะข้อมูลคะแนนที่ได้รับจริงของผู้สอบเท่านั้น 2) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนอง

ข้อสอบจะมีความยืดหยุ่นมากกว่า เพราะว่าสามารถเชื่อมโยงคะแนนระหว่างแบบสอบที่มีสเกลเดียวกัน แบบสอบใหม่สามารถปรับได้ทั้งบางส่วนหรือทั้งหมดของแบบสอบเก่า 3) การเชื่อมโยงข้อทำได้ง่าย ถ้าข้อสอบถูกตัดทิ้ง หรือจำนวนข้อสอบน้อยลงสามารถปรับได้ง่ายบนพื้นฐานของสารสนเทศข้อสอบจากข้อสอบที่มีอยู่ และ 4) การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบของการปรับเทียบก่อน (Pre-equating) มีความเป็นไปได้ โดยที่ Pre-equating เป็นการเตรียมความพร้อมตารางแปลงคะแนนดิบไปเป็นคะแนนมาตรฐาน ก่อนที่มีการทำข้อสอบ จะเป็นไปได้เมื่อข้อมูลข้อสอบรายข้อเป็นตัวแปรที่มีอยู่และสามารถปรับสเกลได้ พารามิเตอร์จะไม่แปรเปลี่ยน เมื่อถูกประยุกต์ใช้กับกลุ่มใหม่ นอกจากนี้ แบบทดสอบที่ใช้ในการเชื่อมโยงคะแนนเป็นได้ทั้งลักษณะของแบบสอบคู่ขนาน และแบบสอบไม่คู่ขนาน (Kolen, 1981 cited in Tian, 2011, p.10)

ดังนั้นการศึกษาค้นคว้าวิจัยใช้รูปแบบการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ เพราะเป็นรูปแบบที่ให้ผลลัพธ์มีความเสมอภาค สมมาตร และไม่แปรผันตามกลุ่มผู้สอบ อีกทั้งสามารถใช้ได้ทั้งแบบสอบที่คู่ขนานและไม่คู่ขนาน โดยมีเทคนิควิธีการที่ใช้ในการเชื่อมโยงคะแนน 2 เทคนิค คือ การเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ มีรายละเอียดดังนี้

1. การเชื่อมโยงคะแนนจริง

การเชื่อมโยงคะแนนโดยใช้คะแนนจริง เป็นการเชื่อมโยงคะแนนที่แปลงคะแนนความสามารถ (θ) ให้อยู่ในรูปของคะแนนจริง (True score) โดยมีวิธีการแปลงคะแนนดังนี้ (Crocker & Algina, 2008, p. 472)

$$T = \sum_{i=1}^N P_i(\theta) \quad ; i \text{ คือ จำนวนข้อสอบ}$$

หลังจากแปลงคะแนนความสามารถเป็นคะแนนจริงเรียบร้อยแล้วก็นำคะแนนจริงที่ได้มาเชื่อมโยงคะแนนระหว่างแบบสอบด้วยเทคนิควิธีการต่าง ๆ (Kolen & Brennan, 2014, pp.192-194)

การการเชื่อมโยงคะแนนนั้น แบบสอบทั้งสองฉบับจะต้องแปลงพารามิเตอร์ข้อสอบให้อยู่บนสเกลเดียวกันโดยการคำนวณค่าคงที่ 2 ค่า คือ ค่า A และค่า B เพื่อใช้ในการปรับพารามิเตอร์ข้อสอบให้อยู่ในสเกลเดียวกัน มีรายละเอียดดังนี้ (Kolen & Brennan, 2014)

$$A = \frac{\sigma(b_T)}{\sigma(b_E)}$$

$$B = \mu(b_T) - A * \mu(b_E)$$

เมื่อ $\sigma(b_T)$ คือ ส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของวิชาภาษาไทย
 $\sigma(b_E)$ คือ ส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของวิชาภาษาอังกฤษ
 $\mu(b_T)$ คือ ค่าเฉลี่ยของพารามิเตอร์ความยากของวิชาภาษาไทย
 $\mu(b_E)$ คือ ค่าเฉลี่ยของพารามิเตอร์ความยากของวิชาภาษาอังกฤษ

1.1 วิธีโค้งคุณลักษณะของแบบสอบ (Test characteristic curves)

คะแนนจริงของแบบสอบฉบับ X ที่สมมูลกับ θ , มีฟังก์ชันคือ

$$T_x = \sum_{j:x} P_{ij}(\theta; a_j, b_j, c_j)$$

คะแนนจริงของแบบสอบฉบับ Y ที่สมมูลกับ θ , มีฟังก์ชันคือ

$$T_y = \sum_{j:y} P_{ij}(\theta; a_j, b_j, c_j)$$

เมื่อ i คือ จำนวนผู้เข้าสอบ ส่วน j คือ จำนวนข้อสอบ ซึ่งฟังก์ชันของทั้งสองแบบสอบนี้เรียกว่า โค้งคุณลักษณะของแบบสอบ โดยโค้งคุณลักษณะของข้อสอบจะเกี่ยวข้องกับความสามารถใน IRT ของคะแนนจริง

สำหรับการใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติกส์ 3 พารามิเตอร์นั้น จะไม่เหมาะสมเมื่อคะแนนจริงมีค่าน้อย เพราะค่า θ ที่เข้าใกล้ $-\infty$ ความน่าจะเป็นในการตอบข้อสอบถูกเข้าใกล้ค่า c และไม่เท่ากับ 0 ดังนั้นคะแนนจริงที่มีความสัมพันธ์กับ θ จะอยู่ในช่วง ต่อไปนี้

$$\sum_{j:x} c_j < T_x < K_x \quad \text{และ} \quad \sum_{j:y} c_j < T_y < K_y \quad ; K \text{ คือ จำนวนข้อสอบ}$$

ซึ่งการเชื่อมโยงที่กล่าวข้างต้นไม่สามารถตรวจสอบความถูกต้องได้อย่างแน่ชัด จึงมีนักศึกษาเพื่อตอบโจทย์ส่วนนี้ เช่น Haebara (1980) นำเสนอการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะของข้อสอบ Stocking and Lord (1983) ได้นำวิธีการของ Haebara มาประยุกต์เป็นการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะของแบบสอบ (Kolen & Brannan, 2014, pp. 184-185) ซึ่งสมการสำหรับการเชื่อมโยงของ Stocking and Lord มีดังนี้

$$\sum_{j-v} P_{ij}(\theta_{ji}; a_{ji}; b_{ji}; c_{ji}) = \sum_{j-v} P_{ij} \left(\theta_{ji}; \frac{a_{li}}{A}; Ab_{li} + B; c_{li} \right)$$

และตรวจสอบคุณภาพของการเชื่อมโยงโดยสมการ

$$SLdiff(\theta_i) = \left[\sum_{j-v} P_{ij}(\theta_{ji}; a_{ji}; b_{ji}; c_{ji}) - \sum_{j-v} P_{ij} \left(\theta_{ji}; \frac{a_{li}}{A}; Ab_{li} + B; c_{li} \right) \right]^2$$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j
 θ_{ji} คือ พารามิเตอร์ความสามารถของคนี่ j ในแบบสอบฐาน
 a_{ji} คือ พารามิเตอร์อำนาจจำแนกของข้อสอบ i ในแบบสอบฐาน
 b_{ji} คือ พารามิเตอร์ความยากของข้อสอบ i ในแบบสอบฐาน
 c_{ji} คือ พารามิเตอร์การเดาของข้อสอบ i ในแบบสอบฐาน
 a_{li} คือ พารามิเตอร์อำนาจจำแนกของข้อสอบ i ในแบบสอบเชื่อมโยง
 b_{li} คือ พารามิเตอร์ความยากของข้อสอบ i ในแบบสอบเชื่อมโยง
 c_{li} คือ พารามิเตอร์การเดาของข้อสอบ i ในแบบสอบเชื่อมโยง

สมการการเชื่อมโยงดังกล่าวใช้สำหรับแบบทดสอบที่ไม่คำนึงถึงอิทธิพลของเทสต์เลท Li, Bolt, and Fu (2005) ได้นำสมการของ Stocking and Lord มาพัฒนาต่อยอดเพื่อใช้เชื่อมโยงแบบสอบที่มีลักษณะของเทสต์เลท สำหรับโมเดลสองพารามิเตอร์ได้สมการดังนี้

$$\sum_{j-v} P_{ij}(\theta_{ji}; a_{ji}; b_{ji}; \gamma_{jd(j)}) = \sum_{j-v} P_{ij} \left(\frac{a_{li}}{A}; \xi_{jd} - Ab_{li} + A * \mu_{\gamma_d} - B \right)$$

โดยที่ $\xi_{jd} = \theta_j - \gamma_{jd(j)}$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j
 $\gamma_{jd(j)}$ คือ อิทธิพลสุ่มของคนี่ j กับข้อสอบ i ภายใต้เทสต์เลท d จากแบบสอบฐาน
 μ_{γ_d} คือ ค่าเฉลี่ยอิทธิพลสุ่มของเทสต์เลท d จากแบบสอบฐาน
 A และ B คือ ค่าคงที่

1.2 กระบวนการปรับเทียบด้วยคะแนนจริง (True score equating process)

การปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบ กรณีที่ให้ค่า θ เดียวกัน จะถือว่าคะแนนจริงของแบบสอบ X ($T_x(\theta_i)$) และคะแนนจริงจากแบบสอบ Y ($T_y(\theta_i)$) สมมูลกันหรือเท่าเทียมกัน ดังนั้นคะแนนจริงจากแบบสอบ Y จะเท่าเทียมกับคะแนนจริงของแบบสอบ X เมื่อ

$$IRT_Y(T_x) = T_y(T_x^{-1}); \quad \sum_{j:X} c_j < T_x < K_x$$

โดยที่ T_x^{-1} แทน θ ที่สอดคล้องกับ T_x และมีขั้นตอนการปรับเทียบคะแนนจริง 3 ขั้นตอน คือ

- 1) กำหนดคะแนนจริง (T_x) ของแบบสอบ X
- 2) ค้นหา θ ที่สอดคล้องกับคะแนนจริง (T_x^{-1})
- 3) ค้นหาคะแนนจริง (T_y) ของแบบสอบ Y ที่สอดคล้องกับ θ ในข้อ 2

จากขั้นตอนที่ 3 นี้จะทำให้คะแนนจริงจากแบบสอบ Y สมมูลกับแบบสอบ X

1.3 วิธีการของ Newton-Raphson (The newton-raphson method)

วิธีการของ Newton-Raphson เป็นวิธีการทั่วไปสำหรับการหารากของฟังก์ชันที่ไม่เป็นเส้นตรง การใช้วิธีนี้ เริ่มต้นด้วยฟังก์ชันที่ถูกกำหนดให้เท่ากับ 0 โดย $\text{func}(\theta)$ ซึ่งเป็นฟังก์ชันของตัวแปร θ ที่ถูกหารด้วย $\text{func}'(\theta)$ และค่าเริ่มต้นสำหรับค่า θ คือ ค่า θ^- โดยสูตรคำนวณของวิธีนี้เพื่อที่จะได้ค่าใหม่ของ θ คือ

$$\theta^+ = \theta^- - \frac{\text{func}(\theta)}{\text{func}'(\theta)}$$

θ^+ จะใกล้เคียงกับรากฐานของสมการมากกว่า θ^- ค่าใหม่นั้นถูกนิยามเป็น θ^- กระบวนการจะถูกทำซ้ำจนกว่า θ^+ และ θ^- จะเท่ากับค่าความแม่นยำที่กำหนด หรือจนกระทั่งค่าเข้าใกล้ 0 ที่ระดับความแม่นยำที่กำหนด ดังนั้นการเลือกวิธีการกำหนดค่าเริ่มต้นมีความสำคัญ เพราะถ้าเลือกไม่ดีก็จะนำไปสู่การแก้ปัญหาที่ผิดพลาดได้ (Press et al, 1989 cited in Kolen & Brennan, 2014, p. 194) ซึ่งวิธีของ Newton-Raphson ที่มีการดัดแปลง จะมีประสิทธิภาพมากกว่าวิธีที่เลือกค่าเริ่มต้นไม่ดี

การใช้วิธี Newton-Raphson ในการปรับเทียบ IRT เริ่มต้นจากนำ T_x เป็นคะแนนจริงที่พบว่าสมมูลหรือเทียบเท่า ลบด้วย θ ณ ที่คะแนนนั้น เท่ากับ 0 ดังสมการ

$$func(\theta_i) = T_x - \sum_{j:K} P_{ij}(\theta_i; a_j, b_j, c_j)$$

จากนั้นหาค่าตัวหารจากสมการ

$$func'(\theta_i) = -\sum_{j:K} P_{ij}'(\theta_i; a_j, b_j, c_j)$$

$$\text{โดยที่ } P_{ij}'(\theta_i; a_j, b_j, c_j) = \frac{1.7a_j(1-P_{ij})(P_{ij}-c_j)}{(1-c_j)}$$

2. การเชื่อมโยงคะแนนสังเกตได้

การเชื่อมโยงคะแนนสังเกตได้ (Observed score) หรือคะแนนดิบ (Raw score) โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการประมาณค่าการแจกแจงของคะแนนสังเกตได้ที่ถูกต้อง (Number-correct score) ของแบบสอบแต่ละฉบับ จากนั้นนำข้อมูลที่ได้มาเชื่อมโยงคะแนน ซึ่งการเชื่อมโยงคะแนนโดยใช้คะแนนสังเกตได้ (Γ) เป็นคะแนนที่อยู่บนสเกลเดียวกับคะแนนจริง (ξ) Kolen and Brennan (2014, pp. 197-200) โดยคำนวณได้จากสูตร

$$\Gamma = \sum_{j=1}^n U_j$$

เมื่อ $U_j =$ คะแนนรายข้อ และ $E(\Gamma) = \xi$

เนื่องจากทฤษฎีการตอบสนองข้อสอบ ได้ให้แนวคิดทางทฤษฎีในการทำนายการแจกแจงคะแนนที่สังเกตได้จากแบบสอบ เมื่อนำแบบสอบสองฉบับไปใช้ ทำให้ทราบการแจกแจงของคะแนนสังเกตได้ของแต่ละฉบับ ซึ่งแบบสอบทั้งสองฉบับไม่จำเป็นต้องมีความยากเท่าเทียมกัน การแจกแจงทางทฤษฎีของคะแนนสังเกตได้ของผู้สอบที่มีความสามารถ θ จากการทำแบบสอบฉบับหนึ่ง สามารถแสดงได้ดังนี้

$$\sum_{r=0}^n f(r/\theta)t^r = \prod_{i=1}^n [Q_i(\theta) - tP_i(\theta)]$$

เมื่อ $\sum_{r=0}^n f(r/\theta)t^r$ เป็นผลรวมความถี่สัมพัทธ์อย่างมีเงื่อนไข

$\prod_{i=1}^n [Q_i(\theta) - tP_i(\theta)]$ เป็นความน่าจะเป็นของการเกิดการแจกแจงคะแนนดิบ

ศิริชัย กาญจนวาสี (2555, หน้า 178-182) ได้สรุปขั้นตอนสำหรับการเชื่อมโยงคะแนนสังเกตได้ไว้ ดังนี้

2.1 นำข้อมูลของผู้สอบจากการทำแบบสอบมาวิเคราะห์เพื่อประมาณ

ค่าพารามิเตอร์ของข้อสอบและผู้สอบ แล้วนำค่า θ ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีการแจกแจงของคะแนนสังเกตได้ $f(\Gamma/\theta)$

2.2 แปลงค่า θ ด้วยสูตร $T(\theta) = \alpha\theta + \beta$ แล้วนำค่า $T(\theta)$ ของผู้สอบมาคำนวณความถี่ตามคะแนนสังเกตได้ทางทฤษฎี $f(\Gamma)$ จะได้ข้อมูลการแจกแจงความถี่ของคะแนนสังเกตได้ทางทฤษฎีของกลุ่มผู้สอบแต่ละฉบับ

2.3 เชื่อมโยงคะแนนระหว่างคะแนนสังเกตได้ของแบบสอบสองฉบับ ประกอบด้วย 4 วิธี คือ วิธีค่าเฉลี่ย วิธีเชิงเส้นตรง วิธีเชิงเส้นตรงคู่ขนาน และวิธีอิกวิเปอร์เซ็นต์ไทล์ (Kolen & Brennan, 2014, pp. 502-504) มีรายละเอียดดังนี้

2.3.1 วิธีค่าเฉลี่ย (Mean method)

สำหรับวิธีค่าเฉลี่ย เป็นการกำหนดให้ค่าเบี่ยงเบนของคะแนนของค่าเฉลี่ยเท่ากัน ดังสมการ

$$x - \mu(X) = y - \mu(Y)$$

จากนั้นแปลงสมการสำหรับประชากรทั้งหมด

$$m_y(x) = x - \mu(X) + \mu(Y)$$

และสมการสำหรับกลุ่มย่อย (h)

$$m_{yh}(x) = x - \mu_h(X) + \mu_h(Y)$$

วิธีค่าเฉลี่ยสามารถใช้ได้ดีในการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีโครงสร้างที่ใกล้เคียงกัน (เนื้อหา) แต่ควรใช้กับกลุ่มตัวอย่างที่มีขนาดใหญ่

2.3.2 วิธีเชิงเส้นตรง (Linear method)

การเชื่อมโยงวิธีนี้มีนิยามว่า คะแนนจากแบบสอบ 2 แบบสอบ คือ แบบสอบฉบับ X และแบบสอบฉบับ Y (วัดสิ่งเดียวกันและมีความเที่ยงสูงพอกัน) จะถือว่าคะแนนจะสมมูลกัน เมื่อคะแนนของแต่ละฉบับมีคะแนนมาตรฐาน (Standard scores) เท่ากัน ซึ่งมีสูตรการคำนวณคะแนนมาตรฐานดังนี้

$$\frac{x-\mu(X)}{\sigma(X)} = \frac{y-\mu(Y)}{\sigma(Y)}$$

สมการแปลงคะแนนจากแบบสอบฉบับ X ให้สมมูลกับแบบสอบฉบับ Y ดังนี้

$$l_Y(x) = \mu(Y) + \frac{\sigma(Y)}{\sigma(X)}[x - \mu(X)]$$

และสมการสำหรับกลุ่มย่อย (h)

$$l_{Yh}(x) = \mu_h(Y) + \frac{\sigma_h(Y)}{\sigma_h(X)}[x - \mu_h(X)]$$

จะเห็นได้ว่าสมการที่ใช้ในการเชื่อมโยงคะแนนระหว่างวิธีค่าเฉลี่ย กับวิธีเชิงเส้นตรงนั้น จะให้ค่าเหมือนกัน เมื่อ $\sigma(Y)/\sigma(X) = 1$ สำหรับทั้งกลุ่ม และ $\sigma_h(Y)/\sigma_h(X) = 1$ สำหรับกลุ่มย่อย

2.3.3 วิธีเชิงเส้นคู่ขนาน (Parallel-linear Method)

วิธีนี้ถูกคิดค้นโดย Dorans and Holland (2000) เป็นการเชื่อมกับกลุ่มหลายกลุ่ม โดยวิธีนี้จะต่างกับวิธีเชิงเส้นตรงในแง่ของสถิติ คือ คะแนนเบี่ยงเบนของแต่ละกลุ่มจะถูกหารด้วย ส่วนเบี่ยงเบนของกลุ่มรวม ดังสมการ

$$\frac{x-\mu_h(X)}{\sigma(X)} = \frac{y-\mu_h(Y)}{\sigma(Y)}$$

และมีสมการแปลงคะแนนสำหรับกลุ่มย่อย (h)

$$pl_{Yh}(x) = \mu_h(Y) + \frac{\sigma(Y)}{\sigma(X)}[x - \mu_h(X)]$$

วิธีนี้จะลดความยุ่งยากในฟังก์ชันการแปลงข้อมูลของกลุ่มย่อย เมื่อค่าคงที่ต่างกัน หรือความชันต่างกัน

2.3.4 วิธีอีควิเปอร์เซ็นต์ไทล์ (Equipercntile method)

วิธีอีควิเปอร์เซ็นต์ไทล์ เป็นวิธีการแปลงคะแนนที่ไม่ใช่เชิงเส้นตรง เป็นการแปลงคะแนนจากแบบสอบ 2 ฉบับ คือ แบบสอบฉบับ X และแบบสอบฉบับ Y จะถือว่าคะแนนจะสมมูล

กัน เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) เดียวกัน ซึ่งคำนวณตำแหน่งเปอร์เซ็นต์ไทล์ได้จากสูตร

$$PR = (cf + \frac{1}{2}f) \frac{100}{N}$$

และคำนวณคะแนนจากแบบสอบ ณ ตำแหน่งเปอร์เซ็นต์ไทล์ที่ต้องการจาก

$$X_{PR} = L_0 + c \left[\frac{\frac{PR \cdot n}{100} - F_b}{f} \right]$$

ซึ่งทั้ง 4 วิธีนี้ วิธีอิควิเปอร์เซ็นต์ไทล์มีข้อดีมากกว่าอีก 3 วิธี คือ 1) ความสมมูลกันของคะแนนอิควิเปอร์เซ็นต์ไทล์อยู่ในช่วงของคะแนนจุดตัดที่เป็นไปได้ โดยหลีกเลี่ยงการเกิดปัญหาข้อมูลอยู่นอกช่วง (Out-of-range) ในวิธีค่าเฉลี่ย วิธีเชิงเส้นตรง และวิธีเชิงเส้นตรงคู่ขนาน 2) ความสัมพันธ์ระหว่างคะแนนของแบบสอบสองฉบับไม่มีข้อตกลงเกี่ยวกับความสัมพันธ์เชิงเส้นตรง 3) พังก์ชันการแจกแจงความถี่สะสมในการแปลงคะแนนของแบบสอบฉบับ X เป็นการประมาณค่าโดยฟังก์ชันการแจกแจงความถี่ของคะแนนจากแบบสอบฉบับ Y และ 4) ค่าที่ได้จากคะแนนแปลง (ค่าเฉลี่ย ความแปรปรวน ความเบ้ และความโด่ง) จะมีลักษณะเดียวกับคะแนนของแบบสอบฉบับ Y แต่วิธีการนี้ไม่เหมาะกับกลุ่มตัวอย่างขนาดเล็ก

3. ขั้นตอนการเชื่อมโยงคะแนน

ขั้นตอนการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ ประยุกต์ใช้จากวิธีของศิริชัย กาญจนวาสี (2541, หน้า 34-38) ได้สรุปไว้ดังนี้

ขั้นที่ 1 เลือกรูปแบบการเชื่อมโยงคะแนน (Linking design) ซึ่งขึ้นอยู่กับธรรมชาติของแบบสอบและกลุ่มผู้สอบ

ขั้นที่ 2 กำหนดโมเดลการตอบสนองข้อสอบ (Item response model) ซึ่งจะมี 2 ประเภทหลัก ๆ คือ ประเภทที่มีการตรวจให้คะแนนรายข้อแบบ 2 ค่า และประเภทที่มีการตรวจให้คะแนนรายข้อมากกว่า 2 ค่า ซึ่งการตัดสินใจเลือกโมเดลนั้นขึ้นอยู่กับลักษณะของการเชื่อมโยงคะแนนและขนาดของกลุ่มตัวอย่าง

ขั้นที่ 3 สร้างเมตริกซ์สเกลร่วมของความสามารถและพารามิเตอร์ข้อสอบ ซึ่งทั้งสองค่านี้มีความสัมพันธ์กันในเชิงเส้นตรง โดยการคำนวณค่าคงที่ในสมการสำหรับการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกัน เพื่อเชื่อมโยง θ ให้อยู่บนสเกลที่สามารถเปรียบเทียบกันได้โดยตรง

ขั้นที่ 4 รายงานคะแนนสอบที่เชื่อมโยงแล้ว

จากแนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบผู้วิจัยสามารถสรุปประเด็นเกี่ยวกับการศึกษาครั้งนี้ได้ว่า การเชื่อมโยงคะแนนระหว่างแบบสอบครั้งนี้อยู่ภายใต้หลักการเชื่อมโยงคะแนนแบบการทำนาย (Predicting) เก็บข้อมูลกับผู้สอบกลุ่มเดียวที่ทำแบบสอบสองฉบับที่ต่างกัน ด้วยวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ การเชื่อมโยงคะแนนจริงวิธีโค้งคุณลักษณะแบบสอบ และการเชื่อมโยงคะแนนสังเกตได้วิธีอิควิเปอร์เซ็นต์ไทล์ ซึ่งการเชื่อมโยงคะแนนแต่ละวิธีมีจุดเด่นและด้อยต่างกัน ผลที่ได้ก็จะต่างกัน ด้วยดังนั้นประสิทธิผลของการเชื่อมโยงคะแนนพิจารณาจากค่าใดได้บ้าง ผู้วิจัยศึกษาเพิ่มเติมโดยมีรายละเอียดในหัวข้อถัดไป

ตอนที่ 3 แนวคิดเกี่ยวกับประสิทธิผลของการเชื่อมโยงคะแนน

ประสิทธิภาพของการเชื่อมโยงคะแนนระหว่างแบบสอบ จะมีวิธีการและเกณฑ์ที่ใช้ในการประเมินประสิทธิผลในการปรับเทียบ โดยมีรายละเอียดในการพิจารณาจากค่าต่าง ๆ ดังนี้

1. ความเสมอภาค (Weak equity)

ความเสมอภาค ใช้เป็นเกณฑ์ในการพิจารณาเปรียบเทียบระหว่างคะแนนจากแบบสอบชุดเดิมกับคะแนนที่ปรับแล้ว โดยมีเกณฑ์พิจารณา 2 ข้อ คือ คะแนนจริงของแบบสอบชุดเดิมกับคะแนนที่ปรับแล้ว และการแจกแจงของคะแนนความคลาดเคลื่อนของแบบสอบชุดเดิมและที่ปรับแล้วจะต้องมีลักษณะการแจกแจงที่เหมือนกัน (Lord, 1980, pp.195-196; Crocker & Algina, 2008, pp. 477-478) นอกจากนี้ Hanson (1991) ได้เสนอเกณฑ์ในลักษณะเดียวกันนี้ใช้ชื่อว่า คะแนนคาดหวังที่เท่าเทียม (Equivalent expected score) โดยมีข้อตกลงเบื้องต้นว่า คะแนนจริงของแบบสอบชุดเดิมกับคะแนนจริงที่ปรับแล้ว มีความสัมพันธ์กันเชิงฟังก์ชันดังนี้

$$E[e(X^*)|\gamma(\tau)] = E(x|\tau) = \tau$$

เมื่อ τ แทน คะแนนจริงจากแบบสอบชุดเดิม

$E[e(X^*)|\gamma(\tau)]$ แทน คะแนนจากแบบสอบที่ปรับแล้ว

$E(x|\tau)$ แทน คะแนนจากแบบสอบชุดเดิม

2. ความเพียงพอของการปรับเทียบ (Adequacy of equating model)

ความเพียงพอเป็นดัชนีเพื่อตรวจสอบประสิทธิภาพของวิธีการปรับเทียบรูปแบบหนึ่ง มีวิธีการตรวจสอบหลายวิธี ดังนี้

2.1 ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (Percentile comparison index) พัฒนาโดย Kolen and Witney เมื่อปี ค.ศ. 1982 (ศิริชัย กาญจนวาสี, 2555, หน้า 184; สุรศักดิ์ อมรรัตนศักดิ์, 2544, หน้า 381) เป็นการนำคะแนนจากกลุ่มตัวอย่างสอบทานผล ซึ่งเป็นกลุ่มตัวอย่างที่มาจากประชากรกลุ่มเดียวกับกลุ่มเปรียบเทียบโดยไม่มีหน่วยตัวอย่างซ้ำกัน ได้รับการทดสอบด้วยแบบสอบเทียบคะแนนทั้งสองชุด คือ ฉบับ X และฉบับ Y ให้แบบสอบฉบับ Y เป็นคะแนนเกณฑ์ แล้วนำแบบสอบฉบับ X ให้อยู่บนสเกลคะแนนของ Y (Y^*) จากนั้นทำการเปรียบเทียบคะแนน Y และคะแนนเปรียบเทียบ Y^* ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน ดังสูตร

$$c = \frac{\Sigma(x-x^*)}{nk}$$

เมื่อ c	แทน	ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์
X	แทน	คะแนนจากแบบสอบฉบับ X ซึ่งเป็นคะแนนเกณฑ์
X^*	แทน	คะแนนเปรียบเทียบของแบบสอบฉบับ Y ที่อยู่บนสเกล X ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน
n	แทน	จำนวนคะแนนดิบของกลุ่มสอบทานผล
k	แทน	จำนวนข้อสอบรวม

ถ้าค่า c มีค่าน้อย หมายความว่า รูปแบบการเปรียบเทียบที่นำมาใช้ในการเปรียบเทียบมีความเหมาะสม และเพียงพอที่จะให้ผลการเปรียบเทียบมีความน่าเชื่อถือ

2.2 ดัชนีความแตกต่าง (Discrepancy index) พัฒนาโดย Petersen et al. เมื่อปี ค.ศ.1982 (สุรศักดิ์ อมรรัตนศักดิ์, 2544, หน้า 382) เป็นดัชนีความเพียงพอ โดยคะแนนเกณฑ์ที่ใช้จะเป็นผลการแปลงคะแนนด้วยรูปแบบอิงทฤษฎีการตอบสนองข้อสอบ โดยมีสูตร ดังนี้

$$\text{total error} = \Sigma f_j d_j^2 | n S_c^2$$

เมื่อ $d_j^2 = (t-t')$; n = จำนวนคะแนนที่ใช้; S_c^2 = ความแปรปรวนของคะแนน จากสูตรของ Kolen และ Petersen นั้น ภาวิณี ศรีสุขวัฒน์นันท์ (2529) ได้นำดัชนีความแตกต่างมาดัดแปลงโดยใช้ความแปรปรวนเป็นตัวถ่วงน้ำหนัก เพื่อให้ได้ค่าที่เป็นมาตรฐาน ได้สูตรดังนี้

$$c = \frac{\Sigma(x-x^*)^2}{n S_x^2}$$

เมื่อ c	แทน	ดัชนีความแตกต่างมาตรฐาน
X	แทน	คะแนนจากแบบสอบถาม X ซึ่งเป็นคะแนนเกณฑ์
X^*	แทน	คะแนนปรับเทียบของแบบสอบถาม Y ที่อยู่บนสเกล X
n	แทน	จำนวนคะแนนดิบของกลุ่มสอบทานผล
S_x^2	แทน	ความแปรปรวนของคะแนนเกณฑ์

3. ดัชนี (Indices)

เกณฑ์นี้เสนอโดย Angoff (อดิศร ศรีบุญวงศ์, 2545, หน้า 41-42; อัญชลี ศรีภักษาณ, 2552, หน้า 86-87) ดัชนีนี้เดิมที่ใช้เพื่อประเมินความคลาดเคลื่อนทางสถิติ เป็นการสรุปภาพรวมของความแตกต่างระหว่างคะแนนแปลงที่ได้จากการเทียบคะแนน 2 วิธี ซึ่งแบ่งออกเป็น 2 วิธี คือ ดัชนีถ่วงน้ำหนัก และดัชนีไม่ถ่วงน้ำหนักด้วยความถี่ของคะแนน มีรายละเอียดแต่ละวิธีดังนี้

3.1 ดัชนีถ่วงน้ำหนักด้วยความถี่ของคะแนน ประกอบด้วย 3 วิธีคือ

ดัชนี RMS (Root mean square) หาได้จากสูตร

$$RMS = \sqrt{\frac{\sum_i f_i (A_i - B_i)^2}{\sum_i f_i}}$$

ดัชนี MAD (Mean Absolute Difference) หาได้จากสูตร

$$MAD = \frac{\sum_i f_i |A_i - B_i|}{\sum_i f_i}$$

ดัชนี MSD (Mean Signed Difference) หาได้จากสูตร

$$MSD = \frac{\sum_i f_i (A_i - B_i)}{\sum_i f_i}$$

เมื่อ A_i แทน คะแนนสมมูล i ได้จากสเกลแบบสอบใหม่ que เทียบให้อยู่บนสเกลของแบบสอบเดิม

B_i แทน คะแนนเกณฑ์ ซึ่งเป็นคะแนนที่เป็นจริง หรือคะแนนสมมูล i จากแบบสอบฉบับใหม่ que เทียบคะแนนไปยังสเกลของฉบับเก่าด้วยวิธีการอื่น

f_i แทน ความถี่ของคะแนนดิบจากแบบสอบชุดใหม่

3.2 ดัชนีไม่ถ่วงน้ำหนักด้วยความถี่ เป็นวิธีการที่ปรับมาจากวิธีดัชนีถ่วงน้ำหนักด้วยความถี่เนื่องจากวิธีนี้ให้ความสำคัญเฉพาะคะแนนที่มีความถี่ ดังนั้นดัชนีไม่ถ่วงน้ำหนักด้วยความถี่จึงพิจารณาความแตกต่างที่เกิดขึ้นตลอดช่วงของมาตรวัด และไม่กำหนดน้ำหนักในแต่ละระดับคะแนน ประกอบด้วย 3 วิธี คือ

ดัชนี RMS_u หาได้จากสูตร

$$RMS_u = \sqrt{\frac{\sum_i (A_i - B_i)^2}{k}}$$

ดัชนี MAD_u หาได้จากสูตร

$$MAD_u = \frac{\sum_i |A_i - B_i|}{k}$$

ดัชนี MSD_u หาได้จากสูตร

$$MSD_u = \frac{\sum_i (A_i - B_i)}{k}$$

เมื่อ k แทน จำนวนของระดับคะแนน (Score Point)

นอกจากนี้ดัชนี RMSD ยังสามารถนำมาแยกเป็นความแปรปรวนของความแตกต่างและความลำเอียงยกกำลังสอง คือ

$$\sum_i \frac{f_i (d_i)^2}{n} = \sum_i \frac{f_i (d_i - \bar{d})^2}{n + \bar{d}^2}$$

เมื่อ d_i แทน $(A_i - B_i)$

\bar{d} แทน ค่าความแตกต่างเฉลี่ย

4. ความคลาดเคลื่อนของการเชื่อมโยงคะแนน (Error of equating)

Cao et al. (2014) ได้สรุปว่าความคลาดเคลื่อนในการปรับเทียบคะแนนจะแบ่งออกเป็น 2 ส่วนคือ ความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) และความคลาดเคลื่อนอย่างสุ่ม (Random error) สามารถตรวจสอบความคลาดเคลื่อนของข้อมูลที่จำลองขึ้นได้ดังนี้

4.1 ความลำเอียง (Bias) เป็นความคลาดเคลื่อนอย่างเป็นระบบของการปรับเทียบคะแนน สามารถคำนวณได้จากสูตร

$$bias(x) = \frac{1}{R} \sum_{r=1}^R |\hat{e}_r(x)| - e(x) = \bar{e}(x) - e(x)$$

ถ้าต้องการถ่วงน้ำหนัก สามารถคำนวณได้จากสูตร $\sqrt{\sum f(x) bias^2(x)}$

4.2 ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) เป็นความคลาดเคลื่อนอย่างสุ่มของการปรับเทียบ สามารถคำนวณได้จากสูตร

$$SEE(x) = \sqrt{\frac{1}{R} \sum_{r=1}^R [\hat{e}_r(x) - \bar{e}(x)]^2}$$

ถ้าต้องการถ่วงน้ำหนัก สามารถคำนวณได้จากสูตร $\sqrt{\sum f(x) SEE^2(x)}$

4.3 ค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error) เป็นการรวมความคลาดเคลื่อนอย่างเป็นระบบ และความคลาดเคลื่อนอย่างสุ่ม สามารถคำนวณได้จากสูตร ดังนี้

$$RMSE(x) = \sqrt{\frac{1}{R} \sum_{r=1}^R [\hat{e}_r(x) - e(x)]^2}$$

ถ้าต้องการถ่วงน้ำหนัก สามารถคำนวณได้จากสูตร $\sqrt{\sum f(x) RMSE^2(x)}$

จากสูตรความคลาดเคลื่อนทั้ง 3 สูตร สามารถนำมาสร้างเป็นสมการความสัมพันธ์ได้ดังนี้

$$RMSE = \sqrt{bias^2 + SEE^2}$$

5. การปรับเทียบคะแนนกลับสู่แบบสอบเดิม (Test equated to itself or Circular equating)

การปรับเทียบวิธีนี้ Lord (1980 อ้างถึงใน อติศร ศรีบุญวงศ์, 2545, หน้า 42) ได้เสนอไว้ ซึ่งการปรับเทียบคะแนนกลับสู่แบบเดิมถูกนำมาใช้บ่อยมากในงานวิจัย ไม่ว่าจะเป็นการปรับเทียบโดยตรง หรือการปรับเทียบผ่านแบบสอบอื่นแล้วปรับกลับสู่แบบสอบเดิม มีการเปลี่ยนแปลง

การปรับคะแนนมากกว่า 1 ครั้ง เช่น การปรับเทียบคะแนนจากแบบสอบชุด A ไปยังชุด B ปรับชุด B ไปยังชุด C และปรับชุด C กลับสู่แบบสอบชุด A เดิม ซึ่งผลการเปลี่ยนแปลงคะแนนจากการปรับเทียบคะแนนกลับสู่แบบสอบเดิมควรเป็นเอกลักษณ์ หรือคะแนนดิบจากแบบสอบชุด A เดิม กับคะแนนจากแบบสอบชุด A ที่ปรับผ่านชุด B และชุด C มีค่าเท่ากัน

จากรายละเอียดของประสิทธิผลของการเชื่อมโยงคะแนนและจากงานวิจัยผู้วิจัย พิจารณาประสิทธิผลของการเชื่อมโยงคะแนนจากค่าความคลาดเคลื่อน 3 ค่า คือ ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ และค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยต่างประเทศ

Zhang (2007) ได้ศึกษาความสามารถของการปรับเทียบคะแนนระหว่างแบบสอบ เทสต์เลทระหว่างโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนสองกับการตรวจคะแนนมากกว่าสองค่า โดยการปรับเทียบคะแนนจริงและคะแนนสังเกตได้จากการประมาณค่าตาม ทฤษฎีการตอบสนองข้อสอบ 3 พารามิเตอร์ (Three-parameter logistic: 3PL) และการให้คะแนน บางส่วน (Generalized partial credit model: GPC) จากข้อมูลจริงที่เก็บจากการทำแบบทดสอบ การพัฒนาการศึกษา (ITED) ที่มี 3 เทสต์เลท และจำลองข้อมูลภายใต้เงื่อนไข จำนวนข้อสอบในแต่ละเทสต์เลท (5 หรือ 8) ระดับความไม่แน่นอนอิสระภายในเทสต์เลท (ต่ำ ปานกลาง และสูง) และมี ข้อมูลที่กำหนดขึ้นที่มีความเป็นอิสระของข้อสอบ ทำการปรับเทียบด้วยวิธีอควิเปอร์เซ็นต์ไทล์ และ ประเมินคุณภาพด้วยความถูกต้องลำดับที่ 1 และลำดับที่ 2 (First- and second- order equity) ผลการวิจัยพบว่า วิธีการประมาณค่าแบบ GPC มีแนวโน้มให้ผลที่ดีกว่าวิธีการประมาณค่าแบบ 3PL กับมิติที่มีความสัมพันธ์กันต่ำหรือปานกลาง แต่ถ้ามิติสัมพันธ์กันสูง วิธีการประมาณค่าแบบ 3PL จะให้ผลที่ดีกว่า ถ้าพิจารณาระดับความสัมพันธ์ภายในและระหว่างเทสต์เลท ที่มีรูปแบบ ความไม่แน่นอนอิสระของข้อสอบไม่เป็นปกติ พบว่า วิธีการประมาณค่าแบบ GPC จะให้ผลการ ปรับเทียบที่มีค่าความแตกต่างมากกว่า และสัดส่วนของความถูกต้องอันดับที่ 1 และอันดับที่ 2 มากกว่า ในกรณีที่มีมิติมีความสัมพันธ์ในระดับปานกลาง ดังนั้น ความสามารถในการปรับเทียบ โดยใช้ GPC สูงกว่าการใช้ 3PL

Li (2009) ได้พัฒนามาตรวัดร่วมสำหรับการประมาณค่าพารามิเตอร์ที่มีแบบสอบ เทสต์เลท ภายใต้กลุ่มไม่เท่ากันใช้ข้อสอบร่วม ได้ทำการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะ ของข้อสอบ ที่มีการประมาณค่าพารามิเตอร์ตามโมเดลการตอบสนองเทสต์เลท (TRT) แบบ 3PL

และแบบ GRM (Graded response model) ซึ่งการวิจัยครั้งนี้เป็นสถานการณ์การจำลองข้อมูลแบบให้คะแนน 2 ค่า (0,1) สำหรับแบบสอบ 2 ฉบับ ฉบับละ 30 ข้อ โดยแต่ละแบบสอบมี 6 เทสต์เลท ในแต่ละเทสต์เลทประกอบด้วย 5 ข้อสอบ ได้จำนวน 1,000 รายการจากแบบสอบหนึ่งฉบับ ภายใต้เงื่อนไข พารามิเตอร์ของข้อสอบ (ค่าความยาก ค่าอำนาจจำแนก) พารามิเตอร์ความสามารถ (แบบสอบฐานมีการกระจายปกติ (0,1) ส่วนแบบสอบปรับมีการ (0.5,1.5²) พารามิเตอร์ของเทสต์เลท (ระดับความไม่แน่นอนอิสระของข้อสอบ 0, 0.5, 1 และ 2) และเกณฑ์ที่ใช้ในการประเมินการเชื่อมโยงคะแนนคือ ความคลาดเคลื่อนเฉลี่ย (Mean square error: MSE) สัดส่วนของค่าเฉลี่ย (Root mean squared deviation: RMSD) และค่าสัมบูรณ์ของความแตกต่างค่าเฉลี่ย (Mean absolute difference: MAD) ได้ผลการวิจัย ที่ค่าอำนาจจำแนกเท่ากับ 1.5 พบว่าที่ระดับความไม่แน่นอนเท่ากับ 2 การประมาณค่าทั้ง 3 วิธีแตกต่างกัน โดยการประมาณค่าแบบ GRM จะให้ค่าเฉลี่ยที่สูงกว่าการประมาณค่าแบบ 3PL แต่กรณีที่ ค่าความยากเท่ากับ 0.5 พบว่าการประมาณทั้ง 3 วิธีจะแตกต่างกัน ที่ระดับความไม่แน่นอนของข้อสอบเท่ากับ 1 และ 2 โดยที่การประมาณค่าแบบ GRM จะให้ค่าเฉลี่ยที่สูงกว่าการประมาณค่าแบบ 3PL และการประมาณค่าพารามิเตอร์แบบสอบเทสต์เลท ถ้าพิจารณาค่า MSE โดยยึดค่าอำนาจเป็นหลัก พบว่า ถ้าค่าระดับความไม่แน่นอนเพิ่มสูงขึ้น ค่า MSE ของการประมาณค่าทั้ง 3 วิธีมีแนวโน้มเพิ่มสูงขึ้น การประมาณแบบ 3PL มีแนวโน้มสูงกว่ารูปแบบอื่น ส่วนถ้ายึดค่าความยากเป็นหลัก พบว่าค่า MSE ที่ระดับความไม่แน่นอนของข้อสอบเท่ากับ 1 จะให้ค่าน้อยกว่าระดับอื่น ๆ แต่ค่า MSE ที่ได้จากการประมาณค่าแบบ GRM จะมีค่าสูงกว่าการประมาณค่ารูปแบบอื่น ส่วนค่า MAD และค่า RMSD มีค่าเพิ่มขึ้น เมื่อระดับความไม่แน่นอนของข้อสอบเพิ่มสูงขึ้น

Puhan et al. (2009) ได้ศึกษาการเปรียบเทียบกลุ่มตัวอย่างขนาดเล็กโดยใช้การออกแบบกลุ่มผู้สอบหนึ่งกลุ่มที่มีแบบสอบใกล้เคียงกัน (SiGNET) โดยมีวัตถุประสงค์ของการวิจัยเพื่อเปรียบเทียบความคลาดเคลื่อนของการเปรียบเทียบและความลำเอียงระหว่างการออกแบบวิธี SiGNET กับการออกแบบวิธีกลุ่มผู้สอบไม่เท่าเทียมกันโดยใช้แบบสอบเทสต์เลท (NEAT) โดยเก็บข้อมูลจำนวน 23,000 คน กับแบบสอบ 120 ข้อ เกี่ยวกับทักษะพื้นฐาน 4 ทักษะ คือ การอ่าน คณิตศาสตร์ สังคม และวิทยาศาสตร์ นำมาสร้างเป็นแบบสอบ 5 ฉบับ ซึ่งแต่ละฉบับจะมี 10 ชุดข้อสอบ ชุดข้อสอบละ 12 ข้อ ซึ่งแต่ละฉบับจะมีข้อสอบที่เหมือนกัน 83% โดยแต่ละเทสต์เลทจะมีความยากเฉลี่ยระหว่าง .63-.68 ทั้งฉบับจะมีความยากเฉลี่ย .65 ซึ่งการออกแบบทั้งสองวิธีจะมีเงื่อนไขเกี่ยวกับจำนวนกลุ่มตัวอย่างต่างกัน ดังนี้ การออกแบบวิธี NEAT จะมีขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 10, 15, 25 และ 50 คน และการออกแบบวิธี SiGNET จะมีขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 50, 100 และ 150 คน ทำการเปรียบเทียบคะแนนระหว่างแบบสอบด้วยวิธีเชิงเส้นตรง แล้ว

เปรียบเทียบผลของการออกแบบทั้งสองวิธีด้วยค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบ (SEE) ค่าเฉลี่ยของความลำเอียง (bias) และค่าเฉลี่ย RMSD ได้ผลการวิจัยดังนี้ การออกแบบการเก็บรวบรวมทั้งสองวิธีจะมีค่าเฉลี่ย SEE และ RMSD ลดลงเมื่อจำนวนกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น แต่ค่าที่ได้จากวิธี SiGNET จะมีค่าน้อยกว่าวิธี NEAT

Rui et al. (2010) ได้ศึกษาวิธีการปรับเทียบคะแนนจากแบบทดสอบทดสอบเดี่ยว โดยมีการวัดคุณสมบัติเพื่อเปรียบเทียบคุณภาพของการปรับเทียบที่มีการประมาณค่าแบบ TRT 2 พารามิเตอร์ (2PTM) และ IRT 2 พารามิเตอร์ (2PLM) และใช้วิธีการปรับเทียบคะแนนระหว่างแบบทดสอบด้วยวิธีโค้งคุณลักษณะตามทฤษฎีการตอบสนองข้อสอบ แล้วคำนวณสัมประสิทธิ์ของการปรับเทียบ (RMSD) มาเปรียบเทียบกับสถิติ Wilcoxon ส่วนการจำลองข้อมูลได้ใช้วิธีมอนติคาร์โล ภายใต้เงื่อนไข 1) พารามิเตอร์ของข้อสอบประกอบด้วยค่าอำนาจจำแนกและค่าความยาก (1/ 6, 1/ 4; 1/ 0, 1/ 6; 1/ 20, 1/ 10 ตามลำดับ) 2) จำนวนผู้สอบ แบ่งเป็น 3 กลุ่มคือ 300, 1,000 และ 5,000 คน และ 3) ระดับของความไม่อิสระกัน แบ่งเป็น 2 กรณี คือ 1) แบบสอบมี 30 ข้อ แบ่งเป็น 6 เทสต์เลท (0,0,0.5,0.5,1,1; 0,0,1,1,2,2; 0,0,2,2,4,4) และ 2) แบบสอบ 20 ข้อ แบ่งเป็น 4 เทสต์เลท (0.25,0.25,0.25,0.25; 0.5,0.5,0.5,0.5; 0.75,0.75,0.75,0.75; 1,1,1,1; 2,2,2,2; 3,3,3,3; 4,4,4,4) ผลการวิจัย พบว่า การประมาณค่าแบบ 2PTM จะให้ผลที่ดีกว่าการประมาณค่าแบบ 2PLM อย่างมีนัยสำคัญทางสถิติ

He et al. (2012) ได้ทำการศึกษาโมเดลที่เหมาะสมสำหรับการปรับเทียบคะแนนแบบทดสอบทดสอบเดี่ยว ภายใต้รูปแบบกลุ่มผู้สอบไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม เป็นการศึกษาจากข้อมูลเชิงประจักษ์ และประยุกต์ใช้การปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบ ซึ่งมีวิธีดำเนินการวิจัยดังนี้ จำนวนกลุ่มตัวอย่างที่ใช้ในการวิจัยแบ่งออกเป็น 2 กลุ่มคือ กลุ่มที่ทำแบบทดสอบจำนวน 3,263 คน และกลุ่มที่ทำแบบสอบปรับจำนวน 3,944 คน ข้อมูลที่ใช้ในการวิจัยได้มาจาก แบบทดสอบการอ่าน State-wide ทั้งแบบสอบที่เป็นฐาน และแบบสอบที่นำมาปรับ โดยแบบสอบทั้งสองฉบับจะมีจำนวนข้อสอบ 50 ข้อ โดยแบบสอบที่เป็นฐานจะมีประกอบด้วย 4 เทสต์เลท (13, 12, 9, 11 ข้อ ตามลำดับ) และ 5 ข้อที่เป็นอิสระกัน ส่วนแบบสอบที่นำมาปรับประกอบด้วย 4 เทสต์เลท (11, 12, 13, 11 ข้อ ตามลำดับ) และ 3 ข้อที่เป็นอิสระกัน โดยแบบสอบทั้งสองฉบับจะมีเทสต์เลทร่วมกัน 1 ชุด จำนวน 12 ข้อ จากนั้นประมาณค่าพารามิเตอร์ของข้อสอบ 6 รูปแบบ คือ IRT 1 พารามิเตอร์ 2 พารามิเตอร์ 3 พารามิเตอร์ TRT 2 พารามิเตอร์ 3 พารามิเตอร์ และ bifactor แล้วทำการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีอิควิเปอร์เซ็นต์ไทล์ และวิธีเชิงเส้นตรง และประเมินคุณภาพการปรับเทียบคะแนนด้วยค่า Weighted bias (WBS), Weighted root measure square error (WRMSE) และค่า Weighted absolute

bias (WABS) ผลการวิจัยพบว่า ค่า WBS, WRMSE และ WABS ของการประมาณค่าแบบ IRT 1 พารามิเตอร์ 2 พารามิเตอร์ที่ได้จากวิธีอควิเปอร์เซ็นไทล์น้อยกว่าวิธีเชิงเส้นตรง ส่วนการประมาณค่า IRT 3 พารามิเตอร์ และ TRT 3 พารามิเตอร์ ที่ได้จากวิธีอควิเปอร์เซ็นไทล์มากกว่าวิธีเชิงเส้นตรง ส่วนการประมาณค่า bifactor และ TRT 2 พารามิเตอร์ ค่า WBS ที่ได้จากวิธีเชิงเส้นตรงน้อยกว่าวิธีอควิเปอร์เซ็นไทล์ แต่ค่า WRMSE และ WABS ที่ได้จากวิธีอควิเปอร์เซ็นไทล์มากกว่าวิธีเชิงเส้นตรง

Cao et al. (2014) ได้ทำการศึกษาผลของทฤษฎีการตอบสนองข้อสอบที่เหมาะสมกับการปรับเทียบคะแนนของแบบสอบทดสอบหลายแบบ โดยมีปัญหาของการวิจัยว่า ถ้าวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบต่างกัน ผลของคะแนนปรับเทียบที่มีและไม่มีการจัดการความไม่เป็นอิสระของข้อสอบจะต่างกันหรือไม่ โดยการวิจัยครั้งนี้ใช้สถานการณ์จำลองข้อมูลให้สอดคล้องกับพารามิเตอร์ของข้อสอบ โดยมีจำนวนกลุ่มตัวอย่างกลุ่มละ 2,000 คน ทำซ้ำ 50 ครั้ง ด้วยโปรแกรม SAS เครื่องมือจะเป็นแบบสอบจำนวน 2 ชุด คือแบบสอบที่เป็นฐานและแบบสอบที่นำมาปรับ ชุดหนึ่งจะมีแบบสอบ 8 ฉบับ ตามเงื่อนไขความเป็นอิสระของข้อสอบและความยากของข้อสอบ โดยในแต่ละแบบสอบจะเป็นแบบเลือกตอบ 40 ข้อ มี 6 เทสต์เลท เทสต์เลทละ 5 ข้อ และเป็นอิสระ 10 ข้อ และตัวแปรที่ใช้ในการวิจัยครั้งนี้ คือ ตัวแปรต้น ประกอบด้วย 1) วิธีการปรับเทียบ แบ่งออกเป็น 2 วิธี คือ การปรับเทียบคะแนนจริงด้วยวิธีของ Newton-Raphson และการปรับเทียบคะแนนสังเกตได้ด้วยวิธีอควิเปอร์เซ็นไทล์ 2) การประมาณค่าพารามิเตอร์ของข้อสอบจะมี 3 รูปแบบ คือ IRT 2 พารามิเตอร์ GRM และ TRT 2 พารามิเตอร์ 3) ระดับความไม่เป็นอิสระของข้อสอบ แบ่งออกเป็น 4 ระดับคือ 0, 0.1-0.5, 0.6-1.0, และ 1.1-1.5 และ 4) ระดับความยากของแบบสอบแบ่งเป็นแบบสอบที่เป็นฐาน (-2.5-2.5) กับแบบสอบที่นำมาปรับเทียบ (0.1,1) ส่วนตัวแปรตาม คือ ประสิทธิภาพของการปรับเทียบคะแนน (bias, SEE, RMSE) ได้ผลการวิจัยดังนี้ การยึดการปรับเทียบคะแนนโดยวิธีอควิเปอร์เซ็นไทล์เป็นฐานในการประเมินพบว่า 1) ค่าน้ำหนักเฉลี่ยของความลำเอียง (Weighted average of bias) ที่ได้จากการประมาณค่าแบบ GRM จะมีค่าความลำเอียงสูงกว่าการประมาณค่ารูปแบบ IRT 2 พารามิเตอร์ และ TRT 2 พารามิเตอร์ ทั้ง 4 ระดับความไม่เป็นอิสระของข้อสอบ 2) ค่าน้ำหนักเฉลี่ยของความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Weighted average of SEE) ที่ได้จากการประมาณค่าแบบ GRM จะมีค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบสูงกว่าเล็กน้อยกับการประมาณค่ารูปแบบ IRT 2 พารามิเตอร์ และ TRT 2 พารามิเตอร์ ทั้ง 4 ระดับความไม่เป็นอิสระของข้อสอบ และ 3) ค่า RMSE ที่ได้จากการประมาณค่าแบบ GRM จะมีค่า RMSE สูงกว่าการประมาณค่ารูปแบบ IRT 2 พารามิเตอร์ และ TRT 2 พารามิเตอร์ ทั้ง 4 ระดับความไม่เป็นอิสระของข้อสอบ

Chen (2014) ศึกษาเปรียบเทียบการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ ภายใต้วิธีการประมาณค่าพารามิเตอร์ 4 รูปแบบ คือ 3PL GRM Bifactor และ TRT โดยใช้ทั้งข้อมูลจริงและจำลองข้อมูลขึ้น ดัชนีที่ใช้ในการเปรียบเทียบประสิทธิภาพของการเชื่อมโยงคะแนน คือค่าดัชนี URMS/WRMS สำหรับแบบทดสอบจริง ส่วนแบบทดสอบที่จำลองขึ้นพิจารณาจากค่าดัชนี Bias SEE และ URMS/WRMS ผลการศึกษาพบว่า ในแบบทดสอบจริงค่าดัชนี URMS/WRMS ที่ได้จากการเชื่อมโยงคะแนนจริงของประมาณค่าพารามิเตอร์รูปแบบ TRT สูงกว่า Bifactor GRM และ 3PL ตามลำดับ เช่นเดียวกับการเชื่อมโยงคะแนนสังเกตได้ เมื่อเทียบระหว่างวิธีการเชื่อมโยงพบว่าการเชื่อมโยงคะแนนสังเกตได้จะมีค่าน้อยกว่าการเชื่อมโยงคะแนนจริง เช่นเดียวกับผลการเชื่อมโยงคะแนนด้วยแบบทดสอบที่จำลองขึ้นค่าดัชนี URMS/WRMS ให้ผล เช่นเดียวกับแบบทดสอบจริง ส่วนดัชนี bias และ SEE ของการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ภายใต้การประมาณค่าพารามิเตอร์ IRT สูงกว่า TRT

Kim (2015) ได้ทำงานวิจัยเกี่ยวกับการเปรียบเทียบพารามิเตอร์ข้อสอบระหว่างการประมาณค่าพารามิเตอร์ข้อสอบตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ทฤษฎีการตอบสนองข้อสอบ (IRT) และทฤษฎีการตอบสนองทดสอบ (TRT) เก็บข้อมูลกับนักเรียนเกรด 5 จำนวน 10,879 คน ที่ทำแบบทดสอบการอ่านเพื่อทำความเข้าใจ โดยแบบทดสอบมี จำนวน 22 ข้อ ประกอบด้วย 7 ข้อความ โดยวัดการอ่านข้อมูลเกี่ยวกับ Common core state standards (CCSS) จำนวน 4 มาตรฐาน คือ มาตรฐานวรรณกรรม มาตรฐานข้อมูลสารสนเทศ มาตรฐานทักษะพื้นฐาน และมาตรฐานภาษา โดยเครื่องมือมีความเชื่อมั่นครอนบาคแอลฟาเท่ากับ 0.79 ผลการวิจัยพบว่า การวิเคราะห์คุณภาพของสอบแบบ IRT ให้สารสนเทศดีกว่า CTT (ค่าความยากอำนาจจำแนก ค่าการเดา) โดยเฉพาะค่าความสามารถของผู้สอบรายบุคคล ส่วนการวิเคราะห์ TRT พบว่าข้อสอบมีความแปรปรวนเกิดขึ้นจากอิทธิพลของทดสอบประมาณ (0.18, 0.19, 0.06) ซึ่งโมเดลนี้ผู้วิจัยได้ปรับเป็นคะแนนมาตรฐาน (ค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1) ผลการเปรียบเทียบพารามิเตอร์ข้อสอบระหว่าง IRT และ TRT พบว่า พารามิเตอร์อำนาจจำแนกและค่าการเดามีความสัมพันธ์กันสูง ยกเว้นพารามิเตอร์ความยากซึ่งพารามิเตอร์ความยากนี้เกี่ยวข้องกับอิทธิพลค่าเฉลี่ยของทดสอบ

Tao and Cao (2016) ได้ทำการศึกษาเกี่ยวกับการขยายการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบไปยังการปรับเทียบคะแนนตามทฤษฎีการตอบสนองทดสอบที่มีการตรวจให้คะแนนแบบ 2 ค่า โดยมีวัตถุประสงค์ของการศึกษาเพื่อขยายผลการปรับเทียบคะแนนด้วยคะแนนจริงและคะแนนสังเกตได้จากแบบสอบที่มีการตรวจให้คะแนนสองค่าตามทฤษฎี

การตอบสนองของเทสต์เลท และตรวจสอบผลของความไม่เป็นอิสระของข้อสอบที่เกิดจากการปรับเทียบ คะแนนถูกต้องตามทฤษฎีการตอบสนองข้อสอบ การศึกษานี้เป็นการจำลองข้อมูลที่มีการออกแบบ การเก็บข้อมูลกับกลุ่มผู้ไม่เท่าเทียมกันมีข้อสอบพร้อม ประมาณค่าพารามิเตอร์ 4 รูปแบบ คือ คะแนนจริง IRT 3PL คะแนนสังเกตได้ IRT 3PL คะแนนจริง TRT 3 PL และคะแนนสังเกตได้ TRT 3PL มีระดับความไม่เป็นอิสระของข้อสอบ 4 ระดับ คือ ไม่มีเลย ต่ำ ปานกลาง และสูง กับแบบ สอบรูปแบบเลือกตอบจำนวน 30 ข้อ มี 6 เทสต์เลท ในแต่ละเทสต์เลทมีจำนวน 5 ข้อ จากนั้นทำการปรับเทียบคะแนนจริงด้วยวิธี Newton-Raphson และคะแนนสังเกตได้ด้วยวิธีอิกวิเปอร์เซนไทล์ และประเมินการปรับเทียบคะแนนด้วยค่าความลำเอียง (bias) และความคลาดเคลื่อนของการปรับเทียบ (SEE) ผลการศึกษาพบว่า เมื่อความไม่เป็นอิสระของข้อสอบอยู่ในระดับต่ำ การใช้โมเดลโลจิสติกประมาณค่าข้อสอบ 3 พารามิเตอร์ไม่มีผลต่อการปรับเทียบคะแนนถูกต้อง แต่เมื่อความไม่เป็นอิสระกันของข้อสอบอยู่ในระดับกลางหรือระดับสูง การใช้โมเดลโลจิสติกประมาณค่าข้อสอบ 3 พารามิเตอร์ จะมีค่าความลำเอียงและความคลาดเคลื่อนของการปรับเทียบมากกว่า เมื่อเทียบกับทฤษฎีการตอบสนองข้อสอบ และการปรับเทียบคะแนนสังเกตได้จะมีความแม่นยำมากกว่าการปรับเทียบคะแนนจริงเมื่อละเมิดข้อตกลงความเป็นอิสระกันของข้อสอบ

Kogar and Kelecioğlu (2017) ได้ทำการศึกษาเปรียบเทียบพารามิเตอร์ข้อสอบและพารามิเตอร์โมเดลการตอบสนองข้อสอบของแบบสอบที่มีลักษณะของเทสต์เลท จำนวน 3 โมเดล คือ UIRT Bifactor และTRT ภายใต้เงื่อนไข แบบทดสอบมีจำนวนข้อ จำนวนเทสต์เลท และจำนวนกลุ่มตัวอย่างต่างกัน ผลการศึกษาพบว่า ความคลาดเคลื่อนมาตรฐานของโมเดล UIRT น้อยกว่า TRT และ Bifactor ในทุกเงื่อนไข อีกทั้งความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ และค่าความคลาดเคลื่อนของพารามิเตอร์ของแต่ละโมเดลมีความสัมพันธ์กันสูง

งานวิจัยในประเทศ

อดิศร ศรีบุญวงศ์ (2545) ได้ศึกษาเกี่ยวกับการพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ โดยมีวัตถุประสงค์ 2 ข้อ คือ 1) เพื่อพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ ด้วยข้อมูลจำลอง จากรูปแบบ การหาเกณฑ์ 2 แบบ คือ การปรับเทียบคะแนนกลับสู่แบบเดิม และการใช้กลุ่มสอบทานผล และ 2) เพื่อตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น ในด้านความตรงเชิงเกณฑ์สัมพันธ์โดยตรวจสอบความสอดคล้องของผลการตัดสินด้วยเกณฑ์ที่พัฒนาขึ้นกับเกณฑ์ของปีเตอร์เซนและคณะ และความสอดคล้องของผลการตัดสินด้วยเกณฑ์ที่พัฒนาขึ้นกับความเสมอภาคของลอร์ด โดยเกณฑ์ที่พัฒนาขึ้นมีทั้งหมด 7 เกณฑ์ คือ เกณฑ์รวมทุกเงื่อนไข เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ เกณฑ์สำหรับการใช้กลุ่มสมมูล เกณฑ์สำหรับการใช้แบบสอบรวม

เกณฑ์สำหรับการเปรียบเทียบกลับสู่แบบสอบเดิม และเกณฑ์สำหรับการใช้กลุ่มสอบทานผล
 เปรียบเทียบคะแนนระหว่างแบบสอนตามทฤษฎีการตอบสนองข้อสอบ ได้แก่ วิธี Same scaling
 convention วิธี Mean and mean วิธี Mean and sigma และวิธี Characteristic curve และ
 เกณฑ์การตัดสินคุณภาพของการเปรียบเทียบพัฒนาจากดัชนีความแตกต่าง 3 วิธีคือ AMD
 (Absoluted mean square) MAD (Mean absolute difference) และ RMS (Root mean
 square) แล้วนำค่าที่ได้ไปหาความสอดคล้องกับเกณฑ์ของปีเตอร์เซนและคณะกับเกณฑ์
 ความเสมอภาคของลอร์ด ผลการวิจัยพบว่า เกณฑ์ที่ผู้วิจัยพัฒนาขึ้นไม่สอดคล้องกับเกณฑ์ของ
 ปีเตอร์เซนและคณะ แต่พบว่าสอดคล้องกับเกณฑ์ความเสมอภาคของลอร์ด

สุนทร เทียนงาม (2551) ได้ทำการศึกษาผลของความไม่เป็นอิสระของข้อสอบที่มีต่อค่า
 ความเที่ยง ค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถของผู้สอบและค่าสารสนเทศของแบบสอบ
 เมื่อมีเงื่อนไขของการทดสอบที่ต่างกัน โดยมีสมมติฐานงานวิจัยจำนวน 2 ข้อ คือ 1) เมื่อกำหนด
 ระดับความไม่เป็นอิสระของข้อสอบเพิ่มสูงขึ้นภายใต้เงื่อนไข จำนวนข้อสอบ ขนาดกลุ่มตัวอย่าง ที่
 แตกต่างกันส่งผลให้ค่าความเที่ยงของแบบสอบสูงขึ้น และ 2) เมื่อกำหนดระดับความไม่เป็นอิสระของ
 ข้อสอบเพิ่มสูงขึ้นภายใต้เงื่อนไขของโมเดลการตอบสนองข้อสอบที่มีจำนวนข้อสอบและขนาด
 กลุ่มตัวอย่างที่แตกต่างกันส่งผลทำให้ค่าพารามิเตอร์ของข้อสอบ อันได้แก่ อำนาจจำแนกสูงขึ้น
 ค่าความยากต่ำลง ค่าการเดาเพิ่มขึ้น และค่าสารสนเทศของแบบสอบลดลง และมีขอบเขตของ
 การวิจัยดังนี้ 1) การศึกษาจะใช้การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลจากโปรแกรมฟอร์แทรน
 และกำหนดจำนวนรอบในการประมาณค่าพารามิเตอร์แต่ละเงื่อนไขจำนวน 1000 รอบ และ
 มีการวิเคราะห์ข้อมูลจริงร่วมด้วย ซึ่งแบบสอบถามที่ใช้มีเนื้อหาที่จะมีโอกาสเกิดความไม่เป็นอิสระ
 ของข้อสอบได้มาก โดยเฉพาะแบบสอบที่ถามเนื้อหาเดียวกันด้วยข้อสอบหลายข้อ หรือแบบสอบที่
 ใช้ตาราง กราฟ แผนภูมิร่วมกันในการตอบข้อสอบ 2) ตัวแปรที่ใช้ในการศึกษา ประกอบด้วย
 ตัวแปรต้น มีดังนี้ ระดับความไม่เป็นอิสระของข้อสอบ 9 ระดับ โดยแบ่งเป็นกลุ่มย่อย 3 กลุ่ม
 (ระดับต่ำ ปานกลาง สูง) โมเดลทฤษฎีการตอบสนองข้อสอบ 3 โมเดล (1PL 2PL 3PL) จำนวน
 ข้อสอบ 3 กลุ่ม (30, 50, 80 ข้อ) และขนาดกลุ่มตัวอย่าง 3 กลุ่ม (400, 800, 1,200 คน) รวม
 จำนวนเงื่อนไขเท่ากับ 243 เงื่อนไข ตัวแปรตามได้แก่ ค่าความเที่ยง ค่าพารามิเตอร์ของข้อสอบ
 ค่าความสามารถของผู้สอบ และค่าสารสนเทศของแบบสอบ และ 3) การประมาณค่าพารามิเตอร์
 ประกอบด้วย การตรวจสอบความไม่เป็นอิสระของข้อสอบ จะใช้ค่า Yen's Q3 ในการคำนวณ และ
 การประมาณค่าพารามิเตอร์ของข้อสอบและการประมาณค่าความสามารถของผู้สอบโดยใช้วิธีการ
 ประมาณค่าของเบส์โดยใช้โมเดลการตอบสนองข้อสอบแบบ 1, 2, 3 พารามิเตอร์ ส่วนการวิเคราะห์
 ข้อมูลจริง จะใช้คะแนนจากแบบสอบวิชาคณิตศาสตร์และภาษาอังกฤษในระดับช่วงชั้นที่ 2

ระดับชั้นประถมศึกษาปีที่ 5 ปีการศึกษา 2550 เขตพื้นที่การศึกษาจังหวัดสุพรรณบุรี เขต 1 ได้ผลการวิจัยดังนี้

1. เมื่อระดับความไม่เป็นอิสระของข้อสอบเพิ่มสูงขึ้น ค่าความเที่ยงจะเพิ่มสูงขึ้นในทุกเงื่อนไขของการทดสอบ โดยเมื่อใช้จำนวนข้อสอบ 30 ข้อ มีค่าความเที่ยงต่ำสุด-สูงสุดเท่ากับ 0.671-0.740 เมื่อจำนวนข้อสอบ 50 ข้อ มีค่าความเที่ยงต่ำสุด-สูงสุดเท่ากับ 0.773-0.828 และเมื่อจำนวนข้อสอบ 80 ข้อ มีค่าความเที่ยงต่ำสุด-สูงสุดเท่ากับ 0.845-0.886

2. เมื่อวิเคราะห์ข้อสอบตามแนวคิดทฤษฎีการตอบสนองของข้อสอบ มีจำนวนผู้สอบ 400 800 และ 1,200 คน ทำข้อสอบ 30 50 และ 80 ข้อ ตามลำดับ 2.1) เมื่อวิเคราะห์โดยใช้โมเดล 1 พารามิเตอร์ พบว่า เมื่อระดับความไม่เป็นอิสระของข้อสอบเพิ่มสูงขึ้น ค่าความยากง่ายและความสามารถไม่แตกต่างกัน และค่าสารสนเทศของแบบสอบส่วนใหญ่มีแนวโน้มคงที่ 2.2) เมื่อวิเคราะห์โดยใช้โมเดล 2 พารามิเตอร์ พบว่า ส่วนใหญ่ความความยากง่ายลดลง แต่ค่าอำนาจจำแนก ค่าความสามารถและสารสนเทศของแบบสอบมีแนวโน้มสูงขึ้น โดยส่วนใหญ่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ 2.3) เมื่อวิเคราะห์โดยใช้โมเดล 3 พารามิเตอร์ พบว่า ค่าความยากง่าย ค่าความสามารถ และสารสนเทศของแบบสอบมีแนวโน้มลดลง แต่สำหรับค่าอำนาจจำแนก ค่าการเดาส่วนใหญ่มีแนวโน้มสูงขึ้นและแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. เมื่อวิเคราะห์ความไม่เป็นอิสระของข้อสอบโดยใช้ข้อมูลจริง ด้วยโมเดล 1, 2, และ 3 พารามิเตอร์ สำหรับวิชาคณิตศาสตร์ พบว่า ค่าเฉลี่ย $|Q3|$ มีเท่ากับ 0.035, 0.026 และ 0.022 ตามลำดับ และสำหรับวิชาภาษาอังกฤษ พบว่า ค่าเฉลี่ย $|Q3|$ มีเท่ากับ 0.033, 0.028 และ 0.022 ตามลำดับ มีค่าใกล้เคียงกับค่าคาดหวังของ $|Q3|$ ซึ่งมีค่าเท่ากับ 0.025 แสดงว่าข้อสอบส่วนใหญ่มีความไม่เป็นอิสระของข้อสอบเล็กน้อย

กิติพงษ์ ลีอนาม (2551) ได้ทำการศึกษาเกี่ยวกับการนำเสนอวิธีการปรับเทียบคะแนนความสามารถโดยประยุกต์ใช้ทฤษฎีการตอบสนองของข้อสอบ มีวัตถุประสงค์ของการวิจัย ดังนี้

1) เพื่อนำเสนอวิธีการปรับเทียบคะแนนความสามารถ โดยประยุกต์ใช้ทฤษฎีการตอบสนองของข้อสอบ 2) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนความสามารถ ระหว่างวิธีที่นำเสนอกับวิธีสมการถดถอย และ 3) เพื่อตรวจสอบการให้ผลการเรียนของนักเรียนโรงเรียนมัธยมศึกษา จากการปรับเทียบคะแนนความสามารถด้วยวิธีที่นำเสนอ โดยมีขอบเขตของการวิจัย ดังนี้ 1) การหาประสิทธิภาพการปรับเทียบคะแนนความสามารถจากข้อมูลเชิงประจักษ์ โดยประชากรที่ใช้ในการวิจัย คือ นักเรียนชั้นมัธยมศึกษาตอนปลายของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ที่สำเร็จการศึกษาประจำปี

การศึกษา 2549 จำนวน 285,796 คน 2) ข้อมูลที่ใช้ในการวิเคราะห์มี 2 กลุ่ม คือ 2.1) แยกแต่ละภาคเรียนทั้ง 6 ภาคเรียน ของนักเรียนทุกคน ในแต่ละโรงเรียนของ 5 สาระการเรียนรู้ ได้แก่ กลุ่มสาระการเรียนรู้คณิตศาสตร์ กลุ่มสาระการเรียนรู้วิทยาศาสตร์ กลุ่มสาระการเรียนรู้ภาษาไทย กลุ่มสาระการเรียนรู้ภาษาอังกฤษ และกลุ่มสาระการเรียนรู้สังคม ศาสนา และวัฒนธรรม 2.2) ผลการตอบข้อสอบรายข้อ รายคน รายวิชา 5 วิชา ได้แก่ วิชาคณิตศาสตร์ วิชาวิทยาศาสตร์ วิชาภาษาอังกฤษ วิชาสังคมศึกษา และวิชาภาษาไทย ของนักเรียนที่สำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย ปีการศึกษา 2549 ที่สอบ O-NET ปี พ.ศ. 2550 3) การประมาณค่าพารามิเตอร์ความสามารถ มี 2 รูปแบบคือ โมเดลแบบ 3 พารามิเตอร์ที่มีการตรวจให้คะแนนแบบ 2 ค่า (0,1) และโมเดล GRM-IRT แบบ 2 พารามิเตอร์ ที่มีการตรวจให้คะแนนมากกว่า 2 ค่า (0, 1, 2, 3, 4) 4) กำหนดความยาวข้อสอบจำนวนไม่น้อยกว่า 5 ข้อ เพื่อประมาณคะแนนความสามารถได้ถูกต้อง 5) เจือจางจำนวนนักเรียนแต่ละกลุ่ม ซึ่งศึกษาเฉพาะโรงเรียน 3 ขนาด คือ กลุ่มขนาดกลาง มีจำนวนนักเรียนที่จบ ม.6 เท่ากับ 100-200 คน กลุ่มโรงเรียนขนาดใหญ่ มีจำนวนนักเรียนที่จบ ม.6 เท่ากับ 201-400 คน และกลุ่มขนาดใหญ่พิเศษ มีจำนวนนักเรียนที่จบ ม.6 เท่ากับ 401 ขึ้นไป ตัวแปรของการวิจัย คือ ตัวแปรต้น ได้แก่ วิธีการปรับเทียบคะแนนความสามารถ 2 วิธี คือ วิธี MASE (Mean ability) และวิธี RE (Regression equating) ส่วนตัวแปรตามคือ ประสิทธิภาพของการปรับเทียบพิจารณาจาก 3 ค่า คือ ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) ความเพียงพอของการปรับเทียบ (Adequacy of equating models) และสัมประสิทธิ์การทำนาย (Coefficient of prediction) ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม Mutilog ซึ่งผลการวิจัยพบว่า 1) ค่าความคลาดเคลื่อนมาตรฐานของวิธีการปรับเทียบปรับเทียบคะแนนความสามารถวิธี MASE มีค่าต่ำกว่าวิธี RE 2) ค่าสัมประสิทธิ์การทำนายของวิธี MASE มีค่าสูงกว่าวิธี RE และ 3) ค่าความเพียงพอของการปรับเทียบทั้งสองวิธีไม่แตกต่างกัน

อัญชลี ศรีกลชาญ (2552) ได้ศึกษาคุณภาพของการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสมโดยการประยุกต์ใช้การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบด้วยวิธีโค้งคุณลักษณะและการปรับเทียบค่าพารามิเตอร์พร้อมกัน มีวัตถุประสงค์ของการวิจัยเพื่อศึกษาผลของการปรับเทียบคะแนนตามแนวทฤษฎีการตอบสนองข้อสอบ สำหรับแบบสอบที่มีรูปแบบผสมระหว่างการตรวจให้คะแนนแบบสองค่าตามโมเดล IRT โลกีสติกแบบ 3 พารามิเตอร์ และการตรวจให้คะแนนแบบหลายค่าตามโมเดล GPC โดยมีวัตถุประสงค์เฉพาะ 2 ข้อ คือ 1) เพื่อวิเคราะห์คุณภาพของการปรับเทียบคะแนนด้วยวิธีโค้งคุณลักษณะและวิธีการปรับค่าพารามิเตอร์พร้อมกัน เมื่อแบบแผนการเก็บรวบรวมข้อมูล และสัดส่วนของข้อสอบทั้ง 2 ชนิดในแบบสอบรูปแบบผสม

และสัดส่วนของข้อสอบทั้ง 2 ชนิดในข้อสอบร่วมของแบบสอบรูปแบบผสมต่างกัน 2) เพื่อเปรียบเทียบคุณภาพผลการเปรียบเทียบคะแนนวิธีโค้งคุณลักษณะ และวิธีการปรับค่าพารามิเตอร์พร้อมกัน เมื่อแบบแผนการเก็บรวบรวมข้อมูลและสัดส่วนของข้อสอบทั้ง 2 ชนิดในข้อสอบร่วมของแบบสอบรูปแบบผสมต่างกันโดยพิจารณาจากดัชนี MSE โดยมีขอบเขตของการวิจัยดังนี้

1) การวิจัยครั้งนี้เป็นสถานการณ์การจำลองข้อมูลให้สอดคล้องกับโมเดลการตอบข้อสอบที่มีการตรวจให้คะแนนแบบสองค่าตามโมเดล IRT แบบ 3 พารามิเตอร์ และโมเดลการตอบข้อสอบที่มีการตรวจให้คะแนนหลายค่า ตามโมเดล GPC 6 ลำดับชั้น ด้วยโปรแกรม WINGEN จำนวนผู้สอบ 1000 คน มีการทำซ้ำ 100 ครั้ง และใช้โปรแกรม MULTILOG ในการวิเคราะห์พารามิเตอร์ของข้อสอบ 2) ประชากรข้อสอบที่ใช้ในการวิจัยครั้งนี้เป็นค่าพารามิเตอร์คุณลักษณะข้อสอบที่สร้างขึ้นด้วยฟังก์ชันคณิตศาสตร์ โดยการจำลองแบบค่าที่เป็นไปได้ของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบจากโมเดลโลจิสติก 3 พารามิเตอร์ ซึ่งประกอบด้วย ค่าความยากของข้อสอบมีค่าอยู่ระหว่าง -2.5 ถึง +2.5 ค่าอำนาจจำแนกของข้อสอบมีค่าอยู่ระหว่าง +0.5 ถึง +2.5 และค่าการเดามีค่าไม่เกิน 0.3 ส่วนโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนหลายค่า เป็นโมเดล Generalized Partial Credit ที่มีรายการคำตอบ 6 ลำดับชั้น (คะแนน 0-5 คะแนน) 3) ตัวแปรที่ใช้ในการวิจัย ประกอบด้วย 1) ตัวแปรต้น มี 4 ตัวแปร คือ 1.1) สัดส่วนของคะแนนระหว่างข้อสอบที่ตรวจให้คะแนนแบบสองค่ากับข้อสอบที่ตรวจให้คะแนนหลายค่า จำนวน 4 สัดส่วนน้ำหนักคะแนน ได้แก่ 50:50, 60:40, 70:30 และ 80:20 จากสัดส่วนนี้นำมาคิดจำนวนข้อจะได้สัดส่วนดังนี้ 50:10, 60:8, 70:6 และ 80:4 สรุปจะได้แบบสอบ 4 ฉบับที่มีจำนวนข้อสอบ 60, 68, 76, และ 84 ข้อ ตามลำดับ 1.2) ข้อสอบร่วม จะมี 2 มิติ คือ จำนวนข้อสอบทั้งสองชนิดในข้อสอบร่วม และจำนวนคะแนนรวมระหว่างข้อสอบทั้งสองชนิดในข้อสอบร่วม โดยแบบสอบแต่ละฉบับจะมีข้อสอบร่วมอย่างน้อย 20% ซึ่งมีมิติของจำนวนข้อสอบจะมีข้อสอบร่วมที่ต่างกัน 3 ระดับคือ 15:2, 20:2 และ 25:2 ข้อ ส่วนในมิติของคะแนนจะมีคะแนนรวมของข้อสอบร่วมเท่ากับ 20 คะแนน โดยมีสัดส่วนต่างกัน 3 ระดับ คือ 15:5, 10:10 และ 5:15 คะแนน 1.3) แบบแผนการเก็บรวบรวมข้อมูล มี 2 แบบแผน คือ กลุ่มเท่าเทียมโดยใช้ข้อสอบร่วม และกลุ่มไม่เท่าเทียมโดยใช้ข้อสอบร่วม 1.4) วิธีการเปรียบเทียบคะแนน 2 วิธี คือ วิธีโค้งคุณลักษณะ และวิธีการปรับค่าพารามิเตอร์พร้อมกัน ส่วน 2) ตัวแปรตาม คือค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (Mean square error: MSE) ได้ผลการวิจัยดังนี้

1. คุณภาพของการเปรียบเทียบคะแนนด้วยวิธีโค้งคุณลักษณะและการปรับค่าพารามิเตอร์พร้อมกัน มีความคลาดเคลื่อนต่ำ และให้ผลที่ใกล้เคียงกันเมื่อมีรูปแบบการเก็บรวบรวมข้อมูลกับกลุ่มผู้สอบที่มีความสามารถเท่าเทียมกัน สำหรับกลุ่มผู้สอบที่มีความสามารถ

ไม่เท่าเทียมกัน วิธีการเปรียบเทียบด้วยการปรับค่าพารามิเตอร์พร้อมกันจะมีความคลาดเคลื่อนในการเปรียบเทียบต่ำกว่าวิธีโค้งคุณลักษณะ

2. คุณภาพของการปรับเทียบคะแนนทั้งวิธีโค้งคุณลักษณะและการปรับค่าพารามิเตอร์พร้อมกัน มีความคลาดเคลื่อนต่ำลงเมื่อ สัดส่วนของข้อสอบทั้งสองชนิดในแบบสอบมีจำนวนข้อสอบที่ตรวจให้คะแนนสองค่าเพิ่มมากขึ้น และเมื่อสัดส่วนของข้อสอบทั้งสองชนิดในข้อสอบร่วมมีจำนวนข้อสอบที่ตรวจให้คะแนนสองค่าลดลง

3. ผลการเปรียบเทียบค่าเฉลี่ย MSE สำหรับการปรับเทียบคะแนนที่มีสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่ากับหลายค่าในแบบสอบที่แตกต่างกัน 4 สัดส่วน พบว่า ค่าเฉลี่ย MSE ที่ได้จากการปรับเทียบคะแนนสำหรับแบบสอบทั้ง 4 สัดส่วนไม่แตกต่างกัน และค่าเฉลี่ย MSE ที่ได้จากการปรับเทียบคะแนนสำหรับข้อสอบร่วมทั้ง 6 สัดส่วนไม่แตกต่างกัน

4. ผลการปรับเทียบค่าเฉลี่ย MSE สำหรับการปรับเทียบแต่ละวิธีที่มีรูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน พบว่า มีอย่างน้อยหนึ่งคู่ที่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 โดยคู่ที่มีความแตกต่างกันมากที่สุด คือ การปรับเทียบคะแนนด้วยวิธีโค้งคุณลักษณะที่เก็บรวบรวมข้อมูลแบบกลุ่มไม่เท่าเทียม กับการปรับเทียบค่าพารามิเตอร์พร้อมกันกับกลุ่มผู้สอบไม่เท่าเทียม

จิระนาฏ ฉวีพัฒน์ และคณะ (2554) ศึกษาเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวคิดโดยการใช้แบบทดสอบร่วมภายในระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยมีวัตถุประสงค์ของการวิจัยจำนวน 2 ข้อ คือ 1) เพื่อเปรียบเทียบค่าความคลาดเคลื่อนของการปรับเทียบคะแนนแนวคิดโดยการใช้แบบทดสอบร่วมภายใน และ 2) เพื่อเปรียบเทียบค่าความเพียงพอของการปรับเทียบคะแนนแนวคิดโดยการใช้แบบทดสอบร่วมภายใน ระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยการวิจัยครั้งนี้ใช้ข้อมูลจริง ซึ่งเก็บข้อมูลกับนักเรียนระดับช่วงชั้นที่ 3 ปีการศึกษา 2552 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากำแพงเพชร เขต 1 จำนวน 1,080 คน ได้จากการสุ่มหลายขั้นตอน และแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มที่ 1 กลุ่มตัวอย่างสำหรับปรับเทียบคะแนน จำนวน 900 คน แบ่งเป็นระดับชั้นม.1 ม. 2 และ ม. 3 จำนวนระดับชั้นละ 300 คน และกลุ่มที่ 2 เป็นกลุ่มสอบทานผล จำนวน 180 คน ระดับชั้นละ 60 คน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ เรื่องสมการ ระดับช่วงชั้นที่ 3 จำนวน 3 ฉบับ ฉบับละ 30 ข้อ ซึ่งในแต่ละฉบับจะมีข้อสอบร่วมฉบับละ 5 ข้อ แต่สำหรับกลุ่มสอบทานผลแบบสอบแต่ละฉบับจะมีจำนวน 80 ข้อ นำไปเก็บรวบรวมข้อมูลแล้วนำข้อมูลมาปรับเทียบคะแนนด้วยวิธีเชิงเส้นตรง และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ นำผลที่ได้มาหาคุณภาพ

และเปรียบเทียบคุณภาพของการปรับเทียบได้ผลการวิจัยดังนี้ 1) ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนด้วยวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้นตรง อย่างมีนัยสำคัญที่ .05 และ 2) ค่าความเพียงพอของการปรับเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีต่ำกว่าวิธีเชิงเส้นตรง อย่างมีนัยสำคัญที่ .05

จากการศึกษางานวิจัยที่ผ่านมา ผู้วิจัยสรุปข้อมูลประเด็นน่าสนใจดังตารางที่ 2-3

ตารางที่ 2-3 สรุปประเด็นที่เกี่ยวข้องกับประสิทธิผลของการเชื่อมโยงคะแนน

ประเด็น	ประเด็น						
	อดิศร ศรีบุญวงษ์ (2545) กิตพงษ์ ลีอนาม (2551) อัญชลี ศรีกลดชาญ (2552) จิระนาฏ ฉวีพัฒน์ และคณะ (2554) Zhang (2007) Li (2009)						
ที่มาข้อมูล	ข้อมูลจริง		1		1		
	จำลอง	1				1	1
การประมาณ	IRT	1,3PL	3PL	3PL	3PL	3PL	3PL
	TRT						
	Polytomous		GRM	GPC		GPC	GRM
การเก็บรวบรวมข้อมูล	กลุ่มเท่าเทียมแบบสอบต่างฉบับ					1	
	กลุ่มเท่าเทียมแบบสอบเท่าเทียม						
	กลุ่มเท่าเทียมข้อสอบร่วม	1		1			
	กลุ่มไม่เท่าเทียมข้อสอบร่วม			1	1		1

ตารางที่ 2-3 (ต่อ)

	ประเด็น	อติศร ศรีบุญวงษ์ (2545) กิตพงษ์ ลือนาม (2551) อัญชติ ศรีกวดชาญ (2552) จิระนาฏ ฉวีพัฒน์ และคณะ (2554)					Zhang (2007)	Li (2009)
วิธีการปรับเทียบ	อีควิเปอร์เซ็นต์ไทล์						1	
	เชิงเส้นตรง						1	
	โค้งคุณลักษณะ	1			1	1		1
	คะแนนจริง							
	Newton-Rhapson							
	MASE & RE			1				
	ปรับพารามิเตอร์พร้อมกัน				1			
ประสิทธิภาพ	ความเสมอภาค							
	ความเพียงพอ			1		1		
	ความคลาดเคลื่อน	1	1	1			1	1
	สัมประสิทธิ์การทำนาย		1					

ตารางที่ 2-3 (ต่อ)

		ประเด็น							
		Puhan et al.(2009)	Rui et al. (2010)	He et al. (2012)	Cao et al. (2014)	Chen (2014)	Tao and Cao (2016)	Kogar (2017)	
ที่มาข้อมูล	ข้อมูลจริง	1		1		1		1	
	จำลอง		1		1	1	1		
การประมาณค่า	IRT		2PL	1-3	2PL	3PL	3PL	2PL	
	TRT		2PT	2-3	2PT	3PL	3PL	2PL	
	Polytomous				GRM	GRM			
การเก็บรวบรวมข้อมูล	กลุ่มเท่าเทียมแบบสอบต่างฉบับ				1	1			
	กลุ่มเท่าเทียมแบบสอบเท่าเทียม	1							
	กลุ่มเท่าเทียมข้อสอบร่วม								
	กลุ่มไม่เท่าเทียมข้อสอบร่วม	1		1			1		
วิธีการปรับเทียบ	อควิเปอร์เซ็นไทล์			1	1	1	1		
	เชิงเส้นตรง	1		1					
	โค้งคุณลักษณะ		1						
	คะแนนจริง								
	Newton-Rhapson				1		1		
	MASE & RE								
	ปรับพารามิเตอร์พร้อมกัน								
ประสิทธิภาพ	ความเสมอภาค								
	ความเพียงพอ								
	ความคลาดเคลื่อน	1	1	1	1	1	1	1	
	สัมประสิทธิ์การทำนาย								

จากผลการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับประสิทธิผลการเชื่อมโยงคะแนนระหว่างแบบสอบทดสอบทดสอบ โดยการศึกษาครั้งใช้ข้อมูลทุติยภูมิ เป็นคะแนนสอบจากการทดสอบ การศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 เนื่องจากว่าแบบสอบที่ใช้วัดนักเรียนนั้นมีลักษณะของทดสอบประกอบอยู่ในสองวิชาคือ วิชาภาษาไทย และวิชาภาษาอังกฤษ ดังนั้นทำการศึกษาเอกสารและงานวิจัยสรุปได้ว่า

การเชื่อมโยงคะแนน หมายถึง การแปลงคะแนนของแบบสอบหนึ่งให้เท่าเทียมกับคะแนนอีกแบบสอบหนึ่ง ซึ่งแบบสอบนั้นจะมีความเป็นคู่ขนานหรือไม่คู่ขนานก็ได้ โดยการศึกษาครั้งใช้หลักการเชื่อมโยงคะแนนรูปแบบการทำนาย (Predictive) เพราะแบบสอบมีโครงสร้างและคุณลักษณะที่แตกต่างกัน ดังนั้นการศึกษาค้นคว้าครั้งนี้เป็นการเชื่อมโยงคะแนนวิชาภาษาอังกฤษให้เท่าเทียมกับคะแนนวิชาภาษาไทย

การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบ จากงานวิจัยในช่วง 10 ปีนี้พบว่า จะมีการศึกษาการเชื่อมโยงคะแนนในรูปแบบของคะแนนจริงและคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แต่แบบสอบที่นำมาใช้ในการศึกษามีลักษณะของความไม่เป็นอิสระกันของข้อสอบ จึงได้มีการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองทดสอบที่มากขึ้น ซึ่งเทคนิควิธีการที่ใช้ส่วนใหญ่ของการเชื่อมโยงคะแนนจริง คือ วิธีโค้งคุณลักษณะ (Characteristic curve) ส่วนเทคนิควิธีการเชื่อมโยงคะแนนสังเกตได้ วิธีอีควิเปอร์เซ็นต์ไทล์ (Equipercentile) ดังนั้น การศึกษาศึกษาค้นคว้าครั้งนี้ผู้วิจัยใช้การประมาณค่าพารามิเตอร์ข้อสอบ 2 วิธี คือ การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ 3 พารามิเตอร์ และการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองทดสอบ 3 พารามิเตอร์ ส่วนวิธีการเชื่อมโยงคะแนนนั้นใช้ 2 วิธีคือ วิธีโค้งคุณลักษณะของแบบสอบ (Test characteristic curve) และวิธีอีควิเปอร์เซ็นต์ไทล์ (Equipercentile)

ประสิทธิผลการของการเชื่อมโยงคะแนน จากงานวิจัยที่ศึกษาพบว่า พิจารณาจากความคลาดมาตรฐานของการปรับเทียบคะแนน ดังนั้นในการศึกษาค้นคว้าครั้งนี้จะใช้ความคลาดเคลื่อนมาตรฐานในการพิจารณาประสิทธิผลเช่นเดียวกัน โดยความคลาดเคลื่อนจะพิจารณา 3 ค่า คือ ค่าความลำเอียง (bias) ความคลาดเคลื่อนมาตรฐาน (SEE) และดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error)

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาครั้งนี้มีวัตถุประสงค์หลักเพื่อเปรียบเทียบประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบทดสอบทดสอบเฉพาะตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีการเชื่อมโยงคะแนนจริงและวิธีการเชื่อมโยงคะแนนสังเกตได้ ภายใต้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน ผู้วิจัยออกแบบขั้นตอนการดำเนินการวิจัยออกเป็น 4 ส่วน คือ ประชากรและกลุ่มตัวอย่าง เครื่องมือที่ใช้ในการวิจัย การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2558 จำนวน 42 เขตพื้นที่การศึกษา รวมจำนวนนักเรียนเป็น 414,287 คน (กระทรวงศึกษาธิการ, 2559)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสังกัดสำนักเขตพื้นที่การศึกษามัธยมศึกษา ปีการศึกษา 2558 โดยผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างจากประชากร เพื่อใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบแบบ 3 พารามิเตอร์ ซึ่งมีนักวิจัยบางกลุ่มให้ข้อเสนอแนะว่าขนาดกลุ่มตัวอย่างจำนวน 700-1,500 คน ก็เพียงพอ (Lam, 2009, p. 14) และบางกลุ่มเสนอแนะว่าขนาดกลุ่มตัวอย่าง 1,000 คน ก็เพียงพอสำหรับการประมาณค่าพารามิเตอร์ข้อสอบ (De Ayala, 2009, p.131; Foley, 2010, p.16) ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยใช้กลุ่มตัวอย่างจำนวน 1,500 คน ได้จากวิธีการสุ่มแบบหลายขั้นตอน (Multi-stage sampling) มีวิธีการสุ่มดังนี้

ขั้นตอนที่ 1 สุ่มตัวอย่างแบบแบ่งชั้น สุ่มสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา มาจากชั้นแต่ละภาคของประเทศตามสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ อันได้แก่ ภาคกลาง (ประกอบด้วย 13 เขตพื้นที่การศึกษา) ภาคเหนือ (ประกอบด้วย 8 เขตพื้นที่การศึกษา) ภาคใต้ (ประกอบด้วย 6 เขตพื้นที่การศึกษา) และภาคตะวันออกเฉียงเหนือ (ประกอบด้วย 15 เขตพื้นที่การศึกษา) สุ่มภาคละ 3 เขตพื้นที่การศึกษา เป็นจำนวนทั้งสิ้น 12 เขตพื้นที่การศึกษา ซึ่งผู้วิจัยทำการสุ่มอย่างเป็นระบบ โดยเรียงลำดับ

เขตพื้นที่การศึกษามัธยมศึกษาเขต 1 ถึงเขต 42 จากนั้นทำการสุ่มตัวเลข 1-4 ขึ้นมาหนึ่งตัว ในที่นี้คือเลข 4 ซึ่งเป็นลำดับที่แรกในการสุ่ม จากนั้นนับต่อไปที่ละ 4 ได้เขตพื้นที่การศึกษาแต่ละภาค ดังนี้ ภาคกลาง ประกอบด้วย สพม.เขต 2 เขต 4 และเขต 8 ภาคเหนือ ประกอบด้วย สพม.เขต 34 เขต 36 และเขต 40 ภาคใต้ ประกอบด้วย สพม.เขต 12 เขต 14 และเขต 16 และภาคตะวันออกเฉียงเหนือ ประกอบด้วย สพม.เขต 20 เขต 28 และเขต 32

ขั้นตอนที่ 2 สุ่มตัวอย่างแบบแบ่งกลุ่ม เมื่อได้เขตพื้นที่การศึกษาตัวอย่างแล้ว ใช้เขตพื้นที่การศึกษาเป็นกลุ่ม แล้วกำหนดตัวแทนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ตามสัดส่วนของประชากร จากนั้นทำการสุ่มอย่างง่าย ได้จำนวนกลุ่มตัวอย่างดังตารางที่ 3-1

ตารางที่ 3-1 จำนวนประชากรและกลุ่มตัวอย่าง

ภาค	เขตพื้นที่	ประชากร(คน)	กลุ่มตัวอย่าง (คน)
กลาง	สพม.2	23,526	268
	สพม.4	11,849	135
	สพม.8	10,545	120
	รวม	45,920	523
เหนือ	สพม.34	7,132	81
	สพม.36	8,762	100
	สพม.40	5,097	58
	รวม	20,991	239
ใต้	สพม.12	14,830	169
	สพม.14	5,240	60
	สพม.16	10,835	124
	รวม	30,905	353
ตะวันออกเฉียงเหนือ	สพม.20	10,160	116
	สพม.28	12,713	145
	สพม.32	10,834	124
	รวม	33,707	385
รวม		131,432	1,500

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้คือ แบบสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาไทย และวิชาภาษาอังกฤษ ซึ่งผ่านกระบวนการสร้างอย่างเป็นมาตรฐาน โดยสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) และแบบสอบทั้งสองฉบับนี้มีลักษณะที่เป็นไปตามเงื่อนไขของการวิจัย คือ ความไม่อิสระในการตอบข้อสอบของแต่ละข้อหรือที่เรียกว่าเทสต์เลท (Testlets) โดยมีรายละเอียดแบบสอบดังนี้

ตารางที่ 3-2 โครงสร้างและข้อสอบที่มีลักษณะเทสต์เลท

แบบสอบ	โครงสร้างแบบสอบ	ข้อที่เป็นเทสต์เลท
วิชาภาษาไทย	เป็นแบบสอบปรนัย จำนวน 50 ข้อ ดังนี้	ข้อที่ 1-10
	1. ข้อที่ 1-40 เป็นแบบปรนัย 4 ตัวเลือก	ข้อที่ 11-15
	คำตอบถูกที่สุดเพียงข้อเดียว ข้อละ 2 คะแนน	ข้อที่ 16-20
	2. ข้อที่ 41-50 เป็นแบบปรนัย 5 ตัวเลือก	ข้อที่ 21-25
	เลือกสองคำตอบที่ถูกต้อง ข้อละ 2 คะแนน	ข้อที่ 26-30
วิชาภาษาอังกฤษ	เป็นแบบสอบปรนัย จำนวน 50 ข้อ ดังนี้	ข้อที่ 2-3
	1. ข้อที่ 1-50 เป็นแบบปรนัย 4 ตัวเลือก	ข้อที่ 4-5
	คำตอบถูกที่สุดเพียงข้อเดียว ข้อละ 2 คะแนน	ข้อที่ 11-15
		ข้อที่ 16-20
		ข้อที่ 21-23
		ข้อที่ 26-30
		ข้อที่ 31-35
		ข้อที่ 36-40

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้ ผู้วิจัยได้ขอความอนุเคราะห์ข้อมูลจากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) เป็นข้อมูลเกี่ยวกับคะแนนสอบรายข้อจากแบบสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาไทย และวิชาภาษาอังกฤษของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558 โดยมีขั้นตอนการเก็บรวบรวมข้อมูลดังนี้

1. ขอนหนังสือจากคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เพื่อขอข้อมูลคะแนนสอบรายข้อวิชาภาษาไทย และวิชาภาษาอังกฤษของนักเรียนที่เป็นกลุ่มตัวอย่างจากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน)
2. ขอความร่วมมือจากเจ้าหน้าที่สถาบันการทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) ในการส่งข้อมูลคะแนนสอบรายข้อของกลุ่มตัวอย่าง ตามคุณลักษณะและขนาดกลุ่มตัวอย่างที่กำหนดในการวิจัย
3. ตรวจสอบความเรียบร้อยของข้อมูล จากนั้นนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

การวิเคราะห์ข้อมูล

ในการวิเคราะห์ข้อมูลนั้น ผู้วิจัยได้ดำเนินการวิเคราะห์ข้อมูลดังนี้

1. คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง เพื่อบรรยายลักษณะการแจกแจงของคะแนนการสอบวิชาภาษาไทย และวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558
2. การประมาณค่าพารามิเตอร์ข้อสอบ ประกอบด้วย ค่าอำนาจจำแนก ค่าความยาก ค่าการเดา และค่าความสามารถผู้สอบตามโมเดลการประมาณค่าทฤษฎีการตอบสนองข้อสอบ และทฤษฎีการตอบสนองเทสต์เลท และประมาณค่าพารามิเตอร์เกี่ยวกับความไม่เป็นอิสระกันของข้อสอบ เพิ่มในส่วนของทฤษฎีการตอบสนองเทสต์เลท ก่อนการประมาณค่าพารามิเตอร์นั้นผู้วิจัยได้นำคะแนนที่ได้จาก สทศ. มาแปลงข้อมูลให้เหมาะสมสำหรับการวิเคราะห์ โดยการแปลงคะแนนข้อสอบที่ตอบถูกให้ 2 คะแนน เป็น 1 คะแนน จากนั้นนำข้อมูลที่แปลงแล้วไปวิเคราะห์ต่อโดยมีรายละเอียดดังนี้

2.1 การประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบตามโมเดลทฤษฎีการตอบสนองข้อสอบ

การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบที่ใช้ในการศึกษาครั้งนี้เป็นการประมาณค่าแบบ 3 พารามิเตอร์ ผู้วิจัยประมาณค่าพารามิเตอร์จากโปรแกรม SCORIGHT โดยมีสมการดังนี้ (Wainer, et al., 2000; Wainer, et al., 2007)

สมการสำหรับข้อสอบที่มีการตรวจให้คะแนน 2 ค่า

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[1.7a_i(\theta_j - b_i)]}}{1 + e^{[1.7a_i(\theta_j - b_i)]}} \quad (1)$$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j
 θ_j คือ พารามิเตอร์ความสามารถของคนี่ j
 a_i คือ พารามิเตอร์อำนาจจำแนกของข้อสอบ i
 b_i คือ พารามิเตอร์ความยากของข้อสอบ i
 c_i คือ พารามิเตอร์การเดาของข้อสอบ i
 e คือ ค่าคงที่ของลอการิทึมธรรมชาติ มีค่าเท่ากับ 2.718

สมการสำหรับข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า

$$P_{ijk} = P_{ijk}^* - P_{ij(k+1)}^*$$

$$P_{ijk}^* = \frac{e^{[1.7a_i(\theta_j - b_{ik})]}}{1 + e^{[1.7a_i(\theta_j - b_{ik})]}} \quad (2)$$

เมื่อ P_{ijk} คือ ความน่าจะเป็นในการเลือกรายการคำตอบ k ข้อสอบที่ i ของคนที่ j
 b_{ik} คือ พารามิเตอร์ความยากของรายการคำตอบ k ข้อสอบที่ i

2.2 การประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบตาม
 โมเดลทฤษฎีการตอบสนองของเทสต์เลท

การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองของเทสต์เลทที่ใช้ในการศึกษาครั้งนี้เป็นการประมาณค่าแบบ 3 พารามิเตอร์ ผู้วิจัยประมาณค่าพารามิเตอร์จากโปรแกรม SCORIGHT โดยมีสมการดังนี้ (Wainer, et al., 2000; Wainer, et al., 2007)

สมการสำหรับข้อสอบที่มีการตรวจให้คะแนน 2 ค่า

$$P_{ij} = c_i + (1 - c_i) \frac{e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}}{1 + e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}} \quad (3)$$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j
 θ_j คือ พารามิเตอร์ความสามารถของคนี่ j
 a_i คือ พารามิเตอร์อำนาจจำแนกของข้อสอบ i

b_i คือ พารามิเตอร์ความยากของข้อสอบ i

c_i คือ พารามิเตอร์การเดาของข้อสอบ i

$\gamma_{jd(i)}$ คือ อิทธิพลสุ่มของข้อสอบ i กับคนที่ j ภายในเทสต์เลข d

e คือ ค่าคงที่ของลอการิทึมธรรมชาติ มีค่าเท่ากับ 2.718

สมการสำหรับข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า

$$P_{ijk} = P_{ijk}^* - P_{ij(k+1)}^*$$

$$P_{ijk}^* = \frac{e^{[1.7a_i(\theta_j - b_{ik} - \gamma_{jd(i)})]}}{1 + e^{[1.7a_i(\theta_j - b_{ik} - \gamma_{jd(i)})]}} \quad (4)$$

เมื่อ P_{ijk} คือ ความน่าจะเป็นในการเลือกการคำตอบ k ข้อสอบที่ i ของคนที่ j

b_{ik} คือ พารามิเตอร์ความยากของรายการคำตอบ k ข้อสอบที่ i

การประมาณค่าพารามิเตอร์ด้วยโปรแกรม SCORIGHT นั้นมีการกำหนดลักษณะ

การแจกแจงของข้อมูล ดังนี้

$$\theta_j \sim N(0, 1)$$

$$\log(a_i) \sim N(\mu_a, \sigma_a^2)$$

$$b_i \sim N(\mu_b, \sigma_b^2)$$

$$\log(c_i) \sim N(\mu_c, \sigma_c^2)$$

$$\gamma_{jd(i)} \sim N(0, \sigma_\gamma^2)$$

2.3 คำนวนสารสนเทศของแบบสอบด้วยการรวมสารสนเทศของข้อสอบ ณ

ความสามารถนั้น ๆ โดยใช้สูตรในการคำนวณสารสนเทศข้อสอบของ Fisher (Wainer, et al., 2007) โดยมีสมการดังนี้

$$I_i(\theta_j) = a_i^2 \left(\frac{e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}}{1 + e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}} \right)^2 \frac{1 - c_i}{c_i + e^{[1.7a_i(\theta_j - b_i - \gamma_{jd(i)})]}}$$

$$I(\theta_j) = \sum I_i(\theta_j)$$

3. การวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์ของการวิจัยเกี่ยวกับการศึกษา

ความสัมพันธ์ของค่าพารามิเตอร์ของข้อสอบระหว่างโมเดลการประมาณค่าตามทฤษฎีการ

ตอบสนองข้อสอบ และทฤษฎีการตอบสนองทดสอบทดสอบ โดยการใช้วิเคราะห์สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's correlation coefficients: r) โดยมีเกณฑ์พิจารณาระดับความสัมพันธ์ (Cohen, 1988 อ้างถึงใน จุฬาลักษณ์ บารมี, 2555, หน้า 160) ดังนี้

$r = 0.01$ ถึง 0.30 แสดงถึงการมีระดับความสัมพันธ์กันต่ำ

$r = 0.31$ ถึง 0.70 แสดงถึงการมีระดับความสัมพันธ์กันปานกลาง

$r =$ สูงกว่า 0.70 แสดงถึงการมีระดับความสัมพันธ์กันสูง

4. การวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์ของการวิจัยเกี่ยวกับการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบทดสอบ ด้วยสถิติ Paired t -test

5. การวิเคราะห์ข้อมูลเกี่ยวกับการเชื่อมโยงคะแนนระหว่างแบบสอบ โดยการปรับคะแนนวิชาภาษาอังกฤษให้เท่าเทียมหรือสมมูลกับคะแนนวิชาภาษาไทย มีวิธีการเชื่อมโยงตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ 1) การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ และ 2) การเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ โดยมีรายละเอียดดังนี้

5.1 การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ

การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะของแบบสอบ สามารถอธิบายขั้นตอนการดำเนินการดังนี้

5.1.1 ผู้วิจัยเริ่มจากคำนวณค่าคงที่ A และ B ด้วยวิธี mean/σ เพื่อใช้ในการเชื่อมโยงคะแนนระหว่างแบบสอบ โดยการนำค่าพารามิเตอร์ความยากของข้อสอบวิชาภาษาไทยและวิชาภาษาอังกฤษที่ประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบทดสอบมาหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน แล้วแทนค่าในสมการ ดังนี้ (Kolen & Brennan, 2014)

$$A = \frac{\sigma(b_T)}{\sigma(b_E)}$$

$$B = \mu(b_T) - A * \mu(b_E)$$

เมื่อ $\sigma(b_T)$ คือ ส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของวิชาภาษาไทย

$\sigma(b_E)$ คือ ส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของวิชาภาษาอังกฤษ

$\mu(b_T)$ คือ ค่าเฉลี่ยของพารามิเตอร์ความยากของวิชาภาษาไทย

$\mu(b_E)$ คือ ค่าเฉลี่ยของพารามิเตอร์ความยากของวิชาภาษาอังกฤษ

5.1.2 คำนวณความน่าจะเป็นในการตอบข้อสอบแต่ละถูกของนักเรียนด้วยสมการ (1) และ (2) ในหัวข้อ 2.1 สำหรับการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบในวิชาภาษาไทยและวิชาภาษาอังกฤษ และคำนวณตามสมการ (3) และ (4) ในหัวข้อ 2.2 สำหรับการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองเทสต์เลทในวิชาภาษาไทยและภาษาอังกฤษ แล้วนำความน่าจะเป็นในการตอบถูกแต่ละข้อมาบวกกัน เรียกว่าคะแนนจริง จะได้ข้อมูลคะแนนจริงทั้งหมด 4 ชุด คือ คะแนนจริงตามโมเดลทฤษฎีการตอบสนองข้อสอบวิชาภาษาไทย (T IRT) คะแนนจริงตามโมเดลทฤษฎีการตอบสนองข้อสอบวิชาภาษาอังกฤษ (E IRT) คะแนนจริงตามโมเดลทฤษฎีการตอบสนองเทสต์เลทวิชาภาษาไทย (T TRT) และคะแนนจริงตามโมเดลทฤษฎีการตอบสนองเทสต์เลทวิชาภาษาอังกฤษ (E TRT)

5.1.3 ดำเนินการเชื่อมโยงคะแนนจริงของแบบสอบภาษาอังกฤษให้เท่าเทียมกับแบบสอบภาษาไทยภายใต้การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบด้วยวิธีโค้งคุณลักษณะแบบสอบของ Stocking and Lord (1983) คำนวณดังสมการ

$$P_{ij} = c_{iEN} + (1 - c_{iEN}) \frac{e^{[1.7a_{iEN}(\theta_{jT} - b_{iEN})]}}{1 + e^{[1.7a_{iEN}(\theta_{jT} - b_{iEN})]}}$$

โดยที่ $a_{iEN} = a_{iE}/A$

$$b_{iEN} = A * b_{iE} + B$$

$$c_{iEN} = c_{iE} - (B/A) * c_{iE}$$

เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j

θ_{jT} คือ พารามิเตอร์ความสามารถคนที่ j จากแบบสอบภาษาไทย

a_{iEN} คือ พารามิเตอร์อำนาจจำแนกข้อสอบ i ของแบบสอบ
ภาษาอังกฤษปรับเท่าเทียมแบบสอบภาษาไทย

b_{iEN} คือ พารามิเตอร์ความยากข้อสอบ i ของแบบสอบ
ภาษาอังกฤษปรับเท่าเทียมแบบสอบภาษาไทย

c_{iEN} คือ พารามิเตอร์การเดาข้อสอบ i ของแบบสอบ

- ภาษาอังกฤษปรับเท่าเทียมแบบสอบภาษาไทย
- a_{iE} คือ พารามิเตอร์อำนาจจำแนกข้อสอบ i ของแบบสอบ
ภาษาอังกฤษ
- b_{iE} คือ พารามิเตอร์ความยากข้อสอบ i ของแบบสอบ
ภาษาอังกฤษ
- c_{iE} คือ พารามิเตอร์การเดาข้อสอบ i ของแบบสอบ
ภาษาอังกฤษ

5.1.4 ดำเนินการเชื่อมโยงคะแนนจริงของแบบสอบภาษาอังกฤษให้เท่าเทียมกับแบบสอบภาษาไทยภายใต้การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองของเทสต์เลทด้วยวิธีได้งคุณลักษณะแบบสอบผู้วิจัยปรับสมการของ Li et al. (2005) ที่พัฒนามาจากสมการของ Stocking and Lord (1983) ดังนี้

$$P_{ij} = c_{iEN} + (1 - c_{iEN}) \frac{e^{[1.7a_{iEN}(\xi_{jd} - b_{iEN})]}}{1 + e^{[1.7a_{iEN}(\xi_{jd} - b_{iEN})]}}$$

โดยที่ $a_{iEN} = a_{iE}/A$

$$b_{iEN} = (A * b_{iE}) + (A * \mu_{\gamma_{dT}}) - B$$

$$c_{iEN} = c_{iE} - (B/A) * c_{iE}$$

$$\xi_{jg} = \theta_{jT} - \gamma_{jdt}$$

- เมื่อ P_{ij} คือ ความน่าจะเป็นในการตอบข้อสอบถูกข้อที่ i ของคนที่ j
- θ_{jT} คือ พารามิเตอร์ความสามารถคนที่ j จากแบบสอบภาษาไทย
- a_{iEN} คือ พารามิเตอร์อำนาจจำแนกข้อสอบ i ของแบบสอบ
ภาษาอังกฤษปรับสเกลเท่าเทียมแบบสอบภาษาไทย
- b_{iEN} คือ พารามิเตอร์ความยากข้อสอบ i ของแบบสอบ
ภาษาอังกฤษปรับสเกลเท่าเทียมแบบสอบภาษาไทย
- c_{iEN} คือ พารามิเตอร์การเดาข้อสอบ i ของแบบสอบ
ภาษาอังกฤษปรับสเกลเท่าเทียมแบบสอบภาษาไทย
- a_{iE} คือ พารามิเตอร์อำนาจจำแนกข้อสอบ i ของแบบสอบ
ภาษาอังกฤษ
- b_{iE} คือ พารามิเตอร์ความยากข้อสอบ i ของแบบสอบ
ภาษาอังกฤษ

c_{iE} คือ พารามิเตอร์การเดาข้อสอบ i ของแบบสอบภาษาอังกฤษ

γ_{jdT} คือ อิทธิพลสุ่มของข้อสอบ i กับคนที่ j ภายในเทสต์เลข d จากแบบสอบภาษาไทย

$\mu_{y_{dT}}$ คือ ค่าเฉลี่ยอิทธิพลสุ่มภายในเทสต์เลข d จากแบบสอบภาษาไทย

5.2 การเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิควิเปอร์เซ็นต์ไทล์

การเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิควิเปอร์เซ็นต์ไทล์นั้น ผู้วิจัยนำข้อมูล

ความน่าจะเป็นในการตอบถูกของนักเรียนที่ได้จากการคำนวณข้อ 5.1.3-5.1.4 มาใช้ในการสร้างคะแนนสังเกตได้ แล้วทำการเชื่อมโยงคะแนนด้วยวิธีอิควิเปอร์เซ็นต์ไทล์ มีขั้นตอนการดำเนินการดังนี้

5.2.1 จัดเตรียมไฟล์ข้อมูลความน่าจะเป็นของการตอบข้อสอบถูก (P) และความน่าจะเป็นของการตอบข้อสอบผิด ($Q = 1-P$) ของแบบสอบภาษาไทย และแบบสอบภาษาอังกฤษที่เชื่อมโยงกับแบบสอบภาษาไทยทั้งจากการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองเทสต์เลข

5.2.2 แปลงความน่าจะเป็นการตอบข้อสอบถูกและการตอบข้อสอบผิด เป็นคะแนนสังเกตได้ แล้วหาความถี่ของแต่ละคะแนนตามหลักการของ Lord and Wingersky (1984) ดังนี้

$$P_r(x|\theta_j) = \begin{cases} P_{r-1}(x) = 1 & ; r = 1 \\ P_{r-1}(x|\theta_j)(1 - P_{jr}) & ; x = 0, r \neq 0 \\ P_{r-1}(x|\theta_j)(1 - P_{jr}) + P_{r-1}(x - 1|\theta_j)(P_{jr}) & ; 0 < x < r \\ P_{r-1}(x - 1|\theta_j)(P_{jr}) & ; x = r \end{cases}$$

ขั้นตอนนี้ผู้วิจัยวิเคราะห์โดยใช้โปรแกรม R package Classify โดยมีตัวอย่างคำสั่งในการวิเคราะห์ดังนี้

```
library("classify")
start_student<-1
finish_student<-1500
file_p=read.table("C:/Users/Ultimate/Desktop/classify/P_Ttrt.txt",sep = "")
file_q=read.table("C:/Users/Ultimate/Desktop/classify/Q_Ttrt.txt",sep = "")
count_cats<-c()
for(i in 1:50){
count_cats[i]=2 }
```

```

summary_prob<-data.frame()
for(student in start_student:finish_student){
data_p<-file_p[1:50,student]
data_q<-file_q[1:50,student]
data<-cbind(data_p,data_q)
probs=matrix(c(data_p,data_q),nrow=50,ncol=2,byrow=FALSE)
data_prob<-wlord(probs,cats=count_cats)
for(i in 1:51){
summary_prob[student,i]<-data_prob[i] }}
write.csv(summary_prob,file="C:/Users/Ultimate/Desktop/classify/
Ttrt_out.csv")

```

5.2.3 นำผลที่ได้จากข้อ 5.2.2 มาเชื่อมโยงคะแนนสังเกตได้วิธีอิควิเปอร์ตีไทร์
ด้วยโปรแกรม RAGE-RGEQUATE ด้วยหลักการเชื่อมโยงคะแนนดังสมการ (Kolen and
Brennan, 2014)

$$Q^{-1}[P(x)] = \frac{P(x)/100 - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + (y_U^* - 0.5) \quad ; 0 \leq P(x) < 100$$

เมื่อ $Q^{-1}[P(x)]$ คือ คะแนนแบบสอบภาษาไทยที่มีตำแหน่งเปอร์เซ็นต์ไทล์
ของคะแนน x ในแบบสอบภาษาอังกฤษ

$P(x)$ คือ ตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนน x ในแบบสอบ
ภาษาอังกฤษ

$G(y_U^* - 1)$ คือ ความถี่สะสมก่อนถึงขั้นที่ตำแหน่งเปอร์เซ็นต์ไทล์แบบสอบ
ภาษาไทยใกล้เคียงกับแบบสอบภาษาอังกฤษ

$G(y_U^*)$ คือ ความถี่สะสมขั้นที่ตำแหน่งเปอร์เซ็นต์ไทล์แบบสอบ
ภาษาไทยใกล้เคียงกับแบบสอบภาษาอังกฤษ

y_U^* คือ คะแนนของแบบสอบภาษาไทยที่ตำแหน่งเปอร์เซ็นต์ไทล์
ใกล้เคียงกับแบบสอบภาษาอังกฤษ

6. การวิเคราะห์เพื่อตอบวัตถุประสงค์การวิจัยเกี่ยวกับประสิทธิผลของการเชื่อมโยงคะแนน โดยพิจารณาความคลาดเคลื่อนของการเชื่อมโยงคะแนนจาก 3 ค่า คือ ค่าความลำเอียง (Bias) ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) และ ค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error) ผู้วิจัยคำนวณโดยใช้โปรแกรม Microsoft office excel โดยมีรายละเอียดดังนี้

6.1 ค่าความลำเอียง (Bias) เป็นความคลาดเคลื่อนอย่างเป็นทางการปรับเทียบคะแนน สามารถคำนวณได้จากสมการดังนี้ (Ojerinde, Popoola, Onyeneho, & Egberongbe, 2016)

$$Bias = \frac{\sum f(E_{iN} - E_i)}{\sum f}$$

เมื่อ Bias คือ ความลำเอียง

f คือ ความถี่ของแต่ละคะแนน

E_{iN} คือ คะแนนภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย

E_i คือ คะแนนภาษาอังกฤษเดิม

6.2 ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (Root mean square error) เป็นการรวมความคลาดเคลื่อนอย่างเป็นทางการ และความคลาดเคลื่อนอย่างสุ่ม สามารถคำนวณได้จากสมการ ดังนี้ (Ojerinde, et al., 2016)

$$RMSE = \sqrt{\frac{\sum f(E_{iN} - E_i)^2}{\sum f}}$$

เมื่อ RMSE คือ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง

f คือ ความถี่ของแต่ละคะแนน

E_{iN} คือ คะแนนภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย

E_i คือ คะแนนภาษาอังกฤษเดิม

6.3 ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) เป็นความคลาดเคลื่อนอย่างสุ่มของการปรับเทียบ สามารถคำนวณได้จากสมการดังนี้ (Ojerinde, et al., 2016)

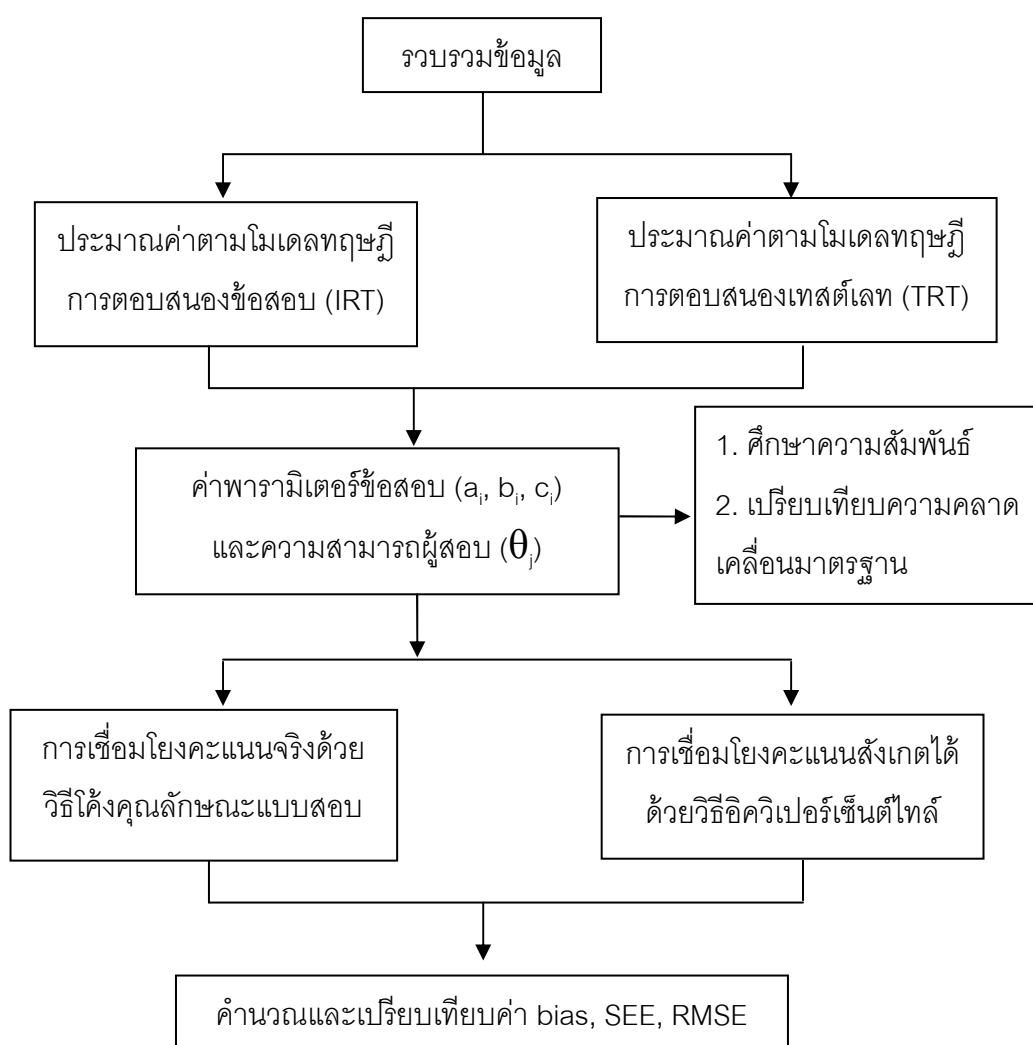
$$SEE = \sqrt{RMSE^2 - Bias^2}$$

เมื่อ *Bias* คือ ความลำเอียง

RMSE คือ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง

SEE คือ ความคลาดเคลื่อนมาตรฐานของการปรับเทียบ

สรุปขั้นตอนการดำเนินการวิเคราะห์ข้อมูลดังแผนภาพ



ภาพที่ 3-1 ขั้นตอนการดำเนินการวิเคราะห์ข้อมูล

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์ 3 ประการ คือ 1) เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบ ระหว่างการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบ (2) เพื่อเปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองทดสอบ และ 3) เพื่อเปรียบเทียบประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีการเชื่อมโยงคะแนนจริง และวิธีการเชื่อมโยงคะแนนสังเกตได้ภายใต้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน ดังนั้นผู้วิจัยนำเสนอผลการวิเคราะห์ข้อมูลตามรายละเอียดดังนี้

ตอนที่ 1 ข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียน

ตอนที่ 2 ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถของผู้สอบ

ตอนที่ 3 ประสิทธิภาพของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบ

สัญลักษณ์ที่ใช้ในการวิจัย

n	หมายถึง	จำนวนข้อ/จำนวนคน
MIN	หมายถึง	คะแนนต่ำสุด
MAX	หมายถึง	คะแนนสูงสุด
$MEAN$	หมายถึง	ค่าเฉลี่ย
SD	หมายถึง	ค่าเบี่ยงเบนมาตรฐาน
$\sigma_{d(i)}^2$	หมายถึง	ค่าความแปรปรวนของอิทธิพลทดสอบ
CV	หมายถึง	สัมประสิทธิ์การกระจาย
Sk	หมายถึง	ความเบ้
Ku	หมายถึง	ความโด่ง
a	หมายถึง	ค่าอำนาจจำแนก
b	หมายถึง	ค่าความยาก
c	หมายถึง	ค่าการเดา

θ	หมายถึง	ค่าความสามารถของผู้สอบ
t	หมายถึง	สถิติทดสอบค่าเฉลี่ย t -test
sig	หมายถึง	ระดับนัยสำคัญในการยอมรับสมมติฐานหลัก
*	หมายถึง	มีนัยสำคัญทางสถิติที่ระดับ 0.05
**	หมายถึง	มีนัยสำคัญทางสถิติที่ระดับ 0.01
r	หมายถึง	สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร
$IRT\ 3PL$	หมายถึง	ทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์
$TRT\ 3PL$	หมายถึง	ทฤษฎีการตอบสนองเทสต์เลท แบบ 3 พารามิเตอร์
TS	หมายถึง	คะแนนจริงผ่าน IRT (IRT True Scores)
OS	หมายถึง	คะแนนสังเกตได้ผ่าน IRT (IRT Observed Scores)
M_SL	หมายถึง	คะแนนเฉลี่ยความแตกต่างกำลังสองของการเชื่อมโยง คะแนนวิธีโค้งคุณลักษณะแบบสอบของ Stocking and Lord
$Bias$	หมายถึง	ความลำเอียงของการเชื่อมโยงคะแนน
$RMSE$	หมายถึง	ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลัง สองของการเชื่อมโยงคะแนน
SEE	หมายถึง	ความคลาดเคลื่อนมาตรฐานของการเชื่อมโยงคะแนน

ตอนที่ 1 ข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียน

ในการศึกษาครั้งนี้ผู้วิจัยใช้ข้อมูลทุติยภูมิจากสถาบันการทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) เป็นข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในรายวิชาภาษาไทย และรายวิชาภาษาอังกฤษ จำนวน 1,500 คน จาก 12 เขตพื้นที่การศึกษามัธยมศึกษา มีรายละเอียดดังตารางที่ 4-1

ตารางที่ 4-1 ข้อมูลคะแนนสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียน

รายการ	<i>n</i>	<i>MIN</i>	<i>MAX</i>	<i>MEAN</i>	<i>SD</i>	<i>CV</i>	<i>Sk</i>	<i>Ku</i>
วิชาภาษาไทย								
ชาย	703	13	67	42.105	9.968	0.237	-0.186	-0.409
หญิง	797	15	70	46.961	8.602	0.183	-0.263	0.191
รวม	1,500	13	70	44.685	9.576	0.214	-0.306	-0.096
วิชาภาษาอังกฤษ								
ชาย	703	12	92	32.057	12.875	0.401	1.511	2.607
หญิง	797	10	90	34.680	14.153	0.408	1.135	0.961
รวม	1,500	10	92	33.450	13.628	0.407	1.297	1.581
ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างวิชาภาษาไทยและภาษาอังกฤษ							0.439**	

จากตารางที่ 4-1 สามารถอธิบายได้ว่า นักเรียนจำนวน 1,500 คน แบ่งเป็นผู้ชาย 703 คน ผู้หญิง 797 คน ทำการสอบ O-NET วิชาภาษาไทย โดยมีคะแนนเต็ม 100 คะแนน ได้คะแนนต่ำสุดเท่ากับ 13 คะแนน คะแนนสูงสุด 70 คะแนน และคะแนนเฉลี่ยเท่ากับ 44.685 คะแนน ส่วนสัมประสิทธิ์การกระจายของข้อมูลเท่ากับ 0.214 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.306 และค่าความโด่งเท่ากับ -0.096 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และผลการสอบ O-NET วิชาภาษาอังกฤษ มีคะแนนเต็ม 100 คะแนน ได้คะแนนต่ำสุดเท่ากับ 10 คะแนน คะแนนสูงสุด 92 คะแนน และคะแนนเฉลี่ยเท่ากับ 33.450 คะแนน ส่วนสัมประสิทธิ์การกระจายของข้อมูลเท่ากับ 0.407 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ 1.297 และค่าความโด่งเท่ากับ 1.581 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และยังพบว่าความสัมพันธ์ระหว่างคะแนนวิชาภาษาไทยและวิชาภาษาอังกฤษสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีขนาดความสัมพันธ์เท่ากับ 0.439

ตารางที่ 4-2 ความเชื่อมั่นของแบบสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ

วิชา	ความเชื่อมั่น
ภาษาไทย	0.581
ภาษาอังกฤษ	0.780

จากตารางที่ 4-2 สามารถอธิบายได้ว่าความเชื่อมั่นของแบบสอบภาษาไทย มีค่าเท่ากับ 0.581 และความเชื่อมั่นของแบบสอบภาษาอังกฤษ มีค่าเท่ากับ 0.780

ตอนที่ 2 ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถของผู้สอบ

ในตอนี่ 2 นี้ผู้วิจัยแบ่งผลการวิเคราะห์ออกเป็น 6 ตาราง คือ ตารางที่ 1 แสดงผลการวิเคราะห์สถิติพื้นฐานของพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และความสามารถของผู้สอบ ตารางที่ 2 แสดงผลการวิเคราะห์ความแปรปรวนของอิทธิพลทดสอบ ตารางที่ 3 แสดงผลการวิเคราะห์ความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) พารามิเตอร์ความสามารถของผู้สอบ ระหว่างพารามิเตอร์ที่การประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์ และตารางที่ 4 และ 5 แสดงผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ข้อสอบระหว่างการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบ รายละเอียดดังตารางที่ 4-3 ถึงตารางที่ 4-7

ตารางที่ 4-3 สถิติพื้นฐานพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถของผู้สอบ

พารามิเตอร์	n	IRT 3PL				TRT 3PL			
		MIN	MAX	MEAN	SD	MIN	MAX	MEAN	SD
วิชาภาษาไทย									
a	50	0.001	7.007	1.125	2.246	0.002	8.984	1.150	2.386
b	50	-1.469	9.944	7.586	3.577	-3.851	9.949	7.963	3.355
c	50	0.000	0.515	0.060	0.114	0.000	0.516	0.060	0.114
θ	1,500	-2.075	2.011	0.022	0.607	-2.232	2.264	0.018	0.673
วิชาภาษาอังกฤษ									
a	50	0.554	7.161	2.680	1.527	0.620	6.291	2.612	1.350
b	50	0.439	3.035	1.661	0.575	0.476	3.726	1.788	0.697
c	50	0.128	0.357	0.227	0.051	0.122	0.352	0.228	0.048
θ	1,500	-1.257	2.786	-0.002	0.882	-1.217	2.794	0.007	0.832

จากตารางที่ 4-3 สามารถอธิบายได้ว่าในวิชาภาษาไทยการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ พบว่า พารามิเตอร์อำนาจจำแนกมีค่าต่ำสุดเท่ากับ 0.001 ค่าสูงสุดเท่ากับ 7.007 และค่าเฉลี่ยเท่ากับ 1.125 พารามิเตอร์ความยากมีค่าต่ำสุดเท่ากับ -1.469 ค่าสูงสุดเท่ากับ 9.944 และค่าเฉลี่ยเท่ากับ 7.586 พารามิเตอร์การเดามีค่าต่ำสุดเท่ากับ 0.000 ค่าสูงสุดเท่ากับ 0.515 และมีค่าเฉลี่ยเท่ากับ 0.060 และพารามิเตอร์ความสามารถของผู้สอบ มีค่าต่ำสุดเท่ากับ -2.075 ค่าสูงสุดเท่ากับ 2.011 และค่าเฉลี่ยเท่ากับ 0.022 ส่วนการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองเทสต์เลท แบบ 3 พารามิเตอร์ พบว่า พารามิเตอร์อำนาจจำแนกมีค่าต่ำสุดเท่ากับ 0.002 ค่าสูงสุดเท่ากับ 8.984 และค่าเฉลี่ยเท่ากับ 1.150 พารามิเตอร์ความยากมีค่าต่ำสุดเท่ากับ -3.851 ค่าสูงสุดเท่ากับ 9.949 และค่าเฉลี่ยเท่ากับ 7.963 พารามิเตอร์การเดามีค่าต่ำสุดเท่ากับ 0.000 ค่าสูงสุดเท่ากับ 0.516 และมีค่าเฉลี่ยเท่ากับ 0.060 และพารามิเตอร์ความสามารถของผู้สอบ มีค่าต่ำสุดเท่ากับ -2.232 ค่าสูงสุดเท่ากับ 2.264 และค่าเฉลี่ยเท่ากับ 0.018

ในรายวิชาภาษาอังกฤษการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ พบว่า พารามิเตอร์อำนาจจำแนกมีค่าต่ำสุดเท่ากับ 0.554 ค่าสูงสุด

เท่ากับ 7.161 และค่าเฉลี่ยเท่ากับ 2.680 พารามิเตอร์ความยากมีค่าต่ำสุดเท่ากับ 0.439 ค่าสูงสุดเท่ากับ 3.035 และค่าเฉลี่ยเท่ากับ 1.661 พารามิเตอร์การเดามีค่าต่ำสุดเท่ากับ 0.128 ค่าสูงสุดเท่ากับ 0.357 และมีค่าเฉลี่ยเท่ากับ 0.227 และพารามิเตอร์ความสามารถของผู้สอบ มีค่าต่ำสุดเท่ากับ -1.257 ค่าสูงสุดเท่ากับ 2.786 และค่าเฉลี่ยเท่ากับ -0.002 ส่วนการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองเทสต์เลท แบบ 3 พารามิเตอร์ พบว่า พารามิเตอร์อำนาจจำแนกมีค่าต่ำสุดเท่ากับ 0.620 ค่าสูงสุดเท่ากับ 6.291 และค่าเฉลี่ยเท่ากับ 2.612 พารามิเตอร์ความยากมีค่าต่ำสุดเท่ากับ 0.476 ค่าสูงสุดเท่ากับ 3.726 และค่าเฉลี่ยเท่ากับ 1.788 พารามิเตอร์การเดามีค่าต่ำสุดเท่ากับ 0.122 ค่าสูงสุดเท่ากับ 0.352 และมีค่าเฉลี่ยเท่ากับ 0.228 และพารามิเตอร์ความสามารถของผู้สอบ มีค่าต่ำสุดเท่ากับ -1.217 ค่าสูงสุดเท่ากับ 2.794 และค่าเฉลี่ยเท่ากับ 0.007

ตารางที่ 4-4 ค่าความแปรปรวนของอิทธิพลเทสต์เลทจำแนกตามวิชา

เทสต์เลท	วิชาภาษาไทย			วิชาภาษาอังกฤษ		
	จำนวนข้อ	$\sigma_{d(i)}^2$	SE	จำนวนข้อ	$\sigma_{d(i)}^2$	SE
1	10	1.317	0.382	2	0.139	0.041
2	5	0.718	0.344	2	0.094	0.020
3	5	0.453	0.143	5	0.084	0.017
4	5	2.651	1.026	5	0.255	0.054
5	5	0.161	0.035	3	0.729	0.168
6	-	-	-	5	0.132	0.030
7	-	-	-	5	0.105	0.018
8	-	-	-	5	0.178	0.047

จากตารางที่ 4-4 สามารถอธิบายได้ว่า แบบสอบวิชาภาษาไทยมีทั้งหมด 5 เทสต์เลท ประกอบด้วยเทสต์เลทที่ 1 มีจำนวน 10 ข้อ เทสต์เลทที่ 2-5 มีจำนวน 5 ข้อ เมื่อพิจารณา ค่าความแปรปรวนในแต่ละเทสต์เลท พบว่า เทสต์เลทที่มีค่าความแปรปรวนอิทธิพลเทสต์เลท สูงที่สุดคือ เทสต์เลทที่ 4 มีค่าความแปรปรวนเท่ากับ 2.651 รองลงมาคือเทสต์เลทที่ 1 มีค่าความแปรปรวนเท่ากับ 1.317 ส่วนเทสต์เลทที่มีค่าความแปรปรวนอิทธิพลต่ำที่สุดคือ

เทสต์เลขที่ 5 มีค่าความแปรปรวนเท่ากับ 0.161

แบบสอบวิชาภาษาอังกฤษมีทั้งหมด 8 เทสต์เลข โดยเทสต์เลขที่ 1-2 มีจำนวน 2 ข้อ เทสต์เลขที่ 5 มีจำนวน 3 ข้อ และเทสต์เลขที่ 3-4, 6-8 มีจำนวน 5 ข้อ เมื่อพิจารณา ค่าความแปรปรวนในแต่ละเทสต์เลข พบว่า เทสต์เลขที่มีค่าความแปรปรวนของอิทธิพลเทสต์เลข สูงที่สุดคือ เทสต์เลขที่ 5 มีค่าความแปรปรวนเท่ากับ 0.729 รองลงมาคือเทสต์เลขที่ 8 มีค่าความแปรปรวนเท่ากับ 0.178 และเทสต์เลขที่มีค่าความแปรปรวนต่ำที่สุดคือเทสต์เลขที่ 3 มีค่าความแปรปรวนเท่ากับ 0.084

ตารางที่ 4-5 ค่าสัมประสิทธิ์สหสัมพันธ์พารามิเตอร์ระหว่างการประมาณค่าตามโมเดลทฤษฎี การตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนองเทสต์เลข แบบ 3 พารามิเตอร์

พารามิเตอร์	วิชา	IRT 3PL			θ
		a	b	c	
TRT 3PL	ภาษาไทย	0.920**	0.881**	1.000**	0.031
	ภาษาอังกฤษ	0.935**	0.983**	0.972**	0.994**

จากตารางที่ 4-5 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองเทสต์เลขทั้งในวิชาภาษาไทยและวิชาภาษาอังกฤษ มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีขนาดความสัมพันธ์ระหว่าง 0.881 ถึง 0.994 ยกเว้นพารามิเตอร์ความสามารถของผู้สอบในรายวิชาภาษาไทยที่มีความสัมพันธ์กันอย่างไม่เป็นนัยสำคัญ มีขนาดความสัมพันธ์เท่ากับ 0.031

ตารางที่ 4-6 ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ระหว่าง
การประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และทฤษฎี
การตอบสนองเทสต์เลทแบบ 3 พารามิเตอร์ ในรายวิชาภาษาไทย

พารามิเตอร์		<i>n</i>	<i>MEAN</i>	<i>SD</i>	<i>t</i>	<i>Sig</i>
อำนาจจำแนก	IRT 3PL	50	2.969	6.066	0.373	0.711
	TRT 3PL	50	3.056	6.325		
ความยาก	IRT 3PL	50	1.348	2.055	3.039*	0.004
	TRT 3PL	50	1.069	1.673		
การเดา	IRT 3PL	50	0.009	0.011	0.741	0.462
	TRT 3PL	50	0.010	0.012		
ความสามารถของผู้สอบ	IRT 3PL	1500	5.736	0.283	1.201	0.230
	TRT 3PL	1500	5.724	0.300		

จากตารางที่ 4-6 พบว่าในรายวิชาภาษาไทยความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และทฤษฎีการตอบสนองเทสต์เลทแบบ 3 พารามิเตอร์ ไม่แตกต่างกันอย่างมีนัยสำคัญ ส่วนพารามิเตอร์ความยากมีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01

ตารางที่ 4-7 ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ระหว่าง
การประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และทฤษฎี
การตอบสนองเทสต์เลทแบบ 3 พารามิเตอร์ ในรายวิชาภาษาอังกฤษ

พารามิเตอร์		<i>n</i>	<i>MEAN</i>	<i>SD</i>	<i>t</i>	<i>Sig</i>
อำนาจจำแนก	IRT 3PL	50	0.729	0.787	2.471*	0.017
	TRT 3PL	50	0.850	0.759		
ความยาก	IRT 3PL	50	0.138	0.066	3.422*	0.001
	TRT 3PL	50	0.151	0.080		

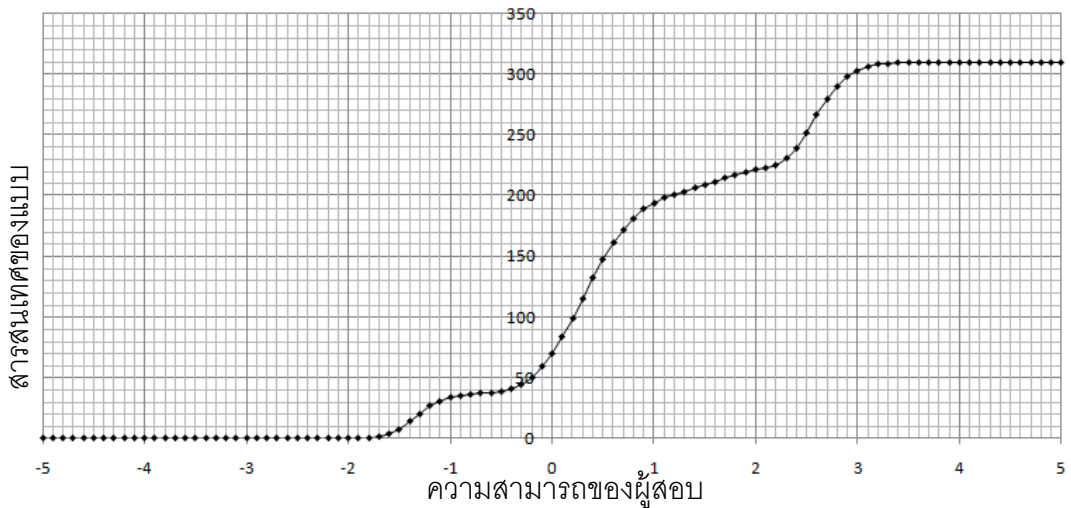
ตารางที่ 4-7 (ต่อ)

พารามิเตอร์		<i>n</i>	<i>MEAN</i>	<i>SD</i>	<i>t</i>	<i>Sig</i>
การเดา	IRT 3PL	50	0.021	0.008	2.315*	0.025
	TRT 3PL	50	0.023	0.011		
ความสามารถของผู้สอบ	IRT 3PL	1500	0.542	0.186	5.644*	0.000
	TRT 3PL	1500	0.552	0.180		

จากตารางที่ 4-7 พบว่าในรายวิชาภาษาอังกฤษความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก และพารามิเตอร์ความเดา ที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และทฤษฎีการตอบสนองเทสต์เลทแบบ 3 พารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ส่วนพารามิเตอร์ความยากและพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์และทฤษฎีการตอบสนองเทสต์เลทแบบ 3 พารามิเตอร์ มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.01

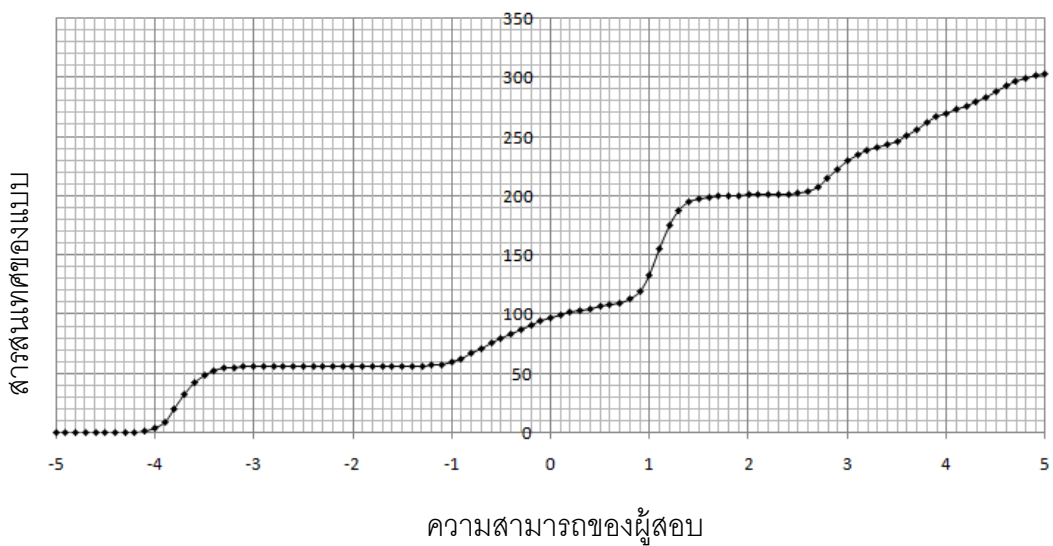
นอกจากนี้ผู้วิจัยได้คำนวณสารสนเทศของแบบสอบ พบว่า แบบสอบภาษาไทยที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบมีความถูกต้องแม่นยำสูงเมื่อผู้สอบมีความสามารถสูงเช่นเดียวกับการประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลทแบบสอบภาษาไทยมีความแม่นยำเมื่อผู้สอบมีความสามารถสูง และแบบสอบภาษาอังกฤษที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบมีความถูกต้องแม่นยำสูงเมื่อผู้สอบมีความสามารถเท่ากับ 2.00 ส่วนการประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลทแบบสอบภาษาไทยมีความแม่นยำเมื่อผู้สอบมีความสามารถตั้งแต่ 1.40 รายละเอียดดังภาพที่ 4-1 ถึงภาพที่ 4-4

กราฟแสดงความสัมพันธ์ระหว่างสารสนเทศของแบบสอบและความสามารถผู้สอบ
วิชาภาษาไทย ตามโมเดลทฤษฎีการตอบสนองข้อสอบ



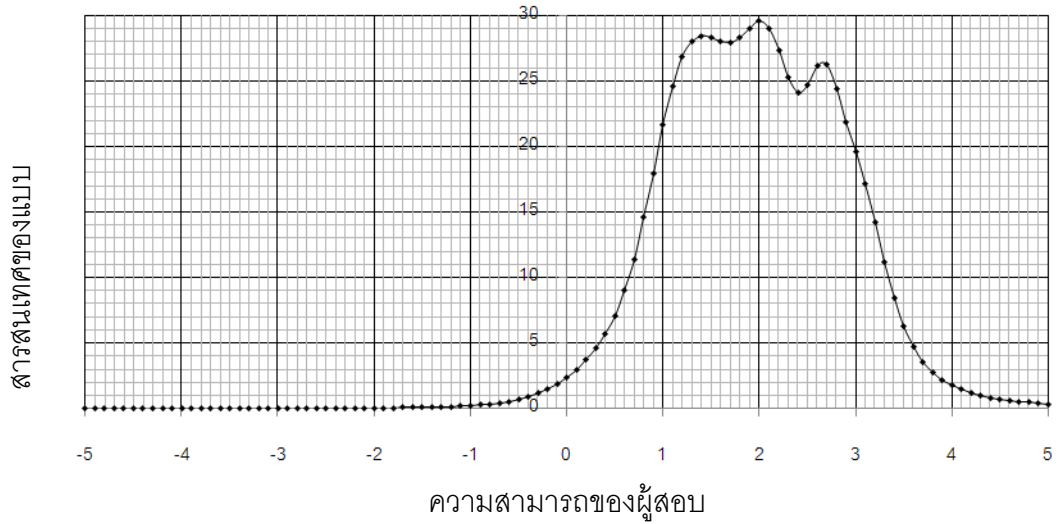
ภาพที่ 4-1 สารสนเทศของแบบสอบภาษาไทยตามโมเดลทฤษฎีการตอบสนองข้อสอบ

กราฟแสดงความสัมพันธ์ระหว่างสารสนเทศของแบบสอบและความสามารถผู้สอบ
วิชาภาษาไทย ตามโมเดลทฤษฎีการตอบสนองของเทสต์เลข



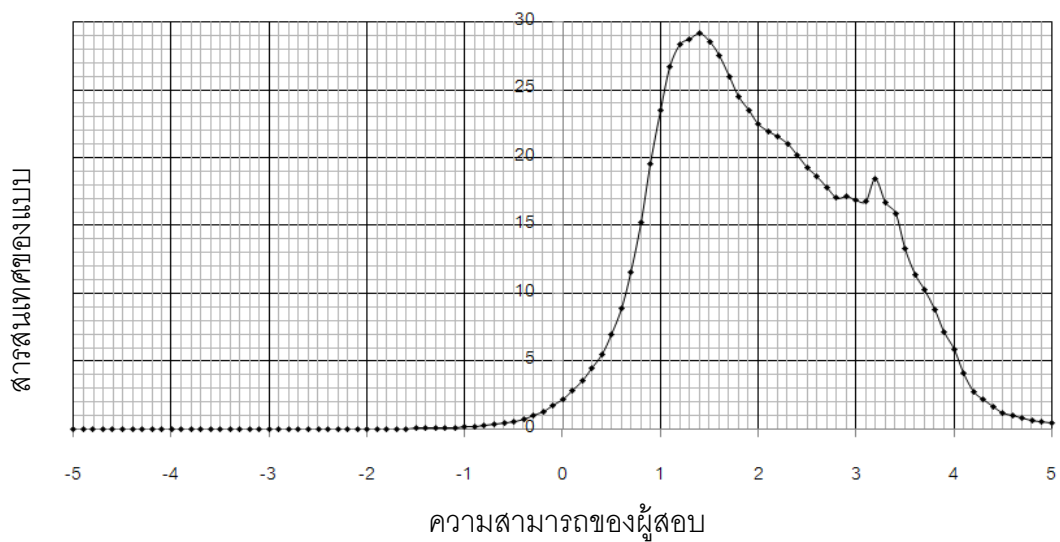
ภาพที่ 4-2 สารสนเทศของแบบสอบภาษาไทยตามโมเดลทฤษฎีการตอบสนองของเทสต์เลข

กราฟแสดงความสัมพันธ์ระหว่างสารสนเทศของแบบสอบและความสามารถผู้สอบ
วิชาภาษาอังกฤษ ตามโมเดลทฤษฎีการตอบสนองข้อสอบ



ภาพที่ 4-3 สารสนเทศของแบบสอบภาษาอังกฤษตามโมเดลทฤษฎีการตอบสนองข้อสอบ

กราฟแสดงความสัมพันธ์ระหว่างสารสนเทศของแบบสอบและความสามารถผู้สอบ
วิชาภาษาอังกฤษ ตามโมเดลทฤษฎีการตอบสนองเทสต์เลท



ภาพที่ 4-4 สารสนเทศของแบบสอบภาษาอังกฤษตามโมเดลทฤษฎีการตอบสนองเทสต์เลท

ตอนที่ 3 ประสิทธิภาพของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบทดสอบ

ในส่วนของตอนที่ 3 ผู้วิจัยแบ่งการนำเสนอผลการวิจัยออกเป็น 3 ส่วน คือ ส่วนที่ 1 ผลการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ ส่วนที่ 2 ผลการเชื่อมโยงคะแนนด้วยวิธีอควิปเปอร์เซ็นต์ไทล์ และส่วนที่ 3 ประสิทธิภาพของการเชื่อมโยงคะแนน รายละเอียดดังนี้

ส่วนที่ 1 ผลการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ

จากการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองทดสอบทดสอบในตอนที่ 2 ผู้วิจัยนำค่าที่ได้ของแต่ละคนมาคำนวณความน่าจะเป็นในการตอบถูกของแต่ละข้อ แล้วนำมารวมกันจะได้คะแนนจริง (True score) จากนั้นนำคะแนนจริงที่ได้มาเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ ตามหลักการของ Stocking and Lord (1983) ได้ผลการวิเคราะห์ดังตารางที่ 4-8 (รายละเอียดคะแนนรายบุคคลนำเสนอในภาคผนวก ง)

ตารางที่ 4-8 คะแนนจริงภาษาไทย คะแนนจริงภาษาอังกฤษ คะแนนเชื่อมโยงด้วยวิธีโค้งคุณลักษณะแบบสอบ

รายการ	MIN	MAX	MEAN	SE	Sk	Ku	M_SL
ทฤษฎีการตอบสนองข้อสอบ (IRT 3PL)							
คะแนนไทย	18.318	27.311	22.276	0.055	0.316	-1.130	-
คะแนนอังกฤษ	11.963	44.363	16.086	0.154	1.988	0.339	-
คะแนนเชื่อมโยง	19.076	22.878	20.648	0.014	0.339	0.232	5.203
ทฤษฎีการตอบสนองทดสอบทดสอบ (TRT 3PL)							
คะแนนไทย	19.287	24.431	21.645	0.027	0.130	-0.043	-
คะแนนอังกฤษ	11.887	42.692	15.831	0.153	2.020	3.382	-
คะแนนเชื่อมโยง	14.077	16.392	14.806	0.008	0.729	0.920	47.309

จากตารางที่ 4-8 สามารถอธิบายได้ว่าคะแนนจริงที่ได้จากการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ในวิชาภาษาไทย พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 18.318 คะแนนจริงสูงสุดเท่ากับ 27.311 คะแนนจริงเฉลี่ยเท่ากับ 22.276 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิจารณาจากค่าความเบ้เท่ากับ 0.316 และค่าความโด่งเท่ากับ -1.130 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) ส่วนในวิชาภาษาอังกฤษ พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 11.963 คะแนนจริงสูงสุดเท่ากับ

44.363 คะแนนจริงเฉลี่ยเท่ากับ 16.086 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติ พิจารณาจากค่าความเบ้เท่ากับ 1.988 และค่าความโด่งเท่ากับ 0.339 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และคะแนนจริงที่เชื่อมโยงคะแนนภาษาอังกฤษ เท่าเทียมกับคะแนนภาษาไทย พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 19.076 คะแนนจริงสูงสุดเท่ากับ 22.878 คะแนนจริงเฉลี่ยเท่ากับ 20.648 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติ พิจารณาจากค่าความเบ้เท่ากับ 0.339 และค่าความโด่งเท่ากับ 0.232 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50)

การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองของเทสต์เลทแบบ 3 พารามิเตอร์ ในวิชาภาษาไทย พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 19.287 คะแนนจริงสูงสุดเท่ากับ 24.431 คะแนนจริงเฉลี่ยเท่ากับ 21.645 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติ พิจารณาจากค่าความเบ้เท่ากับ 0.130 และค่าความโด่งเท่ากับ -0.043 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) ส่วนในวิชาภาษาอังกฤษ พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 11.887 คะแนนจริงสูงสุดเท่ากับ 42.692 คะแนนจริงเฉลี่ยเท่ากับ 15.831 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิจารณาจากค่าความเบ้เท่ากับ 2.020 และค่าความโด่งเท่ากับ 3.382 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และคะแนนจริงที่เชื่อมโยงคะแนนภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย พบว่า มีคะแนนจริงต่ำสุดเท่ากับ 14.077 คะแนนจริงสูงสุดเท่ากับ 16.392 คะแนนจริงเฉลี่ยเท่ากับ 14.806 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิจารณาจากค่าความเบ้เท่ากับ 0.729 และค่าความโด่งเท่ากับ 0.920 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50)

จากตารางยังพบว่า ค่าเฉลี่ยผลต่างกำลังสองของคะแนนเชื่อมโยงตามหลักการ Stocking and Lord ที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ 5.203 ส่วนการประมาณค่าด้วยทฤษฎีการตอบสนองเทสต์เลทมีค่าเท่ากับ 47.309 ซึ่งค่า M_{SL} นี้ ค่าที่น้อยจะประสิทธิภาพในการเชื่อมโยงคะแนนมากกว่า ดังนั้นการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบจะมีคุณภาพมากกว่าการประมาณค่าด้วยทฤษฎีการตอบสนองเทสต์เลท

ส่วนที่ 2 ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์

จากการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองเทสต์เลทในตอนที่ 2 ผู้วิจัยนำค่าที่ได้ของแต่ละคนมาคำนวณความน่าจะเป็นในการตอบถูกของแต่ละข้อ แล้วนำแปลงเป็นคะแนนสังเกตได้ (Observed score) จากนั้นนำ

คะแนนที่ได้มาเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซนส์ไทล์ ได้ผลการวิเคราะห์ดังตารางที่ 4-9 (รายละเอียดคะแนนรายบุคคลนำเสนอในภาคผนวก จ)

ตารางที่ 4-9 คะแนนสังเกตได้ภาษาไทย คะแนนสังเกตได้ภาษาอังกฤษ คะแนนเชื่อมโยงด้วยวิธีอิกวิเปอร์เซนส์ไทล์

รายการ	MIN	MAX	MEAN	SE	Sk	Ku
ทฤษฎีการตอบสนองข้อสอบ (IRT 3PL)						
คะแนนไทย	13.000	39.000	27.707	3.833	-0.136	2.957
คะแนนอังกฤษ	18.000	41.000	29.359	3.470	-0.032	2.918
คะแนนเชื่อมโยง	16.500	34.345	27.706	3.821	-0.130	2.950
ทฤษฎีการตอบสนองทดสอบ (TRT 3PL)						
คะแนนไทย	17.000	39.000	28.353	3.272	-0.029	2.934
คะแนนอังกฤษ	24.000	45.000	35.212	3.212	-0.130	2.940
คะแนนเชื่อมโยง	17.000	39.000	28.349	3.260	-0.034	2.910

จากตารางที่ 4-9 สามารถอธิบายได้ว่าคะแนนสังเกตได้จากการประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ในวิชาภาษาไทย พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 13.000 คะแนนสังเกตได้สูงสุดเท่ากับ 39.000 คะแนนสังเกตได้เฉลี่ยเท่ากับ 27.707 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.136 และค่าความโด่งเท่ากับ 2.957 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) ส่วนในวิชาภาษาอังกฤษ พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 18.000 คะแนนสังเกตได้สูงสุดเท่ากับ 41.000 คะแนนสังเกตได้เฉลี่ยเท่ากับ 29.359 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.032 และค่าความโด่งเท่ากับ 2.918 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และคะแนนสังเกตได้ที่เชื่อมโยงคะแนนภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 16.500 คะแนนสังเกตได้สูงสุดเท่ากับ 34.345 คะแนนสังเกตได้เฉลี่ยเท่ากับ 27.706 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.130 และค่าความโด่งเท่ากับ 2.950 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่

ระหว่าง -2.00 ถึง 3.50)

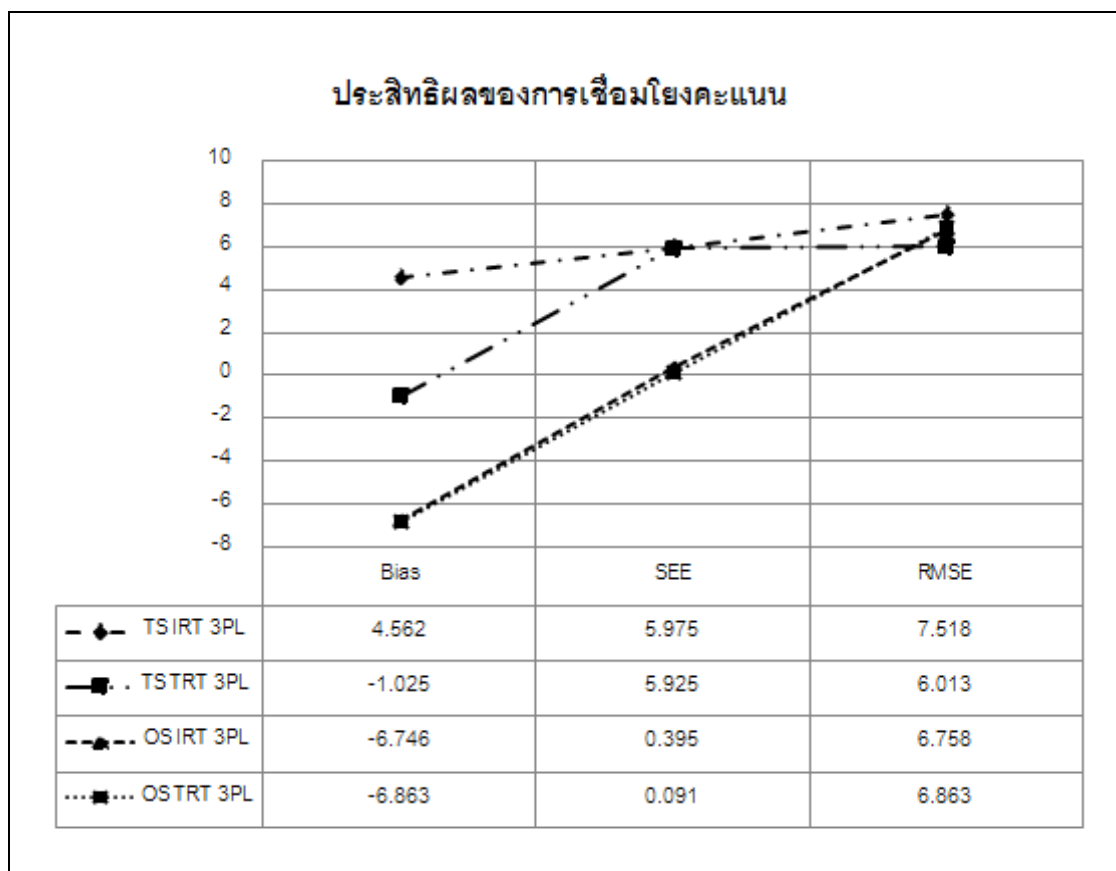
การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองของทดสอบแบบ 3 พารามิเตอร์ ในวิชาภาษาไทย พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 17.000 คะแนนสังเกตได้สูงสุดเท่ากับ 39.000 คะแนนสังเกตได้เฉลี่ยเท่ากับ 28.353 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.029 และค่าความโด่งเท่ากับ 2.934 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) ส่วนในวิชาภาษาอังกฤษ พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 24.000 คะแนนสังเกตได้สูงสุดเท่ากับ 45.000 คะแนนสังเกตได้เฉลี่ยเท่ากับ 35.212 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.130 และค่าความโด่งเท่ากับ 2.940 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50) และคะแนนสังเกตได้ที่เชื่อมโยงคะแนนภาษาอังกฤษเท่ากับคะแนนภาษาไทย พบว่า มีคะแนนสังเกตได้ต่ำสุดเท่ากับ 17.000 คะแนนสังเกตได้สูงสุดเท่ากับ 39.000 คะแนนสังเกตได้เฉลี่ยเท่ากับ 28.349 และการแจกแจงของข้อมูลมีการแจกแจงเป็นปกติพิชารณาจากค่าความเบ้เท่ากับ -0.034 และค่าความโด่งเท่ากับ 2.910 (Lei & Lomax, 2005; ค่าความเบ้และความโด่งอยู่ระหว่าง -2.00 ถึง 3.50)

ส่วนที่ 3 ประสิทธิภาพของการเชื่อมโยงคะแนน

จากผลการเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ และคะแนนสังเกตได้ด้วยวิธีอิกคิวเปอร์เซ็นต์ไทล์ของข้อสอบที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ และการตอบสนองของทดสอบ ผู้วิจัยนำผลที่ได้มาหาประสิทธิภาพโดยพิจารณาจากค่าความลำเอียง (Bias) ค่าความคลาดเคลื่อนมาตรฐานของการเชื่อมโยง (SEE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (RMSE) ได้ผลการคำนวณดังตารางที่ 4-10 และ ภาพที่ 4-5

ตารางที่ 4-10 ประสิทธิภาพของการเชื่อมโยงคะแนน

ประสิทธิภาพ	TS		OS	
	IRT 3PL	TRT 3PL	IRT 3PL	TRT 3PL
Bias	4.562	-1.025	-6.746	-6.863
SEE	5.975	5.925	0.395	0.091
RMSE	7.518	6.013	6.758	6.863



ภาพที่ 4-5 เปรียบเทียบประสิทธิผลของการเชื่อมโยงคะแนน

จากตารางที่ 4-10 และภาพที่ 4-5 สามารถอธิบายได้ว่าประสิทธิผลของการเชื่อมโยงคะแนนที่พิจารณาจาก 3 ค่า คือ ค่าความลำเอียง (Bias) ค่าความคลาดเคลื่อนมาตรฐานของการเชื่อมโยง (SEE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (RMSE) ผลการวิเคราะห์พบว่า

ค่าความลำเอียง (bias) ของการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของเทสต์เลทจะมีค่าต่ำที่สุด มีค่าเท่ากับ -6.863 ส่วนค่าสูงที่สุดคือ ค่าที่ได้จากการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ 4.562 รองลงมาคือการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของเทสต์เลทมีค่าเท่ากับ -1.025 และการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ -6.746 ตามลำดับ

ค่าความคลาดเคลื่อนมาตรฐานของการเชื่อมโยง (SEE) ของการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของเทสต์เลทจะมีค่าต่ำที่สุด มีค่าเท่ากับ

0.091 ส่วนค่าสูงที่สุดคือ ค่าที่ได้จากการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วยโมเดลทฤษฎี การตอบสนองข้อสอบมีค่าเท่ากับ 5.975 รองลงมาคือการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วย โมเดลทฤษฎีการตอบสนองเทสต์เลขมีค่าเท่ากับ 5.925 และการเชื่อมโยงคะแนนสังเกตได้ที่ ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ 0.395 ตามลำดับ

ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (RMSE) ของการเชื่อมโยง คะแนนจริงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองเทสต์เลขจะมีค่าต่ำที่สุด มีค่าเท่ากับ 6.013 ส่วนค่าสูงที่สุดคือ ค่าที่ได้จากการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วยโมเดลทฤษฎี การตอบสนองข้อสอบมีค่าเท่ากับ 7.518 รองลงมาคือการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่า ด้วยโมเดลทฤษฎีการตอบสนองเทสต์เลขมีค่าเท่ากับ 6.863 และการเชื่อมโยงคะแนนจริงที่ ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ 6.758 ตามลำดับ

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์ 3 ประการ คือ 1) เพื่อศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบระหว่างการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองทดสอบ 2) เพื่อเปรียบเทียบความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองทดสอบ และ 3) เพื่อเปรียบเทียบประสิทธิผลของการเชื่อมโยงคะแนนสำหรับแบบสอบทดสอบตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีการเชื่อมโยงคะแนนจริง และวิธีการเชื่อมโยงคะแนนสังเกตได้ภายใต้โมเดลการประมาณค่าพารามิเตอร์ต่างกัน

การดำเนินการวิจัยนั้น ผู้วิจัยใช้ข้อมูลทฤษฎี ซึ่งเป็นข้อมูลผลการสอบรายข้อจากแบบสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาไทย และวิชาภาษาอังกฤษของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558 จำนวน 1,500 คน ส่วนการวิเคราะห์ข้อมูลแบ่งออกเป็น 2 ส่วน คือ ส่วนที่ 1 เป็นการประมาณค่าพารามิเตอร์และหาความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ เริ่มจากประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) และพารามิเตอร์ความสามารถผู้สอบ ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนองทดสอบแบบ 3 พารามิเตอร์ แล้วนำค่าพารามิเตอร์มาหาความสัมพันธ์ด้วยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ด้วยสถิติ Pair t-test และส่วนที่ 2 เป็นการเชื่อมโยงคะแนนและหาประสิทธิผลของการเชื่อมโยงคะแนน ซึ่งการเชื่อมโยงคะแนนนั้นผู้วิจัยเลือกใช้วิธีการเชื่อมโยงคะแนนระหว่างแบบสอบจำนวน 2 วิธี คือ การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ และการเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ส่วนประเด็นการตรวจสอบประสิทธิผลของการเชื่อมโยงคะแนนนั้น ผู้วิจัยพิจารณาจากค่าความลำเอียง (Bias) ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) และค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root mean square error)

สรุปผลการวิจัย

การวิจัยครั้งนี้เก็บข้อมูลจากนักเรียน 1,500 คน เป็นผู้ชายจำนวน 703 คน และผู้หญิงจำนวน 797 คน ที่เข้ารับการทดสอบ O-NET วิชาภาษาไทย และวิชาภาษาอังกฤษ โดยแต่ละวิชามีคะแนนเต็มเท่ากับ 100 คะแนน นักเรียนได้คะแนนสอบวิชาภาษาไทยอยู่ระหว่าง 13-70 คะแนน มีคะแนนเฉลี่ยเท่ากับ 44.685 คะแนน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 9.576 และได้คะแนนสอบวิชาภาษาอังกฤษมีคะแนนอยู่ระหว่าง 10-92 คะแนน มีคะแนนเฉลี่ยเท่ากับ 33.450 คน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 13.628 จากนั้นวิเคราะห์ข้อมูลและสรุปผลการวิเคราะห์ข้อมูลได้ดังนี้

1. การประมาณค่าพารามิเตอร์

1.1 การประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ ในวิชาภาษาไทยพบว่า พารามิเตอร์อำนาจจำแนกมีค่าอยู่ระหว่าง 0.001 ถึง 7.007 พารามิเตอร์ความยากมีค่าอยู่ระหว่าง -1.469 ถึง 9.977 พารามิเตอร์การเดามีค่าอยู่ระหว่าง 0.000 ถึง 0.515 และพารามิเตอร์ความสามารถของผู้สอบมีค่าอยู่ระหว่าง -2.075 ถึง 2.011 ส่วนในวิชาภาษาอังกฤษพบว่า พารามิเตอร์อำนาจจำแนกมีค่าอยู่ระหว่าง 0.554 ถึง 7.161 พารามิเตอร์ความยากมีค่าอยู่ระหว่าง 0.439 ถึง 3.035 พารามิเตอร์การเดามีค่าอยู่ระหว่าง 0.128 ถึง 0.357 และพารามิเตอร์ความสามารถของผู้สอบมีค่าอยู่ระหว่าง -1.257 ถึง 2.786 และผลการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลท ในวิชาภาษาไทยพบว่า พารามิเตอร์อำนาจจำแนกมีค่าอยู่ระหว่าง 0.002 ถึง 8.984 พารามิเตอร์ความยากมีค่าอยู่ระหว่าง -3.851 ถึง 9.949 พารามิเตอร์การเดามีค่าอยู่ระหว่าง 0.000 ถึง 0.516 พารามิเตอร์ความสามารถของผู้สอบมีค่าอยู่ระหว่าง -2.232 ถึง 2.264 และความแปรปรวนของอิทธิพลเทสต์เลทมีค่าอยู่ระหว่าง 0.161 ถึง 2.651 ส่วนในวิชาภาษาอังกฤษพบว่า พารามิเตอร์อำนาจจำแนกมีค่าอยู่ระหว่าง 0.620 ถึง 6.291 พารามิเตอร์ความยากมีค่าอยู่ระหว่าง 0.439 ถึง 3.035 พารามิเตอร์การเดามีค่าอยู่ระหว่าง 0.128 ถึง 0.357 พารามิเตอร์ความสามารถของผู้สอบมีค่าอยู่ระหว่าง -1.217 ถึง 2.794 และความแปรปรวนของอิทธิพลเทสต์เลทมีค่าอยู่ระหว่าง 0.084 ถึง 0.729 นอกจากนี้สารสนเทศของแบบสอบวิชาภาษาไทยที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองเทสต์เลทมีแนวโน้มสูงขึ้น เมื่อความสามารถเพิ่มขึ้น ส่วนวิชาภาษาอังกฤษที่ประมาณค่าด้วยทฤษฎีการตอบสนองข้อสอบเหมาะสมกับผู้สอบที่มีความสามารถเท่ากับ 2.00 ส่วนการประมาณค่าตามโมเดลทฤษฎีการตอบสนองเทสต์เลทเหมาะสมกับผู้สอบที่มีความสามารถ 1.40

1.2 ความสัมพันธ์ของค่าพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถผู้สอบระหว่างการประมาณค่าตามทฤษฎี การตอบสนองข้อสอบและตามทฤษฎีการตอบสนองเทสต์เลท พบว่าในวิชาภาษาไทย พารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และพารามิเตอร์การเดามีความสัมพันธ์กันอย่างมีนัยสำคัญ ที่ระดับ 0.01 โดยมีขนาดความสัมพันธ์เท่ากับ 0.920 0.881 และ 1.000 ตามลำดับ แต่พารามิเตอร์ ความสามารถผู้สอบมีความสัมพันธ์อย่างไม่มีนัยสำคัญ โดยมีขนาดความสัมพันธ์เท่ากับ 0.031 ส่วนในวิชาภาษาอังกฤษ พารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถผู้สอบมีความสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ 0.01 โดยมีขนาด ความสัมพันธ์เท่ากับ 0.935 0.983 0.972 และ 0.994 ตามลำดับ

1.3 ความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามทฤษฎีการตอบสนอง ข้อสอบและตามทฤษฎีการตอบสนองเทสต์เลทแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ในวิชา ภาษาอังกฤษ ส่วนวิชาภาษาไทยพบว่ามีเพียงพารามิเตอร์ความยากที่มีค่าความคลาดเคลื่อน มาตรฐานแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05 ส่วนพารามิเตอร์อำนาจจำแนก พารามิเตอร์ การเดา และพารามิเตอร์ความสามารถผู้สอบความคลาดเคลื่อนมาตรฐานแตกต่างกันอย่างไม่มี นัยสำคัญทางสถิติ

2. การเชื่อมโยงคะแนน

2.1 การเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะของแบบสอบ จากการประมาณ ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนเชื่อมโยงภาษาอังกฤษเท่าเทียมกับ คะแนนภาษาไทย มีคะแนนจริงอยู่ระหว่าง 19.076 ถึง 22.878 จากเดิมที่มีคะแนนอยู่ระหว่าง 11.963 ถึง 44.363 ส่วนการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองเทสต์เลท พบว่า คะแนนเชื่อมโยงภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย มีคะแนนจริงอยู่ระหว่าง 14.077 ถึง 16.392 จากเดิมที่มีคะแนนอยู่ระหว่าง 11.887 ถึง 42.692

2.2 การเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ จากการประมาณ ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนเชื่อมโยงภาษาอังกฤษเท่าเทียมกับ คะแนนภาษาไทย มีคะแนนสังเกตได้อยู่ระหว่าง 18.000 ถึง 41.000 จากเดิมที่มีคะแนนอยู่ระหว่าง 18.000 ถึง 41.000 ส่วนการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองเทสต์เลท พบว่า คะแนนเชื่อมโยงภาษาอังกฤษเท่าเทียมกับคะแนนภาษาไทย มีคะแนนสังเกตได้อยู่ระหว่าง 24.000 ถึง 45.000 จากเดิมที่มีคะแนนอยู่ระหว่าง 24.000 ถึง 45.000

2.3 ประสิทธิภาพของการเชื่อมโยงคะแนนเมื่อพิจารณาจาก ค่าความลำเอียง พบว่า วิธีการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองของเทสต์เลทมีค่าน้อยที่สุด พิจารณาจาก ค่าความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบ พบว่า วิธีการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองของเทสต์เลทมีค่าน้อยที่สุด และ พิจารณาจากค่า ค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน วิธีการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองข้อสอบมีค่าน้อยที่สุด

อภิปรายผลการวิจัย

ในส่วนของผลการอภิปรายผลการวิจัย ผู้วิจัยนำเสนอผลการอภิปรายตามวัตถุประสงค์ของการวิจัยมีรายละเอียดดังนี้

1. การศึกษาความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบที่ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดลทฤษฎีการตอบสนอง เทสต์เลทแบบ 3 พารามิเตอร์

ความสัมพันธ์ของค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก ความยาก การเดา) วิชาภาษาไทยและวิชาภาษาอังกฤษ และค่าพารามิเตอร์ความสามารถผู้สอบวิชาภาษาอังกฤษระหว่างวิธีการประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนอง เทสต์เลทมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และขนาดของความสัมพันธ์ก็อยู่ในระดับมาก แสดงว่าแบบสอบมีความถูกต้องแม่นยำในการวัดเป็นไปตามหลักการที่ว่า เครื่องมือที่วัดคุณลักษณะเดียวกันใช้วิธีการวัดต่างกันจะต้องมีความสัมพันธ์กัน (Campbell & Fiske, 1959) อภิปรายผลการวิจัยในทำนองเดียวกันกับหลักการดังกล่าวได้ว่าการประมาณค่าพารามิเตอร์จากแบบสอบเดียวกันแต่ใช้วิธีการประมาณค่าพารามิเตอร์ต่างกันค่าพารามิเตอร์ที่ได้จึงมีความสัมพันธ์กัน และผลการวิจัยได้สอดคล้องกับการศึกษาของ DeMars (2006) ที่ศึกษาความสัมพันธ์ของค่าพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าพารามิเตอร์ระหว่างวิธีโมเดลการตอบสนองของเทสต์เลทและโมเดลการตอบสนองข้อสอบมีความสัมพันธ์กันอย่างมีนัยสำคัญ และมีขนาดความสัมพันธ์ที่สูงใกล้เคียงหนึ่ง นอกจากนี้ยังสอดคล้องกับการศึกษาของ Eckes (2014), Min and He (2014) และ Kogar and Kelecioğlu (2017) ที่ศึกษาความสัมพันธ์ของพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถผู้สอบระหว่างวิธีการประมาณค่าพารามิเตอร์ต่างกันผลการศึกษาพบว่าค่าพารามิเตอร์มีความสัมพันธ์กันอย่างมีนัยสำคัญและมีขนาดความสัมพันธ์ในระดับสูงมาก แต่การศึกษาครั้งนี้ยังพบว่าความสัมพันธ์ของพารามิเตอร์ความสามารถผู้สอบวิชาภาษาไทยระหว่างวิธีการประมาณค่าตามโมเดลทฤษฎีการตอบสนอง

ข้อสอบและโมเดลทฤษฎีการตอบสนองของเทสต์เลทมีความสัมพันธ์กันอย่างไรไม่มีนัยสำคัญทางสถิติ และขนาดของความสัมพันธ์อยู่ในระดับต่ำ ทั้งนี้อาจเป็นผลมาจากความไม่เป็นอิสระกันของ ข้อสอบ ซึ่งในวิชาภาษาไทยนั้นขนาดของอิทธิพลของเทสต์เลทสูงจึงทำให้ประมาณค่าพารามิเตอร์ ความสามารถผู้สอบตามทฤษฎีการตอบสนองของเทสต์เลทไม่สัมพันธ์กับการประมาณค่าตามทฤษฎี การตอบสนองข้อสอบ สอดคล้องกับหลักการของ Yen (1993) กล่าวว่า ความเป็นอิสระกันของ ข้อสอบเป็นปัจจัยที่สำคัญและเกิดขึ้นบ่อยเกี่ยวกับความสามารถของนักเรียนบางคนในการตอบ ข้อสอบถูกมากกว่าคนอื่น และ Chen (2014) ได้สรุปการศึกษาของนักวิจัยท่านอื่น ๆ ไว้ว่า แบบสอบที่มีลักษณะเทสต์เลทเป็นการละเมิดความเป็นอิสระกันของข้อสอบ จึงทำให้การตอบ ข้อสอบมีอิทธิพลมาจากเทสต์เลทด้วย ส่งผลให้ความสัมพันธ์ของการตอบข้อสอบภายในเทสต์เลท ของนักเรียนที่มีระดับความสนใจหรือความรู้พื้นฐานเดิมเกี่ยวกับสิ่งเร้าต่างจากคนอื่นหรือค่า ผิดปกติ

2. การเปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ข้อสอบ และ พารามิเตอร์ความสามารถผู้สอบที่ประมาณค่าตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดล ทฤษฎีการตอบสนองของเทสต์เลท

ความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถของผู้สอบวิชาภาษาอังกฤษที่ประมาณค่า ตามโมเดลทฤษฎีการตอบสนองข้อสอบและโมเดลทฤษฎีการตอบสนองของเทสต์เลทแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และยังพบว่าในทุกพารามิเตอร์ความคลาดเคลื่อน มาตรฐานที่ประมาณค่าตามโมเดลการตอบสนองของเทสต์เลทสูงกว่าโมเดลการตอบสนองข้อสอบ ซึ่งความคลาดเคลื่อนมาตรฐานเป็นค่าที่บ่งบอกการประมาณค่าข้อมูลมีความคลาดเคลื่อนไปจาก ค่าเฉลี่ยของกลุ่มมากน้อยเพียงใด ค่าที่ได้ควรจะมีค่าน้อย ๆ ดังนั้นการประมาณค่าพารามิเตอร์ ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบวิชาภาษาอังกฤษด้วยโมเดลทฤษฎีการตอบสนอง เทสต์เลทมีความคลาดเคลื่อนสูงกว่าการประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ สรุป ได้ว่าความไม่เป็นอิสระในการตอบข้อสอบหรืออิทธิพลของเทสต์เลทมีผลต่อการประมาณค่า พารามิเตอร์ สอดคล้องกับงานวิจัยของ Kogar and Kelecioğlu (2017) ที่ศึกษาความคลาดเคลื่อน มาตรฐานของพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของผู้สอบที่ประมาณค่าด้วย ทฤษฎีการตอบสนองของเทสต์เลทและประมาณค่าด้วยทฤษฎีการตอบสนองข้อสอบภายใต้เงื่อนไข ของขนาดของกลุ่มตัวอย่าง และจำนวนข้อที่เป็นอิสระต่างกัน พบว่า ในทุกเงื่อนไขความคลาดเคลื่อน มาตรฐานของพารามิเตอร์ที่ประมาณค่าด้วยทฤษฎีการตอบสนองของเทสต์เลทสูงกว่าทฤษฎี การตอบสนองข้อสอบ และยังสอดคล้องกับงานวิจัยของ Eckes (2014) ที่พบว่าความคลาดเคลื่อน

ของพารามิเตอร์ที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ ที่ละเมิดข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อ มีค่าต่ำกว่าความเป็นจริง นั่นก็คือค่าความคลาดเคลื่อนที่ได้มีค่าต่ำกว่าค่าความคลาดเคลื่อนที่ประมาณค่าโดยคำนึงถึงความไม่เป็นอิสระของข้อสอบ แต่ในวิชาภาษาไทย พบว่ามีเฉพาะความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ความยากเท่านั้นที่พบความแตกต่างอย่างมีนัยสำคัญที่ระดับ 0.05 ส่วนความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถผู้สอบไม่พบความแตกต่างอย่างมีนัยสำคัญ แสดงว่าการละเมิดความเป็นอิสระกันของข้อสอบไม่มีอิทธิพลต่อความคลาดเคลื่อนมาตรฐานของพารามิเตอร์อำนาจจำแนก พารามิเตอร์การเดา และพารามิเตอร์ความสามารถผู้สอบ สอดคล้องกับงานวิจัยของ Min and He (2014) ที่เปรียบเทียบความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ข้อสอบระหว่างวิธี UIRT และ TRT พบว่าความคลาดเคลื่อนมาตรฐานของพารามิเตอร์ระหว่างวิธีทั้งสองไม่แตกต่างกัน

3. ประสิทธิภาพของการเชื่อมโยงคะแนนระหว่างแบบสอบ

ประสิทธิผลของการเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบ และการเชื่อมโยงคะแนนสังเกตได้ด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ภายใต้การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองเทสต์เลท ผู้วิจัยพิจารณาประสิทธิผลจากค่าความลำเอียง (Bias) เป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างเบีรนระบบ ค่าความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบ (Standard Error of Equating: SEE) เป็นความคลาดเคลื่อนที่เกิดขึ้นอย่างสุ่ม และค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) เป็นความคลาดเคลื่อนรวม ดังนั้นดัชนีทั้ง 3 ค่านี้จะต้องมีค่าน้อย ๆ จึงจะทำให้ความแม่นยำในการเชื่อมโยงคะแนนสูง โดยมีรายละเอียดของการอภิปรายผลแต่ละดัชนี ดังนี้

3.1 ความลำเอียง (Bias)

ประสิทธิผลของการเชื่อมโยงคะแนนที่พิจารณาจากค่าความลำเอียง พบว่าการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองเทสต์เลทมีค่าความลำเอียงน้อยกว่าการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ โดยที่ความลำเอียงจากการเชื่อมโยงคะแนนสังเกตได้น้อยกว่าการเชื่อมโยงคะแนนจริง แสดงว่าการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลทมีแม่นยำในการเชื่อมโยงมากกว่าวิธีอื่น ๆ สอดคล้องกับงานวิจัยของ Tao and Cao (2016) ที่ศึกษาเปรียบเทียบความลำเอียงที่เกิดขึ้นจากการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ด้วยการประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลท พบว่าความลำเอียงของการเชื่อมโยงคะแนนสังเกตได้มีค่าน้อยกว่าการเชื่อมโยงคะแนนจริง และยังสรุปว่าการเชื่อมโยง

คะแนนสังเกตได้มีความแกร่งมากกว่าการเชื่อมโยงคะแนนจริง นอกจากนี้ยังสอดคล้องกับงานวิจัยของ Chen (2014) ได้จำลองข้อมูลสำหรับแบบสอบมีเทสต์เลขจำนวน 10 เทสต์เลข และมีอิทธิพลของเทสต์เลขต่างกัน พบว่า แบบสอบที่มีอิทธิพลเทสต์เลขเมื่อประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของข้อสอบจะมีค่าความลำเอียงสูงกว่าประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลข

3.2 ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (SEE)

ประสิทธิผลของการเชื่อมโยงคะแนนที่พิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ พบว่า การเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลขมีค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบน้อยกว่าการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของข้อสอบ โดยที่ความคลาดเคลื่อนมาตรฐานของการปรับเทียบจากการเชื่อมโยงคะแนนสังเกตได้น้อยกว่าการเชื่อมโยงคะแนนจริง แสดงว่าการเชื่อมโยงคะแนนสังเกตได้ที่ประมาณค่าตามทฤษฎีการตอบสนองของเทสต์เลขมีแม่นยำในการเชื่อมโยงมากกว่าวิธีอื่น ๆ สอดคล้องกับงานวิจัยของ Tao and Cao (2016) ที่ศึกษาเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานของปรับเทียบที่เกิดขึ้นจากการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ด้วยการประมาณค่าตามทฤษฎีการตอบสนองของเทสต์เลข พบว่า ค่า SEE ของการเชื่อมโยงคะแนนสังเกตได้มีค่าน้อยกว่าการเชื่อมโยงคะแนนจริง และยังสรุปว่าการเชื่อมโยงคะแนนสังเกตได้มีความแกร่งมากกว่าการเชื่อมโยงคะแนนจริง อีกทั้งยังสอดคล้องกับงานวิจัยของ Chen (2014) ได้จำลองข้อมูลแบบสอบประกอบด้วย 10 เทสต์เลข และระดับอิทธิพลของเทสต์เลขต่างกัน พบว่า เมื่อแบบสอบมีอิทธิพลของเทสต์เลข ค่า SEE ของการเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้ด้วยการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของข้อสอบสูงกว่าการประมาณค่าตามทฤษฎีการตอบสนองของเทสต์เลข

3.3 ค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (RMSE)

ประสิทธิผลของการเชื่อมโยงคะแนนที่พิจารณาจากค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน พบว่า การเชื่อมโยงคะแนนจริงภายใต้การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลขมีค่าต่ำที่สุด นั้นแสดงว่าความคลาดเคลื่อนรวมของการเชื่อมโยงคะแนนมีน้อยส่งผลให้มีความแม่นยำในการเชื่อมโยงคะแนน เมื่อนำค่า RMSE ของการเชื่อมโยงคะแนนจริงที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองของเทสต์เลขมาเทียบกับการเชื่อมโยงส่วนการเชื่อมโยงคะแนนสังเกตได้พบว่า การเชื่อมโยงคะแนนจริงด้วยการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลขมีค่า RMSE น้อยกว่า แสดงว่าการเชื่อมโยงคะแนนจริงที่ประมาณค่าตามทฤษฎีการตอบสนองของเทสต์เลขมีแม่นยำในการเชื่อมโยงมากกว่าวิธีอื่น ๆ

สอดคล้องกับงานวิจัยของ Tao and Cao (2016) ที่สรุปว่าค่า URMS/WRMS ของการเชื่อมโยงคะแนนจริงที่ประมาณค่าด้วยทฤษฎีการตอบสนองข้อสอบสูงกว่าการประมาณค่าด้วยทฤษฎีการตอบสนองของเทสต์เลท แต่จะขัดแย้งกับผลของการวิจัยเมื่อเปรียบเทียบระหว่างวิธีการเชื่อมโยงจริงกับการเชื่อมโยงคะแนนสังเกตได้ตามทฤษฎีการตอบสนองของเทสต์เลทเพราะผลการวิจัยของ Tao and Cao นั้นการเชื่อมโยงคะแนนสังเกตได้จะมีความคลาดเคลื่อนน้อยกว่าการเชื่อมโยงคะแนนจริง นอกจากนี้จากการศึกษาของ Chen (2014) ยังพบว่าค่า URMS/WRMS ที่ได้จากทั้งแบบสอบจริงและแบบสอบที่จำลองขึ้นให้ผลไปในทิศทางเดียวกันคือ การเชื่อมโยงคะแนนจริงและการเชื่อมโยงคะแนนสังเกตได้จากการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบจะให้ค่า URMS/WRMS สูงกว่าการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลท

ดังนั้นประสิทธิผลของการเชื่อมโยงคะแนนระหว่างแบบสอบของการศึกษาคั้งนี้ ถ้าต้องการให้เกิดความคลาดเคลื่อนอย่างเป็นระบบและอย่างสุ่มน้อย ควรเลือกวิธีการเชื่อมโยงคะแนนสังเกตได้ภายใต้การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลท และถ้าต้องการให้เกิดความคลาดเคลื่อนรวมน้อยควรเลือกวิธีการเชื่อมโยงคะแนนจริงภายใต้การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของเทสต์เลท แต่การเชื่อมโยงคะแนนคั้งนี้อยู่ภายใต้แบบสอบคนละวิชาที่มีโครงสร้างของเนื้อหาที่ต่างกัน ซึ่งงานวิจัยส่วนใหญ่เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่คู่ขนานกัน ดังนั้นผลการศึกษาคั้งนี้อาจจะเป็นแนวทางในการพัฒนาต่อยอดองค์ความรู้เกี่ยวกับการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีลักษณะต่างกัน

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. จากผลการวิจัยที่พบว่าความสัมพันธ์ระหว่างค่าพารามิเตอร์ข้อสอบที่ประมาณค่าพารามิเตอร์ด้วยทฤษฎีการตอบสนองข้อสอบและทฤษฎีการตอบสนองของเทสต์เลทมีความสัมพันธ์กัน ดังนั้นผู้สร้างหรือผู้ใช้ข้อสอบควรให้ความสำคัญกับการนำแบบสอบไปใช้ให้ตรงกับวัตถุประสงค์และระดับความสามารถของผู้สอบ อีกทั้งการบริหารจัดการสอบต้องเป็นมาตรฐาน
2. จากผลการวิจัยที่พบว่าในวิชาภาษาไทยพารามิเตอร์ความสามารถผู้สอบไม่มีความสัมพันธ์กัน และยังพบว่าความแปรปรวนของอิทธิพลเทสต์เลทสูงด้วยเมื่อเทียบกับวิชาภาษาอังกฤษ ดังนั้นผู้สร้างข้อสอบควรให้ความสำคัญเกี่ยวกับความไม่เป็นอิสระของข้อสอบ ถ้าข้อสอบมีความไม่เป็นอิสระในการตอบข้อสอบสูงจะส่งผลให้เกิดความคลาดเคลื่อนในการประมาณความสามารถของผู้สอบสูงด้วย

3. จากผลการวิจัยที่พบว่าความคลาดเคลื่อนมาตรฐานระหว่างการประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ และโมเดลทฤษฎีการตอบสนองเทสต์เลทมีความแตกต่างกัน ดังนั้นผู้สร้างข้อสอบและผู้ใช้แบบสอบที่มีลักษณะเทสต์เลทหรือไม่เป็นอิสระควรพึงระวังในการนำข้อสอบไปใช้ โดยเฉพาะการใช้กับการทดสอบแบบปรับเหมาะเพราะการเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้สอบจะนำค่าความคลาดเคลื่อนมาใช้ในการพิจารณาตัดสิน

4. จากผลการวิจัยที่พบว่าการเชื่อมโยงคะแนนจริงด้วยวิธีโค้งคุณลักษณะแบบสอบที่ประมาณค่าตามทฤษฎีการตอบสนองเทสต์เลทมีความคลาดเคลื่อนของการเชื่อมโยงทั้ง 3 ค่า คือ ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ และค่าดัชนีรากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนน้อยกว่าการประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ ดังนั้นนักการศึกษาหรือนักวัดผลควรตระหนักถึงความสำคัญเกี่ยวกับการประมาณค่าพารามิเตอร์ที่คำนึงถึงอิทธิพลของความไม่เป็นอิสระของการตอบข้อสอบถูก เมื่อแบบสอบที่สร้างขึ้นหรือนำมาใช้มีลักษณะของความไม่เป็นอิสระในการตอบ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. สำหรับการประมาณค่าพารามิเตอร์

1.1 ควรศึกษาเปรียบเทียบโมเดลการทดสอบเกี่ยวกับการประมาณค่าพารามิเตอร์ที่มีแบบสอบในลักษณะที่แตกต่างออกไป เช่น เป็นแบบสอบอัตนัย เป็นแบบสอบผสมทั้งอัตนัยและปรนัย

1.2 ควรศึกษาเกี่ยวกับการประมาณค่าพารามิเตอร์ข้อสอบเมื่อแบบสอบหนึ่งฉบับมีคุณลักษณะที่แตกต่างกันหรือวัดความรู้/ เนื้อหามากกว่าหนึ่งมิติ

1.3 ควรศึกษาเกี่ยวกับขนาดของความไม่เป็นอิสระกันของการตอบข้อสอบที่เหมาะสมกับความสามารถของผู้สอบระดับความสามารถนั้น ๆ

2. สำหรับการเชื่อมโยงคะแนน

2.1 ควรศึกษาและเปรียบเทียบวิธีการเชื่อมโยงคะแนนรูปแบบในการศึกษานี้กับรูปแบบอื่น เช่น การเชื่อมโยงคะแนนวิธี Newton-Raphson การเชื่อมโยงคะแนนจริง

2.2 ควรศึกษาคุณภาพการเชื่อมโยงคะแนนในประเด็นอื่น ๆ เช่น ความไม่แปรเปลี่ยนของผลการเชื่อมโยงคะแนนด้วยวิธีต่าง ๆ เมื่อกลุ่มผู้สอบมีลักษณะส่วนบุคคลต่างกัน เช่น เพศชายและเพศหญิง โรงเรียนขนาดใหญ่และโรงเรียนขนาดเล็ก เป็นต้น

2.3 ควรศึกษาเปรียบเทียบคุณภาพประเด็นต่าง ๆ ของการเชื่อมโยงคะแนนสำหรับแบบสอบที่มีคุณลักษณะโครงสร้างเนื้อหา คุณลักษณะข้อสอบต่างกันระหว่างการให้ข้อมูลจริงและการให้ข้อมูลจำลอง (Simulation)

บรรณานุกรม

- กระทรวงศึกษาธิการ. (2559). สถิติทางการศึกษาปี 2558. เข้าถึงได้จาก <http://www.bopp-obec.info/home/wp-content/uploads/2015/09/t32.pdf>
- กิติพงษ์ ลีอนาม. (2551). การนำเสนอวิธีการปรับเทียบคะแนนความสามารถโดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์การศึกษาดุษฎีบัณฑิต, สาขาวิชาการวิจัยและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยนเรศวร.
- จิระนาฏ ฉวีพัฒน์, บัณฑิตา อินสมบัติ และธีรยุทธ ภูเขา. (2554). การเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวคิดโดยใช้แบบทดสอบร่วมภายในระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์. วิทยาการวิจัยและวิทยาการปัญญา, 9(1), 78-89.
- จุฬาลักษณ์ บารมี. (2555). สถิติเพื่อการวิจัยทางสุขภาพและการวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS (พิมพ์ครั้งที่ 2). ชลบุรี: บางแสนการพิมพ์.
- ภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (ม.ป.ป.). ภาษาศาสตร์คืออะไร. เข้าถึงได้จาก <http://www.arts.chula.ac.th/~ling/category.php?id=1&>
- ภาวิณี ศรีสุขวัฒนานนท์. (2529). การเปรียบเทียบผลการใช้รูปแบบการเทียบมาตราที่ต่างกันเมื่อแบบทดสอบร่วมมีความยาวต่างกัน. ดุษฎีนิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการศึกษาศาสตร์, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2541). การปรับเทียบคะแนนระหว่างแบบสอบ. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (Modern test theory) (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุนทร เทียนงาม. (2551). ผลของความไม่เป็นอิสระของข้อสอบที่มีต่อค่าความเที่ยงค่าพารามิเตอร์ของข้อสอบค่าความสามารถของผู้สอบและสารสนเทศของแบบสอบเมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน. ดุษฎีนิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- สุรศักดิ์ อมรรัตนศักดิ์. (2544). ทฤษฎีการทดสอบ (Test theory). กรุงเทพฯ: มหาวิทยาลัยรามคำแหง.

- อัษฎาลี ศรีภักษาณู. (2552). *คุณภาพการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสม: การประยุกต์ใช้การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบด้วยวิธีโค้งคุณลักษณะ และการปรับค่าพารามิเตอร์ พร้อมกัน*. ดุษฎีนิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- อดิศร ศรีบุญวงษ์. (2545). *การพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ*. ดุษฎีนิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.
- อรทัย เจริญสิทธิ์. (2559). *การพัฒนาวิธีการปรับเทียบสเกลผลการเรียนสะสมตลอดหลักสูตร มัธยมศึกษาตอนปลายหรือเทียบเท่า โดยการขจัดอิทธิพลของขนาดโรงเรียน และใช้คะแนนผลการสอบทางการศึกษาแห่งชาติขั้นพื้นฐานเป็นคะแนนเชื่อมโยง*. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.
- อริสฟ้า เตห์ลิม. (2559). *การเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมกซิมัมไลค์ลิฮูด วิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลทดสอบ*. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.; pp. 508-600). Washington, DC: American Council on Education.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical Theories of Mental Test Score* (pp. 397-479). Reading, MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bowman, C. M., & Peng, S. S. (1972). *A preliminary investigation of recent advanced psychology tests in the GRE program-an application of a cognitive classification system*. Princeton, NJ.: SAGE.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153-168.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cao, Y., Lu, R., & Tao, W. (2014). *Effect of item response theory (IRT) model selection on testlet-based tests equating*. Princeton, NJ: Educational Testing Service.
- Chen, J. (2014). *Model selection for IRT equating of testlet-based tests in the random groups design*. Doctoral dissertation, Philosophy, University of Iowa.
- Cook, L. L., & Eignor, D. R. (1983). *An investigation of the feasibility of applying item response theory to equate achievement tests. (Paper presented at the annual meeting of the American Educational Research Association, Montreal)*. Princeton, NJ.: SAGE
- Crocker, L., & Algina, J. (2008). *Introduction to classical and model test theory*. New York: Cengage Learning.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford.
- Demars, C. E. (2006). Application of the bifactor multidimensional item response theory model to testlet-based test. *Journal of Educational Measurement*, 43, 145-168.
- Dorans, N. J. (2000). *Distinctions among classes of linkages*. College Board Research Note ZRN-11. New York: The College Board.
- Dorans, N. J. (2004). Equating, concordance and expectation. *Applied Psychological Measurement*, 28(4), 227-246.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281-306.
- Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). *Linking and aligning scores and scales*. New York: Springer.
- Eckes, T. (2014). Examining testlet effects in the TestDaf listening section: A testlet response. *Language Testing*, 31(1), 39-61.

- Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, DC: National Research Council.
- Flanagan, J. C. (1951). Unit, scores, and norms, In E.F. Lindquist (Ed.), *Educational Measurement*. (pp. 695-763). Washington, DC: American Council on Education.
- Foley, B. F. (2010). *Improving IRT item parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique*. Doctoral dissertation, Philosophy, Graduate College at the University of Nebraska.
- Galton, F. (1888). Co-relations and their measurements, chiefly from anthropological data. *Proceedings of the Royal Society of London*, 45, 135-145.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22, 144-149.
- Hanson, B. A. (1991). A note on Levine's formula for equating unequally reliable tests using data from the common item nonequivalent groups design. *Journal of Educational Statistics*, 16, 93-100.
- He, W., Li, F., & Wolfe, E. W. (2012). *Model selection for equating of testlet-based tests in the NEAT design: An empirical study. (Paper presented at the 2012 Annual NCME conference)*. New York, NY: Springer.
- Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds), *Linking and Aligning Score and Scales* (pp. 5-30). New York, NY: Springer.
- Holland, P. W., & Hoskens, M. (2003). *Classical test theory as a first-order item response theory: Application to true-score prediction form a possibly nonparallel test (ETS Educational Testing Service RR-20-02)*. Princeton, NJ: Research & Development Division.
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Kelley, T. L. (1923). *Statistical methods*. New York: Macmillan.

- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: World Book.
- Kogar, E. Y., & Kelecioğlu, H. (2017). Examination of different item response theory models on tests composed of testlets. *Journal of Education and Learning*, 6(4), 113-126.
- Kolen, M. J. (2004). Linking assessments: Concept and history. *Applied Psychological Measurement*, 28, 219-226.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.
- Kim, K. (2015). *Model comparisons among Testlet Response Theory (TRT) on a reading comprehension test*. Doctoral dissertation, Philosophy, College of Graduate Studies at Middle Tennessee State University.
- Lam, W. (2009). *Linking current and future score scales for the AICPA Uniform CPA Exam*. Boston: Technical Report in University of Massachusetts Amherst.
- Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237-255.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25, 357-372.
- Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling*, 12(1), 1-27.
- Li, D. (2009). *Developing a common scale for testlet model parameter estimates under the common-item nonequivalent groups design*. Doctoral dissertation, Graduate School of the University of Maryland, College Park.
- Li, Y., Bolt, D. M., & Fu, J. (2005). A test characteristic curve linking method for the testlet model. *Applied Psychological Measurement*, 29(5), 340-356.
- Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education*, 6(1), 83-102.

- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Min, S., & He, L. (2014). Applying unidimensional item response theory models in testlet-based reading assessment. *Language Testing*, 31(4), 453-447.
- Mislevy, R. L. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: ETS Policy Information Center.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Nickerson, R. S. (1989). On improving thinking through instruction. In E. Z. Rothkopf (Ed.), *Review of Research in Education* (Vol. 15, pp. 3-58). Washington, DC: American Educational Research Association.
- Ojerinde, D., Popoola, O., Onyeneho, P., & Egberongbe, A. (2016). A comparative analysis of pre-equating and post-equating in a large-scale assessment, high stakes examination. *Perspectives in Education*, 34(4), 79-98.
- Puhan, G., Moses, T. P., Grant, M.C., & McHale, F. (2009). Small- sample equating using a single-group nearly equivalent test (SiGNET) design. *Journal of Educational Measurement*, 46(3), 344-362.
- Purya, B., & Hamdollah, R. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicologica*, 37, 85-104.
- Rui, W., Shu-Liang, D., & Deng-Wen, G. (2010). Test Equating with Testlets. *Acta Psychologica Sinica*, 42(3), 343-442.
- Samejima, F. (1969). *Estimating of latent ability using a response pattern of graded scores*. Richmond: Byrd Press.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.

- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Stocking, M. L., & Lord, F. M. (1983) Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimateion in the rasch model. *Journal of Educational Statistics*, 7, 175-191.
- Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimateion in the two-parameter model. *Psychometrika*, 50, 349-364.
- Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimateion in the three-parameter model. *Psychometrika*, 51, 589-601.
- Tao, W., & Cao, Y. (2016). An extension of IRT-based equating to the dichotomous testlet response theory model. *Applied Measurement in Education*, 29(2), 108-121.
- Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, 26, 247-260.
- Tian, F. (2011). *A comparison of equating/ linking using the Stocking-Load method and concurrent calibration with mixed-format test in the non-equivalent group*. Doctoral dissertation, Philosophy degree in Educational Research, Measurement, and Evaluation in the Boston College Lynch School of Education.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel method of equating*. New York: Springer.
- Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 law school admissions test as an example. *Applied Measurement in Education*, 8, 157-186.
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the 3PL model useful in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glass (Eds.), *Computerized adaptive testing: Theory and practice*. (pp. 245-269). Netherlands: Kluwer Academic Publishers.

- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its application*. Cambridge, UK: Cambridge University Press.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlet. *Journal of Educational Measurement*, 24, 185-201.
- Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, 15(1), 22-29.
- Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., & Nelson, L. (2001). Augmented scores "Borrowing strength" to compute scores based on small numbers of items. In D. Thissen & H. Wainer (Eds.), *Test Scoring*. (pp. 343-383). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, 37, 203-220.
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: theory and application. *Applied Psychological Measurement*, 26, 109-128.
- Wang, W., & Wilson, M. (2005b). The raschtestlet model. *Applied Psychological Measurement*, 29, 126-149.
- Wang, X., Bradlow, E. T., & Wainer, H. (2005). *User's guide for SCORIGHT (Version 3.0): A computer program for scoring tests built of testlets including a module for covariate analysis (ETS Research Report RR 04-49)*. Princeton, NJ: Educational Testing Service.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.
- Yi, Cao, Ru, Lu, & Wei, Tao. (2014). *Effect of item response theory (IRT) model selection on testlet-based tests equating*. Princeton, NJ: Educational Testing Service.
- Zhang, J. (2007). *Dichotomous or Polytomous model? Equating of testlet-based tests in light of conditional item pair correlations*. Doctoral dissertation, Philosophy degree in Psychological and Quantitative Foundations (Educational Measurement and Statistics) in the Graduate College of The University of Iowa.

ภาคผนวก

ภาคผนวก ก
หนังสือขอความอนุเคราะห์ข้อมูล

ที่ ศธ ๖๖๒๑/



ศึกษาศาสตร์ มหาวิทยาลัยบูรพา
๑๖๙ ถ.ลงหาดบางแสน ต.แสนสุข
อ.เมือง จ.ชลบุรี ๒๐๑๓๑

๑๒ กรกฎาคม ๒๕๕๙

เรื่อง ขอความอนุเคราะห์ผลสอบรายข้อ

เรียน ผู้อำนวยการสถาบันการทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

สิ่งที่ส่งมาด้วย แบบฟอร์มบันทึกคะแนนนักเรียน

ด้วย นางสาวจงกล บัวแก้ว นิสิตระดับปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชา
วิจัย วัดผลและสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา อยู่ระหว่างดำเนินการสอบเค้าโครง
ดุษฎีนิพนธ์ เรื่อง “ประสิทธิผลของการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีชุดข้อสอบด้วยทฤษฎี
การตอบสนองข้อสอบ” โดยมี รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม เป็นอาจารย์ที่ปรึกษา นั้น

ในการทำวิจัยครั้งนี้ ผู้วิจัยมีความจำเป็นต้องใช้ข้อมูลผลการสอบรายข้อแบบทดสอบทางการศึกษา
ระดับชาติขั้นพื้นฐาน (O-NET) ในระดับชั้นมัธยมศึกษาปีที่ ๓ ในการตอบวัตถุประสงค์ของการวิจัย ข้อมูลที่ผู้วิจัยต้องการ
ขอความอนุเคราะห์จากสถาบันการทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ได้แก่ คะแนนสอบรายข้อรายวิชา
ภาษาไทย และภาษาอังกฤษ ในปีการศึกษา ๒๕๕๘ และข้อมูลพื้นฐานของนักเรียน ได้แก่ เพศ ขนาดโรงเรียน และ
ภูมิภาค ซึ่งผลการวิจัยครั้งนี้จะเป็นประโยชน์เชิงวิชาการในการพัฒนาวิธีการปรับเทียบสเกลระหว่างแบบสอบ เมื่อผู้วิจัย
ทำการวิจัยและสร้างข้อสอบในภาพรวมเรียบร้อยแล้ว จะลบข้อมูลดังกล่าวทิ้งทันที ส่วนระหว่างการวิเคราะห์ข้อมูลจะเก็บ
รักษาข้อมูลดังกล่าวไว้เป็นความลับ และเข้าถึงข้อมูลได้เฉพาะข้าพเจ้าและผู้ร่วมวิเคราะห์ข้อมูลเท่านั้น

จึงเรียนมาเพื่อโปรดทราบและอนุเคราะห์ผลการสอบดังกล่าว คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
หวังเป็นอย่างยิ่งว่าจะได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอแสดงความนับถือ

(ผู้ช่วยศาสตราจารย์ ดร.เชษฐ ศิริสวัสดิ์)

รองคณบดีฝ่ายบัณฑิตศึกษา ปฏิบัติการแทน

คณบดีคณะศึกษาศาสตร์ ปฏิบัติการแทน

ผู้รักษาการแทนอธิการบดีมหาวิทยาลัยบูรพา

ภาควิชาวิจัยและจิตวิทยาประยุกต์

โทรศัพท์ ๐-๓๘๑๐-๒๐๗๖

ผู้วิจัยโทร. ๐๘-๖๙๖๕-๘๓๑๔

ภาคผนวก ข
พารามิเตอร์ข้อสอบ

ตารางภาคผนวกที่ ข-1 ค่าพารามิเตอร์ข้อสอบวิชาภาษาไทย

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
1	0.037, 0.008	9.622, 0.272	0.002, 0.003	0.036, 0.006	9.702, 0.219	0.001, 0.002
2	0.002, 0.003	8.894, 0.420	0.128, 0.028	0.002, 0.003	9.139, 0.371	0.126, 0.032
3	0.002, 0.002	8.846, 0.517	0.122, 0.024	0.003, 0.003	9.148, 0.501	0.122, 0.028
4	0.132, 0.008	9.934, 0.070	0.001, 0.002	0.128, 0.009	9.949, 0.050	0.001, 0.002
5	0.002, 0.003	8.890, 0.450	0.195, 0.030	0.003, 0.003	9.155, 0.413	0.199, 0.024
6	0.004, 0.003	9.132, 0.449	0.017, 0.014	0.005, 0.004	9.409, 0.288	0.014, 0.015
7	0.053, 0.007	9.774, 0.173	0.001, 0.002	0.054, 0.006	9.769, 0.187	0.001, 0.001
8	0.040, 0.005	9.615, 0.228	0.002, 0.002	0.039, 0.005	9.686, 0.212	0.001, 0.002
9	0.001, 0.002	8.623, 0.588	0.515, 0.021	0.002, 0.002	8.956, 0.539	0.516, 0.025
10	0.017, 0.005	9.491, 0.326	0.003, 0.004	0.016, 0.005	9.557, 0.255	0.004, 0.006
11	0.123, 0.007	9.944, 0.053	0.001, 0.001	0.123, 0.009	9.926, 0.070	0.001, 0.001
12	0.029, 0.006	9.593, 0.312	0.002, 0.002	0.031, 0.005	9.657, 0.220	0.002, 0.003

ตารางภาคผนวกที่ ข-1 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
13	0.018, 0.006	9.471, 0.265	0.003, 0.006	0.018, 0.006	9.581, 0.251	0.003, 0.003
14	0.002, 0.002	8.680, 0.574	0.354, 0.028	0.002, 0.002	9.019, 0.442	0.340, 0.017
15	0.041, 0.006	9.660, 0.250	0.001, 0.002	0.040, 0.005	9.717, 0.211	0.001, 0.002
16	0.023, 0.006	9.490, 0.317	0.003, 0.003	0.023, 0.006	9.575, 0.264	0.004, 0.006
17	0.018, 0.005	9.422, 0.358	0.004, 0.006	0.019, 0.005	9.551, 0.231	0.003, 0.003
18	0.002, 0.002	8.811, 0.441	0.229, 0.028	0.002, 0.003	9.105, 0.388	0.237, 0.024
19	0.046, 0.005	9.670, 0.238	0.001, 0.002	0.045, 0.005	9.727, 0.193	0.001, 0.001
20	0.118, 0.009	9.907, 0.065	0.001, 0.002	0.116, 0.008	9.931, 0.068	0.001, 0.003
21	0.024, 0.006	9.508, 0.314	0.002, 0.004	0.025, 0.005	9.582, 0.242	0.003, 0.004
22	0.003, 0.002	9.038, 0.459	0.073, 0.029	0.003, 0.003	9.176, 0.383	0.077, 0.029
23	0.001, 0.002	8.725, 0.564	0.288, 0.028	0.002, 0.004	9.075, 0.467	0.293, 0.028
24	0.110, 0.006	9.898, 0.087	0.001, 0.001	0.110, 0.006	9.915, 0.070	0.001, 0.002

ตารางภาคผนวกที่ ข-1 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
25	0.002, 0.003	8.731, 0.546	0.210, 0.024	0.003, 0.003	9.098, 0.415	0.219, 0.029
26	0.034, 0.006	9.610, 0.272	0.002, 0.002	0.036, 0.006	9.671, 0.225	0.002, 0.003
27	0.003, 0.004	8.840, 0.479	0.132, 0.030	0.003, 0.003	9.170, 0.395	0.128, 0.028
28	0.001, 0.002	8.760, 0.506	0.288, 0.021	0.002, 0.003	9.050, 0.426	0.285, 0.028
29	0.054, 0.006	9.800, 0.167	0.001, 0.002	0.054, 0.006	9.781, 0.164	0.001, 0.001
30	0.023, 0.005	9.541, 0.275	0.003, 0.003	0.022, 0.006	9.570, 0.257	0.003, 0.005
31	0.063, 0.005	9.851, 0.116	0.001, 0.001	0.065, 0.007	9.809, 0.138	0.001, 0.001
32	0.004, 0.003	9.077, 0.409	0.024, 0.021	0.004, 0.003	9.259, 0.350	0.032, 0.022
33	0.006, 0.004	9.268, 0.406	0.009, 0.009	0.007, 0.004	9.419, 0.321	0.010, 0.012
34	0.003, 0.002	8.959, 0.432	0.053, 0.021	0.004, 0.004	9.225, 0.349	0.056, 0.031
35	0.025, 0.005	9.498, 0.290	0.003, 0.004	0.025, 0.005	9.654, 0.242	0.002, 0.003
36	0.024, 0.006	9.540, 0.297	0.003, 0.006	0.024, 0.005	9.637, 0.264	0.003, 0.005

ตารางภาคผนวกที่ ข-1 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
37	0.024, 0.006	9.561, 0.299	0.002, 0.002	0.023, 0.006	9.642, 0.220	0.002, 0.002
38	0.002, 0.003	8.732, 0.515	0.250, 0.030	0.002, 0.003	9.070, 0.396	0.247, 0.028
39	0.057, 0.006	9.720, 0.222	0.001, 0.002	0.057, 0.005	9.778, 0.148	0.001, 0.002
40	0.003, 0.003	8.953, 0.455	0.051, 0.028	0.003, 0.003	9.242, 0.337	0.053, 0.029
41	5.631, 13.508	-0.188, 4.977	0.000, 0.000	6.443, 17.247	2.754, 3.896	0.000, 0.000
42	6.397, 17.492	0.173, 5.592	0.000, 0.000	3.864, 13.319	0.621, 5.755	0.000, 0.000
43	4.315, 13.348	0.664, 5.511	0.000, 0.000	7.490, 19.984	-3.851, 3.285	0.000, 0.000
44	4.417, 12.421	1.505, 6.060	0.000, 0.000	5.242, 15.020	-0.895, 4.315	0.000, 0.000
45	5.214, 14.831	0.474, 5.838	0.000, 0.000	8.984, 19.354	1.010, 6.229	0.000, 0.000
46	6.033, 16.013	2.599, 5.604	0.000, 0.000	5.441, 14.997	4.369, 3.178	0.000, 0.000
47	7.007, 19.100	2.385, 4.600	0.000, 0.000	4.659, 14.956	-0.304, 4.165	0.000, 0.000
48	4.828, 11.894	-0.174, 5.282	0.000, 0.000	2.766, 8.162	5.872, 3.135	0.000, 0.000

ตารางภาคผนวกที่ ข-1 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
49	6.123,	-1.469,	0.000,	5.696,	5.898,	0.000,
	15.141	5.160	0.000	15.187	3.117	0.000
50	5.098,	0.259,	0.000,	5.735,	3.591,	0.000,
	14.519	5.288	0.000	14.399	5.213	0.000

ตารางภาคผนวกที่ ข-2 ค่าพารามิเตอร์ข้อสอบวิชาภาษาอังกฤษ

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
1	3.173,	1.669,	0.262,	2.773,	1.699,	0.255,
	0.660	0.088	0.013	0.595	0.096	0.016
2	4.325,	0.990,	0.243,	5.701,	0.991,	0.233,
	0.682	0.060	0.019	2.333	0.051	0.016
3	2.333,	1.399,	0.211,	2.142,	1.449,	0.203,
	0.453	0.083	0.021	0.383	0.090	0.018
4	1.778,	1.022,	0.206,	1.786,	1.054,	0.205,
	0.250	0.099	0.026	0.380	0.112	0.035
5	2.564,	0.462,	0.226,	2.901,	0.476,	0.227,
	0.381	0.078	0.026	0.716	0.104	0.037
6	2.110,	1.927,	0.217,	1.816,	2.024,	0.209,
	0.414	0.115	0.015	0.463	0.142	0.022
7	4.561,	1.059,	0.243,	4.570,	1.071,	0.246,
	0.737	0.068	0.016	1.003	0.068	0.018

ตารางภาคผนวกที่ ข-2 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
8	2.108, 0.376	1.372, 0.087	0.193, 0.023	2.006, 0.303	1.405, 0.096	0.193, 0.021
9	2.590, 0.561	1.025, 0.084	0.253, 0.028	2.631, 0.485	1.044, 0.078	0.255, 0.023
10	2.372, 0.298	0.439, 0.089	0.224, 0.030	2.476, 0.386	0.487, 0.090	0.240, 0.035
11	2.941, 0.425	1.147, 0.063	0.212, 0.016	2.802, 0.492	1.170, 0.078	0.205, 0.019
12	1.948, 0.423	2.324, 0.197	0.220, 0.016	1.845, 0.641	2.570, 0.292	0.221, 0.014
13	0.554, 0.096	1.072, 0.217	0.154, 0.032	0.620, 0.149	1.365, 0.306	0.206, 0.059
14	4.540, 0.801	1.397, 0.062	0.237, 0.013	5.166, 1.390	1.457, 0.055	0.235, 0.011
15	3.115, 0.486	1.090, 0.074	0.233, 0.018	2.980, 0.565	1.094, 0.076	0.224, 0.022
16	7.161, 3.283	2.627, 0.304	0.341, 0.013	6.291, 2.784	3.230, 0.315	0.342, 0.012
17	2.411, 0.564	1.960, 0.167	0.231, 0.015	3.052, 0.925	2.036, 0.136	0.230, 0.015
18	1.628, 0.396	1.304, 0.101	0.205, 0.039	2.337, 0.754	1.360, 0.108	0.227, 0.028
19	3.961, 1.240	2.135, 0.154	0.256, 0.013	3.450, 1.500	2.564, 0.183	0.257, 0.014

ตารางภาคผนวกที่ ข-2 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
20	5.907, 2.970	3.035, 0.307	0.328, 0.011	5.362, 2.773	3.726, 0.351	0.327, 0.012
21	1.707, 0.275	1.777, 0.114	0.128, 0.018	2.352, 0.454	1.882, 0.124	0.122, 0.014
22	3.182, 1.055	2.232, 0.205	0.228, 0.011	2.940, 1.286	2.972, 0.206	0.232, 0.014
23	0.736, 0.147	1.228, 0.179	0.178, 0.040	1.000, 0.464	1.298, 0.221	0.206, 0.064
24	2.307, 0.603	1.721, 0.119	0.227, 0.021	2.062, 0.585	1.781, 0.099	0.219, 0.024
25	3.390, 0.554	1.474, 0.084	0.243, 0.015	2.927, 0.617	1.498, 0.073	0.234, 0.016
26	1.536, 0.440	2.125, 0.180	0.213, 0.028	1.457, 0.428	2.208, 0.196	0.210, 0.030
27	1.559, 0.394	1.698, 0.144	0.209, 0.029	1.489, 0.395	1.745, 0.120	0.206, 0.025
28	4.478, 1.426	1.846, 0.088	0.307, 0.015	2.902, 1.399	2.063, 0.145	0.291, 0.028
29	2.715, 0.833	2.755, 0.253	0.186, 0.011	2.821, 0.949	2.938, 0.269	0.187, 0.012
30	2.410, 0.478	1.622, 0.086	0.195, 0.018	2.584, 0.554	1.675, 0.089	0.195, 0.016
31	2.687, 0.603	1.784, 0.123	0.289, 0.016	2.823, 1.108	1.877, 0.109	0.287, 0.018

ตารางภาคผนวกที่ ข-2 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
32	4.579, 0.778	1.170, 0.068	0.279, 0.014	4.643, 1.261	1.165, 0.069	0.266, 0.019
33	6.122, 2.576	2.040, 0.142	0.357, 0.013	4.536, 3.250	2.266, 0.209	0.352, 0.016
34	1.154, 0.182	1.636, 0.130	0.173, 0.025	1.093, 0.166	1.778, 0.166	0.172, 0.026
35	1.812, 0.413	1.828, 0.131	0.207, 0.022	1.746, 0.406	1.945, 0.124	0.210, 0.022
36	1.893, 0.408	2.067, 0.152	0.217, 0.019	1.961, 0.685	2.208, 0.163	0.219, 0.018
37	2.125, 0.417	1.281, 0.097	0.233, 0.023	3.240, 1.232	1.341, 0.083	0.251, 0.025
38	3.171, 1.004	2.205, 0.187	0.266, 0.013	2.902, 0.970	2.506, 0.195	0.268, 0.013
39	1.155, 0.191	1.492, 0.153	0.171, 0.024	1.400, 0.269	1.514, 0.119	0.192, 0.025
40	1.675, 0.265	1.829, 0.129	0.168, 0.018	1.558, 0.318	1.956, 0.133	0.162, 0.017
41	1.230, 0.169	1.545, 0.127	0.172, 0.026	1.127, 0.207	1.583, 0.143	0.160, 0.027
42	6.004, 3.295	2.656, 0.280	0.347, 0.013	4.403, 2.546	2.923, 0.348	0.344, 0.013
43	1.328, 0.260	1.432, 0.124	0.196, 0.031	1.273, 0.237	1.454, 0.141	0.192, 0.031

ตารางภาคผนวกที่ ข-2 (ต่อ)

ข้อสอบ	IRT 3PL			TRT 3PL		
	a ,SE	b ,SE	c ,SE	a ,SE	b ,SE	c ,SE
44	0.982, 0.174	1.384, 0.147	0.169, 0.033	1.082, 0.227	1.440, 0.147	0.193, 0.041
45	1.348, 0.545	1.994, 0.173	0.220, 0.043	1.253, 0.446	2.046, 0.184	0.221, 0.034
46	1.879, 0.301	1.158, 0.083	0.173, 0.025	1.802, 0.179	1.162, 0.089	0.167, 0.022
47	4.657, 2.330	3.000, 0.314	0.295, 0.012	4.459, 1.984	3.185, 0.338	0.296, 0.013
48	1.507, 0.265	1.287, 0.094	0.175, 0.025	1.492, 0.241	1.276, 0.111	0.169, 0.029
49	1.102, 0.222	1.368, 0.131	0.188, 0.033	1.130, 0.235	1.424, 0.145	0.198, 0.043
50	1.510, 0.336	1.967, 0.179	0.224, 0.025	1.437, 0.361	2.041, 0.172	0.223, 0.026

ภาคผนวก ค

ตัวอย่างพารามิเตอร์ความสามารถ คะแนนดิบ คะแนนจริงของผู้สอบ

ตารางภาคผนวกที่ ค-1 คะแนนดิบ พารามิเตอร์ความสามารถ คะแนนจริง วิชาภาษาไทย

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
1	39	-0.346	6.149	20.412	-0.013	5.622	21.647
2	39	0.074	5.709	22.044	0.456	6.018	20.786
3	54	-0.226	5.706	20.111	0.730	5.860	24.220
4	54	0.042	6.130	21.033	-0.516	5.744	19.905
5	47	-0.463	5.800	23.373	1.698	5.499	22.755
6	40	-0.174	5.796	21.431	-1.004	5.427	22.182
7	45	0.289	5.684	21.652	0.785	5.844	22.376
8	46	-0.475	5.501	22.585	-0.138	6.058	20.801
9	52	0.995	5.650	19.736	-0.053	5.371	22.079
10	50	-0.488	5.713	24.984	0.375	5.811	20.871
11	44	0.123	5.741	21.529	-0.073	5.637	21.518
12	40	-0.083	5.659	23.751	0.458	5.919	20.842
13	28	-0.037	5.350	19.973	0.564	5.556	22.139
14	51	0.868	5.874	19.956	0.962	5.964	22.425
15	48	-0.071	6.203	25.317	0.317	6.185	24.335
16	44	-0.307	5.911	19.729	0.462	5.850	21.945
17	49	0.144	6.317	20.187	-0.535	4.952	22.281
18	46	0.070	6.001	19.789	-0.252	6.275	20.996
19	54	-0.426	5.810	22.988	0.566	5.846	19.425
20	58	1.094	5.303	21.285	-0.303	5.421	20.377
21	48	-0.417	6.040	20.305	-1.308	5.702	21.930
22	52	0.204	5.796	19.848	0.455	5.512	20.338
23	56	-0.723	5.801	25.295	0.319	5.894	24.125

ตารางภาคผนวกที่ ค-1 (ต่อ)

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
24	54	0.051	5.134	21.183	-0.248	5.096	21.477
25	52	-0.810	5.996	23.585	0.413	5.493	21.618
26	25	1.211	5.317	24.196	0.214	6.153	24.004
27	44	0.194	5.933	19.242	-0.478	5.548	22.076
28	41	0.648	5.425	26.151	-0.474	5.378	21.653
29	56	-0.713	5.391	20.106	-0.299	6.033	21.197
30	32	-0.062	5.996	24.380	-0.115	5.886	21.516
.
.
.
1470	46	-0.106	6.050	24.071	0.311	5.355	21.984
1471	23	-1.008	5.816	19.444	-0.235	5.713	20.617
1472	40	-0.372	5.631	19.738	-0.676	5.651	20.936
1473	28	-0.588	5.590	24.934	0.238	6.210	20.015
1474	47	-0.218	5.658	23.797	-0.635	5.514	20.425
1475	36	1.261	5.574	20.814	0.182	5.633	22.251
1476	41	0.278	5.565	22.196	0.403	5.422	23.121
1477	59	-0.626	5.741	21.160	0.499	5.794	20.666
1478	48	-0.636	5.527	23.714	-0.394	6.049	20.938
1479	36	0.237	6.047	21.051	-0.911	5.702	20.976
1480	47	-0.022	5.687	21.451	0.271	5.998	21.100
1481	51	-0.228	5.625	20.056	-0.430	6.060	21.901
1482	48	0.586	5.488	19.771	-0.341	5.951	20.203

ตารางภาคผนวกที่ ค-1 (ต่อ)

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
1483	38	0.515	6.048	20.515	1.170	6.065	21.904
1484	58	-0.435	6.339	19.579	-1.310	5.936	21.836
1485	38	0.392	6.048	22.150	-0.603	5.433	21.426
1486	46	-0.517	5.585	24.706	0.383	5.653	20.803
1487	36	0.618	6.112	20.337	0.212	6.125	20.541
1488	36	-0.621	5.608	26.278	-0.074	5.774	21.881
1489	48	0.053	5.959	19.840	-0.109	5.814	22.300
1490	51	-0.336	6.200	23.091	0.264	5.749	20.069
1491	41	0.568	6.099	21.721	-0.350	5.354	21.991
1492	51	-0.677	6.242	20.814	1.075	5.282	20.964
1493	38	0.419	5.774	20.005	-1.069	5.479	22.125
1494	42	0.203	5.652	24.880	-0.672	5.773	21.936
1495	26	0.418	5.986	24.690	-0.842	5.531	21.088
1496	33	-0.411	5.246	19.782	0.435	5.873	23.587
1497	65	-1.455	5.683	22.907	0.958	5.701	20.543
1498	47	-0.769	5.629	20.232	-0.032	5.495	22.204
1499	41	-0.699	5.696	18.965	-0.044	6.318	23.211
1500	52	-1.424	5.839	19.020	0.482	5.530	22.293

ตารางภาคผนวกที่ ค-2 พารามิเตอร์ความสามารถแต่ละเทสต์ไลท์ วิชาภาษาไทย

คนที่	เทสต์ไลท์				
	1	2	3	4	5
1	-0.183	-0.190	0.004	0.262	0.012
2	0.225	0.080	-0.029	0.196	-0.064
3	-0.221	0.061	0.001	0.020	0.001
4	0.154	0.017	0.019	0.250	0.016
5	-0.163	-0.007	0.019	0.095	-0.053
6	0.247	0.081	-0.069	0.307	-0.025
7	-0.010	0.038	0.001	-0.034	-0.032
8	-0.075	-0.058	-0.002	-0.103	-0.031
9	0.018	0.071	0.044	-0.423	0.006
10	0.155	0.061	0.051	0.097	0.039
11	-0.318	0.064	0.069	0.073	-0.043
12	0.001	-0.267	0.069	0.070	-0.044
13	0.202	-0.110	0.079	-0.209	0.026
14	0.131	-0.138	0.061	0.004	0.008
15	0.107	0.134	0.072	0.348	0.001
16	0.023	0.088	0.033	-0.206	0.041
17	-0.164	-0.038	0.190	-0.088	-0.009
18	-0.162	0.107	-0.170	-0.074	-0.026
19	-0.040	-0.031	0.033	-0.206	-0.035
20	0.171	0.103	-0.006	-0.097	-0.007
21	0.225	0.084	0.008	0.105	-0.057
22	-0.246	0.006	0.013	-0.050	-0.029
23	-0.015	-0.067	-0.004	-0.019	0.079

ตารางภาคผนวกที่ ค-2 (ต่อ)

คนที่	เทสต์เลข				
	1	2	3	4	5
24	0.053	-0.015	-0.041	-0.032	0.053
25	0.077	-0.164	-0.135	0.089	0.056
26	0.214	0.100	0.056	-0.125	0.009
27	0.139	-0.100	-0.036	0.034	0.002
28	-0.167	0.046	-0.023	0.104	-0.051
29	-0.036	0.078	-0.133	0.103	0.057
30	0.035	-0.076	0.001	0.060	0.009
.
.
.
1470	-0.025	-0.070	-0.025	-0.134	0.048
1471	0.350	0.100	-0.012	-0.113	-0.076
1472	-0.006	0.151	0.107	0.067	-0.045
1473	0.123	0.059	0.073	0.086	-0.025
1474	0.217	-0.054	-0.003	0.250	0.032
1475	0.147	0.021	-0.042	0.009	0.039
1476	0.110	-0.163	0.053	0.238	-0.011
1477	-0.212	0.073	-0.078	-0.059	-0.048
1478	0.070	-0.076	-0.051	0.215	0.028
1479	-0.007	0.100	-0.043	0.173	0.008
1480	0.080	0.080	0.032	0.035	0.087
1481	-0.196	-0.073	0.078	0.393	0.020
1482	-0.065	0.076	0.009	-0.040	0.083

ตารางภาคผนวกที่ ค-2 (ต่อ)

คนที่	เทสต์เลข				
	1	2	3	4	5
1483	0.014	0.015	0.087	-0.059	-0.008
1484	0.123	-0.157	0.089	0.345	-0.055
1485	0.054	-0.051	-0.117	0.229	-0.020
1486	-0.148	-0.007	0.025	-0.671	0.053
1487	-0.088	-0.013	0.103	-0.058	0.020
1488	0.266	0.019	0.037	-0.184	0.014
1489	0.173	0.136	-0.008	-0.457	-0.020
1490	0.079	-0.017	0.088	0.279	0.001
1491	0.004	0.001	0.038	0.022	0.007
1492	0.012	-0.116	-0.155	0.167	0.016
1493	-0.021	-0.026	0.144	0.056	0.003
1494	-0.252	-0.208	0.048	0.185	0.037
1495	0.013	-0.081	-0.026	0.370	0.077
1496	0.057	-0.002	0.056	-0.516	-0.070
1497	-0.084	0.213	0.039	-0.051	-0.068
1498	-0.023	0.113	0.084	-0.131	-0.015
1499	-0.114	-0.047	-0.029	-0.048	0.056
1500	-0.246	0.024	0.050	-0.218	0.001

ตารางภาคผนวกที่ ค-3 คะแนนดิบ พารามิเตอร์ความสามารถ คะแนนจริง วิชาภาษาอังกฤษ

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
1	20	-1.053	0.654	12.106	-1.060	0.767	11.979
2	26	1.316	0.191	13.088	1.363	0.235	12.818
3	30	0.093	0.547	12.924	0.022	0.632	12.996
4	36	-0.329	0.522	13.744	-0.359	0.619	14.020
5	34	-0.409	0.782	14.876	-0.270	0.683	15.350
6	16	-0.078	0.559	12.342	0.102	0.514	12.064
7	30	0.216	0.748	14.903	0.393	0.500	14.210
8	28	-0.004	0.539	15.151	-0.191	0.562	14.935
9	42	0.063	0.618	17.089	0.222	0.570	17.318
10	40	-0.517	0.796	17.627	-0.462	0.681	16.157
11	30	1.780	0.199	12.786	1.782	0.241	12.552
12	42	-0.804	0.703	20.805	-0.959	0.690	20.193
13	32	0.222	0.634	13.244	0.139	0.609	13.337
14	34	0.364	0.563	14.096	0.392	0.554	13.835
15	42	-0.720	0.712	17.434	-0.730	0.710	16.405
16	30	0.754	0.294	13.538	0.774	0.304	13.328
17	26	0.272	0.497	12.667	0.308	0.425	12.349
18	42	-1.150	0.678	21.529	-0.960	0.632	20.458
19	34	-0.676	0.756	15.397	-0.649	0.615	15.370
20	30	-0.085	0.505	12.574	-0.141	0.551	12.623
21	30	1.791	0.203	12.606	1.977	0.268	12.724
22	30	0.864	0.305	12.941	0.914	0.305	12.481
23	32	0.579	0.409	14.184	0.652	0.353	13.622

ตารางภาคผนวกที่ ค-3 (ต่อ)

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
24	46	0.058	0.637	20.966	0.062	0.623	19.976
25	56	-1.207	0.790	23.216	-1.145	0.667	24.198
26	24	0.090	0.909	12.324	0.089	0.856	12.229
27	22	0.052	0.609	13.030	0.249	0.574	12.606
28	32	0.646	0.263	12.678	0.510	0.526	12.542
29	28	-0.412	0.652	12.768	-0.526	0.676	12.543
30	24	-0.485	0.750	12.629	-0.507	0.691	12.263
.
.
.
1470	24	-0.801	0.630	12.811	-0.935	0.658	12.588
1471	16	0.818	0.309	12.193	0.798	0.323	12.092
1472	16	-0.130	0.732	12.110	-0.093	0.614	12.039
1473	32	0.248	0.397	13.254	0.202	0.429	12.808
1474	20	-0.286	0.651	12.288	-0.027	0.591	12.229
1475	36	0.041	0.638	17.568	-0.113	0.771	17.136
1476	34	1.322	0.197	15.877	1.307	0.246	15.193
1477	36	-0.625	0.643	15.475	-0.617	0.689	15.318
1478	42	-0.789	0.656	17.323	-0.746	0.638	17.146
1479	52	0.018	0.620	24.776	0.081	0.609	24.588
1480	34	-0.677	0.786	13.655	-0.678	0.616	13.329
1481	56	-0.161	0.735	26.432	-0.058	0.615	25.946
1482	32	-0.101	0.669	14.291	-0.071	0.720	14.348

ตารางภาคผนวกที่ ค-3 (ต่อ)

คนที่	คะแนนดิบ	IRT 3PL			TRT 3PL		
		θ	SE	คะแนนจริง	θ	SE	คะแนนจริง
1483	44	2.786	0.296	21.774	2.794	0.323	21.686
1484	28	0.337	0.440	12.936	0.207	0.438	12.630
1485	32	0.225	0.457	14.273	0.229	0.524	14.420
1486	26	0.908	0.244	13.276	0.957	0.264	12.892
1487	32	-0.645	0.678	13.588	-0.491	0.608	13.401
1488	34	2.119	0.228	14.145	2.203	0.263	13.356
1489	22	0.018	0.506	12.566	-0.064	0.660	12.396
1490	36	-0.698	0.737	14.064	-0.634	0.746	13.945
1491	26	-0.698	0.608	12.495	-0.774	0.698	12.315
1492	32	-0.641	0.684	13.499	-0.582	0.703	13.501
1493	50	1.033	0.311	20.255	1.136	0.252	20.289
1494	26	1.324	0.211	12.538	1.385	0.248	12.573
1495	20	0.930	0.219	12.468	1.023	0.248	12.216
1496	24	-0.365	0.703	12.544	-0.315	0.794	12.436
1497	50	-0.502	0.580	25.831	-0.382	0.658	25.822
1498	32	-0.684	0.847	13.012	-0.704	0.694	12.881
1499	28	-0.598	0.698	12.757	-0.721	0.747	12.747
1500	56	1.353	0.220	26.251	1.358	0.241	25.428

ตารางภาคผนวกที่ ค-4 พารามิเตอร์ความสามารถแต่ละเทสต์ไเลท วิชาภาษาอังกฤษ

คนที่	เทสต์ไเลท							
	1	2	3	4	5	6	7	8
1	0.011	0.025	0.026	0.060	0.133	-0.010	-0.004	-0.012
2	0.013	-0.099	-0.015	0.106	-0.150	0.033	0.013	0.058
3	0.011	0.027	-0.012	0.082	-0.139	-0.102	-0.011	0.083
4	-0.139	0.079	0.018	0.092	0.266	0.032	0.002	-0.039
5	-0.130	0.120	0.000	0.129	0.268	0.012	-0.052	0.189
6	0.023	0.077	0.014	0.032	-0.090	-0.039	0.017	-0.033
7	0.025	-0.074	0.054	0.034	-0.094	-0.035	0.062	0.014
8	-0.084	-0.162	-0.060	-0.064	0.276	0.069	0.061	-0.098
9	0.150	-0.126	0.087	0.152	-0.607	0.047	0.060	0.007
10	-0.096	-0.082	0.114	0.059	-0.104	0.114	0.087	-0.135
11	-0.058	-0.008	0.080	0.018	0.034	0.054	0.062	-0.018
12	-0.076	-0.171	0.104	0.107	0.308	-0.124	0.164	-0.294
13	-0.004	0.134	0.016	-0.077	0.106	-0.050	0.046	0.089
14	-0.002	0.087	-0.016	0.088	0.090	-0.087	0.036	-0.038
15	0.102	-0.009	0.142	-0.103	0.035	0.004	0.057	0.066
16	-0.124	0.062	0.030	0.082	-0.100	-0.040	-0.066	0.053
17	0.050	0.088	-0.043	-0.100	0.075	-0.037	-0.042	0.069
18	-0.063	0.040	-0.019	0.205	0.393	-0.015	-0.007	0.083
19	0.130	0.033	-0.003	0.177	0.018	-0.016	0.097	-0.090
20	0.070	0.099	-0.018	0.016	0.110	-0.030	-0.006	-0.027
21	0.042	0.005	0.009	-0.036	0.343	0.027	-0.011	-0.086
22	0.000	0.037	0.023	-0.069	-0.063	0.073	0.029	0.058
23	0.093	-0.093	0.081	-0.039	-0.065	0.000	0.096	0.016

ตารางภาคผนวกที่ ค-4 (ต่อ)

คนที่	เทสต์เลข							
	1	2	3	4	5	6	7	8
24	0.148	-0.096	-0.044	0.198	0.116	-0.110	-0.091	0.032
25	0.039	-0.013	0.136	-0.315	-0.938	-0.005	-0.253	-0.339
26	0.035	-0.003	0.001	0.113	0.040	0.023	0.015	0.073
27	-0.025	0.047	0.019	0.027	-0.186	-0.011	-0.002	0.075
28	0.035	0.116	-0.006	0.004	-0.200	0.037	-0.004	0.011
29	0.053	0.023	0.017	0.052	0.006	-0.076	-0.012	0.024
30	0.039	0.017	-0.048	0.073	-0.036	0.010	0.003	0.076
.
.
.
1470	0.033	0.082	0.000	-0.023	-0.165	-0.011	0.001	-0.125
1471	-0.068	0.030	0.040	-0.019	-0.218	0.027	0.066	0.021
1472	-0.024	-0.011	0.012	0.053	0.190	0.055	0.025	0.036
1473	-0.123	0.030	-0.001	-0.055	-0.085	-0.031	0.038	0.008
1474	0.002	0.081	0.019	-0.106	-0.057	0.034	0.057	-0.012
1475	0.163	0.011	0.013	0.081	0.329	0.144	0.008	-0.144
1476	0.111	0.005	0.141	-0.189	0.236	0.067	0.071	-0.007
1477	-0.193	0.139	0.140	-0.086	-0.075	0.137	-0.086	0.100
1478	0.163	0.029	0.030	-0.019	-0.291	0.044	-0.084	-0.141
1479	-0.140	-0.035	0.142	-0.312	0.463	0.025	-0.245	-0.101
1480	-0.097	-0.068	0.073	0.036	0.007	0.087	0.026	-0.025
1481	-0.173	-0.047	-0.190	-0.517	0.497	0.141	-0.254	-0.342
1482	0.056	-0.101	0.010	0.100	0.322	0.065	-0.119	-0.243

ตารางภาคผนวกที่ ค-4 (ต่อ)

คนที่	เทสต์เลข							
	1	2	3	4	5	6	7	8
1483	-0.016	0.087	-0.196	0.005	0.314	-0.039	0.184	-0.089
1484	0.068	0.032	0.029	0.021	-0.151	-0.036	-0.011	-0.005
1485	-0.060	-0.065	0.078	-0.175	-0.180	0.103	0.050	-0.127
1486	0.050	0.051	-0.087	0.011	0.113	-0.057	0.035	0.031
1487	0.116	-0.091	-0.015	0.074	0.184	-0.004	-0.002	0.073
1488	-0.063	-0.028	0.027	-0.134	0.124	0.028	0.023	-0.043
1489	-0.081	-0.069	0.026	-0.008	0.125	-0.020	0.007	0.080
1490	0.044	-0.013	-0.018	-0.083	0.326	-0.023	0.004	-0.011
1491	0.064	0.000	0.034	0.098	0.222	0.014	0.035	-0.021
1492	-0.027	-0.044	0.001	0.020	0.192	-0.043	-0.075	0.057
1493	-0.068	-0.024	0.028	-0.240	0.307	-0.068	-0.188	-0.116
1494	-0.059	0.015	0.031	0.030	0.089	0.093	-0.052	0.036
1495	0.028	0.069	0.047	-0.094	0.139	0.044	0.058	-0.022
1496	-0.014	0.048	0.048	0.023	0.169	0.064	-0.002	0.022
1497	0.051	-0.059	-0.170	-0.065	0.433	0.136	-0.004	0.284
1498	-0.033	-0.079	0.002	-0.173	0.174	-0.022	0.022	-0.027
1499	0.018	-0.052	-0.006	-0.048	0.260	0.041	0.021	-0.158
1500	0.256	0.059	-0.132	-0.127	0.665	0.030	0.000	0.056

ภาคผนวก ง

ตัวอย่าง ผลการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ

ตารางภาคผนวกที่ ง-1 ผลการเชื่อมโยงคะแนนด้วยวิธีโค้งคุณลักษณะแบบสอบ

คนที่	IRT 3PL			TRT 3PL		
	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง
1	20.412	12.106	20.289	21.647	11.979	14.779
2	22.044	13.088	20.630	20.786	12.818	14.501
3	20.111	12.924	20.192	24.220	12.996	15.883
4	21.033	13.744	20.437	19.905	14.020	14.352
5	23.373	14.876	20.863	22.755	15.350	15.203
6	21.431	12.342	20.517	22.182	12.064	14.911
7	21.652	14.903	20.559	22.376	14.210	15.029
8	22.585	15.151	20.725	20.801	14.935	14.534
9	19.736	17.089	19.984	22.079	17.318	14.894
10	24.984	17.627	21.224	20.871	16.157	14.524
11	21.529	12.786	20.536	21.518	12.552	14.757
12	23.751	20.805	20.934	20.842	20.193	14.537
13	19.973	13.244	20.132	22.139	13.337	14.889
14	19.956	14.096	20.124	22.425	13.835	15.018
15	25.317	17.434	21.336	24.335	16.405	16.156
16	19.729	13.538	19.978	21.945	13.328	14.842
17	20.187	12.667	20.219	22.281	12.349	15.028
18	19.789	21.529	20.025	20.996	20.458	14.589
19	22.988	15.397	20.795	19.425	15.370	14.191
20	21.285	12.574	20.488	20.377	12.623	14.423
21	20.305	12.606	20.258	21.930	12.724	14.809
22	19.848	12.941	20.065	20.338	12.481	14.464

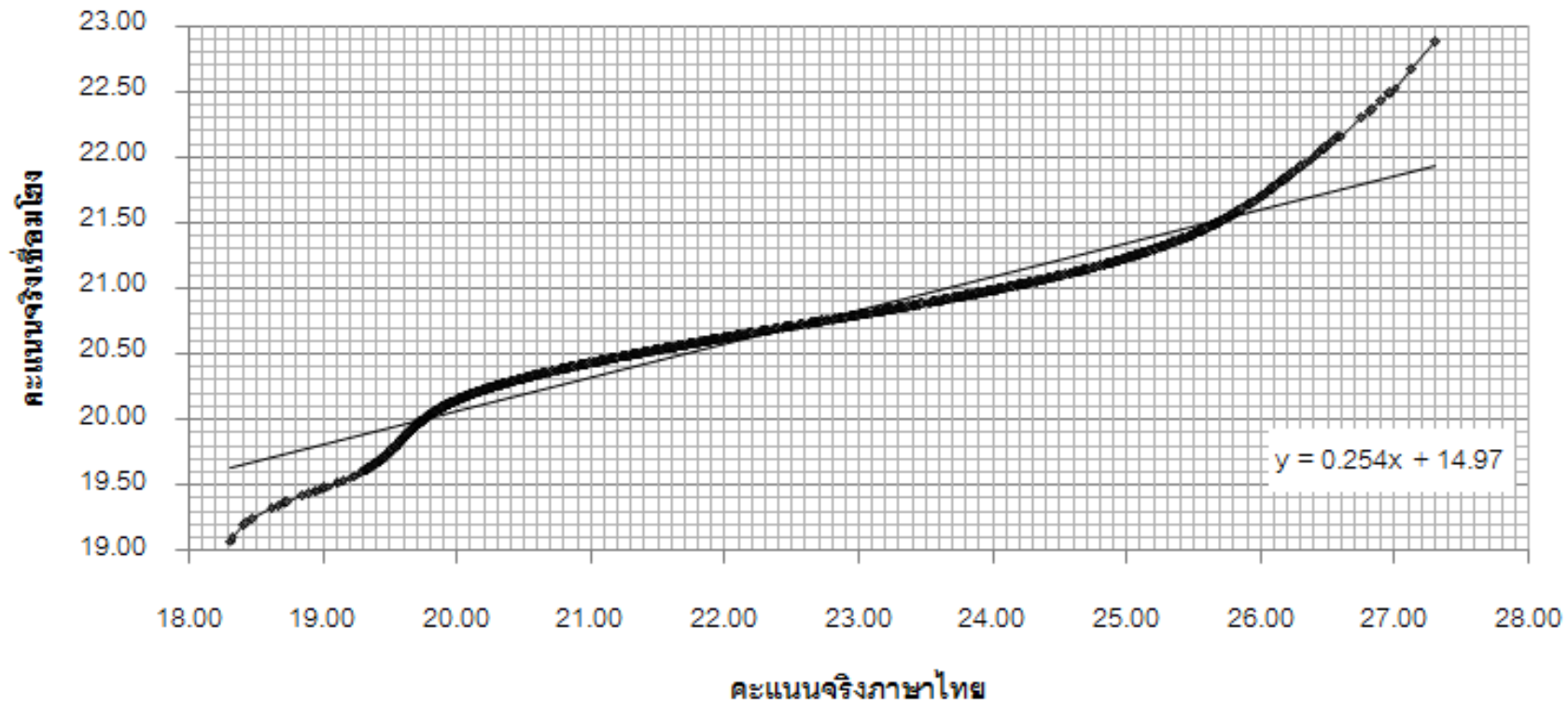
ตารางภาคผนวกที่ ง-1 (ต่อ)

คนที่	IRT 3PL			TRT 3PL		
	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง
23	25.295	14.184	21.328	24.125	13.622	15.636
24	21.183	20.966	20.468	21.477	19.976	14.675
25	23.585	23.216	20.902	21.618	24.198	14.717
26	24.196	12.324	21.025	24.004	12.229	15.438
27	19.242	13.030	19.572	22.076	12.606	14.881
28	26.151	12.678	21.810	21.653	12.542	14.772
29	20.106	12.768	20.190	21.197	12.543	14.620
30	24.380	12.629	21.066	21.516	12.263	14.692
.
.
.
1470	24.071	12.811	20.998	21.984	12.588	14.869
1471	19.444	12.193	19.697	20.617	12.092	14.450
1472	19.738	12.110	19.986	20.936	12.039	14.557
1473	24.934	13.254	21.209	20.015	12.808	14.369
1474	23.797	12.288	20.943	20.425	12.229	14.436
1475	20.814	17.568	20.389	22.251	17.136	14.948
1476	22.196	15.877	20.657	23.121	15.193	15.219
1477	21.160	15.475	20.463	20.666	15.318	14.525
1478	23.714	17.323	20.927	20.938	17.146	14.551
1479	21.051	24.776	20.441	20.976	24.588	14.568
1480	21.451	13.655	20.521	21.100	13.329	14.580

ตารางภาคผนวกที่ ง-1 (ต่อ)

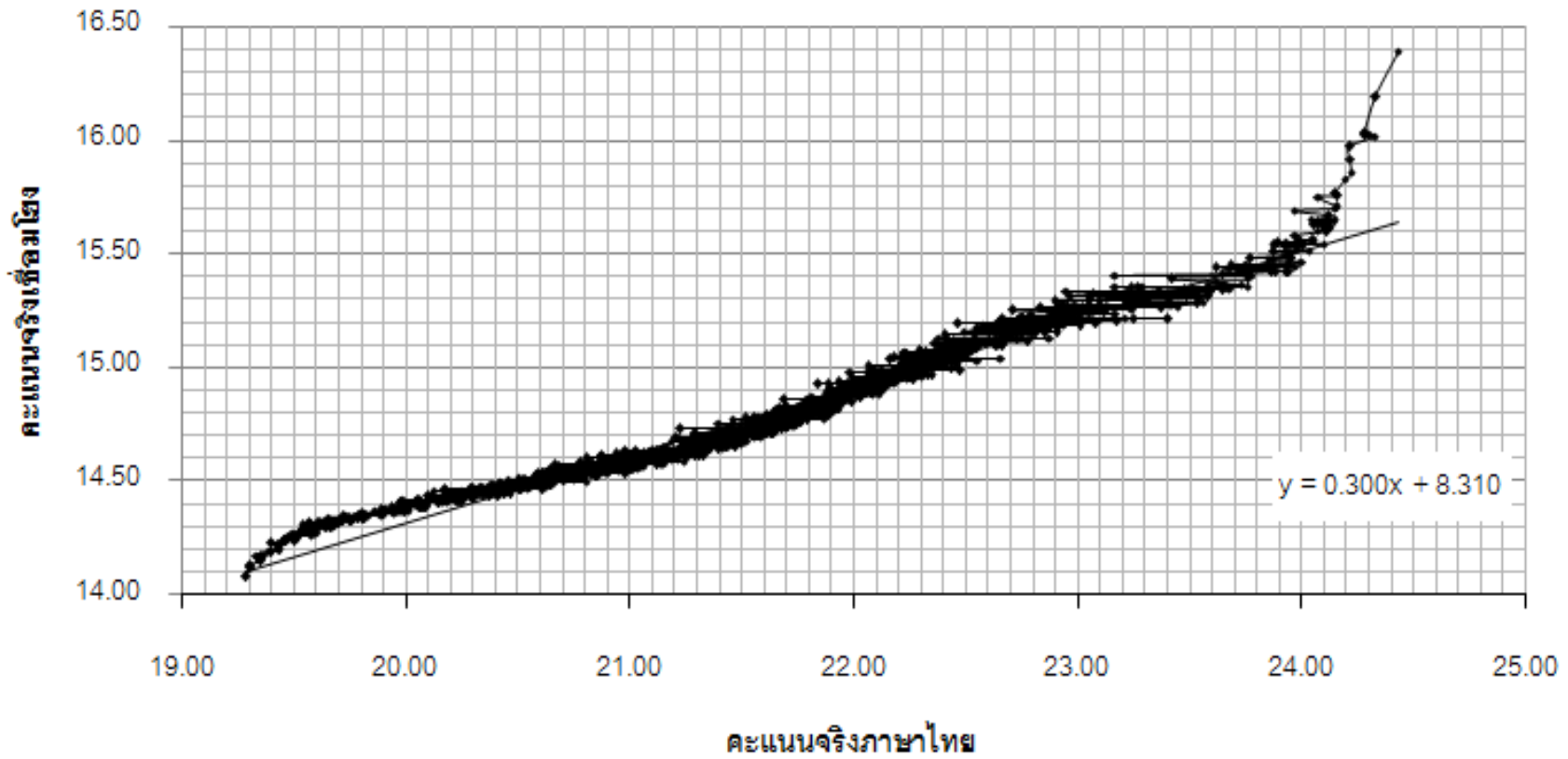
คนที่	IRT 3PL			TRT 3PL		
	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง	คะแนน ไทย	คะแนน อังกฤษ	คะแนน เชื่อมโยง
1481	20.056	26.432	20.169	21.901	25.946	14.886
1482	19.771	14.291	20.012	20.203	14.348	14.417
1483	20.515	21.774	20.317	21.904	21.686	14.831
1484	19.579	12.936	19.828	21.836	12.630	14.798
1485	22.150	14.273	20.649	21.426	14.420	14.668
1486	24.706	13.276	21.146	20.803	12.892	14.538
1487	20.337	13.588	20.268	20.541	13.401	14.483
1488	26.278	14.145	21.909	21.881	13.356	14.775
1489	19.840	12.566	20.060	22.300	12.396	14.951
1490	23.091	14.064	20.813	20.069	13.945	14.388
1491	21.721	12.495	20.572	21.991	12.315	14.871
1492	20.814	13.499	20.389	20.964	13.501	14.563
1493	20.005	20.255	20.147	22.125	20.289	14.934
1494	24.880	12.538	21.193	21.936	12.573	14.907
1495	24.690	12.468	21.142	21.088	12.216	14.596
1496	19.782	12.544	20.020	23.587	12.436	15.326
1497	22.907	25.831	20.781	20.543	25.822	14.483
1498	20.232	13.012	20.234	22.204	12.881	14.961
1499	18.965	12.757	19.459	23.211	12.747	15.298
1500	19.020	26.251	19.480	22.293	25.428	15.050

กราฟแสดงความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ



ภาพภาคผนวกที่ ง-1 ความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ

กราฟแสดงความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของทดสอบ



ภาพภาคผนวกที่ ง-2 ความสัมพันธ์ระหว่างคะแนนจริงวิชาภาษาไทยกับคะแนนเชื่อมโยงที่ประมาณค่าด้วยโมเดลทฤษฎีการตอบสนองของทดสอบ

ภาคผนวก จ

ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซนส์ไทล์

ตารางภาคผนวกที่ จ-1 ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (IRT 3PL)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
0	0.00	0	0	0	0	0	0	0	0	0	0
1	1.00	0	0	0	0	0	0	0	0	0	0
2	2.00	0	0	0	0	0	0	0	0	0	0
3	3.00	0	0	0	0	0	0	0	0	0	0
4	4.00	0	0	0	0	0	0	0	0	0	0
5	5.00	0	0	0	0	0	0	0	0	0	0
6	6.00	0	0	0	0	0	0	0	0	0	0
7	7.00	0	0	0	0	0	0	0	0	0	0
8	8.00	0	0	0	0	0	0	0	0	0	0
9	9.00	0	0	0	0	0	0	0	0	0	0
10	10.00	0	0	0	0	0	0	0	0	0	0
11	11.000	0	0	0	0	0	0	0	0	0	0

ตารางภาคผนวกที่ จ-1 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
12	12.000	0	0	0	0	0	0	0	0	0	0
13	13.000	1	1	0.001	0.001	0.030	0	0	0	0	0
14	14.000	1	2	0.001	0.001	0.100	0	0	0	0	0
15	15.000	1	3	0.001	0.002	0.170	0	0	0	0	0
16	16.000	1	4	0.001	0.003	0.230	0	0	0	0	0
17	16.500	3	7	0.002	0.005	0.370	0	0	0	0	0
18	16.500	6	13	0.004	0.009	0.670	1	1	0.001	0.001	0.030
19	16.500	12	25	0.008	0.017	1.270	2	3	0.001	0.002	0.130
20	16.500	22	47	0.015	0.031	2.400	5	8	0.003	0.005	0.370
21	16.500	35	82	0.023	0.055	4.300	10	18	0.007	0.012	0.870
22	16.500	53	135	0.035	0.090	7.230	19	37	0.013	0.025	1.830
23	16.500	73	208	0.049	0.139	11.430	33	70	0.022	0.047	3.570
24	17.000	96	304	0.064	0.203	17.070	53	123	0.035	0.082	6.430

ตารางภาคผนวกที่ จ-1 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny*g(y)$	$Ny*G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx*f(x)$	$Nx*F(x)$	$f(x)$	$F(x)$	$P(x)$
25	18.500	118	422	0.079	0.281	24.200	78	201	0.052	0.134	10.800
26	19.500	136	558	0.091	0.372	32.670	107	308	0.071	0.205	16.970
27	20.357	148	706	0.099	0.471	42.130	135	443	0.090	0.295	25.030
28	21.233	153	859	0.102	0.573	52.170	157	600	0.105	0.400	34.770
29	22.179	148	1007	0.099	0.671	62.200	170	770	0.113	0.513	45.670
30	23.133	134	1141	0.089	0.761	71.600	169	939	0.113	0.626	56.970
31	24.099	114	1255	0.076	0.837	79.870	155	1094	0.103	0.729	67.770
32	25.079	89	1344	0.059	0.896	86.630	131	1225	0.087	0.817	77.300
33	26.068	64	1408	0.043	0.939	91.730	101	1326	0.067	0.884	85.030
34	27.069	42	1450	0.028	0.967	95.270	72	1398	0.048	0.932	90.800
35	28.083	25	1475	0.017	0.983	97.500	47	1445	0.031	0.963	94.770
36	29.103	14	1489	0.009	0.993	98.800	28	1473	0.019	0.982	97.270
37	30.134	7	1496	0.005	0.997	99.500	15	1488	0.010	0.992	98.700

ตารางภาคผนวกที่ จ-1 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
38	31.173	3	1499	0.002	0.999	99.830	7	1495	0.005	0.997	99.430
39	32.220	1	1500	0.001	1.000	99.970	3	1498	0.002	0.999	99.770
40	33.279	0	1500	0.000	1.000	100.000	1	1499	0.001	0.999	99.900
41	34.345	0	1500	0.000	1.000	100.000	1	1500	0.001	1.000	99.970
42	35.435	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
43	36.500	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
44	37.750	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
45	39.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
46	46.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
47	47.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
48	48.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
49	49.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
50	50.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000

ตารางภาคผนวกที่ จ-2 ผลการเชื่อมโยงคะแนนด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ที่ประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองของทดสอบเลขแบบ 3 พารามิเตอร์ (TRT 3PL)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny*g(y)$	$Ny*G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx*f(x)$	$Nx*F(x)$	$f(x)$	$F(x)$	$P(x)$
0	0.000	0	0	0	0	0	0	0	0	0	0
1	1.000	0	0	0	0	0	0	0	0	0	0
2	2.000	0	0	0	0	0	0	0	0	0	0
3	3.000	0	0	0	0	0	0	0	0	0	0
4	4.000	0	0	0	0	0	0	0	0	0	0
5	5.000	0	0	0	0	0	0	0	0	0	0
6	6.000	0	0	0	0	0	0	0	0	0	0
7	7.000	0	0	0	0	0	0	0	0	0	0
8	8.000	0	0	0	0	0	0	0	0	0	0
9	9.000	0	0	0	0	0	0	0	0	0	0
10	10.000	0	0	0	0	0	0	0	0	0	0
11	11.000	0	0	0	0	0	0	0	0	0	0

ตารางภาคผนวกที่ จ-2 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
12	12.000	0	0	0	0	0	0	0	0	0	0
13	13.000	0	0	0	0	0	0	0	0	0	0
14	14.000	0	0	0	0	0	0	0	0	0	0
15	15.000	0	0	0	0	0	0	0	0	0	0
16	16.000	0	0	0	0	0	0	0	0	0	0
17	16.500	1	1	0.001	0.001	0.030	0	0	0	0	0
18	16.500	1	2	0.001	0.001	0.100	0	0	0	0	0
19	16.500	3	5	0.002	0.003	0.230	0	0	0	0	0
20	16.500	7	12	0.005	0.008	0.570	0	0	0	0	0
21	16.500	15	27	0.010	0.018	1.300	0	0	0	0	0
22	16.500	28	55	0.019	0.037	2.730	0	0	0	0	0
23	16.500	49	104	0.033	0.069	5.300	0	0	0	0	0
24	17.000	76	180	0.051	0.120	9.470	1	1	0.001	0.001	0.030

ตารางภาคผนวกที่ จ-2 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
25	18.500	108	288	0.072	0.192	15.600	2	3	0.001	0.002	0.130
26	19.500	140	428	0.093	0.285	23.870	4	7	0.003	0.005	0.330
27	20.357	166	594	0.111	0.396	34.070	8	15	0.005	0.010	0.730
28	21.233	180	774	0.120	0.516	45.600	16	31	0.011	0.021	1.530
29	22.179	179	953	0.119	0.635	57.570	30	61	0.020	0.041	3.070
30	23.133	161	1114	0.107	0.743	68.900	50	111	0.033	0.074	5.730
31	24.099	133	1247	0.089	0.831	78.700	77	188	0.051	0.125	9.970
32	25.079	100	1347	0.067	0.898	86.470	109	297	0.073	0.198	16.170
33	26.068	68	1415	0.045	0.943	92.070	141	438	0.094	0.292	24.500
34	27.069	42	1457	0.028	0.971	95.730	169	607	0.113	0.405	34.830
35	28.083	23	1480	0.015	0.987	97.900	184	791	0.123	0.527	46.600
36	29.103	12	1492	0.008	0.995	99.070	182	973	0.121	0.649	58.800
37	30.134	5	1497	0.003	0.998	99.630	164	1137	0.109	0.758	70.330

ตารางภาคผนวกที่ จ-2 (ต่อ)

คะแนน	คะแนนเชื่อมโยง	วิชาภาษาไทย					วิชาภาษาอังกฤษ				
		$Ny \cdot g(y)$	$Ny \cdot G(y)$	$g(y)$	$G(y)$	$Q(y)$	$Nx \cdot f(x)$	$Nx \cdot F(x)$	$f(x)$	$F(x)$	$P(x)$
38	31.173	2	1499	0.001	0.999	99.870	133	1270	0.089	0.847	80.230
39	32.220	1	1500	0.001	1.000	99.970	98	1368	0.065	0.912	87.930
40	33.279	0	1500	0.000	1.000	100.000	64	1432	0.043	0.955	93.330
41	34.345	0	1500	0.000	1.000	100.000	37	1469	0.025	0.979	96.700
42	35.435	0	1500	0.000	1.000	100.000	19	1488	0.013	0.992	98.570
43	36.500	0	1500	0.000	1.000	100.000	8	1496	0.005	0.997	99.470
44	37.750	0	1500	0.000	1.000	100.000	3	1499	0.002	0.999	99.830
45	39.000	0	1500	0.000	1.000	100.000	1	1500	0.001	1.000	99.970
46	46.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
47	47.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
48	48.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
49	49.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000
50	50.000	0	1500	0.000	1.000	100.000	0	1500	0.000	1.000	100.000

ประวัติย่อผู้วิจัย

ชื่อ-สกุล	นางสาวจงกล บัวแก้ว
วัน เดือน ปีเกิด	25 ตุลาคม พ.ศ. 2526
สถานที่เกิด	จังหวัดสงขลา
สถานที่อยู่ปัจจุบัน	7/2 หมู่ 9 ตำบลปริก อำเภอสะเดา จังหวัดสงขลา 90120
ประวัติการทำงาน	
พ.ศ. 2551-2552	ครูผู้สอน (คณิตศาสตร์) โรงเรียนทานตะวันไตรภาษา
พ.ศ. 2554-2557	อาจารย์ประจำคณะครุศาสตร์ มหาวิทยาลัยราชภัฏสงขลา
ประวัติการศึกษา	
พ.ศ. 2551	ศึกษาศาสตรบัณฑิต (เทคโนโลยีสารสนเทศและประเมินผลการศึกษา) มหาวิทยาลัยสงขลานครินทร์
พ.ศ. 2554	ครุศาสตรมหาบัณฑิต (การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2560	ปรัชญาดุษฎีบัณฑิต (วิจัย วัดผลและสถิติการศึกษา) มหาวิทยาลัยบูรพา